

Metaによる Misinformation及び Disinformationへの対応について

27 March 2024

mis

意図せずして共有されることがあるもの
削除、抑制、情報共有の基本方針で臨む コンテンツ

vs

dis

ミスリード又は操ろうとする 計画的なもの
コミュニティ規定違反として対応する 行為

ABCフレームワーク(Actor, Behavior, Content)

行為者に係る問題

偽アカウント／濫用的アカウント

乗っ取り

なりすまし

虚偽表示

行為に係る問題

いじめやいやがらせ

現実性の高い暴力による脅迫

チャイルドグルーミング

金融詐欺

スパム

不正なエンゲージメント稼ぎ

組織的偽装行為

コンテンツに係る問題

Misinformation

ヌード・ポルノ

自殺自傷

暴力や過激な描写

ヘイトスピーチ

児童に対する性的搾取

MISINFORMATION



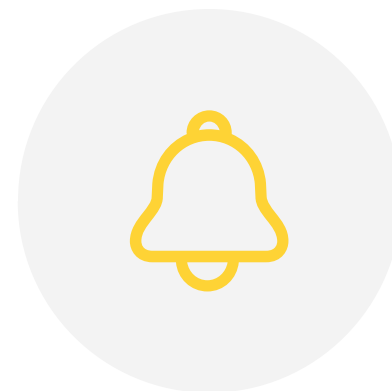
削除

コミュニティ規定に違反するコンテンツやアカウント(有害な誤情報を含む。)を削除



抑制

質の低いコンテンツ(誤情報を含む)の配信を抑制



情報提供

利用者が何を読み、何を信頼し、何を共有すべきかを判断できるようにより多くのコンテキストを情報提供

削除



差し迫った身体的危害を及ぼす危険性のある誤情報



投票行動抑制及び投票妨害



コミュニティ規定の他のカテゴリに違反するもの

なぜすべてのmisinformationを削除しないのか。

1. 我々は真実の裁定者ではないし、誰もそのような存在になるべきはない、と考えます。
2. 仮に真実の裁定者たろうとすると、すべての事象について真実が何かを知っていることを求められる。しかし、そのようなことは不可能です。
3. すべての情報が厳密に正誤が付けられるわけではない。真実には程度の問題もあるし、何が真実かについてもさまざまな意見があります。

MISINFORMATION

削除



差し迫った身体的危害を及ぼす危険性のある誤情報



投票行動抑制及び投票妨害



コミュニティ規定の他のカテゴリに違反するもの

なぜすべてのmisinformationを削除しないのか。

1. 我々は真実の裁定者ではないし、誰もそのような存在になるべきはない、と考えます。
2. 仮に真実の裁定者たろうとすると、すべての事象について真実が何かを知っていることを求められる。しかし、そのようなことは不可能です。
3. すべての情報が厳密に正誤が付けられるわけではない。真実には程度の問題もあるし、何が真実かについてもさまざまな意見があります。

選挙に係るポリシー

投票行動抑制と投票に係る不正行為

- 投票、有権者登録、国勢調査の**実施日・時間、場所、方法**に関する詐称
- **有権者の定義、投票要件、投票が集計されるかどうか、投票するためにどのような情報及び資料を提供しなければならないか**についての詐称
- **立候補の有無**に関する詐称
- 現金や贈答品で**票を売買する**申し出に関するもの
- 投票プロセスへの**不正な関与**の意図、指示、または明確な意図を示す発言

協調的干渉

- 個人が選挙に参加できなくなるような影響を与えるための**協調的干渉**の呼びかけ
- **投票を行うことが法執行の対象**(逮捕、国外退去、投獄など)をもたらすおそれがある旨の内容
- **投票、有権者登録、または選挙結果に起因する相当程度重大な暴力**に係る意図的な声明、行動の呼びかけ、または意図を含む内容

MISINFORMATION

抑制

コンテンツ配信 ガイドライン

コミュニティ規定でどのようなコンテンツがFacebookから削除されるかを説明しているのと同様に、コンテンツ配信ガイドラインでは、どのようなコンテンツが、問題あるコンテンツまたはクオリティの低いコンテンツとしてFacebook上の配信制限の対象となるのかについて説明しています。

