

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（2023年4月13日）

社名	Meta Platforms, Inc.	
1. 前提（サービス概要）		
①	対象サービス名	Facebook
	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ー数又は書き込み数	<p>【日本の数値】 Facebook 月間アクティブ利用者数：2,600万人（2019年7月）</p> <p>【グローバルの数値】 Facebook 月間アクティブ利用者数 29億6,000万人（2023年12月時点）</p> <p>【（質問に答えられない場合）参考となる数値】</p>
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	<p>弊社の misinformation に対する考え方は、表現の自由とバランスをとりながら、正確で有益なコンテンツを人々に提供するという原則に基づいています。</p> <p>Misinformation は、何が禁止されているかを包括的に明示する方法がないため、弊社のコミュニティ規定で扱われている他のタイプのスピーチとは異なります。例えば、暴力や過激な描写を含むコンテンツやヘイトスピーチの場合、弊社のポリシーは、禁止することとなる内容を明記していますが、それは弊害、比例性、必要性の原則、そしてコミュニティスタンダードの背景にある価値（自らの考えを主張できること、信頼できること、安全が保たれていること、プライバシーが確保されること、そして尊厳が保たれることなど）を考慮した上でのことになります。</p> <p>Misinformation について考えるとき、政府、政策立案者、市民団体、学術研究社、そして一</p>

		<p>般の人々が、Misinformation とは一体何かについて同意していない、という本質的に困難な問題をはらんでいるという課題があります。ある人が misinformation であると考えてるものであっても、それは別の人の意見であったり、風刺やパロディであったりすることもある、ということ です。</p> <p>さらに、意図せずに共有される misinformation と、騙すために意図的に共有される misinformation—一般に「disinformation」と呼ばれる—ものとの間には重要な違いがあります。何が misinformation (意図的でないもの) であるかを定義することは、非常に難しいことです。さらに、何が真実でないかどうかを一体誰が判断するのか (すなわち、誰が、あるいは何が真実を語る資料 (the source of truth) なのか)、そしてどのような罰則が必要なのか、ということも、この課題に拍車をかけています。これらの問題に対処しようとするルールは、合法的な言論を一貫性なく恣意的に制限してしまう危険性があります。一部の政府は、mis/disinformation を有害なコンテンツのカテゴリーとみなしています。しかし、何が有害であるかという判断は、依然として文脈に大きく左右され、定義が難しく、文化に依存した主観的であることが多く、法的にも曖昧なものになります。したがって、有害コンテンツに関するルールは、有害コンテンツの削除と表現の自由やその他の基本的権利の保護のバランスを取る必要性を認識する必要があります。</p> <p>弊社は、差し迫った物理的な危害のリスクを直接助長する可能性が高い misinformation を削除します。また、政治プロセスの機能の妨害を直接助長する可能性のあるコンテンツや、非常に欺瞞的な加工されたメディアも削除します。このようなカテゴリーにおける misinformation の要素を判断するにあたり、弊社は、知識と専門性を有する独立した専門家と連携し、コンテンツの真実性や、差し迫った危害のリスクを直接助長する可能性が高いかどうかを評価します。例えば、当事国で活動する人権団体と連携して内戦に関する噂の真偽を判断したり、新型コロナウイルス感染症の世界的流行時に保健機関と連携したりすることが挙げられます。</p>
--	--	---

その他のすべての misinformation については、弊社は、その情報の表示頻度の抑制や、生産的な対話を促す環境作りに力を入れています。弊社は、例えば、人々は大げさなことを言うとき(例:「このチームの成績はスポーツ史上最悪だ!」)や、ユーモアや風刺を伝えるとき(例:「私の夫は『ハズバンド・オブ・ザ・イヤー』を受賞した!」)に無害な方法で misinformation を利用することが多々あることを理解しています。また、不正確な情報を含んだストーリーを交えて体験談をシェアすることもあります。場合によっては、自分では深く信じていても、他人にとっては虚偽だと思える意見や、自分では真実だと信じているが、他人には不完全であったり誤解をまねいたりすると捉えられる情報を伝えることもあります。

弊社は、このような言論がいかにより一般的であるかを認識し、デマや広がりやすい misinformation の拡散を遅らせ、利用者を権威ある情報へ導くことに注力しています。権威ある情報についての弊社の戦略は、コミュニティが特定の脅威(COVID-19 健康危機など)に直面し、その脅威について misinformation が広まり、利用者が混乱するリスクが高く、参照できる権威あるソースや情報が広く合意されているような場合に、特設サイト(新型コロナウイルス感染症情報センター、気候学センターなど)を立ち上げることが中心になります。弊社は、権威ある情報をより見やすく、アクセスしやすくすることで、こうした脅威に対する人々の行動、態度、知識を変えたいと考えています。そうすることで、misinformation の拡散を抑え、不確実性につけこんで言論を操作しようとする悪意のあるネットワークの効力を弱めようとするものです。

https://www.facebook.com/coronavirus_info/

<http://www.facebook.com/climatescienceinfo>

また、利用者が自分で何を読み、何を信頼し、何を共有するかを決められるように、メディアとデジタルのリテラシーを高めるためのリソース(下記参照)を提供しています。

<https://wethinkdigital.fb.com/>

		<p>最後に、misinformationの拡散と重なることが多い、その他の分野でのコンテンツや行動も禁止しています。例えば、弊社のコミュニティ規定は、いわゆる偽アカウント、不正及び詐欺、組織的偽装行為を禁止しています。</p> <p>https://transparency.fb.com/en-gb/policies/community-standards/account-integrity-and-authentic-identity/</p> <p>https://transparency.fb.com/en-gb/policies/community-standards/fraud-deception/</p> <p>https://transparency.fb.com/en-gb/policies/community-standards/inauthentic-behavior/</p> <p>オンラインとオフラインの環境が変化し、進化するにつれて、弊社はこれらのポリシーを進化させ続けます。以下に挙げる misinformation を繰り返し共有するページ、グループ、プロフィール、Instagram アカウ​​ントは、コンテンツが削除されるだけでなく、配信数の減少、広告掲載の制限、または弊社のプラットフォームから削除される場合があります。Facebook がコンテンツを削除する際の追加情報は、こちらをご覧ください。</p> <p>https://www.facebook.com/help/260743102021762</p>
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		<p>【②健康・医療関係（コロナ関係等）】</p> <p>パンデミック開始以来、新型コロナウイルス感染症及びワクチンの misinformation に関するポリシーに違反したとして、日本での 23,000 件以上を含め、全世界で 2600 万件以上のコン</p>

		テンツを Facebook および Instagram から削除しています (2022 年 3 月時点)。
④		【③選挙・政治関係】
⑤		<p>【④全般・その他】</p> <p>弊社の提供するコミュニティは、全世界で 37 億人を超える利用者からなり、またデジタルコミュニケーションは国境を越えています。従って、弊社のコンテンツモデレーションポリシーとその実務は、グローバルに運用され、拡張される必要があります。そのため、弊社の施行にあたっての努力を測定するアプローチもグローバルなものとなっています。弊社のコミュニティ規定施行レポートは、その進捗状況をより効果的に追跡し、利用者にとって Facebook と Instagram を安全にするための弊社の継続的な取り組みを示すために、四半期ごとに発行されています。</p> <p>https://transparency.fb.com/data/community-standards-enforcement/</p>
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	下記 8. 参照。
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	<p>数十億人にも及ぶ弊社プラットフォームの利用者全員と協力関係を構築することは不可能と いってよいです。そのため、弊社では市民団体や活動家グループなどの他者の利害を代表する 組織、ならびにデジタル、公民権、差別禁止法、言論の自由、その他の基本的人権の分野 における専門家と協力し合う機会を模索しています。</p> <p>https://transparency.fb.com/en-gb/policies/improving/stakeholders-help-us-develop-community-standards/</p> <p>https://transparency.fb.com/en-gb/policies/improving/our-stakeholders</p>

		<p>市民団体は、メタが当社のプラットフォームの影響と、当社が世界中で事業を展開する多様なコミュニティの背景をよりよく理解するための重要なパートナーです。私たちは、このような市民社会とのパートナーシップを促進し、現地組織のソーシャルメディア監視能力を強化し、私たちのプラットフォームで利用者の安全を守るためのポリシー、実施プロセス、製品を改善するために、信頼できるパートナープログラムを構築しました。私たちの信頼できるパートナーのネットワークには、世界 113 カ国から 400 を超える NGO、人道支援機関、人権擁護者、研究者が参加しています。パートナーは、豊富な知識と経験を持ち、私たちのコンテンツモデレーションの取り組みに役立っています。</p> <p>弊社は、Facebook や Instagram のコンテンツについて質問や懸念を持っている利用者の声や経験を代表するような専門家組織と提携しています。</p> <p>信頼できるパートナーの選定にあたっては、ソーシャルメディアモニタリングの経験があり、弊社のコンテンツポリシーについて学ぶことに関心があり、オンラインコミュニティを安全に保つことに尽力し、有害コンテンツの影響を不当に受け、社会的に不利な立場にあるグループの声を代弁することができる組織を求めています。弊社は、地域の状況、言論の傾向、差し迫った被害のシグナルをよりよく理解するために役立つ、専門的な市民団体とのパートナーシップを大切にしています。</p> <p>また、弊社は、社会的コミュニケーション技術に関連する misinformation や分断・分極化といった社会的課題の解決に関心を持つ研究者との関係構築やパートナーシップを構築する取り組みを行っています。</p>
②	具体的な役割	<p>弊社のインテグリティへの取り組みは、「人と人がより身近になる世界を実現する」という会社のミッションの中核をなすものです。なぜなら、ソーシャルテクノロジーを、世界中の人々が自由に、公正に、そして安全に自己表現できる場としたいと願っているからです。ここ数</p>

