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		Facebook Japan 株式会社
１．前提（サービス概要）		
①	対象サービス名	Facebook/Instagram
	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】 Facebook 月間アクティブ利用者数：2,600 万人（2019 年 3 月） Instagram 月間アクティブアカウント数：3,300 万（2019 年 3 月）
		【グローバルの数値】 Meta が提供するプラットフォーム利用者数：月間アクティブ利用者数 35 億 9,000 万人（2021 年 12 月時点）
		【（質問に答えられない場合）参考となる数値】
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	誤情報は、包括的な禁止事項を明示する方法がないため、弊社のコミュニティ規定で取り扱う他の種類の発言とは異なります。例えば、過度な暴力描写やヘイトスピーチについては、弊社が禁じている発言はポリシーに規定されているため、そのポリシーに賛成しない人でも従うことができます。しかし誤情報については、そのような方針を提供することができません。世界は絶え間なく変化し続けているため、ある時点では真実であっても、次の瞬間には真実でなくなることがあります。また弊社は、自身の周りの世界について異なるレベルの情報を有しているため、真実でない情報も真実だと信じてしまうことがあります。「誤情報」を単純に禁止するポリシーでは、情報への完全なアクセスができないため、弊社のサービスを利用する人びとに役立つ通知を提供することができず、ポリシーに強制力を持たせることもできません。 代わりに、弊社のポリシーでは、誤情報についてさまざまなカテゴリーを明確にし、対象となる発言を弊社が見つけたときの対処法を示した明確なガイダンスを設けるよう努めています。それぞれのカテゴリーにおいて、弊

		社のアプローチには、表現、安全、尊厳、真正性、プライバシーに対する弊社の価値観のバランスをとる試みが反映されています。 弊社は、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誤情報を削除します。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、非常に紛らわしい特定の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誤情報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。 その他のすべての誤情報については、弊社は、その情報の表示頻度の抑制に力を入れています。その取り組みとして、弊社は第三者のファクトチェック団体と連携し、弊社のプラットフォームで最も拡散しやすいコンテンツの正確性についてレビューおよび評価を行っています（弊社のファクトチェックプログラムの仕組みについては こちら をご覧ください）。また、利用者が閲覧、信頼、共有するコンテンツを自分自身で決められるよう、メディアとデジタルのリテラシー向上のための リソースも提供しています。 最後に、弊社は、誤情報の拡散と重複することが多い他の分野でのコンテンツや言動を禁止しています。例えば、弊社のコミュニティ規定により、偽アカウント、不正行為、組織的な偽装行為は禁じられています。
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】
⑤		【④全般・その他】
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	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です。そのため、弊社では市民社会組織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。

