

## 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（2021年3月30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	コンテンツ共有サービス
②	加入者数、月間アクティブユーザー数又は書き込み数	【日本の数値】
		【グローバルの数値】
		【（質問に答えられない場合）参考となる数値】 収益化可能な日時平均アクティブユーザー数（mDAU）：1億9,200万（2020年第4四半期；グローバル）
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	
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）】
③		【②健康・医療関係（コロナ関係等）】
④		【③選挙・政治関係（米国大統領選等）】
⑤		【④全般・その他】

⑥	研究者への情報提供、利用条件	<p>無償の学術研究プロダクトトラックを提供しています。</p> <p>新しい学術研究プロダクトトラックにより、許可された研究者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。</p> <ul style="list-style-type: none"> <li>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。</li> <li>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。</li> <li>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、データクレンジング要件を最小限に抑えます。</li> <li>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。</li> </ul> <p>(参考：ブログ「Twitter API による学術研究の未来」  <a href="https://blog.twitter.com/developer/ja_ip/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api.html">https://blog.twitter.com/developer/ja_ip/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api.html</a>)</p>
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	Disinformation 対策フォーラム ( <a href="https://www.saferinternet.or.jp/anti-disinformation">https://www.saferinternet.or.jp/anti-disinformation</a> ) に構成員として参加しています。
②	具体的な役割	
③	議論を踏まえた取組	
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	<p>【①社会的混乱関係（災害等）】</p> <p><b>暴言や脅迫、差別的言動に対する Twitter のポリシー</b></p>

		<p>Twitterでは、法的または社会的に守られるべき特定の 카테고리に属する人々について、恐怖心を扇動したり、危険または非合法的な活動に参加する傾向が強いと主張することを禁止しています。</p> <p>特に社会的混乱の最中、例えば災害の直後などは、同様の主張は特定の 카테고리に属する人々の深刻な身体的危害に繋がりがねないことを考慮し、Twitter上で広めることは認められておりません。</p> <p>(参考：暴言や脅迫、差別的言動に対する Twitter のポリシー  <a href="https://help.twitter.com/ja/rules-and-policies/hateful-conduct-policy">https://help.twitter.com/ja/rules-and-policies/hateful-conduct-policy</a>)</p> <p>【②健康・医療関係（コロナ関係等）】  <b>COVID-19 について誤解を招く情報に関するポリシー</b></p> <p>Twitter のサービスを使って、損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。</p> <p>(参考：COVID-19 について誤解を招く情報に関するポリシー  <a href="https://help.twitter.com/ja/rules-and-policies/medical-misinformation-policy">https://help.twitter.com/ja/rules-and-policies/medical-misinformation-policy</a>)</p> <p>【③選挙・政治関係（米国大統領選等）】  <b>市民活動の阻害に関するポリシー</b></p> <p>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。</p> <p>(参考：市民活動の阻害に関するポリシー <a href="https://help.twitter.com/ja/rules-and-policies/election-integrity-policy">https://help.twitter.com/ja/rules-and-policies/election-integrity-policy</a>)</p> <p>【④全般・その他】  <b>合成または操作されたメディアに関するポリシー</b></p>
--	--	---

		<p>何らかの損害につながる可能性が高い、合成または操作されたメディアを、ユーザーを欺くことを意図してプロモーションすることは禁止されています。さらに、Twitterでは合成されたメディアや操作されたメディアが含まれるツイートをラベル付けし、ユーザーがツイートの信憑性や文脈を把握するのを支援する場合があります。</p> <p>(参考：合成または操作されたメディアに関するポリシー <a href="https://help.twitter.com/ja/rules-and-policies/manipulated-media">https://help.twitter.com/ja/rules-and-policies/manipulated-media</a>)</p>
	(ii)削除等の対応	<p><b>【①社会的混乱関係（災害等）】</b></p> <p><b>暴言や脅迫、差別的言動に対する Twitter のポリシーに違反した場合</b></p> <p>この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。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、最終的にはアカウントが永久凍結される場合があります。アカウントの活動が主に攻撃的な行為ばかりである場合、または強烈な身体的脅迫を共有したとみなされる場合、Twitter は最初の調査でアカウントを永久凍結する場合があります。</p> <p>執行オプションの範囲についての詳細 (<a href="https://help.twitter.com/ja/rules-and-policies/enforcement-options">https://help.twitter.com/ja/rules-and-policies/enforcement-options</a>) をご覧ください。</p> <p>アカウントが誤って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。 (参考：<a href="https://help.twitter.com/ja/rules-and-policies/hateful-conduct-policy">https://help.twitter.com/ja/rules-and-policies/hateful-conduct-policy</a>)</p> <p><b>【②健康・医療関係（コロナ関係等）】</b></p> <p><b>COVID-19 について誤解を招く情報に関するポリシーに違反した場合</b></p> <p>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</p>

