

SNS等における権利侵害情報・災害に対する取組に関するヒアリング事項 X Corp. 回答(日本語)

事業者名: X Corp.(対象サービス: X)

1. 権利侵害情報等への対応について

Q1-1. 権利侵害情報等への対応に係る利用規約等の内容・実態について以下の点をご教示ください。

(1) 権利侵害情報等への対応に関して、利用規約等において、削除以外のモデレーション(表示順位の低下等)としてどのような対応を行うことと定めているのか。行為や情報の類型ごとに取り得る対応が異なる場合、その旨を分かりやすく整理しご回答ください。

【回答】

Xの目的は、公共の場における会話に寄与することです。オープンな会話を促進するという当社のミッションに沿い、当社は本プラットフォーム上で多様な視点を奨励しています。嫌がらせおよびその他同様の行為は、人々の自己表現を妨げ、最終的には世界的な公共の会話の価値を低下させます。このため当社は、すべての人が自由に、安心して公共の会話に参加できるよう設計された、明確なルールおよびポリシーを設けています。

権利侵害情報に対するXのアプローチは、「攻撃的な行為および嫌がらせ」(<https://help.x.com/ja/rules-and-policies/abusive-behavior>)ならびに「暴言や脅迫、差別的言動」(ヘイト行為)(<https://help.x.com/ja/rules-and-policies/hateful-conduct-policy>)に関する各ポリシーに定められており、段階的かつ危害の程度に基づいて適用されます。

「攻撃的な行為および嫌がらせ」に関するポリシーのもと、Xは、他者への嫌がらせや侮辱、または他者を貶める行為およびコンテンツを禁止しています。「暴言や脅迫、差別的言動」に関するポリシーのもと、Xは、個人または集団を、ヘイト行為への言及、扇動、中傷や差別的揶揄、人間性の抹消、ヘイト表現を伴う画像、またはヘイト表現を伴うプロフィールによって標的とするアカウントに対して措置を取ります。

これらのポリシーへの違反に対し、Xは、ポストの表示制限、ポストの削除、アカウントの凍結を含む、段階的な強制的対応の選択肢を設けています。

処分を判断する際、Xは、違反の重大さ、標的となった人がいるかどうか、当該個人の過去のルール違反歴などさまざまな要因を考慮します。

詳細は、当社ヘルプセンターの「日本に関するリソース」ページ(<https://help.x.com/ja/rules-and-policies/japan-resources>)で公開しています。

(2) 権利侵害情報にあたると思われるもののうち、利用規約等に基づく対応が困難と考えるものはどのようなものか。また、そうした情報の削除を行うのはどのような場合があり得るか。特に名誉毀

損への対応、侮蔑的表現への対応に関する考え方については必ずご回答ください。また、裁判による判決が無ければ削除しないものがあれば具体的にご回答ください。

【回答】

Xは、Xルールに基づく対応が困難な権利侵害情報の類型として特定しているものではありません。争点は、強制的対応の運用上の困難さではなく、求められる判断の性質にあります。

Xルールは、暴力的な脅迫や、特定の人物を標的とした攻撃的な行為または嫌がらせなど、明確に定義された行為を禁止しています。ポストがXルールに違反する場合、Xは、裁判所の命令がなくても強制的対応を取ることがあります。強制的対応は違反の程度に比例したものであり、違反の重大さ、標的となった人がいるかどうか、当該アカウントの過去の違反歴に応じて、ポストの表示制限、ポストの削除要請、またはアカウントの凍結が含まれることがあります。

侮辱的または侮蔑的な表現は、「攻撃的な行為および嫌がらせ」に関するポリシーに該当する場合には、同ポリシーのもとで対応されます。Xは、他者への嫌がらせや侮辱、または他者を貶める行為およびコンテンツを禁止しています。これには、他の人を侮辱したり、貶めたりするために、特定の個人を悪意を持って一方的に攻撃の標的にする行為や、他者を標的とした暴言または冒涇の使用が含まれます。侮辱的または不快な表現のすべてが、この基準を満たすわけではありません。当該行為が特定の人物に向けられたものかどうか、その重大さ、正当な公共の関心事に関わるものかどうかを含め、文脈が重要となります。