		https://transparency.fb.com/en-gb/policies/improving/stakeholders-help-us-develop-community-standards/ https://transparency.fb.com/en-gb/policies/improving/our-stakeholders 例えば、新型コロナウイルス感染症やワクチンについて、主要な保健機関が否定している誤った主張を削除し、事実確認機関が虚偽と判断したコンテンツの配信を減らし、ワクチン受容の向上に役立つ権威ある情報を提供しています。また、新型コロナウイルス感染症情報センターと Facebook や Instagram での教育用ポップアップを通じて、世界中の 20 億人を保健専門家のリソースと結びつけています。 https://transparency.fb.com/en-gb/features/approach-to-misinformation/ https://about.fb.com/news/2021/08/community-standards-enforcement-report-q2-2021/ また、弊社のインテグリティへの取り組みは、世界をより身近にするという会社のミッションの中核をなすものです。なぜなら弊社は、ソーシャルテクノロジーが、世界中の人々が自由に、公平に、安全に自分自身を表現できる場所であることを望んでいるからです。ここ数年、弊社は、人々がプラットフォーム上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、人材と技術への投資を増やしてきました。
②	具体的な役割	弊社は、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上における誤情報、分断・両極化、情報の質及び紛争に関する課題について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 万米ドルの資金を提供します。 ２０２１年６月、弊社は第三回 Foundation Integrity Research プログラムにおいて、誤情報及び分極化についての研究のプロポーザルの公募をいたしました。この公募に対し、446 もの質の高いプロポーザルが世界中の 288 の大学・研究機関から提出されました。選ばれた 19 のプロポーザルに基づく研究は、21 カ国に渡ります。 https://research.facebook.com/blog/2021/09/announcing-the-2021-recipients-of-research-awards-in-misinformation-and-polarization/
③	議論を踏まえた取組	上記 2 で示した賞の目的は、これらの分野における科学者コミュニティの発展を支援し、ソーシャルテクノロジー企業が自社のプラットフォーム上で社会問題にどのように対処すればよいか、業界全体で共通の理解を得ることにあります。研究対象は、弊社のアプリやテクノロジーに限りません。
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・アカウンタビリティの確保」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）】 弊社は、専門知識を有するパートナーが、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した場合、その誤情報および検証できない噂を削除します。誤情報とは、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されます。検証できない噂とは、専門知識を有するパートナーによる情報元の特定が極めて困難または不可能な主張、信頼できる提供元がない主張、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されます。
		【②健康・医療関係（コロナ関係等）】 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誤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誤情報には次のようなものがあります。 - ワクチンに関する誤情報。弊社は、主にワクチンに関する誤情報について、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。 - 公衆衛生上の緊急事態の間の誤情報。誤情報については、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。 - 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。
		【③選挙・政治関係】 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誤情報を削除します。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、方法に関する誤情報 - 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誤情報 - 候補者が立候補するか否かに関する誤情報
		【④全般・その他】 誤情報は、包括的な禁止事項を明示する方法がないため、Facebook のコミュニティ規定で取り扱う他の種類の発言とは異なります。例えば、過度な暴力描写やヘイトスピーチについては、弊社が禁じている発言はポリシーに規定されているため、そのポリシーに賛成しない人でも従うことができます。しかし誤情報については、そのような方針を提供することができません。世界は絶え間なく変化し続けているため、ある時点では真実であっても、次の瞬間には真実でなくなることがあります。また弊社は、自身の周りの世界について異なるレベルの情報

