

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	Facebook Japan 株式会社	
1. 前提（サービス概要）		
1	対象サービス名	Facebook
	サービス分類	コンテンツ共有サービス
2	加入者数、月間アクティビューザ数又は書き込み数	【日本の数値】 月間アクティブアカウント数：2,600万人（2019年7月）
		【グローバルの数値】 月間アクティブアカウント数：28億人（2020年6月）
		【（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	
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	
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	例えば、Facebook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、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、私たちは、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	Facebook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

		<p>深めることを目的とした研究提案に対して、総額200万ドルの資金を提供します。</p> <p>https://research.fb.com/blog/2020/08/announcing-the-winners-of-facebooks-request-for-proposals-on-misinformation-and-polarization/</p>
3	議論を踏まえた取組	<p>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、Facebook社のアプリやテクノロジーに限られません。</p>
<p>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</p>		
<p>(1) 偽情報等に関するポリシー</p>		
1	(i) 禁止行為	<p>【①社会的混乱関係（災害等）】</p> <p>コミュニティ規定では、例えば、「暴力と扇動」という以下の定めを置いております。</p> <p>Facebookでは、Facebookのコンテンツに関連して起こりうるオフラインでの危害を防止するための取り組みを行っています。人は不満や他人との意見の相違を示すため、実行するつもりのない暴力の脅しや呼びかけを、深く考えずに行うことがあります。Facebookは、深刻な暴力行為を煽ったり、促したりする発言を削除します。人身に実際の危害を及ぼす、または公共安全を直接脅かすおそれがあると思われる場合には、Facebookはコンテンツを削除し、アカウントを停止し、法執行機関に協力します。気軽な発言と、公共または個人の安全を実際に脅かすおそれがあると思われるコンテンツとを区別するために、Facebookでは言葉や文脈について考慮するように努めています。また、起こりうる脅威かどうかを判断する際に、Facebookは対象の個人の知名度や身の危険が及ぶリスクなどの追加情報を考慮することもあります。</p> <p>【②健康・医療関係（コロナ関係等）】</p> <p>コミュニティ規定に以下のような定めを置いております。</p> <p>世界中の人々が、前例のない公衆衛生上の緊急事態に直</p>

		<p>面しています。このためFacebookでは、コミュニティ規定で、新型コロナウイルス感染症に関連した有害なコンテンツや新たな不正利用から利用者を守れるようにしたいと考えています。弊社は、現実世界での危害が生じるリスクを助長するコンテンツを削除するように取り組んでいます。例えば、危害を加えるための計画、医療用マスクと関連商品の販売、ヘイトスピーチ、いじめや嫌がらせ、差し迫った暴力または身体的危害のリスクにつながる偽情報を禁止するポリシーを設け、それに沿って対応しています。</p> <p>状況の変化に対応しつつ、弊社は今後もプラットフォーム上のコンテンツの確認、発言の動向の見極めや関係者との連携を続け、このような緊急事態において弊社コミュニティの皆さまの安全を守るために、適時ポリシーに関する指針を追加します。</p> <p>【③選挙・政治関係（米国大統領選等）】</p> <p>コミュニティ規定は、例えば、「危害を加えるための計画および犯罪の宣伝」という以下のような定めを置いております。</p> <p>オフラインでの危害と模倣行為を防止および阻止するため、Facebookでは、人、ビジネス、財産、動物を対象とする特定の犯罪行為や有害行為について、これらを助長、計画、宣伝すること、またはそのような行為を告白することを禁止しています。</p> <p>（例）</p> <ul style="list-style-type: none"> ・ 投票者または国勢調査への干渉 ・ 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、方法に関する不実表示 ・ 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不実表示 ・ 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不実表示 <p>米国移民関税執行局(ICE)が投票所に待機していると主張しているコンテンツ</p> <ul style="list-style-type: none"> ・ 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主張
--	--	---

	<p>【④全般・その他】 フェイクニュースに関して、コミュニティ規定に以下のような定めを置いております。</p> <p>Facebook上のフェイクニュース(偽ニュース)の拡散防止は、Facebookが責任を持って真摯に対策を講じている事項の1つです。Facebookはこれがセンシティブで難しい課題であることを認識しています。Facebookでは、人々の間の生産的な議論を妨げることなく、利用者が有益な情報を得ることができる環境を作りたいと考えています。フェイクニュースと風刺や意見との線引きも、また難しい問題です。このような理由により、Facebookではフェイクニュースをなくすのではなく、ニュースフィードでの表示回数を下げることでフェイクニュースの配信を大幅に減らす措置を取っています。</p>
(ii) 削除等の対応	<p>【①社会的混乱関係(災害等)】 コミュニティ規定に照らして違反する場合には削除等の措置を講じます。</p> <p>【②健康・医療関係(コロナ関係等)】 例えば、危害を加えるための計画、医療用マスクと関連商品の販売、ヘイトスピーチ、いじめや嫌がらせ、差し迫った暴力または身体的危害のリスクにつながる偽情報を禁止するポリシーを設け、それに沿って対応しています。</p> <p>【③選挙・政治関係(米国大統領選等)】 コミュニティ規定に照らして違反する場合には削除等の措置を講じます。</p> <p>【④全般・その他】 Facebookでは利用者が情報にアクセスできるコミュニティを構築しながら、フェイクニュースの拡散を抑えるべく、さまざまな取り組みを行っています。取り組みの事例は以下のとおりです。</p> <ul style="list-style-type: none"> ・ 偽情報を広める利用者、ページ、ドメインに対する経済的なインセンティブの提供を断つ ・ コミュニティからのフィードバックを含むさまざまなシグナルを利用して、どの記事がフェイクであるかを予測する機械学習モデルに通知する

