参考資料2

# 第23回会合における
# プラットフォーム事業者からの主な発言

2021年３月１７日

事　　務　　局

# プラットフォーム事業者からの主な発言①



## 1 ヤフー株式会社

- 前回の研究会から進んだ部分は、昨年６月の有識者会議（プラットフォームサービスの運営の在り方検討会）の設置と、有識者会議の結果として、同年12月に有識者からの提言書に従い今後の取組を示したこと。今後、これに従って対応進めていきたい。
- 研究会の中身は、個人に対する誹謗中傷等を内容とする投稿への対応を検討するとしており、特に、透明性の確保、あるいはAIを用いた対策の実効性をどうやって担保していくのかといったことについて、課題として取り上げた。
- 提言書は、特に誹謗中傷投稿の抑止について、AIを用いて直接的に投稿を削除していくことや、優良な投稿について奨励していく環境を整備することによって、間接的に誹謗中傷を減らしていこうという取組、透明性を高めていくための透明性レポートを策定していくことなどが含まれている。
- ヤフーとしては、提案書に沿ってポリシーや削除基準を定めていく、あるいは措置ユーザーに対しての窓口や、措置理由を開示していくような手当てをきちんとしていくことが重要だと考えている。
- 今後、提言を踏まえ、知恵袋におけるポリシーの見直しなど、提言を受けた対策をより強化していきたい。
- 事業者として今回の課題は非常に重要だと思うため、引き続き総務省も含めて連携していきたい。

※なお、前回の研究会で発表した取組み（利用規約における禁止行為の規定や、機械と人的パトロールによる投稿削除等）も継続的に行っている。

# プラットフォーム事業者からの主な発言②



## 2 Facebook Japan 株式会社

- フェイスブックではコミュニティ規定、インスタグラムではコミュニティガイドラインにおいて、してはいけないことを明確に定めております。その中には、いじめや嫌がらせも含まれており、この規定に違反するコンテンツ、書き込みは、我々のコンテンツモデレーションの中で削除という対応を取っている。
- 暴力的なコンテンツ、ヘイト、嫌がらせ、いじめなどの不適切なコンテンツについて、AIと人員の組合せで削除対応している。全世界で約３万5,000人が安心、安全を保つために取り組んでいる。日本語を含む50以上の言語に対応しており、365日、24時間体制で監視を行っている。
- 透明性レポートは年２回、公開しており、四半期に一度、コミュニティ規定施行レポートを公開をしている。なお、コミュニティ規定施行レポートは、最近は日本語でも公開している。
- 2020年第４四半期（10月から12月）のコミュニティ規定施行レポートでは、フェイスブックでは全世界で630万件、インスタグラムでは500万件のいじめと嫌がらせに関するコンテンツに対して措置を講じている。いずれも、前四半期と比べて対応は増加している。
- 一般ユーザーからの不適切なコンテンツの報告は重要な情報ソースの一つ。誰でも簡単に報告できるよう工夫を凝らしている。なお、第４四半期には、一般ユーザーからいじめや嫌がらせのコンテンツの報告がある前におよそ半分は対応している。インスタグラムでも同様。
- インスタグラムにおけるいじめ関連コンテンツに対応する仕組み・テクノロジーとして、①制限、②タグ付けとメンション、メッセージを許可する相手を選択、③コメントの管理、④ポジティブなコメントを固定などがある。これらは「保護者のためのInstagramガイド」の中で詳細を説明している。
- テクノロジーだけではなく、啓発・普及活動も行っている。昨年「＃インスタANZENカイギ」を開催し、インスタグラムにおけるクリエーターの方に協力してもらい、嫌がらせ、いじめは悪いことであるとユーザーに伝える活動を行った。また、昨年12月、フェイスブック社がグローバルに展開しているデジタルリテラシー教育のプログラム「We Think Digital」、日本語でいうと「みんなのデジタル教室」を立ち上げ、教育の現場で中高生を対象にデジタルリテラシーの向上、いじめや嫌がらせの防止を含む内容の出前授業を全国で行っている。