		を有しているため、真実でない情報も真実だと信じてしまうことがあります。「誤情報」を単純に禁止するポリシーでは、情報への完全なアクセスができないため、弊社のサービスを利用する人びとに役立つ通知を提供することができず、ポリシーに強制力を持たせることもできません。 代わりに、弊社のポリシーでは、誤情報についてさまざまなカテゴリーを明確にし、対象となる発言を弊社が見つけたときの対処法を示した明確なガイダンスを設けるよう努めています。それぞれのカテゴリーにおいて、弊社のアプローチには、表現、安全、尊厳、真正性、プライバシーに対する弊社の価値観のバランスをとる試みが反映されています。 詳細については下記 URL を御参照ください。 https://transparency.fb.com/en-gb/policies/community-standards/misinformation/
	(ii)削除等の対応	【①社会的混乱関係（災害等）】 弊社は、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誤情報を削除します。このようなカテゴリーにおける誤情報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。
		【②健康・医療関係（コロナ関係等）】 同上
		【③選挙・政治関係】 同上
		【④全般・その他】 その他のすべての誤情報については、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、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！」)に無害な方法で誤情報を利用することが多々あります。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、誤情報の拡散と重複することが多い他の分野でのコンテンツや言動を禁止しています。例えば、Facebook のコミュニティ規定により、偽アカウント、不正行為、組織的な偽装行為は禁じられています。
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	Facebook のコミュニティ規定は生きた文書であり、弊社のポリシーもオンラインにおける人々の行動の変化に合わせて進化しています。例えば、かつては無害だった言葉が害のある言葉になることはありますし、その逆もあり得ます。また、特定のグループが急に攻撃の対象になることもあり得ます。 弊社のコンテンツポリシーチームは、２週間毎にポリシーフォーラムという会議を開き、コミュニティ規定及び広告ポリシーについて、調整の必要がないか、議論します。議論は、セーフティチーム、サイバーセキュリティチーム、カウンターテロリズムの専門家、グローバルオペレーションチーム、プロダクトマネジャー、リサーチャー、公共政策担当、法務担当、人権とダイバーシティチームなど、社内の幅広い専門家を集めて行われ、時にはジャーナリストやアカデミアの方にオブザーバーとして参加していただくこともあります。 詳細については下記 URL をご参照ください。 https://transparency.fb.com/en-gb/policies/improving/policy-forum-minutes/ 外部専門家との連携 弊社のコンテンツポリシー/ステークホルダーエンゲージメントチームは社外ステークホルダーと定期的に連携し、彼らの世界を網羅する知識と助言をポリシー策定プロセスに取り入れてポリシーをより強固なものにするために尽力しています。 さらに、弊社のポリシーが与える影響を深く理解するため、NGO や学術機関などの専門家との関係を築いたり、世界中のさまざまな市民社会団体と連携したりしています。 ステークホルダーとの連携への弊社の取り組みについては下記 URL をご参照ください。 https://transparency.fb.com/ja-jp/policies/improving/input-from-external-stakeholders/ これに加えて、2020 年、弊社が行う最も重要で困難なコンテンツに関する決定について、独立したチェックを行うために、監督委員会が設立されました。監督委員会は、コンテンツに関する決定が Facebook や Instagram のポリシーや価値観、さらには国際的な人権規範の枠組みの中で表現の自由を守るというコミットメントと一致しているかどうかを審査します。そして、監督委員会は、これらの原則及び利用者と社会への影響に基づいて決定を下します。これらの決定には拘束力があり、監督委員会の会則で、弊社は、監督委員会の決定の実施が法律に違反するおそれがない限り、実施しなければならないこととされています。また実際に弊社は、監督委員会の決定を全て、その公表後速やかに実施しています。

		https://transparency.fb.com/oversight/
（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】 弊社は四半期ごとにコミュニティ規定施行レポート(CSER)を発行しています。このレポートには、当社のポリシーに違反するコンテンツの防止と対策の詳細が記載されています。 https://transparency.facebook.com/community-standards-enforcement なお、CSER 第 10 版（2021 年８月）では、Facebook と Instagram において全世界で 2400 万件以上のコンテンツを新型コロナウイルス感染症に関する誤情報に関連するポリシーに違反していると判断し、削除した旨公表しています。
		【（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】 弊社は四半期ごとにコミュニティ規定施行レポート(CSER)を発行しています。このレポートには、当社のポリシーに違反するコンテンツの防止と対策の詳細が記載されています。 https://transparency.facebook.com/community-standards-enforcement なお、CSER 第 10 版（2021 年８月）では、3,000 以上のアカウント、ページ、グループを新型コロナウイルス感染症およびワクチンの誤情報拡散に関連するポリシーに繰り返し違反していると判断し、削除した旨公表しています。