<https://transparency.fb.com/en-gb/features/approach-to-ranking/types-of-content-we-demote>

利用者の直接的なフィードバックへの対応

弊社は、Facebookで見たいもの、見たくないものについて、利用者からのフィードバックを常に歓迎しており、それに応じてフィードに変更を加えています。

- ・ 迷惑広告
- ・ クリックベイトのリンク
- ・ 報告されたり非表示にされたりする可能性が高いコメント
- ・ エンゲージメントベイト
- ・ 不必要なユーザーデータを要求するウェブサイトへのリンク
- ・ 低品質なコメント
- ・ 低品質なイベント
- ・ 低品質な動画
- ・ オリジナルではない動画を投稿している
- ・ スпамだと推定されるページ
- ・ 扇動的な健康関連コンテンツや健康に関する投稿

クリエイターによる高品質で正確なコンテンツへの投資の奨励

弊社では、利用者に長期間、興味深い新しいコンテンツを楽しんでもらいたいと考えており、こうしたコンテンツの作成を促すインセンティブを設けています。

- ・ オリジナルコンテンツが少ないドメイン
- ・ ファクトチェックされた偽情報
- ・ 不正なシェア
- ・ 「クリックギャップ」が大きいドメインやページへのリンク
- ・ 広く信頼されていないニュースパブリッシャーからの投稿
- ・ 配信数を人為的に水増ししているページからの投稿

安全なコミュニティの構築

コンテンツの中には、その意図の有無にかかわらず、コミュニティにとって問題となるものがあります。弊社は、このようなコンテンツが、積極的に見ようとしなない人の目に届きにくいようにしています。

- ・ コミュニティ規定のボーダーライン上のコンテンツ
- ・ コミュニティ規定に違反している可能性が高いコンテンツ
- ・ 弊社のポリシーに繰り返し違反している利用者が投稿したコンテンツ
- ・ バイラリティが疑わしいことを示す投稿
- ・ 自殺に関する不確かな報告

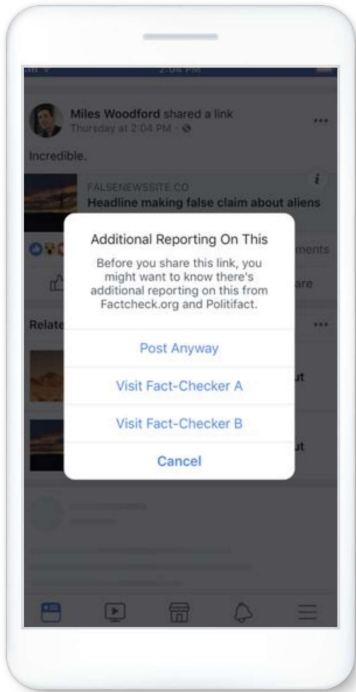
る利用者からの投

MISINFORMATION

情報提供

ラベル付けと通知

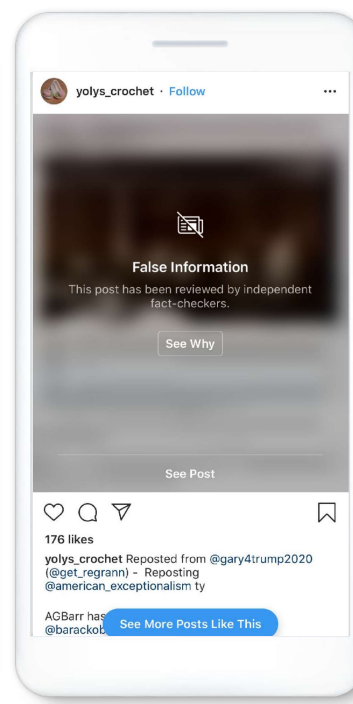
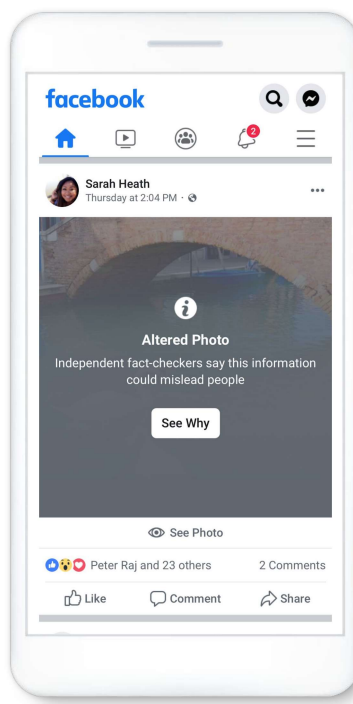
シェアする際の通知



