

# フェイクニュースや偽情報等に対する取組の フォローアップについて

---

2022年3月28日

事 務 局

## 1 自主的スキームの尊重

- ・民間による自主的な取組を基本とした対策を進めていくとともに、総務省はモニタリングと検証評価を継続的に行っていくことが必要

## 2 我が国における実態の把握

- ・PF事業者の認識や実態把握と調査結果とのギャップが生じていることから、プラットフォーム事業者は、自らのサービス上で生じている我が国における偽情報の問題について適切に実態把握を行い、研究者が分析を行うために必要な情報の無償で情報提供が行われることが望ましい

## 3 多様なステークホルダーによる協力関係の構築

- ・「Disinformation対策フォーラム」「Innovation Nippon」等において継続的に議論・研究が行われることが望ましい

## 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

- ・プラットフォーム事業者は、リスク分析・評価に基づき偽情報への対応を適切に行い、それらの取組に関する透明性・アカウントビリティ確保を進めていくことが求められる
- ・総務省は、これらの取組に関するモニタリングと検証評価を継続的に行っていくことが必要。どのような方法や情報により偽情報への適切な対応が図られているかどうかを評価することが可能かについて引き続き検討が必要

## 5 利用者情報を活用した情報配信への対応

- ・広告の種類・対応に応じてリスクや問題の差異を分析したうえで、特に、偽情報を内容とする広告の配信やターゲティング技術の適用については、そのリスクを踏まえ、より注意深い対応と、それに伴う透明性・アカウントビリティ確保が求められる

## 6 ファクトチェックの推進

- ・プラットフォーム事業者・ファクトチェッカー・ファクトチェック推進団体・既存メディア等が連携し、取組がさらに進められることが期待される
- ・我が国におけるファクトチェック結果を積み重ねて分析を行うことにより、偽情報の傾向分析やそれを踏まえた対策の検討が行われることが望ましい

## 7 情報発信者側における信頼性確保方策の検討

- ・現代のメディア環境に対応した情報の信頼性の確保の在り方について、既存メディア・ネットメディア・プラットフォーム事業者など関係者の間で検討を深めていくことが望ましい
- ・ミドルメディアを中心とした偽情報の生成・拡散・流通メカニズムに関する実態把握と分析も踏まえ、検討を深めていくことが望ましい

## 8 ICTリテラシー向上の推進

- ・偽情報の特徴を踏まえながら引き続きICTリテラシー向上施策が効果的となるよう取り組むことが必要

## 9 研究開発の推進

- ・ディープフェイク等に対抗するための研究開発や事業者の対応が進められることが望ましい

## 10 国際的な対話の深化

- ・偽情報に関する政策について国際的な対話の深化を深めていくことが望ましい



・次回以降のモニタリングにおいて、偽情報への対応に関して、事業者が自主的な報告を行わない場合や、我が国における透明性・アカウントビリティ確保が実質的に図られない場合には、透明性・アカウントビリティの確保方策に関する行動規範の策定及び遵守の求めや法的枠組みの導入等の行政からの一定の関与について、具体的に検討を行うことが必要

## プラットフォームサービス事業者における偽情報等への対策状況ヒアリング 質問項目

## 1. 前提(サービス概要)

- ① 対象サービス名・サービス分類
- ② 加入者数、月間アクティブユーザ数又は書き込み数

## 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 **グローバル**・日本における偽情報等の発生・拡散状況、結果の**分析**・公表
- ③ 研究者への情報提供、利用条件
- 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)
- ⑤ 不正な申告や削除要請への対策の方法・仕組み、対応件数

## 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 **Disinformation対策フォーラム**などの産学官民の協力関係の構築
- ② **多様なステークホルダーによる協力関係の構築**に向けての具体的な役割
- ③ 議論を踏まえた取組

## 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 (ii)削除等の対応
- 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**

## (2) 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
- ② 対応決定時における通知の内容、理由の記載の程度
- 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
- 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
- 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