		【（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】 パンデミック発生以来、弊社は新型コロナウィルス感染症とワクチンの誤情報に関するポリシーに違反するコンテンツを全世界で 2400 万件以上削除してきました。また、同ポリシーに繰り返し違反したアカウント、ページ、グループを 3,000 以上削除しました。
		【（質問に答えられない場合）参考となる数値】 措置を講じた違反コンテンツのうち、利用者の報告を受ける前に弊社が検出したコンテンツの数は、上記 CSER の中において「事前対応率」として報告しています。
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）	Meta では独自のテクノロジーを利用して違反コンテンツを積極的に検出し、利用者によって報告される前にその大半を削除しています。エンジニア、データサイエンティスト、審査チームはこのテクノロジーを更新してさらに洗練されたものにするために協力して取り組んでいます。こうした中、審査チームはこのテクノロジーを活用して、コンテンツの審査優先順位の決定にも役立てています。 弊社が Facebook と Instagram で削除している違反投稿・違反アカウントの数は 1 日あたり数百万件に達しています。これらの大半は、背後にあるテクノロジーにより自動的に実行され、しかもそのほとんどが利用者に表示される前に処理されています。ときには、違反の可能性のあるコンテンツが検出された後、確認と措置の実施のために審査チームに転送されることもあります。 違反コンテンツに対する取り組みに終わりはありません。弊社のテクノロジーによる監視の目を巧妙に逃れようとする行為は尽きることがないため、弊社は絶え間なくテクノロジーを改良し続ける必要があります。 テクノロジーによる違反検出のしくみ等についての詳細は下記 URL をご参照ください。 https://transparency.fb.com/ja-jp/enforcement/
⑤	削除以外の取組 (i)警告表示	下記参照
	(ii)表示順位の低下	下記参照
	(iii)その他の取組内容	弊社では虚偽の情報の拡散抑制に全力で取り組んでいます。一部の国では、虚偽の情報を識別して審査するために、国際ファクトチェックネットワーク（IFCN）が認証した独立した第三者ファクトチェック団体と連携しています。ファクトチェック団体は事実確認を行い、情報の正確性を評価します。

		虚偽の情報の配信の抑制 ● フィードでの誤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誤情報と評価すると、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、フェイクニュース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。Meta では、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 ● 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。 ● フェイクニュース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、Facebook から通知が送付されます。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す。 ● 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、こちらでご確認いただけます。 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、シェアするべきかを事実に基づいて判断できるようになります。 ● フェイクニュースだと思う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。 ● 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。 https://www.facebook.com/help/1952307158131536 https://transparency.fb.com/ja-jp/features/approach-to-misinformation/
⑥		【日本の数値】

	不正な申告や削除要請への対策の方法・仕組み、対応件数	【グローバルの数値】 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。 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および利用規約に違反する可能性があります。弊社は、このような行為を制限し、また、このような行為の過度の違反を検知する技術を有しています。 利用者のネットワークが組織的にこの機能を悪用し、弊社の制限や執行を逃れようとしていることが判明した場合、弊社はそれらの利用者に対して措置を講じます。
		【（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	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。 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。 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。 コミュニティ規定をどのように実施し、何を掲載すべきか、何を削除すべきかをどのように決定するかについての詳細は、下記 URL をご覧ください。 https://about.fb.com/news/2018/04/comprehensive-community-standards/ https://about.fb.com/news/2019/03/inside-feed-vanity-fair-policy-team/ https://about.fb.com/news/2018/08/hard-questions-free-expression/ https://about.fb.com/news/2018/04/community-standards-examples/ 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て用いています。 コミュニティからの報告

