

誹謗中傷等の違法・有害情報への対策に関するワーキンググループ 検討アジェンダ（案）

全体の検討を通じて留意すべき事項

「誹謗中傷等の違法・有害情報への対策に関するワーキンググループ」（以下「本WG」という。）においては、次の事項について十分留意しつつ、検討を進める。

情報の流通によって権利が侵害された被害者の救済の観点と発信者の表現の自由の観点、プラットフォームサービスを提供する事業者（以下「プラットフォーム事業者」という。）が措置を講じることによって生じる言論空間への影響や経済的負担の観点に、それぞれ十分留意する。

誹謗中傷等の違法・有害情報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ことに十分留意する。

情報の流通によって権利を侵害された被害者、こうした情報をプラットフォームサービス上に流通せしめた発信者、及びプラットフォーム事業者の三者間の関係性について、それぞれの主体の法的責任や、どのような場合に、どの主体がどの主体に対して何を請求することが可能かといったことや、どの主体間での問題解決が適切かといったことに十分留意する。

ユーザ数や投稿数が多い大規模なサービスを提供するプラットフォーム事業者の多くが、海外に拠点を置くプラットフォーム事業者であることを踏まえ、検討にあたっては、海外プラットフォーム事業者を含めた業界全体の健全化を目指すスキームを作ること、国内での法人登記義務の履行など、国内外の事業者が、法律上のみならず、実務面、執行面も含めて等しく扱われることといった内外無差別の原則について、準拠法や国際裁判管轄権に関する一般的理解も踏まえつつ、十分留意する。

プラットフォームサービス以外のインターネット上の情報流通に係るサービス（例えば、Web 検索サービス、電子掲示板、まとめサイト等のサービスや、SMS、メール等の一般非公開のサービス）の特性や、これらのサービスを提供する事業者の役割や責任にも十分留意する。

具体的な検討事項

1. プラットフォーム事業者の責務

プラットフォームサービスは、匿名で気軽に書き込みを行うことができる仕組みや、誰もが容易に情報を拡散し得る仕組みを有しており、情報の閲覧者が多くなるほどより多くの関心を集める機能を通じて、関心を得ることを目的とした刺激的な情報の発信を誘引する仕組みなどを備えており、こうした仕組みを背景に、ひとたび被害を生じさせる情報の送信や拡散が行われた場合には、被害が即時かつ際限なく拡大し、甚大になりやすく、プラットフォームサービスがインターネット上の違法・有害情報の流通の主要な舞台の一つとなっていると考えられるが、どうか。

こうした点に加え、ユーザが投稿するコンテンツを流通させることを通じて収益を上げるといふプラットフォーム事業者の収益構造、プラットフォーム事業者が自らのサービス上の違法・有害情報の流通に最も実効的に関与し得る点を踏まえれば、プラットフォーム事業者には、表現の自由を確保しつつ、違法・有害情報の流通に対応し、自らのプラットフォームサービス上の情報流通の適正化を図る責務があると考えられるが、どうか。

また、プラットフォーム事業者がこのような責務を果たすための環境整備として、以下の2.から4.までに記載された事項が検討項目として考えられるが、どうか。

2. 透明性・アカウントビリティの確保方策の在り方

違法・有害情報に対応するにあたり、表現の自由を確保しつつ情報流通の適正化を図るためには、プラットフォーム事業者による自主的な取組を促進することが重要と考えられるが、どうか。具体的には、特定の要件を満たすプラットフォーム事業者に対し、予めコンテンツモデレーション¹に関する運用方針を策定・公表するとともに (Plan)、運用結果を公表し (Do)、運用結果について自己評価を実施・公表し (Check)、必要に応じて運用方針を改定する (Act) ことを求めることにより、プラットフォーム事業者による自主的な改善サイクル (PDCA サイクル) を確立することについて、どう考えるか。

また、個別のコンテンツモデレーションに関する手続の適正性を確保するために、特定のプラットフォーム事業者に対して、コンテンツモデレーションに関する申請窓口等の透明化や、コンテンツモデレーションの実施又は不実施の判断に係る理由の説明等の一定の措置を求めることについて、どう考えるか。

さらに、違法・有害情報に対する効果的かつ継続的な取組を確保する観点から、違法・有

¹ 本検討アジェンダにおいて、「コンテンツモデレーション」は、例えば、問題となる投稿の削除・表示順位の低下、そのような投稿を行ったアカウントの凍結・停止、投稿に注意喚起を促すラベルの付与、当該サービスにおける検索結果からの排除・表示順位の低下、特定の地域での表示制限といった、違法・有害情報の流通を抑止するために講じる措置を指す。これには、投稿を行ったユーザ自身に気づかれない形で実施される投稿の非表示又は表示順位低下等の措置を含む。