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) COVID-19 ウイルスの特性または治療に関して誤解を招く情報や、(2) 悪意のある勢力や強い権力による意図的な陰謀を思い起こさせるパンデミックや COVID-19 ワクチンなど、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 ストライクとしてカウントされます。

### 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。

- 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、おすすぬから除外したりする
- 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、上記すべての措置が取られます。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。有害と判断され、ラベル付けされたツイートは、1 ストライクとしてカウントされます。

		<p><b>アカウントのロックと永久凍結</b></p> <p>COVID-19 について誤解を招く特定の説（または一連の説）をツイートしたり広めたりするためだけにアカウントが利用されていると Twitter が判断した場合も、凍結の根拠となります。</p> <p>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。</p> <p>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。</p> <ul style="list-style-type: none"> <li>• 1 ストライク: アカウントレベルでの対応なし</li> <li>• 2 ストライク: 12 時間にわたるアカウントのロック</li> <li>• 3 ストライク: 12 時間にわたるアカウントのロック</li> <li>• 4 ストライク: 7 日にわたるアカウントのロック</li> <li>• 5 ストライク以上: 永久凍結</li> </ul> <p>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せん。</p> <p>(参考 : <a href="https://help.twitter.com/ja/rules-and-policies/medical-misinformation-policy">https://help.twitter.com/ja/rules-and-policies/medical-misinformation-policy</a>)</p> <p><b>【③選挙・政治関係（米国大統領選等）】</b></p> <p><b>市民活動の阻害に関するポリシーに違反した場合</b></p> <p>市民活動の阻害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、以下などが挙げられます。</p>
--	--	--

### 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ストライクとしてカウントされます。

### 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せん。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せん。

(参考: <https://help.twitter.com/ja/rules-and-policies/election-integrity-policy>)

#### 【④全般・その他】

##### 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た場合

ツイートで共有されているメディアが、ユーザーを欺くことを意図して、大幅に改ざんまたはねつ造されていると判断する根拠がある場合、Twitterは、Twitter上に表示される当該メディアを共有しているツイートに文脈を追加して表示することがあります。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



		<ul style="list-style-type: none"> <li>• ツイートの拡散を抑制したり、おすすりから除外したりする</li> <li>• Twitter のモーメントやランディングページなどに、補足説明のためのリンクを表示する</li> </ul> <p>多くの場合、ラベル付け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。</p> <p>合成または操作されていて、ユーザーを欺くことを意図した方法で共有されており、何らかの損害につながる可能性が高いものは Twitter 上で共有することが禁止されており、削除の対象となります。当ポリシーの違反を繰り返したアカウントや、当ポリシーに対して重大な違反を行ったアカウントは永久凍結される場合があります。</p> <p>(参考 : <a href="https://help.twitter.com/ja/rules-and-policies/manipulated-media">https://help.twitter.com/ja/rules-and-policies/manipulated-media</a>)</p>
(2) 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】
		【グローバルの数値】
		【(質問に答えられない場合) 参考となる数値】
②	(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】 (コンテンツ削除件数 : 2020 年上半期) COVID-19 について誤解を招く情報に関するポリシー : 4,647 市民活動の阻害に関するポリシー : 2,710 (参照 : Twitter 透明性センター <a href="https://transparency.twitter.com/en/reports/rules-enforcement.html#2020-jan-jun">https://transparency.twitter.com/en/reports/rules-enforcement.html#2020-jan-jun</a> )