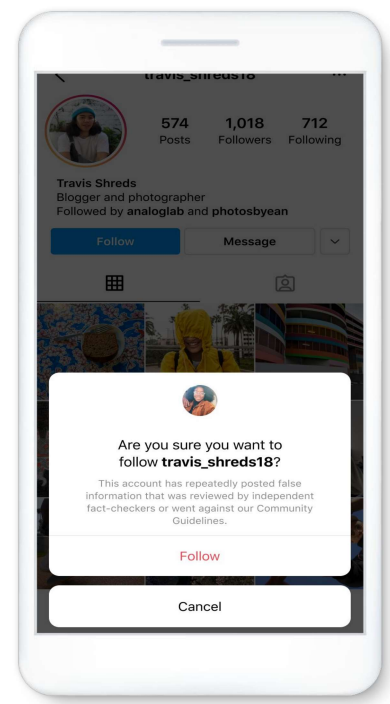
利用者への通知



警告・注意ラベル



ファクトチェックの通知



敵対的脅威となる勢力の探知・排除及び偽アカウントの摘発



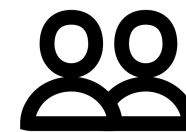
正真性 (Authenticity)

Metaのセキュリティシステムはバックグラウンドで毎秒数百万回の探知を実行。



偽アカウントの特定

AIにより、報告を受ける前に99.3%以上の偽アカウントを特定・削除



人の目による監視

安心・安全に係るスタッフを40,000人以上に三倍増。

敵対的脅威となる勢力の探知と排除



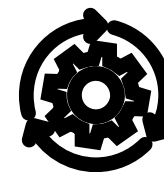
探知

悪影響を及ぼす選挙関連の行為は人の目により警戒探知



対抗措置

疑わしい活動に対してセキュリティ対応チームが対応・違反アカウントの削除



適応化・順応化

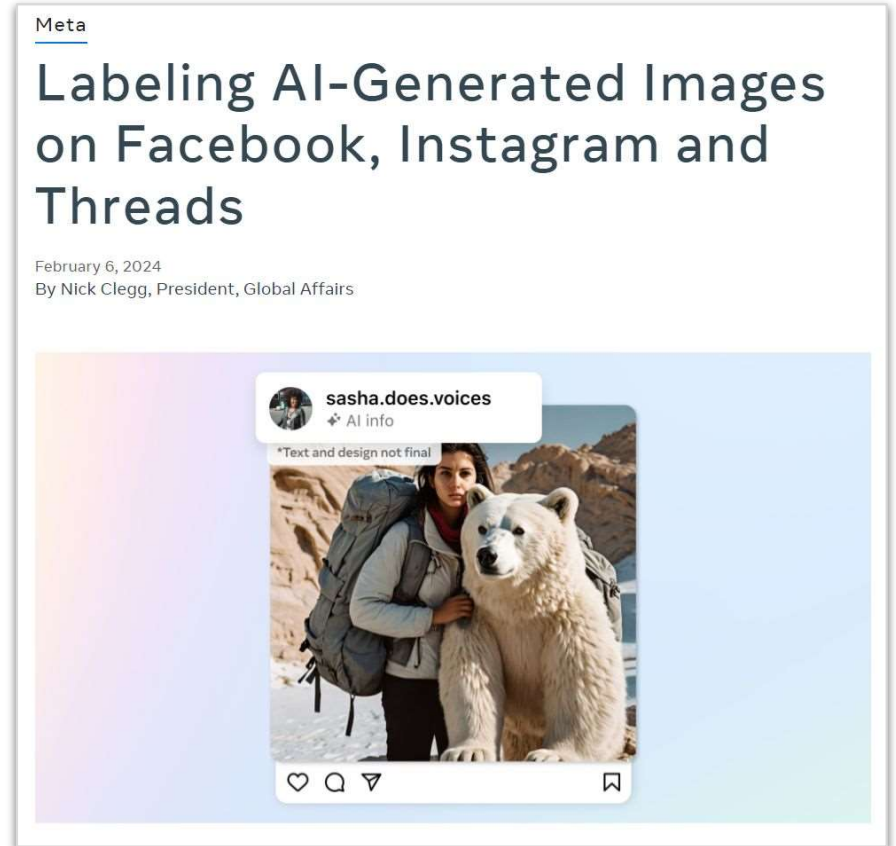
世界規模で行われる事象から得られる知見を活かして継続的に改善

コスト引き上げと利得の引き下げ

- 該当勢力による作戦実施を困難にする
- 作戦の一挙殲滅
- 採った措置について情報公開
- プラットフォームを横断した協力関係

生成AIによるコンテンツに対する弊社のアプローチ

- AIによるコンテンツを識別するための共通の技術標準について業界パートナーと協力
- 利用者がFacebook、Instagramに投稿した画像がAIによって生成されたものであることを示す業界標準の指標を検出できた場合、その画像にラベリング
- Meta AIを使用して作成されたフォトリアリスティックなイメージにそれが識別可能なように「Imagined with AI」とラベリング
- AIによって生成されたコンテンツや広告を投稿する利用者のための自己開示のためのツール



TRANSPARENCY

AI システム カード

- AIを説明するための新しいツール
- 弊社のAIシステムがコンテンツをランク付けする方法に関する情報
- どのコンテンツが利用者に最も関連性があるかを判断するために行う各システムについて説明
- 体験をカスタマイズするために使用できるコントロールを利用者に譲与

Facebook recommender systems

Filter by...

<p>FACEBOOK Feed</p> <p>When you view and interact with Facebook, one of the underlying AI systems delivers the connected content you see in your Feed, which is content you've chosen to see.</p> <p>→</p>	<p>FACEBOOK Feed Ranked Comments</p> <p>When you view and interact with Facebook, one of the underlying AI systems shows you comments on posts in your Feed that are ranked in order of relevance to you.</p> <p>→</p>	<p>FACEBOOK Feed Recommendations</p> <p>When you view and interact with Facebook, one of the underlying AI systems delivers suggested content to your Feed on the Facebook Home tab.</p> <p>→</p>
