



第3回会合における構成員からのご質問 (追加質問回答等)

2023年4月7日
事務局

伊藤構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：プラットフォームが、誹謗中傷への取組に対して、透明性やアカウントビリティを果たす際、政府やユーザーその他ステークホルダーへ説明をする場合に参考にする
 - ①ガイドラインの必要性の有無
 - ②ガイドラインの項目について各者様にお聞きできれば幸いです。

ヤフー株式会社

- ①（必要性の有無）については、事業者間での解釈のばらつきを防止し、公開内容の比較可能性を担保していく上では、参考として何らかのガイドラインをお示しいただくことが望ましいと考えています。
- ②（項目の設定）については、コンテンツモデレーションの実務の実態を十分に踏まえる必要があると考えます。ヒアリングの場でも述べさせていただきましたとおり、CGMサービスには様々な形態があり、これに応じてコンテンツモデレーションの取組も異なっています。同一の項目に関するデータがサービスによって様々に異なる含意を持つことも想定されますし、特定のデータの公表を求めることによって事業者のインセンティブに不測の影響が生じることも考えられます。したがって、ガイドラインの項目については、少なくともガイドライン等において具体的かつ詳細なデータ等の公表を一律に求めることは適当ではなく、いかなるサービスであっても共通して求められるような高次の目標を項目としていただくことが適当と考えています。

LINE株式会社

- ①これまでも総務省の「プラットフォームサービスに関する研究会 第二次とりまとめ」や公益社団法人商事法務研究会の「インターネット上の誹謗中傷をめぐる法的問題に関する有識者検討会 取りまとめ」等を参考にして社内の取組みを検討して参りましたので、コンテンツモデレーションに関する一定の考え方があれば参考となります。なお、誹謗中傷等への対策にあたっては、プラットフォーム事業者が自らのサービスの特性に応じて、効果的な取組を自律的な創意工夫により実施することが原則であると考えおりますので、ガイドライン等を定める場合には考慮していただきたいと思います。
- ②判例集や事例集等など、削除請求を受けた際の判断に資する事例が整理されたものがあれば、プラットフォーム事業者による削除の判断がしやすくなるものと考えております。



伊藤構成員	<ul style="list-style-type: none">■ 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.■ 質問内容：プラットフォームが、誹謗中傷への取組に対して、透明性やアカウントビリティを果たす際、政府やユーザーその他ステークホルダーへ説明をする場合に参考にする<ul style="list-style-type: none">①ガイドラインの必要性の有無②ガイドラインの項目 <p>について各者様にお聞きできれば幸いです。</p>
--------------	---

Google LLC

--

伊藤構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
 - 質問内容：プラットフォームが、誹謗中傷への取組に対して、透明性やアカウントビリティを果たす際、政府やユーザーその他ステークホルダーへ説明をする場合に参考にする
 - ①ガイドラインの必要性の有無
 - ②ガイドラインの項目
- について各者様にお聞きできれば幸いです。

Meta Platforms, Inc.

- 多くのプラットフォームがそれぞれのコンテンツポリシーの透明性を提供しています。Metaは、すべてのポリシーを一括してこのサイトで提供しています。 https://www.facebook.com/policies_center/
- コミュニティスタンダードでは、ポリシーの根底にある原則を公開しています。 <https://transparency.fb.com/en-gb/policies/community-standards/>
- これには当社のサービス上で何が許され、何が許されないかについての詳細な事例も示されています。
- また、Transparency Centerにおいて、私たちのポリシーの施行実施方法に関する透明性のあり方について説明しています。 <https://transparency.fb.com/enforcement/>
- 政府が透明性に関するガイドラインを設定する際の課題として、プラットフォームによって運営方法が大きく異なるため、業界間で共通の基準を設定することが困難であることが挙げられます。日本で利用可能なソーシャル・ネットワーキング・サービスの多様性を考えると、すべてのプラットフォームに適用できる単一の基準を設けることは現実的ではないと考えます。

Twitter, Inc.

- ① Twitterは真実の裁定者ではなく、何が真実で何が誤りかを判定する立場にはありません。コンテンツが違法とみなされる国においてのみ、コンテンツへのアクセスを制限します。
- ② Twitterでは、特定のコンテンツに対して、ユーザーが身の危険を感じる場合があります。Twitterサービスでは、すべての利用者に自由な自己表現を認めていますが、嫌がらせや脅し、又は恐怖を与えて他の利用者が発言できないようにする行為は禁止しています。さらに、他のユーザーの個人情報、明確な許可を受けずに公開または投稿することは禁じられています。

上沼主査代理

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：透明性について、公開そのものが自己目的化しないことへの懸念や、公開が意味があることを要する、公表データの悪用などの意見を頂戴した。公表に関し、どのような内容であれば、上記懸念が払拭されるのかについて、具体的にうかがいたい。

ヤフー株式会社

- 伊藤構成員への回答のとおりです。（以下、再掲）
- ①（必要性の有無）については、事業者間での解釈のばらつきを防止し、公開内容の比較可能性を担保していく上では、参考として何らかのガイドラインをお示しいただくことが望ましいと考えています。
- ②（項目の設定）については、コンテンツモデレーションの実務の実態を十分に踏まえる必要があると考えます。ヒアリングの場でも述べさせていただきましたとおり、CGMサービスには様々な形態があり、これに応じてコンテンツモデレーションの取組も異なっています。同一の項目に関するデータがサービスによって様々な異なる含意を持つことも想定されますし、特定のデータの公表を求めることによって事業者のインセンティブに不測の影響が生じることも考えられます。
したがって、ガイドラインの項目については、少なくともガイドライン等において具体的かつ詳細なデータ等の公表を一律に求めることは適当ではなく、いかなるサービスであっても共通して求められるような高次の目標を項目としていただくことが適当と考えています。

LINE株式会社

- コンテンツモデレーションの判断基準を公表することとすると、公表情報の悪用（判断基準のすり抜け、判断基準に定義されていない違法・有害情報に関する投稿等）の懸念がありますが、コンテンツモデレーションの透明性の確保や、情報の投稿を行うユーザーの予見可能性を高めるため、措置の対象となる情報や行為、取り得る措置の内容を公表することは適当であると考えております。
- なお、LINE オープンチャットにおいては、誹謗中傷に該当する可能性のある行為や誹謗中傷等にあたる判断した場合の措置等を「誹謗中傷に関する詳しいガイドライン」 (<https://lin.ee/LlvPTfu/gyjp>) として定めて2022年12月に公表するとともに、ルームに入室する際に適宜ポップアップで警告し注意喚起に努めています。



上沼主査代理

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：透明性について、公開そのものが自己目的化しないことへの懸念や、公開が意味があることを要する、公表データの悪用などの意見を頂戴した。公表に関し、どのような内容であれば、上記懸念が払拭されるのかについて、具体的にうかがいたい。

Google LLC

上沼主査代理

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：透明性について、公開そのものが自己目的化しないことへの懸念や、公開が意味があることを要する、公表データの悪用などの意見を頂戴した。公表に関し、どのような内容であれば、上記懸念が払拭されるのかについて、具体的にうかがいたい。

Meta Platforms, Inc.

