



第23回会合における構成員等からの主なご意見 (追加質問回答)

2021年5月13日
事務局

山口構成員

- 質問先：Facebook、Google、Twitter
- ヒアリングシートのご回答内容について、貴重な情報提供に御礼を申し上げます。その上で、日本での制度設計の今後の方向性に関して、一点、ご質問をさせていただきます。昨年末に公表された、欧州委員会によるDigital Services Act規則案26条のいわゆる「リスク評価」では、システミックリスクとして表現の自由や選挙過程にもかかわる広範な事項に関するリスクの特定・分析・評価が求められるところ、これに相応するリスク評価がもし日本で〈法的義務〉として求められる場合には、対象となる事業者の方々は、日本で具体的にどのような情報提供を予定され、かつ、そこでの最大の問題は何かについて、率直なご意見をお伺いできればと存じ、ご教示いただけますようお願いいたします。

Facebook Japan 株式会社

- デジタルサービス法についてはその草案が公開されたばかりであり、欧州委員会は今後さらに詳細について議論することになっていると理解しています。当社としてはコメントすることを控え、現段階ではさらなる議論を見守る必要があると考えています。また当社としては、今後数ヶ月の間にEUの機関と関わり、同法案の策定過程に貢献していきたいと考えています。
- 当社は、政府、企業、市民社会が責任を共有し、安全なオンラインコミュニティを促進するための明確な方法を示すことで、適切に設計された規制の枠組みがインターネットの継続的な成功に貢献できると信じています。私たちは、これらの重要な問題について日本政府と意見を交換する機会を得られることを歓迎し、協力していきたいと考えています。

Twitter Japan

- 当局と連携のうえ、法的義務を果たすための対応を行います。

山口構成員（続き）

- 質問先：Facebook、Google、Twitter
- ヒアリングシートのご回答内容について、貴重な情報提供に御礼を申し上げます。その上で、日本での制度設計の今後の方向性に関して、一点、ご質問をさせていただきます。昨年末に公表された、欧州委員会によるDigital Services Act規則案26条のいわゆる「リスク評価」では、システミックリスクとして表現の自由や選挙過程にもかかわる広範な事項に関するリスクの特定・分析・評価が求められるところ、これに相応するリスク評価がもし日本で〈法的義務〉として求められる場合には、対象となる事業者の方々は、日本で具体的にどのような情報提供を予定され、かつ、そこでの最大の問題は何かについて、率直なご意見をお伺いできればと存じ、ご教示いただけますようお願いいたします。

Google Japan

- 今後、規制がさまざまな技術やサービスの変化に対応することは重要であり、すべての人にとってより責任感があり、革新的で有用なインターネットを構築するための欧州の取り組みをGoogleは支援しています。そのためGoogleは、欧州のデジタルサービス法（DSA）に対して、既存の法律を改善する方法を提案するとともに、新しい規則が不適切に設計された場合のリスクを警告するための文章 <https://blog.google/documents/89/Googles_submission_on_the_Digital_Services_Act_package_1.pdf> を提出しています。
- DSAは、基本的な権利の保護を強化することを目的としています。私たちは、表現の自由や事業活動の自由に影響を与えるような条項には引き続き懸念を抱いています。特に、第26条が基本的権利を危険にさらす形で、合法的なコンテンツの規制につながる可能性があることを懸念しています。
- 第27条により、規制当局は、第26条で特定されたリスクについて、合法的なコンテンツであったとしても（「市民の言論活動に悪影響を及ぼす」可能性のあるコンテンツなど）、超大規模オンラインプラットフォームに実際に義務付けられている行動規範を通じて、影響力を及ぼすことができる幅広い権限を有しています。

山口構成員（続き）

- 質問先：Facebook、Google、Twitter
- ヒアリングシートのご回答内容について、貴重な情報提供に御礼を申し上げます。その上で、日本での制度設計の今後の方向性に関して、一点、ご質問をさせていただきます。昨年末に公表された、欧州委員会によるDigital Services Act規則案26条のいわゆる「リスク評価」では、システミックリスクとして表現の自由や選挙過程にもかかわる広範な事項に関するリスクの特定・分析・評価が求められるところ、これに相応するリスク評価がもし日本で〈法的義務〉として求められる場合には、対象となる事業者の方々は、日本で具体的にどのような情報提供を予定され、かつ、そこでの最大の問題は何かについて、率直なご意見をお伺いできればと存じ、ご教示いただけますようお願いいたします。

