

プラットフォームサービス事業者におけるフェイクニュースや偽情報への 対策状況ヒアリング

【前提】

御社における、一般のユーザが書き込み可能かつ他のユーザからその書き込みが見られるような SNS 等のプラットフォームサービスについて、以下の質問にお答えください。回答は別添のヒアリングシートに記載をお願いします。

回答を記載いただく場合に、取組等の対象となるフェイクニュースや偽情報（以下、「偽情報等」といいます。）については、「①社会的混乱関係（災害等）」、「②健康・医療関係（コロナ関係等）」、「③選挙・政治関係（米国大統領選等）」に分類することが可能な場合、それぞれに分けてお答えをお願いします（※一部の質問については、あらかじめ回答欄をそれぞれの分類に分けて回答いただくよう設計しております）。それらに共通の取組や、これらの枠に当てはまらない場合の取組については、「④全般・その他」としてお答えください。

回答に当たっては、原則として、日本における取組または数値についてお答えください。その上で、日本では行っていないが他国（米国・欧州等）で行っている取組がある場合や、グローバルでサービス全体における対応件数等の数値を把握している場合には、【他国の取組・数値であることを必ず明記の上】、併せてその点もお答えください。なお、日本における取組又は数値が示せない場合や、質問に回答できない場合、その理由をお答えください。

直接的に質問に答えられない場合には、どのような対象についての数値であるか明記の上、参考となる数値をお答えください。

数値を回答いただく場合、対象となる集計期間についてもお答えください。最新の集計期間の数値以外にも、過去の数値に関する推移も把握されている場合、それらについても可能な限りお答えください。

昨年2月の「最終報告書」の公表以降、新たに開始された取組や新たに把握された数値について記載いただく場合、「〇年〇月より実施・把握開始」等の記載を行うなどにより、その旨の強調をいただくようお願いいたします。

このヒアリングは任意でのお願いです。回答結果は、研究会で配付資料とした上で、総務省 HP にて公表させていただきます。

【質問事項】

< 1. 前提（サービス概要） >

- ① 以下の回答において対象とする御社のサービス名をお答えください。もし可能な場合には、サービスごとにシートを分けてご回答ください。それらのサービスについて、(i)コンテンツ共有サービス（一般ユーザ含む第三者

の投稿を共有するサービス)、(ii) ニュース配信サービス、(iii) 検索サービスのいずれに当たるかをお答えください。

- ② 当該サービスの日本における加入者数(アカウント数)、月間アクティブユーザー数や書き込み数など、サービスの規模が分かる情報をご回答ください。

< 2. 「我が国における実態の把握」関係 >

- ① 御社のサービスにおいて、偽情報等の発生・拡散状況を把握できる体制を整え、具体的に分析・調査を行っているか、お答えください。
- ② 御社のサービスにおいて、日本における偽情報等の発生・拡散状況についてお答えください。その結果を公表しているかお答えください。
- ③ 偽情報等の流通状況について、研究者等が調査分析を行う際に必要な情報の提供を行っているのかお答えください。行っている場合、無償か有償か、利用条件等についてもお答えください。

< 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 偽情報等への対応を進めていく目的で、産学官民の多様なステークホルダーとの協力関係を構築し、取組について議論・検討を行っているか、お答えください。
- ② 協力関係構築及び議論を行っている場合、その場において、御社がどのような役割を果たしているか、お答えください。
- ③ 協力関係構築及び議論を踏まえて、御社においてどのような取組を進めているか、お答えください。

< 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」関係 >

(1) 偽情報等に関するポリシー

- ① 偽情報等への対応に関して、ポリシーや約款において、(i) どのような種類・性質の情報や行為を禁止としており、(ii) どのような対応 (削除・閲覧制限・アカウント停止・表示順位低下等、いわゆるコンテンツモデレーション) を行うことと定めているのかお答えください。