<p>FACEBOOK Reels</p> <p>When you view and interact with Facebook, one of the underlying AI systems delivers reels (short-form video content).</p> <p>→</p>	<p>FACEBOOK Stories</p> <p>When you view and interact with Facebook, one of the underlying AI systems delivers stories to you.</p> <p>→</p>	<p>FACEBOOK People you may know</p> <p>When you view and interact with Facebook, one of the underlying AI systems delivers personalized recommendations of people you may know.</p> <p>→</p>
<p>FACEBOOK Notifications</p> <p>When you view and interact with Facebook, one of the underlying AI systems delivers notifications to you, such as comments, recommendations and suggested content from friends.</p> <p>→</p>	<p>FACEBOOK Marketplace</p> <p>When you view and interact with Facebook, including Facebook Marketplace feed, one of the underlying AI systems recommends relevant Marketplace listings.</p> <p>→</p>	<p>FACEBOOK Video</p> <p>When you view and interact with Facebook Video, one of the underlying AI systems delivers a range of video types that may match your preferences.</p> <p>→</p>
<p>FACEBOOK Search</p> <p>When you view and interact with Facebook, one of the underlying AI systems delivers results when you search for content.</p> <p>→</p>	<p>FACEBOOK Groups Feed</p> <p>When you view and interact with Facebook Groups Feed, one of the underlying AI systems shows you status updates, photos, videos and other content from groups you follow and from groups you might be interested in.</p> <p>→</p>	<p>FACEBOOK Individual Group Feed</p> <p>When you visit a group and view and interact with the Individual Group Feed, one of the underlying AI systems delivers status updates, photos, videos and other content that has been posted to that specific group.</p> <p>→</p>
<p>FACEBOOK Suggested Groups</p> <p>When you view and interact with Facebook, one of the underlying AI systems suggests public or private groups to join that you may be interested in.</p> <p>→</p>	<p>FACEBOOK Pages You May Like</p> <p>When you view and interact with Facebook, one of the underlying AI systems delivers personalized Page recommendations to you.</p> <p>→</p>	

