

フェイクニュースや偽情報への対策状況 ヒアリングシート（2022 年 3 月 28 日）

社名		Twitter Japan 株式会社
1. 前提（サービス概要）		
①	対象サービス名	Twitter
	サービス分類	ソーシャルメディア
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】
		【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】 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）
2. 「我が国における実態の把握」関係		
①	偽情報等の発生・拡散状況を把握できる体制 分析・調査の有無	有
②	日本における偽情報等の発生・拡散状況、結果公表	【①社会的混乱関係（災害等）】 災害の情報収集：Twitter は一番早く「今」起こっている災害状況を知ることのできる場であると同時に、必要な情報を瞬時に伝えることができる重要なコミュニケーションツールとなっています。 災害時に偽情報の拡散を防ぐために様々な取り組みを実施しています。 例えば、9 月 1 日の「防災の日」および 8 月 30 日～9 月 5 日の「防災週間」に際し、様々な取り組みを実施：Twitter ライフライン（@TwitterLifeline）の 47 都道府県ごとにまとまった

	リストや東京都防災 (@tokyo_bousai) の防災・災害時用リストを提供。お住まいの地区の災害情報・緊急情報をすぐに確認することが出来ます。また、直近では Yahoo! リアルタイム検索と連携し、Twitter に投稿された災害状況を「災害マップ」に表示する機能が導入された(1/11/2022) https://blog.twitter.com/ja_jp/topics/company/2021/2021-disaster-prevention
③	【②健康・医療関係（コロナ関係等）】 Twitter は日本でも COVID-19 に関するオープンな会話を保護する取り組みについて最新情報を公開しました。COVID-19 に関して誤解を招く非常に有害な情報を Twitter 上から削除する継続的な取り組みに加え、COVID-19 ワクチンに関する誤解を招く情報を含むツイートへのラベルを適用します。 また、Twitter は、さらなる措置の必要性を判断するためのストライク制の導入を進めています。このストライク制は、Twitter のポリシーに関する一般の皆さんの理解を促し、Twitter 上での有害または誤解を招く恐れのある情報の拡散（特に度重なる中程度および重度のルール違反）を抑制する効果が期待されます。 https://blog.twitter.com/ja_jp/topics/company/2020/covid-19 https://blog.twitter.com/ja_jp/topics/company/2021/updates-to-our-work-on-covid-19-vaccine-misinformation https://transparency.twitter.com/ja/reports/covid19.html#2021-jan-jun
④	【③選挙・政治関係】 選挙やその他の市民活動の操作や妨害を目的として Twitter のサービスを利用することを禁じます。これには、市民活動への参加を抑圧するようなコンテンツや、市民活動に関する日時、

	場所、方法について誤解を招くコンテンツの投稿や共有が含まれます。また、市民活動について誤った情報や誤解を招く情報が含まれるツイートには、背景情報を提供するために、ラベルを付けたり、拡散を抑制する場合があります。 Twitter で生まれる公共の会話は、選挙やその他の市民イベントの期間中は特に重要です。Twitter のサービスの清廉性を損なういかなる行為も、私たちの基本的権利の対極にあり、表現の自由の基本原則という Twitter が基盤としている価値を損なうものです。 Twitter は、妨害や操作を阻止し、公共の会話の清廉性を守る責任があると考えています。そのため、Twitter では、Twitter のサービスを使用して市民活動进行操作したり、妨害したりする行為を禁じています。これには、市民活動の手続きや環境に関する誤った情報や、誤解を招く情報による妨害行為が含まれます。誤解を招く情報によって、市民活動が直接的に操作されたり、妨害することを目的としていないものの、Twitter のサービスに混乱を引き起こしたりする場合は、背景情報を提供するためにツイートにラベルを付ける場合があります。選挙に関する重要な情報についての重大な混乱のリスクを考慮して、ツイートに風刺的またはユーモア溢れる要素が含まれている（または含まれようとしている）場合でも、これらの措置を取ることがあります。 https://help.twitter.com/ja/rules-and-policies/election-integrity-policy Twitter は、特に危機的な状況においてこそ、社会に対する責任を果たさなければなりません。私たちは Twitter ルールを積極的に適用し、Twitter へのアクセスを保護し、信頼性の高い情報を提供し、利用者のプライバシーと安全を確保し、開かれた会話の場を操作しようとする企みを阻止しています。 ウクライナ情勢をめぐるツイートに関しては、日本を含む全世界で様々な対策を行っています。
--	---

		情報の信頼性の向上 オンラインコンテンツを推奨し拡散する上で、アルゴリズムが重要な役割を担っていることを考慮し、私たちは「残すか消すか」の二択に囚われずにコンテンツの監視業務を拡大するべきと考えます。現在、次のようなアプローチをとっています。 • コンテンツがTwitter ルールに違反している場合、ツイートレベルまたはアカウントレベルで対処します（コンテンツの削除など）。 • ツイートにより即座に危害が生じるリスクは低いですが、文脈を明確にしなければ誤解が生じる恐れがある場合、当該ツイートをタイムラインに積極的に拡散せず、ツイートへのリーチを減らすことに注力します。コンテンツの拡散を防いで露出を減らし、ラベルを付与して重要な文脈を付け加えます。 以下の取り組みにより、ホームタイムラインでも検索画面でも信頼性の高い最新情報を簡単に見つけられるようにし、安全性を強化しています。 • モーメントで、ウクライナでの戦争に関するさまざまな言語のリアルタイムのニュースやリソースをキュレートして発信しており、これまでに 80 億 8000 万回以上閲覧されています。 • 英語、ウクライナ語、ロシア語で検索画面などにプロンプトを表示し、デジタル上の安全とセキュリティに関する重要なリソースを提供しています。 • 専用のイベントページを 12 か国語、67 か国以上で提供しており、さらに拡大する予定です。日本語はこちらです。 • ホームタイムラインで常に最新情報を入手するために役立つトピックと、信頼性の高い情報源をまとめたリストを作成しました。日本語はこちらです。 • ウクライナとロシアの利用者に対して、フォローしていない利用者の一部のツイートがおすすめとしてホームタイムラインに表示されないようにし、不健全なコンテンツが広
--	--	--

		がるリスクを低減しています。 さらに、トピック、リスト、スペースなど信頼性が重視される機能について、安全性を高め、悪用されるのを防ぐ対策を広範に行っています。 国家当局関係メディアに関する取り組み 戦争や紛争のさなかでは、事実に基づいた信頼性の高い情報にアクセスできるようにし、どのアカウントが当事国によってコントロールされているかを明確にすることが、極めて重要です。 2020 年以降、Twitter は一部の国の国家当局関係（政府系）メディアアカウントの推奨や拡散は行わず、「ロシア州関係メディア」などのラベル付けを行ってきました。先月にはこの取り組みを拡大し、ロシア政府系メディアのウェブサイトへのリンクを共有するツイートにもラベル付けを開始しました。これら政府系メディアからのコンテンツの多くは、Twitter がラベル付けしたアカウントだけでなく、こうしたリンクシェアからも発信されていたからです。 https://twitter.com/yoyoel/status/1498343849273425921?s=20&t=RSoSrTNRYSPI45MYq6cWUw 2 月 28 日以降、このカテゴリーにラベル付けされたツイートは、61,000 件以上に達しています。ラベル付けの基準に従い、そのようなツイートは拡散の対象外とされ、Twitter のホームタイムラインや通知などの画面でおすすめとして表示されません。これらの取り組みにより、この種のコンテンツのリーチは 30%減少しました。 さらにこの度、ベラルーシにある複数の政府系メディアアカウントとウクライナにある 1 つの政府系メディアアカウントへのラベル付けを開始しました。戦争が、そしてオンラインでの会話が進むなかで、利用者の皆様には文脈を正確に把握し、正しい情報に基づいて Twitter を利用していただきたいと思います。 利用者には、国家とメディアアカウントが結びついていることを知る権利があると Twitter は考えています。私たちは今後、既存の対象国に加え、国家間紛争に関わる国々の政府と政府系
--	--	---

	メディアにもラベルを付与します。この方針により、国際紛争に関する会話に公平に文脈が付け加えられるようになります。Twitter はこの取り組みを拡大し、徐々にラベル付けの対象国を増やしていく予定です。 また、先日発表された欧州連合（EU）による制裁措置により、Twitter は EU 加盟国内で一部の政府系メディアのコンテンツを差し控えることが法的に求められ、対応を進めています。EU 圏外でのグローバルな取り組みでは、引き続き Twitter 上でそのような政府系メディアのコンテンツの拡散を防止し、ラベル付けによって重要な文脈を明らかにすることに注力します。 政府アカウントラベルに関する取り組み Twitter は、ツイートの文脈を明確にするために、対象の政府アカウントにラベル付けを行っており、今週ウクライナ政府アカウントへのラベル付けを開始する予定です。 ポリシーに基づく強制的対応の拡大 世界的な大規模事案の際と同様に、Twitter の専門チームは潜在的なリスクについて積極的に監視しています。特に被害に遭いやすいのは、ジャーナリスト（特に女性ジャーナリスト）、活動家、政府機関、政府関係者などの知名度の高いアカウントです。Twitter は、標的を定めたアカウントの乗っ取りや操作による被害を軽減する取り組みを進めており、英語、ウクライナ語、ロシア語でアカウントのセキュリティに関するリソースを公開しました。また、Trust and Safety Council のメンバーを含む世界中のパートナーと協力して、支援を必要としているジャーナリストや活動家などからの緊急の支援要請を特定し、段階的に拡大して対応しています。 Twitter は引き続き、不正行為などのルール違反の審査に積極的に取り組みます。ウクライナでの戦争が始まってから、75,000 件以上のアカウントをプラットフォームの操作とスパムに関するポリシーへの違反のため削除しました。そのようなアカウントは、この機に乗じた金銭目当てのスパムなど、さまざまな形で Twitter を悪用しようとしています。現在のところ、そ
--	--

