

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）回答との主な項目の比較【ヤフー株式会社】

参考資料4-1

※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。

ヤフー株式会社		ヒアリングシート(2021年2月25日)回答		ヒアリングシート(2022年3月7日)回答	
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		Yahoo!知恵袋、Yahoo!ニュース(コメント欄のみ)、Yahoo!ファイナンス
1-②	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】	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		2021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1,100万件 ■ファイナンス掲示板 総投稿数:約230万件
1-④	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				随時、必要に応じて、ポリシーの見直しを行っており、ポリシーを改定した場合は、サービスの内のお知らせページやオウンドメディア等で告知している。また、場合によっては、有識者会議への諮問や、外部有識者に意見を伺う場合もある。
2-①	(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		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じて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。		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じてカスタマーサポート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。 また、プロバイダ責任制限法に基づく請求を受け付ける専用の書面受付窓口も用意している。
	(ii)主体的な誹謗中傷等の情報の検知態勢		24時間365日、専門チームによるパトロールのほか、最先端の技術を駆使して不適切な投稿への対策を行っている。また、ユーザーより「違反申告」ボタンから違反投稿の申告をいただいたものについて、専門チームが1件ずつ確認し、対応している。 ■ヤフーニュース 「Yahoo!ニュース コメント」においては、サービス開始当初より、マルチポスト対策、複数アカウントからの投稿の制限、不適切コメントの自動検知など、さまざまな機械的な対応を行っている。また、2014年から機械学習による不適切投稿への対応(AI)を開始。 また、深層学習を用いた自然言語処理モデル(AI)による検知を通して、1日平均約2万件の不適切な投稿(記事との関連性の低いコメントや誹謗中傷等の書き込みなど)の削除を行っている。 (参考) ・Yahoo!ニュースの「不適切コメント対策」最前線(2020.02.07掲載) https://news.yahoo.co.jp/newshack/technology/comment_kukai.html ・Yahoo!ニュース コメントの健全化に向けた取り組み(2020.06.24リリース) https://about.yahoo.co.jp/info/blog/20200624/newscomment.html ・Yahoo!ニュースにおける建設的コメント順位付けモデルの導入(2018.11.15掲載) https://news.yahoo.co.jp/newshack/inside/yjnews_comment2018.html また、2020年10月には、違反投稿を繰り返す投稿者への対策として、一度「投稿停止措置」を受けたユーザーが、当初のID作成時に利用した携帯電話番号と同じ番号を利用してYahoo! JAPAN IDを再取得した場合であっても、コメントの投稿を制限する取り組みを開始した。 (参考) Yahoo!ニュース、違反コメント投稿者に対する投稿停止措置を強化 https://about.yahoo.co.jp/pr/release/2020/10/14a/		カスタマーサポートチームが24時間365日、目視によるパトロールを行い、違反コメントに対して、削除対応などを行っている。知恵袋では、機械学習による低品質な投稿の判定・自動削除を実施、ヤフーニュースでは、機械学習による不適切なコメントの解析・検知・自動削除を実施している。
2-③	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				原則として、申し立て順に対応している。もっとも、申し立て内容の権利侵害の重大性に鑑み、緊急的に対応することもある。
2-④	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	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	全サービス共通の数値ではあるが、パトロ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たる。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たる。		投稿の削除等のパトロール業務を専門とするパトロール部隊には、約70名が在籍し、24時間365日、シフトを組んで対応にあたる。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たる。
		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	日本国内の拠点:有 日本における責任者:有		日本国内の拠点:有 日本における責任者:有
2-⑤	(i)対応決定時の申立て者への通知の有無とその理由説明内容		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面の発送を行う。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。		問い合わせ方法により異なり、違反申告フォームからの申告等については申し立て者に対する返答・通知は行っていないが、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、申し立て者に対し回答書面の発送を行う。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。
	(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		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。		知恵袋については、削除された質問・回答に対して、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。 (参考) https://support.yahoo-net.jp/PccChiebukuro/s/article/H000008062
2-⑥	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		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じて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、後述の社内有識者会議において、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。		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じて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。 特に、ヤフーニュースについては、削除に疑義が生じた場合に、運営者に削除理由などをお問い合わせいただくための専用手続きを用意している。このプロセスにおいては、理由の通知のみならず、措置理由が合理的であったか否かの判断も含めて確認がなされる。 他のサービスについては、後述の社内有識者会議において、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異議申し立てについて、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。
2-⑧	不正な申告や削除要請への対策の方法・仕組み、対応件数	【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、パトロ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。		特になし。 申告や要請が正当か正当でないかにかかわらず、カスタマーサポート窓口へ寄せられた申告や違反申告フォームへ寄せられた申告については、削除基準書にのっとり、判断している。
		【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
3-①	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】	各サービスにおいて、特定の一か月間の数値を算出。 ・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		以下、各サービスにおいて、特定の一か月間の数値を算出。 ・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。 (1)一般ユーザ ■知恵袋 違反報告のうち、「誹謗中傷」等を含む違反申告数:約5.7万件 ※他人を攻撃・不快にさせる投稿のほか、なりすまし行為や自作自演等を理由とする申告も含む。 ■ヤフーニュース 「誹謗中傷」「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21.8万件 ■ファイナンス掲示板 不快(アダルト・グロ・罵詈雑言)の違反申告数:約2.7万件 (2)司法機関 裁判における削除決定/判決件数:6件 ※集計対象:2021年4月1日～12月31日までに名誉毀損を理由とした請求があったもののうち、2022年2月21日時点で結論が出ているもの
	(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	【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		(1)一般ユーザ ■知恵袋 上記の申告数のうち、「誹謗中傷」という削除理由で削除した件数:約0.5万件 ■ヤフーニュース 総違反申告数約21.8万件のうち、「不快」を理由に削除された件数:約1.2万件 ※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」ものについて集計している。 ■ファイナンス掲示板 不快(アダルト・グロ・罵詈雑言)の削除件数:約1.3万件 (2)司法機関 裁判において削除決定/判決が下され、最終した場合には速やかに削除対応を行っている。
	(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	【日本の数値】			■ヤフーニュース 2021年10月19日～2021年12月18日の2か月間でコメント欄が非表示となった記事数:計216件、1日あたり平均3.5件 ※Yahoo!ニュース、コメント欄非表示機能導入後の状況を公表(2022年1月20日) https://about.yahoo.co.jp/pr/release/2022/01/20a/ ※申告や削除要請に対する非表示対応は行っていないが、AI判定による非表示対応を行っているサービスはある。(③(ii)に記入)
	(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	【日本の数値】			(1)一般ユーザ ■知恵袋 約5.2万件 ■ヤフーニュース 約20.6万件 ■ファイナンス掲示板 約1.4万件 (2)司法機関 裁判において削除決定/判決が下された場合は速やかに削除対応を行っている。

ヤフー株式会社		ヒアリングシート(2021年2月25日)回答		ヒアリングシート(2022年3月7日)回答	
	(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】 【(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 申告窓口やパトロールの手法(人かAIか)、事案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、人によるパ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両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、一担当者からの判断によらず、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。	回答なし 回答なし 申告窓口やパトロールの手法(人かAIか)、事案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、人によるパ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両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、一担当者からの判断によらず、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。	
3-③	(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】	・各サービスにおいて、特定の一か月間のパトロール数値を算出。 ・また、サービスごとに利用しているパ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。 ・効率的なパトロールとなるよう、重点的に対応すべきパトロール対象、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	・各サービスにおいて、特定の一か月間の削除件数を算出。 ・また、サービスごとに利用しているパ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。 ・効率的なパトロールとなるよう、重点的に対応すべきパトロール対象、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。 ■知恵袋 約2.6万件 ※違反申告以外で、積極的なパトロールやAIにより「誹謗中傷」として削除した件数 ■ヤフーニュース 約3万件 ※違反申告以外で、「不快」で削除した件数 ※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 ※加えて、AI「不適切投稿判定モデル」による投稿削除数:約14.9万件 ■ファイナンス 約1.7万件 ※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	
	(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】		■知恵袋 約12万件 ※知恵袋では、申告による非表示は行っていない。機械判定により「低品質な投稿」と判定され、非表示になった回答の件数。 ※低品質投稿とは、アダルト関連や文意不明(質問/回答になっていない)投稿、誹謗中傷を含む一般人が不快と感じる投稿を指す。	
	(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】		回答なし	
3-④	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)	【日本の取組】	■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。 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出も実施。 ・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。 ■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。 ■ファイナンス掲示板 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。	24時間365日、専門チームによるパトロールのほか、最先端の技術を駆使して不適切な投稿への対策を行っている。また、ユーザーより「違反申告」ボタンから違反投稿の申告をいただいたものについて、専門チームが1件ずつ確認し、対応している。 ■ヤフーニュース 「Yahoo!ニュース」では、以前より誹謗中傷などの内容を含む投稿を禁止し、コメントポリシーに違反投稿の具体例をわかりやすく明示したうえで、コメント投稿フォームにも記載している。また、24時間体制の専門チームによる人的パトロールに加えて、自社で開発した「深層学習を用いた自然言語処理モデル(AI)」や、独自に開発した深層学習特化型スーパーコンピュータ「kukai」を活用するなど、テクノロジーを駆使しながら、誹謗中傷をはじめとする違反コメントを1日平均約2万件削除するなどの対策を行っている。加えて、誹謗中傷などの違反コメントを複数にわたって投稿したユーザーに対しては、「投稿停止措置」を行っている。一方で、情報の流通基盤であるプラットフォーム事業を営む企業としては、投稿者の表現の自由とのバランスを慎重に見極めながら適正に対応することも重要であると考えている。Yahoo! JAPANでは、人権救済の観点より誹謗中傷投稿に対して厳正に対処していくとともに、表現の自由にも配慮したインターネット上の言論空間全体の健全化に向けて、誹謗中傷対策の強化や透明化を進めるために取り組んでいる。例えば、一定以上コメント数がある記事のコメント欄を対象に、AIが判定した違反コメント数などの基準に従い、コメント欄を自動的に非表示にする機能を実装している。 ※Yahoo!ニュース、コメント欄の健全化に向けた取り組みを強化(2021年10月19日) https://about.yahoo.co.jp/pr/release/2021/10/19a/ ※Yahoo!ニュース、コメント欄非表示機能導入後の状況を公表(2022年1月20日) https://about.yahoo.co.jp/pr/release/2022/01/20a/	
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	(i) 裁判上の開示請求件数: 23件 ※集計対象: 2021年4月1日～12月31日までに名誉毀損を理由とした請求があったもののうち、2022年2月21日時点で結論が出ているもの(件数は事件単位) (ii) 裁判外の開示請求件数: 28件 ※集計対象: 2021年4月1日～12月31日までに名誉・信用・プライバシー侵害を理由とした請求があったもののうち、2022年2月21日時点で結論が出ているもの(件数は受領した書面単位)	
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)	(i) 裁判上の開示件数: 18件(うち一部開示が6件) (ii) 裁判外の開示件数: 2件 ※集計対象等は11に同じ	
5-①	透明性レポートの公開 日本語で閲覧可能か		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。	2021年12月に、2020年度の透明性レポートを公開した。日本語で閲覧可能。 https://about.yahoo.co.jp/common/transparencyreport/	
5-②	誹謗中傷等の取組の効果分析		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。 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。 ■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) 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 ヤフー株式会社が提供する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。	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。 ■ニュースコメント 投稿時注意メッセージの取り組みについて、効果を検証しその結果を公開した。 Yahoo!ニュースの「コメント機能」 AIを活用した投稿時注意メッセージ掲出の効果について https://news.yahoo.co.jp/newshack/information/comment_alert_20210216.html	

