



日本の首相に関するコメント

2024-027-TH-UA

概要

Threadsに投稿された、日本の首相および脱税疑惑に関するコメントに対する、あるユーザーの返信をめぐる今回の事例において、当該コンテンツを削除するという措置は、必要でもなければ、Metaの人権に関する責任に合致するものでもありませんでした。この事例は、Metaが比喩的な脅迫と実際の暴力の脅迫をどのように区別すべきかという問題を投げかけています。委員会は、比喩的な脅迫に対する過度な措置について繰り返し指摘してきました。これは、Metaの暴力と扇動に関するポリシーにおいて、実際の脅迫と比喩的な脅迫が明確に区別されていないことに関係しています。この事例における政治指導者への脅迫は、政治批判を目的とした文字通りでない表現であり、強い言葉を用いて、持ち上がった不正疑惑への注意を喚起するものでした。このような行為は日本のソーシャルメディアでは珍しいことではなく、実害を引き起こす可能性は低いものでした。この事例に関与していた2名のモデレーターは日本語を話すことができ、日本の社会政治的な背景も理解していましたが、誤って当該コンテンツを削除しました。そのため、Metaは表現と現地の背景を評価する方法について、審査担当者に追加のガイダンスを提供するとともに、内部ガイドラインとポリシーの基本理念の整合性を図る必要があります。

この事例について

2024年1月、Threadsに岸田文雄首相についての新しい記事を共有するコメントが投稿されました。その記事は、自民党が関与した政治資金疑惑に関する岸田首相の対応を説明したものでした。投稿には、岸田首相の脱税を批判するキャプションが含まれていました。あるユーザーがこの投稿に公開で返信し、「へえ」という語を添えて、国会で説明すべきだとコメントしました。そして首相を脱税者と呼び、「死ね」という言葉も使っていました。この言葉は複数のハッシュタグに含まれていたほか、ユーザーの返信にはメガネをかけた人を揶揄する表現も含まれていました。

このユーザーの返信に「いいね」は付かず、Metaのいじめと嫌がらせに関するルールに基づいて一度報告が行われました。3週間後、人間の審査担当者が、当該コンテンツは暴力と扇動に関するルールに違反するとの判断を下しました。このユーザーは不服を申し立てたものの、別の審査担当者にも違反と判断されたため、最終的に委員会に上



訴しました。委員会がこの事例を取り上げたのち、Metaは当初の決定が誤りであったと判断し、Threadsの返信内容を復元しました。

このThreads投稿と当該ユーザーによる返信があった頃、自民党議員らは政治資金不記載の容疑で刑事告発されましたが、岸田首相は不起訴処分となりました。安倍晋三元首相が暗殺された2022年以降、日本では政治的暴力の懸念がくすぶっています。

岸田首相は最近、2024年9月27日に実施される自民党の次期総裁選に出馬せず、首相を退任することを表明しました。

主な所見

委員会では、「死ね」という言葉は信憑性のある脅迫ではなく、「死につながる可能性のある暴力の脅迫」を禁止した、暴力と扇動に関するルールへの違反にはあたらないと考えます。専門家は、この言葉は嫌悪感や非難を示す文脈で比喩的な表現として広く使われているとしています。当該コンテンツが比喩的な表現である理由として、興味や皮肉を表す「へえ」という語が添えられていることもあげられます。

ただ、Metaの暴力と扇動に関するルールでは、高リスク者に対し「～に死を」という言葉を使って死を求める内容の投稿を禁止していますが、このルールは明快さに欠けます。Metaのポリシーの基本理念では、脅迫を評価する際には文脈が重要とされますが、委員会が前回の事例で指摘したように、大規模に配置される人間の審査担当者には、脅迫の意図や現実性を評価する権限がありません。そのため、投稿に「～に死を」などの脅迫的な発言と標的（すなわち「標的に対する暴力の行動喚起」）が含まれる場合、その投稿は削除されます。委員会は[2022年の勧告](#)を再提示しつつ、Metaに対して、「～に死を」という語を使った誇張的な脅迫表現は高リスク者に向けられた場合を除き、一般に認められる旨の説明を公開ポリシーの文言に追加すること、また誇張的な政治的言論を保護するため、国家元首に向けられた脅迫的な発言が許可される場合の基準を設定することを求めています。

このポリシーでは、「公人・著名人」と「高リスク者」の扱いの違いもわかりにくくなっています。現在、深刻度が中程度の暴力の脅迫は、公人・著名人に対するものであれば「信憑性が高い」場合にのみ削除されますが、その他の個人に対しては「信憑性の有無にかかわらず」削除されます。さらにまぎわらしいことに、このポリシーには高リスク者に「追加的保護」を提供するという文言も存在します。審査担当者への内部ガイドランス（非公開）には、高リスク者に対する「～に死を」を使った表現は削除すべきと明記されています。Metaは委員会からの質問に対し、このポリシーは公人・著名人