		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。 人間によるレビュー ● 現在、弊社では 4 万人が安全・セキュリティに取り組んでいます。このチームはグローバルに活動しており、日本語を含む数十の言語で 24 時間 365 日コンテンツをレビューしています。 ● 弊社のコンテンツモデレーターはさまざまなバックグラウンドを持っていますが、弊社が人材の採用及び最適化において最も重要と考えているのは、言語と文化的背景です。 ● （人間の）コンテンツモデレーターは、公正かつ正確にコミュニティ規定を実施するために不可欠であり、特にコンテンツを取り巻く文脈が重要な場合には、その重要性が増します。 〇 例えば、ヘイトスピーチなどです。弊社のシステムは、ヘイトスピーチとしてよく使われる特定の言葉を認識できますが、それを使う人の意図は必ずしも認識できません。そのため、審査チームがこのコンテンツをレビューします。 テクノロジーを使って有害なコンテンツを主体的にに識別 ● Meta では主に人工知能（AI）を利用して、Facebook や Instagram 上の違反コンテンツを特定しています。弊社のテクノロジーは、コンテンツの一部が Facebook コミュニティ規定や Instagram のコミュニティガイドラインに違反していることを確信して、（多くの場合、誰かの目に触れる前に）自動的に削除することがあります。 ● 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。 〇 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。 〇 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。 〇 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。 https://transparency.facebook.com/community-standards-enforcement
②	対応決定時における通知の内容、理由の記載の程度	Facebook のコミュニティ規定や Instagram のコミュニティガイドラインに違反するコンテンツがあると判断した時点で、弊社はそれを削除します。その際には コンテンツを共有した利用者には、削除した理由を理解してもらい、今後違反するコンテンツを投稿しないようにする方法を理解してもらうために通知します。この情報は、その人のサポート受信箱でも確認できます。また、弊社が何か間違っていると思われる場合には、その措置を不服として訴えることができるプロセスも用意されています。 コンテンツの削除の流れについては下記 URL をご参照ください。 https://transparency.fb.com/enforcement/taking-action/taking-down-violating-content/#takedown-experience
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】 Meta の審査チームは、数多くの職務の一環としてコンテンツ審査を担当する弊社の社員と、弊社のパートナーに従事するコンテンツ審査担当者によって構成されています。Meta が提供するプラットフォームのコミュニティの多様性の豊かさを反映して、審査チームには退役軍人や法律専門家に加え、児童の安全、ヘイトスピーチ、対テロ対策などの各種ポリシー分野に精通する規定施行の専門家など、さまざまな背景と職歴を持つ人たちが集結しています。 また、現在、弊社では 4 万人が安全・セキュリティに取り組んでいます。Meta の審査チームはグローバル体制の下、24 時間年中無休でコンテンツを審査しています。ドイツ、アイルランド、ポルトガル、スペイン、ポルトガ

		ルフィリピン、米国など、弊社が全世界に展開している 20 以上の拠点でコンテンツの審査を行うこれらのチームは全体で日本語を含む数十種類に及ぶ言語に対応しています。 https://transparency.fb.com/ja-jp/enforcement/detecting-violations/people-behind-our-review-teams/
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	Meta の審査チームはグローバル体制の下、24 時間年中無休でコンテンツを審査しています。ドイツ、アイルランド、ラトビア、スペイン、ポルトガル、フィリピン、米国など、弊社が全世界に展開している 20 以上の拠点でコンテンツの審査を行うこれらのチームは全体で日本語を含む数十種類に及ぶ言語に対応しています。 詳細については下記 URL をご参照ください。 https://transparency.fb.com/ja-jp/enforcement/detecting-violations/how-review-teams-work/
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	利用者が Facebook や Instagram で報告した場合、その報告の状況を Facebook ではサポート受信箱や Instagram ではサポートリクエストから確認することができます。 報告を受け取った日時、報告の理由の確認、内容についてのレビューの結果を通知しています。 弊社は、コンテンツがコミュニティ規定に反すると判断した場合、コンテンツを削除します。また、そのコンテンツが弊社のポリシーに違反していない場合には、その旨をお知らせします。その際、コンテンツの掲載を継続するという弊社の判断にご同意いただけない場合には、通常、再審査をリクエストする機会を提供しています。 利用者が再審査をリクエストされた場合、コミュニティ運営の専門チームがコンテンツを再度審査します。間違いが見つかった場合には、利用者にお知らせし、報告されたコンテンツは非表示または削除されます。最も有害な種類のコンテンツ（児童搾取画像など）の一部については、再審査を申請することはできません。 再審査のリクエストについては下記 URL をご参照ください。 www.facebook.com/help/134552198624586/?helpref=uf_share 利用者のアカウントが日本語に設定されている場合、すべての通知やレポートは日本語で提供されます。 また、下記 URL にて利用者にとってポリシーの施行がどのようなものとなるかの例をご覧いただけます。 https://transparency.fb.com/policies/community-standards/bullying-harassment/#user-experiences
（４）透明性・アカウンタビリティの確保		
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・アカウンタビリティ確保方策	