(2) 削除等の対応

- ① 偽情報等に関する情報や行為について、申告や削除要請があった数をお答えください。
- ② 偽情報等に関する情報や行為について申告や削除要請があった場合に、(i) 削除等を行った数をお答えください。また、(ii) アカウントの停止 (偽情報を拡散する bot アカウント含む) を行った数をお答えください。

- ③ 上記②以外に、申告や削除要請に基づかずに、御社において主体的に削除等を行った数（例えば、AIを用いた自動検知機能の活用等）をお答えください。
- ④ 上記③について、削除を行う場合の方法・仕組み（例えば、AIを用いた自動検知機能の活用等）を導入している場合、具体的な仕組みについてお答えください。
- ⑤ 偽情報等に関する情報や行為について、削除以外に、(i)警告表示、(ii)表示順位の低下、(iii)その他、御社のサービス上の仕組みとして対策を行っている場合、当該機能の内容や取組内容（後述のファクトチェック結果の表示を除く）をお答えください。行っている場合にその数値をお答えください。
- ⑥ 不正な申告や削除要請（明らかに偽情報とは言えない書き込みに対する削除要請など濫用的な申告や削除要請）への対策を行っているかお答えください。行っている場合、対応件数や、具体的な方法・仕組みについてお答えください。

(3) 削除要請や苦情に関する受付態勢・プロセス

- ① 偽情報等に関する情報や行為について、一般ユーザ、行政機関、国際機関、ファクトチェック機関等からの申告や削除要請に関してどのような受付窓口・受付態勢を定めているのかをお答えください。また、一般ユーザから申告や削除要請があった場合、どのようなプロセスにより、どのように対応が行われるのか、具体的にお答えください。特に、申告や削除要請に対して人間がレビューしているのか、機械的にレビューしているのかについてお答えください。また、社内のチームにおいてレビューを行っているのか、あるいは第三者的なファクトチェック機関等がレビューを行っているのかについてお答えください。
- ② 偽情報等に関する情報や行為について、削除や警告画面表示等の可否が決定された時点で、申立者及び発信者に対して、どのような通知が行われるのか（特に、海外事業者の場合、日本語で通知されるのか）お答えください。特に、削除や警告画面表示等の対応がされた・されなかった理由をどの程度詳細に通知・説明しているのかをお答えください。
- ③ 偽情報等に関する情報や行為について、日本（日本語の投稿又は日本ユーザ）における一般ユーザ、行政機関、国際機関、ファクトチェック機関等からの申告や削除要請に対応する部署・チームの規模や人数をお答えください。その他、偽情報等に関する情報や行為について何らかの対応を行っている部署・チームがある場合には、その内容・規模・人数についてお答えください。

- ④ 上記④の部署・チームが日本国内の拠点に置かれているかお答えください。さらに、それらの部署・チームに関して、日本における専属の責任者が置かれているかお答えください。
- ⑤ 削除等についての対応可否に関して、申立者や当該情報の発信者から当該対応の決定に関する苦情や問い合わせがあった場合に、どのような苦情受付態勢及び苦情処理プロセスを定めているのかお答えください。

(4) 透明性・アカウントビリティの確保

- ① 偽情報等に関する情報や行為について、機械学習を含む AI によるアルゴリズムを活用してコンテンツモデレーションを行っている場合、それらの基準や考え方を明らかにするなど、当該アルゴリズムに関する透明性を確保し、アカウントビリティを果たす方策を行っているか、お答えください。特に、「人間中心の AI 社会原則」、「国際的な議論のための AI 開発ガイドライン案」や「AI 利活用ガイドライン」に沿って、透明性・アカウントビリティの確保がどのように図られているか、お答えください。
- ② 偽情報等に関する情報や行為への対応に関して、日本における削除等の取組に関する数値等を含む取組状況について、透明性レポート等により公開しているか、お答えください。また、透明性レポートは日本の取組に係る全ての情報が日本語で閲覧可能になっているか、お答えください。
- ③ 偽情報等に関する情報や行為を減らすための取組の効果について分析しているかお答えください。行っている場合、その効果をお答えください。