名誉毀損は、これとは異なります。ある表現が名誉毀損にあたるかどうかは、事実関係、同定可能性、反真実性および名誉の毀損に関する認定、ならびに裁判所が判断する権限を有する法的評価などによって決まります。Xは、争いのある言論の真偽を判断することはなく、また、さらなる証拠がないまま名誉毀損の主張のみをもってXルール違反として扱うこともありません。名誉毀損であると主張されたコンテンツがそれ自体としてXルールに違反しない場合、Xは、有効な裁判所の命令またはその他の有効な法的手続に基づいて対応します。特定の法域において違法であると主張されるもののXルールには違反しないコンテンツに対する当社のアプローチと同様に、通常の措置は、当該法域において、特定されたコンテンツを表示制限することです。

(3) 申立ての対象となった投稿について、申立て理由と投稿内容が一致していないが、当該投稿内容が他の利用規約上の禁止事項に該当する場合は、当該他の利用規約に基づく対応は行われるか。

(例：誹謗中傷を受けたとして申立てがあった場合であって、当該投稿は誹謗中傷には該当しないと考えられるが、個人を特定する情報が記載されているような場合はプライバシーに関する規約等に基づいて対処されるか)

【回答】

Xは、報告されたポストについて、報告で示された理由以外の違反も含め、Xルール全般に照らして審査を行うことがあります。違反が確認された場合、Xは、該当するルールに基づいて措置を取ることがあります。措置は、違反の性質および重大さ、標的となった人がいるかどうか、当該アカウントの過去の違反歴に応じて決まり、違反の程度に比例したものとなります。

例えば、嫌がらせとして報告されたポストであっても、当該コンテンツを投稿したアカウントがスパム行為を行っている場合には、「信頼性」に関するポリシーに照らして審査されることがあります。違反

が確認された場合には、報告されたカテゴリーが最も適切なものでなかったとしても、同ポリシーに従い、必要に応じて削除やアカウントの凍結を含む強制的対応が取られます。

(4) 権利侵害情報について、削除以外のモデレーションを含めた対応の実施の判断基準及び意思決定プロセス等、貴社のガバナンス体制について。また、公表されている場合は、その公表場所も併せてご回答ください。

【回答】

権利侵害情報に対するXのアプローチは、「攻撃的な行為および嫌がらせ」ならびに「暴言や脅迫、差別的言動」(ヘイト行為)に関する各ポリシーに定められており、段階的かつ危害の程度に基づいて適用されます。

これらのポリシーへの違反に対し、Xは、ポストの表示制限、ポストの削除、アカウントの凍結を含む、段階的な強制的対応の選択肢を設けています。

処分を判断する際、Xは、違反の重大さ、当該個人の過去のルール違反歴などさまざまな要因を考慮します。

これらのポリシーおよび強制的対応の選択肢は、Xルールおよびヘルプセンター(「攻撃的な行為および嫌がらせ」、「暴言や脅迫、差別的言動」(ヘイト行為)、「強制的対応の適用レベルと適用範囲」、「日本に関するリソース」等)において公開しています。

Xルールの日常的な適用は、上記のポリシーのもと、当社の強制的対応のためのシステムおよびチームによって行われています。利用者は、ヘルプセンターに記載のとおり、異議申し立てを行うことができます(場合によっては、ポストの表示制限に関するラベルに対する異議申し立てを含みます)。日本からの情報流通プラットフォーム対処法(情プラ法)に基づく申出は、専用のワークフローに従って処理されます。