	れらが国家に関連する明確な組織的キャンペーンであるとは考えていません。 また、古い戦争の映像を現在ウクライナで起こっていることのように共有するなど、虚偽の、あるいは誤解を招く、もしくは不正確な文脈で情報を発信するコンテンツの量が大幅に増加しています。Twitter は合成または操作されたメディアに関するポリシーに基づいて、50,000 件以上のコンテンツに対してラベル付けや削除を実施しました。 Twitter チームは、ヘイト行為に関するポリシーと暴力の扇動を禁止するルールへの違反を積極的に監視しており、今後も状況の変化に応じて特定の国籍やグループの人を標的として攻撃するアカウントやツイートに対する措置を講じる予定です。また、攻撃的な行為に関するポリシー（特に多数の死傷者が発生した事件の否定ルール）に基づき、戦争中の暴力行為やその被害者について否定や誤解を招く主張を行った数多くのアカウントに対して強制的対応を講じました。 Twitter はこの取り組みにおいて透明性を重視しており、数か月以内に Twitter Moderation Research Consortium (TMRC) にこの一連の対応に関するデータを提供し、さらなる分析を行う予定です。 広告に関する取り組み 2月21日の週から、Twitter はウクライナとロシアでの広告を一時的に停止しました。重要な公共の安心・安全に関する情報がより伝わりやすく、広告が会話の妨げにならないようにしています。 戦争について議論したり取り上げたりするコンテンツや、Twitter ルールに基づいて虚偽または誤解を招くと見なされるコンテンツは、Twitter を通じた収益化の対象となりません。また、戦争に関する検索語句を収益化の対象外とし、特定の語句の検索結果ページに広告が表示されないようにしました。
--	---

		Twitter はこれまでも、広告に関する取り組みを行ってきました。2017 年以降、Twitter は（ロシア国営メディアである）ロシア・トゥデイ（RT）とスプートニクが所有する全アカウントからの広告を禁止し、デジタルリテラシーの向上、偽情報の防止、公正な選挙のための取り組みを行っている NGO に収益を寄付してきました。さらに、2019 年には、国が支援するニュースメディアの広告や政治広告を全面禁止しています。 現地コミュニティへの支援 Twitter は、現地で活動する人道支援団体と関わり、支援を今後も引き続き継続していきます。Twitter の従業員からの寄付に会社が上乘せし、ウクライナ難民を直接支援していると認められた団体、パートナー団体へ直接寄付しています。 また、これらの団体の重要な活動を支援し、彼らのメッセージを Twitter 上で広く届けるため、サービスや Ads for Good の助成金を提供します。Twitter 上では、これらのキャンペーンのインプレッション数は、すでに 117,463,000 回を超えています。 今後の予定 ウクライナ情勢が進展するなか、Twitter は潜在的な危害を減らし、信頼性の高い情報を提供する取り組みを継続していきます。Twitter チームからの最新情報は、@TwitterSupport、@TwitterSafety、@Policy でご確認ください。 この内容はロシア語とウクライナ語でも発信しています。 https://blog.twitter.com/ja_jp/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine
--	--	---

⑤		【④全般・その他】 Twitter における誤報への対処方法をまとめた新しいヘルプページをご用意しました。 https://help.twitter.com/ja/resources/addressing-misleading-info
⑥		(網羅的な回答が難しい場合、4. において回答するポリシー違反として偽情報を処理した具体的なケースについて)
⑦	研究者への情報提供、利用条件	無償の学術研究プロダクトトラックを提供しています。 新しい学術研究プロダクトトラックにより、許可された研究 は過去にリリースされたすべてのバージョン2のエンドポイントへのアクセスと、以下のことが可能 になります。 ・ フルアーカイブサーチエンドポイントを介したオープンな会話のすべての履歴への無償アクセス。この機能の利用は、これまで有償の Premium API または Enterprise API のカスタマーに限定されていました。 ・ Twitter 開発者プラットフォームへのより高レベルの無償アクセス。これには、月次のツイートボリュームの大幅なキャパシティ 1,000 万件 (Standard API で現在利用できる量の 20 倍) が含まれます。 ・ すべてのバージョン2 エンドポイントを対象とした、よりきめ細やかなフィルタリング機能。データ収集の対象を研究テーマに限定し、データクリーニング 要件を最小限に抑えます。 ・ 研究の成功を最大化する、新しい技術的および方法論的ガイド。 https://blog.twitter.com/developer/ja_jp/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api.html
3. 「多様なステークホルダーによる協力関係の構築」 関係		
①	産学官民の協力関係の構築	Twitter は様々なステークホルダーと協力しています。 Twitter は、一般社団法人セーファーインターネット協会が設立した Disinformation 対策フォ

		ーラムに加盟しています。 https://www.saferinternet.or.jp/anti-disinformation それ以外の産学官民の協力関係につきましては他の質問への回答をご参照ください。
②	具体的な役割	Twitter は上記 Disinformation 対策フォーラムに構成員として参加しています。 Twitter では、信頼できるパートナー（AP およびロイターとのパートナーシップ、その他の報道機関、公共保健機関、政府など）と密接に連携して当ポリシーを適用しています。Twitter のチームは、さまざまなパートナーとオープンな対話を行い、助言を求めたり、さまざまなメディアや主張を検証してもらったりしています。 https://help.twitter.com/ja/rules-and-policies/manipulated-media
③	議論を踏まえた取組	Twitter の具体的な取り組みについては他の質問への回答をご参照ください。
4. 「プラットフォーム事業者による適切な対応及び透明性・アカウントビリティの確保」関係		
(1) 偽情報等に関するポリシー		
①	(i) 禁止行為	【①社会的混乱関係（災害等）】 Twitter の目的は、公共の場における会話に寄与することです。Twitter では、すべての人にそれぞれの意見や考えを共有していただきたいと考えています。しかし、他の人が自らを表現することを阻止したり、危害が及ぶリスクにさらしたりするような行為も存在します。Twitter ルールは、すべての人が自由かつ安全に、公共の場における会話に参加できるように存在しています。禁止されているコンテンツや行為の種類を説明する、個別のポリシーが含まれています。 【②健康・医療関係（コロナ関係等）】

		Twitter のサービスを使って、健康被害をもたらす可能性のある、COVID-19 についての誤った情報や誤解を招く情報を共有することは禁止されています。 COVID-19 のパンデミックについての科学的な理解が深まり続けているにもかかわらず、根強い陰謀説、調査や信頼できる報告に基づかない人騒がせな言い回し、広範囲に及ぶ虚偽の談話や根拠のないうわさが出現しているため、背景情報が示されないと人々が自身の健康について情報に基づく決断を下せなくなっており、個人、家族、コミュニティが危険にさらされています。 明らかに誤ったコンテンツ、誤解を招くコンテンツ、重大な損害（ウイルスへの暴露を増やす、公衆衛生システムに悪影響もたらす、など）をもたらす可能性のあるコンテンツは Twitter 上で共有することが禁止されています。これには、COVID-19 ウイルスの性質、感染防止策、症状の緩和または治療の方法、またはその他の予防措置の効果性や安全性、健康についての勧告に伴う政府の規制、制限、除外規定、COVID-19 に関連するウイルスのまん延または感染や死亡のリスクについて、誤解を招く可能性のあるコンテンツを共有することが含まれます。 【③選挙・政治関係】 選挙やその他の市民活動の操作や妨害を目的として Twitter のサービスを利用することを禁じます。これには、市民活動への参加を抑圧するようなコンテンツや、市民活動に関する日時、場所、方法について誤解を招くコンテンツの投稿や共有が含まれます。また、市民活動について誤った情報や誤解を招く情報が含まれるツイートには、背景情報を提供するために、ラベルを付けたり、拡散を抑制する場合があります。 Twitter で生まれる公共の会話は、選挙やその他の市民イベントの期間中は特に重要です。Twitter のサービスの清廉性を損なういかなる行為も、私たちの基本的権利の対極にあり、表現の自由の基本原則という Twitter が基盤としている価値を損なうものです。
--	--	---

		Twitter は、妨害や操作を阻止し、公共の会話の清廉性を守る責任があると考えています。そのため、Twitter では、Twitter のサービスを使用して市民活動を操作したり、妨害したりする行為を禁じています。これには、市民活動の手続きや環境に関する誤った情報や、誤解を招く情報による妨害行為が含まれます。誤解を招く情報によって、市民活動が直接的に操作されたり、妨害することを目的としていないものの、Twitter のサービスに混乱を引き起こしたりする場合は、背景情報を提供するためにツイートにラベルを付ける場合があります。選挙に関する重要な情報についての重大な混乱のリスクを考慮して、ツイートに風刺的またはユーモア溢れる要素が含まれている（または含まれようとしている）場合でも、これらの措置を取ることがあります。 https://help.twitter.com/ja/rules-and-policies/election-integrity-policy
		【④全般・その他】 利用者を欺いたり、混乱させたりして、損害をもたらす可能性のある、合成または操作されたメディアや、文脈から切り離されたメディア（「誤解を招くメディア」）を共有することは禁止されています。さらに、Twitter では誤解を招くメディアが含まれるツイートにラベル付けを行い、利用者がツイートの信ぴょう性や文脈を把握するのを支援する場合があります。 https://help.twitter.com/ja/rules-and-policies/manipulated-media
	(ii) 削除等の対応	【①社会的混乱関係（災害等）】 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。Twitter が行う場合のある強制的対応の例を以下に示します。