< 5. 「利用者情報を活用した情報配信への対応」関係 >

- ① 広告に関するポリシーや約款において、偽情報等に関する情報や行為に対する広告表示の制限（収益化の停止や配信停止など）等の規定を設けているか、お答えください。
- ② 広告に関するポリシーや約款において、偽情報等に関する広告や政治広告に関する出稿制限等を設けているかお答えください。
- ③ 広告に関するポリシーや約款において、偽情報等に関する広告や政治広告におけるターゲティング技術の適用に関する規定を設けているかお答えください。特に、マイクロターゲティングに関する規定を設けているか、お答えください。
- ④ 偽情報等に関する広告や政治広告に関するターゲティング技術のアルゴリズムについて、それらの基準や考え方を明らかにするなど、当該アルゴリズムに関する透明性を確保し、アカウントビリティを果たす方策を行っているか、お答えください。特に、「人間中心の AI 社会原則」、「国際的な議

論のための AI 開発ガイドライン案」や「AI 利活用ガイドライン」に沿って、透明性・アカウントビリティの確保がどのように図られているか、お答えください。

- ⑤ 広告に関するポリシーや約款において、偽情報等に関する広告や政治広告について、出稿者の情報や資金源の公開について規定を設けているか、お答えください。
- ⑥ 広告とコンテンツの分離に関するポリシーや約款、具体的な取組についてお答えください。
- ⑦ その他、偽情報等に関する広告や政治広告に関する透明性を高めるための仕組みや、ユーザに対するツールの提供等を行っているか、お答えください。行っている場合、具体的な取組についてお答えください。

< 6. 「ファクトチェックの推進」関係 >

- ① 御社のサービスにおいて、偽情報等に関する情報や行為に、関連するファクトチェック結果を表示させる等の取組を行っているか、お答えください。行われている場合、具体的な仕組みや、掲載に当たっての基準についてお答えください。
- 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等を行っているか、お答えください。行われている場合、具体的な取組についてお答えください。
- ③ ファクトチェックを実施する人材の育成を行っているか、お答えください。行われている場合、具体的な取組についてお答えください。
- ④ その他、ファクトチェック機関との連携が行われているか、お答えください。

< 7. 「ICT リテラシー向上の推進」関係 >

- ① ユーザが偽情報等に対抗する ICT リテラシーを身につけるための普及啓発について、御社においてどのような取組や投資を行っているのかお答えください。
- ② ユーザが偽情報等に対抗する ICT リテラシーを身につけるための普及啓発について、他のステークホルダーとどのような連携・協力・投資等を行っているのかお答えください。

< 8. 「研究開発の推進」関係 >

- ① 偽情報等に関する情報や行為に関するコンテンツモデレーションの対応に関して、機械学習を含む AI を活用した技術に関する研究開発を行っているか、お答えください。

- ② いわゆる「ディープフェイク」に対抗する技術に関する研究を推進しているか、お答えください。

< 9. 「情報発信者側における信頼性確保方策の検討」関係 >

- ① 公共機関や国際機関、メディア等による信頼性の高い情報をサービスの目立つ場所に表示させる仕組みを導入していますか。導入している場合、具体的にどのような取組かお答えください。
- ② ニュース配信プラットフォームサービスを提供している場合、ニュースや情報に関する選別・編集について、透明性やアカウンタビリティの確保方策を行っているか、お答えください。
- ③ 伝統的なメディアやネットメディア等の関係者間で連携体制を構築しているか、お答えください。また、信頼できる質の高い情報をどのようにサービス上に増やしていくかについて、どのような検討を行い、どのような取組を行っているか、お答えください。
- ④ 偽情報等への対策として、情報源のトレーサビリティを確保するための仕組みや、なりすましを防ぐ認証の仕組みを設けているか、お答えください。