		【（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】 （アカウント凍結件数：2020 年上半期） COVID-19 について誤解を招く情報に関するポリシー：1,751 市民活動の阻害に関するポリシー：0 （参照：Twitter 透明性センター <a href="https://transparency.twitter.com/en/reports/rules-enforcement.html#2020-jan-jun">https://transparency.twitter.com/en/reports/rules-enforcement.html#2020-jan-jun</a> ）
		【（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）	コンテンツがポリシーに違反していると担当チームが判断した場合、ツイートにラベルがつけられます。そのような判断は自動化ツールに反映されるとともに、サービス上の同様のコンテンツを自主的に発見しラベル付けする能力の向上にも利用されます。自動化ツールと人間によるレビュー双方を活用し、COVID-19 ワクチンについて誤解を招く情報に対処していくことを目指しています。機械学習や言語処理の自動化が実行を示すまでには時間がかかります。そのため、まず英語のコンテンツへの対応から開始し、その後他の言語や文化的コンテンツにも拡大させる予定です。

⑤	削除以外の取組 (i)警告表示	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、当該ツイートにラベルや警告メッセージを適用したり、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ことがあります。
	(ii)表示順位の低下	ツイートの表示順位を下げたり、「おすすめ」から除外したりすることがあります。
	(iii)その他の取組内容	いいね、返信、リツイートが無効にしたり、まとめられたランディングページや関連する Twitter のポリシーなどに、補足説明のためのリンクを表示することがあります。
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】
(3) 削除要請や苦情に関する受付態勢・プロセス		
①	一般ユーザからの申告・削除要請への受付窓口・受付態勢、対応プロセス	<p>(参考：Twitter ルールおよび利用規約に違反している可能性のある事例の報告方法)</p> <p>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。それ以外の種類の違反を報告する方法については、「特定の種類の違反を報告する方法」(<a href="https://help.twitter.com/ja/rules-and-policies/twitter-report-violation">https://help.twitter.com/ja/rules-and-policies/twitter-report-violation</a>) の該当セクションをご覧ください。</p> <p>Twitter で不安な思いをしている利用者が、問題のある投稿を報告するという負担を背負うべきではありません。以前は、利用者から報告された場合にのみ悪意のあるツイートかどうか調査をしていました。ただ、それでは十分ではないと理解し、利用者からの報告に頼るだけでなく、積極的な強制執行対応を優先して取り組んでいます。</p>
②	対応決定時における通知の内容、理由の記載の程度	(参考：攻撃的な行為について、有効な報告を Twitter が受け付けた場合の処理)

		<p>報告を送信すると、報告を受け付けたことを通知する Twitter からの確認メッセージが表示されます（メッセージが表示されるまでに最長で 24 時間かかります）。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。報告した利用者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、その旨を利用者に通知します。</p> <p>報告したツイートの元のコンテンツは、報告済みであることを示すお知らせに置き換えられます。必要に応じ、クリックしてツイートを表示することができます。</p> <p>利用者が報告したアカウントに対して Twitter が何らかの対応を行った場合、その利用者に Twitter 内で通知されます。この対応は、その利用者が行った報告に関係していない場合もあります。</p> <p>(<a href="https://help.twitter.com/ja/safety-and-security/report-abusive-behavior">https://help.twitter.com/ja/safety-and-security/report-abusive-behavior</a>)</p>
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】

④	④の部署・チームに関する日本国内の拠点の有無、日本における責任者の有無	
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	
(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	
	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	2012年7月以来、年2回更新される「Twitter 透明性センター」( <a href="https://transparency.twitter.com/ja.html">https://transparency.twitter.com/ja.html</a> )を通じて、Twitterに寄せられる法的請求、知的財産に関する異議申し立て、Twitterルールに基づく強制的対応、プラットフォームの操作の状況等を公表しています。 日本語訳も提供しています。最新版(2020年上半期)は2021年3月現在翻訳作業中。
③	取組の効果分析	
< 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	誤解を招く情報のラベルがついたコンテンツは、Amplify パブリッシャープログラム( <a href="https://business.twitter.com/ja/help/troubleshooting/amplify-publisher-faqs.html">https://business.twitter.com/ja/help/troubleshooting/amplify-publisher-faqs.html</a> )により自動的に収益化を無効にされます。また、不適切なコンテンツのポリシーのもと、広告として配信することができません。 (参考: 不適切なコンテンツに関するポリシー <a href="https://business.twitter.com/ja/help/ads-policies/ads-content-policies/inappropriate-content.html">https://business.twitter.com/ja/help/ads-policies/ads-content-policies/inappropriate-content.html</a> )