Google Japan（続き）

- これは、基本的権利に深刻な影響を与え、合法的なコンテンツを保護するという欧州委員会の表明した意図に反するものであると考えています。欧州委員会は、DSAの説明資料の中で、合法的ではあるが有害なコンテンツについては、「表現の自由の保護に深刻な影響を及ぼすデリケートな分野であるため、削除義務の対象とすべきではない」と指摘していますが、この規定はまさにその結果を招く危険性があります。また、イノベーションを阻害し、ビジネスを行うための基本的な権利を害する可能性もあります。
- DSAは、EUの民主的プロセスではなく、裏口から合法的な行為を規制することにつながるような規定を設けることは避けるべきです。
- こちらが26条に関する私たちの懸念ですが、DSAに関する私たちのより広い意見はこちら
<https://blog.google/documents/89/Googles_submission_on_the_Digital_Services_Act_package_1.pdf>
に記載されておりますのでご参照ください。私たちは、先日総務省の委員会のヒアリングに参加させていただいたように、20年以上にわたってさまざまな技術を生み出してきた経験を生かし、今後も日本における制度設計に関する議論に貢献していきたいと思っております。

生員構成員

- 質問先：Facebook、Twitter
- 昨年の緊急提言においては、「プラットフォーム事業者が上記の誹謗中傷への対応の必要性和表現の自由への萎縮効果のバランスを考慮した対応を実施するための方策として、コンテンツの削除等だけでなく、ユーザの選択に応じた形でのAIによる表示順位・頻度抑制等のコンテンツモデレーションや、規約やポリシーに基づくサービス設計技術（アーキテクチャ）の工夫による何らかの仕組みについて、事業者の創意工夫により、導入を検討することが期待される。（例えば、投稿内容について投稿時等に再考・再検討を行う機会を設ける機能や、ユーザの選択に応じた、コンテンツフィルタリング機能、一定の短期間の中に大量の誹謗中傷が集まった場合に自動的に検知を行い一時的に非表示にする機能など）」と記載している。再考再検討を促す機能やコンテンツフィルタリング機能についてはある程度実装されていると思うが、「一定の短期間の中に大量の誹謗中傷が集まった場合に自動的に検知を行い一時的に非表示にする機能など」に関連するサービス上の創意工夫を行っている点があるか、また、これらについてどうお考えか、ご説明いただきたい。

Facebook Japan 株式会社

- 当社は、利用者の皆様にとってより良い、より安全なプラットフォームとなるよう、様々なご意見に耳を傾け、エンジニアリング・技術的な側面を考慮しながら、サービスの向上に努めています。今後とも様々な意見を採用できるよう努力していきたいと思っております。

Twitter Japan

- 強制執行措置をとった攻撃的なコンテンツのうち38%（2019年）は、利用者からの報告ではなく、テクノロジーを駆使しTwitterの専門のチームによって積極的に特定されたものでした。

山本構成員

- 質問先：ヤフー、Facebook、Google、LINE、Twitter
- コンテンツモデレーションは、表現の自由などに関わる難しい判断を行う業務だが、コンテンツモデレーションに対応するスタッフへの研修・教育（トレーニング）をどのようにどれくらい実施しているのか。

ヤフー株式会社

- 約款の内容に即した内部のスタッフに向けた運用マニュアル（各禁止行為について具体的な削除例・保留例を記載）を作成し、担当者ごとに判断がぶれることのないよう努めている。
- スタッフは、配属時に1ヶ月以上の基礎研修（座学、モニタリング、OJT等）がある
- 定期的な勉強会（判断のすり合わせ）を実施している
- 疑義が生じた投稿についてはパトロールチーム内での協議や、サービス担当者、法的な判断を行う専門部署へのエスカレーションを実施している

Facebook Japan 株式会社

- コンテンツのモデレーションを行うスタッフには、各種のトレーニングやコーチングを行い、業務に役立てています。加えて、
- ポリシーの策定にあたっては、レビュアーの判断プロセスを可能な限り客観的にするために、違反するコンテンツと違反しないコンテンツを明確に区別し、非常に厳格に運用しています。これにより、主観や偏見を排除するようにしています。
- レビューアーがコミュニティ規定に基づいて客観的な判断を下せるように、「性的に挑発的な」コンテンツとは何かをレビューアーに判断させるのではなく、明確に定義しています。
- 毎週、各レビューアーの判断をサンプル調査し、その判断の正確性とコミュニティ規定を一貫して実施しているかどうかを確認しています。レビューアーがミスを行っている場合は、個々のレビューアーを再度教育します。また、審査員の監査も行っています。
- 質の高い監査を行っていても、常に正しい判断ができるわけではないため、プロフィール、ページ、またはグループが削除された場合には、その決定に対して利用者が異議を申し立てをする機会を設けてきました。

山本構成員（続き）

- 質問先：ヤフー、Facebook、Google、LINE、Twitter
- コンテンツモデレーションは、表現の自由などに関わる難しい判断を行う業務だが、コンテンツモデレーションに対応するスタッフへの研修・教育（トレーニング）をどのようにどれくらい実施しているのか。