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）回答における主な項目の比較【Google合同会社】

※総務省において要約したものを。詳細は各社の回答原本を参照。

参考資料4-2

Google合同会社		ヒアリングシート(2021年2月25日)回答	ヒアリングシート(2022年3月7日)回答
1-①	対象サービス名	YouTube, Google 検索, Google マップ	Google 検索, YouTube, Google マップ
1-②	加入者数、月間アクティブユーザ数又は書き込み数	回答なし ● YouTube: 毎月20億人以上のログインユーザーがYouTubeを訪れ、毎日10億時間以上の動画を視聴しています。 ● Google検索: 毎年、Google検索では何兆回もの検索が行われています。 ● Googleマップ: 10億人以上の人々に道案内をし、人々が探索し、用事を達成するのに役立っています。	回答なし ● Google 検索: 毎年、Google 検索では何兆回もの検索が行われています。 ● YouTube: 毎月20億人以上のログインユーザーがYouTubeを訪れ、毎日10億時間以上の動画を視聴しています。 ● Google マップ: 10億人以上の人々の道案内をし、人々が探索し、用事を達成するのに役立っています。
1-④	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		これらのポリシーの作成にあたり、私たちは企業としての価値観や文化だけを検討するのではなく、運用面、技術面、ビジネス面から考慮すべき事項についても検討します。私たちは、ポリシーが「現実」に沿ったものとなるように、ネット上のトレンドや慣行、業界の規範、規制の変化を定期的に確認しています。また、表示される広告の種類についてのユーザーのフィードバックや懸念に耳を傾けています。また、新たなコンテンツポリシーの策定や新機能の検討にあたっては、子供の発達、情報の質、またヘイトスピーチや健康に関する問題などの分野の専門家に相談しています。 例えば2019年、嫌がらせに対するポリシーのアップデートをしましたが、そこに至るまでには日本政府や、日本でこの問題に取り組んでいる日本国内のパートナーや専門家ともミーティングを行い日本の状況に関するインサイトをいただきました。これらの努力が実り、2019年初頭以降、1日あたりのヘイトスピーチコメント削除数はグローバルで46倍にも至っています。 また最近では、日本でも見られた「特定可能な個人の死や深刻な負傷を喜んだり、嘲笑したりするコンテンツ」もハラスメントとネットいじめに関するポリシー違反の対象として明記されるなど、嫌がらせの新しいトレンドへの対応を続け、その情報わかりやすく提供するようにしています。 https://support.google.com/youtube/answer/2802268?hl=ja&ref_topic=9282436&zippy=、このポリシーに違反する他の種類のコンテンツ
2-①	(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	Googleでは問題のあるコンテンツを報告するためのツールやプロセスをユーザーに提供しています。以下は、当社のサービスに関するヘルプセンター等の案内を抜粋したものです。 [Google 検索] Google検索では、悪意あるやり方で、または嫌がらせを目的として連絡先情報が掲載されているページの URL や、不当な削除方針が示されているサイトのコンテンツの削除をリクエストするためのウェブフォームをユーザーに提供しています。 [YouTube] YouTubeでは、これらのポリシーが遵守されるよう尽力する何千もの従業員に加え、YouTubeのユーザー自らが不適切と思われるコンテンツを報告することができます。コンテンツの報告は匿名で行われるため、誰が動画を報告したかは他のユーザーに開示されません。また、問題を報告しても、自動的にコンテンツが削除されるわけではありません。報告されたコンテンツは、コミュニティガイドラインに沿って審査されます。不適切なコンテンツを報告する方法については、ヘルプページで詳しく説明しています。 https://support.google.com/youtube/answer/2802027?hl=ja&ref_topic=9387085_topic=9387085 [Google マップ] Google のクチコミに関するポリシーに違反するクチコミは、Google のビジネス プロフィールから削除することができます。こちらで順を追って説明しています。 https://support.google.com/business/answer/4596773?co=GENIE.Platform%3DDesktop&hl=ja [法律に基づく削除] Google製品に関するポリシーは、世界中で適用されますが、各地域の法律を遵守するためのプロセスもあります。Google製品全般に共通のこちらのフォームで、法律に基づく削除リクエストを提出することができます。 https://support.google.com/legal/troubleshooter/1114905?hl=ja	Googleでは問題のあるコンテンツを報告するためのツールやプロセスをユーザーに提供しています。以下は、当社のサービスに関するヘルプセンター等の案内を抜粋したものです。 [Google 検索] Google検索では、悪意あるやり方で、または嫌がらせを目的として連絡先情報が掲載されているページの URL や、不当な削除方針が示されているサイトのコンテンツの削除をリクエストするためのウェブフォームをユーザーに提供しています。 [YouTube] YouTubeでは、これらのポリシーが遵守されるよう尽力する何千もの従業員に加え、YouTubeのユーザー自らが不適切と思われるコンテンツを報告することができます。コンテンツの報告は匿名で行われるため、誰が動画を報告したかは他のユーザーに開示されません。また、問題を報告しても、自動的にコンテンツが削除されるわけではありません。報告されたコンテンツは、コミュニティガイドラインに沿って審査されます。不適切なコンテンツを報告する方法については、ヘルプページで詳しく説明しています。 https://support.google.com/youtube/answer/2802027?hl=ja&ref_topic=9387085_topic=9387085 [Google マップ] Google のクチコミに関するポリシーに違反するクチコミは、Google のビジネス プロフィールから削除することができます。こちらで順を追って説明しています。 https://support.google.com/business/answer/4596773?co=GENIE.Platform%3DDesktop&hl=ja [法律に基づく削除] Google製品に関するポリシーは、世界中で適用されますが、各地域の法律を遵守するためのプロセスもあります。Google製品全般に共通のこちらのフォームで、法律に基づく削除リクエストを提出することができます。 https://support.google.com/legal/troubleshooter/1114905?hl=ja
	(ii)主体的な誹謗中傷等の情報の検知態勢		私たちは、プラットフォーム上の問題のあるコンテンツを検出するために、テクノロジーを使用しています。以前は、ユーザーによる報告が非常に多かったのですが、現在では、Google製品で削除されるコンテンツの大半がまず機械によって識別されます。 [Google 検索] Google 検索では、たとえば、できるだけ多くのクエリに対して質の高い検索結果が得られるようにランキングシステムを設計していますが、一部のクエリは悪意ある第三者を受容してしまいやすいため、特別なソリューションが必要です。そのような例の一つが不当な削除方針が示されているサイトで、コンテンツを削除するために支払いを要求するサイトです。Googleは、2018年から導入したポリシーで、ユーザーが自分に関する情報を含むページをGoogle 検索結果から削除することをリクエストできるようにしています。 Google 検索にこれらのページが表示されないようにするだけでなく、このような削除をGoogle 検索における降格シグナルとして使用し、こうした不当な方針が示されているサイトの順位が検索結果の中で下がるようになります。このソリューションは業界の中でも先駆けており、このようなサイトが行う誹謗中傷の被害を受けている人々への影響を低減するのに効果的に機能しています。 [YouTube] YouTubeには毎分数百時間分にも上る新しいコンテンツがアップロードされるため、人間と機械学習を組み合わせ、問題のあるコンテンツを大規模に検出しています。機械学習は、パターンの検出に適しているため、視聴される前のものであっても、他のすでに削除したコンテンツと類似したコンテンツを見つけることができます。今では四半期ごとに1,000万本近くの動画（ハラスメントやネットいじめ、ヘイトスピーチ関連のみならず全てのポリシー領域）が削除されており、その多くは再生回数が10回に達する前に機械によって削除されています。私たちは、ユーザーが安心して使えるように、こうした技術を用いより良いものにする取り組みを続けていきます。 また、さまざまな有害コンテンツに精通した100以上の外部組織からなる公認報告者プログラムも設けています。このプログラムに協力いただいているのは、コミュニティガイドラインに違反するコンテンツを特定する訓練を受け、医療情報、自殺の防止、ヘイトスピーチや嫌がらせなどの分野に精通した約300の政府系パートナーやNGOです。彼らによる報告は、当社のポリシーに基づいて、ポリシー適用の担当者が優先的に審査し、対処します。公認報告者が持つ価値ある専門知識は、私たちの自動検出システムを補完し、機械では検出の難しいコンテンツの特定に大きな役割を担っています。 https://support.google.com/youtube/answer/7554338?hl=ja 日本では昨年4月に法務省人権擁護局にこのプログラムにご参加いただきました。それ以来、法務省人権擁護局との実りある対話を続けており、共有していただいた情報は、Googleにとって日本独自のトレンドやその背景、今までの議論への理解をさらに深めるために非常に有益なものです。貴重なパートナーシップに心から感謝します。人と機械の両面からアプローチしていくことの重要性が改めて認識できた良い事例と言えます。
2-③	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		たとえばYouTube では、現実社会で生じうる被害が大きいと想定されるケースを優先しています。また、ポリシーを適切に適用することはとても重要で、それを大規模に実行可能とするために人間による審査と最先端のテクノロジーを組み合わせ使用しています。また同時に、背景事情も重要だと考えています。私たちは、各動画の教育的、ドキュメンタリー、科学的、芸術的な意図を考慮します。表面的にはYouTubeのポリシーに違反する可能性があるように見える動画でも、コンテンツが視聴者の目に見える形で背景説明を行っており、説得力のある理由を提示している場合は、YouTubeで許可されることがあります。 また、さまざまなポリシー領域に精通した100以上の外部組織からなる「YouTube 公認報告者」プログラムも実施しています。公認報告者による報告は、YouTube のポリシーとその施行に関するスペシャリストが優先的に審査し、必要に応じて対処します。「YouTube 公認報告者」の貴重な専門知識は、私たちの自動検出システムを補完するものです。
2-④	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	回答なし 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ます。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ます。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ます。 具体的な人数などは非公開となっておりますが、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しています。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います。そのために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働いています。こうした体制でグローバルに統一して適用するコミュニティガイドラインに基づいて対策を執行しています。例えば、ある言葉がコミュニティガイドライン違反の誹謗中傷に当たるかどうかといったことは、言語の理解、そして誹謗中傷されている対象の理解、そのビデオないしコメントがなされたタイミングや背景など総合的な理解が求められます。そうした理解の必要性をYouTube でも強く認識しており、適切な体制（日本語のネイティブのレビュアー等）を整えています。	回答なし 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ます。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ます。現在ではGoogle 全体で20,000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ます。一方で、日本の案件の審査と承認は、日本語話者だけではなく、さまざまな部署のグローバルなチームがそれぞれの専門性を持って対応しておりますため、日本の案件だけに従事する人間の数を正確にお示しすることができませんが、日本をカバーする人数は必要に応じて拡充してまいります。
2-⑤	(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	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す。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ます。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ます。削除するかどうかの決定の結果についても、メールで伝えられます。濫用を防ぐために結果に至った理由は詳細には伝えられませんが、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ます。	[Google検索] 晒し行為に関するポリシーに沿って報告されたリクエストへの対応を行った際には、ユーザーにお知らせします。こちらのリンクの晒し行為に関するヘルプセンターでは、削除リクエストを送信した後の流れを説明しています。 ● 自動確認メールが届きます。これにより、リクエストが受領されたことがわかります。 ● Google による審査が行われます。各リクエストは上記の要件に基づいて評価されます。 ● 必要に応じて、より詳細な情報を収集します。場合によっては、詳細な情報の提供をお願いすることがあります。不明な URL など、評価するのに十分な情報がリクエストに含まれていない場合は、具体的な手順をお伝えし、リクエストを再送信するようお願いしています。 ● 対応を行った際には、お知らせいたします。 ○ リクエストが削除要件を満たしていない場合は、簡単な説明も記載されます。リクエストが拒否されたものの、その後、対応するのに役立つ追加資料がある場合は、リクエストを再送信できます。 ○ 報告されたURL がGoogle のポリシーの適用範囲にあると判断された場合、そのURL は、申立人の名前またはその他の識別情報（エイリアス、ソーシャルメディアのハンドル名など）が含まれるクエリの検索結果からのみ除外されます。 不当な削除方針が示されているサイトに関するポリシーに沿って報告されたリクエストへの対応を行った際には、ユーザーにお知らせします。ヘルプセンター ページであるこちらのリンクでは、削除リクエストを送信した後の流れを説明しています。 ● 自動確認メールが届きます。これにより、リクエストが受領されたことがわかります。 ● Google による審査が行われます。各リクエストは上記の要件に基づいて評価されます。 ● 必要に応じて、より詳細な情報を収集します。場合によっては、詳細な情報の提供をお願いすることがあります。不明な URL など、評価するのに十分な情報がリクエストに含まれていない場合は、具体的な手順をお伝えし、リクエストを再送信するようお願いしています。
			対応を行った際には、お知らせいたします。リクエストが削除要件を満たしていない場合は、簡単な説明も記載されず、リクエストが拒否されたものの、その後、対応するのに役立つ追加資料がある場合は、リクエストを再送信できます。 [YouTube] 1) コミュニティによる報告について、ユーザーはいつでも報告履歴ページで確認することができます。詳しくは、こちらのヘルプページをご覧ください。さらにコンテンツの削除を行った場合は、報告したユーザーにメールでお知らせします。このプロセスがより良いものとなるよう、新しい方法を検討中です。 2) 公認報告者は、報告したコンテンツの審査結果を表示したり、ポリシー担当者や連絡を取って報告に関する説明やフィードバックを得ることができます。 3) 法律に基づく削除リクエストの場合、リクエストした人はメールで結果を受け取ることができます。 https://www.youtube.com/reporthistory https://support.google.com/youtube/answer/7687979?user_flagging_history&hl=ja&visit_id=637814252500116316-4041843369&rd=1