	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	コミュニティ規定施行レポートを日本語に翻訳し公表しています。 https://about.fb.com/ja/news/2020/11/community-standards-enforcement-report-nov-2020/ https://about.fb.com/ja/news/2021/02/community-standards-enforcement-report-q4-2020/ https://about.fb.com/ja/news/2021/05/community-standards-enforcement-report-q1-2021/ https://about.fb.com/ja/news/2021/08/community-standards-enforcement-report-q2-2021/
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	コンテンツ収益化ポリシーは、第三者ファクトチェック団体により虚偽と判断されたコンテンツは収益化の対象として認めていません。
②	広告出稿制限	広告ポリシーにおいて以下のように定めています。 Meta では、第三者ファクトチェック団体によって虚偽であると証明された主張、また場合によっては特定の専門性を持つ組織によって虚偽であると証明された主張を含む広告を禁止しています。虚偽と判断された情報を繰り返し投稿する広告主は、Meta が提供するプラットフォームでの広告掲載に制限がかけられる可能性があります。 https://www.facebook.com/policies/ads/prohibited_content/misinformation
③	ターゲティング技術の適用に関する規定	広告ポリシーにおいて以下の定めを置いています。 7. ターゲット設定 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・アカウンタビリティ確保方策	Meta は表示する広告を利用者にとってできるだけ魅力的で有益なものにしたいと考えています。以下に、表示される広告を決定する際に使用する基準の例を示します。 ● Facebook での利用者のアクティビティ（ページへの「いいね！」や、表示された広告へのクリックなど）。

		● Facebook アカウントのその他の情報(年齢、性別、所在地、Facebook へのアクセスに使用しているデバイスなど)。 ● 広告主の情報、広告主のパートナーの情報、マーケティングパートナーが Facebook と共有している情報(メールアドレスなど)。 ● 未成年者に表示される広告に対しては、追加のポリシーが定められています。詳しくはこちらのヘルプセンター記事をご覧ください。 ● Facebook 外のウェブサイトとアプリでのアクティビティ。広告設定でこれをオフにする方法について、詳しくはこちらをご覧ください。 なお、 ● Meta は、利用者本人の許可がない限り、利用者個人を特定できる情報(氏名またはメールアドレスなど、それ自体を利用者への連絡または個人の特定に利用できる情報)を共有しません。Meta がどのような情報を受け取り、どのように使用するかについて、詳しくは Facebook のデータに関するポリシーと Cookie ポリシーをご覧ください。 ● Meta では、利用者が提供する情報、利用者が Meta の各プラットフォームで実行したアクション、利用者が他のウェブサイト、アプリ、店舗で実行してサードパーティーの企業がシェアしたアクションを使用します。ただし、プロフィールに追加された特別保護対象の個人情報は例外とします。 Facebook 広告のしくみについて、詳しくは下記 URL をご参照ください。 https://www.facebook.com/about/ads https://transparency.fb.com/ja-jp/features/ranking-and-content/
	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	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、表示された広告に関する情報を確認できます。 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。

		https://www.facebook.com/help/259468828226154
⑥	広告とコンテンツの分離	
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	誤情報に対するポリシーや、どのような誤報を削除するかなど、詳細については下記 URL をご参照ください。 https://transparency.fb.com/en-gb/policies/community-standards/misinformation/ また、Transparency Center において、誤情報に対するアプローチを詳しく説明しておりますので、下記 URL をご参照ください。 https://transparency.fb.com/en-gb/features/approach-to-misinformation/
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	プログラムの重要なステップは下記のとおりです。 誤情報を特定:利用者からのフィードバックなどに基づいて誤情報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うえで審査する場合もあります。 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、事実を検証したうえ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。 誤情報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。 誤情報を目にする利用者の数を抑制: ファクトチェック団体が「虚偽」、「改変」、「一部虚偽」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、誤情報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。 誤情報を繰り返し配信する違反者への措置: 「虚偽」または「改変」と評価された誤情報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。