		<ul style="list-style-type: none"> ・独立した第三者ファクトチェッカーによりフェイクと評価されたコンテンツの配信を減らす ・記事の背景を示し、利用者のニュースリテラシーを高めることで、何を読み、信じ、シェアすべきかを利用者自身で決められるようサポートする ・この困難な問題を解決するため、学術団体やその他の団体と協力する
(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】
		<p>【（質問に答えられない場合）参考となる数値】</p> <p>Facebook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。</p> <p>https://transparency.facebook.com/community-standards-enforcement</p>
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】
		<p>【（質問に答えられない場合）参考となる数値】</p> <p>Facebook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。</p> <p>https://transparency.facebook.com/community-standards-enforcement</p>
	(ii) アカウントの停止数	【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>コミュニティ規定に違反するコンテンツに対して講じた措置件数については、上記CSERで報告しています。</p>
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	【日本の数値】
		【グローバルの数値】

	能の活用等)	【（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）	Facebook社では、コンテンツをシェアする利用者の安全を確保するために以下のように管理しています。 Facebookには、安全とセキュリティに取り組むスタッフが35,000人以上おり、24時間体制で60以上の言語のコンテンツをチェックし、またコミュニティからの報告に対応しています。 人の目に触れる前にフェイクニュースが拡散するのを抑え、Facebookのポリシーに違反している可能性のあるコンテンツを特定するためのテクノロジーに投資しています。
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	Facebookは、利用者の皆様がどのコンテンツを閲覧し、信頼してシェアできるのかを判断するために、より多くの情報を知らせる機能と製品に投資しています。2018年、コンテキストボタンでニュースフィードの記事の背景情報を提供することを開始しました。このボタンにはパブリッシャーについてのウィキペディア記事、Webサイトの年数、Facebookで記事がシェアされた回数やシェアされた場所などが表示されます。コミュニティ規定に違反して削除された投稿や、第三者ファクトチェッカーによって「虚偽」、「虚実混交」または「虚偽の見出し」と評価された投稿をページオリティタブに示すことで、ページの所有者がコンテンツの質を改善することを支援しました。 偽情報やクリックベイトなど、問題のあるコンテンツであってもコミュニティ規定の削除基準を満たさないものがあります。利用者の皆様からは、この種類のコンテンツは見たくないという声をいただくことが多く、私たちはこれらのコンテンツの投稿を許可する一方で、拡散されないようにしています。 Facebookはフェイクニュースの拡散抑制に全力で取り組んでいます。その方法として、偽アカウントを削除したり、偽情報を発信する個人やページの経済的誘因を取り

		除いたりしています。また、コミュニティからのフィードバックなどのシグナルを利用して、虚偽の可能性がある記事を特定しています。Facebookが独立した第三者ファクトチェッカーと連携している国では、ファクトチェッカーが虚偽と評価した記事はニュースフィードの下のように表示されます。繰り返し偽情報を作成またはシェアするページやドメインに対しては、そこからのコンテンツ配信を大幅に削減したり、広告を出せなくしたりします。
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】 Facebook上のコンテンツを簡単に報告できるようにし、コミュニティには報告ツールを使って違反していると思われるコンテンツを報告することを奨励しています。Instagramを利用している人は、各コンテンツに表示されるオプションを含め、さまざまな方法で違反行為を報告することができます。また、利用者からの重複報告を認識する技術を使用しており、1,000人があるコンテンツを報告しても、1,000人が同じコンテンツをレビューすることがないようにしています。また、AIは、コンテンツレビュアーがレビューするケースに優先順位をつけて、最も有害で時間的制約のあるコンテンツを最初に処理するのにも役立っています。
(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	Facebook上のコンテンツを簡単に報告できるようにしています。また、コミュニティ規定に違反していると思われるコンテンツを報告するために、報告ツールを使用することを奨励しています。Facebookの利用者は、各コンテンツに表示されるオプションを含め、さまざまな方法でコミュニティ規定違反を報告することができます。「報告」の方法については、以下を参照してください。 https://www.facebook.com/help/1380418588640631/ 利用者が問題のあるコンテンツを「報告」すると、