害情報への対応に関する取組状況の共有等の継続的な実施について、どう考えるか。また、こうした情報について、一般に公表することと、2-6において述べるような取組状況の共有の場に限って報告することの違いについて、どう考えるか。

2-1 透明性・アカウントビリティの確保が求められる事業者

本来、なるべく多くのプラットフォーム事業者において、透明性・アカウントビリティの確保が図られることが望ましいと考えられる。一方で、透明性・アカウントビリティの確保には事業者にながら経済的負担が生じることや、ユーザ数や投稿数が多く、利用する時間が長いサービスにおいて、違法・有害情報の流通とそれに触れる機会が多いと考えられること等の諸点を踏まえて、まずは、違法・有害情報の流通の多い大規模なサービスから、サービスの特性等を踏まえ、透明性・アカウントビリティの確保を求めることについて、どう考えるか。

2-2 運用方針及び運用結果の公表

運用方針及び運用結果を公表することについて、次の項目を検討する。

(1) コンテンツモデレーションの運用方針（実施基準、実施の手続）

プラットフォーム事業者におけるコンテンツモデレーションの実施の基準となる運用方針として、措置の対象となる情報や行為、取り得る措置の内容やその適用にあたっての判断基準、措置の実施のために経ることを要する手続を公表することについて、どう考えるか。また、公表する運用方針は、日本語で、明確かつ分かりやすい表現を用いることについて、どう考えるか。

(2) コンテンツモデレーションの運用結果

上記(1)の運用方針に沿って、どのようにコンテンツモデレーションが運用されたかを公表することについて、どう考えるか。例えば、削除²等のコンテンツモデレーションの実施に係る端緒の件数、端緒ごとに措置が実施された又は実施されなかった件数、根拠となるポリシーや法的根拠ごとに行われた措置件数、苦情申立ての件数やそれに対する対応結果ごとの件数、措置に要した時間の中央値等を公表することについて、どう考えるか。また、これらの数値は、日本のユーザに関連する投稿等に関するものとするについて、どう考えるか。

² 本検討アジェンダにおいて、「削除」は、サーバ内のデータの抹消ではなく、情報の送信を防止する措置を念頭に置いている。

2-3 プラットフォーム事業者による評価、運用方針の改善

プラットフォーム事業者が、コンテンツモデレーションの運用について、自ら評価を行いその結果を公表することについて、どう考えるか。また、必要に応じて、プラットフォーム事業者が、自己評価結果に基づいて運用方針を改定することについて、どう考えるか。その際、自己評価には、ともすればいわゆるお手盛り評価となる可能性があり、評価の客観性や実効性を高める工夫を行うことが考えられるが、具体的にどのような工夫が考えられるか。プラットフォーム事業者によっては、コンテンツモデレーションの判断の際、助言・審査等一定の関与を行う機関を設けるなど、判断の客観性向上に資する工夫を行っているところ、こうした取組状況を公表することについてどう考えるか。

このほか、自己評価の前提条件の把握のために、プラットフォーム事業者が、自らのサービス上に流通している違法・有害情報とそれによって生じている被害の実態についても、把握・公表することについて、どう考えるか。

2-4 その他透明性・アカウントビリティの確保が求められる事項（運用体制等）

コンテンツモデレーションに従事する人員の数や監督者の有無など、実施に係る人的体制の整備や育成状況について、プラットフォーム事業者に公表を求めることについて、どう考えるか。また、これらの整備状況は、日本のユーザに関連する投稿等に関するものとするについて、どう考えるか。

また、プラットフォーム事業によっては、AI 等による自動処理といった、コンテンツモデレーションの迅速化に資する工夫を行っているところ、こうした取組状況を公表することについてどう考えるか。

さらに、プラットフォーム事業者は、例えば投稿時に再考・再検討を促す機能といった、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いるところ、そうした取組状況について公表することについて、どう考えるか。

その他、プラットフォーム事業者に、透明性・アカウントビリティの確保が求められる事項はあるか。

2-5 手続の適正性確保のために透明性・アカウントビリティの確保が求められる事項

(1) コンテンツモデレーションの措置申請窓口

プラットフォーム事業者による適切な対応につなげるため、被害者等が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（以下「コンテンツモデレーションの申請等」という。）を行うにあたって、申請や請求の理由を容易かつ十分に説明できるよう手続を整備することについて、どう考えるか。