年、弊社は、利用者が私たちのプラットフォームで遭遇するネガティブな体験の影響を最小限に抑えるために、人とテクノロジーへの投資を増やしてきました。こうした考えから、弊社は、多様なステークホルダーと様々な方法で築き上げ、育んできた関係において積極的な役割を果たしており、その関係性の範囲に加えて、それが弊社の意思決定や行動にどのように役立ったかについて、以下に説明します。

第三者によるファクトチェック団体

ファクトチェックの仕組みと、その関係、パートナーシップにおける弊社の役割のプロセスは、以下の通りです：

Misinformation の特定：

(a) 弊社のテクノロジーは、人々の反応やコンテンツの拡散速度など、さまざまなシグナルに基づいて misinformation の可能性が高い投稿を検出することができます。また、利用者がコンテンツに「偽ニュース」のフラグを立てるかどうかが、不信感を示す投稿にコメントがあるかどうかを考慮します。(b) ファクトチェック団体は独自にレビューするコンテンツの特定も行います。

コンテンツの審査：

ファクトチェック団体は、コンテンツを審査し、その正確性を評価します。このプロセスは弊社から独立して行われ、情報源への問い合わせ、公開データの参照、画像やビデオの確証作業などが含まれます。ファクトチェック団体が行うことになる評価は、False、Altered、Partly false、Missing context、Satire、True です。これらの評価は、下記サイトでより詳細に定義されています。

<https://transparency.fb.com/en-gb/features/content-ratings-fact-checkers-use/>

これらの評価に基づいて弊社が取る行動は、以下のとおりです。

False または Altered と評価されたコンテンツは、最も不正確なコンテンツであるため、最

も積極的な対応が採られます。「Satire」または「True」と評価されたコンテンツは、ラベルや制限を受けません。

Misinformation のラベルを付け、それを利用者に知らせる：

コンテンツがファクトチェック団体によって評価された場合、そのコンテンツに通知を追加し、利用者が文脈を踏まえて読むことができるようにします。False または Altered と評価されたコンテンツには最も強い警告のラベルを、Partly false または Missing context と評価されたコンテンツにはより軽いラベルを適用しています。「Satire」または「True」と評価されたコンテンツにはラベルを付けませんが、Facebook 上の投稿にファクトチェックの記事を添付しています。また、このコンテンツを共有しようとする人や、過去に共有したことがある人には、事前に通知します。

Misinformation と評価されたコンテンツを利用者が見れなくなるようにする：

ファクトチェック団体が「False」「Altered」「Partly False」と評価したコンテンツ、または弊社が「ほぼ同一」と検出したコンテンツは、フィード内で下位に表示されます。弊社は、False と Altered の投稿の配信を劇的に減らし、Partly false の配信をより少なく抑えます。これにより、それを目にする人の数を大幅に減らすことができます。Missing context については、ファクトチェック団体からの情報をより多く提供することに重点を置いています。また、ファクトチェック団体から False、Altered、Partly false、Missing context と評価されたコンテンツの広告を拒否し、このコンテンツは推奨しません。

連続違反者に対する対応：

False または Altered と評価されたコンテンツを繰り返し共有するページ、グループ、プロフィール、ウェブサイト、Instagram アカウントは、一定期間、いくつかの制限を受けることとなります。これには、弊社が利用者に示す「おすすめ」からの削除、配信の減少、収益化と広告の能力の排除、ニュースページとしての登録の削除が含まれます(ファクトチェックプログラムの仕組みについては、下記参照。)

<https://transparency.fb.com/en-gb/features/how-fact-checking-works/>

信頼できるパートナー/市民団体:

信頼できるパートナーの関係形成を行ってきており、彼らに弊社にコンテンツを報告してもらい、また、弊社の取組みが利用者の安全を守ることを保証するために、弊社のコンテンツポリシーとその運用について重要なフィードバックを提供してもらっています。

信頼できるパートナーが持つ特定課題の専門家の知識及び地域の実情に精通した専門家の知識は、弊社のコンテンツモデレーションの際に有用な情報を提供し、さまざまな視点を考慮することを可能にすることを通じて、弊社のポリシーを強化するのに役立ちます。例えば、信頼できるパートナーと協議することで、ある種の misinformation がどのような状況や文脈で、差し迫った暴力や身体的危害のリスクにつながるか、社会的緊張を激化させるか、暴力を誘発するか、民主的プロセスを弱めるかについて地域レベルで洞察を得て、有害な misinformation について理解しようとしています。このような協議により、弊社は、危機的な状況下でもコミュニティの安全を守ることができるような、適切な政策を開発することができます。

研究者/学術関係者:

弊社は、社会的コミュニケーション技術に関連する misinformation や分断・分極化といった社会的課題の探求に関心を持つ研究者に資金を提供しています。例えば、2021年には、ソーシャルメディアやソーシャルテクノロジーのプラットフォームにおける misinformation に関する課題の理解を深めることを目的とした研究提案に対して、総額 200 万ドルの資金を投じました。

2021年6月、弊社は misinformation と分断・分極化に関する第3回 Foundational Integrity Research の提案公募 (RFP) を発表しました。この RFP には、世界中の 288 の大学や研究所

		<p>から 446 件の質の高い提案が集まりました。受賞した 19 の提案では、21 カ国での研究がカバーされています。</p> <p>この RFP を通じて、ソーシャルメディアやソーシャルテクノロジーのプラットフォームにおける misinformation、偏向報道、情報の質、社会的対立の分野における科学的知見の発展を支援することを目的としています。また、ソーシャルテクノロジー企業がプラットフォーム上の社会的問題にどのように対処すればよいか、より広い科学界とテクノロジー業界全体の共通理解に貢献することも目標に掲げました。なお、この研究は、弊社のアプリやテクノロジーに焦点を当てたものに限定されるものではありません。</p> <p>https://research.facebook.com/blog/2021/09/announcing-the-2021-recipients-of-research-awards-in-misinformation-and-polarization/</p>
③	議論を踏まえた取組	上記参照。
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>【②健康・医療関係（コロナ関係等）】</p> <p>【③選挙・政治関係】</p> <p>【④全般・その他】</p> <p>弊社は以下に掲げる種類の偽情報を削除します。</p> <p>I. 実際の危害や暴力</p>

	<p>弊社は、専門家であるパートナーが人々に対する差し迫った暴力または実際の危害のリスクを直接助長する可能性が高いと判断する、偽情報および検証できない噂を削除します。偽情報とは、信頼できる第三者が虚偽であると判断する主張を含むコンテンツと定義しています。検証できない噂とは、専門家のパートナーによる情報源の特定が極めて困難または不可能な主張、信頼できる情報源がない主張、その内容を証明するための具体性が不十分な主張、またはその内容があまりに信じがたい、もしくは不合理で信用できない主張と定義しています。</p> <p>弊社は、一見無害に思われる偽情報でも、特定の文脈では、死や深刻な怪我といった実際の危害の高いリスクに繋がりがかねない暴力的な脅しなど、オフラインでの危害のリスクを助長する可能性があることを認識しています。弊社は、こうした各地域の動向について専門知識を有する非政府組織(NGO)、非営利団体、人道支援組織、国際機関のグローバルネットワークと連携しています。</p> <p>II. 有害な健康関連の偽情報</p> <p>弊社では、主要な保健機関に相談し、公衆衛生および安全への差し迫った危険を直接助長する可能性が高い健康関連の偽情報を特定しています。弊社が削除する有害な健康関連の偽情報には次のようなものがあります。</p> <ul style="list-style-type: none"> ● ワクチンに関する偽情報。弊社は、主にワクチンに関する偽情報について、保健当局がその情報は虚偽であり差し迫ったワクチン接種の拒否を直接助長する可能性が高いと判断した場合は、その情報を削除します。これには次のようなものがあります。 <ol style="list-style-type: none"> a. ワクチンは自閉症を誘発する(例: 「最近自閉症の子どもが多いのはワクチン接種の増加が原因である」) b. ワクチンは乳児突然死症候群(SIDS)を誘発する(例: 「ワクチンがSIDSを引き起こすことをご存じないのですか?」)
--	---

		<p>c. ワクチンは予防しようとしている病気を引き起こす、またはワクチンを接種した人はその病気にかかりやすくなる(例: 「ワクチンには病気の株が含まれているため、接種するとその病気にかかりやすくなります。気をつけて!」)</p> <p>d. ワクチンまたはその成分は致命的、有毒、有害または危険である(例: 「もちろんワクチンを受けてもかまいません。体に毒を入れるのに抵抗がないのであれば」)</p> <p>e. 自然免疫のほうがワクチンによる獲得免疫よりも安全である(例: 「ワクチンを接種するよりもその病気にかかるほうが安全である」)</p> <p>f. 複数のワクチンを短期間に接種することは、医師に推奨されたタイミングであっても危険である(例: 「複数のワクチンを同時に接種しないようにしましょう。医者が何と言おうと同時接種は危険です!」)</p> <p>g. ワクチンには予防しようとしている病気を防ぐ効果がない。ただし、新型コロナウイルス感染症、インフルエンザ、マラリアのワクチンに関して、これらのワクチンが当該ウイルスに感染するのを防ぐ効果がないとする主張については削除しません。(例: 「ポリオワクチンは病気を予防するために何の役にも立ちません」は削除。「ワクチンは病気を予防するために何の役にも立ちません」は削除。ただし、「ワクチンを打っても、新型コロナウイルス感染症にかかるのを防ぐことはできません。だからこそ、周りに人がいるときには引き続き社会的距離をとって、マスクを着用する必要があるのです」は許可。)</p> <p>h. 麻疹にかかっても死ぬことはない(追加の情報や背景情報が必要)(例: 「麻疹にかかっても命にかかわることはないので心配いりません」)</p> <p>i. ビタミンCはワクチンと比較して、そのワクチンで予防できる病気の予防に同程度の効果がある。</p> <p>● 公衆衛生上の緊急事態の間の偽情報。 偽情報については、保健当局がその情報は虚偽であり差し迫った実際の危害のリスクを直接助長する可能性が高いと判</p>
--	--	---