# プラットフォーム事業者からの主な発言③



## 3 Google Japan

- 違法かつ有害なコンテンツは弊社のプラットフォーム及びサービス上のコンテンツの非常に小さな部分しか占めていないが、ユーザーのためにプロダクト、製品の安全性を保つ責任を持つ取組には真摯に取り組んでいる。そして、グーグルの20年間にわたる経験及びコンテンツモデレーションによる継続的な改善の下①Remove、②Raise、③Reduce、④Rewardの４つの行動指針に基づき取り組んでいる。
- 普及啓発活動については、誹謗中傷の問題だけでなく、インターネットセーフティーの分野について、引き続き、オンライントレーニングや、学校の先生に対する教材の提供していきたい。
- プロダクトについてもしっかりと説明をしていくことが大事だと思っており、私たちが設けているルールをクリエーターやユーザーと共有しながら、よりよいコミュニティの在り方を引き続き考えていきたい。今、日本のユーチューブクリエイターと協力した企画を検討中で、誹謗中傷の問題について広く、若い世代も含めて議論していくことに取り組みたい。
- AI全般に対しての弊社の考えを共有したい。弊社においてAIと人によるレビューの組合せを利用している。弊社はAIの最大のサポーター的な立場を取っている。それは、今後、AIはさらなる多くの可能性があると私たちは見ているからである。それに加え、機械学習も日々向上しており、ネット上での悪質な行為が、それが投稿されてオンライン上で人々に害を与える前に検知できるところまで、向上しています。ただ、その技術、テクノロジーは、さらに改善していく必要があるとも考えている。なぜなら、AIは機械であり、例えばニュアンスが重要になってくる事柄に関して、例えばヘイトスピーチやハラスメントなど、果たしてその内容が違法なのか、そうではないのか、本当に境界線があるような内容のものに関しては、やはり人の判断が必要になってくる。重要なのは、AIはまだ文脈、脈絡を完全に正しく理解するのには苦労している段階なので、例えばあるコンテンツがオンライン上に投稿された際、その内容が学習的なものなのか、ドキュメントなのか、あるいは芸術なのか、それとも悪質な行為の内容なのかを見極めるには、やはりまだ人の目が必要と考えている。
- DSAとEUに関しては、ヨーロッパにおいては最近では規制のほうが多く、イノベーションは少ないという印象を受ける。つまり、日本はAIの分野においては一歩先んじているので、この研究会においても、さらにAIに関して研究、議論を重ね、今後、さらに倫理的にAIを活用していけるのか、そして、協働、協力した形でどのようにAIを活用していくかということを議論していけば、非常に役立つ事柄にたどり着くのではないかと考えている。決してEUで行われていることを見るだけではなく、私たちの間で討議をして、何かを見いだしていくことも重要ではないか。

# プラットフォーム事業者からの主な発言④



## 4 LINE 株式会社

- リテラシー向上のための啓発を一丁目一番地と位置づけ、特に注力してきた全国の学校等への講演活動については、2019年時点で講演活動が累計で１万回を超えている。また、「LINEの安心安全ガイド」をまとめ、いわゆる禁止行為など投稿に関するルールや、利用上の注意点などを平易な形で、どなたにも分かりやすく提供し、引き続きその注意喚起を継続している。
- LINEというサービスの特性を生かし、関係省庁や地方公共団体の皆様と、いじめをはじめとする子供向けの悩み相談を手がけている。昨年８月には、専門家の皆様と連携して、誹謗中傷を受けるなどして心の傷を負われた方などへのLINE相談窓口を開設した。
- 削除に関するポリシーの明示については、誹謗中傷をはじめとするサービス上の禁止行為、また、禁止行為が行われた場合の対応などについて、利用規約に明確に定めた上、周知徹底を行っている。
- 透明性の確保については、透明性報告書（ LINE Transparency Report ）を公表している。具体的な内容は、違反投稿への対応として、弊社におけるコンテンツモニタリングの仕組みや対応実績を公開している。また、ユーザーからの削除申告への対応として、名誉毀損やプライバシー侵害、これらに関する対応実績を公開している。今後とも、引き続き透明性を高めながら、事業者としての説明責任を最大限果たしていきたい。
- 業界団体への主体的参画については、ソーシャルメディア利用環境整備機構（SMAJ）という業界団体に参画をし、会員各社の皆様と連携しながら、その運営の一翼を担っている。昨年の緊急声明の発出や、総務省・法務省との連携による啓発キャンペーンなど、これからも個社の取組と併せ、業界団体を通じて貢献していきたい。
- モニタリングの充実を図り、ガイドライン違反の可能性があるコンテンツへの対応を徹底するとともに、お問合せ窓口の拡充を図った上で、専用フォームを通じた削除依頼等に対しても、引き続きモニタリングと連携しながら対応に努めている。
- AIの活用については、既にAIを活用した違反画像の検知は行っている。誹謗中傷などの違反テキストのAIによる検知の実装を計画しており、具体的な機能としてガイドラインに違反する可能性があるテキストを投稿しようとする場合、その投稿前に事前に検知し、ユーザーに対して警告を通知する機能の開発を進めている。この警告ポップアップにより再考を促し、勢いに任せた投稿が公開されることを未然に防ぐなど、抑止力となることを企図している。これは、2021年上半期に弊社サービスのタイムラインへの実装、そして下半期には他のサービスへの実装を目指している。