に対する中程度の深刻度の脅迫を含むユーザーの言論に幅広い保護を提供していると回答しました。誇張的な表現を用いた不満の表明には、多くの場合暴力的な意図はないからです。しかし高リスク者に死を求めるなど、深刻度の高い暴力の脅迫は、オフラインでの危害につながるリスクがより大きくなります。この事例では、Metaは岸田首相が両方のカテゴリに該当すると判断しました。委員会は、ポリシーにおける「公人・著名人」と「高リスク者」の定義がユーザーにとって十分に明確でなく、両方の境界線があいまいであることを大変懸念しています。

Metaは委員会による前回の勧告を受け、ポリシーにおいて暴力的言論と政治的表現のバランスをより適切に取るための作業を行いました。しかし、「高リスク者」の定義はいまだに明確にされていません。委員会では、コミュニティ規定の中で例を交えながら基本的な定義を説明することで、どのような職業、政治活動、公務に従事する人物が保護されるのかをユーザーが理解できるのではないかと考えています。委員会は、2022年の[イランにおける抗議スローガン](#)に関する事例で、そのようなリストを提供しています。

監督委員会の決定

委員会は、当該コンテンツを削除するとしてMetaの当初の決定を覆します。

委員会はMetaに対し、次のとおり勧告します。

- 暴力と扇動に関するポリシーを更新し、「高リスク者」の基本的な定義を掲載して、暗殺などの暴力に見舞われるリスクの高い政治指導者などが含まれることを明確にする。わかりやすいように例も添える。
- 大規模に配置される審査担当者向けの内部ガイドラインで、「高リスク者」に対し「～に死を」という言葉を使って死を求める投稿についての内容を更新し、深く考えない思いつきの暴力的脅迫によって不満や他人との意見の相違を示した投稿を認めることを明示する。現地の言語や背景についても考慮する必要がある。
- 暴力と扇動に関するポリシー、および公人・著名人について言及されているその他のコミュニティ規定に、公人・著名人に対するいじめや嫌がらせの定義へのハイパーリンクを設置する。

*事例の概要は、事例の内容をまとめたものであり、先例としての価値を有しません。

この事例に関する決定



1. この事例の説明と背景

2024年1月、あるThreadsユーザーが、ニュース記事のスクリーンショットを含む投稿に返信しました。その記事には、自民党の所属議員が関与している不記載の政治資金収入をめぐる岸田文雄首相の声明が含まれていました。その声明で岸田首相は、この金額は「ありのままであり、不正に操作されたものではない」旨を述べています。Threadsの元の投稿には、岸田首相の画像と脱税を批判するキャプションが含まれていました。このユーザーはその投稿に対する返信の中で、「へえ」という間投詞を添えて、国会で説明すべきだとコメントしています。さらに、「死ね」という言葉を含む複数のハッシュタグを使って、岸田首相を脱税者と呼び、メガネをかけている人を揶揄しました（#死ね脱税メガネ、#死ねゴミ汚物メガネ など）。コメントはすべて日本語で書かれています。

元の投稿と返信が行われたのは、岸田首相がこの収入過少申告疑惑について国会で声明を発表した時期でした。2021年10月から首相を務める岸田氏は最近、2024年9月27日に実施される自民党の次期総裁選に出馬しないことを[表明](#)しました。

このユーザーの返信には、「いいね」もコメントも付きませんでした。この返信は[いじめと嫌がらせ](#)に関するポリシーに基づき、公人・著名人に対して「死を求めた」として、一度報告されました。人間のモデレーターがこの内容を審査したのは、未処理の事例がたまっていたためおよそ3週間後になりました。モデレーターは[暴力と扇動](#)に関するポリシーに違反していると判断し、Threadsからこの返信を削除しました。このユーザーはMetaに申し立てを行いました。2人目の審査担当者も、当該コンテンツが暴力と扇動に関するポリシーに違反していると判断しました。このユーザーは最終的に委員会に上訴しました。委員会がこの事例を取り上げたのち、Metaは返信内容を削除するという当初の決定が誤りであったと判断し、Threadsに復元しました。

監督委員会は次の背景を検討し、決定を下しました。



Threadsの元の投稿に対する返信が行われた2024年1月当時、検察が自民党議員らを政治資金不記載の容疑で刑事告発しています。岸田首相は不起訴処分とされました。

委員会が委託した調査によると、Threadsでは首相の脱税疑惑に対する非難や批判的な感情が広がっており、さまざまな投稿に「死ね」という言葉が含まれていました。委員会の依頼を受けた専門家は、日本の人々はよくソーシャルメディアで政治を批判する内容を投稿すると述べています。以前はネット掲示板が、気兼ねなく社会的な不満を表現するための匿名のプラットフォームとして機能していました（一般の意見（PC-29594およびPC-29589）を参照）。