## (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
- 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、(ii)アカウントの停止数
- 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)
- 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)
- 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
- 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

## (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
- 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
- ③ ターゲティング技術の適用に関する規定
- 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、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
- ⑥ 広告とコンテンツの分離
- 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

## 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
- 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 **日本におけるファクトチェック推進に向けた取組**
- 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
- ② 他のステークホルダーとの連携・協力・投資

## 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

## 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
- ④ 情報源のトレーサビリティ確保、なりすまし防止・認証

## 10. その他

- ① 意見・補足

※赤字は昨年度のヒアリング項目から追加、変更した質問項目。

(参考)プラットフォーム事業者等による対応のモニタリング結果  
偽情報への対応に関する透明性・アカウントビリティ確保状況に対する評価の概要

- プラットフォーム事業者の偽情報への対応及び透明性・アカウントビリティ確保の取組の進捗は限定的。
- 多様なステークホルダーによる協力関係の構築、ファクトチェック推進、ICTリテラシー向上に関しては、まだ十分とは言えないものの、我が国においても取組が進められつつある。

偽情報対策に関するプラットフォーム事業者における透明性・アカウントビリティ確保状況の主な評価項目の比較

我が国における実態の把握(2. 関係)	・すべての事業者について、我が国における偽情報の実態把握及び結果の分析・公開は未実施
多様なステークホルダーによる協力関係の構築(3. 関係)	・「Disinformation対策フォーラム」や「Innovation Nippon」により一定の進捗がみられる
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(4. 関係)	<ul style="list-style-type: none"> <li>・ヤフー(CGMサービス)は偽情報を直接禁止するポリシーが設けられておらず、対応件数等も非公開</li> <li>・LINEは、虚偽の情報の流布を包括的に禁止している。対応件数は一部公開</li> <li>・Googleはコロナ関連偽情報に関してポリシーがあり、我が国における削除件数等について一部公開</li> <li>・Facebook・Twitterは、削除ポリシーの整備や削除結果の公表等がグローバルでは進められているが、我が国における取組状況やその効果については非公開</li> </ul>
利用者情報を活用した情報配信への対応(5. 関係)	<ul style="list-style-type: none"> <li>・いずれの事業者も、偽情報に関する広告や政治広告について、広告配信先の制限や、広告内容に関する何らかの制限を規定するなど、一定の対応は行われている</li> <li>・LINE及びTwitterでは政治広告禁止。Googleでは日本において選挙広告は禁止</li> <li>・政治広告についてのターゲティング技術の適用については明確になっていない</li> </ul>
ファクトチェックの推進(6. 関係)	<ul style="list-style-type: none"> <li>・ヤフーやgoogleなど一部のプラットフォーム事業者において、プラットフォーム上の情報へのファクトチェック結果の紐付けや、資金提供等の取組が開始されている</li> <li>・Facebookでは、グローバルな取組は進められているが、我が国において国際ファクトチェックネットワーク(IFCN)に加盟しているファクトチェッカーがいないため、ファクトチェッカーとの連携は行われていない</li> <li>・LINE、Twitterでは、取組は未実施(LINEは台湾では取組を実施)</li> </ul>
情報発信者側における信頼性確保方策の検討(9. 関係)	<ul style="list-style-type: none"> <li>・「Disinformation対策フォーラム」において、既存メディアや有識者との情報共有や協議が進展</li> <li>・いずれの事業者も、新型コロナウイルス感染症関係の政府(内閣官房や厚労省等)の情報をサービス内で優先表示させる仕組み等を積極的に実施しているが、認知度は低い</li> <li>・Googleでは、ジャーナリズムの支援策を実施</li> </ul>
ICTリテラシー向上の推進(7. 関係)	<ul style="list-style-type: none"> <li>・Facebook・Google・LINEでは、我が国において偽情報の問題に対応したリテラシー教育に関する取組を実施</li> <li>・Twitterの取組は、偽情報対策に資する内容となっているか不明。ヤフーは今後実施予定</li> </ul>
研究開発の推進(8. 関係)	<ul style="list-style-type: none"> <li>・Facebook、Google、Twitterではグローバルな取組としてディープフェイク対策のための研究開発が行われている</li> <li>・ヤフー、LINEではディープフェイク対策の研究開発は行われていない</li> </ul>