- 私たちは、透明性が説明責任への第一歩であると考えます。しかし、透明性を絶対的な善として扱うべきではありません。透明性の目的と意義は、リスクやコストと比較して検討される必要があり、有意義なものでなければなりません。公共の利益に役立つ重要かつ関連性の高い問題を明らかにし、社会的信頼を高め、啓蒙することが目的である必要があります。
- 透明性に関するルールやガイドラインの策定は、以下の原則を考慮する必要があります：
 - どのような問題に答えようとしているのか対象を明確にする
透明性は、社会的対話の促進に資する重要な課題・問題への対応や 情報提供を目指すことで、意義あるものとなるべきです。プラットフォームに特定の指標を公表するよう求める場合、その指標がどのような問いに情報を提供し、回答するのか、そしてなぜそのような問いに価値があるのかを明確にすることが重要です。
 - 柔軟性を持たせるべきである。
デジタルプラットフォームの種類は増え、それぞれ異なる利用者と使用例があります。プラットフォームの形態や 取り組みの多様性を考慮した上で、質問項目を設定することで、企業はそれぞれのプラットフォームに最も適した方法で透明性の要求に対応することが可能になり、また柔軟な対応をする余地が生じます。
 - 異なる価値観のバランスを考慮する
プラットフォームは、表現の自由、安全性、信頼性、プライバシーなど、しばしば緊張関係にあるさまざまな価値観を、大規模に運用しながら両立させる必要があります。データの透明性は、これらすべての価値を考慮した上で情報を提供し、利用者がプラットフォームがコンテンツモデレーションにおいて実施した全体的な価値観のトレードオフを理解できるようにする必要があります。

上沼主査代理

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：透明性について、公開そのものが自己目的化しないことへの懸念や、公開が意味があることを要する、公表データの悪用などの意見を頂戴した。公表に関し、どのような内容であれば、上記懸念が払拭されるのかについて、具体的にうかがいたい。

Meta Platforms, Inc. (続き)

- 競合する要求とプラットフォームの能力に対する制約を考慮する必要がある。
価値観のバランスの問題に関連して、正確な評価と透明性は、実施、利用者のプライバシー、ストレージの制限、リソースのトレードオフ、実用性など、追加的で時に競合する懸念に配慮した方法で実施されなければなりません。
- 成果の指標に焦点を当てるべきである。
成果に焦点を当てた指標と、成果につながる構成要素に焦点を当てた指標を区別することが重要です。デジタルプラットフォームの多様性を考慮すると、成果につながる構成要素は大きく異なり、定義することが困難な場合があります。基本的に重要な質問に答えることに根ざしたデータ透明化アプローチでは、構成要素ではなく、成果に焦点を合わせます。
- このような観点から、Metaは9つの透明性レポートを発行しています：<https://transparency.fb.com/ja-jp/>

Twitter, Inc.

- 透明性に関するレポートにおいて、日本の内訳を公表しています。しかし、個人に関する情報については安全性やセキュリティの観点から非公開。
(<https://transparency.twitter.com/en/reports/countries/jp.html>)
- Twitterの使命は、公共の会話を促進し、保護することです。つまり、インターネットの広場になることです。透明性はTwitterにおける重要な指針の1つです。そのため、Twitterはコードの一部をGitHubに公開しています。
(https://blog.twitter.com/en_us/topics/company/2023/a-new-era-of-transparency-for-twitter)

上沼主査代理

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：ある投稿の権利侵害性の有無の判断が難しいとの意見があったが、プロバイダ責任制限法には、意見照会を行い、7日以内に返答がない場合には対発信者の免責が得られるという形で、権利侵害性の有無に踏み込まず、外形的な手続によって免責を得ることができる仕組みがある。この制度は活用しているか。活用していないとすれば、その理由は何か。

ヤフー株式会社

- 当社においては、投稿の削除の請求があった場合は、必要に応じて意見照会を行うこととしています。
- なお、請求の中には、理由の記載や証拠の添付が不十分であるなど、実体要件充足性について判断しがたいものもあります。意見照会をすること自体が、場合によっては発信者に対する萎縮効果を生じさせることもあり得ることを踏まえると、請求のあった全件ではなく、プロバイダ事業者において一定の実体判断を行った上で意見照会を行うことが適当と考えています。

LINE株式会社

- プロバイダ責任制限法に基づく対応をする際、発信者へ連絡することが可能な場合には発信者へ申立者から送信防止措置依頼がきていること及び同依頼内容について通知を行っています。この際、送信防止措置に不同意の場合には同意しない旨を返信することを依頼するとともに、不同意の返信がない場合にはLINE側で該当する投稿等を削除する旨を通知し、回答期日（7日以内）までに返答がない場合には削除しております。

上沼主査代理

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：ある投稿の権利侵害性の有無の判断が難しいとの意見があったが、プロバイダ責任制限法には、意見照会を行い、7日以内に返答がない場合には対発信者の免責が得られるという形で、権利侵害性の有無に踏み込まず、外形的な手続によって免責を得ることができる仕組みがある。この制度は活用しているか。活用していないとすれば、その理由は何か。

Google LLC

上沼主査代理

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：ある投稿の権利侵害性の有無の判断が難しいとの意見があったが、プロバイダ責任制限法には、意見照会を行い、7日以内に返答がない場合には対発信者の免責が得られるという形で、権利侵害性の有無に踏み込まず、外形的な手続によって免責を得ることができる仕組みがある。この制度は活用しているか。活用していないとすれば、その理由は何か。

Meta Platforms, Inc.

- 弊社では、利用者から削除請求を受けた際に件のコンテンツを発信した利用者に通知をするプロセスは設けておりません。弊社プラットフォームの利用者数の大きさに鑑み、全ての削除請求に対して意見照会を行うことは運用上困難です。また、利用者からの通報には、弊社のポリシーや法律には違反しておらず、ただ単に気に入らないとか同意できないというコンテンツの通報も数多くあります。従って、利用者からの削除請求に基づいて、削除請求の内容の検討なしに自動的にコンテンツを削除するような仕組みを採用することは、ポリシーの過剰な施行、ひいては表現の自由の抑制に繋がるものと危惧します。
- 削除請求が正当なものかを判断するには、それぞれの請求がなされた文脈を理解する必要があります。そういった判断は、コンテンツを発信した利用者に意見照会をする前に行う必要があります。さらに、削除請求が共有されたコンテンツに対してなされた場合、元のコンテンツの内容を必ずしも反映していない場合もあります。
- その代わり弊社一般的に、私たちは、弊社私たちのポリシー違反またはに抵触して削除された、または違法との判断によりされ投稿を削除した場合、その投稿の発信投稿者/作成者に通知を送ります。しかしが、その通知が該当の企業または個人に危険をもたらす場合はその限りではありません。（例えば、当社のコミュニティ規定にある危険な人物および団体に関するポリシーに違反するコンテンツについては、セキュリティ上のリスクをもたらす可能性があるため、投稿者/作成者に通知を送らない場合があります。）利用者は、削除に同意しない場合、アピールすることができます。

Twitter, Inc.

- 意見照会により発信者情報開示の民事訴訟で有用な情報を投稿者から得られる可能性は極めて低いため、活用はしていない。

上沼主査代理

- 質問先 : Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容 : Good Samaritanが保護される法制度を希望とのことだったが、日本では削除対応等を行うことが編集と同様として、ホスティングするコンテンツについての責任追及につながるとはされていないと解している。寧ろ、プロバイダ責任制限法では、「情報の流通」を知っていることが責任を問われる要件であることから「知っていて対応しない」などが責任の根拠となっており、善意の行為が責任の根拠とはされていない。日本の法制度において、Good Samaritanへの対応が不十分とされるのはどのような点なのか具体的にうかがいたい。

Google LLC

上沼主査代理

- 質問先：Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：Good Samaritanが保護される法制度を希望とのことだったが、日本では削除対応等を行うことが編集と同様として、ホスティングするコンテンツについての責任追及につながるとはされていないと解している。寧ろ、プロバイダ責任制限法では、「情報の流通」を知っていることが責任を問われる要件であることから「知っていて対応しない」などが責任の根拠となっており、善意の行為が責任の根拠とはされていない。日本の法制度において、Good Samaritanへの対応が不十分とされるのはどのような点なのか具体的にうかがいたい。

Meta Platforms, Inc.

- 善きサマリア人による保護は、企業が責任リスクを負うことなく、不快なコンテンツを検出し削除するための措置を自主的に実施することを可能にします。
- 有害なコンテンツに対して自主的に対策を講じることは、プラットフォームにとって利益となることがしばしばあります。
- プラットフォームはそうすることを義務付けられるべきではありませんが、PLLAに基づくセーフハーバーの権利を犠牲にする恐れなしに、そうする権限を与えられるべきです。
- PLLAの通知および削除の制度と互換性のある法的枠組みに善きサマリア人の原則を組み込むことは、プラットフォームによる自発的な措置が不完全であることが判明した場合にその責任が追及されるリスクを減らすことになり、結果としてプラットフォームによる自発的な措置を取ることを可能とすると考えます。

Twitter, Inc.