Google 合同会社		ヒアリングシート(2021年2月25日)回答	ヒアリングシート(2022年3月7日)回答
(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>[YouTube] コミュニティガイドラインに違反するコンテンツが発見され、違反警告が発行されると、クリエイターにメールで通知されます。またクリエイターは、モバイルとパソコンの通知設定やチャンネル設定にて、通知を受信するように選択することもできます。 クリエイターにお知らせする内容は以下のとおりです。</p> <p>[ポリシーに基づく削除] - 削除されたコンテンツ - 違反したポリシー - チャンネルへの影響 - 対処方法</p> <p>[法律に基づく削除] - 削除されたコンテンツ - どの国の Youtube のコンテンツがブロックされたか - 制限が誤って加えられたと考える場合、クリエイターはどこにフィードバックをするのか</p> <p>YouTube 側に誤りがあると思う場合には、YouTube Studio を通じて再審査請求を行うことができます。詳しくは、セクション2-Q6 の回答をご覧ください。詳細は次のリンクでご確認いただけます。 https://support.google.com/youtube/answer/185111</p> <p>[Google 検索] ユーザーがGoogle 検索エンジンに投稿することはないため、「投稿」は検索には該当しません。むしろ、URL はクロールされ、インデックス化されます。</p>
2-⑥ 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>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ります。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できます。 https://policies.google.com/terms?hl=ja#toc-removing https://policies.google.com/terms/service-specific?hl=ja</p> <p>例えば、YouTube では再審査請求を認めています。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ます。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ます。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ます。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ます。詳細はこちら https://support.google.com/youtube/answer/185111?hl=ja</p> <p>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できます。 https://support.google.com/local-guides/contact/local_guides_appeal?hl=ja</p>	<p>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ります。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できます。 https://policies.google.com/terms?hl=ja#toc-removing https://policies.google.com/terms/service-specific?hl=ja</p> <p>[Google 検索] 検索エンジンは一般的に、コンテンツ制作者と直接的な契約関係を結びません。多くの場合、特定のウェブサイトに関して Google が保有する唯一の情報は、ウェブブラウザで誰もが知ることのできる情報です。この点で、ソーシャルメディアサービスで一般的に理解されている再審査請求が Google 検索と同じように機能することはありません。ユーザーがサインインしてコンテンツを投稿するホスト型プラットフォームとは異なり、ウェブサイトの管理者はサインインして URL を「投稿」するわけではありません。代わりに Google 検索は、他の検索エンジンと同様に、オープンウェブをクロールしてインデックスに追加します。このため、検索エンジンでの再審査請求の提供は、ホスト型サービスよりも困難です。ポリシーの適用により影響を受けるサイトが登録されている場合、Google 検索を行った対応について、私たちは可能な限りウェブサイトの管理者向けの Google Search Console を通じてお知らせします。 https://search.google.com/search-console/about</p> <p>[YouTube] 例えば、YouTube では再審査請求を認めています。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ます。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ます。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ます。クリエイターには、再審査結果を知らせるメールが送られます。詳細はこちら https://support.google.com/youtube/answer/185111?hl=ja</p> <p>[Google マップ] 私たちは、「不適切あるいは、露骨な性的表現を含む」と分類された報告を、人間のオペレーターによって確認しています。人間のオペレーターと機械学習システムを組み合わせて、不適切/憎悪(ヘイト)/ハラスメントのコンテンツが公開される前に検出しています。また、これらの検出システムを使用して、Google マップに投稿された後のレビューを継続的に評価し、Google のポリシーに違反するコンテンツを削除しています。</p>
2-⑧ 不正な申告や削除要請への対策の方法・仕組み、対応件数	【具体的な方法・仕組み】	<p>Google の製品でポリシー違反のコンテンツを見つけたユーザーから報告をいただけることはとてもありがたいことです。例えば YouTube では、YouTube 公認報告者プログラムを設け、コミュニティガイドラインに違反しているコンテンツの報告に関して特に貢献度の高い個人、政府機関、非政府組織 (NGO) に強力なツールを提供しています。日本では自殺予防に取り組む日本の非営利団体 OVA と連携し、公認報告者として、プラットフォーム上の問題のあるコンテンツをフラグしていただいております。公認報告者から報告された動画は、YouTube のコミュニティガイドラインに基づいて審査されます。公認報告者から報告されたコンテンツは、自動的に削除されたり、異なるポリシーによって対処されたりするわけではありません。他のユーザーから受け取った報告と同じ基準が適用されます。ただし、その高い正確性により、YouTube 公認報告者からの報告は優先的に審査されます。</p> <p>一方で、究極的には、YouTube の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と私たちは考えています。このため、YouTube を含む様々なプロダクトで、当社の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ます。</p> <p>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ます。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ます。</p> <p>こうした取り組みに投資を継続し、ユーザーから報告されなくとも有害なコンテンツからプラットフォームやユーザーを守ることができるようにすることを目指しています。</p>	<p>[Google 検索] Google 検索に関しては上記 Q2-6 の回答をご覧ください。</p> <p>[YouTube] Google の製品でポリシー違反のコンテンツを見つけたユーザーから報告をいただけることはとてもありがたいことです。例えば YouTube では、YouTube 公認報告者プログラムを設け、コミュニティガイドラインに違反しているコンテンツの報告に関してその分野の知識が多い個人、政府機関、非政府組織 (NGO) にツールを提供しています。日本では自殺予防に取り組む日本の非営利団体 OVA や、法務省人権擁護局と連携し、公認報告者として、プラットフォーム上の問題のあるコンテンツをフラグしていただいております。公認報告者から報告された動画は、YouTube のコミュニティガイドラインに基づいて審査されます。公認報告者から報告されたからといって、自動的に削除されたり、異なるポリシーによって対処されたりするわけではなく、他のユーザーから受け取った報告と同じ基準が適用されます。ただし、その高い正確性により、YouTube 公認報告者からの報告は優先的に審査されます。 https://support.google.com/youtube/answer/7554338?hl=ja https://ova-japan.org/ https://www.moj.go.jp/JINKEN/jinken03_00084.html</p> <p>しかし究極的には、ポリシー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と私たちは考えています。このため、YouTube を含む様々なプロダクトで、当社の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ます。</p> <p>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だと考えています。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ます。</p> <p>こうした取り組みに投資を継続し、有害なコンテンツからプラットフォームやユーザーを守ることができるようにすることを目指しています。</p>
3-① 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】	<p>「誹謗中傷に関連する一般ユーザーからの報告件数」</p> <p>Google は、2019 年に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 を約 7000 件受け取っています。</p> <p>「誹謗中傷に関連する司法機関、行政機関からの削除要請の件数」 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) で、政府から受けた依頼全体の 85% を占めています。政府からのコンテンツ削除依頼に関する情報は、こちらの Google の透明性レポートで確認できます。 https://transparencyreport.google.com/government-removals/by-country/JP?hl=ja</p>	<p>「誹謗中傷に関連する一般ユーザーからの報告件数」</p> <p>[Google 検索] Google は、2021 年 1 月から 6 月にかけて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 を約 2300 件受け取っています。 https://support.google.com/legal/troubleshooter/1114905</p> <p>[YouTube] YouTube では、世界中のユーザーや YouTube 公認報告者から、コミュニティガイドラインの違反が疑われる動画に対する報告を受け取っています。日本は世界中で 7 番目にこうした報告が多い国です。詳細は透明性レポートでご確認ください。 https://transparencyreport.google.com/youtube-policy/flags?hl=ja</p> <p>[Google マップ] 2021 年 1 月 - 10 月の間に不適切/露骨な表現に該当する報告を合計 218,870 件受けました。複数のユーザーから報告を受けるレビューがあるため、報告の総数は、報告を受けたコンテンツ数 (175,881 件) より多くなっています。こうした報告は人間のオペレーターが目視確認しています。</p> <p>「名誉毀損に関連する司法機関、行政機関からの削除要請の件数」 2021 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 29 件 (アイテム数 76 件) で、政府から受けた依頼全体の 32.22% を占めています。政府からのコンテンツ削除依頼に関する情報は、こちらの Google の透明性レポートで確認できます。 https://transparencyreport.google.com/government-removals/by-country/JP?hl=ja</p> <p>[Google マップ] Google マップでは、2021 年 1 月から 10 月の間に、政府からの公式現地レビューの削除要請が 1 件だけありましたが、ヘイトやハラスメントとは関係ありません。</p>
3-② (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	【日本の数値】	<p>「誹謗中傷に関連する一般ユーザー等からの申告や削除要請に対する削除件数」 2020 年 10 月から 12 月までの間に、日本で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は 49,180 本です。これは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数字です。(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% です。</p> <p>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、2019 年に Google は法的理由で約 1700 件の URL を検索結果から非表示にしています。非表示としたものは約 24% となります。投稿された URL のうち、約 55% が非表示となりませんでした。提出された URL のうち約 20% は、Google から報告者に追加情報の提供を求める必要がありました。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、などが挙げられます。</p> <p>「誹謗中傷に関連する司法機関、行政機関からの削除要請に基づく削除件数」 2020 年 1 月から 6 月にかけて、日本の政府からの請求や裁判所の命令に基づき、449 件を Google 製品から削除しています。(この数字は名誉毀損以外の請求などに関するものも含みます。)</p>	<p>「誹謗中傷に関連する一般ユーザー等からの申告や削除要請に対する削除件数」</p> <p>[YouTube] 透明性レポートで国別に公開されている通り、2021 年 10 月から 12 月までの間に、日本でアップロードされた動画の内、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は 24,153 本です。これはユーザーからの報告や機械によって検知されたものを含む数字です。グローバルでは、削除された約 370 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% に当たる 322,627 件です。日本はヘイトスピーチが原因で削除された件数は 620 件、嫌がらせやネットいじめが原因で削除された件数は 626 件でした。 https://transparencyreport.google.com/youtube-policy/removals?hl=ja&videos_by_country=period:2021Q4;region:JP&u=videos_by_country</p> <p>[Google マップ] 2021 年 1 月から 10 月の間に 32,190 件 (そのうちヘイト、不適切、ハラスメントなど: 3,648 件) をユーザーからの報告を受けて削除しています。</p> <p>「誹謗中傷に関連する司法機関、行政機関からの削除要請に基づく削除件数」 2021 年 1 月から 6 月にかけて、日本の政府からの請求や裁判所の命令に基づき、90 件の依頼 (アイテム数 287 件) を受け取っています。(この数字は名誉毀損以外の請求などに関するものも含みます。) 政府からのコンテンツ削除依頼に関する情報は、こちらの Google の透明性レポートで確認できます。</p> <p>Google マップについては、セクション 3-Q1 の回答をご覧ください。</p> <p>「ユーザーウェブフォーム申請」 (※こちらに記載されている数字はすべて概数となります) 2021 年 1 月から 6 月にかけて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 2300 件の URL のうち、Google は法的理由で約 300 件の URL を検索結果から非表示にしています。非表示としたものは報告された全体の約 15% となります。報告された URL のうち、コンテンツが確認できなかったり、すでに非表示となっていたものも含め、約 34% が非表示となりませんでした。報告された URL の約 51% は、Google から報告者に追加情報の提供を求める必要がありました。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場合や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場合などが挙げられます。 https://support.google.com/legal/troubleshooter/1114905</p>
	【グローバルの数値】	回答なし	回答なし