②	<p>広告出稿制限</p>	<p><b>政治に関するコンテンツ</b></p> <p>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。</p> <p>Twitter では、政治に関するコンテンツとは、候補者、政党、選出または任命された政府関係者、選挙、国民投票、住民投票、法律、規制、指令、または裁判に関する結果に言及するコンテンツであると定義しています。</p> <p>このポリシーでは、政治に関するコンテンツへの言及を含んだ投票の呼びかけ、財政支援の要請、または、上記の類の政治に関するコンテンツに対する賛成または反対意見などを含む広告を禁止しています。</p> <p>また、候補者、政党、選出または任命された政府関係者によるいかなる種類の広告も禁止しています。</p> <p>(参考：政治に関するコンテンツのポリシー <a href="https://business.twitter.com/ja/help/ads-policies/ads-content-policies/political-content.html">https://business.twitter.com/ja/help/ads-policies/ads-content-policies/political-content.html</a>)</p> <p><b>不適切なコンテンツ</b></p> <p>また、不適切なコンテンツに関するポリシーは、危険または搾取的なコンテンツ、虚偽なコンテンツ、誤解を招く、合成または操作されたコンテンツを禁止しています。</p> <p>(参考：不適切なコンテンツに関するポリシー <a href="https://business.twitter.com/ja/help/ads-policies/ads-content-policies/inappropriate-content.html">https://business.twitter.com/ja/help/ads-policies/ads-content-policies/inappropriate-content.html</a>)</p>
③	<p>ターゲティング技術の適用に関する規定</p>	<p>Twitter 広告には、広告主のコンテンツに関心がありそうな利用者にリーチするためのさまざまなターゲティングオプションが用意されています。タイムラインでのキーワードターゲティングを使用する場合、利用者の信頼を損なわずにキャンペーンの目標を達成することが重要です。広告主がタイムラインでのキーワードターゲティングを使用する場合、キーワー</p>

		<p>ドターゲティングに関するポリシーおよび他のすべての Twitter 広告ポリシーに従う必要があります。</p> <p>(参考: キーワードターゲティングに関するポリシー  <a href="https://business.twitter.com/ja/help/ads-policies/campaign-considerations/policies-for-keyword-targeting.html">https://business.twitter.com/ja/help/ads-policies/campaign-considerations/policies-for-keyword-targeting.html</a>)</p> <p>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。</p> <p>(参考: 興味関心ターゲティングのオプトアウトに関するポリシー  <a href="https://business.twitter.com/ja/help/ads-policies/product-policies/interest-based-opt-out-policy.html">https://business.twitter.com/ja/help/ads-policies/product-policies/interest-based-opt-out-policy.html</a>)</p>
④	<p>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</p> <p>AI 原則・ガイドライン等の参照</p>	
⑤	出稿者の情報や資金源の公開	<p>Twitter は 2019 年 11 月に政治広告を全面的に禁止しました。2018 年 5 月 24 日から 2019 年 11 月 22 日にかけて配信された政治広告と、2018 年 8 月 8 日から 2019 年 11 月 22 日にかけて配信された社会問題広告については、以前開設していた Twitter 広告の透明性センター (ATC) データのアーカイブ版を提供しています。</p> <p>(参考: 広告の透明性 <a href="https://business.twitter.com/ja/help/ads-policies/product-policies/ads-transparency.html">https://business.twitter.com/ja/help/ads-policies/product-policies/ads-transparency.html</a>)</p>
⑥	広告とコンテンツの分離	<p>広告主のプロモツイートにはすべて、「プロモーション」というラベルが明示されます。</p> <p>(参考: プロモツイートとは何ですか?  <a href="https://business.twitter.com/ja/help/overview/what-are-promoted-tweets.html">https://business.twitter.com/ja/help/overview/what-are-promoted-tweets.html</a>)</p>