- Twitterは、ユーザーの表現の自由と安全性を最優先にしています。イーロン・マスク氏曰く「言論の自由は確保するが、リーチの自由は確保しない」。<https://twitter.com/elonmusk/status/1593673339826212864>
- Twitterは、最近更新した暴力的な発言に関するポリシーのように、常にポリシーの進化を注力しています。<https://help.twitter.com/ja/rules-and-policies/violent-speech>

上沼主査代理

- 質問先：LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：削除義務について、表現の自由への考慮が必要とのことであったが、具体的にはどのような考慮が必要ということか。削除義務の対象が「違法情報」に限った場合でも同様か。

LINE株式会社

- 当社の提供サービスにおいては、違法行為や公の秩序に反するおそれのある行為を一切許容しておらず、利用規約において禁止事項として定めており、違法行為等が確認された場合には、投稿の削除や閲覧制限、LINEアカウントの停止等の措置を行っています。対応に当たっては、画一的な対応を求めることとなれば、実際には権利侵害情報ではない疑わしい情報を全て削除することにつながりかねず、表現の自由に著しい萎縮効果をもたらす可能性があるため、「被害者救済」と発信者の「表現の自由」という重要な権利・利益のバランスに配慮しつつ、削除等の適切な対応を行う必要があると考えております。



上沼主査代理

- 質問先：LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：削除義務について、表現の自由への考慮が必要とのことであったが、具体的にはどのような考慮が必要ということか。削除義務の対象が「違法情報」に限った場合でも同様か。

Google LLC

上沼主査代理

- 質問先：LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：削除義務について、表現の自由への考慮が必要とこのことであったが、具体的にはどのような考慮が必要ということか。削除義務の対象が「違法情報」に限った場合でも同様か。

Meta Platforms, Inc.

- コンテンツが非常に有害であり、削除される必要がある場合についても、このワーキンググループが表現の自由の重要性について慎重に検討していることを歓迎します。
- 有害なコンテンツに対する規制は、安全性と表現のバランスをとることが必要です。
- 表現の自由は、情報の自由な流れを含み、それを可能にする基礎的な人権でもあります。世界中の多くの政府がオンライン上で言論の自由を取り締まり、報道の自由が脅かされているのを目の当たりにすると、私たちはこのことがいかに重要であるかを思い知らされます。
- 何が有害であるかという判断は、非常に文脈依存的であり、定義が難しく、しばしば文化的主観が入り、法的にもあいまいな場合があります。政府は、コンテンツ（違法なコンテンツ）の制限は、ICCPR、関連する権威ある指針、または（緊急事態における）シラクサ原則で定義された、他の権利、国家の安全、公的秩序または公衆衛生への害を守るために、合法性、必要性、比例性の制約内でしか行えないことを明確にする必要があります。
- また、多くのハイテク企業がグローバル・ネットワーク・イニシアチブ（GNI）のメンバーであり、市民的及び政治的権利に関する国際規約（ICCPR）やビジネスと人権に関する国連指導原則などの国際人権条約に強く取り組んでいることを認識していただくことも重要です。GNIのメンバーであることにより、企業は定期的な人権監査の対象となり、コンテンツモデレーションの実践が人々の表現の自由に対する権利を侵害しないことを証明する必要があります。

Twitter, Inc.

- Twitterは、すべての利用者が安心して自由に表現できるよう取り組みを進めています。そのために、表現の自由と規制とのバランスを取っています。Twitterでは、利用者の表現を尊重していますが、該当する現地法に基づいて表示制限を実施することがあります。利用規約に違反するコンテンツの削除またはコンテンツへのアクセスの制限が行われます。

清水構成員

- 質問先：ヤフー株式会社
- 質問内容：ADRが簡易迅速な解決が可能になると考えているということであるが、
 - ①判断は誰が担うのが適当であると考えるか。
 - ②判断に納得ができない場合は裁判に移行することになるため、当初から裁判手続の方がよいのではないかという考え方についてはどう考えるか。
 - ③ADRだと迅速になり、裁判手続（特に仮処分）では迅速にならないというのは、どういう理由からか（ヤフーにおける事情のみでよいです）。

ヤフー株式会社

- 発表時間の関係でご説明が不足しており恐縮ですが、ADRの仕組みさえあれば、直ちに簡易迅速な解決が可能になると考えているものではございません。前提として、違法・有害投稿を巡る紛争の数は膨大であることから紛争解決はあくまで各事業者を中心として行われるべきであり、ADR機関のみで簡易迅速な解決を行うことは現実的ではないと考えております。また、最終的な紛争解決は裁判手続によって行われるべきであり、ADRが介在することによって裁判手続への移行が阻害されることは望ましくないと考えております。
- 一方で、仮処分手続においては、決定の理由が十分に示されていなかったり射程が明確ではない等の事情により、後々の基準として活かせるような解釈が生まれづらいという状況があります。また、被害者にとっては裁判費用が必要となるという問題もあります。こうしたことから、各事業者が参照できるような判断事例をなるべく多数、事例に即して蓄積・整理することによって、その判断の簡易化・迅速化を図っていくことが重要と考えます。
- ADRについては様々な形態があると認識しており、個別の紛争についてみた場合、仮処分手続とどちらが簡易・迅速であるかは区々であると思われませんが、例えば国民生活センターの紛争解決委員会のように類型性・多発性があったり論点が複雑であるような重要紛争を中心に取扱い、同種紛争の解決にもつなげる指針として処理結果の概要を公開するものも存在していると承知しています。

清水構成員

- 質問先：Google LLC
- 質問内容：名誉毀損について事実関係の情報収集をすることができない結果、判断することが困難である、判断する立場にないといった発言を度々されていたところ、意地悪な見方をすれば、「原則として放置する」という宣言にも聞こえる。そのため、
 - ① どういった場合であれば対応していくべきと考えているのか。
 - ② 逆にどういった情報提供があれば対応し得るのか、についての情報提供をオープンにしていくということは可能と考えるか。

Google LLC

- 質問 1
- 質問 2

清水構成員

- 質問先：Meta Platforms, Inc.
- 質問内容：プロセスではなく、表示頻度によって判断するべきという趣旨の発表があったが、
 - ①「表示頻度」というのはどういう趣旨になるか。表示頻度というと、広告表示の問題のようにも感じられ、各ユーザーの投稿はいつでも閲覧できる状態になっているはずであり、何を意味するものか説明してもらえるとありがたい。
 - ②ある投稿がデータとしては残っているものの、報告をした人には見えないようにする、というもののか。
 - ③それとも、報告があれば全体として当該投稿等は表示されないようにするというのか。それともそれ以外か。

Meta Platforms, Inc.

- インターネット上でコンテンツが害を及ぼすには、それが閲覧されることです。
- インターネットの性質上、コンテンツが閲覧される回数は均等ではありません。ほんの小さなコンテンツが非常に短い時間で拡散されることもあれば、他のコンテンツは長い間インターネット上にありながら誰にも見られずにいることもあります。有害コンテンツの取締りを理解するために用いる指標は、その点を考慮する必要があります。
- そのため、FacebookやInstagramで違反行為が人々にどのような影響を与えるかを測定するため、私たちは表示頻度を重要な指標とみなしています。私たちは、私たちの基準に違反するコンテンツが、Facebookで閲覧される総回数に対して、実際にどれくらいの頻度で見られているかを最も重要視しています。
- これは、私たちが呼吸する空気中の汚染物質の濃度を測定することに似ています。空気の質を測定する際、環境規制当局は、空気の何パーセントが二酸化窒素であるかを調べ、どの程度が人に有害であるかを判断します。表示頻度とは、インターネットにおけるそれに相当するもので、有害なものを目にした回数が何パーセントになるかを測定するものです。
- 私たちの目標は、ポリシー違反が私たちのサービスを利用する人々に与える影響を最小限に抑えることです。私たちは、この目標に対して私たちがどのような役割を果たしているかを測定するために、違反コンテンツの表示頻度を測定しています。
- 普及率は、FacebookやInstagram上のすべてのコンテンツの閲覧を考慮し、それらの閲覧のうち、違反コンテンツの閲覧数の推定割合を測定します。この指標は、違反コンテンツがもたらす影響が、そのコンテンツの閲覧回数に比例すると仮定しています。
- また、違反コンテンツの閲覧を防げなかった数（違反コンテンツの早期発見や見逃しなど）も、表示頻度の指標となります。