Google 合同会社	ヒアリングシート(2021年2月25日)回答		ヒアリングシート(2022年3月7日)回答
(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)一般ユーザー、(2)司法機関・行政機関	【日本の数値】 【グローバルの数値】		例えば、Google マップでは、計測した期間(2021年1月-10月)において、日本ではおよそ120万件のレビューが公開前に機械学習によってブロックされています。 回答なし
(iii)対応を行わなかった件数 (1)一般ユーザー、(2)司法機関・行政機関	【日本の数値】 【グローバルの数値】		(1)一般ユーザー 例えば、Google マップでは、2021年1月から10月の間に約143,000件の日本のユーザーからの報告が、審査の結果、ポリシー違反はないと判断されました。 (2)司法機関・行政機関 例えば、透明性レポートで公開しているように、2021年1月~2021年6月に、日本政府や裁判所から受けた報告の内訳を見ると、現地の法律に基づいて(この数字は名誉毀損以外の請求などに関するものも含みます)削除されたものが120件(41.8%)、Googleのポリシーに基づいて削除されたものが38件(13.2%)となっています。一方、情報不足で56件(19.5%)、コンテンツ不足で28件(9.8%)、すでに削除されたコンテンツで9件(3.1%)については、アクションを起こすことができなかった。 レビューの結果、36件(12.5%)はアクションを取らなかった。 このように、Googleが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さまざまです。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ません。上記に示す情報で一部は削除割合を示してはいますが、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ください。
(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)一般ユーザー、(2)司法機関・行政機関	【日本の数値】 【グローバルの数値】	回答なし 回答なし	たとえば、Google マップでは、ヘイトスピーチやハラメントに関するユーザーからの報告をすべて評価し、タイムリーに対処するための人材とトレーニングに投資しています。「不適切/露骨な」カテゴリに関するユーザーからの報告をすべて24時間以内に見直し、中には数時間で対処するものもあります。 YouTubeでは、グローバルプラットフォームとしての責任を果たしているかどうかを測る指標として、違反コンテンツの視聴率(VVR)を調査しています。VVRは、簡単に言うと、YouTube上の全体の動画視聴の中で、YouTubeのコミュニティガイドラインに違反する動画が何回視聴されたかを集約したものです。ポリシーの適用のために必要な人材やテクノロジーへの投資を拡大するにつれ、VVRは低下しています。最新のVVR(2021年第4四半期)は0.12~0.14%で、これはYouTubeにおける視聴1万回のうち、12~14回が違反コンテンツによるものということになります。こちらのVVRデータも今年から、透明性レポートの中で公開し、四半期ごとに更新していきます。 https://transparencyreport.google.com/youtube-policy/views?hl=ja https://transparencyreport.google.com/youtube-policy/views?hl=ja
(i)主體的な削除件数(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】	回答なし 2020年10月から12月の間に、YouTubeは930万本以上の動画をコミュニティガイドラインに違反したことを理由に削除しました。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が立てられました。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回以下の視聴回数うちに削除されています。 同期間にYouTubeは12億件以上のコメントを削除しました。削除されたコメントの99.6%が自動的に検出されました。	例えばGoogle マップでは、日本での計測期間(2021年1月-10月)において、およそ120万件のレビューが公開前に機械によりブロックされています。私たちの自動モデレーションによって「関連性のない」とみなされるものを含め、ポリシーに違反するコンテンツを削除しています。120万件のレビューのうち、ヘイト、ハラメント、不適切な発言と分類されたものは、約9%に過ぎませんでした。そのうち1,342件は、機械によりヘイトと分類され、疑わしいコンテンツとタグ付けられ、人間のオペレーターを助けました。 2021年10月から12月の間に、YouTubeは370万本以上の動画をコミュニティガイドラインに違反したことを理由に削除しました。これらの動画の92%に当たる3,451,691件は、人間ではなく機械によって最初にフラグが立てられました。人工知能によって検出された動画のうち、31.7%は一度も視聴をされる前に、38%は1~10回以下の視聴回数うちに削除されています。 同期間にYouTubeは12億件以上のコメントを削除しました。削除されたコメントの99.5%が自動的に検出されました。
(ii)主體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】	回答なし 回答なし このような取り組みの個別件数については、グローバルに開示しておりません。一方で、ユーザーの安全を守るために、どのような仕組みが機能しているかについては、さまざまな形で開示しています。 Google 検索について詳細は、How Search Works(Google 検索の仕組みを確認する)をご覧ください。 https://www.google.com/intl/ja/search/howsearchworks/ YouTube についての詳細は、How YouTube Works (YouTube の仕組みとは)をご覧ください https://www.youtube.com/howyoutubeworks/ Google マップについての詳細は、Google のクチコミの仕組みについてのブログをご覧ください。 https://japan.googleblog.com/2022/02/google.html
(iii)主體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】	Googleマップは、機械学習による自動システムを定期的に更新し、対象となるコンテンツが公開される前に検出、そして削除される精度を高めています。まれに自動システムによる検知がなかったコンテンツがあった場合には、ユーザーからの違反報告に応じて速やかに対応し、1日以内に削除します。 回答なし G3-2(iv)の回答をご覧ください。
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)	【日本の取組】	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実際には違反していないため、削除されない場合があります。私たちはこれをボーダーラインのコンテンツと呼んでいます。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%にも満たない数ではありますが、これについてもYouTubeは真剣に取り組んでいます。YouTubeのおすすめ機能は、ユーザーがより多くのお気に入りの動画を発見するのに役立ちます。YouTubeではそのようなボーダ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しています。	YouTubeは、誰でも動画をアップロードして世界中の人と共有できる、オープンな動画プラットフォームです。このオプション性(開かれた場であること)には、希薄の可能性と重大な責任が伴います。そのためYouTubeは、創造的な表現と、有害なコンテンツからコミュニティを保護する責任とのバランスを取るよう常に取り組んでいます。私たちのアプローチは、4つの柱とされている4つの「R」(Rから始まる4つの単語)に沿っています。1つ目の柱は、違反コンテンツを削除(remove)することです(詳細は上記をご覧ください)。 2つ目の柱は、ニュース速報時など、ユーザーがニュースや情報を探しているときに、信頼できる情報を見つけやすく(raise)することです。これについては、誤情報についての次の研究会ヒアリングで詳しく説明させていただきますと思います。 3つ目の柱は、YouTubeのガイドラインに違反するかどうか際どい、ボーダーライン上のコンテンツの拡散を減らす(reduce)ことです。2011年以降、センセーショナルなコンテンツやクリックバイトコンテンツ、質の低いコンテンツの質の低い閲覧されないようにするため、リコメンデーションシステムを改善してきました。この種のコンテンツの拡散を制限するため、2019年1月には当該システムの変更を発表しました。これによりほとんどのユーザーは、当該チャンネルに登録するといった積極的な選択をしない限り、このようなコンテンツを目にする可能性はるかに低くなりました。 アルゴリズムを変更して稼働させるには時間がかかりますが、私たちの目標は、ガイドラインにぎりぎり違反しないボーダーライン上のコンテンツの、リコメンデーションシステムを通じた視聴(チャンネル登録者へのおすすめは除く)を0.5%未満にすることです。この数字をゼロにすることが目標ですが、完璧なシステムはありません。この数字をさらに低下させるための対策が意図しない悪影響をもたらし、正当な表現がおすめられなくなってしまう可能性がある事にも留意が必要です。 4つ目の柱は、収益化プログラムを通じて、信頼できるクリエイターに利益を還元する(reward)ことです。私たちはYouTubeパートナープログラムを導入し、YouTubeでどのようなチャンネルが収益化できるかという基準を厳格化しました。このプログラムでは、収益化はコミュニティガイドラインと広告掲載に適切なコンテンツのガイドラインに従った信頼できるクリエイターのみが対象であると明確にしています。ハラメントやネットいじめに関するポリシーへの違反を繰り返すチャンネルは、YouTubeパートナープログラムへの参加が停止され、YouTubeでの収益化が無効になります。このようなチャンネルは違反警告を受けたり(コンテンツの削除に至る可能性があります)、アカウントが停止されたりする場合もあります。 https://support.google.com/youtube/answer/72851?hl=ja https://support.google.com/youtube/answer/6162278?hl=ja
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】 【グローバルの数値】	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】 政府からの情報開示リクエストに対して、Googleがどのような対応をとったかについて透明性を持たせることを主目的としている弊社の透明性レポートでは、その目的に沿って裁判所を含む政府機関(民事訴訟を除く)におけるユーザー情報の開示についてのデータのみ提供しております。 こちらで記載しております通り、2019年7月~2019年12月の日本におけるユーザー情報の開示リクエストは合計241件、アカウント数は464件、なんらかの情報の開示に至ったリクエストの割合は87%です。 また裁判外の開示請求について、日本法人への請求については、運営主体であるGoogle LLCに連絡するように通知しています。運営主体であるGoogle LLCへの請求はほぼないと理解していますが、仮にあった場合には裁判手続きを経なければGoogle LLCは開示できない旨が米国の法律で定められているため、それに従った対応をすることになっています。	回答なし 回答なし 政府からの情報開示リクエストに対して、Googleがどのような対応をとったかについて透明性を持たせることを主目的としている弊社の透明性レポートでは、その目的に沿って裁判所を含む政府機関(民事訴訟を除く)におけるユーザー情報の開示についてのデータのみ提供しております。 こちらで記載しております通り、2020年7月~2020年12月の日本におけるユーザー情報の開示リクエストは合計74件、アカウント数は135件、なんらかの情報の開示に至ったリクエストの割合は54%です。 訴訟以外の開示要求に関しては、当社の日本の子会社から、運営事業体であるGoogle LLCに連絡するよう請求者に対して通知しました。運営事業体としてのGoogle LLCへの要求はほぼ無いに等しいと理解していますが、もしあった場合、Google LLCは、米国の法律が規定する司法手続きを経ることなしに情報を開示することはできませんが、相応に対応していきます。 Google 検索は、インデックス(索引)であり、ホスティングを行うプラットフォームではありません。ユーザーは、検索に対してコンテンツを「投稿」するわけではありません。URLは検索され、見出しを付けられ、整理されます。詳細は、こちらをご覧ください。 https://www.google.com/search/howsearchworks/