	ツイートレベルの対応 ツイートレベルで対応する際は、これまでに違反がなく、間違えて Twitter ルールに違反したアカウントに対して厳しくなりすぎないようにしています。ツイートレベルで実施される対応をいくつか以下に紹介します。 真偽が問われている情報や誤解を招く情報を含む可能性のあるツイートへのラベルの適用：誤解を招く情報に関する現在の Twitter ポリシーは、合成または操作されたメディア、COVID-19、市民の清廉性に対応しています。ツイートに、Twitter ポリシーの定める誤解を招く情報や真偽が問われている情報が含まれており、実害につながりうると判断された場合、Twitter は背景情報や追加情報を提供するため、そのコンテンツにラベルを適用する場合があります。 ツイートの表示の制限：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たり、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たりして Twitter 上のコンテンツ表示を減らします。また、メールやおす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。やり取りの性質やコンテンツの種類に見られるさまざまなサインに基づいて、ツイートの表示が制限されます。 ツイートの削除を求める：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。 削除されるまで違反ツイートを非表示にする：Twitter がツイートの削除を要請してから実際に削除されるまでの間、ツイートが公開されないよう非表示にし、元のコンテンツを、Twitter
--	--

	ルールに違反しているためツイートを表示できないことを示す警告に置き換えます。この警告はツイートの削除後も 14 日間表示されます。 この警告が表示される場所： Twitter for iOS、Twitter for Android、twitter.com の違反ツイート詳細ページ Twitter for iOS、Twitter for Android、twitter.com の違反アカウントのプロフィールページ この警告が表示されない場所： ホームタイムライン、通知、検索など、プロフィールとツイートの詳細ページ以外の場所 公共の利益にかなう場合の例外措置： Twitter のルールに違反するツイートのうち、公共の利益にかなうものについては、例外的に Twitter 上でのアクセスが認められる場合があります。公共の利益にかなう場合の例外措置については、こちらをご覧ください。 こうした例外措置を取る場合、ツイートは非表示となり、これが例外であることを説明する警告が表示されます。この警告上でツイートを表示するかどうかを選択できます。 警告を表示する場合は、そのツイートの拡散を抑えるために、以下の措置を取ります。 返信、リツイート、いいねなどのエンゲージメントを無効にします。当該ツイートを話題にしたい場合、引用ツイートすることはできます。 ツイート上のエンゲージメント数を非表示にします（いいねや返信の数など）。ただし、以前
--	---

	のエンゲージメントを取り消すことはできます（いいねやリツイートの取消しなど）。以前の返信はツイートの詳細に表示されなくなります。 さらに、有害だと思われるコンテンツを最小限に抑えるために、警告付きのツイートは、Twitter の次の領域において表示されません。 「トップツイート」ホームタイムラインに表示されるツイート セーフサーチ プッシュ通知と [通知] タブに表示されるおすすめ メールとテキストによるおすすめ リアルタイムのイベントタイムライン [話題を検索] タブ ダイレクトメッセージレベルの対応 報告された違反者と報告者のアカウントの会話を停止する：ダイレクトメッセージによるプライベートな会話では、会話の相手の違反を報告した場合、Twitter は違反者が報告者にメッセージを送れなくします。また、報告者の受信箱から当該会話が削除されます。ただし、報告者が違反者にダイレクトメッセージを送ると、会話が再開されます。 ダイレクトメッセージを非表示にする：ダイレクトメッセージによるグループ会話では、違反しているダイレクトメッセージは、グループのメンバー全員が見られないように、非表示になることがあります。 アカウントレベルの対応 アカウントが Twitter ルールに著しく違反している、または通知後も違反を繰り返していると判断した場合、Twitter はアカウントレベルで対応します。
--	---

		メディアまたはプロフィールの編集を求める：アカウントのプロフィールまたはメディアコンテンツが Twitter ポリシーに準拠していない場合、Twitter は該当プロフィールまたはメディアを一時的に非公開にしたうえで、ポリシーに準拠するようにメディアまたはプロフィールの情報を編集するよう違反者に要請します。プロフィールまたはメディアコンテンツがどのポリシーに違反しているかについても説明します。 アカウントを読み取り専用モードに切り替える：これまでに違反のなかったアカウントが突然攻撃的な行為をしている場合は、該当アカウントを一時的に読み取り専用モードに切り替えて、落ち着くまでツイート、リツイート、いいねをできないように制限することがあります。読み取り専用モードでは、タイムラインを見ることがと、フォロワーにダイレクトメッセージを送ることができます。 他のアカウントは、読み取り専用モードのアカウントを見たり、会話したりできます。読み取り専用モードが継続する期間は、違反の内容に応じて 12 時間から 7 日間までの間で異なります。 アカウント所有権を認証する：違反者が Twitter の匿名性を悪用して、Twitter プラットフォーム上で他のアカウントに対して嫌がらせ行為をするのを防ぐため、電話番号またはメールアドレスによる所有権の認証をアカウント所有者に要請することがあります。これは、攻撃目的で複数のアカウントを利用している違反者を特定し、そうしたアカウントに対応するのに役立ちます。アカウントがロックされ、異議申し立てがまだ完了していない（たとえば、電話番号を認証する必要がある）場合は、電話番号を認証するまでそのアカウントはフォロワー数、リツイート、いいねにカウントされません。 永久凍結する：これは、Twitter による最も厳しい対応です。アカウントが永久凍結されるとグローバルで閲覧対象から削除され、違反者は新しいアカウントを作成できなくなります。Twitter はアカウントを永久凍結する場合、攻撃的な行為による違反が見られるためにアカウントを凍結した旨を通知し、どのポリシーに違反したか、どのコンテンツが違反したかについて
--	--	---

		て説明します。 違反者は、Twitter の判断が間違っていると思う場合は、永久凍結に対して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こちらから行うことができます。Twitter は異議申し立てを受けても凍結が妥当であると判断した場合、アカウントが違反しているポリシーについて説明することにより異議に対応します。 違反ではないコンテンツに対して行うことがある対応 ツイートを非表示にする：成人向けコンテンツや写実的な暴力描写など、センシティブな画像/動画を非表示にし、クリックするとそうした画像/動画が表示されることを利用者に忠告することがあります。この対応により、見たくない人がいるかもしれないセンシティブなコンテンツが含まれる可能性があることがわかります。センシティブなメディアの表示を管理する方法については、こちらをご覧ください。 年齢に応じてツイートの表示を制限する：成人向けコンテンツなどのセンシティブなメディアの特定のフォームを、18 歳未満の閲覧者またはプロフィールに生年月日を登録していない閲覧者に対しては表示しないようにします。成人向けコンテンツとは、合意のもとで製作され配信された、ポルノをはじめとする性的興奮の喚起を目的としたあらゆるメディアをいいます。年齢制限のあるコンテンツの場合、手前に警告メッセージが表示されます。これについては、こちらで確認できます。 特定の国でツイートまたはアカウントを表示制限する：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請求を受け取った場合、その国で特定のコンテンツを表示制限することがあります。また、コンテンツを表示制限した場合は、該当する商品内で表示制限されている旨を明記します。 https://help.twitter.com/ja/rules-and-policies/enforcement-options
--	--	--

		【②健康・医療関係（コロナ関係等）】 Twitter のサービスを使って、健康被害をもたらす可能性のある、COVID-19 についての誤った情報や誤解を招く情報を共有することは禁止されています。 COVID-19 のパンデミックについての科学的な理解が深まり続けているにもかかわらず、根強い陰謀説、調査や信頼できる報告に基づかない人騒がせな言い回し、広範囲に及ぶ虚偽の談話や根拠のないうわさが出現しているため、背景情報が示されないと人々が自身の健康について情報に基づく決断を下せなくなっており、個人、家族、コミュニティが危険にさらされています。 明らかに誤ったコンテンツ、誤解を招くコンテンツ、重大な損害（ウイルスへの暴露を増やす、公衆衛生システムに悪影響もたらす、など）をもたらす可能性のあるコンテンツは Twitter 上で共有することが禁止されています。これには、COVID-19 ウイルスの性質、感染防止策、症状の緩和または治療の方法、またはその他の予防措置の効果性や安全性、健康についての勧告に伴う政府の規制、制限、除外規定、COVID-19 に関連するウイルスのまん延または感染や死亡のリスクについて、誤解を招く可能性のあるコンテンツを共有することが含まれます。 このポリシーへの違反とはどのようなものですか？ COVID-19 に関連するコンテンツが、このポリシーに違反していると判断されるための条件は、次のとおりです。 • 明確な言い回しを使って事実であると主張している。 • 広範に入手可能な信頼できる情報源によると明らかに誤っているか、誤解を招いている。 • 公共の安全に影響を及ぼしたり、深刻な健康被害をもたらしたりする可能性が高い。
--	--	---