断した場合、公衆衛生上の緊急事態の間、弊社はその情報を削除します。これには、個人が有害な疾病に感染したり、それを拡散させたりするリスクや、関連するワクチンを拒否するリスクを助長するものが含まれます。弊社では、世界と地域の保険機関と連携して公衆衛生上の緊急事態を特定しています。これには現在、ウイルスの存在または深刻度、治療や予防の方法、ウイルスの感染経路、免疫のある人に関する、専門の保健当局によって検証された新型コロナウイルス感染症関連の偽の主張や、新型コロナウイルス感染症に関する正しい衛生習慣（検査を受けること、社会的距離を置くこと、マスクの着用、新型コロナウイルス感染症ワクチンの接種など）を阻害する偽の主張が含まれます。**新型コロナウイルス感染症およびワクチンについて、弊社が許可していない偽情報の種類をまとめたルール一覧については [こちら](#)をご参照ください。**

<https://www.internmc.facebook.com/help/230764881494641>

- **健康上の問題に対する有害な「奇跡的な治療法」の宣伝または擁護**これには、推奨される使い方が健康上の文脈では深刻な怪我や死のリスクを直接助長する可能性が高い治療法、および正当な健康上の用途を有しない治療法（漂白剤、消毒剤、黒軟膏、苛性ソーダなど）が含まれます。

III. 投票者または国勢調査への干渉

選挙や国勢調査の健全性を促進する取り組みとして、弊社は、このようなプロセスに参加する人びとを妨害するリスクを直接助長する可能性が高い偽情報を削除します。これには次のようなものが含まれます。

- 投票や有権者登録、国勢調査への参加に関する日付、場所、時間、方法に関する偽情報
- 投票できる人物、投票資格、投票の有効性、投票するために提供しなければならない情報や書類に関する偽情報

	<ul style="list-style-type: none"> ● 候補者が立候補するか否かに関する偽情報 ● 国勢調査への参加資格や、国勢調査へ参加するために提供しなければならない情報や書類に関する偽情報 ● 個人の国勢調査の情報が別の(国勢調査の実施を担当していない)政府機関と共有されると述べる事(ただし該当する場合)など、政府の国勢調査への関与についての偽情報 ● 米国移民関税執行局(ICE)が投票所に待機しているとの偽りの主張をしているコンテンツ ● 投票プロセスに参加すると新型コロナウイルス感染症(または別の伝染病)に感染するとする明確な偽の主張 <p>弊社では、暴力の呼びかけ、違法な参加の助長、および選挙への計画的な干渉の呼びかけを対象とする追加ポリシーを設けており、これらは弊社のコミュニティ規定の他のセクションに記載されています。</p> <p>IV. 加工されたメディア</p> <p>メディアはさまざまな方法で編集することができます。多くの場合、このような変更は、芸術的な理由による切り取りもしくは短縮、または音楽の追加など、無害なものです。ただ、特に動画コンテンツの場合、中には加工の有無がはっきりせず誤解を与えてしまう可能性が生じることもあります。このようなコンテンツはすぐに拡散されやすく、専門家の助言によれば、加工されたメディアに関する誤った思い込みはさらなる議論で修正できないことが多々あるため、弊社ではこのようなコンテンツを削除します。</p> <p>次の一定の条件を満たす動画については、本ポリシーに基づき削除します。(1)明瞭さや画質・音質の調整にとどまらず、動画の被写体が実際には発言していない言葉を言ったと一般の利用者に誤解させるような編集または合成が、一般の利用者にはわからない形でされている動</p>
--	---

		<p>画、および(2)ディープラーニング技術を含む人工知能または機械学習(例:人為的なディープフェイク)によって制作され、特定の動画に他のコンテンツを統合、結合、置換、重ね合わせるなどして動画が本物に見えるように作成されている動画。</p>
	(ii)削除等の対応	<p>【①社会的混乱関係(災害等)】</p>
		<p>【②健康・医療関係(コロナ関係等)】</p>
		<p>【③選挙・政治関係】</p>
		<p>【④全般・その他】</p> <p>弊社はテクノロジーと審査チームを通じて、Facebook と Instagram で毎日何百万ものコンテンツを検出し、審査し、措置を採っています。</p> <p>違反の検出</p> <p>テクノロジーと審査チームは、違反の可能性があるコンテンツやアカウントの検出と審査を行っています。弊社のテクノロジーは、違反コンテンツの大部分を、誰かが報告する前に積極的に検出し、削除しています。エンジニア、データサイエンティスト、審査チームが協力して、この技術を長期的に更新・改善しています。一方、弊社のテクノロジーは、審査チームがコンテンツに優先順位をつけるのに役立っています。弊社は、Facebook コミュニティ規定と Instagram コミュニティガイドラインを施行するためにテクノロジーを使用しています。私たちのチームは互いに協力して、このテクノロジーを構築し、洗練させています。その仕組みは以下の通りです。詳しくは、以下の下記をご覧ください：</p> <p>テクノロジーが違反を検出する仕組み</p> <p>https://transparency.fb.com/en-gb/enforcement/detecting-violations/technology-detects-violations/</p> <p>エンフォースメントテクノロジーの仕組み</p>

	<p>https://transparency.fb.com/en-gb/enforcement/detecting-violations/how-enforcement-technology-works/ メタが技術を鍛える方法</p> <p>https://transparency.fb.com/en-gb/enforcement/detecting-violations/training-technology/ メタのテクノロジーへの投資方法</p> <p>https://transparency.fb.com/en-gb/enforcement/detecting-violations/investing-in-technology/</p> <p>講じる対応 弊社は、コンテンツへの対応に関して、「削除」「抑制」「情報提供」の3つのアプローチを採っています。詳しくは、下記をご覧ください。</p> <p>https://transparency.fb.com/en-gb/enforcement/</p> <p>コミュニティ規定をどのように実施し、何を上位に表示させるべきか、表示順位を下げるべきかを決定するかについての詳細は、下記のページをご覧ください。</p> <p>https://about.fb.com/ja/news/2023/03/metas-integrity-and-transparency-reports-q4-2022/ https://about.fb.com/ja/news/2022/12/community-standards-enforcement-report-q3-2022/ https://about.fb.com/ja/news/2022/08/community-standards-enforcement-report-q2-2022/</p> <p>上記に加えて、上記4(1)(i)の対応をご参照ください。</p>
--	---

<p>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</p>	<p>弊社では細心の注意を払って、さまざまな考え方や信条、特に、関心を寄せなければ見過ごされたり取り残されたりしてしまう可能性のある人々やコミュニティの考え方や信条を盛り込んで、ポリシーを策定しています。</p> <p>ポリシーフォーラムという会議を定期的に行い、弊社のコミュニティ規定、コミュニティガイドライン、広告ポリシーあるいは商品ポリシーの将来的な変更について話し合います。この会議では、コンテンツポリシーチームから参加した対象分野の専門家が新しいポリシーの追加または既存のポリシーの改訂を提案します。コンテンツポリシーチームにとってこの会議は、許容されることの文化的な違いを考慮しつつ、安全と表現に関する幅広い視点や、Metaのポリシーが世界中のコミュニティに与える影響をより深く理解するのに役立ちます。また、さまざまな社内ステークホルダーも会議に参加します。これには安全とサイバーセキュリティに関するポリシー、グローバルオペレーション、公民権・人権、法律、コミュニケーション、ダイバーシティといった各種分野を担当するチームからのメンバーだけでなく、テロ対策専門家や製品マネージャー、公開ポリシーに携わるさまざまなリーダーが含まれます。</p> <p>コンテンツポリシーチームは、通常、会議で注意喚起とポリシー助言の2種類のプレゼンテーションを行います。注意喚起は、社内外の意見を聞きながら、チームが取り組む予定の問題を紹介する短いプレゼンテーションです。チームが意見を聞き、関連するデータを分析し、ポリシーを更新するための選択肢を準備した後、専門家が推奨事項を提示し、より大きなグループで議論することができます。</p> <p>弊社のポリシーは、こうした会議からのフィードバックに加え、社会規範の変化、言語、製品のアップデートなどに基づいて、時間の経過とともに進化していきます。</p> <p>ポリシーフォーラムに続いて、ポリシーの変更は実施に向けて準備されますが、これには時間がかかるため、実際の開始日は異なります。ポリシーが開始されると、ポリシーフォーラムの議事録を公開し、コミュニティ規定における「変更履歴」で変更内容を確認することができます。</p> <p>https://transparency.fb.com/en-gb/policies/improving/policy-forum-minutes/</p>
---	--

		<p>上記のポリシーフォーラムに加え、弊社は以下のような外部コンサルテーションを実施しています。</p> <p>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、より幅広いステークホルダーからのインプットを取り入れたポリシーとするために尽力しています。</p> <p>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/</p>
<p>(2) 削除要請や苦情に関する受付態勢・プロセス</p>		
<p>①</p>	<p>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>	<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、特定的かつ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、およびテクノロジーを組み合わせ用いています。違反コンテンツを事前に検出し、対策を講じるための AI の機能を向上させたことで、利用者からの報告への依存度は大幅に低下しています。しかし、毎週、世界中の人々が、弊社のポリシーに違反していると思</p>

われるコンテンツを何百万件も当社に報告しています。

弊社は、利用者が私たちのポリシーに違反していると思われるコンテンツを報告しやすくするために、報告ツールの改善に努めていますが、限界もあります。例えば、デジタルリテラシーの低い地域では、利用者は報告するという選択肢をあまり意識していない可能性があります。また、利用者は、自分が嫌いなコンテンツや同意できないコンテンツを報告することがよくありますが、それは弊社のポリシーに違反するものではありません。例えば、ライブルとなるスポーツチームのコンテンツや、自分が同意できない意見を報告する利用者もいます。また、報告される前に多くの人に見られてしまうコンテンツもあるため、利用者の報告だけを頼りにすることはできません。通報方法については、下記サイトで説明しています。