例えば、コンテンツモデレーションの申請等の窓口の所在を分かりやすく公表することについて、どう考えるか。また、コンテンツモデレーションの申請等を受けた場合に、プラットフォーム事業者が当該申請等を行った者に申請等の受付に関する通知を行うことについて、どう考えるか。プラットフォーム事業者が当該申請等の受付に関する通知を行ったときには、申請等に係る情報の流通についてプラットフォーム事業者に認識があったものとみなすことについて、どう考えるか。

また、プラットフォーム事業者が措置を実施するか否かの判断に必要な期間をあらかじめ明らかにすることについて、どう考えるか。

(2) 個別のコンテンツモデレーションの実施又は不実施に関する理由

プラットフォーム事業者がコンテンツモデレーションの申請等に対して措置を実施しなかった場合に、申請等を行った者に対して、措置を実施しなかった事実及びその理由を説明することについて、どう考えるか。

プラットフォーム事業者がコンテンツモデレーションを実施した場合に、発信者に対して、コンテンツモデレーションを実施した事実及びその理由を説明することについて、どう考えるか。とりわけ、アカウントの停止・凍結やアカウントの再作成の制限等については、コンテンツモデレーションの中でも特に将来の投稿を制限する点で影響が大きいと考えられるが、これらの措置の実施にあたって発信者に対する手続的保障について、どう考えるか。

(3) コンテンツモデレーションに関する苦情処理

コンテンツモデレーションの申請等が認められなかった被害者等や、コンテンツモデレーションが行われた投稿の発信者に対して、苦情申出の機会を設けるため、処理手続や窓口について開示するとともに、こうした苦情への誠実な対応をプラットフォーム事業者に求めることについて、どう考えるか。

その他、個別具体の措置申請や措置に関する手続の適正性を確保する観点から、透明性・アカウントビリティの確保が必要な事項として、何があるか。

2-6 取組状況の共有等の継続的な実施

プラットフォームサービスに関する研究会では、これまで、個別のプラットフォーム事業者や個別のサービスのみならず、日本のユーザに関連する違法・有害情報の全体の流通状況を俯瞰するとともに、プラットフォーム事業者をはじめとする各ステークホルダーにおける取組状況の共有を行ってきたところ、こうした取組について、引き続き、産官学民が協力して、継続的に実施することについて、どう考えるか。

3. プラットフォーム事業者が果たすべき積極的な役割

プラットフォームサービス上では、ひとたび被害を生じさせる情報の送信や拡散が行われた場合、被害が即時かつ際限なく拡大し、甚大になりやすく、現に、誹謗中傷等の被害が発生し続けている。こうした点を踏まえ、表現の自由を確保しつつ、プラットフォーム事業者のより積極的な役割を検討する。具体的には、問題となる投稿の検知、削除の要請・請求、削除の実施といった各フェーズに応じて、以下のとおり検討を行う。

3-1 投稿のモニタリングのフェーズ

(1) 権利侵害情報の流通の網羅的なモニタリング

プラットフォーム事業者に対し権利侵害情報の流通を網羅的にモニタリングすることを法的に義務づける場合、検閲に近い行為を強いることとなり、表現の自由や検閲の禁止の観点から問題が生じうると考えられ、また、事業者によっては、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ことについて、どう考えるか。

(2) 繰り返し多数の権利侵害情報を投稿するアカウントのモニタリング

インターネット上の権利侵害は、スポット的な投稿によってなされるケースも多い一方で、そのような投稿を繰り返し行う者によってなされているケースも多いとの指摘がある。権利侵害情報の流通を網羅的にモニタリングすることをプラットフォーム事業者に対し法的に義務づけることには前述した問題があるとしても、繰り返し多数の権利侵害情報を投稿するアカウントを対象を限定した上でこれを継続的にモニタリングすることは、権利侵害情報の流通を効果的に抑制する上で有効と考えられるか。その際、そうしたアカウントによる投稿については、プロバイダ責任制限法第3条が定める責任制限の要件に関して、プラットフォーム事業者の情報流通の認識があったとみなすことで、プラットフォーム事業者