Google合同会社		ヒアリングシート(2021年2月25日)回答		ヒアリングシート(2022年3月7日)回答	
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】	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
		【(質問に答えられない場合)参考となる数値】	刑事(民事は含まない)87%	Q4-1の回答をご覧ください。	
5-①	透明性レポートの公開 日本語で閲覧可能か		はい、日本語で公開しております。こちら https://transparencyreport.google.com/?hl=ja	はい、日本語で公開しております。こちら https://transparencyreport.google.com/?hl=ja	
5-②	誹謗中傷等の取組の効果分析	<p>私たちは、YouTube上のボーダ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されて視聴する動画の再生時間を70%削減(米国)することに成功しました。この様なシステムの拡充を進めていく中で、他の地域へのインパクトについても、より多くの情報を共有できるようにしていきたいと考えています。また、私たちは今後もチームや新機能への投資も続けていきます。</p> <p>また第2項の質問2と3で説明したように、私たちは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しました。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、私たちのチームは200万本以上のコンテンツをレビューしました。私たちは、自動化されたフラグシステムの品質と有効性を継続的に向上させるために取り組んでいます。</p> <ul style="list-style-type: none"> ● 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回未満でした。 ● 2018年第1四半期には、この数字は50%以上に達しました。 ● 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ました。 ● 2020年第4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ました。削除されたコメントの99.6%が自動的に検出されました。 	<p>2019年、ガイドラインのボーダーライン上のコンテンツの拡散を制限するため、リコメンデーションシステムの変更を発表しました。その結果、同年米国では、ガイドラインに違反しないがボーダーライン上のコンテンツに関して、おすすめ動画等からの視聴時間が(チャンネル登録者へのおすすめは除く)70%減少しました。他の市場でも、おすすめ動画をきっかけとして見つけたボーダーライン上のコンテンツの視聴時間が低下しました。</p> <p>アルゴリズムを変更して稼働させるには時間がかかり、ガイドラインに違反するかどうかの際どいボーダーライン上のコンテンツの消費は増減する可能性があります。私たちの目標は、こうしたコンテンツにおけるおすすめ動画等からの視聴(チャンネル登録者へのおすすめは除く)を0.5%未満にすることです。現在、私たちが事業を展開しているすべての地域で、ボーダーライン上のコンテンツの視聴を減らすためのリコメンデーションシステムの変更を導入しています。もちろん、これには日本も含まれます(上記セクション3-Q4の回答をご参照ください。)</p> <p>セクション3-Q3でご説明したとおり、YouTube上の過激主義コンテンツを検出する機械学習テクノロジーを2017年6月に導入しました。機械学習を訓練するにあたって、YouTubeチームは200万以上のコンテンツを審査しました。私たちは、自動で報告を行うシステムの品質と有効性を向上させる取り組みを継続しています。</p> <p>YouTubeは、プラットフォームから有害コンテンツを削除する取り組みについて説明責任を果たすため、「YouTubeコミュニティガイドラインの適用について」というレポートを発表しました。このレポートでは、当社のポリシーに違反した動画やチャンネル、コメントのグローバル全体の削除、再審査請求とそれによって元に戻された動画、人間と機械による報告、に関するデータを提供しています。また、ハラスメントやネットいじめ、暴力的過激主義、ヘイトスピーチといった主要なコンテンツポリシー領域について詳細な分析も提供しています。たとえば、当社の透明性レポートでは、次のような内容を報告しています。</p> <p>https://transparencyreport.google.com/youtube-policy/removals?hl=ja</p> <ul style="list-style-type: none"> ● 2017年第1四半期には、暴力的または危険なコンテンツに関する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回未満でした。 ● 2018年第1四半期には、この数字は50%以上に達しました。 ● 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ました。 ● 2021年第4四半期には、ハラスメントやネットいじめ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において、YouTubeは12億件以上のコメントを削除しました。削除されたコメントの99.5%が自動的に検出されました。 		

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）回答との主な項目の比較【LINE株式会社】

参考資料4-3

※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。

LINE株式会社		ヒアリングシート(2021年2月25日)回答		ヒアリングシート(2022年3月7日)回答	
1-①	対象サービス名		LINE タイムライン、LINE Openchat、LINE LIVE、LINE BLOG、livedoor Blog		LINE VOOM、LINEオープンチャット、LINE LIVE、LINE BLOG、livedoor Blog
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】	国内MAU 8,600万人(2020年9月末時点)		国内 MAU 8,900万人(2021年9月末時点)
	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	【グローバルの数値】	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)		主要4か国(日本・台湾・タイ・インドネシア)MAU 1億7,300万人(2021年9月末時点)
2-①	(i)一般ユーザーからの申告・削除要請への受付窓口・受付態勢		お問い合わせフォーム等を通じてご連絡いただき、スタッフが対象となるコンテンツの確認や本人確認などを行い、判断します。 [タイムライン][Openchat][LIVE] CSフォームからの問い合わせという形でご連絡いただき、CSの対応マニュアルに則り対応を進めます。 [BLOG] [livedoor Blog] 以下の流れに沿って対応を進めます。 1)入信フォームや郵送(プロバイダ責任制限法の依頼書)で削除等の依頼 2)削除の場合、本人・第三者に限らずモニタリングで対応が可能かどうかエスカレーション(削除が完了すればその旨申し立て者に通知) 3)モニタリングで対応不可の場合は、弊社の方針(プロバイダ責任制限法)に基づく対応を行う旨を伝え必要書類を送るようご案内		お問い合わせフォーム等を通じてご連絡いただき、スタッフが対象となるコンテンツの確認や本人確認などを行い、判断します。 [LINE VOOM][LINEオープンチャット][LINE LIVE] CSフォームからの問い合わせという形でご連絡いただき、CSの対応マニュアルに則り対応を進めます。 [LINE BLOG] [livedoor Blog] 以下の流れに沿って対応を進めます。 1)入信フォームや郵送(プロバイダ責任制限法の依頼書)で削除等の依頼 2)削除の場合、本人・第三者に限らずモニタリングで対応が可能かどうかエスカレーション(削除が完了すればその旨申し立て者に通知) 3)モニタリングで対応不可の場合は、当社の方針(プロバイダ責任制限法)に基づく対応を行う旨を伝え必要書類を送るようご案内
	(ii)主体的な誹謗中傷等の情報の検知態勢		一部のLINEサービスで適用されている機械的なチェックでは、事前にシステムに登録している禁止用語やルールと照合し、規約や法令に反した投稿になっていないかを確認します。これに該当する場合には、投稿後ただちに自動で投稿が非表示にされる仕組みになっています。次に、人の目によるチェックでは、機械的なチェックでは判断のつかなかったケースについて、事前に用意した判断基準や事例集と照合して規約や法令に反する投稿がないかを一件ずつ担当者が確認します。モニタリングチームがコミュニティ規約や法令に反すると判断した場合には、担当者がその投稿を非表示にします。 ・AIを活用した違反“画像”の検知 検知対象: わいせつ、出会い系、不快画像など 現状: 違反画像検知AIをLINE全サービスで適用中 展開: 精度向上の対応継続中 ・AIを活用した違反“テキスト”の検知 検知対象: 誹謗中傷、スパムなど 現状: テキスト違反検知AIを特定サービスでテスト中 展開: 2021年上半年期 タイムラインに実装予定 2021年下半年期 その他全てのサービスに実装予定		一部のLINEサービスで適用されている機械的なチェックでは、事前にシステムに登録している禁止用語やルールと照合し、規約や法令に反した投稿になっていないかを確認します。これに該当する場合には、投稿後ただちに自動で投稿が非表示にされる仕組みになっています。次に、人の目によるチェックでは、機械的なチェックでは判断のつかなかったケースについて、事前に用意した判断基準や事例集と照合して規約や法令に反する投稿がないかを一件ずつ担当者が確認します。モニタリングチームがコミュニティ規約や法令に反すると判断した場合には、担当者がその投稿を非表示にします。 ・AIを活用した違反“画像”の検知 検知対象: わいせつ、出会い系、不快画像など 現状: 違反画像検知AIをLINE全サービスで適用中 展開: LINE全サービスで共通の違反に対する自動処理を実装予定 ・AIを活用した違反“テキスト”の事前検知(ユーザーへのポップアップ警告) 検知対象: 誹謗中傷、スパムなど 現状: テキスト違反検知AIをテスト中 展開: 各サービスにおいて逐次開発予定
2-③	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				・入信フォームや郵送で削除等の依頼があったものから、順次対応しています。なお、捜査機関から削除要請(被害予告等)や情報開示請求を受けた場合、投稿記事削除及び発信者情報開示仮処分命令申立書を受けた場合には優先して対応しています。 ・モニタリングにおいては、ユーザーのアクセスが多いチャットやコメント欄等における投稿から優先的に対応しており、投稿内容に応じてポイントを付与し、その高いものから順に対応しています。
2-④	一般ユーザー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	一般ユーザー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	各サービスにつき、10名程度		・ユーザーからの削除要請については、各サービスにつき10名程度で対応しています。モニタリングについては、各サービスにつき15名～90名程度で対応しています。
		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	日本国内に拠点が有り。日本における責任者を置いている。		・東京(LINE株式会社)及び福岡(LINE Fukuoka株式会社)に拠点があります。また、日本における責任者を置いています。
2-⑤	(i)対応決定時の申立て者への通知の有無とその理由説明内容		削除可否等について通知。		・削除された場合、申立て者に「送信防止措置が講じられています。」との通知をしています。また、削除されなかった場合には発信者から送信防止措置に同意しない旨の回答があったことを伝え、プロバイダ責任制限法の当該条文を提示し、権利侵害が明白とは判断できないため削除できないことを通知しています。 ・送信防止措置の不同意に関して異議申し立てがある場合、もしくはプロバイダ責任制限法以外の対応を求められた場合には、状況に応じて、弁護士等への相談のほか、インターネット安心・安全相談(警察庁)やインターネット・ホットラインセンターを案内しています。
	(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				プロバイダ責任制限法に基づく対応をする際、発信者へ連絡することが可能な場合には発信者へ申立て者から送信防止措置依頼がきていること及び同依頼内容について通知を行っています。この際、送信防止措置に不同意の場合には同意しない旨を返信することを依頼するとともに、不同意の返信がない場合にはLINE側で該当する投稿等を削除する旨を通知しています。
2-⑥	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		[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。		[LINE VOOM][LINEオープンチャット][LINE 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します。 [LINE 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します。
2-⑧	不正な申告や削除要請への対策の方法・仕組み、対応件数	【具体的な方法・仕組み】	回答なし		基本的にはどなたでも自由に申告等を行える仕組みとしていますが、ヘルプやWebページにおいて啓蒙活動を実施する(LINE BLOG)、第三者からの申告については「内容を確認するが個別返信は行わない場合がある」旨をお問い合わせフォーム上で明確に表示する(livedoor Blog)等の取り組みを行っています。
		【日本の数値】	[livedoor Blog] 857件 ※主に違法性がないと思われる投稿等への濫用的な申告の件数。		(2021年1月～12月) [LINE VOOM] 53件 [LINEオープンチャット] 809件 [LINE LIVE] 9件 [LINE BLOG] 0件 [livedoor Blog] 1,316件 ※主に違法性がないと思われる投稿等への濫用的な申告の件数。
3-①	誹謗中傷等に関する申告や削除要請の件数 (1)一般ユーザー、(2)司法機関・行政機関	【日本の数値】	(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		(2021年1月～12月) (1)一般ユーザー [LINE VOOM] 693件 [LINEオープンチャット] 4,701件 [LINE LIVE] 52件 [LINE BLOG] 9件 [livedoor Blog] 3,644件 ※ LINE VOOM/LINEオープンチャット/LINE LIVEについては、迷惑行為全般に関する件数。LINE 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。 (2)司法機関・行政機関 [LINE VOOM] [LINEオープンチャット] [LINE LIVE] [LINE BLOG] 0件 [livedoor Blog] 11件
3-②	(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)一般ユーザー、(2)司法機関・行政機関	【日本の数値】	(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		(2021年1月～12月) (1)一般ユーザー [LINE VOOM] 20件 [LINEオープンチャット] 56件 [LINE LIVE] 9件 [LINE BLOG] 0件 [livedoor Blog] 0件 (2)司法機関・行政機関 [LINE VOOM] [LINEオープンチャット] [LINE LIVE] [LINE BLOG] [livedoor Blog] 0件
	(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)一般ユーザー、(2)司法機関・行政機関	【日本の数値】			(2021年1月～12月) (1)一般ユーザー [LINE VOOM] 2件 [LINEオープンチャット] 0件 [LINE LIVE] 0件 [LINE BLOG] 9件 [livedoor Blog] 1,364件 (2)司法機関・行政機関 [LINE VOOM] [LINEオープンチャット][LINE LIVE][LINE BLOG] 0件 [livedoor Blog] 1件(法務局からの削除依頼に基づき、記事の修正を発信者に依頼通知)、9件(記事の非表示化)
	(iii)対応を行わなかった件数 (1)一般ユーザー、(2)司法機関・行政機関	【日本の数値】			(2021年1月～12月) (1)一般ユーザー [LINE VOOM] 10件 [LINEオープンチャット] 40件 [LINE LIVE] 43件 [LINE BLOG] 0件 [livedoor Blog] 13件(送信防止措置不同意7件、一部修正の不同意6件) (2)司法機関・行政機関 [LINE VOOM] [LINEオープンチャット][LINE LIVE][LINE BLOG][livedoor Blog] 0件
	(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)一般ユーザー、(2)司法機関・行政機関	【日本の数値】	(i)一般ユーザー 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		(1)一般ユーザー モニタリングで削除可能な場合は平均24時間以内、プロバイダ責任制限法の対応の場合には発信者に意見照会をする必要があるため7日間前後 (2)司法機関・行政機関 1か月～3か月程度