		ツイートの削除 Twitter は利用者に、このポリシーに違反し、深刻な健康被害をもたらしていることが判明したツイートを削除するように要請する場合があります。また、今後ツイートや情報共有ができないように、アカウントを一時的にロックすることもあります。こうしたツイートは、以下に説明する Twitter の「ストライクポリシー」に則り、2 警告として数えられます。Twitter が利用者に削除を要請するツイートの例は以下のとおりです。 • COVID-19 について、悪意のある勢力や強い権力による意図的な陰謀を想起させる、以下などの誤った情報 ◦ パンデミックはでっち上げである、またはパンデミックは人口を抑制する意図的な施策の一部である、あるいは 5G 無線通信技術が COVID-19 を引き起こしている。 ◦ COVID-19 は本物の病気ではない。 ◦ 予防接種は国際的監視網、人口抑制、または人口減少のための取り組みの一環として行われている。 ◦ （COVID-19 ワクチンに限らず、一般的に）ワクチンは危険であり、その副反応は政府や医療業界によって隠蔽されている。 ◦ 社会的弱者（妊婦、高齢者、または子供）が被験者にさせられている。 ◦ COVID-19 ワクチンは、接種した人に磁気反応を引き起こす。 ◦ 保健機関で承認されたワクチン（米国ではファイザー社のコミルナティワクチンなど）は、実際には完全な承認・認可を受けておらず、そのワクチンは未試験であり、「実験的」であり、何らかの安全性を欠いている。 • 特定の集団や人々（または人口統計学的に識別可能なその他の帰属性を持つ人々）の方が感染しやすい/しにくい、または重症化しやすい/しにくいといった、特定の集団への帰属性を根拠とした主張。 • 健康被害につながり得る、承認されていない治療法または予防法に関する虚偽の主張や
--	--	---

		誤解を招く主張（二酸化塩素やポビドンヨードを COVID-19 の予防や治療に用いることができるとする主張など）。 - 健康についての勧告に伴う政府の規制、制限、除外規定についての誤った情報や誤解を招く情報。 - ワクチン接種証明書（またはその他のデジタル記録物）など、いわゆる「隔離免除書」の偽造物の販売を奨励、宣伝、ほう助したり、その偽造方法を指南したりするあらゆる行為。 - PCR 検査ではウイルスを検出できないとする情報など、広く信頼性が認められている検査方法に関する虚偽の情報。 - ワクチンに致死成分や重篤な健康被害をもたらす成分が含まれていることを示唆する虚偽の主張。 - 所属の詐称 - 所属を偽ったアカウントや、開業医、公衆保健機関当局者、公衆衛生当局、調査機関への所属を偽って表明するコンテンツか、または COVID-19 の問題についての専門知識を偽って示唆するコンテンツを共有するアカウント。 ツイートのラベル COVID-19 に関する誤解を招く情報がツイートに含まれている場合、その主張に関する正しい情報が掲載されたラベルを当該ツイートに適用する場合があります。誤解を招く主張に関連する健康被害が起こり得ると判断された場合には、誤解を招く情報の拡散を防止するため、リツイート、引用ツイート、またはその他のエンゲージメント機能を他のアカウントが利用できないようにします。こうしたツイートは、以下に説明する Twitter の「ストライクポリシー」に則り、1 警告として数えられます。 また、（科学的、またはその他の）権威筋の見解が時間の経過とともに変わる可能性があるか、または変わりつつある場合や、地域の状況が重要となる場合、あるいは直接的または差し迫った健康被害が及ぶ可能性が低い場合には、背景情報を提供するためにラベルを適用することも
--	--	--

		あります。さらに、Twitter に直接投稿すると Twitter ポリシーへの違反と見なされる、第三者ウェブサイトのコンテンツへのリンクを掲載しているツイートについても、ラベルを適用する場合があります。こうした条件に該当するツイートには、信頼のおける権威筋の情報が記載されたラベルが適用されますが、以下に説明する Twitter の「ストライクポリシー」に則り、警告は与えられません。 Twitter がラベルを適用する場合があるツイートの例は以下のとおりです。 • 感染を防止するために実践できる予防策に関する誤った情報や誤解を招く情報（マスクが低酸素症や細菌性肺炎を引き起こすとする主張や、マスクには感染を抑制したり、COVID-19 を防御したりする効果はないとする主張など）。 • 承認されていない何らかの治療法が、COVID-19 の治療に効果があることを示唆する誤った情報や誤解を招く情報。 • 承認または認可された COVID-19 ワクチンの安全性や科学的根拠に関する、以下などの誤った情報や誤解を招く情報。 ◦ ワクチンのせいで病気になる、ワクチンによってウイルスが拡散される、またはワクチンは COVID-19 に感染するよりも有害な可能性がある。 ◦ COVID-19 ワクチンの成分や含有物について恐怖心を扇動したり、虚偽の情報を伝えたりするツイート。 ◦ mRNA（メッセンジャーRNA）ワクチンとその仕組みの特徴や科学的知識について誤った説明をしているツイート。 ◦ ワクチンが遺伝情報を書き換えると主張するツイート。 ◦ 公式な報告ツールや統計について、虚偽の情報を伝えたり、誤用したりしているツイート。 ◦ ワクチンを接種した人は、ワクチン未接種の人に対して、ウイルス（または症状や免疫）を拡散または伝播することができるとする、虚偽の主張や誤解を招く主張。
--	--	---

	- 保健機関の主張に反論するため、ワクチンの予防効果を不正確に伝える虚偽の情報または誤解を招く情報。疾病の重篤性、ウイルスのまん延、または有効性が広く認められている予防策、治療、ワクチンの効果に関する調査または統計的知見を不正確に伝える主張。 - ツイートにラベルが適用されると、通常は以下が実施されます。 - 当該ツイートへの警告メッセージの表示。 - 当該ツイートの共有やいいねを行おうとする利用者に対する、さらなる警告メッセージの表示。 - ツイートの拡散の抑制や、おすすりめからの除外。 - いいね、返信、リツイートの無効化。 - まとめられたランディングページや関連する Twitter のポリシーなど、補足説明のページへのリンクの表示。 ストライク制 COVID-19 について誤解を招く情報に関するポリシーへの違反に対する措置は、その違反の程度や種類、違反者の過去の違反歴によって決まります。このポリシーへの違反を繰り返すアカウントについては「ストライク制」を適用して、さらなる強制的対応を適用すべきかどうか判断します。こうした制度は、特に Twitter ルールへの違反の程度が深刻な場合、損害につながったり、誤解を招いたりするおそれのある情報の拡散を抑制するのに役立つと Twitter は考えています。 このポリシーへの違反が繰り返された場合、その違反によって課された累積ストライク数に応じた強制的対応が取られます。 1 ストライク：アカウントレベルでの対応なし 2 ストライク：12 時間にわたるアカウントのロック
--	---

	3 ストライク：12 時間にわたるアカウントのロック 4 ストライク：7 日にわたるアカウントのロック 5 ストライク以上：永久凍結 アカウントが誤ってロックまたは凍結されたと思われる場合は、異議申し立てを提出できます。 COVID-19 の誤報に関連するコンテンツの拡散の抑制 Twitter では、COVID-19 の誤報に関するポリシーに違反していると確信をもって判断されたツイートまたはアカウントについては、その拡散を抑制する場合があります。この枠組みに基づいたツイートやアカウントの表示制限は、一定の期間が過ぎると自動的に解除されますが、当該アカウントがポリシーへの違反を継続していると判断された場合、表示制限が手動または自動で再度適用される場合があります。Twitter では、以下の方法により拡散を抑制する場合があります。 • Twitter 商品の特定の場所（検索結果の上位など）に、当該アカウントによるツイートやリツイートが表示されないようにします • 当該アカウントによる返信を、会話の下部に表示します • メールやおすすめから、当該アカウントとそのツイートを除外します • 永久凍結 次の違反行為が見られるアカウントは、即時かつ永久に凍結する場合があります。 所属の詐称：当該アカウントが所属を偽っていると判断された場合や、開業医、公衆保健機関当局者、公衆衛生当局、調査機関への所属を偽って表明するコンテンツか、または COVID-19 の問題についての専門知識を偽って示唆するコンテンツを共有していると判断された場合。 違反行為の繰り返し：当該アカウントが、30 日間に COVID-19 の誤報に関するポリシーに繰り返し
--	---

	返し違反していると判断された場合や、COVID-19 に関する虚偽情報や誤解を招く情報をツイートすることを明確に目的として掲げ、用意されたものであると判断された場合。 https://help.twitter.com/ja/rules-and-policies/medical-misinformation-policy 【③選挙・政治関係】 市民の清廉性に関するポリシーの違反に対する措置は、その違反の程度や種類、違反者の過去の違反歴によって決まります。このポリシーへの違反を繰り返すアカウントについては「ストライク制」を適用して、さらなる強制的対応を適用すべきかどうか判断します。こうした制度は、特に Twitter ルールへの違反の程度が深刻な場合、実害につながったり、誤解を招いたりするおそれのある情報の拡散を抑制するのに役立つと Twitter は考えています。 Twitter が取る措置としては、次のものがあります。 ツイートの削除 (1) 選挙やその他の市民活動への参加方法について誤解を招いたり、(2) 参加の抑圧や脅迫を行ったりするなど、このポリシーへの違反の程度が深刻な場合、Twitter は該当するコンテンツを削除するように要請します。また、アカウントを一時的にロックしてツイートできないようにします。ツイートの削除は2 警告として数えられます。 プロフィールの修正 このポリシーの違反がプロフィールの情報（自己紹介など）に関するものである場合、該当するコンテンツを削除するように要請します。また、アカウントを一時的にロックしてツイートできないようにします。最初の警告を受けた後に再度このポリシーに違反した場合、アカウントは永久に凍結されます。 ラベル付け
--	--