<https://www.facebook.com/help/263149623790594/>

利用者／コミュニティからの報告

弊社は、濫用アカウントが使用する可能性のある新しいトレンドやテクニックに対抗するために、技術の革新と改善を続けていますが、利用者が私たちのポリシーに違反していると思われるコンテンツやアカウントを個別に報告することも可能です。弊社の一般的なアプローチで述べたように、利用者からの報告は、私たちが違反するコンテンツや行動のシグナルを得ることができる3つの柱のうちの1つです。

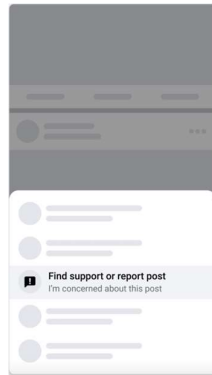
- Facebook では、ページ、グループ、プロフィール、投稿、写真、ビデオ、コメント、広告など、すべてのコンテンツを報告することができます。また、コミュニティ規定に違反していると思われるコンテンツは、誰でも弊社に報告することができます。
- Instagram でも同様に、アプリ内のフィード投稿、ビデオ、ストーリーズ、リール、コメントなど、あらゆるコンテンツを簡単に報告することができます。
- また弊社は、世界各地の市民団体とパートナーとしてのネットワークを構築しており、彼らは専用のチャンネルを通じて弊社に連絡を取り、新たな問題を知らせたり、弊社のチームが知らないような重要な情報を提供してくれます。

- 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。

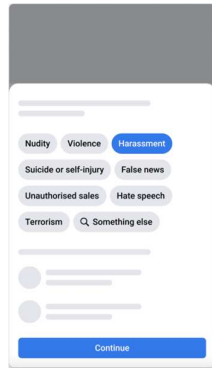
利用者からの報告に関するユーザーエクスペリエンス

Facebook に載せるべきでないと思われるものを報告した場合、Facebook 上の利用者がどのような手順をたどるのかについて、以下にその例を挙げます。

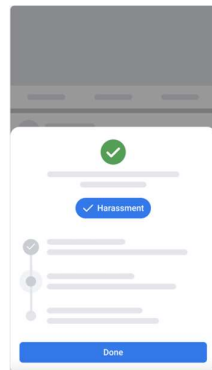
1. ユニバーサルなエントリーポイント： 投稿、コメント、ストーリー、メッセージ、その他を問わず、報告するためのオプションを用意しています。
2. はじめに： 弊社は、利用者が私たちのプラットフォーム上にあるべきでないと思うものを報告するのをお手伝いします。



3. 問題を選択： 何が問題なのか、もっと詳しく教えてくださいとお願いしています。これにより、適切な担当にレポートを送ることができます。



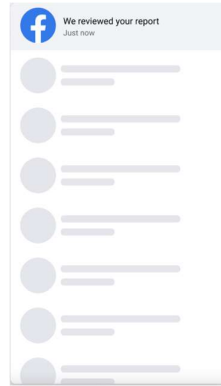
4. 報告書の提出：これらのステップを経て、報告書を提出します。また、今後どのようなこと手順を辿るかについても説明します。



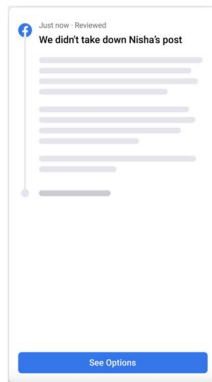
テクノロジーを使って有害なコンテンツを事前に識別

- 弊社では主に人工知能（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

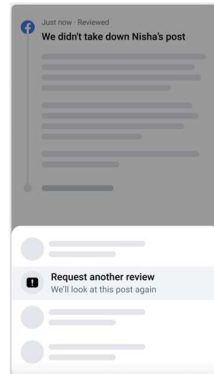
		<p>ニティ規定や Instagram のコミュニティガイドラインに違反していることに自身がある場合、(多くの場合、誰かの目に触れる前に) 自動的に削除することがあります。</p> <ul style="list-style-type: none"> ● また、AI は、弊社のプラットフォームで審査するコンテンツの量が多いため、弊社のコンテンツモデレーターが審査する案件の優先順位付けにも役立っており、最も有害で一刻を争うコンテンツを優先的に審査しています。優先順位付けは、以下のようないくつかの要素に基づいて行います。 <ul style="list-style-type: none"> ○ バイラリティ： 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。 ○ 重大性： 現実の被害に関連するコンテンツは、他のカテゴリーよりも優先的に審査されます。 ○ 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。 ● 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>https://transparency.facebook.com/community-standards-enforcement</p>
②	対応決定時における通知の内容、理由の記載の程度	<p>利用者からの報告か事前検知であるかを問わず、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を認識した場合、そのコンテンツを削除します。</p> <p>報告後のユーザーエクスペリエンスの概要は以下のとおりです：</p> <ol style="list-style-type: none"> 1. 通知による更新： 報告を確認した後、その報告を行った利用者に通知を送ります。



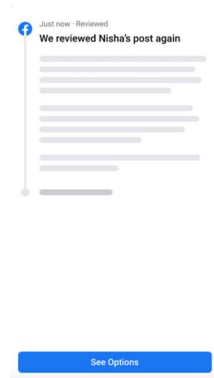
2. サポート受信箱の詳細：審査の結果の詳細については、「サポート受信箱」で共有します。この情報があることを通知し、そのリンクを送ります。



3. 再審査の機会：弊社の判断が誤っていると思われる場合は、再調査を依頼することができます。



4. 再審査後のやりとり：コンテンツの再審査を行った後、最終的な回答を再度サポート受信トレイに送信します。



また、コンテンツを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方法を理解してもらうために、コンテンツを共有した人にも通知します。この情報は、その人のサポート受信箱で見ることができます。

www.facebook.com/support

		<p>さらに、弊社が何か間違っただけをしたと思う場合、利用者が強制措置に異議を唱えることができるようにするためのプロセスも用意しています。このようなコンテンツ削除の体験の流れは、こちらも御参考にさせていただきます。</p> <p>https://transparency.fb.com/enforcement/taking-action/taking-down-violating-content/#takedown-experience</p>
③	<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>【日本の数値】 日本国内の利用者から寄せられた報告や、その結果としての執行の決定や措置は、弊社のコンテンツモデレーション業務の一貫性を確保するために、明確で原則的、運用可能で説明可能なポリシーに基づいています。コンテンツモデレーションの決定は、社内の複数のグローバルチームによる検討、協議、承認を経て行われます。</p> <p>【グローバルの数値】 弊社は、違反するコンテンツ、アカウント、または疑わしい行動を検出・特定するために、人とテクノロジーの両方に多大な投資を行ってきました。弊社の執行実務のうち、人間によるレビューに焦点を当てたものについては以下の通りです：</p> <p>人の目による審査</p> <ul style="list-style-type: none"> - 弊社では、4万人以上が安全・セキュリティに取り組んでいます。前述の通り、これはグローバルなチームであり、日本語を含む多数の言語で24時間365日コンテンツをレビューしています。 - 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、最も重要であり、私たちが採り入れ、最適化しようとしているのは、言語と文化的背景です。 - 人間の審査担当は、弊社のコミュニティ規定を公正かつ正確に実施するために不可欠であり、特にコンテンツの一部を取り巻く文脈が重要である場合に重要です <ul style="list-style-type: none"> - 例えば、ヘイトスピーチについて弊社のシステムは、ヘイトスピーチとしてよ

		<p>く使われる特定の言葉を認識できますが、その言葉を使う人の意図が必ずしもわかっているとは限りません。そこで、弊社のチームはそうしたコンテンツを審査するのです。</p>
		<p>【(質問に答えられない場合) 参考となる数値】</p>
<p>その他の対応に関する部署 やチームの内容・規模・人数</p>	<p>【日本の数値】 フェイクニュース対策など、コミュニティ規定やコミュニティガイドラインに基づくコンテンツモデレーションの決定は、日本だけの専門チームによって行われるものではありません。これらの決定は、社内の複数のグローバルチームによる検討、協議、承認を経て行われるものです。日本に専門チームを持っていても、内部システムによりこのプロセスが変更されることはありません。</p>	
	<p>【グローバルの数値】 弊社のプラットフォームで人々の安全と安心を守るために、4 万人以上の者が献身的に働いています。これには、安全、サイバーセキュリティ、人権、テロ対策、社会科学、ローカルマーケット（地域や国など）の専門家チームにより構成されています。弊社のコンテンツ審査担当は、さまざまな文化的背景を持ち、弊社のコミュニティの多様性を反映し、さまざまな職業経験を有しています。また、社会的・文化的背景を提供し、コミュニティ規定の執行システムの運用能力を向上させるために、マーケットスペシャリストを採用しています。ドイツ、アイルランド、ラトビア、スペイン、リスボン、フィリピン、米国など、世界中に 20 以上の拠点があり、これらのチームが日本語を含む数十カ国語のコンテンツを審査しています。 また、2016 年以降、安全・安心のために 130 億米ドル以上を投資しており、2021 年には 50 億米ドル以上を費やしました。</p>	
		<p>【(質問に答えられない場合) 参考となる数値】 弊社のレビューチームは、より大きな責任の一部としてコンテンツをレビューする正社員と、</p>