Instagram recommender systems

Filter by...

<p>INSTAGRAM Feed</p> <p>When you view and interact with Instagram, one of the underlying AI systems delivers the connected content you see in your feed, which is content from accounts that you follow.</p> <p>→</p>	<p>INSTAGRAM Stories</p> <p>When you view and interact with Instagram Stories, the underlying AI system automatically determines the order in which stories from people you follow show up in your stories.</p> <p>→</p>	<p>INSTAGRAM Explore</p> <p>Instagram Explore shows you content recommendations such as photos and reels from accounts you don't follow.</p> <p>→</p>
<p>INSTAGRAM Reels Chaining</p> <p>When you view and interact with Instagram, one of the underlying AI systems delivers reels (short-form video content) in the Reels tab.</p> <p>→</p>	<p>INSTAGRAM Search</p> <p>When you view and interact with Instagram, one of the underlying AI systems delivers results when you search for content.</p> <p>→</p>	<p>INSTAGRAM Feed Recommendations</p> <p>When you view and interact with Instagram, one of the underlying AI systems delivers suggested content to your feed.</p> <p>→</p>
<p>INSTAGRAM Suggested Accounts</p> <p>When you view and interact with Instagram, one of the underlying AI systems delivers suggested accounts you might want to follow.</p> <p>→</p>	<p>INSTAGRAM Notifications</p> <p>When you view and interact with Instagram, one of the underlying AI systems delivers notifications about new accounts to follow, activity from your connections, suggested content and more.</p> <p>→</p>	

<https://transparency.fb.com/features/explaining-ranking/>

Big Tech companies sign accord to combat AI-generated election disinformation



By Associated Press

Published on 17/02/2024 - 13:35



IFES PRESS RELEASE

NEWS AND UPDATES | Press Release

IFES Announces Voluntary Election Integrity Guidelines for Technology Companies

PUBLISHED
March 19, 2024

Google, Meta, Microsoft, The Rockefeller Foundation, Snap, and TikTok among supporters

<https://www.euronews.com/next/2024/02/17/big-tech-companies-sign-accord-to-combat-ai-generated-election-disinformation>

<https://www.ifes.org/news/ifes-announces-voluntary-election-integrity-guidelines-technology-companies>

MISINFORMATION

Metaは、世界60カ国語以上に対応する90以上のファクトチェック団体と協力し、Metaのプラットフォーム上に流れる誤情報を特定、検証、評価しています。

2024年2月、Metaは日本国内の適格なファクトチェック団体に第三者ファクトチェック・プログラムへの申請を呼びかけています。



90+

Fact-checking
organizations



60+

Languages

MISINFORMATION

THE FACT-CHECKING PROCESS



ファクトチェッカーは、広告、記事、写真、動画、テキストのみのコンテンツなど、FacebookやInstagramで公開された投稿を審査し、評価します。表現の自由の精神に基づき、通常、意見や討論、明らかに風刺的またはユーモラスなコンテンツ、ビジネス上の紛争には干渉しません。

優先される対象

- 選挙または危機的状況に関する主張
- 特定の民族、社会集団、宗教団体を標的とした主張
- 医薬品の使用や効果に関する主張
- 大きな金銭的損失につながる危険がある商品又はビジネスに関する主張

ファクトチェックの対象外

- 有名人のゴシップ、エンターテイメント、スポーツに関するニュース
- 歴史上の人物の言葉だと間違っているとらえられている引用
- エンターテイメントを扱っているメディア

Learn. Share. Grow.