< 10. その他 >

- ① その他、本件質問に関して意見や補足があればお答えください。

【参考】

プラットフォームサービスに関する研究会 最終報告書（2020年2月）抜粋

第6節 我が国におけるフェイクニュースや偽情報への対応の在り方

1. 自主的スキームの尊重
2. 我が国における実態の把握
3. 多様なステークホルダーによる協力関係の構築
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
5. 利用者情報を活用した情報配信への対応
6. ファクトチェックの推進
7. ICTリテラシー向上の推進
8. 研究開発の推進
9. 情報発信者側における信頼性確保方策の検討
10. 国際的な対話の深化

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

（2）プラットフォーム事業者による透明性・アカウントビリティの確保

利用者が安心・信頼してプラットフォームサービスを利用することができるよう、上記（1）で記載したプラットフォーム事業者による自律的な情報の削除等の対応に加えて、それらの取組が適切に行われていることが利用者や社会全体に対して明らかにされることが望ましい。また、利用者の表現の自由を確保する観点から、プラットフォーム事業者によって過剰な削除や不当なアカウント停止等の行き過ぎた対応が行われていないかという点についても明らかにされることが望ましい。

以上の観点から、プラットフォームサービスの提供に当たって、いわばプラットフォーム事業者によるガバナンスが適切に機能しているか否かについて、利用者や社会全体が把握することができるようにすることが重要であることから、プラットフォーム事業者は、自らの取組の透明性やアカウントビリティを確保する方策についても、上記の削除等の対応と同時に積極的に取り組むことが適当である。

プラットフォーム事業者による透明性やアカウントビリティの具体的な確保方策としては、例えば、

- ① どのような種類・性質の情報又はアカウントに対して、どのような対応を行うのか、自らが提供するサービスの全体的な考え方や具体的な対応に関するポリシーをあらかじめ明確に定めて公開すること
 - ② 自らが実際に行った取組の結果を公開すること
 - ③ 取組の効果について分析を行い公開すること
 - ④ 取組の効果や偽情報の流通状況について外部の研究者等が調査分析を行う際に必要な情報を提供すること
 - ⑤ 削除やアカウント停止等の対応に関して利用者からの苦情や問合せ等がある場合に備え、苦情受付態勢及び苦情処理プロセスを適切に定め、利用者に対してわかりやすく公開し、適切に運用を行うこと
- などの取組を実施することが望ましい。

なお、このうちプラットフォーム事業者自らが実際に行った取組の結果の公開に関しては、削除等の個別の対応全てに関して逐一公開することはプラットフォーム事業者の負担の面でも、また、利用者にとってのわかりやすさの面でも現実的ではないと考えられる。この点、各プラットフォーム事業者においては、一部のプラットフォーム事業者が実施している透明性レポートの作成の取組なども参考にしつつ、一定期間内に実施した取組を定期的に取りまとめて公表するなど、自らが実施した取組全体についてのアカウントビリティを果たすとともに、個別の対応について利用者から苦情や問合せがあった場合には、苦情処理プロセスの中で適切に個々の利用者に対してもアカウントビリティを果たすことが望ましい。

上記の①～⑤を始めとした透明性やアカウントビリティの確保方策について、グローバルにサービスを提供している国外のプラットフォーム事業者においては、米国や欧州の英語圏のみで実施しており、必ずしも我が国では実施されていない場合や、日本で実施されているとしても日本の利用者向けにわかりやすく日本語で説明されていない場合がある、という指摘がなされている。

したがって、国外のプラットフォーム事業者においては、第4節における記載のとおり、既に米国や欧州において多くの取組を実施しているところ、これらの取組について、欧米と我が国との間の偽情報の現状、社会状況、法制度等の違いに留意しつつ、可能な限り我が国でも実施されることが望ましい。