# プラットフォーム事業者からの主な発言⑤



## 5 Twitter Japan

- ①ルールの適用、②目にする情報のコントロール、③安全性向上のためのパートナーの３本柱で進めている。①については、ツイッタールールやポリシーを定めて、それを皆様に持っていただく、そして違反があれば厳正な対処をする。②については、ユーザーの皆さん自身が目にする情報、それからユーザーの方が目にされることをコントロールしていただくためのツールを提供している。③については、１社だけではできることにも限界があるため、多くのパートナー、あるいは業界団体と連携をした上で、こういった取組により努力している。
- 幅広い分野をカバーしたツイッタールール、そしてポリシーを用意しており、暴力や攻撃的な行為、合意のない裸体の描写、センシティブなコンテンツ、なりすましなどの幅広いルールを用意している。
- 違反の報告に関しては、様々なカテゴリーからできるようにしており、問題があるツイートから、あるいは専用のウェブフォームからも報告ができるようになっている。また、攻撃をされた本人だけではなく、第三者からも報告ができるようになっている。受け付けた報告については、24時間、365日体制で、グローバルで対応するようになっており、皆様をお待たせすることなく、迅速な対応をするよう取り組んでいる。
- 表示される情報をコントロールするために、ブロック機能、ミュート機能、セーフサーチ機能など様々な機能も提供している。ミュート機能については、アカウント単位、あるいは自身で設定した言葉、そういった様々な条件ごとに表示を停止する機能も用意している。ぜひこうした機能を使って、自身のより快適な利用環境をカスタマイズしていただきたい。
- 自分のツイートに返信ができる相手を選択する機能もあり、デフォルトでは誰もが返信できるようになっているが、自分が指定した相手だけ、あるいは自分のフォロワーだけ、そういった返信できる範囲を選択することで、会話の健全性をより高めることができる。
- 開かれたプラットフォームという性質上、外部のパートナーとの連携を非常に重視している。NPOや、有識者、一般のユーザーから様々な発言をしていただくことで、会話の広がり、そして会話の健全性が高められると考えている。例えば、NPOには、有料の広告枠を無償提供することで、より積極的な情報発信を手伝っている。また、ハッシュタグや絵文字を利用した啓発キャンペーンなども積極的に実施している。また、トラスト・アンド・セーフティーカウンシルを用意をし、セキュリティーの有識者や、NPOなどの第三者諮問機関から意見をいただき、ツイッタールール、ツイッターポリシーを定期的に見直して、常に最新、かつ情勢に合ったものにしていくような取組も進めている。
- Twitter透明性センターで、年２回、透明性に関する取組の公表を行っている。英語が先に公表され、その後、翻訳作業を行った後、日本語についても公表している。今現在、日本語については英語よりも１期遅い内容が公表されているが、多くの方にこうした内容もぜひ日本語で御覧いただくように、我々としても迅速な取組を進めている。

# プラットフォーム事業者への質疑応答①



**削除等の基準について各社のお考えを聞かせて欲しい。考え方として以下３つがあるかと思うが、どれに該当するのか。【寺田構成員】**
**①あくまでも個々の企業の自主的な規制によるべき**
**②民間の団体などを通じて共通のガイドラインを策定し、それらをベースにすべき**
**③法律によって枠組みや規制の前提が作られるべき**