FactChecking Fundamentals

- Metaの支援を受けて、IFCNがアジア太平洋州のジャーナリストを対象として提供するもの。
- 日本語を含め、15か国語に対応。
- 無料。
- 本コース修了した受講生には、その成果を称える修了証を授与。

内容

- (1)ファクトチェックの概要、
- (2)検証、デバンキング(虚偽の暴露)について、
- (3)健康・保健に関する mis/disinformation

<https://www.poynter.org/shop/fact-checking/fact-checking-fundamentals-japanese/>



リテラシー向上に向けて



www.wethinkdigital.fb.com

みんなのデジタル教室

日本での取り組み

日本では2020年12月に開始、NPO法人企業教育研究会と共同で日本に合わせたコンテンツを制作

全国の中学校、高校で2トピックの授業を提供し、**24,000名**以上が受講

86%の受講者がインターネットやアプリ、SNSの利用の仕方について考えが変わったと回答

「リソース」ページでは、安心安全のためのInstagramの機能の設定方法を紹介する動画なども提供

Meta

教育プログラム「みんなのデジタル教室」にデジタルシチズンシップの新授業を追加、都内中学校にて公開授業を実施

2023年07月18日

Metaは7月18日、国内で中高生向けに展開するデジタルリテラシー教育プログラム「みんなのデジタル教室」に、本日から新たな授業「デジタルシチズンシップと情報発信」を導入したことを発表しました。また、これに際して中央区立佃中学校の3年生・約40名を対象に公開授業を行い、授業の内容を検討するにあたって助言をいただいた総務省 情報流通行政局の担当者にも視察いただきました。



公開授業を中央区立佃中学校で実施

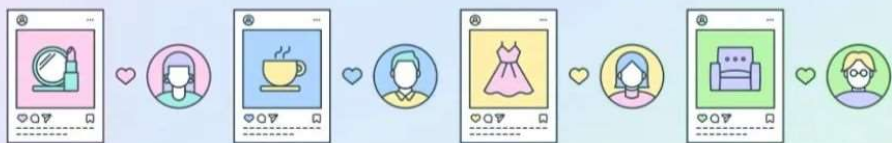
本日7月18日には、中央区立佃中学校の3年生・約40名を対象に新授業「デジタルシチズンシップと情報発信」を実施し、メディアにも公開しました。



授業では、端末を利用しながら、自分が住む地域で災害が起きたというシナリオで少人数のグループワークを実施。地域の人たちがどんな情報を必要としているのかを話し合ったあと、アレルギー対応の食品配給についての周知や復旧作業のボランティアの募集など、地域の人々の役に立つ投稿を作成しました。

仕組みを知ろう! パーソナライズ広告

フィードも広告も、自分好みにカスタマイズ



Meta

Metaは2月6日（日本時間）、InstagramやFacebook上で利用者の興味関心に合わせてパーソナライズされたコンテンツや広告を表示する仕組みについて学べるキャンペーン「仕組みを知ろう！パーソナライズ広告」を開始しました。インターネット上の安全に関する啓発など様々な取り組みが一斉に行われるセーファーインターネットデーに合わせてローンチするもので、人気動画クリエイターの古川優香さん（@iamyukaf）が、パーソナライズ広告の仕組みやアルゴリズム、自分好みに利用体験をカスタマイズする方法について学ぶ動画を特設サイトに公開します。

広告体験を自分好みにカスタマイズ

私たちは、パーソナライズ広告とプライバシーの保護は両立できると考えており、利用者の皆さんが自分自身でプライバシーや広告体験を管理しやすくなるような機能の開発や改善に取り組んでいます。

広告の右上にある[...]をクリックすると「この広告が表示されている理由」というオプションが表示され、広告主の配信設定などを確認することができます。また、広告トピックの設定を自分で変更したり、見たくないものがあれば非表示にもできます。

Metaは、利用者がより自分にあった広告が表示されるよう、ご自身で体験をカスタマイズできる仕組みづくりに力を入れています。



[詳しくはこちら](#)