LINE株式会社		ヒアリングシート(2021年2月25日)回答	ヒアリングシート(2022年3月7日)回答
3-③	(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】 [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。	(2021年1月～12月) [LINE VOOM] [LINEオープンチャット] [LINE LIVE] [LINE BLOG] [livedoor Blog] 0件
	(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】	(2021年1月～12月) [LINE VOOM] 約780,000件 [LINEオープンチャット] 約1,800,000件 [LINE LIVE] 約30,000件 [LINE BLOG] 約1,700件 [livedoor Blog] 約160,000件
	(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】	10分程度
3-④	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)	【日本の取組】 ・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。 ・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、機能制限が設けている。 ・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。 ・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。 ・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。 ・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年上半期にタイムラインに実装し、2021年下半期にOpenchat等の他サービスへ実装する予定。	LINEでは、友だちのブロック機能、年齢認証によるLINE ID検索の制限、友だち以外からのメッセージの受信拒否機能、友だち以外から送られたURLを開く際に警告を出す機能、通報機能が設けられています。 [LINEオープンチャット] 誹謗中傷や嫌がらせに対応するため、以下の活動を強化しています。 ・事前に設定した誹謗中傷関連のNGキーワードの自動非表示(当該チャットの管理者がNGキーワードを追加することも可能) ・安心安全ガイドライン(https://openchat-jp.line.me/other/guideline)を基本とした啓発活動の一環として、誹謗中傷や嫌がらせを禁止行為と明示し、ガイドラインページのフッターに厚労省のSNS相談サイトや#NoHeartNoSNS特設サイトへのリンクを掲載 ・ユーザが特定の検索ワードを入力した際に相談窓口へ誘導するバナーを表示 [LINE VOOM] ・コンテンツ及びコメントに対する通報機能、LINE VOOM上でのブロック機能(LINE VOOMでブロックした場合、お互いにフォローできなくなり、投稿を閲覧することができなくなる機能)が設けられています。
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 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)	(2021年1月～12月) (i) 裁判上の開示請求件数 14件(LINEオープンチャット 5件、livedoor Blog 9件) (ii) 裁判外の開示請求件数 24件(LINE LIVE 2件、LINE BLOG 1件、livedoor Blog 21件)
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 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)	(2021年1月～12月) (i) 裁判上の開示件数 8件(LINEオープンチャット 1件、livedoor Blog 7件) (ii) 裁判外の開示件数 9件(LINE BLOG 0件、livedoor Blog 9件)
5-①	透明性レポートの公開 日本語で閲覧可能か	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。	LINE Transparency Report (https://linecorp.com/ja/security/transparency/top)にて以下を公開 しています。 ・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しています。
5-②	誹謗中傷等の取組の効果分析	・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。 ・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。 ・「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。	・LINEみらい財団は、青少年のネット利用実態把握を目的とした調査を実施し、小中高生に対するネット上のコミュニケーションに関する教育の効果について分析して2021年4月にその成果を取り纏めた報告書をHP上に公開しています。調査の結果、教育を実施することで、ネット上での他人の悪口の書き込みや嫌がらせのメッセージの送信等が、その内容によっては違法であることを理解する小中高生の増加に寄与すること等が明らかになりました。 (https://linecorp.com/ja/csr/newslst/ja/2021/367)

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）回答との主な項目の比較【Meta】

参考資料4-4

※総務省において要約したものを。詳細は各社の回答原本を参照。

Meta		ヒアリングシート(2021年2月25日)回答		ヒアリングシート(2022年3月7日)回答
1-①	対象サービス名	Facebook	Instagram	Facebook/Instagram
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】 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)	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)
1-④	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	【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)	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) 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せんし、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。 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/ 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。 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や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、実施しなければなりません。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/
2-①	(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	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。「報告」の方法については、以下を参照してください。 https://www.facebook.com/help/1380418588640631/ 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。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。	Instagramでは、コンテンツを簡単に報告できるようにしています。また、コミュニティガイドライン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。Instagram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ガイドライン違反を報告することができます。「報告」の方法については、例えば以下を参照してください。 https://help.instagram.com/547601325292351?help=uf_permalink 利用者が問題のあるコンテンツを「報告」すると、Instagramはそのコンテンツを審査し、コミュニティガイドラインに準拠していないコンテンツをすべて削除します。Instagramが問題の原因となった人物と連絡を取った場合、ユーザー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。利用者がInstagramのコミュニティガイドラインに違反していることを「報告」した場合、利用者は利用者のサポートリクエスト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リクエストを見ることができるようになります。	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。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決定にも役立っています。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。 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/
	(ii)主体的な誹謗中傷等の情報の検知態勢	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。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。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。	FacebookやInstagramのコンテンツを簡単に報告できるようにし、コミュニティには報告ツールを使って違反していると思われるコンテンツを報告することを奨励しています。Instagramを利用している人は、各コンテンツに表示されるオプションを含め、さまざまな方法で違反行為を報告することができます。また、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なっています。また、AIは、コンテンツレビュー者がレビューするケースに優先順位をつけて、最も有害で時間的制約のあるコンテンツを最初に処理するのにも役立っています。「報告」の方法については、以下を参照してください。 https://help.instagram.com/477434105621119 利用者が問題のあるコンテンツを「報告」した場合、当社はそのコンテンツを検査し、当社のコミュニティガイドラインに準拠していないコンテンツをすべて削除します。問題の原因となった人物に連絡を取った場合、ユーザー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。利用者が違反を「報告」した場合は、サポートリクエストで報告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リクエストを見ることが出来ます。	

Meta		ヒアリングシート(2021年2月25日)回答	ヒアリングシート(2022年3月7日)回答
2-③ 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>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、特定の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。</p> <p>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> <p>コミュニティからの報告については、<input type="checkbox"/> 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。</p> <p>-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。</p> <p>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。</p> <p>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</p> <p>-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。</p> <p>-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。</p> <p>-(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。</p> <p>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すが、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。</p> <p>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</p> <p>-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認して、(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。</p> <p>-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われます。</p> <p>○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。</p> <p>○重大性 現実の被害に関連するコンテンツは、他のカテゴリよりも優先的に審査されます。</p> <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。</p> <p>https://transparency.facebook.com/community-standards-enforcement</p>
2-④ 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>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</p> <p>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】 回答なし</p> <p>【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(スードなど)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。</p> <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。</p>	<p>回答なし</p> <p>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々が集結しています。</p> <p>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して協力しています。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び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>
2-⑤ (i) 対応決定時の申立て者への通知の有無とその理由説明内容		<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。</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ずれも、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。</p> <p>場合によっては、(アダルトス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。</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。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、利用者に知らせず、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。</p> <p>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>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/#take-down-experience</p>
2-⑥ 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。コンテンツが誤って削除された場合、削除されたコンテンツは復元される。</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、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ほかが入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、利用者にその旨お知らせし、投稿、写真、ビデオは復元されます。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</p> <p>異議申し立ての手続きについては下記URLをご参照ください。</p> <p>https://transparency.fb.com/en-gb/policies/improving/appealed-content-metric/</p> <p>コンテンツがFacebookやInstagramから削除され、異議申し立ての手続きを経ている場合は、監督委員会に申し立てることによってその決定に異議を唱えることができます。</p> <p>詳しくは下記URLをご参照ください。</p> <p>https://oversightboard.com/news/833880990682078-the-oversight-board-is-now-accepting-cases/</p>
2-⑧ 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】</p> <p>【日本の数値】 回答なし</p> <p>【グローバルの数値】 回答なし</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ない。</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レビューすることがないようにしている。</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る。</p> <p>回答なし</p> <p>回答なし</p>	<p>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悪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。</p> <p>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悪用とみなされ、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>回答なし</p> <p>回答なし</p>