清水構成員

- 質問先：Meta Platforms, Inc.
- 質問内容：プロセスではなく、表示頻度によって判断するべきという趣旨の発表があったが、
 - ①「表示頻度」というのはどういう趣旨になるか。表示頻度というと、広告表示の問題のようにも感じられ、各ユーザーの投稿はいつでも閲覧できる状態になっているはずであり、何を意味するものか説明してもらえるとありがたい。
 - ②ある投稿がデータとしては残っているものの、報告をした人には見えないようにする、というものか。
 - ③それとも、報告があれば全体として当該投稿等は表示されないようにするということか。それともそれ以外か。

Meta Platforms, Inc. (続き)

- 違反コンテンツの表示頻度は、FacebookまたはInstagram全体からのコンテンツ閲覧のサンプルを使用して推定されます。違反コンテンツの推定閲覧数を、FacebookまたはInstagramの全コンテンツの推定閲覧数で割った値として計算されます。仮に成人向けヌードや性的行為の表示頻度が0.18%~0.20%だった場合、1万件のコンテンツビューのうち、平均18~20件は成人向けヌードや性的行為に関する当社の基準に違反するコンテンツが表示されたこととなります。
- 弊社は、四半期ごとに発行する「コミュニティ規定施行レポート」で普及率を公表しています。
- 表示頻度の詳細については、こちらをご覧ください。<https://transparency.fb.com/ja-jp/policies/improving/prevalence-metric/>

清水構成員

- 質問先：SIA
- 質問内容：
 - ①削除について新制度を作ることについては賛成方向ではないといった発言があったが、その理由についてお教えいただきたい。
 - ②新制度を作るにしても、特に海外事業者への執行ベースでの強制力を持たせることが必要という指摘があったが、具体的にはどういったことが考えられるか。できれば、間接強制以外の方法や、仕組みについてアイデアがあれば伺いたい。

SIA

- ① 具体案が示されるまでは賛否については留保するという趣旨です。
- ② 現時点では具体的なアイデアはございません。

藤原構成員

- 質問先：Google LLC（及び、もし何かご意見があれば全事業者）
- 質問内容：貴社からは、透明化の対象情報が明確でないと言務の履行に支障がある旨、ご説明をいただきました。透明化を図るべき対象情報としては、違法情報（権利侵害情報及び法令違反情報）のみとすることが適切であると考えているのでしょうか。それとも、現時点では法令上定義が定かではない有害情報のようなものであっても、新たに定義を置くことによって明確化される限り含まれ得る余地があるとお考えでしょうか。ご教示いただけますと幸いです。

Google LLC

藤原構成員

- 質問先：Google LLC（及び、もし何かご意見があれば全事業者）
- 質問内容：貴社からは、透明化の対象情報が明確でない義務の履行に支障がある旨、ご説明をいただきました。透明化を図るべき対象情報としては、違法情報（権利侵害情報及び法令違反情報）のみとすることが適切であると考えているのでしょうか。それとも、現時点では法令上定義が定かではない有害情報のようなものであっても、新たに定義を置くことによって明確化される限り含まれ得る余地があるとお考えでしょうか。ご教示いただけますと幸いです。

ヤフー株式会社

- どのような情報が有害情報に該当するかは、情報の受け手となるユーザーの年齢等の属性、発信されているプラットフォームの属性等により変動するため、仮に、有害情報も透明化の対象とするのであれば、対象範囲が明確となるよう、有害情報の定義を明確化すべきと考えます。

Meta Platforms, Inc.

- 上沼主査代理からのご質問（「透明性について、～」）で述べたように、透明性に関するルールやガイドラインは、以下の原則を考慮する必要があります：
 - 私たちが解決しようとしている重要な問題（または課題）に答えるものであること。
 - 柔軟性を持たせること。
 - 異なる価値観のバランスを保つこと。
 - 競合する要求や、プラットフォームが測定する能力の制限を考慮する必要があること。
 - 成果指標に焦点を当てること。
- このような理由から、Metaでは、幅広いカテゴリーで定期的に透明性レポートを発行しています：
 - コミュニティ規定施行レポート
 - 知的財産権の行使
 - 政府によるユーザーデータの提供要請
 - コンテンツの制限
 - 敵対的脅威に関するレポート
- これらは通常の透明性確保のツールや情報などに加えて実施されているものです。

藤原構成員

- 質問先：Google LLC（及び、もし何かご意見があれば全事業者）
- 質問内容：貴社からは、透明化の対象情報が明確でないと言務の履行に支障がある旨、ご説明をいただきました。透明化を図るべき対象情報としては、違法情報（権利侵害情報及び法令違反情報）のみとすることが適切であると考えているのでしょうか。それとも、現時点では法令上定義が定かではない有害情報のようなものであっても、新たに定義を置くことによって明確化される限り含まれ得る余地があるとお考えでしょうか。ご教示いただけますと幸いです。

Twitter, Inc.

- Twitterは法律に従い、その上法律にないところまでポリシーに含まれてます。現在のTwitterの透明性に関するレポートは、世界各地から寄せられる法的請求の報告・対応やTwitterルールに基づく強制的対応が記載されています。国別のレポートを作成するには精度・背景・有用性など様々な課題があります。有害情報の定義は明確ではないが、Twitterはポリシーの継続的な進化を推進しています。

藤原構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：ユーザからのポリシー違反に関する申告及び法的請求について、迅速に削除するためあるいは迅速に削除の可否を判断するために、申告や請求にはどのような情報が含まれるべきで、どのような証拠が提出されるべきとお考えでしょうか。また、実際に申告を行うユーザに対して、情報提供や証拠提出について具体的にどのような指示や依頼をしているのでしょうか。ご教示いただけますと幸いです。

ヤフー株式会社

- ポリシー違反の内容により、削除の可否の判断にあたって必要となる情報は異なるため、一概にどのような情報や証拠が必要ということを上記することは困難です。例えば、苛烈な表現を用いた誹謗中傷であれば、表現自体で削除の可否が判断でき、ユーザーからの情報提供や証拠提出は要しないため、申告のみでの判断が可能です。他方、名誉権侵害の申告では真実性が争点となる場合が多く、その場合、申告フォームからの限られた情報では判断が困難であることも少なくありません。現状の多くは、書面による投稿削除請求の際に、申告者から虚偽であることを裏付ける証拠等の提出を受け判断しているため、申告時にそうした情報や証拠提出があれば、削除の可否の判断が迅速になる可能性があります。
- また、申告者からの情報提供や証拠提出が充実するほど、削除の可否の判断は容易になるとは思われますが、下記①の違反投稿の検知手段という意味では、申告時の情報提供や証拠提出を多く求めるとかえって申告件数の減少につながるおそれがあるため、バランスが難しいところと考えています。
- ユーザーからの違反申告には、①違反投稿の検知手段、②被害救済の申出という2つの機能があると考えており、当社は①の手段として間口を広くするため、例えば「Yahoo!ニュースコメント」では、違反申告の理由は必須としているものの、選択式かつ匿名での簡易な申告を可能としています。他方、書面による投稿削除請求の場合は、侵害されたとする権利や権利が侵害されたとする理由を具体的かつ詳細に記載するように依頼しているほか、本人確認のため、身分証等の写しの添付を求めています。

藤原構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：ユーザからのポリシー違反に関する申告及び法的請求について、迅速に削除するためあるいは迅速に削除の可否を判断するために、申告や請求にはどのような情報が含まれるべきで、どのような証拠が提出されるべきとお考えでしょうか。また、実際に申告を行うユーザに対して、情報提供や証拠提出について具体的にどのような指示や依頼をしているのでしょうか。ご教示いただけますと幸いです。