委員会の依頼を受けた専門家によると、日本ではここ数十年、政治的暴力はほとんど発生していません。そのため、2022年に遊説中の安倍晋三元首相が暗殺された事件は日本中に衝撃を与えました。2023年4月には、演説中の岸田首相を狙って男性がパイプ爆弾を使用し、首相に怪我はなかったものの、2名の聴衆が負傷。この事件を受けて政治的暴力に対する懸念が高まりました。

委員会の依頼を受けた言語の専門家によると、投稿に使われた言葉は攻撃的であり、強い非難や不満を示すために広く用いられているものです。「死ね」という言葉は文字どおりの脅迫として用いられる場合もありますが、一般には怒りを比喩的に表現するために用いられ、そこに真の脅威はありません（Ayako Hatano氏による一般の意見（PC-29588）を参照）。

2017年、表現の自由に関する国連特別報告者は、日本における表現の自由について懸念を表明しました。懸念の理由として、政府関係者によるメディアへの直接的かつ間接的な圧力の行使、歴史的出来事に関する議論の場の制限、国家安全保障を名目にした情報へのアクセス制限強化をあげています。

Article 19 は、2024年のGlobal Expression Reportにおいて、日本を161か国中30位と位置づけています。Freedom Houseは、2023年のインターネットの自由度評価で日本を「自由」に分類しましたが、政府によるオンラインメディアエコシ



システムへの介入、独立した規制機関の欠如、インターネット上の誹謗中傷を規制する[改正法](#)における明確な定義の欠如について、懸念を提起しました。ただ、同組織の世界の自由度[レポート](#)で、日本は政治および市民の自由において100ポイント中96ポイントを獲得しています。日本はまた、民主主義と法の支配の指標でも一貫して高い評価を受けています。世界正義プロジェクトの2023年の法の支配指標では、日本は142か国中14位に[ランク](#)されました。

2. ユーザーの主張

当該の返信を投稿したユーザーは、委員会への陳述において、単に脱税の容認と助長という疑惑の行為について自民党政権を批判したにすぎないと主張しました。そしてMetaが公人・著名人の批判を禁じて投稿を削除したことは、日本における言論の自由を妨害するものだとしています。

3. Metaのコンテンツに関するポリシーと主張

I. Metaのコンテンツに関するポリシー

委員会による分析は、意見の表明に対するMetaの取り組み（Metaはこれを「何よりも重要なもの」と表現）と、安全性を重視する価値観を受けて行われたものです。Metaは、[暴力と扇動](#)および[いじめと嫌がらせ](#)に関するポリシーに基づいて当該コンテンツを評価した結果、当初は暴力と扇動に関するポリシーに基づいてこれを削除していました。委員会がこの事例を対象に審査を行ったのち、Metaは当該コンテンツがどちらのポリシーにも違反していないとの判断を下しました。

暴力と扇動に関するコミュニティ規定

[暴力と扇動に関するコミュニティ規定](#)におけるポリシーの基本理念で、Metaは「弊社プラットフォームのコンテンツに関連して起こりうるオフラインでの暴力行為を防止する」ための取り組みを行っているとし、「人は不満や他人との意見の相違を示すため、深く考えずに思いつきで暴力的脅迫を行うことがよくある」



と説明しています。また、「文脈が重要であるため、Metaは暴力的脅迫による非難または認知度向上、脅迫の標的となっている人物の知名度や脆弱性など、さまざまな要素を考慮する」としています。

このポリシーでは、死につながる可能性のある暴力（または他の形態による深刻度の高い暴力）の脅迫について、すべての個人を対象とした普遍的な保護を定めています。脅迫には、「標的に対する暴力の意図、願望、行動喚起を表す発言やビジュアル」が含まれます。

2024年4月以前、このポリシーでは、高リスク者を含む特定の人や集団に対する「重傷につながる可能性のある暴力（深刻度が中程度の暴力）の脅迫、および過去の暴力の容認」が禁止されていました。2024年4月、Metaはこのポリシーを更新し、このような脅迫について、その信憑性にかかわらず、あらゆる個人を対象とした普遍的な保護を提供するようになりました。ただし、「公人・著名人に対する」脅迫については、「信憑性」が必要になります。現バージョンのポリシーでは、「高リスク者」については深刻度の低い脅迫に関連する箇所でのみ言及されており、そこでも「一般の成人、すべての未成年、高リスク者、および保護特性を持つ個人や集団に対する追加的保護」が引き続き認められています。