		https://www.facebook.com/business/help/2593586717571940 ファクトチェック団体による評価の種類には、虚偽、改変、一部虚偽、背景の説明不足、風刺、事実があり、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。
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	弊社の技術は、利用者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、誤解を招く可能性のある投稿を検出することができます。また、Facebook や Instagram の利用者は、ファクトチェック団体が投稿を詳しく見るために、コンテンツの一部にフラグを立てることができます。その他にも、虚偽の情報を特定するためのシグナルとして、以下のようなものがあります。 - 投稿に対する不信感を示すコメント - 機械学習モデルから得られるもの（誤情報を予測する能力を継続的に向上させています） - ファクトチェック団体による経験（自らコンテンツを特定し、レビューしています） 国際ファクトチェッキングネットワーク（IFCN）の認証を受けているファクトチェック団体は、CrowdTangle（メタ社のツール）にもアクセスできます。このツールは、ソーシャルメディアにおける公開コンテンツのパフォーマンスに関する洞察を提供し、誤情報を含む投稿を特定するのに役立ちます。出版社、ジャーナリスト、研究者、学術関係者も、CrowdTangle を使用して、ソーシャルメディア上の公開コンテンツを追跡、分析、報告することができます。
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	弊社では、Facebook と Instagram における誤情報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サードパーティーファクトチェック団体と提携しています。
7．「ICT リテラシー向上の推進」関係		
①	普及啓発の取組・投資	【みんなのデジタル教室】 弊社は、アジア太平洋地域の専門家と協力して、Facebook や Instagram といったオンライン上での嫌がらせやいじめなどに対処し、責任あるデジタル市民によるグローバルコミュニティを構築するためのリソースを提供するオンライン出張プログラム「みんなのデジタル教室」を立ち上げました。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出張授業を国内の中学校・高等学校などで行いましたす。

		2021 年 12 月末までに 13,000 名以上の学生が授業を受講しており、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。 https://wethinkdigital.fb.com/jp/ja-jp/ https://about.fb.com/ja/news/2020/12/we_think_digital/ また、新型コロナウィルス感染症に関する誤情報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。 https://fightcovidmisinfo.com/japanese/
②	他のステークホルダーとの連携・協力・投資	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。
８．「研究開発の推進」関係		
①	AI 技術に関する研究開発	弊社が誤情報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を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、サードパーティーの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。 https://about.fb.com/news/2018/05/hard-questions-false-news/ 詳しくは、下記 URL をご参照ください。 https://ai.facebook.com/
②	「ディープフェイク」対策の研究開発	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数名の組織の横断的な連合が参加しています。 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。

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	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誤情報と戦い、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 弊社は、新型コロナウイルス感染症情報センターのほか、Facebook および Instagram 上での通知画面の表示を通じて、20 億人の利用者を専門家からの信頼できる情報につなぎました。また、米国においてはワクチン検索ツールを通じてワクチン接種を申し込めるようにし、実際に 400 万人が利用するなど、ワクチン接種を促すために信頼できる情報を提供してきました。
②	ニュースの選別・編集に関する透明性・アカウンタビリティ確保方策	
③	メディアとの連携体制構築 具体的検討・取組	
④	情報源のトレーサビリティ確保、なりすまし防止・認証	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。 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。
１０．その他		
①	意見・補足	ウクライナにおける戦争で影響を受けたすべての人に、弊社は思いを寄せています。弊社は、コミュニティの安全を確保し、ウクライナや世界中で弊社のサービスを利用する利用者をサポートするために、アプリ全体で大規模な措置をとっています。

		特に、誤情報の拡散を減らすことについて、弊社のサービス上での拡散に対抗するために広範な措置を講じ、外部の専門家と協議を続けています。詳細については下記 URL をご参照ください（継続的に更新されていきます。）。 https://about.fb.com/news/2022/02/metas-ongoing-efforts-regarding-russias-invasion-of-ukraine/