LINE株式会社

■ 質問前段

- 違法・有害情報を目撃したユーザーからの通報により、プラットフォーム事業者が違法・有害情報を認知し、速やかな削除等の対応を可能とするためには、サービス画面上に通報ボタンを設けるなど、簡便に通報できるような手段を設けておくことが有効であると考えています。（複雑な通報フォームにすると、かえって通報をためらうことも考えられるため。）
- 一方で、違法・有害情報の投稿による権利侵害の被害者が当該情報の削除を求める場合には、プラットフォーム事業者のポリシーに基づくコンテンツモデレーションの申請や法的な削除請求を行うにあたって、お問い合わせフォームを通じて申請や請求の理由を十分に説明していただくことが必要であると考えています。

■ 質問後段（「また、～」の部分）

- 通報をする際には、通報する理由（「スパム／宣伝目的」、「性的いやがらせ／出会い目的」、「迷惑行為」及び「その他」）をユーザが指定し、通報があった場合、通報理由に加え、通報するメッセージより前にやり取りしたメッセージ最大10件、通報するユーザに関する情報等が当社に送信されます。（LINE オープンチャットの場合）
- お問い合わせフォームを通じて権利侵害情報の削除を申請する際には、権利侵害を受けた者、通報の有無、ご希望の対応、問題コンテンツの投稿日時、問題コンテンツの投稿者、権利侵害の具体的な内容（問題が確認できるスクリーンショットの添付を含む。）等のうちユーザに記載していただいた情報が当社に送信されます。

藤原構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：ユーザからのポリシー違反に関する申告及び法的請求について、迅速に削除するためあるいは迅速に削除の要否を判断するために、申告や請求にはどのような情報が含まれるべきで、どのような証拠が提出されるべきとお考えでしょうか。また、実際に申告を行うユーザに対して、情報提供や証拠提出について具体的にどのような指示や依頼をしているのでしょうか。ご教示いただけますと幸いです。

Google LLC

藤原構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：ユーザからのポリシー違反に関する申告及び法的請求について、迅速に削除するためあるいは迅速に削除の可否を判断するために、申告や請求にはどのような情報が含まれるべきで、どのような証拠が提出されるべきとお考えでしょうか。また、実際に申告を行うユーザに対して、情報提供や証拠提出について具体的にどのような指示や依頼をしているのでしょうか。ご教示いただけますと幸いです。

Meta Platforms, Inc.

- 私たちは、利用者が報告した場合にその進捗の確認や対処状況について情報を提供できるように、報告手順について案内するテンプレートを提供しています。
- 例えば、Instagramの投稿やプロフィールを報告する方法について、ヘルプセンターの項目で詳細を説明しています。
(<https://help.instagram.com/192435014247952>)
- 法的な苦情に関しては、ヘルプセンターでは、必要な情報を確保するため違反の種類ごとに個別のテンプレートを用意しています。例えば、名誉毀損に関する苦情の報告用テンプレート (<https://help.instagram.com/contact/653100351788502>) と、著作権侵害に関する苦情の報告用テンプレート (<https://help.instagram.com/contact/552695131608132>) を用意しています。

Twitter, Inc.

- 質問前段
 - Twitterルールまたは利用規約に違反しているコンテンツを見つけた場合は報告できます。Twitterの決定が誤りだと思われる場合は異議申し立てをすることもできます。現在、報告方法を改善する事を検討しています。
- 質問後段（「また、～」の部分）
 - Twitterは、裁判所命令などの有効な法的請求に応じます。しかし、その場合Twitterは相手方当事者になり、訴訟中と利益相反になるため、必要な情報や証拠については控えます。

水谷構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：先日のヒアリングで違法有害情報への取り組みがPF自身のビジネスにとっても不可欠であるという点は重要だと思いました。また透明化の各社の取り組みも理解できました。他方で、透明化したデータの「意味」が重要というお話や、事業者によってサービスは多様でありOne Size Fits Allアプローチは適当でない旨の発言があったと思いますが、それならばコンテンツモデレーションの運用方針とその結果の適正性を検証可能にするために、自社のサービスについて具体的にどのような数値等の透明化を行えば良いとお考えでしょうか。

ヤフー株式会社

- 伊藤構成員への回答のとおりです。（以下、再掲）
- ①（必要性の有無）については、事業者間での解釈のばらつきを防止し、公開内容の比較可能性を担保していく上では、参考として何らかのガイドラインをお示しいただくことが望ましいと考えています。
- ②（項目の設定）については、コンテンツモデレーションの実務の実態を十分に踏まえる必要があると考えます。ヒアリングの場でも述べさせていただきましたとおり、CGMサービスには様々な形態があり、これに応じてコンテンツモデレーションの取組も異なります。同一の項目に関するデータがサービスによって様々に異なる含意を持つことも想定されますし、特定のデータの公表を求めることによって事業者のインセンティブに不測の影響が生じることも考えられます。
したがって、ガイドラインの項目については、少なくともガイドライン等において具体的かつ詳細なデータ等の公表を一律に求めることは適当ではなく、いかなるサービスであっても共通して求められるような高次の目標を項目としていただくことが適当と考えています。

水谷構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：先日のヒアリングで違法有害情報への取り組みがPF自身のビジネスにとっても不可欠であるという点は重要だと思いました。また透明化の各社の取り組みも理解できました。他方で、透明化したデータの「意味」が重要というお話や、事業者によってサービスは多様でありOne Size Fits Allアプローチは適当でない旨の発言があったと思いますが、それならばコンテンツモデレーションの運用方針とその結果の適正性を検証可能にするために、自社のサービスについて具体的にどのような数値等の透明化を行えば良いとお考えでしょうか。

LINE株式会社

- プラットフォーム事業者としての透明性を高めるため、LINE Transparency Report (<https://linecorp.com/ja/security/transparency/top>) にて、違反投稿への対応状況やユーザからの削除申告への対応状況等を公開しています。
- 一方、サービス上における違法・有害情報の総流通量を測ることは困難であるとの事情から、例えば、評価期間中にコンテンツモデレーションの実施件数が増加した場合、ある側面では確実にコンテンツモデレーションを実施しているとの評価になり得るが、別の側面では違法・有害情報の流通量が増加しているともいえるため、評価基準を一様に定めることには一定の限界があるものと考えています。そのため、評価基準については、他のプラットフォーム事業者の数値と比較可能なものとするのが適当であり、その時々違法・有害情報の流行、システムの利用者による通報等の閾値の変化、システムの検知精度など、様々な要素が影響することを踏まえつつ、実効性のある評価スキームの確立が必要であると考えています。

水谷構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：先日のヒアリングで違法有害情報への取り組みがPF自身のビジネスにとっても不可欠であるという点は重要だと思いました。また透明化の各社の取り組みも理解できました。他方で、透明化したデータの「意味」が重要というお話や、事業者によってサービスは多様でありOne Size Fits Allアプローチは適当でない旨の発言があったと思いますが、それならばコンテンツモデレーションの運用方針とその結果の適正性を検証可能にするために、自社のサービスについて具体的にどのような数値等の透明化を行えば良いとお考えでしょうか。

Google LLC

水谷構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：先日のヒアリングで違法有害情報への取り組みがPF自身のビジネスにとっても不可欠であるという点は重要だと思いました。また透明化の各社の取り組みも理解できました。他方で、透明化したデータの「意味」が重要というお話や、事業者によってサービスは多様でありOne Size Fits Allアプローチは適当でない旨の発言があったと思いますが、それならばコンテンツモデレーションの運用方針とその結果の適正性を検証可能にするために、自社のサービスについて具体的にどのような数値等の透明化を行えば良いとお考えでしょうか。

Meta Platforms, Inc.

- 清水構成員からのご質問への回答で述べた理由から、私たちは、表示頻度は私たちのコミュニティスタンダード執行の有効性を考慮する際に有用な指標であると考えます。
- さらに、事前検知率も有用な指標であると考えます。これは、有害なコンテンツが人々に閲覧される前に私たちが削除した量だからです。具体的には、事前検知率は、Instagramを利用する人々が報告する前に私たちが発見して対処したコンテンツの数を、私たちが対処したコンテンツの総数で割った値です。例えば、2022年第4四半期時点のInstagram上のいじめ&嫌がらせコンテンツについては、ユーザーからの報告を受けて削除したコンテンツは14.6%であるのに対し、当社の事前検知率は85.4%となりました。
- ただし、プラットフォームごとにユーザーコミュニティ、ユースケース、したがってリスクが異なるため、あるプラットフォームにとって適切で有意義な指標は、別のプラットフォームではそうでない場合があることに注意する必要があります。プラットフォームは、コンテンツモデレーションプラクティスの有効性を測定するための最も適切な指標を決定する柔軟性を与えられることが望ましいです。

Twitter, Inc.