## ヤフー株式会社

- 当然、誹謗中傷に該当して違法になるものについては、現状の法律でも手当てがされているところで、それについて削除等の対応を行っていくというところは、各社共通していると思うが、違法ではないものについて、どこまで対応していくのか、共通認識をどこまで持っていくのかというところは、引き続き検討していくところになるかと思う。
- 現状、弊社として不適切だと判断したものについては、自主的に対応している。それを業界で共通認識を持っていくのか、あるいは政府と共同で取組をしていくのかについては、ケース・バイ・ケースになると思うが、問題の投稿、問題となる類型に合わせてやっていくのがよいのではないか。少なくとも、違法であるものについては間違いなく排除していくというところは共通していると思っている。

## Facebook Japan 株式会社

- 当社の立場として取っているのは①自主規制によるべきであるというところ。２つ理由があり、一つ目は、当社はグローバルに展開していることもあり、グローバルで統一した基準を定めて、それを運用していくということが望ましいと考えている。二つ目は、コミュニティ規定、コミュニティガイドラインは不断の見直しを行っており、現場、さらには外部の専門家の意見を取り入れて、絶えず更新・見直しをしているため、自主的な動きを尊重していただければ、より迅速に問題に対応することができないかと考えている。

## Google Japan

- 法的な義務、つまり、違法なコンテンツに関しては適切に対処しており、政府からの要請、あるいは裁判所命令を受けたら、それに対応している。そして、自社のルールやポリシーを設定していく上で、第三者の専門家のインプットを常に取り入れております。例えば、ヘイトスピーチであったり、ほかの悪質な行為になど。
- インターネットの世界、それを利用しているユーザーの皆さんは、進化のスピードが非常に速いため、我々としても、それに適切に、迅速に対応していく必要があると考える。つまり、変化に見合ったガイドラインであったり、ポリシーを常に更新しながら設定していくということによって、悪質な行為等に関する最新のトレンドに常に柔軟に対応していけると考えている。

# プラットフォーム事業者への質疑応答②



**（再掲）**
**削除等の基準について各社のお考えを聞かせて欲しい。考え方として以下３つがあるかと思うが、どれに該当するのか。【寺田構成員】**
**①あくまでも個々の企業の自主的な規制によるべき**
**②民間の団体などを通じて共通のガイドラインを策定し、それらをベースにすべき**
**③法律によって枠組みや規制の前提が作られるべき**

## LINE　株式会社

- 法的な対応に関しては、最低限、当然にして果たすべき責務だろうと考えている。
- その上で大切なのは、各事業者が個社ごとの経営理念、あるいは削除のポリシー、こういったものをしっかり自己の責任において見定めた上で、いかにユーザーに周知し、モニタリング等を含めた対応を徹底できるかどうかということだと思う。
- ただ、自社、個社の自由、ある種の経営理念が優先され過ぎてしまうと、結果、取り残されるリスク、実際に被害を負うリスクがあるのはユーザーのため、先ほどSMAJから説明があった、個社の過去からの知見、ノウハウ、ポリシーといったものを寄せ合って、どういった今後の対策がよりよいユーザー環境を提供できることに資するのかといったことを、法的な規制によらない、あるいは依存しないような環境を自らがつくり上げていくためにも、業界を通じてお互い健全に議論することが重要になると考えている。

## Twitter Japan

- 法律による規制は非常に重視をしており、各国における法律については遵守をしている。
- ただ、表現の自由というものを信じており、過度な規制によるインターネット上の自由や、柔軟性が失われることには懸念を持っているため、個々の企業による規制、ツイッターでいえばツイッタールール、ツイッターポリシー、こういったものを明確に打ち出すことで、ユーザーの皆様に安心、安全な環境を提供するよう取組を進めている。

# プラットフォーム事業者への質疑応答③



**ツイッターとフェイスブックに対しての質問。前回のヒアリングでも申し上げたが、一般ユーザーからの削除の要請、それへの対応件数、国内の数字、モデレーションの体制が無回答となっている。日本の数字を出せないのはどういう理由か【生貝構成員】**
**透明性レポートが、日本における動きが分かりかねる部分がある。公表できないとしても、日本における利用者の動向について把握されていることがあれば教えていただきたい【大谷構成員】**