(5) 利用者が知るべき利用規約等の情報について、利用者が容易に認識・理解できるように実施している工夫について。

【回答】

当社は、Xルールおよび関連ポリシーを、ヘルプセンターにおいて日本語で公開しています。日本向けの情報(情プラ法に基づく権利侵害の申出など)は、「日本に関するリソース」ページに集約しています。また、ヘルプセンターでは、アプリ内での報告、強制的対応および異議申し立て、ならびに利用者が自分に表示される内容や他の利用者との関わり方を管理する方法(例:「Xエクスペリエンスのカスタマイズ」)についても説明しています。

Q1-2. 権利侵害情報等の発信・拡散への事前の対応について、次の点をご教示下さい。

(1) 権利侵害情報等を発信・拡散する行為を事前に抑止するための貴社のサービス設計・機能上の工夫等。(ご発表いただいたもの以外にある場合)

【回答】

Xの目的は、公共の場における会話に寄与することです。オープンな会話を促進するという当社のミッションに沿い、当社は本プラットフォーム上で多様な視点を奨励しています。嫌がらせおよびその他同様の行為は、人々の自己表現を妨げ、最終的には世界的な公共の会話の価値を低下させます。このため当社は、すべての人が自由に、安心して公共の会話に参加できるよう設計された、明確なルールおよびポリシーを設けています。

Xは、人工知能モデルを活用することにより、「攻撃的な行為および嫌がらせ」および「暴言や脅迫、差別的言動」に関する各ポリシーのもとで、特定の違反ポストを能動的に検知し、強制的対応を適用する能力を強化しています。この仕組みは、ポストを適時に審査できるため、強制的対応が迅速に適用され、リーチを抑制することができます。

(2) 誹謗中傷を含むネガティブな情報がトレンドに掲載される等、プラットフォーム側の仕組みを一つの要因として拡散される場合、それを抑制するための取組や問題意識。

【回答】

Xルールに違反するポストについては、例えば、トレンド、検索結果、タイムラインから除外するなどして、ポストの表示制限を行うことがあります（「強制的対応の適用レベルと適用範囲」を参照）。これはXルールに沿った対象を絞った措置であり、当社は、適法な言論全般をトレンドやランキングから包括的に除外するような仕組みは設計していません。

(3) (2)も含む、権利侵害情報の発信・拡散に係るシステミックリスクへの対応に関する取組や問題意識。

【回答】

侮蔑的な表現が「攻撃的な行為および嫌がらせ」に関するポリシーに該当する場合には、同ポリシーのもとで対応されます。Xは、他者への嫌がらせや侮辱、または他者を貶める行為およびコンテンツを禁止しています。これには、他の人を侮辱したり、貶めたりするために、特定の個人を悪意を持って一方的に攻撃の標的にする行為や、他者を標的とした暴言または冒涇の使用が含まれます。「暴言や脅迫、差別的言動」に関するポリシーのもと、Xは、個人または集団に対して、ヘイト行為や、扇動、中傷、差別的揶揄、人間性の抹消を行ったり、ヘイト表現を伴う画像、またはヘイト表現を伴うプロフィールによって標的とするアカウントに対して措置を取ります。

ポストがXルールに違反する場合、Xは、裁判所の命令がなくても強制的対応を取ることがあります。強制的対応は違反の程度に比例したものであり、違反の重大さ、標的となった人がいるかどうか、当該アカウントの過去の違反歴に応じて、ポストの表示制限、ポストの削除要請、またはアカウントの凍結が含まれることがあります。

名誉毀損は民事上の請求であり、Xルールへの違反とは区別されます。ある表現が名誉毀損にあたるかどうかは、事実関係、同定可能性、反真実性および名誉の毀損に関する認定、ならびに裁判所が判断する権限を有する法的評価などによって決まります。Xは、争いのある言論の真偽を判断することはなく、また、さらなる証拠がないまま名誉毀損の主張のみをもってXルール違反として扱うこともありません。名誉毀損であると主張されたコンテンツがそれ自体としてXルールに違反しない場合、Xは、有効な裁判所の命令またはその他の有効な法的手続に基づいて対応します。