- Twitter2.0は改革しています。ツイートインプレッションやブックマーク機能など、透明性を改善するための取り組みに引き続き注力しています。今後数週間から数ヶ月のうちにまたアップデートを共有します。



水谷構成員	<ul style="list-style-type: none">■ 質問先 : Google LLC/Meta Platforms, Inc./Twitter, Inc.■ 質問内容 : グローバル企業のGoogle、Meta、Twitterに関しては、上記に加えて、「日本の文脈での透明化の対応」を自主的にどこまで取り組んでいただけるのかという点をもっと明確にしていいただければと思います。例えば、検討アジェンダにあるコンテンツモデレーションの運用方針、運用結果、運用体制について、日本における固有の状況が明らかになることは、ユーザやユーザ以外の者に対する説明責任を果たす上で、最低限必要であるように考えられます。もしそうではないとするならば、どのような数値等の透明化が必要になるのでしょうか。
--------------	---

Google LLC

--

水谷構成員

- 質問先 : Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容 : グローバル企業のGoogle、Meta、Twitterに関しては、上記に加えて、「日本の文脈での透明化の対応」を自主的にどこまで取り組んでいただけるのかという点をもっと明確にしていいただければと思います。例えば、検討アジェンダにあるコンテンツモデレーションの運用方針、運用結果、運用体制について、日本における固有の状況が明らかになることは、ユーザやユーザ以外の者に対する説明責任を果たす上で、最低限必要であるように考えられます。もしそうではないとするならば、どのような数値等の透明化が必要になるでしょうか。

Meta Platforms, Inc.

- 私たちは、私たちのポリシー、技術、そして執行が日本にとって適切であることを確認するために、専門チームを編成しています。各チームは必要な言語能力を備えており、日本市場に関連する最新の情報を関係者と適時に共有できるように調整されています。また、各チームは世界各地に配置されており、24時間体制でコンテンツモデレーションの要望に応えられるようになっています。
- また、日本のパートナーや関係者と協力し、当社のポリシーやツール、リソースをより広く知ってもらうために、現地にチームを置いています。

Twitter, Inc.

- Twitterでは、さまざまな手段でより多くの情報を利用できるように取り組んでおり、サイト上でもわかるようにしています。
- また、重要な変更が発生したときにそれを理解するためのさまざまな方法にも取り組んでいます。独自調査をご覧ください：
 - <https://www.sprinklr.com/blog/identify-toxic-content-with-leading-analytical-ai>
 - <https://twitter.com/TwitterSafety/status/1638255718540165121>

水谷構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./SIA/SMAJ/Twitter, Inc.
- 質問内容：オンラインハラスメント対策という点でいうと、削除以外にも、アーキテクチャやナッジによる解決という方法も重視する必要があると思います。例えばRethink（警告表示）機能なども注目を集めています。またMetaから、AIによる検知のお話がありましたが、これも不可欠だと理解しています。そして、こうした仕組みのうちどのようなものが最善かは、やはりPF自身が検討とテストを重ねてもらう必要があると思います。アーキテクチャやアルゴリズム改善の取り組みをそれぞれの社で自律的に取り組んでいただけるようなインセンティブ設計という観点から、例えば日本政府や業界団体ができること、やってもらいたいことは何かあるでしょうか。

ヤフー株式会社

- プラットフォームごとの特性や規模に応じて有効な方策が異なることから、事業者自身が検討とテストを重ねる必要がありますが、導入コストや検証までの時間がかかることから、ベストプラクティスの共有や議論の場があることが望ましいと考えます。あわせて、透明性・アカウントビリティの確保策の検討に当たっては、こうした各事業者の創意工夫が阻害されないようご配慮いただけますと幸いです。

LINE株式会社

- 各事業者において、様々なアーキテクチャ上の工夫による違法・有害情報の被害低減に取り組んでおられますので、業界団体における意見交換や本ワーキンググループのような政府会合における情報共有が重要であると考えております。

水谷構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./SIA/SMAJ/Twitter, Inc.
- 質問内容：オンラインハラスメント対策という点でいうと、削除以外にも、アーキテクチャやナッジによる解決という方法も重視する必要があると思います。例えばRethink（警告表示）機能なども注目を集めています。またMetaから、AIによる検知のお話がありましたが、これも不可欠だと理解しています。そして、こうした仕組みのうちどのようなものが最善かは、やはりPF自身が検討とテストを重ねていってもらう必要があると思います。アーキテクチャやアルゴリズム改善の取り組みをそれぞれの社で自律的に取り組んでいただけるようなインセンティブ設計という観点から、例えば日本政府や業界団体ができること、やってもらいたいことは何かあるでしょうか。

Google LLC

水谷構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./SIA/SMAJ/Twitter, Inc.
- 質問内容：オンラインハラスメント対策という点でいうと、削除以外にも、アーキテクチャやナッジによる解決という方法も重視する必要があると思います。例えばRethink（警告表示）機能なども注目を集めています。またMetaから、AIによる検知のお話がありましたが、これも不可欠だと理解しています。そして、こうした仕組みのうちどのようなものが最善かは、やはりPF自身が検討とテストを重ねてもらう必要があると思います。アーキテクチャやアルゴリズム改善の取り組みをそれぞれの社で自律的に取り組んでいただけるようなインセンティブ設計という観点から、例えば日本政府や業界団体ができること、やってもらいたいことは何かあるでしょうか。

Meta Platforms, Inc.

- 過去数年間、私たちはネット上のいじめや嫌がらせに対処するために多くの投資を行い、その結果、Metaはいくつもの改善を行いました。
- ネット上のいじめや嫌がらせに対処するための基本的な問題は、プラットフォームのルールに反しないにもかかわらず、受け手が不快感や危険性を感じるようなやり取りがしばしば発生することです。そのため、プラットフォームの取り組みには、検知だけでなく、オンライン体験を自分で決められるようにするためのユーザー向けコントロールの手段が必要です。
- 私たちの取り組みには、AIベースの検出の改善も含まれています。外部監査を受けている四半期ごとのコミュニティ規定施行レポートでは、いじめとハラスメントについてグローバルで進歩を遂げています。2019年には、誰かが報告してくる前に、違反するいじめやハラスメントのコンテンツの36.2%を事前に対処することができましたが、現在ではそのようなコンテンツの85.4%を事前に対処できるようになっています。2023年2月の四半期報告書はこちらからご覧ください。
<https://transparency.fb.com/data/community-standards-enforcement/>
- また、私たちは、コミュニティのニーズやプラットフォームでのさまざまな経験について広く相談し、その結果、同じ時期に、ブロックやミュートといった従来の機能を超えて、制限、マルチブロック、非表示ワード、制限といったユーザーコントロールを開始し、世界中で利用できるようにしました。
- これまでお伝えしてきたように、ユーザーの属性やインタラクションの方法はプラットフォームによって異なるため、万能なソリューションは存在しません。業界は、それぞれのプラットフォームに適した仕組みづくりを進めるとともに、その進捗を共有することが望まれます。

水谷構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./SIA/SMAJ/Twitter, Inc.
- 質問内容：オンラインハラスメント対策という点でいうと、削除以外にも、アーキテクチャやナッジによる解決という方法も重視する必要があると思います。例えばRethink（警告表示）機能なども注目を集めています。またMetaから、AIによる検知のお話がありましたが、これも不可欠だと理解しています。そして、こうした仕組みのうちどのようなものが最善かは、やはりPF自身が検討とテストを重ねてもらう必要があると思います。アーキテクチャやアルゴリズム改善の取り組みをそれぞれの社で自律的に取り組んでいただけるようなインセンティブ設計という観点から、例えば日本政府や業界団体ができること、やってもらいたいことは何かあるでしょうか。