Google Japan

- 例えば、YouTubeでは、ポリシー策定の段階から適切な判断が行えるように慎重に検討を行っています。ハラスメントやヘイトスピーチのような新しいポリシーの策定には、何ヶ月もかかることがあります。私たちのポリシーは、一貫性があり、偏りがなく、十分な情報に基づき、世界中のコンテンツに適用できるよう、一つひとつ慎重に検討されています。ポリシーは、外部の業界やそのポリシーに関連する領域の専門家、YouTubeクリエイターなど、さまざまな人たちの協力を得ながら策定しています。社内では、世界中から集まった専門チームが、ポリシー開発と一貫性を持った執行の実現に取り組んでいます。
- ポリシーは、4段階の承認プロセスからなる熟考されたプロセスを通じて審議され、最終的にはYouTube内の関連各部門の上級幹部が、それぞれ異なる視点から意見を出し合いながら決められます。
- 続いて、この新しいポリシーが、世界中にいる何千人もの異なる審査担当者によって一貫して実施されることを確認する必要があります。ポリシーが実際に運用される前のステージング環境にいる審査担当者たちが、非常に高い確率で一貫して同じ決定を下せることを確認する必要があります。それができなければ、トレーニングや社内ガイドラインを改訂して明確にし、同じプロセスを繰り返します。そして、許容できるレベルの精度に達して初めて、そのポリシーを一般に公開することができるのです。
- また、審査担当者は、すべてのYouTube コミュニティ・ガイドラインに関する定期的な必須トレーニングと認定を受けています。

山本構成員（続き）

- 質問先：ヤフー、Facebook、Google、LINE、Twitter
- コンテンツモデレーションは、表現の自由などに関わる難しい判断を行う業務だが、コンテンツモデレーションに対応するスタッフへの研修・教育（トレーニング）をどのようにどれくらい実施しているのか。

LINE 株式会社

- コンテンツモデレーションに関しては、主に以下のような人材育成の取り組みを実施しています。

[初期段階]

- ・モニタリング業務そのものの意義・使命といった部分の研修(座学)
- ・各サービスオペレーションの業務研修(座学)
- ・実践研修・テスト(e-learning)
- ・実対応研修(OJT)

[通常時]

- ・ガイドライン、サブ基準(事例集)におけるナレッジ集約
- ・QCおよびフィードバック(手動・自動)
- ・新事例共有会・対応検討会
- ・モニタリングに関わるニュースの共有会

[その他]

- ・定期的なe-learningを用いた理解度テスト

Twitter Japan

- 着任時の集中的なトレーニングに加え、スタッフには定期的にトレーニングを実施しています。

寺田構成員

- 質問先：SIA、SMAJ
 - 各団体にお伺いしたいと思います。
 団体に参加しておらず、また適切な対応を行っていただけない事業者への対策はどうあるべきとお考えでしょうか？ 例えば下記のようなことは考えられるでしょうか？
 - ①団体からの働きかけに対する各社の対応の一覧を定期的に公表する
 - ②一歩踏み込んでブラックリスト化して公表
 - ③法的なエンフォースメントの制定を
- ※ SMAJ様は会員以外への働きかけを行っていないように思いますが、今後、行うのか否か、行う場合には上記のようなことは考えられるのかについてお伺いできれば。

SIA

- SIAは会員社に適切な対応を促すだけでなく、「誹謗中傷ホットライン」を設置し、一般国民からの連絡を端緒に、会員であるかないかにかかわらず、外国も含めて全てのサイトを対象に、各社の利用規約に沿ったご対応を促すものです。現在のところ具体的に検討をしているわけではありませんが、一定の条件に該当する運営者について事実を公表するのは一案かもしれません。他方で、合理的理由があって削除措置を講じない運営者もある可能性を踏まえた対応が必要かと存じます。

SMAJ

- SMAJ会員以外への対応を含め、政府・関係団体と連携し、ソーシャルメディアの利用環境整備の推進を図るとともに、政府内で検討されている各種事項について協力して参ります。
- また、SMAJおよびSMAJ会員各社の取組内容を公表するなどの情報発信を通して、業界全体の取組を推進して参ります。
- ※参考：ソーシャルメディア上の名誉毀損や侮辱等を意図したコンテンツの投稿行為等に対する取り組みについて
 (https://smaj.or.jp/news/pressrelease_20210225_1.pdf)

大谷構成員

- 質問先： Facebook、Google、Twitter
- Twitter、Facebook及びGoogleに対して、（私も含めて）コンテンツモデレーション関係の質問が相次いだ研究会でしたが、もう少しだけ正確に教えていただきたいと思ひまして、この3社に質問させていただきます。

AIによる前捌きにあたり、映像は国によって違いは少ないと思いますが、文字情報については、様々な言語での対応が必要になると思ひます。日本語による表現を対象としたAIはいつからどの程度整備されているのでしょうか。