自身による適切な投稿の削除の促進に資すると考えられるか。その一方で、限定されたアカウントを対象とする場合であっても、プラットフォーム事業者に対し個別の権利侵害情報の流通をモニタリングすることを法的に義務づけた場合、表現の自由に萎縮効果をもたらす可能性があることについて、対象となる発信者のプライバシーへの影響も踏まえつつ、どう考えるか。さらに、悪質な侵害者は次々にアカウントを作成することでモニタリングを逃れることが可能であり、また、モニタリングの対象とするアカウントの範囲を法律で明確に規定することも困難であることを踏まえて、どう考えるか。

その他、繰り返し多数の権利侵害情報を投稿するアカウントへの対応として、アカウントの停止・凍結等を行うことは、将来の権利侵害の抑止に有効と考えられるか。このようなケースでは、同一人が複数のアカウントを用いて権利侵害情報の投稿を行う場合も考えられることから、繰り返し多数の権利侵害情報を投稿するアカウントの保有者が新たなアカウントを作成することや別のアカウントを用いた投稿をすることを制限する措置をとることは有効と考えられるか。一方で、アカウントの停止・凍結やアカウントの再作成の制限等については、将来の投稿を制限する点においてその影響が非常に大きく、繰り返し多数の権利侵害情報を投稿するアカウントへの対応であっても、公法上義務付けることについては極めて慎重であるべきとも考えられるが、どうか。

3-2 要請・請求のフェーズ

(1) 削除請求権

人格権を侵害する投稿の削除をプラットフォーム事業者に求める権利は、判例法理によって認められているところ、かかる権利を明文化することは、一定の要件において被害者がプラットフォーム事業者に対して差止請求を行うことが可能であるという事実を明確化し、被害者による削除請求に基づく権利侵害情報の削除の促進に資すると考えられるがどうか。一方で、判例法理を明文化するだけでは、現状とあまり変わらず、必ずしも被害者による削除請求に基づく権利侵害情報の削除の促進に資さない可能性もあることについて、どう考えるか。

また、営業権などの排他性を有しない財産上の権利を侵害する投稿も見受けられるとの指摘があり、こうした権利を侵害する情報について、削除を求める権利を創設することは有効と考えられるか。その一方で、こうした投稿の削除を求める権利が一般に認められるかについては、実務上あるいは学説上も明らかではなく、こうした権利の創設には慎重な検討を要すると考えられることについて、どう考えるか。

さらに、個々の投稿に違法性はないものの全体として人格権を侵害している投稿群の事案（いわゆる「炎上事案」）があるところ、このような事案について、現行法では削除請求

できるか必ずしも明らかではないため、炎上事案においても削除請求を可能とする規定を定めることは有効と考えられるか。その一方で、削除できる投稿の範囲、個別には違法性がない投稿の削除の可否について、このような投稿を行った者の被害拡大への甚大な影響を考慮しつつも表現の自由との関係を検討する必要があると考えられ、検討すべき課題は多く慎重な検討を要すると考えられることについて、どう考えるか。

(2) プラットフォーム事業者による権利侵害性の有無の判断の支援

プラットフォーム事業者は、被害者から投稿の削除の請求を受けた際に、投稿内容の権利侵害性について一定の判断を行うことが求められるものの、プラットフォーム事業者は、特定の情報が真実かどうか、ひいては権利侵害性があるかどうかを判断することに困難を伴うことが多いことについて、どう考えるか。これに関して、削除の判断の迅速化を図るため、例えば、プラットフォーム事業者において、削除請求や苦情があった場合は、外形的な判断基準を満たしているときは、いたずらや嫌がらせと判断できるものでない限り、可及的速やかに削除の処置を行うこととし、発信者に削除の対象になったことを通知し、発信者から異議の申立てなどを受け付けた場合には再表示する環境を整備することについて、どう考えるか。

また、プラットフォーム事業者による権利侵害性の有無の判断を支援するための環境を整備することについて、どう考えるか。例えば、プラットフォーム事業者の判断を支援するための環境整備として、公平中立な立場からの削除要請等の法的位置づけや、要請を受けたプラットフォーム事業者に求められる対応を明確化することについて、どう考えるか。

そのほか、かかる公平中立な立場からの要請に対して、その実効性を高めるため、例えば、応答義務を課すことや、要請された投稿を削除した場合の免責を定めることについて、どう考えるか。一方で、要請の実効性を担保する仕組み、要請をする者の判断の確からしさや中立性を担保するための要件や仕組み、要請をする者の運営に関する事項、要請に法的効果を与える場合の根拠など、検討すべき課題は多く慎重な検討を要すると考えられることについて、どう考えるか。