	このポリシーに違反するコンテンツを削除しない場合は、Twitter 上に表示されているコンテンツを共有するツイートに背景情報を追加する場合があります。具体的には、以下の措置が取られる場合があります。 • Twitter 商品に表示される当該コンテンツにラベルや警告メッセージを適用する • 当該コンテンツの共有やいいねを行おうとするユーザーに警告を表示する • Twitter 利用者が当該ツイートに返信、リツイート、またはいいねを行う機能を無効にする • 当該ツイートの拡散を抑制したり、おすすりめから除外したりする • Twitter のモーメントや関連するポリシーなどに、補足説明のためのリンクを表示する • いいね、返信、リツイートを無効にする 多くの場合、ラベル適用の対象となるツイートに対し、上記すべての措置が取られます。さらに当該アカウントがツイートに返信、リツイート、またはいいねを行う機能を無効にする場合もあります。Twitter 上の誤解を招くコンテンツが大きな注目を集め、公衆に混乱が生じる場合には、優先的に Twitter モーメントを作成します。ツイートへのラベルの適用は1ストライクとしてカウントされます。 アカウントのロックと永久凍結 このポリシーへの違反が重大であったり、違反が繰り返されたりする場合、アカウントを永久に凍結します。 このポリシーへの違反が繰り返された場合、その違反によって課された累積ストライク数に応じた強制的対応が取られます。 • 1 ストライク：アカウントレベルでの対応なし • 2 ストライク：12 時間にわたるアカウントのロック • 3 ストライク：12 時間にわたるアカウントのロック • 4 ストライク：7 日間にわたるアカウントのロック • 5 ストライク以上：永久凍結
--	--

		アカウントが誤ってロックまたは凍結されたと思われる場合は、異議申し立てを提出できます。 https://help.twitter.com/ja/rules-and-policies/election-integrity-policy 【④全般・その他】 利用者を欺いたり、混乱させたりして、損害をもたらす可能性のある、合成または操作されたメディアや、文脈から切り離されたメディア（「誤解を招くメディア」）を共有することは禁止されています。さらに、Twitter では誤解を招くメディアが含まれるツイートにラベル付けを行い、利用者がツイートの信ぴょう性や文脈を把握するのを支援する場合があります。 Twitter では、信頼できるパートナー（AP およびロイターとのパートナーシップ、その他の報道機関、公共保健機関、政府など）と密接に連携して当ポリシーを適用しています。Twitter のチームは、さまざまなパートナーとオープンな対話を行い、助言を求めたり、さまざまなメディアや主張を検証してもらったりしています。 オーストラリア、韓国、米国において、利用者が誤解を招くと思われるツイートを報告できる新しい報告機能のテストを開始しました。この実験/検証の一環として、[問題を報告する]を選択すると、「It's misleading（誤解を招いています）」という文言がオプションに表示されるようになります。 Twitter の合成または操作されたメディアに関するポリシーの違反に対する処分は、その違反の重大さによって異なります。 ツイートの削除 誤解を招くメディアに、個人やコミュニティに損害をもたらす深刻な危険性がある場合など、
--	--	---

	当ポリシーへの違反の程度が深刻な場合、当該コンテンツの削除を要請します。 ラベル付け 当ポリシーに違反するコンテンツを削除しない場合は、Twitter 上に表示されている誤解を招くメディアを共有するツイートに背景情報を追加する場合があります。具体的には、以下の措置が取られる場合があります。 • 当該ツイートにラベルや警告メッセージを適用する • 当該ツイートの共有やいいねを行おうとするユーザーに警告を表示する • ツイートの拡散を抑制したり、おすすりめから除外したりする • いいね、返信、リツイートを無効にする • まとめられたランディングページ（Twitter モーメント）や関連する Twitter のポリシーなどに、補足説明のためのリンクを表示する 多くの場合、ラベル適用の対象となるツイートに対し、上記の措置のいくつかが組み合わされて執行されます。Twitter 上の誤解を招くコンテンツが大きな注目を集め、公衆に混乱が生じる場合には、優先的に Twitter モーメントを作成します。 アカウントのロック 特定のアカウントが、誤解を招く有害な説明を流布しており、それが Twitter の合成または操作されたメディアに関するポリシーに違反していると判断された場合、そのアカウントは一時的にロックまたは凍結される場合があります。 アカウントが誤ってロックまたは凍結されたと思われる場合は、異議申し立てを提出できます。 https://help.twitter.com/ja/rules-and-policies/manipulated-media
--	---

②	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	ポリシーの新規策定や変更を行う場合、オンラインでの言動の傾向を綿密に調べることで、許可される行為をわかりやすい明確な言葉で表現できるようにすること、膨大な数のツイートに適用できる調査担当者向けの対応ガイドラインを作成することが必要になります。 ポリシーの文言の草案をまとめる際は、社内のさまざまなチームや Trust & Safety Council から意見を募ります。この作業は、さまざまな文化や社会環境における Twitter ルールの適用や解釈の方法をはじめ、移り変わるオンラインでの発言の性質に対する世界各地の見解を考慮するうえで不可欠です。最後に、世界各地の調査チームのトレーニングを実施し、Twitter ルールを更新して、新しいポリシーの適用を開始します。 https://help.twitter.com/ja/rules-and-policies/enforcement-philosophy
(2) 削除要請や苦情に関する受付態勢・プロセス		
①	一般ユーザからの申告・ 削除要請への受付窓口・ 受付態勢、対応プロセス	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）。 https://help.twitter.com/ja/rules-and-policies/twitter-report-violation
②	対応決定時における通知 の内容、理由の記載の程 度	ユーザーが報告したアカウントに対して Twitter が何らかの対処を行った場合、そのユーザーに Twitter 内で通知されます。この対処は、そのユーザーが行った報告に関係していない場合もあります。 https://help.twitter.com/ja/safety-and-security/report-abusive-behavior
③	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】
		【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】

	その他の対応に関する部署やチームの内容・規模・人数	【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】 社内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タクララ原則および複数関係者がか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。
④	③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	社内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タクララ原則および複数関係者がかかわる

		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。
⑤	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	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
(3) 削除等の対応		
①	偽情報等に関する申告や削除要請の件数	【日本の数値】 2021 年 1 月～6 月 法的要請による削除要請件数は 18,518 ※誤情報に限らない
		【グローバルの数値】 2021 年 1 月～6 月 法的要請による削除要請件数は 43,387 ※誤情報に限らない
		【（質問に答えられない場合）参考となる数値】
②	(i) 偽情報等に関する申告や削除要請に対する削除件数	【日本の数値】
		【グローバルの数値】 2021 年 1 月～6 月 COVID-19 を巡る誤解を招く情報（対応したアカウント）：27,935

		市民の清廉性（対応したアカウント）：581 COVID-19 を巡る誤解を招く情報（削除されたコンテンツ）：33,761 市民の清廉性（削除されたコンテンツ）：593 https://transparency.twitter.com/en/reports/rules-enforcement.html#2021-jan-jun
		【（質問に答えられない場合）参考となる数値】
	(ii) アカウントの停止数	【日本の数値】
		【グローバルの数値】 <u>2021 年 1 月～6 月</u> COVID-19 を巡る誤解を招く情報（凍結されたアカウント）：617 市民の清廉性（凍結されたアカウント）：23 https://transparency.twitter.com/en/reports/rules-enforcement.html#2021-jan-jun
		【（質問に答えられない場合）参考となる数値】
③	偽情報等に関する主体的な削除件数（AI を用いた自動検知機能の活用等）	【日本の数値】
		【グローバルの数値】
		【（質問に答えられない場合）参考となる数値】

④	③についての削除の方法・仕組み（AI を用いた自動検知機能の活用等）	【日本の数値】 ポリシーに違反する行為を行っているすべてのアカウントに対しては、日本語を含む世界の言語で 24 時間 365 日、グローバルに対応する専任のスペシャリストチームが、ポリシーに則り、削除も含めたさまざまな強制的対応を用いて公正かつ公平にルールを執行しています。 チームは利用者からの報告を受けるだけでなく、攻撃的な行為をより効率的に検出するための技術的に優れたツールも活用しています。その結果、現在では、利用者からの報告がなくとも、プラットフォームに投稿される攻撃的なコンテンツの 65%以上が、スペシャルチームによる目視確認が必要として、システムに検知されるまでになっています。 ポリシー違反行為を検知するためのシステムへの投資は、継続的に行っています。また公正さや一貫性をもってルールを適用するため、担当チームが利用者からの報告を受けた際、十分な準備を整えた上で対応できるよう、長期間にわたる徹底したトレーニングプログラムを確立しています。
⑤	削除以外の取組 (i) 警告表示	コンテンツが Twitter のポリシーに違反しているかどうか、Twitter のチームが確認し、違反していると判断された場合にラベル提示が適用されます。この評価は自動検知ツールに共有され、Twitter 内で同様のコンテンツを特定しラベルを掲示する一連のプロセスをより能動的を行うために利用されます。 そして自動レビューと人間の目視によるレビューの両方を駆使して、COVID-19 ワクチンに関する誤った情報について定めた Twitter ルールに違反するコンテンツに対処することを最終目標としています。機械学習と自動言語処理の効果があらわれるまでには時間がかかります。そのため、まずは英語のコンテンツを対象とし、徐々に他の言語や文化的文脈に対して同じプロセスを適用していきます。