さらに、問題となる情報の分類及び具体的な対応に関するポリシーの策定、透明性レポート等の作成・公開、苦情受付態勢の整備などに関しては、我が国の利用者に対して、わかりやすく、我が国における個別事情に応じた対応が行われることが期待される。具体的には、

- ① 日本語で我が国の利用者にもわかりやすい形でポリシーや透明性レポートなどの情報を公開すること
- ② 透明性レポートを公開する際には、グローバルな対応件数の総数だけでなく、我が国の国内における対応件数についても併せて公開すること
- ③ 日本語を正しく理解できるスタッフを十分確保した上で、日本語で手続可能な適切な苦情受付態勢および苦情処理プロセスを整備するとともに、裁判手続を含めた国内での迅速な救済メカニズムを確保すること
- ④ 偽情報も含め、日本特有のプラットフォーム上の情報流通の問題にも適切に対応できるポリシーを策定すること

などの取組を実施することが望ましい。

なお、プラットフォーム上では大量の情報が流通することから、プラットフォーム事業者は、機械学習を含む AI によるアルゴリズムを活用して情報の流通をコントロールすることが一般的になっている。この点、機械学習を含む AI 技術により情報の削除等の対応を行う場合には、AI により誤った判定がなされ、本来は削除されるべきでない情報まで過剰に削除されてしまう可能性や、アルゴリズムにより不当に表示順位が低くなってしまふなどの可能性がある点が課題と考えられる。

したがって、プラットフォーム事業者は、機械学習を含む AI によるアルゴリズムの活用により情報の流通をコントロールしたり削除等の対応を行う場合には、どのような考え方に基づいて問題となる情報に対処するのかを明らかにする観点から、アルゴリズムに関する透明性を確保したり、アカウントビリティを果たすことが重要である。その際、「人間中心の AI 社会原則」、「国際的な議論のための AI 開発ガイドライン案」や「AI 利活用ガイドライン」などを参考にすることが期待される。

また、AI 技術により削除等の対応を行った結果として利用者から苦情や問合せがあった場合には、適切なアカウントビリティを確保する観点から、苦情処理プロセスの中で人による確認や対応を行うことが望ましい。

以上のとおり、プラットフォーム事業者が自主的に取組を実施し、それらの取組に関する透明性及びアカウントビリティの確保を図るとともに、プラットフォーム事業者自身による対応状況等の公開・説明やフォーラムの開催等を通じて、国民（利用者）やメディア等に対して取組の効果や課題などが明らかになることで社会全体としてのモニタリング機能が果たされ、それらの反応を踏まえてプラットフォーム事業者による更なる取組が進められていく、というサイクルが回っていくことが期待される。

また、政府は、こうしたプラットフォーム事業者による透明性やアカウントビリティの確保方策の状況について、フォーラム等の場を通じて随時適切に把握することが適当であり、仮にこれらの自主的スキームが達成されない場合、あるいは将来的に偽情報の拡散等の問題に対して効果がないと認められる場合には、透明性・アカウントビリティの確保方策に関して、プラットフォーム事業者に対する行動規範の策定や対応状況の報告・公表など、行政からの一定の関与も視野に入れて検討を行うことが適当である。

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

社名		
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	【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. 「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	
②	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	【①社会的混乱関係（災害等）】
		【②健康・医療関係（コロナ関係等）】
		【③選挙・政治関係（米国大統領選等）】
		【④全般・その他】
	(ii) 削除等の対応	【①社会的混乱関係（災害等）】
		【②健康・医療関係（コロナ関係等）】
		【③選挙・政治関係（米国大統領選等）】
		【④全般・その他】
(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】
		【(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】
		【(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】
		【(質問に答えられない場合) 参考となる数値】
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)	【日本の数値】
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	【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】
(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	
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	【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】

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	
(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	
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	
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	
10. その他		
①	意見・補足	