例えばFacebookでのAIの改善については以下のようなリリースがありますが、英語によるものであり、日本語の言語に対応する状況については、確認することができませんでした。

<https://about.fb.com/news/2021/02/update-on-our-progress-on-ai-and-hate-speech-detection/>

Facebook Japan 株式会社

- 私たちが開発するテクノロジーのほとんどは、世界中のあらゆる国で動作するという意味で、地理にとらわれないものです。毎日何億ものコンテンツが共有されるグローバルなプラットフォームにおいて、複数の市場で動作する技術を構築することが重要です。これにより、私たちのコミュニティ規定の施行の規模を拡大することができ、より迅速な対応が可能になります。
- 私たちは、日本語のコンテンツを含むコンテンツのモデレーションに高度なAI技術を活用することで、利用者にとってより良い、より安全なプラットフォームとなるようサービスの向上に努めてきました。特に力を入れているのは、言語や文化、地域を超えてコンテンツを文脈に沿ってレビューすることができるAIをさらに進化させることです。このように、AIの能力は着実に向上しており、その勢いは衰えることはありません。今後も、すべての利用者と同じサービスを提供できるよう努力していきたいと思ひます。

大谷構成員（続き）

- 質問先：Facebook、Google、Twitter
- Twitter、Facebook及びGoogleに対して、（私も含めて）コンテンツモデレーション関係の質問が相次いだ研究会でしたが、もう少しだけ正確に教えていただきたいと思いますので、この3社に質問させてください。

AIによる前捌きにあたり、映像は国によって違いは少ないと思いますが、文字情報については、様々な言語での対応が必要になると思います。日本語による表現を対象としたAIはいつからどの程度整備されているのでしょうか。

例えばFacebookでのAIの改善については以下のようなリリースがありますが、英語によるものであり、日本語の言語に対応する状況については、確認することができませんでした。

<https://about.fb.com/news/2021/02/update-on-our-progress-on-ai-and-hate-speech-detection/>

Google Japan

- Google は、さまざまな言語のユーザーに適した機械学習システムを構築するために、多くの研究開発を行っています。例えば、ニューラルネットを使った機械翻訳システムを搭載したGoogle 翻訳は、現在 100以上の言語で利用可能です。私たちは、日本語を含む英語以外の言語でのパフォーマンスを向上させるために常に努力しています。例えば、転移学習と呼ばれる手法を用いて、英語やスペイン語のように多くの人に使われている言語の学習内容を、アムハラ語やインドネシア語、日本語のように比較的使用人口の少ない言語の翻訳に適用しています。また、日本語の翻訳を改善するためのデータを得るために、コミュニティからの協力も得ています。
- Google は10年以上前からAI を使ってグローバルにコンテンツモデレーションを強化してきましたが、これらの様々な研究やGoogle 翻訳のような特定の製品から得られた結果は、各言語でのコンテンツモデレーションのためのシステムを改善するのにも役立っています。

大谷構成員（続き）

- 質問先：Facebook、Google、Twitter
- Twitter、Facebook及びGoogleに対して、（私も含めて）コンテンツモデレーション関係の質問が相次いだ研究会でしたが、もう少しだけ正確に教えていただきたいと思ひまして、この3社に質問させていただきます。

AIによる前捌きにあたり、映像は国によって違いは少ないと思いますが、文字情報については、様々な言語での対応が必要になると思ひます。日本語による表現を対象としたAIはいつからどの程度整備されているのでしょうか。

例えばFacebookでのAIの改善については以下のようなリリースがありますが、英語によるものであり、日本語の言語に対応する状況については、確認することができませんでした。

<https://about.fb.com/news/2021/02/update-on-our-progress-on-ai-and-hate-speech-detection/>

Twitter Japan

- 2018年に強制執行措置を取られた悪意のあるコンテンツのうち、積極的に特定されたものは0%でしたが、2019年4月の時点で、強制対応された攻撃的なコンテンツの38%は、利用者からの報告ではなく、テクノロジーを活用し、Twitterの専門のチームが積極的に特定したものです。これらには、攻撃的な行為、ヘイト行為、自殺や自傷行為の助長や扇動、脅迫や、具体的な暴力的脅迫などが含まれます。
- スпам行為や、プラットフォームの不正利用、その他のTwitterルールの違反を検出するものと同じテクノロジーが、攻撃的なツイートの特定に役立っています。常に先んじて迅速に対応し、利用者の皆さんがTwitterを安心してお使いいただけるよう、主要な分野や地域において対応チームを増員しています。利用者からの報告は会話の文脈を理解し、コンテンツを調査する上で重要な手助けとなりますが、Twitterを安全な場所にするにはそれだけでは十分ではありません。私たちはさらに積極的に対応する必要があり、現時点ではまだ初期段階ではありますが、この取り組みは将来に向けて健全性の向上を約束するものです。