		ラベルは、設定されている表示言語で表示され、Twitter モーメントのガイドラインと方針や正式な公衆衛生または Twitter ルールにリンクされています。この機能は、COVID-19 に関する追加の背景情報や正式な情報を提供することを目的としています。 ストライク制の導入により、コンテンツがルール違反とされる理由について人々の理解が深まり、利用者が自らの行動とオープンな会話に与える影響について考える機会がもたらされることが期待しています。このストライク制は、先日更新された市民活動の阻害に関するポリシーに定められているストライク制と類似しているものです。ラベル適用またはツイート削除要求がアカウントレベルでの追加措置に発展した場合、利用者に直接通知されます。COVID-19 に関するポリシーへの違反が繰り返された場合、その違反によって課された累積ストライク数に応じた強制的対応が取られます。 1 ストライク： アカウントレベルでの対応なし 2 ストライク： 12 時間にわたるアカウントのロック 3 ストライク： 12 時間にわたるアカウントのロック 4 ストライク： 7 日間にわたるアカウントのロック 5 ストライク以上： 永久凍結 これまでと同様、アカウントが誤ってロックまたは凍結されたと思われる場合は、異議申し立てを提出できます。 https://blog.twitter.com/ja_jp/topics/company/2021/updates-to-our-work-on-covid-19-vaccine-misinformation
	(ii) 表示順位の低下	ツイートの拡散を抑制したり、おすすりめから除外したりする。

	(iii) その他の取組内容	いいね、返信、リツイートが無効にする、まとめられたランディングページ（Twitter モーメント）や関連する Twitter のポリシーなどに、補足説明のためのリンクを表示する。
⑥	不正な申告や削除要請への対策の方法・仕組み、対応件数	【日本の数値】
		【グローバルの数値】
		【（質問に答えられない場合）参考となる数値】
（４）透明性・アカウントビリティの確保		
①	コンテンツモデレーションのアルゴリズムに関する透明性・アカウントビリティ確保方策	社内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タクラ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。
	AI 原則・ガイドライン等の参照	
②	透明性レポート 日本語で閲覧可能か	日本語の透明性レポートはこちらをご参照ください。 https://transparency.twitter.com/ja.html

③	取組の効果分析	世界的な大規模事案の際と同様に、Twitter の専門チームは潜在的なリスクについて積極的に監視しています。特に被害に遭いやすいのは、ジャーナリスト（特に女性ジャーナリスト）、活動家、政府機関、政府関係者などの知名度の高いアカウントです。Twitter は、標的を定めたアカウントの乗っ取りや操作による被害を軽減する取り組みを進めており、英語、ウクライナ語、ロシア語でアカウントのセキュリティに関するリソースを公開しました。また、Trust and Safety Council のメンバーを含む世界中のパートナーと協力して、支援を必要としているジャーナリストや活動家などからの緊急の支援要請を特定し、段階的に拡大して対応しています。 Trust & Safety Council https://about.twitter.com/en/our-priorities/healthy-conversations/trust-and-safety-council また、Twitter は昨年、プラットフォーム上の誤情報対策として、誤解を招く可能性のあるツイートにほかの利用者が意見や情報ソースを追加できる新機能「Birdwatch」を米国で試験的にリリースしました。日本では現時点で導入されていませんが、アメリカでの実証実験の結果と機能の拡充について、日本の皆様にもいち早くお知らせします。 Birdwatch は誤解を招くと思われるツイートに対して、人々が共同で文脈を追加していく新しい方法です。Twitter は、人々が共同で作業することでさまざまな視点を持つ人々が、役に立つ背景情報をもたらす文脈を追加できると考えています。 今回、Birdwatch への参加者や一般の人々、および有識者からのフィードバックにもとづいて多くの点を改善し、期待どおりの成果が出ました。米国で Twitter を利用している人々を無作為に抽出した調査では、次のような知見が得られました。
---	---------	--

		• 大多数の人々が、Birdwatch ノート（具体的には、Birdwatch の参加者が「役に立つ」と指定したノート）は実際に役に立つと答えました。これには、さまざまな政治的立場を持つ人々が含まれています。 • 調査対象となった人々のうち、誤解を招く可能性のあるツイートに関するノートを読んだ人々は、ノートを読まずにツイートだけを読んだ人々に比べて、その内容に同意する割合が 20% から 40% 低くなっていました。 • これは、Birdwatch が Twitter 利用者により役に立つ正確な情報をもたらしていることを示す、明るい材料です。この結果をもって、より多くの Twitter 利用者からフィードバックを集め、さらなる改善ができるようテストを拡大します。 https://blog.twitter.com/ja_jp/topics/product/2022/building-a-better-birdwatch_2022
＜ 5. 「利用者情報を活用した情報配信への対応」 関係 ＞		
①	広告表示先の制限	COVID-19（またはその派生語）に関連する不快な表現は禁じられています。 センセーショナルまたはパニックを引き起こしかねないコンテンツは禁じられています。 COVID-19 に関連した商品の値上げを誇張して伝えることは禁じられています。 COVID-19 関連のニュースは、政治に関する広告コンテンツのポリシーが適用除外されているメディアパブリッシャーからのみ許可されます。こうした適用除外に興味があり、なおかつその条件を満たしている方は、AM/AE または、Twitter にお問い合わせください。 除菌用ローションの広告は許可されるようになりますが、クリエイティブ内で COVID-19 について言及することはできません。 https://business.twitter.com/ja/help/ads-policies/ads-content-policies/inappropriate-content.html

②	広告出稿制限	Twitter では、政治に関するコンテンツを広告に掲載することを全世界で禁止しています。この決定は、政治に関するメッセージのリーチは購入するものではなく獲得するものであるという信念に基づいています。 詳しくはこちらをご参照ください。 https://business.twitter.com/ja/help/ads-policies/ads-content-policies/political-content.html Twitter では、不適切なコンテンツを含む広告を全世界で禁止しています。Twitter 上のすべてのコンテンツは Twitter ルールに沿ったものでなければなりません、Twitter では広告コンテンツに別途制限を設けています。 詳しくはこちらをご参照ください。 https://business.twitter.com/ja/help/ads-policies/ads-content-policies/inappropriate-content.html
③	ターゲティング技術の適用に関する規定	Twitter 広告には、広告主様が自社コンテンツに興味関心を持つお客様にリーチしやすいよう幅広いターゲティングオプションが用意されています。 詳しくはこちらをご参照ください。 https://business.twitter.com/ja/help/ads-policies/campaign-considerations/targeting-of-sensitive-categories.html Twitter はユーザーのプライバシーを尊重しているため、ユーザーはサポートされている端末で、Twitter の興味関心ターゲティングによる広告サービスをオプトアウトできます。 https://business.twitter.com/ja/help/ads-policies/product-policies/interest-based-opt-out-policy.html

④	広告のアルゴリズムに関する透明性・アカウント ビリティ確保方策	
	AI 原則・ガイドライン等 の参照	
⑤	出稿者の情報や資金源の 公開	連邦選挙や総選挙に関する広告は、現在は禁止されています。 https://business.twitter.com/ja/help/ads-policies/product-policies/ads-transparency.html
⑥	広告とコンテンツの分離	広告主が有料で Twitter に投稿したことがわかるよう、“プロモ広告”には「プロモーション」というラベルが明示されます。 https://business.twitter.com/ja/help/overview/what-are-promoted-ads.html
⑦	その他の透明性・アカウン タビリティ確保方策 ユーザへのツール提供	Twitter は、広告パートナーと連携することで、興味関心に基づいた広告配信を行い、ユーザーの役に立ち興味を引くと思われる広告を表示しています。たとえば、広告パートナーから提供される情報を利用して広告をカスタマイズすることがあります。これにより、Twitter 広告がユーザーにとってさらに役立つものになるよう努めています。 詳しくはこちらをご参照ください。 Twitter による Twitter 内外での興味関心に基づいた広告表示を希望しない場合は、いくつかの方法でこの機能をオフにできます。 Twitter の設定で [カスタマイズとデータ] 設定に移動して、[広告をカスタマイズ] 設定を変更します。

		- twitter.com を利用している場合、ブラウザで optout.aboutads.info を開き、Digital Advertising Alliance による消費者選択ツールを使って、Twitter からの興味関心に基づいた広告配信を停止します。 https://help.twitter.com/ja/safety-and-security/privacy-controls-for-tailored-ads Twitter では、Twitter 広告に関連する報告のためのフォームをご用意しています。また、問題をすぐに解決できる手段もあります。 - ツイートを非表示にする：今後表示したくないプロモツイートが表示される場合は、そのツイートを非表示にするとタイムラインからすぐに削除されます。この操作により、その広告主に関するフィードバックが Twitter に提供され、以後その特定のプロモツイートはタイムラインに表示されなくなります。プロモツイートを非表示にする方法に関する詳細はこちらをご覧ください。 - 広告主に連絡する：プロモツイートもツイートであることに変わりはないため、他のツイートと同様に、相手とのやり取りに使うことができます。広告主に返信したり、広告主があなたをフォローしている場合にはダイレクトメッセージを送信したりして、広告主に問題への注意を喚起することができます。 - 広告主をブロックする：広告を配信しているアカウントをブロックすると、以後そのプロモツイートはタイムラインに表示されなくなります。またこの操作により、そのアカウントはおすすめのプロモアカウントとして表示されなくなります。
6. 「ファクトチェックの推進」関係		
①	ファクトチェック結果の表示 具体的な仕組み・基準	Twitter では、信頼できるパートナー（AP およびロイターとのパートナーシップ、その他の報道機関、公共保健機関、政府など）と密接に連携して当ポリシーを適用しています。Twitter