		<p>パートナーに雇用されたコンテンツ審査担当で構成されています。彼らは様々なバックグラウンドを持ち、多様なコミュニティを反映し、退役軍人から法律の専門家、子どもの安全、ヘイトスピーチ、テロ対策などの政策分野の執行専門家まで、様々な職業経験を持っています。</p>
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	<p>上記のとおり、弊社のコンテンツ審査チームは、世界 20 カ所の拠点到分散しており、人間の審査員も含まれています。</p> <p>コンテンツ・モデレーション・チームとそのスキルや専門性については、上記 (3) と (4) をご参照ください。</p>
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	<p>弊社は、コンテンツに対する対応を決定する際に、常に正しい判断を下せるわけではないことを認識しています。アカウントやコンテンツに対して採った対応に誤りがあったと考える場合、ほとんどの状況で、利用者には不服申し立ての権利が提供され、利用者は再審査をリクエストすることができます。</p> <p>Facebook で決定を不服とする場合、利用者は、コンテンツが削除された、または警告で覆われたと私たちが通知した後、「再審査をリクエストする」というオプションを選択します。</p> <p>決定が不服申し立ての対象となる場合、利用者は、自分のコンテンツが削除された、または警告が付されたという通知を受け取った後、弊社にもう一度見てもらうよう依頼するオプションが与えられています。異議申し立てがあった場合、弊社はその投稿を再度確認し、コミュニティ規定に沿っているかどうかを判断します。このプロセスは、弊社が間違いを犯したと思う場合に、それを知らせてくれるものであり、公正なシステムを構築するために不可欠なものです。</p> <p>https://www.facebook.com/help/2090856331203011</p>

Meta の異議申し立てプロセスの仕組み

誰かが公開した投稿について、弊社ポリシーへの違反を理由として Facebook から削除することを弊社が決定したとします。投稿者には、この決定が通知されるとともに、この決定に同意するか、または同意せずに再審査をリクエストする選択権が与えられます。

決定に同意しない場合、コンテンツが再審査されます。再審査中は、Facebook 上で他の利用者がそのコンテンツを見ることはできません。

審査担当者が元の決定に同意した場合、そのコンテンツは Facebook から削除されたままになります。審査担当者が当初の審査結果に同意せず、削除すべきでない判断した場合、当該コンテンツは別の審査担当者によって再審査にかけられます。この審査担当者の判断によって、コンテンツが Facebook 上に残されるべきか否かが決まります。

<https://transparency.fb.com/en-gb/policies/improving/appealed-content-metric/>

異議申し立ての対象

現在、Facebook および Instagram 上の大多数の違反タイプについて異議申し立ての機会が与えられています。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。弊社では、スタッフによる審査とテクノロジーを組み合わせ、異議申し立てのリクエストを処理しています。混雑時は、弊社の審査能力により、必ずしもすべてを審査できない場合があります。

弊社では、措置が講じられたコンテンツに関する異議申し立てだけでなく、報告されたものの措置が講じられなかったコンテンツに関する異議申し立ての機会も提供しています。

Facebook や Instagram からコンテンツが削除され、その人が弊社の再審査プロセスをも経た場合、監督委員会に上訴することでこの決定に異議を唱えることができます。弊社は、オン

		<p>ライン上の表現の自由をめぐる最も難しい問題、すなわち、何を削除し、何を残すか、そしてなぜ残すかを解決するために監督委員会を設立しました。</p> <p>この委員会の目的は、Facebook や Instagram における弊社のコンテンツに係る措置の決定に同意できない場合、利用者が訴えることができる独立した外部機関を持つことです。弊社は、この委員会によるコンテンツに係る決定を実行することを約束しており、彼らの勧告は、弊社のポリシーの管理方法を形作るのに役立つものと考えています。監視委員会の詳細については、下記 URL を参照してください。</p> <p>https://oversightboard.com/news/833880990682078-the-oversight-board-is-now-accepting-cases/</p>
(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】
		【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】
		<p>【グローバルの数値】</p> <p>弊社は、Facebook と Instagram が、利用者が声をあげられる場所であってほしいと願っています。誰もが気持ちよく自己表現できる状況を作るためには、利用者の安全、プライバシー、尊厳、正真性を守ることも必要です。そのため、Facebook コミュニティ規定と Instagram コミュニティガイドラインを設け、コミュニティで何が許され、何が許されないかを定義しています。</p>

		<p>弊社は、より効果的に進捗状況を把握し、Facebook と Instagram を安全で包括的なものにするという私たちの継続的な取り組みを示すために、四半期ごとにコミュニティ規定施行報告書を発行しています。この報告書には、ポリシーに違反するコンテンツの防止と対策について、弊社がどのように取り組んでいるかが詳細に記載されています。</p>
		【(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】
		<p>【グローバルの数値】</p> <p>弊社は、より効果的に進捗状況を把握し、Facebook と Instagram を安全で包括的なものにするという弊社の継続的な取り組みを示すために、四半期ごとにコミュニティ規定施行レポートを発行しています。このレポートでは、いわゆる偽アカウントに対するアクション率など、ポリシーに違反するコンテンツの防止と対策について、弊社がどのように取り組んでいるかを詳細に説明しています。</p> <p>例えば、2023 年 2 月に発表されたレポートでは、2022 年第 4 四半期に 13 億の偽アカウントに措置を講じたことが報告されています。</p>
		【(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】
		<p>【グローバルの数値】</p> <p>パンデミックが始まって以来、弊社は新型コロナウイルス感染症とワクチンに係る Misinformation に関するポリシーに違反したコンテンツを全世界で 2400 万個以上削除して</p>

		<p>きました。また、COVID-19 やワクチンの Misinformation を拡散させないという弊社のルールに繰り返し違反したアカウント、ページ、グループを 3,000 以上削除しています。</p> <p>【(質問に答えられない場合) 参考となる数値】</p> <p>弊社のコミュニティ規定施行報告書には、事前対応率が記載されており、これは、ポリシーに違反するとして削除または無効化したコンテンツやアカウントを総数のうち、利用者から報告される前に弊社が発見し、フラグを立てたものの割合を示すものです。この指標は、私たちがいかに効果的に違反を検知しているかの指標として使用しています。</p>
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)	<p>人工知能は、有害なコンテンツから利用者を保護するための重要なツールです。人工知能は、人間の専門家の作業をスケールアップし、問題のある投稿やコメントが利用者に危害を加える前に、事前に行動を起こすのに役立ちます。</p> <p>弊社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の決定にも役立てています。</p> <p>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 日当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。</p> <p>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。</p> <p>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。</p>

		https://transparency.fb.com/ja-jp/enforcement/ さらに、有害コンテンツに対応するための AI システムについては、下記 URL をご参照ください。 https://about.fb.com/news/2021/12/metas-new-ai-system-tackles-harmful-content/
⑤	削除以外の取組	
	(i) 警告表示	
	(ii) 表示順位の低下	
	(iii) その他の取組内容	<p>Misinformation に対する弊社が採るアプローチは、表現の自由とバランスを取りながら、正確で有益なコンテンツを利用者に提供すべきだという原則に基づいています。利用者は、私たちのプラットフォームで質の高いコンテンツを見たいと思っていますし、弊社としても同じ思いです。そのため、弊社の misinformation 対策は、「削除」「抑制」「情報提供」の3つのパートで構成されています。以下のサイトに詳細が記載されています。</p> <p>https://about.fb.com/news/2019/04/remove-reduce-inform-new-steps/</p> <p>この戦略では、弊社のポリシーに違反するコンテンツを削除し、ポリシーに違反しないものの、プラットフォームの信頼性を損なうような問題のあるコンテンツの拡散を抑え、追加情報を提供して、利用者にクリック、閲覧、共有するのかわを選択できるようにします。</p> <p>虚偽の情報の配信の抑制</p> <ul style="list-style-type: none"> フィードでの虚偽の情報の表示順位を下げる：ファクトチェック団体が記事を False と評価すると、フィード上で該当する写真、動画、テキスト、リンクの表示順位が下がります。これにより、それを見る利用者の数を大幅に抑制できます。

- 繰り返し虚偽の情報を配信する違反者に対するアクション：繰り返し虚偽の情報を配信するページやウェブサイトに対しては、配信量の抑制や広告機能の停止の措置をとります。
- 虚偽の情報のコピーを見つけるテクノロジーを使用する：Facebook 上には情報のコピーが何千も存在している可能性があり、それらは写真のトリミングなど、若干の違いがある場合もあります。弊社は、機械学習テクノロジーを使用してこのようなコピーを検出することにより、ファクトチェック団体が新しい情報に集中できるようにしています。

虚偽の情報に関する追加情報の提供

- Misinformation のラベル：Facebook や Instagram のコンテンツで、ファクトチェック団体によって False または Partly false と評価されたものは、目立つようにラベル付けされ、利用者は何を読み、信頼し、共有するかを自分で判断できるようになっています（2020 年の米国選挙における弊社のラベル付け取り組みについては、こちらに詳述されています。）。
<https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/>
- 虚偽の情報に関する補足情報の提供：ファクトチェック団体が補足記事を作成した場合、表示された通知をクリックして理由を確認することができます。

<https://www.facebook.com/journalismproject/programs/third-party-fact-checking/new-ratings>

- 虚偽の情報をシェアした利用者への通知：虚偽の情報を含む投稿をシェアしようとした場合や、過去にシェアしたことがある場合、弊社から通知が送付されません。虚偽の情報を含む投稿をシェアしたページの管理者にも通知が送付されません。

- ファクトチェック団体による評価の種類：独立した第三者ファクトチェック団体による評価の種類について、詳しくはこちらをご覧ください。各評価に該当するコンテンツのガイドラインと例についても、下記 URL でご確認いただけます。