→今回のモニタリングの質問項目について、事業者が「回答を控えた理由」及び「今後の対応方針」は次ページの通り

(参考)プラットフォーム事業者による対応のモニタリング結果(2021年)  
 (参考)「回答を控えた理由」及び「今後の対応方針」

事業者名	回答を控えた理由及び今後の対応方針
ヤフー	4. (2)削除等の対応 理由:「偽情報」という切り口から対応件数をカウントしているサービスはないため。 対応方針:対処すべき投稿がどのようなものかを含め、偽情報への対応につき有識者の意見も聞きながら検討を進める。 8. 研究開発の推進 理由:現在は行っていないため。 対応方針:今後の事情に応じて、研究の可否も検討していく。
Facebook	(誹謗中傷関連回答の再掲) ・コミュニティ規定施行レポートにおいて、国別の内訳を公表していない理由: ①悪意のある利用者が意図的にサービスを悪用する場合、VPNを使って発信元の国を隠すことが多いため、国レベルのデータは信頼性が低くなる ②1つのコンテンツが複数の場所で見られる可能性があるため、コンテンツの数や事前対応率を国別に報告することは困難 ③国レベルのデータではサンプルサイズが小さくなり、データの精度や信頼性が低下する可能性がある ・データの報告義務を規定すると、企業が直面する様々な技術的制限、利用可能なリソース、法的およびプライバシーに関する考慮事項(各国のデータ保護法を含む)を適切に考慮できない可能性がある。特定のデータセットを追跡し報告するための過度に規定的で義務的な要件は、企業側、そして規制当局の両方にとってコストがかかり負担となる可能性があり、そのプラットフォームの安全性を正確に反映していない可能性もある。その結果、イノベーションと成長が損なわれ、テクノロジーの圧倒的な活用をもたらす利益が制限されることになる ・利用者の皆様のためにサービスを向上させ、利用者の皆様の安全性を確保するための投資を継続的に行っていく
Google	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(4. 関係) ・Googleでは、コンテンツを「偽情報」として特別に分類をしておらず、Googleのプラットフォーム上での詐欺的または低品質のコンテンツを防止することを目的に、長年にわたり多数のコンテンツポリシーを設けており、例えば、スパム、なりすまし、意図的な情報操作に関するポリシーを定めている。このため、「偽情報」という分類でのデータを有していない ・偽情報への対処を目的とした政策立案・実施のため、どのような透明性の確保が有益なのか等について、プラットフォームサービス研究会、有識者等の意見を考慮しつつ、検討を行っていく
Twitter	・アルゴリズムに関する透明性・アカウントビリティについて:「責任ある機械学習イニシアティブの開始」というブログ記事を公表 ( <a href="https://blog.twitter.com/ja_jp/topics/company/2021/introducing-responsible-machine-learning-initiative">https://blog.twitter.com/ja_jp/topics/company/2021/introducing-responsible-machine-learning-initiative</a> ) (以下、誹謗中傷関連回答の再掲) ・日本における一般ユーザからの誹謗中傷や偽情報に関する申告・削除要請の件数ならびに削除件数については、今後順次公開する予定 ・なお、データの集計頻度や分類方法、公開範囲については社内で検討中 ・一般ユーザからの申告や削除要請に対応する部署の規模については、非公開。サービスごとに直面する課題や受け取る申告・要請は異なっており、求められる対応も異なるため、サービス間の担当者の体制や人数を比較するのは文脈上適切でなく、また、偏った比較となる ・対応部署の拠点については安全性やセキュリティの観点から非公開

※LINEはヒアリングにおいて日本における取組に関する全ての項目に回答済み