Meta		ヒアリングシート(2021年2月25日)回答		ヒアリングシート(2022年3月7日)回答
3-① 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】	サービス全ての合計、誹謗中傷以外を含む (i)一般ユーザ 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)		回答なし
3-② (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	【日本の数値】	(誹謗中傷以外を含む) (i)一般ユーザ 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)		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。 (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/
	【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)	(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。
(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	【日本の数値】			回答なし
	【グローバルの数値】			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
(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	【日本の数値】			回答なし
	【グローバルの数値】			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
(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】	回答なし		回答なし
	【グローバルの数値】	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】	「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。		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気づき、報告して下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。
3-③ (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】	回答なし		回答なし
	【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)	回答なし
(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】			回答なし
	【グローバルの数値】			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
(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】			回答なし
	【グローバルの数値】			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
3-④ 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)	【日本の取組】	(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。 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。 ・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、防止することが可能。 ・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。 ・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。 ・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。 ・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。 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。 「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。 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。		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。 -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。 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/131671940421729 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。 さらに、以下のような機能も備えています。 -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、防止する助けをする 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。 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://www.facebook.com/help/1182883832161405 -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取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。そのために、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なり、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が行う施策は、「利用者の直接的なフィードバックへの対応」、「ハ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かなくなるようにしています。
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】	回答なし		回答なし
	【グローバルの数値】	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
4-② 開示を行った件数 (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】	回答なし		回答なし
	【グローバルの数値】	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
5-① 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/		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/

Meta		ヒアリングシート(2021年2月25日)回答	ヒアリングシート(2022年3月7日)回答
5-②	誹謗中傷等の取組の効果分析	<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、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。</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た。</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>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>

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）回答における主な項目の比較【Twitter japan株式会社】

※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4-5

Twitter japan株式会社		ヒアリングシート(2021年2月25日)回答	ヒアリングシート(2022年3月7日)回答
1-①	対象サービス名	Twitter	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】 【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】 回答なし 回答なし 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)	回答なし 回答なし グローバルの収益化可能な日時平均アクティブユーザー数(mDAU):2億1700万(2021年第4四半期の平均)
1-④	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		Twitterは新しい方針について協議しています。2018年には、人間性を否定する表現に関する新しい規則について広く意見を求め、2週間で30か国以上から8,000以上の回答が寄せられました。これに基づいて2019年に、合成本または操作されたメディアに関する新しいルールのドラフトについて公開調査を実施し、世界中から6,500を超える回答を集めました。2021年の3月に、世界のリーダーのための政策の枠組みの開発に情報を提供するために、公開調査への回答を求めました。グローバルで49,000人近くが、世界のリーダーからのコンテンツを私たちのサービスでどのように処理すべきかについてのフィードバックを共有するために時間を費やしました。このような取り組みは、様々な文化および社会的文脈でルールがどのように適用および解釈されるかなど、オンラインスピーチの性質の変化に関するグローバルな視点を確実に検討するために不可欠です。これらのポリシーは、社会情勢に応じて断続的に見直し、改訂しています。 新しいポリシーや機能を開発する一方で、様々な社内のチームやTrust & Safety Councilからフィードバックを収集しています。2016年には、Twitter Trust and Safety Councilを設立しました。Trust & Safety Councilは、専門家やNPO等の団体を集めて、プロダクト、プログラム、およびポリシーの開発に役立つアドバイスを提供します。2019年には、このグループを拡大および再構築し、より多様な意見を取り入れ、メンバーを組織してより深い会話を行えるようになりました。Trust & Safety Councilには、東京自殺予防センターやLightHouseなどの日本のパートナーも加盟しています。
2-①	(i)一般ユーザーからの申告・削除要請への受付窓口・受付態勢	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す(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)。 Twitterでは人による調査とテクノロジーを組み合わせてTwitterルールの徹底に取り組んでいます。 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。 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。 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。 (参考: https://help.twitter.com/ja/rules-and-policies/twitter-report-violation , https://help.twitter.com/ja/safety-and-security/report-a-tweet)	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す(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)。 Twitterでは人による調査とテクノロジーを組み合わせてTwitterルールの徹底に取り組んでいます。 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。 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。 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。 https://help.twitter.com/ja/rules-and-policies/twitter-report-violation https://help.twitter.com/ja/safety-and-security/report-a-tweet
	(ii)主体的な誹謗中傷等の情報の検知態勢		Twitterでは、誰もがそれぞれの意見や考えを、安心して発信できると感じられるはずで、Twitterは表現の自由を尊重し、開かれた話し合いを行うべきだと考えています。ただし、そのような理念も、人々が恐怖に怯えているために何も言えないという状況では何の意味も成しません。 Twitterでの健全な対話を促進し、利用者が安心してさまざまな意見や信念を表現できるよう、Twitterは、嫌がらせや脅し、または他人を辱めたり、侮辱したりするといった行為を禁じています。人々の安全性にリスクをもたらすだけでなく、攻撃的な行為は、影響を受けた人たちに、物理的および感情的な苦境をもたらす可能性があります。 適用範囲 ● 個別のツイートをとり上げる攻撃的に見えても、会話全体の流れの中で見るとそうではない場合があります。Twitterがこの種の内容を調査する際、個人を攻撃するものなのか、合意の上での会話なのか明確でない場合があります。Twitterチームが会話の流れを理解し、執行措置を実施する前に必要な情報を確実に得るために、Twitterは標的となった人から直接話を聞かなければならない場合があります。 ● 個人またはグループを標的とする、ツイートまたはダイレクトメッセージによる、以下のいずれかの行為が見られるアカウントに対しては、検討のうえで執行措置を行います。プロフィール上で攻撃的な行為を行っているアカウントについては、攻撃的プロフィール情報を参照してください。人種、民族、国籍、性的指向、性別、性同一性、信仰している宗教、年齢、身体障害、または深刻な疾患を理由として他者を標的とする行為は、Twitterの暴言や脅迫、差別的言動に対するTwitterのポリシーに違反している可能性があります。 強烈な身体的脅迫 ● Twitterは、特定可能な標的に対する、強烈な身体的脅迫を伴うコンテンツを禁止しています。強烈な身体的脅迫とは、「殺してやる」などの言葉で、個人に重篤かつ持続的な身体的危害を加え、死に至らしめるか、重傷を負わせる意図の表明のことです。 ● 注記: Twitterのポリシーでは、いかなる強烈な身体的脅迫も容認されません。個人または特定の集団が深刻な損害を被ることを願う、希望する、または要求するなど、強烈な身体的脅迫を共有していると見なされたユーザーは、ただちにアカウントが永久凍結されます。 個人または特定の集団が深刻な損害を被ることを願う、希望する、または要求する ● Twitterは、個人または特定の集団に対して、死亡、深刻な身体的危害、重大な疾患を願う、希望する、促進する、扇動する、またはそのような願望を表す内容を禁止しています。これには以下の内容が該当しますが、限定されません。 ○ 深刻な疾患によって誰かが死亡することを願う。例:「がんになって死ねばいいのに」 ○ 誰かが深刻な事故に巻き込まれることを願う。例:「次しゃべったら、車にひかれてしまえ」 ○ 特定の集団が深刻な身体的危害を被るべきだと言う。例:「この抗議団体が黙らないのなら、撃たれるべきだね」 危害を望む発言に対するTwitterでの例外措置について ● Twitterは、重大な暴力について信頼できる形で告発された特定の個人に関する会話や、怒りとそれに伴って危害を望む発言を引き起こす可能性があることを認識しています。このような限られた場合には、Twitterは、アカウントのペナルティ、ストライク、永久凍結を行うことなく、ユーザーにツイートの削除を要求します。例としては以下のものも挙げられますが、これらに限定されません。 ○ 「すべての強姦犯は死ねばいいのに」 ○ 「児童虐待者は絞首刑にすべきです」 不適切な性的誘いかげ ● Twitterは、一部の合意の上での裸体や成年向け内容は許可していますが、求められていない性的な誘いかげや、合意なく個人を性的な対象として見る内容は禁止しています。これには以下の内容が該当しますが、限定されません。 ○ 画像、動画、GIFを含む、一方的な望まれていない成人向け内容を誰かに送る ○ 誰かの体に関する、望まれていない性的な話し合い、 ○ 性的な行為の要求、 ○ 合意なく個人を性的な対象とみなすその他の内容 他人を苦しめるまたはおびえさせる目的で侮辱、冒涇、中傷する ● Twitterは、他者を標的とした侮辱、冒涇、中傷などを使用することに対して措置を取ります。他者への嫌がらせまたは威圧を意図した侮辱、中傷(これらに限定されませんが)がひどく繰り返し使用されている場合には、Twitterはツイートの削除を要請することがあります。他者への嫌がらせまたは威圧を意図した侮辱や冒涇的な内容(これらに限定されませんが)が中程度の重大度であり、1回のみの場合には、Twitterは以下に示すようにツイートの表示を制限することがあります。個人によっては特定の言葉を攻撃的だと見なす場合がありますが、Twitterは侮辱的な言葉が使われたすべての事例に対応するわけではありません。 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 ● Twitterは、特定の個人または集団にいやがらせをする、または個人や集団を攻撃的な行為で標的にするよう、他人に呼びかける行為を禁止しています。オンラインで攻撃的な行為または嫌がらせ行為で他人を標的にするよう呼びかける、身体的嫌がらせなどのオフラインでの行動を取るよう催促する行為が含まれますが、これに限定されません。 多数の死傷者が発生した事件の否定 ● Twitterは、大量殺人など多数の死傷者を伴う事件について、その発生が確認できる場合、こうした事件の発生を否定するようなコンテンツを攻撃的な意図で共有することを禁止しています。これには、事件が「でっち上げ」である、または被害者が偽者または「俳優」であるとほめかす発言が含まれます。これには、ホロコーストや校内銃撃、テロ攻撃、自然災害などが該当しますが、これらに限定されません。 Twitterルールの違反について調査を依頼できるのは、こういったコンテンツの標的になった当事者だけですか？ ● いいえ、そのような内容の当事者と第三者、両方の報告を考慮します。 措置 ● 本ポリシーの違反に対する処分を判断する際に、Twitterは、その違反の重大さや違反者の過去の違反歴など、さまざまな要因を考慮します。以下は、本ポリシーに違反したコンテンツに対して適用される可能性のある強制的対応の一覧です。 ○ Twitterユーザーがツイート作成者をフォローしている場合を除き、返信ツイートの表示を規制します。 ○ 検索結果の上位やツイート作成者をフォローしていないユーザーのタイムライン上にツイートが拡散されないようにします。 ○ メールやおすめから当該ツイートおよびアカウントが除外されます。 ○ ツイートの削除を要請します。 ■ たとえば、Twitterは違反者に対して、違反しているコンテンツの削除を要請し、再びツイートできるようになるまで、一定期間読み取り専用モードで利用するよう求める場合があります。その後の違反については、読み取り専用モード期間が延長され、最終的には永久凍結される場合があります。 ○ このポリシーで定義されているように、攻撃的な行為に関与したり、強烈な身体的脅迫を共有したりすることが主な用途となっていると確認されたアカウントは凍結されます。 https://help.twitter.com/ja/rules-and-policies/abusive-behavior
2-②	申立時における申立者への通知の有無	有(日本語)	Twitterのヘルプセンターがアプリから報告の送信を完了すると、Twitterを快適に使うために推奨される対策が表示されます。 https://help.twitter.com/ja/rules-and-policies/twitter-report-violation 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もできます。
2-③	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		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。そうした対応を行うのは、Twitterルールに違反する行為があった場合です。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。 https://help.twitter.com/ja/rules-and-policies/enforcement-options