Meta Platforms, Inc.（続き）

- また、国際標準やベストプラクティスを確立するために、すでに行われている業界の取り組みを評価し、連携することが重要です。例えば、Digital Trust and Safety Partnership（DTSP）は、世界的に適用可能なコンテンツモデレーションのベストプラクティスを定め、内部および独立した第三者機関の評価を通じて、企業がそのベストプラクティスを実践しているのかを検証することを目的としています。DTSPは、規模もビジネスモデルも異なる10社以上のグローバルなハイテク企業が、インターネット上の「信頼と安全」を高めるための共通のアプローチに参加する、世界初の取り組みです。独立した、業界の枠にとらわれない、オンライン上の信頼と安全を高めるためのベストプラクティスを提供する、専門家グループです。これらのベストプラクティスは、消費者・ユーザー擁護団体、政策立案者、法執行機関、関連NGO、業界全体の様々な専門家の意見を聞き、製品開発のライフサイクル全体を見据えて、協力して開発されています。<https://dtspartnership.org/best-practices/>
- DTSPの詳細については、こちらをご覧ください。<https://dtspartnership.org/>

水谷構成員

- 質問先：ヤフー株式会社/LINE株式会社/Google LLC/Meta Platforms, Inc./SIA/SMAJ/Twitter, Inc.
- 質問内容：オンラインハラスメント対策という点でいうと、削除以外にも、アーキテクチャやナッジによる解決という方法も重視する必要があると思います。例えばRethink（警告表示）機能なども注目を集めています。またMetaから、AIによる検知のお話がありましたが、これも不可欠だと理解しています。そして、こうした仕組みのうちどのようなものが最善かは、やはりPF自身が検討とテストを重ねてもらう必要があると思います。アーキテクチャやアルゴリズム改善の取り組みをそれぞれの社で自律的に取り組んでいただけるようなインセンティブ設計という観点から、例えば日本政府や業界団体ができること、やってもらいたいことは何かあるでしょうか。

SIA

- 先生ご指摘のとおり、基本的にはユーザーに安心してサービスをご利用いただくためにも、各PFがそれぞれの特質に応じて機能開発していくべきものと考えており、実際上も各社継続的に努力し実装しているものと推測しております。一方で、そのような努力の結果として実際にどのような効果が出ているかについて、客観的に評価する仕組みがあると良いと思います。

SMAJ

- 各事業者において、様々なアーキテクチャ上の工夫による違法・有害情報の被害低減に取り組んでおられますので、業界団体における意見交換や本ワーキンググループのような政府会合における情報共有が重要であると考えております。

Twitter, Inc.

- Twitterは現在、対応の検討を進めています。会社にとって前例のない変化ですので、しばらくの忍耐と時間が必要です。さらに、ユーザーの安全には1つまたは複数のプラットフォームの取り組みだけでは不十分であり、あらゆる年齢層のユーザーに社会全体の追加教育とデジタルリテラシーへの投資が必要だと思えます。

水谷構成員

- 質問先：Twitter, Inc.
- 質問内容：報道を通じて、Twitterの現CEOが、社内会議で「日本が中心」とおっしゃられたと聞いています。日本は政治コミュニケーションにおいて実際にもTwitterに大きく依存しています。その一方で、特定の政治家のアカウント復活の是非や、凍結アカウントの「恩赦」をアンケート投票機能で決定したり、理由もよくわからない大量のアカウント凍結措置が起きるといったことが散見されます。こうした事象は、モデレーションの仕組みが恣意的に運用されているのではいかという疑念を抱かざるを得ないのですが、こうした点について今後どのように「自主的なガバナンス」を働かせていくおつもりか、その見通しについてのご説明をWGの一員として強く期待しております。

Twitter, Inc.

- Twitterは、日本でのサービスとユーザーに力を入れており、ルールとポリシーも整っています。ルールとポリシーの進化スピードについても公開しています。Twitterは、正かつ公平な方法でそれらを適用するための合理的な業務努力を精進しています。

山根構成員

- 質問先：Twitter, Inc.
- 質問内容：「コミュニティノートの作成を考えている。該当するかどうかを判断するためのテクノロジー・・・」との発言があったと思うが、どのようなものか知りたい。

Twitter, Inc.

- コミュニティノートは、Twitterでより正確な情報を入手できるようにすることを目的に作られた機能です。この機能により、誤解を招く可能性があるツイートに、Twitterユーザーが協力して背景情報を提供することができます。協力者はあらゆるツイートにコメントを残すことができ、十分な数の協力者がさまざまな視点から「役に立った」と評価すると、そのコメントがツイートに表示されます。協力者になりたい方は、こちら(<https://twitter.com/i/flow/join-birdwatch>)からお申し込みいただけます。
- コミュニティノートは、Twitterの見解を表すものではありません。また、Twitterの社内チームが、コミュニティノートを編集または改変することはできません。コミュニティノートが追加されたツイートは、Twitterルール(<https://help.twitter.com/ja/rules-and-policies/twitter-rules.html>)、利用規約(<https://twitter.com/ja/tos>)、またはTwitterのプライバシーポリシー(<https://twitter.com/ja/privacy>)への違反が認められた場合を除き、Twitterによるラベル付け、削除、またはその他の措置の対象となることはありません。規則の違反に対しては、コミュニティノートへのアクセス権の剥奪や、その他の違反防止措置が取られることがあります。
- 日本ではコミュニティノートが使えます。参加者の募集も開始しました。
- 最新情報は<https://twitter.com/HelpfulNotes>をご覧ください。

山口構成員

- 質問先：Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：お忙しい中、貴重なご報告をいただきまして、誠に有難う御座います。本日のGoogle LLCのご担当者様からの資料に基づくご報告では、「有害」(harmful)、「オンラインハラメント」に焦点を当てられていたと存じます。また、ご報告の最後の方で、「有害」な内容をいかに定義するか、という問題が一番重要とおっしゃっておられたかと存じます。Meta Platforms, Inc.のご担当者様からの資料に基づくご報告では、コミュニティ規定施行レポート等に関し、「いじめと嫌がらせ」(bullying and harassment)にかなり重点を置かれていたと存じます。Twitter, Inc.のご担当者様からの口頭でのご報告では、聞き逃したところがあるかと思ひ恐縮ですが、特にharmful/harassmentに言及されていたかと存じます。ただ、本WGにおいて冒頭に掲げているのは、「違法」、しかも日本法の下で「違法」とされる情報への対策です。もちろん、名誉毀損など、違法性の判断が難しい場合が多々あることは、皆様からの的確にご指摘いただきました。これまでも繰り返し指摘され、本WGの親となるプラットフォームサービスに関する研究会の方でも、検討が重ねられてきたところですので、ご質問として、
 - ①特に「違法」な内容に対する御社の対策に、より明確に焦点を当てた形で、今後、さらに具体的に、新たな事実状況についてご教示いただけることはございますでしょうか？ぜひ前向きにご検討くださいますと幸いです。
 - ②日本法の下での「違法」な情報への対策において、日々ご苦勞されながらもさまざまな工夫をされて、既に公表されている点以外に、新たに顕著な成果(outcome)があった点などは、ございますでしょうか？日本市場におけるユーザー等の関係の方々への、良いメッセージとなると思ひます。ぜひご教示いただけますと幸甚です。

Google LLC

山口構成員

- 質問先：Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：お忙しい中、貴重なご報告をいただきまして、誠に有難う御座います。本日のGoogle LLCのご担当者様からの資料に基づくご報告では、「有害」(harmful)、「オンラインハラズメント」に焦点を当てられていたと存じます。また、ご報告の最後の方で、「有害」な内容をいかに定義するか、という問題が一番重要とおっしゃっておられたかと存じます。Meta Platforms, Inc.のご担当者様からの資料に基づくご報告では、コミュニティ規定施行レポート等に関し、「いじめと嫌がらせ」(bullying and harassment)にかなり重点を置かれていたと存じます。Twitter, Inc.のご担当者様からの口頭でのご報告では、聞き逃したところがあるかと思ひ恐縮ですが、特にharmful/harassmentに言及されていたかと存じます。ただ、本WGにおいて冒頭に掲げているのは、「違法」、しかも日本法の下で「違法」とされる情報への対策です。もちろん、名誉毀損など、違法性の判断が難しい場合が多々あることは、皆様からの的確にご指摘いただきました。これまでも繰り返し指摘され、本WGの親となるプラットフォームサービスに関する研究会の方でも、検討が重ねられてきたところですので、ご質問として、
 - ①特に「違法」な内容に対する御社の対策に、より明確に焦点を当てた形で、今後、さらに具体的に、新たな事実状況についてご教示いただけることはございますでしょうか？ぜひ前向きにご検討くださいますと幸いです。
 - ②日本法の下での「違法」な情報への対策において、日々ご苦勞されながらもさまざまな工夫をされて、既に公表されている点以外に、新たに顕著な成果(outcome)があった点などは、ございますでしょうか？日本市場におけるユーザー等の関係の方々への、良いメッセージとなると思ひます。ぜひご教示いただけますと幸甚です。

Meta Platforms, Inc.