さらに、削除の仮処分手続について、申立てから発令まで数ヶ月を要することが通例であり、その間に被害が拡大してしまうとの課題が考えられるところ、民事保全手続よりも簡易・迅速な、削除に特化した手続（裁判外紛争解決手続（ADR）等）を創設することについて、どう考えるか。一方で、ADRの制度設計によっては、プラットフォーム事業者が手続に参加しなければ実効性が伴わないとの課題が考えられるが、ADRの有効性について、どう考えるか。

(3) 行政庁からの削除要請を受けたプラットフォーム事業者の対応の明確化

現状、法務省の人権擁護機関や警察庁の委託事業であるインターネット・ホットラインセ

ンター等の行政庁からプラットフォーム事業者に対して、違法・有害情報の削除要請が行われており、一定の実効性が認められるところ、かかる削除要請を受けたプラットフォーム事業者が取ることが求められる対応を明確化することは、更なる実効性の向上を図る上で有効と考えられるか³。

一方、この要請に応じることをプラットフォーム事業者に義務付けることは、検閲の禁止の規定の趣旨や表現の自由への影響を踏まえると困難であると考えられるが、どうか。さらに、行政庁からの削除要請については、その要請に強制力は伴わないとしても、プラットフォーム事業者に対し何らかの対応を求めるのであれば、さらなる透明性の確保が求められると考えられるが、どうか。そのためには、どのような制度的対応が考えられるか。

3-3 削除等の判断・実施のフェーズ

(1) プラットフォーム事業者による削除等の義務付け

プラットフォーム事業者に対して、権利侵害など一定の条件を満たす投稿について削除等の措置を行うことを公法上義務付けることは、この義務を背景として、当該プラットフォーム事業者によって、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など投稿の過度な削除等が行われ表現の自由への著しい萎縮効果をもたらすおそれがあることから、極めて慎重であるべきと考えられるが、どうか。

(2) 裁判外の請求への誠実な対応

プラットフォーム事業者によっては、裁判外での投稿の法的な削除請求に応じないケースもあるところ、裁判外の削除請求や削除要請に関して権利侵害性の有無の真摯な検討などの誠実な対応を行うことをプラットフォーム事業者に求めることは有効と考えられるか。

4. その他

4-1 検討対象となる情報の範囲

受信者の属性や文脈によって評価が変化し、法律上の定義が一義的に定まらない有害情報については、プラットフォーム事業者や発信者、受信者との間の権利義務関係の変動をともなう措置や行政庁からの強制力を伴う削除要請等によって対応することには極めて慎重であるべきであり、まずは、プラットフォーム事業者によるポリシーの策定及びその運用と、

³ 例えば、警察庁によれば、一部のプラットフォーム事業者における、当該プラットフォーム事業者が自主的に用意した一定の要件を満たした団体等からの通報について優先的にポリシー違反や違法性の有無の審査を行うなどの仕組みにおいて、警察やインターネット・ホットラインセンターからの要請について対応してもらうという考え方がある。

それらの透明性・アカウントビリティの確保を通じた対応の適正性を担保する仕組みによって対応されることが必要と考えられるが、どうか。

4-2 行政の体制や手続

2. や 3. に記載された事項についてプラットフォーム事業者に対応を求めるに当たって、行政の体制や手続について留意すべき点があるか。

4-3 相談対応の充実

インターネット上の違法・有害情報による被害の救済に関して、インターネット上の違法・有害情報に関する相談対応の充実を図ることが重要であると考えられるが、どうか。相談対応の充実を図るにあたって、改善すべき現状の課題や留意すべき点はあるか。

4-4 発信者情報開示請求

DM（ダイレクトメッセージ）など一対一の通信における侮辱、差別表現、暴力的表現、脅迫により権利侵害が生じた際に、こうした情報の発信者が不明であることを踏まえ、これらの対象についても発信者情報開示請求を可能とすることについて、どう考えるか。一方で、現行の発信者情報開示制度は、情報が拡散され被害が際限なく拡大するおそれがあることに着目して不特定の者に受信されることを目的とする通信を対象としていることや、仮にそのようにした場合に、開示請求を受けた事業者において、開示の可否を判断するために、公になっていない一対一の通信の内容について確認することとなることについて、どう考えるか。また、DM などの機能については、プラットフォーム事業者の中には、受信する側のアカウントにおいて、受信する範囲を選択することができる機能（例：友達の友達以外の者からの DM は受け取らないように設定可能とする等）を設けるなどの工夫を講じていることについて、どう考えるか。

以上