Twitter japan株式会社		ヒアリングシート(2021年2月25日)回答		ヒアリングシート(2022年3月7日)回答	
2-④	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	【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】	回答なし 回答なし 回答なし	回答なし 回答なし 回答なし	社内でのグローバルチームは、日本語を含むTwitterの全サポート言語において、年中無休でTwitterルールの強制的対応を管理しています。Twitterでは、Twitterルールを客観的に、そして一貫性を持って適用することを目指しています。Twitterルールに違反すると判断されるコンテンツには強制的な措置が取られます。 Twitterは適切なプロセスを提供し、公正で偏りがなく、バランスが取れ、人権を尊重する方法でTwitterルールが適用されるように力を注いでいます。その基盤となるのが、コンテンツモデレーションでの透明性と説明責任に関するサントクラ原則および複数関係者がかわるその他のプロセスの精神です。今後も、Twitterルールの適用方法に関して将来の報告書で提供できる情報範囲の拡大にリソースを投じていきます。 https://transparency.twitter.com/ja/reports/rules-enforcement.html#2021-jan-jun
2-⑤	(i) 対応決定時の申立てへの通知の有無とその理由説明内容 (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		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。	上記2、3をご確認ください。 Twitterは、アカウント情報開示請求について特定のアカウント所有者に通知します。ただし、通知を禁止されている場合、またはTwitterの「ユーザーへの報告に関するポリシー」の例外に該当する場合は除きます。 Source: https://transparency.twitter.com/ja/reports/information-requests.html#2021-jan-jun ツイートがTwitterルールに違反していると判断した場合、Twitterは違反者に違反ツイートの削除を要請します。違反者は削除するまで新しくツイートすることはできません。違反ツイートと、どのポリシーに違反しているかをメール通知で違反者に連絡します。違反者は、違反ツイートを削除する手順を実行するか、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必要があります。 https://help.twitter.com/forms/general?subtopic=suspended Twitterアカウントが凍結される理由の詳細: ● スпам、アカウントが凍結される原因のほとんどは、スパム行為または明らかな偽装であり、それらがTwitterやTwitterを利用しているすべてのアカウントにとってセキュリティ上のリスクを招いていると判断されることです。こうしたアカウントは、Twitterルールに違反します。本人のアカウントであるにもかかわらず、手違いで凍結されることもまれにあります。そのような場合は、本人の協力の下、Twitterはアカウントを凍結解除するための措置を講じます。 ● セキュリティが危険な状態にあるアカウント: アカウントが乗っ取られたりハッキングされたりした疑いがある場合、乗っ取りによる悪質な行為のリスクを軽減するために、アカウントのセキュリティが確保されて所有者の管理下に戻るまでの間、アカウントを凍結する場合があります。 ● 攻撃的なツイートや行動: 嫌がらせに関するTwitterルールに抵触しているという報告があった場合にアカウントを凍結することがあります。他人に対する脅迫や他のアカウントへのなりすましなど、嫌がらせに関与しているアカウントは、一時的に、または場合によっては恒久的に凍結することがあります。 注記: Twitterプロフィールにアカウントの凍結が表示されるだけでなく、アカウント凍結の理由の詳細が追加されることがあります。 https://help.twitter.com/ja/managing-your-account/suspended-twitter-accounts/why-might-Twitter-account-suspended	
2-⑥	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た場合、画面の指示で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。	アカウントが誤って凍結されたと思われる場合は、自分で凍結解除できる場合があります。 Twitterアカウントの凍結解除方法ログインしたときに、携帯電話番号の入力またはメールアドレスの確認を求められた場合、画面の指示でアカウントを凍結解除することができます。 アカウントがロックされているというメッセージが表示される場合は、特定の手順に従う必要があります。スパム行為や攻撃的な行動がみられるという報告を受けてアカウントを一時的に利用できない状態にすることがあります。その場合、一定期間ツイートできなくなったり、自分自身に関する特定の情報を認証するよう求められたりすることがあります。アカウントをロック解除する手順については、こちらをご覧ください。 異議申し立てをすることによって、アカウントを凍結解除できる場合があります。上記の手順でも自分ではアカウントを凍結解除できない場合があります。手違いでアカウントが凍結またはロックされていると考えられる場合には、異議申し立てをすることができます。まず、凍結されているアカウントにログインしてください。その後、ブラウザで新しいタブを開き、異議申し立ての詳細を入力します。 https://help.twitter.com/ja/managing-your-account/suspended-twitter-accounts	
2-⑧	不正な申告や削除要請への対策の方法・仕組み、対応件数	【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】	回答なし 回答なし 回答なし 回答なし	回答なし 回答なし 回答なし 回答なし	
3-①	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】	(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)	Twitterからのコンテンツ削除を求める法的要請や、世界各地の現地法に基づくその他のコンテンツ削除請求に関する分析情報です。 裁判所命令: 15 その他の法的要請: 18,503 合計: 18,528 *2021年1月1日から6月30日の期間: 誹謗中傷以外の削除要請の件数も含む https://transparency.twitter.com/ja/reports/removal-requests.html#2021-jan-jun	
		【グローバルの数値】	(i)一般ユーザ(2020年上半期) 暴力をほめめかす脅迫: 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) (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	(i)一般ユーザ(2020年上半期) 暴力をほめめかす脅迫: 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	
3-②	(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	【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) 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	回答なし	
		【グローバルの数値】	(i)一般ユーザ(2020年上半期) 暴力をほめめかす脅迫: 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	ルールに基づく強制的対応: グローバルで削除されたコンテンツの件数: 5,913,337 (2021年1月-6月、誹謗中傷以外の削除件数も含む)	
		【(質問に答えられない場合)参考となる数値】		ルールに基づく強制的対応(誹謗中傷関連、グローバル、2021年1月-6月): 攻撃的な行為・嫌がらせ: 1,043,525 児童の性的搾取: 456,146 ヘイト行為: 1,108,722 なりすまし: 216,846 合意のない裸体の描写: 29,635 個人情報: 30,714 センシティブな内容のメディア: 1,630,554 https://transparency.twitter.com/ja/reports/rules-enforcement.html#2021-jan-jun	
	(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	【日本の数値】		回答なし	
		【グローバルの数値】		2021年1月から6月: グローバルで対応されたアカウント数(何らかの強制的対応が行われたアカウント数): 4,826,539 2021年1月から6月: グローバルで凍結されたアカウント数: 1,240,148 * 誹謗中傷以外のものも含む https://transparency.twitter.com/ja/reports/rules-enforcement.html#2021-jan-jun	
	(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	【日本の数値】		回答なし	
		【グローバルの数値】		回答なし	
	(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】	回答なし	回答なし	
		【グローバルの数値】	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	
3-③	(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】	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。	回答なし 回答なし テクノロジーの活用による攻撃コンテンツの特定、違反の検出	
	(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】		回答なし 回答なし 回答なし	
	(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】		回答なし 回答なし 回答なし	

Twitter japan株式会社		ヒアリングシート(2021年2月25日)回答		ヒアリングシート(2022年3月7日)回答	
3-④	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)	【日本の取組】	<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。 ・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。 	<p>Twitterを安心してご利用していただくための機能:</p> <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。 ● 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。 <p>Twitterは、2021年第3四半期に、利用者の皆様により安心してTwitterをお使いいただくための機能を紹介する「デジタルセーフティーブック」の日本語版(全26ページ)を公開しました。ブックはどなたでも無料でダウンロードいただけます。プレイブックの発行は今回が初めてで、今後も更新してまいります。</p> <p>プレイブックの概要: プレイブックはTwitterをより安心して使っていただくための手引となるもので、「安心して使う」「安全を確保する」「自分のデジタルフットプリント(Twitter上のデータ)を管理する」の3つの目的ごとに、Twitterが提供している様々な安全機能を網羅的に紹介しています。</p> <ul style="list-style-type: none"> ● 安心して使う:自分のアカウントや投稿に不測の事態が迫った場合に取らうる手段として、「報告」「ブロック」「ミュート」「フォロワーの削除(ブロックせずに削除)」「セーフティーモード」の各機能の使い方や特長をまとめています。 ● 安全を確保する:タイムライン上の会話やメッセージに不快な内容があった場合などに、表示内容を設定できる「表示するコンテンツの管理」「My DM」「返信を非表示にする」「会話の設定(返信できる人を管理する)」の設定方法や使い方をまとめています。 ● 自分のデジタルプリントを管理する:Twitterの自分のデータを管理して自分自身を守るための「2要素認証」「ツイートの保護」「自分の公開情報のコントロール」などの各設定方法について紹介しています。 <p>プレイブックは日本語のほか、英語、フランス語、ヒンディー語、スペイン語など7言語で作成されています。 https://blog.twitter.com/ja_jp/topics/company/2021/playbook-for-safety 世界各地の政府、執行機関・捜査機関、第三者からのアカウント情報の開示を求める法的要請に関する情報です。 2,298 (2021年1月1日～6月30日、日本) https://transparency.twitter.com/ja/reports/information-requests.html#2021-jan-jun</p>	
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】	回答なし	【グローバルの数値】	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】	回答なし	【グローバルの数値】	回答なし
5-①	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を公表しています。 ・日本語も提供しています。最新版(2020年上半期)は2021年2月現在翻訳作業中。 		<p>Twitterは2012年に業界で最初に透明性に関するレポートを立ち上げたうちの1社となります。2012年7月2日以来、年2回発行する「Twitterの透明性に関するレポート」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、メールのプライバシーに関するベストプラクティスの状況を公表しています。 https://transparency.twitter.com/ja.html</p>
5-②	誹謗中傷等の取組の効果分析		<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しました。(2019年4月時点) 		<p>Twitterは、Twitterルールの適用方法についてより多くの背景情報と詳細を共有する方法を引き続き模索しています。その一環として、特定のツイートの削除が必要な強制的対応に関する新しい測定基準、インプレッションを導入しています。インプレッションは削除前にツイートが獲得した閲覧数を表します。</p> <p>Twitterでは2021年1月1日～6月30日に、Twitterルールに違反した470万件のツイートを削除しました。削除されたツイートのうち、削除前のインプレッションが100未満のものは全体の68%、100～1,000のツイートは24%をそれぞれ占めました。インプレッションが1,000を超える削除ツイートはわずか8%にとどまりました。今回の報告期間におけるルールに違反するツイートのインプレッションが全ツイートのインプレッションに占める割合は0.1%未満でした。</p> <p>前回報告書から顕著に変化した項目を以下にピックアップしています。</p> <ul style="list-style-type: none"> ● 対応したアカウント件数を前回報告期間と比較した増加率(グローバル): +36% ● 凍結されたアカウント数を前回報告期間と比較した増加率(グローバル): +23% ● 削除されたアカウント件数を前回報告期間と比較した増加率(グローバル): +32% <p>https://transparency.twitter.com/ja/reports/rules-enforcement.html#2021-jan-jun</p>