(4) 権利侵害情報の発信・拡散に係るシステミックリスク等の実態把握のための、国、研究者等へのデータアクセス提供による協力に関する現在の取組。

【回答】

権利侵害情報がどのように投稿・拡散されているかについての理解を支援する取組の一環として、当社は、公開の枠組みを通じて情報へのアクセスを提供しています。コミュニティノートについては、ノート、評価および協力者に関するデータを毎日公開しており、ノートのランキングアルゴリズムのコードも公開しています。[おすすめ]タイムラインの基盤となるレコメンデーションアルゴリズムのコードも、同様に公開リポジトリで公開しています。加えて、有料(従量課金制)のAPIにより、公開ポストへのプログラムによるアクセスが可能です。

(5) 上記(1)～(3)で記載いただいた取組を含めた、①その効果の評価・分析 ② ①を踏まえた改善の仕組み。また、これらの公表状況。

【回答】

当社は、当社のポリシーおよび強制的対応の措置が、一貫性があり、合理的で、比例的かつ実効的なものとなることを目指しています。そのために、当社は、利用者の安全と表現の自由のバランスを取ることに重点を置いたポリシー策定プロセスを構築しています。当社の運用部門とポリシー部門は、インシデント対応の一環として、連携して課題を特定し、ポリシーおよび強制的対応のガイドラインを更新しています。

表現の自由を尊重しつつコミュニティを保護することに取り組むプラットフォームとして、Xは、システムリスクを軽減するためのさまざまな管理策を導入しています。Xは、サービス、利用者の行動および脅威の状況の変化に応じて、これらの措置を継続的に更新・改善しています。

当社は、強制的対応に関するポリシーおよび措置について、有意義な透明性を提供することを目指しています。これには、当社の強制的対応に関する利用者への通知のほか、ヘルプセンターの記事および@Safetyアカウントを通じた、ポリシーがいつどのように更新されたか、潜在的な違反をどのように報告し審査を受けられるか、強制的対応がいつ行われるか、ならびに利用者による異議申し立ての手段についての周知が含まれます。

当社の運用ハンドブックおよび能動的なヒューリスティクスは、新たな中傷表現、有害な用語および言い回しを含むよう更新されました。これにより、変化し続ける状況や、保護対象のカテゴリーに属する人々を標的とする言葉の使われ方を正確に捕捉できるようにしています。

(6) 一般的監視(網羅的監視)とはどのようなものと考えているか。

【回答】

当社は、「一般的監視(網羅的監視)」とは、違法または不適切なコンテンツを探し出すために、サービス上の利用者コンテンツの全部または実質的に全部を、能動的かつ継続的に審査および／または監視する義務を意味するものと理解しています。これに対し、①報告または通知に基づいて、公開されたルールに違反するコンテンツに対応すること、および②有効な法的手続および裁判所の命令に従うことは、これとは区別されるものです。

当社は、「攻撃的な行為および嫌がらせ」に関するポリシーおよび「暴言や脅迫、差別的言動」に関するポリシーに違反する一定の侮辱的表現および中傷表現について、自動化された表示制限の仕組みを導入しています。

(7) 上記(1)～(3)で記載いただいた貴社における取組も含めて、権利侵害情報の発信・拡散を未然に防ぐ取組についてどのような場合が一般的監視(網羅的監視)にあたると考えられるか。また、その理由。

【回答】

名誉毀損は、この議論において「違法」な言論と一括りにされることがよくあります。ある表現が名誉毀損にあたるかどうかは、事実関係、同定可能性、反真実性および名誉の毀損に関する認定、ならびに裁判所が判断する権限を有する法的評価などによって決まります。Xは、争いのある言論の真偽を判断することではなく、また、さらなる証拠がないまま名誉毀損の主張のみをもってXルール違反として扱うこともありません。名誉毀損であると主張されたコンテンツがそれ自体としてXルールに違反しない場合、Xは、有効な裁判所の命令またはその他の有効な法的手続に基づいてのみ対応します。