- 名誉毀損を有害なコンテンツと同列に考えるべきではありません。
- 名誉棄損とオンラインでの安全や有害なコンテンツに関する法とは目的が異なるため、両者を分けて考えるよう注意することが必要です。ヘイトスピーチ、いじめ、ハラズメント、性的搾取や虐待など、対象となるグループや個人に現実に害を及ぼす可能性があるものとは異なり、名誉毀損は（公共の利益ではなく）風評や個人の問題として位置づけられます。プラットフォームは、コンテンツが誹謗中傷であるかどうかを判断することができず、それらの判断は、裁判所や名誉毀損に特化して設けられた法的枠組みを通じて裁定されることを要求します。

山口構成員

- 質問先：Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：お忙しい中、貴重なご報告をいただきまして、誠に有難う御座います。本日のGoogle LLCのご担当者様からの資料に基づくご報告では、「有害」(harmful)、「オンラインハラスメント」に焦点を当てられていたと存じます。また、ご報告の最後の方で、「有害」な内容をいかに定義するか、という問題が一番重要とおっしゃっておられたかと存じます。Meta Platforms, Inc.のご担当者様からの資料に基づくご報告では、コミュニティ規定施行レポート等に関し、「いじめと嫌がらせ」(bullying and harassment)にかなり重点を置かれていたと存じます。Twitter, Inc.のご担当者様からの口頭でのご報告では、聞き逃したところがあるかと思い恐縮ですが、特にharmful/harassmentに言及されていたかと存じます。ただ、本WGにおいて冒頭に掲げているのは、「違法」、しかも日本法の下で「違法」とされる情報への対策です。もちろん、名誉毀損など、違法性の判断が難しい場合が多々あることは、皆様からの的確にご指摘いただきました。これまでも繰り返し指摘され、本WGの親となるプラットフォームサービスに関する研究会の方でも、検討が重ねられてきたところですので、ご質問として、
 - ①特に「違法」な内容に対する御社の対策に、より明確に焦点を当てた形で、今後、さらに具体的に、新たな事実状況についてご教示いただけることはございますでしょうか？ぜひ前向きにご検討くださいますと幸いです。
 - ②日本法の下での「違法」な情報への対策において、日々ご苦勞されながらもさまざまな工夫をされて、既に公表されている点以外に、新たに顕著な成果(outcome)があった点などは、ございますでしょうか？日本市場におけるユーザー等の関係の方々への、良いメッセージとなると思います。ぜひご教示いただけますと幸甚です。

Meta Platforms, Inc. (続き)

- Metaでは、私たちのサービス上のコンテンツについて、誹謗中傷と思われる懸念がある場合、こちらの報告フォームから報告していただくことができます。<https://www.facebook.com/help/3894625317215414> 私たちは、これらの報告を日本の法律に従って検討し、日本の法律に基づいて誹謗中傷と思われるコンテンツへのアクセスを制限する処置をとります。
- また、日本の法令に違反する可能性のあるコンテンツへのアクセスを制限するために、日本の機関と協力し、現地法に抵触するコンテンツのレポート(<https://transparency.fb.com/data/content-restrictions/>)で開示します。

山口構成員

- 質問先：Google LLC/Meta Platforms, Inc./Twitter, Inc.
- 質問内容：お忙しい中、貴重なご報告をいただきまして、誠に有難う御座います。本日のGoogle LLCのご担当者様からの資料に基づくご報告では、「有害」(harmful)、「オンラインハラスメント」に焦点を当てられていたと存じます。また、ご報告の最後の方で、「有害」な内容をいかに定義するか、という問題が一番重要とおっしゃっておられたかと存じます。Meta Platforms, Inc.のご担当者様からの資料に基づくご報告では、コミュニティ規定施行レポート等に関し、「いじめと嫌がらせ」(bullying and harassment)にかなり重点を置かれていたと存じます。Twitter, Inc.のご担当者様からの口頭でのご報告では、聞き逃したところがあるかと思い恐縮ですが、特にharmful/harassmentに言及されていたかと存じます。ただ、本WGにおいて冒頭に掲げているのは、「違法」、しかも日本法の下で「違法」とされる情報への対策です。もちろん、名誉毀損など、違法性の判断が難しい場合が多々あることは、皆様からの確にご指摘いただきました。これまでも繰り返し指摘され、本WGの親となるプラットフォームサービスに関する研究会の方でも、検討が重ねられてきたところですので、ご質問として、
 - ①特に「違法」な内容に対する御社の対策に、より明確に焦点を当てた形で、今後、さらに具体的に、新たな事実状況についてご教示いただけることはございますでしょうか？ぜひ前向きにご検討くださいますと幸いです。
 - ②日本法の下での「違法」な情報への対策において、日々ご苦勞されながらもさまざまな工夫をされて、既に公表されている点以外に、新たに顕著な成果(outcome)があった点などは、ございますでしょうか？日本市場におけるユーザー等の関係の方々への、良いメッセージとなると思います。ぜひご教示いただけますと幸甚です。

Twitter, Inc.

- ①利用規約に違反するコンテンツは削除するか、コンテンツへのアクセスを制限などします。しかし、真実性や相当性については、プロバイダ等の立場では判断できません。

Twitter, Inc. (続き)

■ ②Twitterは継続的にポリシーの改善に努めています。以下のように常に更新しています。

攻撃的な行為(<https://help.twitter.com/ja/rules-and-policies/abusive-behavior>) (最終更新2023年2月)

特定の人物や他の利用者を標的とした嫌がらせを禁じます。

Twitterでは、誰もが皆それぞれの意見や考えを、安心して発信できると感じられるはずです。Twitterは表現の自由を尊重し、開かれた話し合いを行うべきだと考えています。そしてTwitterでの健全な対話を促進し、ユーザーが安心してさまざまな意見や信念を表現できるよう、Twitterは、他者への嫌がらせ、侮辱、名誉毀損といった行為を禁じています。攻撃的な行為は、人々の安全性にリスクをもたらすだけでなく、影響を受けた人たちに、物理的および感情的な苦境をもたらす可能性があります。

個人情報および私的なメディアに関するポリシー(<https://help.twitter.com/ja/rules-and-policies/personal-information>) (最終更新2022年12月)

他の利用者の個人情報を、明確な同意や許可なく公開または投稿することは禁止されています。個人情報を公開すると脅迫する行為、または他者にこれを促す行為も禁止されています。

また、私人である個人の画像/動画などの私的なメディアを本人の同意を得ずに共有することは禁止されています。ただしTwitterは、報道価値のある出来事の一部として、または公益性のある問題や出来事に関する公共の議論を即移審するために、公人ではない個人の個人的な画像/動画を利用者が共有する場合があることを認識しています。そのような場合、Twitterは、そのメディアをTwitter上に残す可能性があります。他者の個人情報を本人の許可なくオンライン上で共有することは、晒し行為と呼ばれることがあり、プライバシーの侵害に当たり、Twitterルール (<https://help.twitter.com/ja/rules-and-policies/twitter-rules.html>) の違反とみなされます。