## Facebook Japan 株式会社

- 今回のヒアリングシートで、御指摘事項は非開示とさせていただいている。理由は、トランスペアレンシーレポートにおいて、個別の、刑事ではない、民事のケースは報告をしておりませんので、それに準じた形での報告とさせていただいている。
- また、コンテンツモデレーションの部隊が日本でどの程度用意されていて、実務で運用しているかという件については、これも社の方針として、モデレーションに当たる人員のセーフティーを守るという観点から、個別の詳細に関しては公表を控えておりますので、それに準じた形とさせていただいている。
- 大谷構成員からの質問については、ヒアリングシート中で、AIを使った自動検知の活用という点に関しての御質問と受け止め、例えば、フェイスブックで申し上げますと、事前対応率48.8%となっている。これは、ユーザーの方々からの報告の前にAIが探知した数字が48.8 %ということ。いじめ、嫌がらせやヘイトスピーチ、偽アカウントの事前対応率はそれぞれ違っておりますけれども、今回の質問のフォーカスポイントは、いじめ、嫌がらせに当たるものだと認識し、その48.8 %という数字を紹介した。

## Twitter Japan

- モデレーション体制は、先ほどのフェイスブックと同じ理由。セキュリティーの観点からも、ロケーション（どこの国にあるのか）については非開示となっています。また、人数だけではなくて、弊社の場合、人間とテクノロジーを組み合わせた対応を取っており、スタッフの教育にも非常にリソースを割いている。人数だけを出してしまうと、その人数だけが独り歩きしてしまい、質の部分やテクノロジーの部分が考慮されないといった懸念もあるため、非開示としている。
- 透明性レポートの国内の数字についても、同じく非開示としているが、以前から研究会でも、ほかの政府機関からも、日本における実情、モデレーション体制についての情報を開示するよう、かなり長い間、複数回にわたってリクエストをいただいているので、こちらについては引き続き社内で、どういったものをお出しできるのか検討させていただければと思う。

# プラットフォーム事業者への質疑応答④



**グーグルに対しての質問。国内の数字を含めてかなり出していただいているが、重要な数字が構成員限りになっている。公にできない理由があれば教えていただきたい【生貝構成員】**

## Google Japan

- コンテクストの中で、文脈が分かった上で理解して、トランスペアレンシーレポートを見ていただかなければならないため、様々な注意であったり、限定を要する項目がある。その正しい理解をしていただくため、どういった形で、どのようにリリースをしていくのかという検討に時間をかけているところ。これから、どのように、どのような形で開示をしていくかということを検討していく。
- また、どういったデータ、情報が役に立つかを聞きたい。ユーザーの観点から見ても、何が起こっているかを知らせるために、どういった情報やデータが役立つかということを知りたいのはもちろんだが、政策立案側である総務省やこの研究会に対してどういったデータを提供すれば、どのように現状把握に役立つのか。そして、それに基づいて公共の政策を策定する、何らかのソリューションを考えていく上で、何が役立つのかということをお知らせいただければと思う。

# プラットフォーム事業者への質疑応答⑤



**フェイスブックに対しての質問。ツイッターから、日本の数字を出すことを検討しているというお話があったが、フェイスブックについてはどうか。次回まで待っても、やはり今と変わらず日本の数字は非公表になるのか。日本語の投稿がどうかという話で、日本語の分かる方、日本人の方で対応していただくしかないのではないか。また、ヒアリングシートに日本以外の他国にはチームが存在するとあるが、日本にチームが存在するのか。また、その人数は。【森構成員】**

## Facebook Japan 株式会社

- 国別の数字を出しづらい理由は２つある。１つ目は、利用者がVPNを使っていると、どこから投稿したのかが分からない、どこの地域から投稿したものなのか特定しにくいということ。２つ目は、１の地域で発言したものが、２の地域で閲覧され、３の地域で報告される場合、どこでどうカウントするのかという統計の難しさがあること。
- また、コンテンツモデレーションの透明性を図るために幾つかの指標を公開しており、その一つとして、プリバレンス、どの程度バッドコンテンツが表示されるか、こういった指標を重要視している。これをはじき出すためにはある程度まとまりが必要と理解していただきたく、それを国別に出すと母数が集まらず正確な統計が取れないといった制約がある。
- その上で、国別の数字が出せるかどうかについては、今日の議論を持ち帰って検討したい。
- 場所を明かすことによって、例えば事務所に対して抗議が行われたり、そういうことが世界では起こり得る。働く社員の安全が脅かされるということはあり得る。そうしたことを防ぐために、コンテンツモデレーションを行っている人、レビュワーが働いている所は公開しないというのが社の方針。
- 日本語でコンテンツモデレーションを行っている部隊はいる。人数については手元に数字がなく、社として公表していないので、この場での説明はできない。