		のチームは、さまざまなパートナーとオープンな対話を行い、助言を求めたり、さまざまなメディアや主張を検証してもらったりしています。 オーストラリア、韓国、米国において、利用者が誤解を招くと思われるツイートを報告できる新しい報告機能のテストを開始しました。この実験/検証の一環として、[問題を報告する]を選択すると、「It's misleading（誤解を招いています）」という文言がオプションに表示されるようになります。 https://help.twitter.com/ja/rules-and-policies/manipulated-media
②	ファクトチェックを容易にするツールの開発及び提供	Twitter は昨年、プラットフォーム上の誤情報対策として、誤解を招く可能性のあるツイートにほかの利用者が意見や情報ソースを追加できる新機能「Birdwatch」を米国で試験的にリリースしました。日本では現時点で導入されていませんが、アメリカでの実証実験の結果と機能の拡充について、日本の皆様にもいち早くお知らせします。 Birdwatch は誤解を招くと思われるツイートに対して、人々が共同で文脈を追加していく新しい方法です。Twitter は、人々が共同で作業することでさまざまな視点を持つ人々が、役に立つ背景情報をもたらす文脈を追加できると考えています。 今回、Birdwatch への参加者や一般の人々、および有識者からのフィードバックにもとづいて多くの点を改善し、期待どおりの成果が出ました。米国で Twitter を利用している人々を無作為に抽出した調査では、次のような知見が得られました。

		• 大多数の人々が、Birdwatch ノート（具体的には、Birdwatch の参加者が「役に立つ」と指定したノート）は実際に役に立つと答えました。これには、さまざまな政治的立場を持つ人々が含まれています。 • 調査対象となった人々のうち、誤解を招く可能性のあるツイートに関するノートを読んだ人々は、ノートを読まずにツイートだけを読んだ人々に比べて、その内容に同意する割合が 20% から 40% 低くなっていました。 • これは、Birdwatch が Twitter 利用者により役に立つ正確な情報をもたらしていることを示す、明るい材料です。この結果をもって、より多くの Twitter 利用者からフィードバックを集め、さらなる改善ができるようテストを拡大します。 https://blog.twitter.com/ja_jp/topics/product/2022/building-a-better-birdwatch_2022
③	ファクトチェックを実施する人材の育成	3 月 3 日（米国時間）より、米国の Twitter 利用者の小規模（かつ無作為に抽出された）グループが、一部のツイートについての Birdwatch ノートを直接閲覧できるようになりました。これらのユーザーはノート进行评估することもでき、Birdwatch がさまざまな立場の人に役立つ文脈を追加する能力を向上させるための情報を提供します。 ツイート上にノートを表示させるには、まず異なる視点を持つ Birdwatch の参加者によって十分な数の「役に立つ」という評価を得る必要があります。参加者の視点の違いは、属性ではなく過去にどのようにノート进行评估してきたかによって判断されます。評価法についての詳しい情報はこちらをご覧ください。 https://blog.twitter.com/ja_jp/topics/product/2022/building-a-better-birdwatch_2022
④	日本におけるファクトチェック推進に向けた取組	Twitter は一般社団法人セーファーインターネット協会の Disinformation 対策フォーラムに加盟しています。

		https://www.saferinternet.or.jp/anti-disinformation また、今後 Birdwatch 機能も日本で導入する予定です。
⑤	ファクトチェック機関との連携	1 年前に Birdwatch をスタートさせたとき、Twitter は Birdwatch を公開し、Twitter の利用者、Birdwatch の参加者、アカデミックアドバイザーおよび研究者からインプットを得ようと思いました。私たちは、Birdwatch に投稿されたすべてのデータを一般公開したほか、ノートの評価に使用されたデータのコードも公開し、現在もこれを続けています。 Twitter コミュニティからのフィードバックとインサイトにより、Birdwatch は以下のように改善されました。 多様な人々から「役に立つ」と評価されたノートを識別する機能を強化しました。Birdwatch の参加者を保護し、人々がより快適に投稿できるようにするために、パイロット版では全員に自動生成されたエイリアスで投稿する方法を導入しました。 参加者が自分の Birdwatch への投稿がどのような影響を与えているのかを把握できるようにするために、自分のノートが「役に立つ」と評価されているかどうか、あるいは自分が評価したノートが他の人に閲覧されているかどうかを通知するようにしました。 アプリ内の通知機能を通じて、ノート内に情報ソースと明確な説明を記述することを参加者に推奨するようにしました。 多様なフィードバックを必要とするノートをより簡単に見つけられるようにするために、テストの参加者向けにパーソナライズされた「Needs Your Help」タブを追加しました。 Twitter は先頃、Birdwatch の参加者が閲覧している情報の質をチームでも評価できるようにするために、AP 通信およびロイター通信との連携を発表しました。さらに、誤情報、分断、ネットの悪用およびハラスメントを専門とする学者および研究者で構成されたアドバイザリー

	ボードを立ち上げました。アドバイザーは、MIT、ワシントン大学、ミシガン大学情報学部（UMSI）などの団体に所属しています。 「オンラインで誤情報の拡散が続いており、偽りの情報、あるいはミスリードを目的とした情報にしっかりと対処するには、多方面からのアプローチが必要となります。Birdwatch アドバイザリーボードのメンバーとして、役に立つ背景情報を明示するランキングシステムといったコミュニティベースのツールを活用してツイートに背景情報を追加するという、Birdwatch 独自のアプローチを高く評価しており、このコミュニティが成長するのを楽しみにしています」 - エイミー・チャン（Amy Zhang）（ワシントン大学教授） Birdwatch が今日あるのは、パイロット版へ参加者した一万人のみなさんのおかげです。Birdwatch 参加者は、ノートの執筆および評価に時間と労力を費やしてくれただけでなく、Birdwatch 体験についての貴重なフィードバックを一貫して提供し、さらに Birdwatch の働きや改善点について、思慮深く率直なフィードバックも共有してくれました。 Twitter は今後も学び、反復し、オープンにしながらテストを拡充し続けます。最新のアップデートについての情報、ならびに当社の活動についてのフィードバックは、@Birdwatch をフォローしてください。 https://blog.twitter.com/ja_jp/topics/product/2022/building-a-better-birdwatch_2022 Twitter は、ルールの執行だけでなく、専門家と協力し、利用者が信頼できる医療情報を取得できるようにすることを重視しています。そのため、地方、国、世界の公共医療機関と世界中で緊密に協議しながら、取り組みを進めています。直近では、以下のような取り組みを行いました。
--	---

		• 信頼できる情報の取得：2020 年 1 月以降、Twitter には COVID-19 専用の検索表示機能が導入されています。これにより、COVID-19 に関する情報を検索すると、信頼できる正式なコンテンツが検索結果の上位に表示されます。この機能は世界 80 か国以上に展開され、現在では 29 の言語で利用できます。また、一部の国では、これに COVID-19 のワクチンに関する情報にリンクするボタンが追加されています。 • 無料広告：Twitter は、「Ads for Good」プログラムを通じて、世界中の非営利団体や NGO と継続的に協力し、COVID-19 に関する正式な情報を積極的に提供しています。コロナ禍において、Twitter は 35 か国以上でこの取り組みで無償支援してきました。さらに、プロモトレンドやファーストビューなどのプレミアム広告を寄付し、ワクチン接種や吹雪時の緊急支援場所に関するアメリカ合衆国連邦緊急事態管理庁 @FEMA のメッセージなど、重要な公共医療情報を積極的に伝えています。 • 国際協力：2021 年 2 月、WHO のポリシーに関する国際協議に参加し、「ソーシャルメディアセクターにおいて誤った医療情報に対処するための革新的な方法とベストプラクティス」の導入について意見を述べました。また、Twitter は今後も、#AskWHO で WHO のウィークリーQ&A ライブのイベントページをホストします。 • Twitter 上での一般の質問への対応：今月、Twitter は@WHCOVIDResponse の Q&A イベントを開催しました。この Q&A には、バイデン大統領の首席医務官である Anthony Fauci 博士をはじめ、ホワイトハウスの COVID-19 対策チームのメンバーが出演しました。インドでは、保健省と協力し、Twitter 上でワクチン専門家が一般の質問に回答するウィークリートークイベントである Vaccine Vartha を企画しました。 • 健全な会話の促進：1 月、Team Halo、ユニセフ、NHS、Vaccine Confidence Project と協力して、絵文字ハッシュタグ「#ワクチン完了」「#vaccinated」の利用を開始し、ワクチン接種への支持を表明しました。これは、# #StayHome、#手を洗おう #WashHands、#マスクをしよう #WearAMask によって人々を励ましてきた活動が発展したものです。
--	--	--