⑦	その他の透明性・アカウントビリティ確保方策 ユーザへのツール提供	Twitterの利用者は誰でもTwitter広告をアプリ内またはwebフォームから報告することができます。また、特定の広告を非表示にしたり、広告主をブロックすることもできます。 (参考: Twitter 広告を報告する <a href="https://help.twitter.com/ja/safety-and-security/reporting-twitter-ads">https://help.twitter.com/ja/safety-and-security/reporting-twitter-ads</a> )
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	
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	
③	ファクトチェックを実施する人材の育成	
④	ファクトチェック機関との連携	<p>パンデミックの期間中、Twitterではパートナーシップを構築し、公共の会話を保護し、利用者が信頼性の高い公衆衛生情報を見つけやすくするための取り組みを進めてきました。また利用者が正しい情報源から正しい情報を得られるよう、広告枠の無償提供を行なっています。この取り組みはファクトチェック機関に限定されたものではありません。</p> <p>世界保健機関、多数の政府機関や公共保健機関、世界中の当局者など、各国の関連するステークホルダーとオープンな対話を行っており、それらのステークホルダーがアカウントの問題をトラブルシューティングしたり、所属する専門家の認証を取得したり、損害を軽減するためにTwitterを活用する際に戦略的なアドバイスを求めたりできるようにしています。</p> <p>これまでに実施してきた #StayHome (#ステイホーム)、#WashHands (#手を洗おう)、#WearAMask (#マスクをしよう) を呼びかける取り組みの成果を踏まえ、2021年1月には Team Halo、UNICEF、NHS、Vaccine Confidence Project と合同でワクチン接種への指示</p>



		を表明する #vaccinated (#ワクチン完了) 絵文字・ハッシュタグキャンペーンを立ち上げています
7. 「ICT リテラシー向上の推進」関係		
①	普及啓発の取組・投資	<ul style="list-style-type: none"> <li>Twitter プラットフォーム上における公式アカウントからの情報提供</li> <li>ヘルプセンターページでの情報や動画等の提供</li> </ul>
②	他のステークホルダーとの連携・協力・投資	<ul style="list-style-type: none"> <li>セーファーインターネットデーやその他国際的な記念日に際して、国際機関・各国政府・NPO 等と連携し啓発キャンペーンをプラットフォーム上で実施。期間限定の絵文字や共通ハッシュタグを提供。</li> <li>NPO に対する広告枠の無償提供プログラム「Ads for Good」の実施。</li> <li>UNESCO とメディア情報リテラシーに関するパートナーシップを締結し、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」を提供。</li> <li>UNESCO ならびに欧州委員会とのパートナーシップ による、COVID-19 に焦点を当てたメディア情報リテラシーキャンペーンの実施。</li> <li>SNS 運営企業や学識経験者によって設立された業界団体であるソーシャルメディア利用環境整備機構（SMAJ）への参画。</li> </ul>
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	<p>Adobe、ニューヨークタイムズ、Twitter はデジタルコンテンツの信頼性確保を目的とした業界標準開発のためのイニシアティブを発表しました。このイニシアティブは、各メンバーがその深い技術的およびビジネス知識をソリューションに反映させることを目的としています。長期的なソリューションに貢献するメンバーのエコシステムを構築し、フレームワークを採用し、消費者が誰を、何を信頼すべきかを理解できるよう支援していきます。教育もまた、デジタル・インテグリティ・ソリューションの重要な要素です。このイニシアティブは、消費者がデジタルメディアをよりよく理解し、評価できるようにするための教育活動に取り組んでいます。パートナー企業は、消費者のメディアリテラシーに関する意識とスキルを向上させるために協力していきます。</p>

9. 「情報発信者側における信頼性確保方策の検討」関係	
①	<p>信頼性の高い情報の表示</p> <p><b>#ThereIsHelp</b></p> <p>Twitter は、公共の場での会話の健全性に全力で取り組んでいます。このミッションを達成するためには、利用者が信頼性の高い情報源から信頼できるコンテンツを得られることが重要です。Twitter では特定分野のキーワードを検索すると、信頼性の高い公的機関や NPO 等の連絡先が表示される #ThereIsHelp を提供しています。</p> <p>COVID-19 に関連するキーワードを検索すると、日本では厚生労働省の「新型コロナウイルス感染症について新型コロナウイルス感染症について」のページが案内されます。この取り組みは 78 カ国・地域で実施されています。</p> <p>COVID-19 以外では、自殺防止の #ThereIsHelp は日本を含む 30 カ国・地域で、ワクチン情報については 39 カ国・地域で、ジェンダーに基づく暴力対策では 27 カ国・地域で提供されています。</p> <p><b>キュレーションによる Twitter モーメント</b></p> <p>キュレーションチームは、テキスト、動画、GIF 画像やライブ放送などの優れたツイートを見つけ、紹介することを目指しています。こうしたツイートは利用者に見識を与え、Twitter 上で「いま」交わされている会話の背景情報をもたらします。キュレーターは、オリジナルコンテンツを作ったり記事形式で紹介したりするのではなく、Twitter に既に存在する魅力的なコンテンツをまとめ、提示する役割を担っています。こうしたコンテンツはモーメントや、トレンドの説明コンテンツ、リストなどに表示されます。</p> <p>キュレーターはグローバルな多言語チームを構成しており、ニュース、スポーツ、エンターテイメント、ユーモアといったカテゴリーを横断しながら、Twitter でまさに「いま」起きていることを探し求めています。Twitter では現在、5 つの言語（英語、日本語、アラビア</p>