		<p>Facebookはそのコンテンツを審査し、コミュニティ規定に準拠していないコンテンツをすべて削除します。Facebookが問題の原因となった人物と連絡を取った場合、利用者名などの個人情報が開示されたり、通知されたりすることはありません。</p> <p>利用者がFacebookのコミュニティ規定に違反していることを「報告」した場合、利用者はサポート受信箱で報告の状況を確認することができます。報告した利用者だけがサポート受信箱を見ることができるようになります。</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>Facebookでは、虐待や嫌がらせであると報告されたコンテンツを評価する際には、常に、その報告をした人にサポート受信箱でその人の言語で最新情報を提供することで、明確なコミュニケーションをとっています。報告者には、当社が報告を評価したことだけでなく、当社がいつ、どのような措置をとったかがわかります。</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】
		<p>【グローバルの数値】</p> <p>コミュニティオペレーションチームには、日本語を含む50以上の言語でFacebookやInstagramのコンテンツをレビューするコンテンツレビュアーがおり、世界中のサイトでグローバルコミュニティにサービスを提供しています。</p> <p>過去数年の間に、Facebook社は安全、セキュリティ、コンテンツレビューに特化したチームを35,000人以上に拡充し、そのうち約15,000人をコンテンツレビュアーに充てています。</p>
	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】
		<p>【グローバルの数値】</p> <p>【（質問に答えられない場合）参考となる数値】</p>
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	<p>当社の審査は、ドイツ、アイルランド、ラトビア、スペイン、リスボン、米国、フィリピンなどの国の世界20以上のサイトに拠点を置いています。</p>

	ける責任者の有無	<p>いくつかのレビューサイトの所在地は公表していますが、安全性とセキュリティ上の理由から、多くのサイトやそこで働く人々についての具体的な詳細は公開していません。</p> <p>私たちの人の目による審査の担当者は、様々なバックグラウンドを持った人たちです。最も重要なのは、言語と文化的背景です。当社のチームは、日本語を含む当社のコミュニティで最も広く使われている言語をカバーし、50以上の言語で当社のポリシーに照らし合わせてコンテンツを審査しています。</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>Facebookでは、いじめや嫌がらせ、ヘイトスピーチなどのコミュニティ規定違反で削除された決定に異議申し立てする機会を提供しています。コンテンツが誤って削除されたと判断された場合、削除されたコンテンツは復元されます。</p> <p>詳しくは以下のサイトをご参照ください。</p> <p>https://about.fb.com/ja/news/2018/04/comprehensive-community-standards/</p> <p>https://transparency.facebook.com/community-standards-enforcement/guide</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	
	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を日本語に翻訳し公表しました。</p> <p>https://about.fb.com/ja/news/2021/02/community-standards-enforcement-report-q4-2020/</p>
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	
2	広告出稿制限	<p>広告ポリシーにおいて以下のように定めています。</p> <p>Facebookでは、サードパーティファクトチェッカーによ</p>

		<p>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、Facebookでの広告掲載に制限がかけられ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。 <p>https://www.facebook.com/policies/ads</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>広告ライブラリでは、Facebook製品に掲載されている広告を検索ことができ、表示された広告に関する情報を確認できます。</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Facebook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	
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	
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	<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>フェイクニュースを特定: Facebookの利用者からのフィードバックなどに基づいて偽情報の可能性のあるコンテンツを特定し、ファクトチェッカーに表示します。また、ファクトチェッカー自身がコンテンツを特定したうえで審査する場合があります。</p> <p>コンテンツを審査: ファクトチェッカー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Facebook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>偽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へ通知します。</p> <p>偽情報を目にする利用者の数を抑制: ファクトチェッカーが「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、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。また、ファクトチェッカー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>フェイクニュースを繰り返し配信する違反者への措置: 「虚偽」または「改変」と評価された偽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。</p> <p>https://www.facebook.com/business/help/2593586717571940</p>

		<p>ファクトチェッカー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カーが独自にコンテンツの審査と評価を行います。Facebook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	
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>Facebook社では、FacebookとInstagramにおける偽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、中立的な国際ファクトチェックネットワーク(IFCN)に認定された独立した第三者ファクトチェッカー団体と提携しています。</p>
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>Facebook社は、アジア太平洋地域の専門家と協力して、いじめ防止やオンラインでの嫌がらせなどのリソースを提供するプログラム「みんなのデジタル教室」を立ち上げました。</p> <p>日本の中学生・高校生を対象に、フェイクニュースが発信される動機や、情報を受け取る側の視点を考え、フェイクニュースを見分けるための様々なポイントについて、アクティビティを通じて学ぶ授業を展開しています。</p> <p>https://wethinkdigital.fb.com/jp/ja-jp/</p> <p>また「みんなのデジタル教室」の授業のエッセンスをInstagram上で体験・学習できるコンテンツ『タグでたどる物語』を展開しています。専用のInstagramアカウント(@minnano_digital_kyoshitsu_1)を訪れ、物語を読み進めながら、プライバシー保護やフェイクニュースへの対応などについて理解を深めていただくことを目的としています。</p> <p>https://about.fb.com/ja/news/2020/12/we_think_digital/</p>

2	他のステークホルダーとの連携・協力・投資	上記プログラムでは、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	
2	「ディープフェイク」対策の研究開発	Facebook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9月に「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、コーネル工科大学、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。
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	
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	Facebook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Facebookが認めたことを示します。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。 詳しくは下記URLをご参照ください。 https://www.facebook.com/help/1288173394636262/?ref

		=share
10. その他		
1	意見・補足	