<https://www.facebook.com/help/publisher/341102040382165>

正確で信頼性の高い情報と利用者を結びつける

- 弊社は、正確で信頼できる権威ある情報を利用者に伝えるための新しい方法を探し続けています。これは、misinformation と戦うための私たちの戦略の中核をなすものです。なぜなら、私たちは、何を読み、何を信頼し、何を共有するかを決定する手段を利用者に提供できるようにしたいからです。弊社は、新型コロナウイルス感染症情報センターのようなページを立ち上げて、トピックに関する信頼できる正確な情報と利用者を結びつけています。
- **利用者が共有するコンテンツについて、より多くのコンテキストの提供**
 - 弊社の目標は、利用者がタイムリーで信頼性が高く、自分にとって最も価値のあるコンテンツを簡単に特定できるようにすることです。そのため、利用者が何を読み、何を信頼し、何を共有するかを決めるのに役立つ、より多くのコンテキストを提供するツールを導入しています：
 - 共有しようとしているニュース記事が 90 日以上前のものである場合、通知画面によりその旨お知らせします。その記事がまだ重要であると判断すれば、利用者は共有を継続することができます。特にニュース出版社は、古い記事が最新のニュースとしてソーシャルメディアで共有され、時事問題の状況を誤解させることに懸念を表明しています。
 - 新型コロナウイルス感染症に関連するリンクを共有しようとする際に、通知画面により、より詳細な情報を提供することができます。また、信

頼できる情報にアクセスできるように、新型コロナウイルス感染症情報センターへの誘導も行っています。

- **ニュース記事について、より詳細な情報の提供**

- 投稿にあるボタンをクリックすると、出版社の Wikipedia の項目からの情報、出版社のページをフォローするボタン、出版社の他の記事、Facebook でその記事がどのように利用者に共有されているかについての情報など、Facebook や他のソースから引き出された文脈的な情報を得ることができます。場合によっては、その情報が利用できない場合、弊社は利用者に知らせますが、それも有益な文脈となりえます。詳細はこちらをご覧ください。

<https://about.fb.com/news/2018/04/news-feed-fyi-more-context/>

- **虚偽の主張を繰り返し共有するページへの警告**

- misinformation をシェアしているページを「いいね！」しようとしている人に警告を出し、そのページをフォローするかどうか、利用者が十分な情報を得た上で判断できるようにします。警告には、そのページが虚偽の情報を共有していることを示すメッセージなど、追加の文脈情報へのリンクが含まれています。有害な虚偽の主張の共有が繰り返されると、最終的に私たちのプラットフォームからそのページが完全に削除される可能性があります。

- **コミュニティ規定に繰り返し違反するグループに対する警告**

- コミュニティ規定の違反があるグループに参加しようとしている利用者に警告を発し、参加前に十分な情報を得た上で判断してもらえるようにします。このようなグループへの招待通知を制限し、参加する可能性を低くします。既存のメンバーには、そのグループのコンテンツの配信を減らし、フィードの下位に表示されるようにします。

虚偽の情報を識別し、フィードバックする手段の提供

- 虚偽の情報の特定方法を確認する：何に注意すべきかを理解することで、どんな記事を読み、信頼し、シェアすべきかを事実に基づいて判断できるようになります。こちらにて詳細をご覧ください。
<https://www.facebook.com/help/188118808357379>
- フェイクニュースだと思ふ投稿についてフィードバックをお寄せください：虚偽と思われる投稿があればお知らせください。虚偽の情報を識別するための基準の1つとして使用させていただきます。こちらにて詳細をご覧ください。
<https://www.facebook.com/help/572838089565953>
- ファクトチェック団体によるコンテンツの評価に異議を申し立てる：評価後にコンテンツを修正した場合や、コンテンツのファクトチェックに問題があると思われる場合は、ファクトチェック団体にお知らせください。

クリティカルシンキングのためのデジタルリテラシー向上に係る取組み

- 弊社は、オフラインでもさまざまなプログラムに投資し、デジタルとニュースのリテラシーという根本的な問題に取り組むために、市民団体や業界のパートナーとさまざまな取り組みを行っています。弊社の代表的なプログラムのひとつである「みんなのデジタル教室」は、人々が批判的に考え、思慮深くオンラインで共有できるようにするための短いレッスンや学習モジュールのコレクションです。すぐに使えるオンライン・モジュールは、アジア太平洋地域のオンライン・セーフティの専門家と共同で開発されました。現在、18カ国、15言語（日本語を含む）で提供されています。

<https://wethinkdigital.fb.com/>

<https://www.facebook.com/help/1952307158131536>

<https://transparency.fb.com/ja-jp/features/approach-to-misinformation/>

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】
		<p>【(質問に答えられない場合) 参考となる数値】</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濫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濫用とみなされ、弊社のコミュニティ規定及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>
(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	<p>弊社は、安全性とセキュリティのリスクをもたらさない限りにおいて、安全性と完全性に関連するポリシー、プロセス、システムについて、透明性を確保することを約束します。透明性を確保することで、プラットフォームの取り組みが世間の注目を浴び、ひいては弊社が下す決定に対して責任を負うことになるため、説明責任を果たすことができると考えています。</p> <p>この透明性へのコミットメントに基づき、弊社は、事前検出とコンテンツモデレーションに使用されているAIシステムとモデルについて、何度も公表してきました。</p> <p>例えば、Misinformationを検出するためにAIをどのように活用しているかについて https://about.fb.com/news/2021/12/metas-new-ai-system-tackles-harmful-content/</p> <p>これに加えて、有害なコンテンツのさまざまなカテゴリに対する事前検出率も公開してお</p>

		<p>り、四半期ごとのコミュニティ規定施行レポート（CSER）で公表しています。</p> <p>弊社はまた、利用者が、自分が見るコンテンツやオンラインでの経験ややり取りをよりよくコントロールし、情報を得ることができるようにすることにも取り組んでいます。</p> <p>弊社は、オンラインの安全性と、misinformation や disinformation を含む有害なコンテンツに対処する最も効果的な方法は、利用者が何を読み、信頼し、共有するかを自分自身で批判的に決定できるようにするツールとリソースを提供することによって、リジリエントなデジタル社会を構築することだと信じています。弊社は、利用者により一層の透明性とコントロールを提供し、情報に基づいた意思決定に役立つ情報を提供し、メディアとデジタルリテラシーを向上させることによって、これを実現します。</p>
	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	<p>コミュニティ規定施行レポートは日本語に翻訳されてます。</p> <p>https://about.fb.com/ja/news/2023/03/metas-integrity-and-transparency-reports-q4-2022/</p> <p>https://about.fb.com/ja/news/2022/12/community-standards-enforcement-report-q3-2022/</p>
③	取組の効果分析	<p>企業が説明責任を果たし、より良い活動を行うためには、独立した監査や評価が不可欠です。そのために、弊社は2つの重要な取組みを行っています：</p> <ol style="list-style-type: none"> 1. データ透明性アドバイザーグループ（DTAG）（計測、統計、法律、経済、ガバナンスの国際的な専門家グループ）による2019年の独立した公開評価では、弊社が共有する指標が有用で、健全な手法に基づき、プラットフォームで起こっていることを正確に反映しているかどうかを評価しました。DTAGは、弊社が正しい指標を見ていることに大筋で同意し、審査がどのように処理されたかを知ることができる手続の整備を含む改善のためのいくつかの提案を行いました。

		コミュニティ規定施行レポートが正しく測定され報告されているかどうかを検証するため、アーンスト・アンド・ヤング社（EY）による独立監査を実施しました。この監査でEYは、同レポートの指標の計算は正確であり、公正に記載されており、内部統制が適切に設計され、有効に機能していることを確認しました。
＜ 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	<p>弊社は、クリエイターやパブリッシャーがコンテンツから収益を得るための機能や製品を提供しています。このコンテンツは、一定のルールに従わなければなりません。クリエイターとパブリッシャーは、コンテンツから収益を得るために、弊社のパートナー・マネタイゼーション・ポリシーとコンテンツ・マネタイゼーション・ポリシーを遵守する必要があります。</p> <p>このポリシーでは、misinformation、すなわち第三者のファクトチェック団体によってFalseと評価されたコンテンツを禁止しています。</p> <p>https://www.facebook.com/business/help/1348682518563619?id=2520940424820218</p>
②	広告出稿制限	<p>弊社は、第三者ファクトチェック団体によって否定された主張、または特定の状況において、特定の専門知識を持つ組織によって否定された主張を含む広告を禁止しています。Falseとみなされる情報を繰り返し投稿する広告主は、プラットフォームでの広告掲載に制限を受ける可能性があります。</p> <p>https://www.facebook.com/policies/ads/prohibited_content/misinformation</p>
③	ターゲティング技術の適用に関する規定	<p>広告ポリシーにおいて以下の定めを置いています。</p> <p>7. ターゲット設定</p> <ol style="list-style-type: none"> 1. 利用者を差別、侮辱、挑発、攻撃する目的で、または略奪的な広告活動を行う目的でターゲット設定オプションを使用してはいけません。 2. カスタムオーディエンスを広告のターゲットにする場合は、オーディエンス作成時に利用規約に準拠する必要があります。

		https://www.facebook.com/policies/ads
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	<p>パーソナライズドランキングを採用する理由</p> <p>弊社は、最も重要な人、最も重要なもの、つまり、皆さんの知人・友人、興味、そしてあなたとつながりのある世界を一か所に集めることができます。フィードを閲覧しているとき、検索で特定のものを探しているとき、グループでコミュニティとつながっているとき、マーケットプレイスで取引を探しているとき、Watch でビデオを見ているときなど、利用者にとって最も価値のあるコンテンツを確実に見ることができるようにすることが私たちのゴールです。そのために、弊社は、利用者が最も個人的に興味があると思われるものを各サーフェスの最上部に表示することを目的として、表示されるすべてのコンテンツを配置しています。</p> <p>つまり、Facebook でコンテンツに出会ったとき、その表示方法は、利用者一人一人に合わせてカスタマイズされていることが多いのです。利用者がつながりを持った友達、参加したグループ、「いいね！」したページ、付したコメント、「この投稿は時間を費やす価値がありますか」などのフィード調査に残した回答はすべて、利用者にとって何が最も重要であるかを示すインプットであり、弊社が使用するアルゴリズムは、このすべての活動に基づいて利用者が最も興味を持ちそうなものを予測することによって何を表示すべきかを決定しています。このパーソナライゼーションにより、利用者は積極的に参加するという経験を得ることができるようになります。</p> <p>フィードアルゴリズムの仕組み</p> <p>フィードに表示される投稿は、利用者にとって最も価値があると思われるものに基づいてランク付けされています。フィードは、利用者がつながりを持つことを選択した友達、グループ、ページからの投稿で主に構成されています。ほとんどの人が1回のセッションで閲覧できないほど多くのコンテンツをフィードに登録しているため、弊社はアルゴリズムを使って、</p>