		<p>語、スペイン語、ポルトガル語)で16の市場に向けてキュレーションサービスを提供しています。キュレーターは正確性と公平性を期して、高品質のコンテンツを見つけるためのトレーニングを定期的に受けています。</p>
②	<p>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</p>	<p><b>公平性、正確性、基準</b></p> <p>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。</p> <p>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。</p> <p><b>公平性</b>：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。</p> <p><b>正確性</b>：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。</p> <p><b>誤情報の訂正</b>：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。</p>

		<p><b>基準</b>：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。</p> <p><b>利害関係の対立の回避</b></p> <p>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。</p> <p>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。</p> <p>（参考：Twitterモーメントのガイドラインと方針 <a href="https://help.twitter.com/ja/rules-and-policies/twitter-moments-guidelines-and-principles">https://help.twitter.com/ja/rules-and-policies/twitter-moments-guidelines-and-principles</a>）</p>
③	メディアとの連携体制構築 具体的検討・取組	<p>メディアパートナーは「話題を検索」タブに掲載されるモーメントを提出することができます。このパートナーシップは無償のものです。</p> <p>（参考：Twitterモーメントのガイドラインと方針 <a href="https://help.twitter.com/ja/rules-and-policies/twitter-moments-guidelines-and-principles">https://help.twitter.com/ja/rules-and-policies/twitter-moments-guidelines-and-principles</a>）</p>
④	情報源のトレーサビリティ 確保、なりすまし防止・認証	<p><b>認証プログラム</b></p> <p>Twitterのブルーの認証済みバッジは、著名人のアカウントなど、世間の関心を集めるアカウントが本物であることを示します。</p>

		<p>Twitter では 2021 年の早期に、新たな一般公募手続きを含む認証の再開を予定しています。現時点では認証を申請する方法はありません。</p> <p>(参考：ブログ「Twitter 認証プログラム再開の計画と今後の予定について」  <a href="https://blog.twitter.com/ja_jp/topics/company/2020/our-plans-to-relaunch-verification-and-whats-next.html">https://blog.twitter.com/ja_jp/topics/company/2020/our-plans-to-relaunch-verification-and-whats-next.html</a>)</p> <p><b>政府および国家当局関係メディアアカウントの識別</b></p> <p>Twitter は人々が今、世界で何が起きているのかを目にして、自国の政府や政府高官の考えを知ることができる場所です。私たちは、安全と表現の自由とは密接に結びついていると考えており、特に国家の代表者や関連機関とやり取りする際には、Twitter 上で人々が目しているものに背景情報を追加することで、人々のより多くの情報に基づく体験に役立つと考えています。</p> <p>Twitter では、1) 主要な政府関係者のアカウント、2) 国家当局に関係したメディア、の 2 つのカテゴリーにラベルを適用しています。Twitter がフォーカスしているのは、国外に向けて国家の見解を表明する政府高官、国家代表者および機関です。私たちはこれが重要な措置だと確信しており、人々が別の国の地政学的な問題を話し合う際に、国家との関連性に関する背景情報や、そのアカウントがどの国を代表しているのか、より多くの情報を得ることができます。私たちは、国家代表者下にある海外に政策的視点を提供するそれぞれの政権内の人々にも焦点を当てています。</p> <p>(参考：ブログ「政府および国家当局関係メディアアカウントの識別に関する取り組みの拡大」  <a href="https://blog.twitter.com/ja_jp/topics/company/2021/account-labels-phase2.html">https://blog.twitter.com/ja_jp/topics/company/2021/account-labels-phase2.html</a>)</p>
10. その他		
①	意見・補足	