2. 災害発生時の対応について

Q2-1. 災害時の対応について以下の点をご教示ください。

(1) 利用規約等において、特に災害時の情報流通に関して定めている事項。

【回答】

当社は、災害に特化した個別の利用規約は設けていません。災害時の情報流通についても、同様に利用規約／Xルールを適用し、強制的対応を行っています。虚偽の救助要請、詐欺的な寄付依頼およびこれらに類するコンテンツは、「信頼性」に関するポリシー(偽アカウント、スパム、詐欺などを対象とするもの)に基づいて対応し得ます。

(2) 信頼できる情報源からの情報の優先表示等、災害時に実施している、もしくは実施を検討している取組。(ご発表いただいたもの以外にある場合)

【回答】

当社は、災害を含む現実世界の重大な事象に対するインシデント対応プロトコルを整備しています。7月の熊本地震の際、虚偽の救助要請、詐欺的な寄付依頼、「科学的根拠のない正確な予測があった」と主張するポストがプラットフォーム上に現れました。2024年1月の能登半島地震からの教訓も踏まえ、利用規約／Xルールを適用し、早期の段階で必要な対応を取りました。報道ではそのようなポストが現れたことが指摘されましたが、記事が掲載される時点では、多くがすでにプラットフォームによって対応されていたことも報じられました。また、利用者がプラットフォーム上で見つけた情報について追加の背景情報を確認できる取り組みとして、コミュニティノートとGrok Analyze(分析)機能があり、活用されました。

(3) 災害に備えて平時より実施している取組。

(利用者向けの啓発活動、災害時の情報確認方法に関する案内等)

【回答】

平時より、Xルールおよびヘルプセンターを通じて、禁止される行為や報告方法を周知しています。利用者が自分に表示される内容や他の利用者との関わり方を管理するための機能については、「X

エクスペリエンスのカスタマイズ」などのページで説明しています。また、コミュニティノートにより、利用者はプラットフォーム上で見つけた情報について追加の背景情報を確認することができます。加えて、ポスト上のGrokアイコンからアクセスできる機能(Grok Analyze(分析)機能／「Explain this post」)により、利用者はポストの内容をより深く掘り下げることができ、災害時を含め、利用者が情報を確認する助けとなり得ます。

Q2-2. その他

事業者様によっては、地震等の災害関連への対応として、日本にローカライズした独自の取組をされていると伺いました。災害に限らず、このように日本にローカライズしたお取組みがございましたらご教示ください。また、ローカライズに関する貴社のお考えをご教示ください。

【回答】

ローカライズの具体例として、いわゆる「闇バイト」対策の一環として、2024年12月3日より、特定ワードを用いた検索の結果画面に注意喚起を表示しています。オーバードーズ(市販薬の過剰摂取など)に関連する検索についても、同様の仕組みを実装しています。ローカライズに関する考え方としては、特定の問題が提起された場合に、関係者と協力し、対話を通じて効果的なリスク管理と解決策を模索することが望ましいアプローチであると考えています。当社は、公開されているXルールおよび適用法令を超える、日本独自の言論ルールは設けていません。

3. その他

ADR手続や事業者の判断に対する不服申立て等の事業者と利用者間の紛争処理の手続について、貴社の所感をご教示ください。

【回答】

当社は、ヘルプセンターに記載のとおり、アプリ内での報告機能と、一定の強制的対応に対する異議申し立ての経路を提供しています。日本の情プラ法に基づく申出は、「日本に関するリソース」ページに記載の専用のプロセスに従って処理されます。

X Corp.と利用者との間の紛争解決手続は、X利用規約に定められています。

以上