	<p>利用者が見ることができるすべての投稿の順番を決定しています。最も意味のある投稿をフィードの上部に優先的に表示するために、アルゴリズムは4つのステップで動作します:「インベントリ」「シグナル」「予測」「関連性」です。</p> <p>フィードの中で利用者にとって最も価値のあるものを予測するためにランキングを使用する方法については、下記 URL を参照してください。</p> <p>https://about.fb.com/news/2021/01/how-does-news-feed-predict-what-you-want-to-see/</p> <p>逆に、利用者が見たくないと言った投稿や、有害であると広く理解されている投稿もあります。そのため、過激な暴力、ヘイトスピーチ、偽の新型コロナウイルス感染症治療薬など、実際に害を及ぼすリスクがある場合は、プラットフォームからコンテンツを完全に削除するように努めています。弊社が有害なコンテンツを禁止するコミュニティ規定を設けているのも、有害なコンテンツを迅速に特定し対処する方法を開発するために多額の投資を行っているのも、このためです。また、フィードランキングを利用して、利用者が不快に思うコンテンツを含む可能性のある投稿の配信を減らしていますが、必ずしも私たちのポリシーに基づく削除の基準を満たすものではありません。例えば、misinformation、センセーショナルな健康上の主張、クリックベイトなどを含む可能性が高い投稿は、低い価値スコアを受け取り、その結果、フィードで下位に表示されます。弊社は、このような新しいタイプの問題あるコンテンツを監視し、それらを検出し、強制するための新しいアルゴリズムを開発しています。しかし、私たちは、増え続ける利用者により利用されるプラットフォームがもたらす課題に直面しています。そのため、弊社のチームは日々、適応に努め、生じる可能性のあるギャップには、思慮深く慎重な方法で対処しています。</p> <p>https://transparency.fb.com/features/ranking-and-content/</p>
--	--

		<p>コンテンツ配信ガイドライン</p> <p>コンテンツ配信ガイドラインでは、第三者によるファクトチェックで False と評価された投稿、ボーダーライン又はポリシー違反の可能性がある投稿、低品質または扇情的な投稿、報告または非表示にされる可能性が高いコメントなど、配信を縮小するコンテンツに関する方針を詳しく説明しています。降格させるコンテンツの全リストは、こちらで下記 URL をご参照ください。</p> <p>https://transparency.fb.com/ja-jp/features/approach-to-ranking/types-of-content-we-demote/</p> <p>レコメンデーションガイドライン</p> <p>弊社は、サービスを利用する方々が新しいコミュニティやコンテンツを発見するのに助けるために、「おすすめ」を行います。Facebook と Instagram の両方が、利用者がまだフォローしていないコンテンツ、アカウント、エンティティを推薦することがあります。弊社のレコメンデーション体験の例としては、Instagram Explore、Accounts you may like、Reels タブなどがあります。</p> <p>弊社の目標は、それを見る一人ひとりにとって適切で価値のある「おすすめ」を作ることです。そのためには、「おすすめ」をパーソナライズすること、つまり、利用者一人ひとりにとってユニークな「おすすめ」を作成することが必要です。例えば、Instagram でレストランや書店と交流する場合、私たちは食べ物、レシピ、本、読書に関するコンテンツをおすすめすることがあります。</p>
	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	広告ライブラリでは、Meta のプロダクトに掲載されている広告を検索することができ、表示

	開	<p>された広告に関する情報を確認できます。</p> <p>社会問題、選挙または政治に関連する広告の場合、アクティブではない広告(すでに Meta 製品に掲載されていない広告)も検索できます。選挙への介入を防ぐためには透明性が重要であるため、広告ライブラリには広告費用の出資者、広告費の金額範囲、広告がリーチした利用者層など広告についての追加情報も表示されます。これらの広告は、ライブラリに 7 年間収録されます。</p> <p>https://www.facebook.com/help/259468828226154</p>
⑥	広告とコンテンツの分離	<p>弊社では、コンテンツと広告について別々のポリシーを設けています。</p> <p>広告基準</p> <p>広告基準は、弊社が許可する広告コンテンツの種類と、禁止する広告コンテンツの種類に関するポリシーの詳細とガイダンスを提供しています。広告主が広告を発注する際、個々の広告は弊社のポリシーに照らして審査されます。また、広告基準では、ビジネスアカウントまたはそのアセット（広告アカウント、ページ、ユーザーアカウント）に対して広告の制限を設ける可能性のある広告主の行動に関するガイダンスも提供しています。</p> <p>コミュニティ規定/コミュニティガイドライン</p> <p>弊社のポリシーは、提供するテクノロジーを用いる上で何が許され、何が許されないかを定義しています。もしコンテンツが弊社のポリシーに反する場合、私たちはそれに対して行動を起こします。</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	<p>Misinformationに関する弊社のポリシーは、プラットフォームから削除されるさまざまな種類の misinformation を利用者に知らせ、よりよく知ってもらうために公開されました。詳細については、下記 URL を御参照ください。</p> <p>https://transparency.fb.com/en-gb/policies/community-standards/misinformation/</p>

		<p>さらに、弊社のプラットフォームへのアクセスを制限するダイナミクスを監視しながら、広範なポリシー、misinformationを含むポリシーの執行方法、政府からのデータ要求への対応、知的財産の保護について可視化する透明性センターを構築しました。</p> <p>詳しくは下記 URL を御参照ください。 https://transparency.fb.com/en-gb/features/approach-to-misinformation/</p>
<p>6. 「ファクトチェックの推進」関係</p>		
<p>①</p>	<p>ファクトチェック結果の表示 具体的な仕組み・基準</p>	<p>プログラムの重要なステップは下記のとおりです。</p> <p>Misinformation を特定:利用者からのフィードバックなどに基づいて misinformation の可能性のあるコンテンツを特定し、ファクトチェック団体に表示します。また、ファクトチェック団体自身がコンテンツを特定した上で審査する場合があります。</p> <p>コンテンツを審査:ファクトチェック団体がコンテンツを審査し、事実を検証した上でその正確性を評価します。これは弊社からは独立して行われ、発信元や公開データを参照したり、動画と画像を認証したりなどの作業が含まれます。</p> <p>Misinformation を明確にラベル付けして利用者へ周知: ファクトチェックパートナーに審査されたコンテンツにラベル付けを行い、利用者が追加の背景情報を閲覧できるようにします。また、ラベル付けされたコンテンツをシェアしようとする利用者や過去にシェアした利用者に通知します。</p> <p>Misinformation を目にする利用者の数を抑制: ファクトチェック団体が「False」、「Altered」、「Partly false」と評価したコンテンツは、ニュースフィードでの表示順位が下がり、Instagram での発見タブの表示対象から除外され、フィードやストーリーズで目立たないようになります。これにより、misinformation を見る利用者の数を大幅に抑制できます。また、ファクトチェック団体によって評価されたコンテンツの広告は却下されます。</p> <p>Misinformation を繰り返し配信する違反者への措置: 「False」または「Altered」と評価さ</p>

		<p>れた misinformation を繰り返し配信するページやウェブサイトに対して、配信数の抑制などの制限措置を実施します。また、それらのページやウェブサイトは収益化や広告に関する機能へのアクセスやニュースページとしての登録が一定期間取り消されます。</p> <p>https://www.facebook.com/business/help/2593586717571940</p> <p>最終的にはファクトチェック団体が独自にコンテンツの審査と評価を行います。弊社が評価を変更することはありません。</p>
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	<p>弊社のテクノロジーは、利用者からの報告や不信感を持ったコメントなど、さまざまなシグナルに基づいて、誤解を招く可能性が高い投稿を検出することができます。</p> <p>また、ファクトチェック団体は、自らレビューすべきコンテンツを特定することも可能です。</p>
③	ファクトチェックを実施する人材の育成	<p>弊社は、世界で 90 以上の独立した第三者ファクトチェック団体と提携し、60 以上の言語に対応しています。アジア太平洋地域では、インドネシア、シンガポール、マレーシア、香港、ミャンマー、タイ、スリランカ、バングラデシュ、パキスタン、モンゴル、カザフスタン、韓国、台湾、インド、フィリピン、オーストラリア、ニュージーランド、太平洋諸島のファクトチェック団体が含まれています。</p> <p>https://www.facebook.com/journalismproject/programs/third-party-fact-checking/partner-map</p> <p>近年では、Facebook や WhatsApp からの助成金付与など、ファクトチェックに係るコミュニティへの支援を拡大しています。これは、大きな影響を受けた地域の第三者ファクトチェック団体が、この重要な業務を行う際の能力を高めるための支援です。</p> <p>https://www.facebook.com/journalismproject/coronavirus-grants-fact-checking</p>