		世界中の医療機関が COVID-19 やワクチン接種プログラムに対する理解を深めるなか、Twitter は引き続き最新の正式な情報を広めていきます。この重要な時期に、全社一丸となってオープンな会話を守るという Twitter の役割を果たし、Twitter の取り組みに関する最新情報を随時お伝えしていきます。 https://blog.twitter.com/ja_jp/topics/company/2021/updates-to-our-work-on-covid-19-vaccine-misinformation
7. 「ICT リテラシー向上の推進」関係		
①	普及啓発の取組・投資	Twitter は常にインターネット上の情報流通と表現の自由を支援してきました。その一環として、メディア情報リテラシーを巡る議論にも積極的に取り組んでいます。 私たちは Twitter のサービスを利用する人たちが、真偽を見抜く目を持ってコンテンツに関われるようにしたいと考えており、10 月 24 日から @UNESCO の「Global Media and Information Literacy Week 2019（グローバルメディア情報リテラシーウィーク 2019）」が始まるにあたって、教育者向けの新しいハンドブック「Twitter を活用した教育と学習」の提供を開始しました。 この教育者向けのハンドブックは読みやすく、有益な内容になっており、若い世代にメディアリテラシーのスキルを身につけてもらい、インターネット上の様々なコンテンツに対して的確な疑問を持ち、Twitter 上のニュースや情報に対しても適切な分析ができるスキルを備え、活用してもらうことを狙いとしています。 ハンドブックには UNESCO のメディアリテラシーに関するベストプラクティスのガイドラインが含まれているほか、UNESCO のプログラム・スペシャリストによって選定された文献リスト

	も掲載されており、このトピックに関する最新の教材を通じて、教育者にとっての指針になるよう考えられています。 世界各国に幅広く行き渡るよう、このハンドブックはまず、英語、日本語、フランス語、スペイン語、ドイツ語、ポルトガル語、アラビア語、スウェーデン語、ヒンズー語の9カ国語に翻訳され、将来的に他の言語にも翻訳される予定です。 また、Twitter は UNESCO 関連の NGO ネットワークや各国の政府教育機関、欧州ネット安全センター、日本では国連広報センター（UNIC）やアジア太平洋メディア情報リテラシー教育センター（AMILEC）など、世界中にある数多くのパートナーと連携して、学校教材としても配布して行きます。 この取り組みは、誤った情報に関する Twitter ポリシーの策定プロセスと密接に関連しており、合成もしくは悪意を持って操作された画像や映像メディアに対抗するために新たに制定するポリシーに関し、利用者がフィードバックできるパブリックコメントの募集をする予定です。 デジタルクラスルーム ハンドブックには学校や家庭で、学習ツールとして Twitter の利点を最大限に活用したいと考える教育者や保護者のためのヒントも含まれており、考え方や学習プラン、活用事例など役に立つ情報が入っています。 良いデジタルシチズンになることの意味を総合的な観点からとらえ、ハンドブックではオンライン上の安全性にも触れており、教育者としてネットいじめにどのように対処し、ご自身のデジタルフットプリントをどのように管理すればよいかも学べるようになっています。ハンドブックの詳細についてはこちらをご覧ください。
--	--

		UNESCO と Twitter のパートナーシップ この教育者向けのハンドブックは Twitter と UNESCO が結んでいるメディア情報リテラシーに関する 18 カ月間のグローバルパートナーシップによる成果です。この期間、Twitter は Ads For Good プログラムを通じてメディアリテラシーに取り組む NGO を支援すると同時に、Twitter @Nonprofits など UNESCO の @MILClicks を支援します。 Twitter のその他の取り組み • #世界メディア情報リテラシーウィーク、#シェアする前に考えよう、#クリックする前に考えよう の 3 つのハッシュタグで期間限定の絵文字を作成。 • 全米メディアリテラシー教育協会（National Association for Media Literacy Education、NAMLE）とパートナーシップを組み、ハッシュタグ#MediaLitWk を使って米国におけるメディアリテラシーウィークを支援 • @UNESCO と関連するの NGO に無償広告支援プログラム「Ads For Good」を提供して、メディアリテラシーウィークの認知度を向上。 https://blog.twitter.com/ja_jp/topics/company/2019/twitter-launches-new-media-literacy-handbook-for-schools
②	他のステークホルダーとの連携・協力・投資	Twitter は様々なステークホルダーと連携しています。事例につきましては、他の質問への回答をご参照ください。
8. 「研究開発の推進」関係		
①	AI 技術に関する研究開発	

②	「ディープフェイク」対策の研究開発	Adobe, Twitter, NYT がディープフェイクの対策の取り組みを開始した。誰が画像や動画を作成し、どのような変更が加えられたかを明確にするための取り組み。 https://blog.adobe.com/en/publish/2019/11/04/content-authenticity-initiative#gs.v2xzs.j
9. 「情報発信者側における信頼性確保方策の検討」関係		
①	信頼性の高い情報の表示	利用者が適切な情報に出会えるよう、2020 年 1 月以降、Twitter には COVID-19 専用の検索表示機能が導入されています。「#ThereIsHelp」とよばれる機能で、検索結果の最上位に各国の公的保健機関へのリンクを掲載するという機能「#ThereIsHelp」を実装しています。（日本は厚労省のサイトへのリンク）。この機能は世界 80 か国以上に展開され、現在では 29 の言語で利用できます。また、一部の国では、これに COVID-19 のワクチンに関する情報にリンクするボタンが追加されています。
②	ニュースの選別・編集に関する透明性・アカウントビリティ確保方策	Twitter では、キュレーションの正確性、公平性、公正性について、高い基準を維持することを目指しています。 モーメントでは、さまざまな内容の魅力的なオリジナルコンテンツを取り上げます。モーメントに掲載するツイートや動画の選定では、内容や画像/動画に加え、投稿者のプロフィール画像やユーザー名も検討の対象となります。 公平性：議論を呼びそうな話題に関するツイートを選定する際には、データに基づいて決定を行い、すでに Twitter で大きな反応を呼んでいるツイートを見つけます。公の議論に関するトピックの場合、可能な限り多様な視点を提示するように務めます。キュレーターは、Twitter でのありのままの会話を客観的に映し出すことを目指し、公平性を期すためのトレーニングを継続的に受けています。

	正確性：キュレーションの目的は、正確な情報を提示する質の高いツイートを紹介することにあります。キュレーションチームは、ソーシャルメディアの検証方法に関して定期的にトレーニングを受けています。また、ほとんどのモーメントは、レビューのプロセスを経てから取り上げられます。 誤情報の訂正：万が一、取り上げた情報が誤りだったことが判明した場合には、モーメントに目に見える形で訂正を加え、その旨をツイートで発表します。ただし、稀に当該モーメントを削除し、関連する国の Twitter アカウントを使用して、誤情報を撤回する旨をツイートする場合があります。 基準：ニュース性の高い会話を紹介するために必要な場合を除いて、冒涇、暴力、裸体など、センシティブであるとみなされるコンテンツは選定の対象から外されます。また違法行為を描写したり、助長したりするようなコンテンツは採用しません。センシティブであると受け取られる可能性のあるコンテンツがモーメントに含まれる場合、警告を表示します。 利害関係の対立の回避 Twitter のキュレーションチームは、収益を伸ばしたり、アカウント数を増やしたり、Twitter のパートナー関係を管理したりする責任は負いません。ツイートや動画の紹介や、モーメントのトピックの選択は、Twitter のユーザーにとっての利益を最も重視して行われます。こうした決定が、広告主、パートナー、または Twitter の事業関係者の利益の影響を受けることはありません。 キュレーションチームは基本的に、Twitter が属する業界や自社、Twitter の競合企業を扱ったモーメントは作成しません。Twitter や Twitter の競合企業が Twitter 上の会話の中で重要な要素となった場合（たとえば、当該企業の何らかの問題がトレンドトピックになるなど）、
--	--

		Twitter のモーメント編集チームは、当該企業からのリアクションがあればそれも含め、事実に基づく概要をまとめることがあります。 https://help.twitter.com/ja/rules-and-policies/twitter-moments-guidelines-and-principles
③	メディアとの連携体制構築 具体的検討・取組	東日本大震災の発生から 10 年となった 2021 年より、毎年 3 月に国内の報道機関をはじめとするパートナーと連携し、大震災の記憶を語り継ぎ、その教訓を未来へつなぐキャンペーンを展開しています。 2022 年は、「#防災いまできること」「#PrayForTohoku」の共通ハッシュタグを使い、3 月 10 日・11 日にはパートナーが同じタイミングでツイートします。Twitter スペースも活用しながら、一人でも多くの方の目に留まることを願い、より「わかりやすく」「見つけやすく」メッセージを発信しました。 https://blog.twitter.com/ja_jp/topics/company/2022/disasterprevention また、ウクライナ情勢につきましては、キュレーションチームがメディアからの正確の情報をまとめています。 https://twitter.com/i/events/1499255716128837632 https://twitter.com/i/events/1497404496284033029

④	情報源のトレーサビリティ確保、なりすまし防止・認証	Twitter は 2021 年に認証プログラムを再開する計画を発表し、新たな認証ポリシー案を公開して、今後のポリシー構築に向けて皆さんからのフィードバックを募集しました。この新しいポリシーは将来の運用の基盤となるもので、今後、認証の意味や資格、プロセスをより公正なものにするために、アカウントが認証を失う可能性がある理由などを定義します。 皆さんからのフィードバックは、Twitter のポリシー開発プロセスの重要な部分を占めます。それは世界中の人々にオープンなサービスとして、確実に利用者の声を Twitter のポリシーに反映したいからです。フィードバックの募集を始めて 2 週間で 22,000 件以上の回答があり、ポリシーの改善方法について非常に多くのヒントを得ました。 詳しくはこちらをご参照ください。 https://blog.twitter.com/ja_jp/topics/company/2020/our-plans-to-relaunch-verification-and-whats-next
10. その他		
①	意見・補足	