		<p>さらに、最新の取り組みとして、IFCN と弊社は、ファクトチェック入門コースを無料で提供することになりました。このコースは、ファクトチェック、検証及び虚偽の暴露、健康関連の mis/disinformation に焦点を当てた 3 つのモジュールで構成され、ファクトチェック可能な主張を見つける方法、ファクトチェックの方法、それらを支援するツールやテクニックについて、ジャーナリストをガイドするものです。</p> <p>https://www.poynter.org/shop/fact-checking/fact-checking-fundamentals-japanese/</p> <p>このコースは、アジア太平洋地域のジャーナリスト、学者、ファクトチェッカー志望者を対象とした自習型のコースです。このコースを修了した参加者には修了証が発行されます。</p>
④	日本におけるファクトチェック推進に向けた取組	<p>弊社では、Facebook と Instagram における misinformation の拡散防止に真摯に取り組んでいます。多くの国と地域で、この種のコンテンツの識別、審査および措置の遂行のために、国際ファクトチェックネットワーク (IFCN) に認証された独立した第三者ファクトチェック団体と提携しています。</p> <p>また</p> <p>https://www.facebook.com/formedia/mjp</p> <p>また、上記ファクトチェック入門コースの普及に努めてまいります。</p>
⑤	ファクトチェック機関との連携	
7. 「ICT リテラシー向上の推進」関係		
①	普及啓発の取組・投資	<p>みんなのデジタル教室／デジタル・シチズンシップ</p> <p>デジタルにレジリエントな社会を構築するためには、利用者が情報を批判的に評価し、情報に基づいた意思決定を行い、自ら間違いを正すことができるようになることが重要です。</p>

	<p>オンライン上で遭遇する可能性のある脅威が数多くあることを考えると、デジタル環境を安全かつ知的に渡っていくためには、デジタルやメディアのリテラシーにとどまらない幅広いスキルとコンピテンシーが必要です。利用者が必要なスキルを身につけるには、批判的思考やメディアリテラシーだけでなく、デジタル社会に安全に、敬意を持って、責任を持って参加する方法を学ぶ、総合的なアプローチが必要です。</p> <p>デジタルシチズンシップの育成は、社会全体の責任であり、特定のステークホルダーが達成できるものではありません。政府から産業界、市民団体、教育者、市民自身まで、あらゆるステークホルダーを巻き込んだ学際的な戦略が必要です。弊社は、アジア太平洋地域の専門家と協力し、Facebook と Instagram でオンラインいじめやハラスメント対策などのリソースを提供する「みんなのデジタル教室」を立ち上げ、デジタル社会のためのスキルを備えた責任あるデジタルシチズンのグローバルコミュニティを構築しています。</p> <p>今回、日本での「みんなのデジタル教室」プログラムの開始にあたり、Facebook Japan は、企業とのコラボレーションによる授業づくりを得意とする NPO 法人企業と教育の会をパートナー団体として招き、日本の教育現場のニーズに合わせたコンテンツを共同制作しています。</p> <p>例えば、誰かを騙すためや冗談で投稿された情報は、すぐに拡散し多くの人に影響を与えます。ここでは、そうした「フェイクニュース」の動機や、情報を受け取る側の視点について考えてみます。「偽ニュースの見分け方」では、偽ニュースを見抜くための様々なポイントをアクティビティを通じて学びます。また、偽ニュースの拡散を防ぐためのヒントも学びます。</p> <p>これまでに 24000 人以上の生徒がみんなのデジタル教室の講座を受講し、そのうち「偽ニュースの見分け方」を受講した生徒の 90%以上が「授業を受けて、インターネットやアプリ、SNS への関心が高まった」と回答しています。</p> <p>https://wethinkdigital.fb.com/jp/ja-jp/ https://about.fb.com/ja/news/2020/12/we_think_digital/</p>
--	--

		<p>また、新型コロナウイルス感染症に関する misinformation から身を守るために必要な知識を共有するために、「新型コロナウイルス感染症に関する誤情報に対処するための 6 つのヒント」を展開しています。</p> <p>https://fightcovidmisinfo.com/japanese/</p>
②	他のステークホルダーとの連携・協力・投資	<p>上記「みんなのデジタル教室」では、日本では特定非営利活動法人企業教育研究会の協力のもと、デジタルリテラシーに関するオンライン出張授業を提供しています。</p>
8. 「研究開発の推進」関係		
①	AI 技術に関する研究開発	<p>偽のニュースは、世界中で混乱と破壊的な結果をもたらしています。弊社には重要な責任があり、それに応えるためにやるべきことがたくさんあることもわかっています。</p> <p>弊社が misinformation への対策を強化すればするほど、悪意のある者は回避しようとし続けるでしょう。弊社は彼らの一歩先に行く必要がありますが、これは弊社だけではできません。弊社は AI 研究チームと協力し、学者から学び、第三者ファクトチェック団体とのパートナーシップを拡大し、他の組織（他のプラットフォームを含む）と協力する方法について話し合っています。</p> <p>https://about.fb.com/news/2018/05/hard-questions-false-news/</p> <p>前述のように、弊社は世界中の研究者との提携にも投資し、彼らの発見や学習が、私たちのプラットフォームにおける misinformation の蔓延と拡散を緩和し防止するための弊社の取り組みにさらに役立つよう、賞や助成金を提供しています。</p> <p>詳しくは、以下の URL をご覧ください。</p> <p>https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-</p>

		misinformation/
②	「ディープフェイク」対策の研究開発	<p>「操作 (manipulation)」には、Photoshop のような単純な技術から、人工知能や “ディープラーニング” の技術を使った通常 “ディープフェイク” と呼ばれる現実を歪める動画を作成する高度なツールまであります。後者について、こうした動画はインターネット上ではまだ稀ですが、その利用が増えるにつれ、この業界や社会にとって大きな課題となっています。</p> <p>ディープフェイクとあらゆるタイプの操作されたメディアの両方にどのように対処しているかについていえば、弊社のアプローチは、AI が生成したコンテンツや偽アカウントのような欺瞞的な行動の調査から、学界、政府、産業界と連携して、こうした取り組みの背後にいる人々を暴くことまで、いくつかの要素を含んでいます。</p> <p>世界中で、技術、政策、メディア、法律、市民団体活動、学術的背景を持つ 50 人以上の世界的な専門家と対話を進め、弊社の政策開発に情報を提供し、操作されたメディアを検出する科学を向上させてきたのです。</p> <p>弊社は、操作されたコンテンツの特定にも取り組んでいますが、その中でもディープフェイクは検出が最も難しいものです。そのため、2019 年には「Deep Fake Detection Challenge」を立ち上げ、世界中の人々がディープフェイクを検出するためのより多くの研究やオープンソースのツールを生み出すことに拍車をかけています。1000 万米ドルの助成金で支援されたこのプロジェクトには、Partnership on AI、コーネル大学、カリフォルニア大学バークレー校、MIT、WITNESS、マイクロソフト、BBC、AWS など、市民社会やテクノロジー、メディア、学術のコミュニティに所属する複数の組織の横断的な連合が参加しています。</p> <p>https://ai.facebook.com/datasets/dfdc/</p> <p>これとは別に、世界最大のマルチメディア・ニュースプロバイダーであるロイターと提携し、</p>

		<p>無料のオンライントレーニングコースを通じて、世界中のニュースルームがディープフェイクや操作されたメディアを識別できるよう支援する取り組みを行っています。報道機関は、大量の画像や動画をサードパーティに依存することが増えており、操作されたビジュアルを識別することは重要な課題となっています。本プログラムは、この作業を行おうとするニュースルームを支援することを目的としています。</p> <p>https://www.reuters.com/manipulatedmedia/ja/</p> <p>2021年、弊社はミシガン州立大学（MSU）と共同で、ディープフェイクの検出・帰属の研究手法を新たに発表しました。これは、AIが生成した1枚の画像から、その画像を生成するために使用した生成モデルをリバースエンジニアリングするものです。この方法により、ディープフェイク画像そのものが検出器の唯一の情報であることが多い実世界でのディープフェイク検出とトレースが容易になりました。</p> <p>これらのパートナーシップと私たち自身の洞察が進化するにつれて、操作されたメディアに対する弊社のポリシーも進化していくでしょう。その一方で、弊社は引き続き必要な投資を行い、この分野の他のステークホルダーと協力して、真のインパクトをもたらす解決策を見出すことに全力を尽くします。</p>
<p>9. 「情報発信者側における信頼性確保方策の検討」関係</p>		
①	信頼性の高い情報の表示	上記 2.1(1) 参照。
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	
③	メディアとの連携体制構築 具体的検討・取組	メタ・ジャーナリズム・プロジェクトは、世界中の出版社と協力し、ジャーナリストとコミュニティとのつながりを強化するために活動しています。また、ニュース業界が抱えるビジネス上の中核的な課題への対処も支援しています。弊社のプロジェクトは、ニュースを通じ

		<p>てコミュニティを構築すること、世界中のニュースルームを訓練すること、ニュース出版社や非営利団体と提携して misinformation と戦い、ニュースリテラシーの促進、新しい取り組みへの資金提供、弊社のプラットフォームでのジャーナリズムを改善することの3点で活動しています。https://www.facebook.com/journalismproject</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>いわゆるなりすましを軽減するための弊社のアプローチの1つは、認証バッジ機能を利用者のプロフィールやページに提供することです。弊社では認証バッジを発行しており、これは、公人・著名人、有名人、グローバルブランドの真正性が確認されたアカウントであることを弊社が認めたことを示します。</p> <p>例えば、検索とプロフィールで Facebook ページまたはアカウント名の横に表示されます。発行に当たっては、Facebook アカウントを審査する際に複数の情報を考慮して、社会的関心が高く認証の条件を満たしているかどうかを判断します。</p> <p>https://www.facebook.com/help/1288173394636262/?ref=share</p>
10. その他		
①	意見・補足	<p>弊社は、オンライン上で気候学に係る事実を曲げるために使われる手法の種類についてのブログ記事を発表しました。</p> <p>https://sustainability.fb.com/blog/2022/10/24/climate-science-literacy-initiative/</p> <p>これは、モナシュ気候変動コミュニケーション研究ハブ、ケンブリッジ社会意思決定研究所、エール大学気候変動コミュニケーションプログラムとの共同研究によるものです。</p> <p>この記事は、特に気候変動に関する misinformation に対抗することを目的としていますが、その手法は、健康上の misinformation や陰謀論など、他のタイプの misinformation にも応用できると考えています。</p>