公開されているポリシーに「高リスク者」の定義はありません。しかし、審査担当者向けのMeta内部のガイドラインには高リスク者のリストがあり、国家元首、元国家元首、国家元首の立候補者および元立候補者、国政選挙および超国家組織の選挙での立候補者（落選した場合、選挙後最大30日間）、暗殺未遂に遭った人物、活動家、ジャーナリストなどが含まれています（[イランにおける抗議のスローガン](#)に関する決定を参照）。

いじめと嫌がらせに関するコミュニティ規定

Metaの[いじめと嫌がらせに関するコミュニティ規定](#)では、個人に向けられたさまざまな形の暴力（「脅迫」など）を禁止し、「公人・著名人と一般の個人を区別」して、「報道で取り上げられたり社会的に注目を集めたりする人について



は、批判的な意見も含めた議論を認め」ています。このポリシーは、公人・著名人に対する「重大な」攻撃や、「投稿やコメントに公人・著名人を直接タグ付けする」などの「意図的にさらす」行為を禁じています。

このポリシーでは、「公人・著名人」を、「州・国家レベルの政府職員およびその立候補者、ソーシャルメディアで100万人を超えるファンまたはフォロワーを有している人物、ニュース報道で大々的に取り上げられる人物」と定義しています。

II. Metaの主張

Metaは委員会に対し、この事例におけるハッシュタグの「死ね」という語は、各種ポリシーの違反にはならないと報告しました。Metaはこの語を、信憑性のある死の呼びかけではなく、比喩的な表現による政治的発言とみなしています。Metaは、信憑性のある死の脅迫を含む発言と、政治的意見を述べるための比喩的な表現は十分に区別できない場合が多く、それが当該コンテンツを当初削除した理由だと説明しました。

Metaは委員会に対し、岸田首相はMetaの暴力と扇動に関するポリシーおよびいじめと嫌がらせに関するポリシーにおいて公人・著名人とみなされる一方、投稿に返信したユーザーは公人・著名人とはみなされないと述べました。また、岸田首相は暴力と扇動に関するコミュニティ規定による「高リスク者」とみなされることも報告しました。

暴力と扇動に関するコミュニティ規定

Metaは[暴力と扇動](#)に関するポリシーにおいて、「死につながる可能性のある暴力（または他の形態による深刻度の高い暴力）の脅迫」を禁止しています。人間の審査担当者に対する非公開のガイドラインでは、Metaは高リスク者の死を求める呼びかけについて、「～に死を」という言葉が使われている場合は削除するとしています。Metaは委員会に対し、高リスク者という概念はこのポリシーのみに



限定されるものであり、暗殺などの暴力にさらされるリスクの高い政治指導者を含むと述べています。

Metaは、「～に死を」および「死ね」という言葉を個別の事例で区別するのは難しく、特に日本語で両者の意味が重なる場合は困難であるとしました。この事例では、コンテンツ内のハッシュタグに、「～に死を」ではなく「死ね」という言葉が使われています（#死ね脱税メガネ、#死ねゴミ汚物メガネ）。さらにMetaは、「死ね」と「～に死を」は同様の意味合い（死を求める）として扱うものの、エスカレーションされた当該コンテンツは文字どおりの脅迫ではなく、「ポリシーの精神」に違反していないため、削除しないと述べました。ポリシーを厳格に解釈することでポリシーの意図に反する結果が生じる場合は、「ポリシーの精神」による猶予が適用され、コンテンツは許可されます（[スリランカの医薬品](#)に関する決定を参照）。ハッシュタグや返信自体に含まれる他の文言では、国会でのヒアリングを通じた政治的な説明責任について触れられていることから、Metaは脅迫に文字どおりの意図はないと判断しました。政治指導者に国会で説明責任を果たすよう求める呼びかけであることから、死の脅迫に文字どおりの意図はなく、比喩的な表現であることがうかがえます。このような理由により、Metaは、当該コンテンツは暴力と扇動に関するポリシーに違反するものではないと判断しました。

いじめと嫌がらせに関するコミュニティ規定

Metaは委員会に対し、当該コンテンツは公人・著名人を「意図的にさらす」内容ではないため、[いじめと嫌がらせ](#)に関するポリシーに違反していないと報告しました。このユーザーは岸田首相によるコメントにタグ付けや返信をしておらず、岸田首相のページに関する内容も投稿していませんでした。そのためMetaは、当該コンテンツは岸田首相を意図的にさらす内容ではなく、脅迫が文字どおりの意味だとしても、いじめと嫌がらせに関するポリシーに違反するものではないと判断しました。



委員会はMetaに書面で19の質問をしました。主な内容は、Metaの日本における施策とリソースについて、大規模に配置される人間の審査担当者にどのようなトレーニングを提供し、現地の背景をどのように盛り込んでいるかについて、大規模に運用しているポリシーのエスカレーションのプロセスについて、エスカレーションされた場合のみ高リスク者に対する死の脅迫を禁止するポリシーの実行可能性について、Threadsの審査バックログと自動検出機能についてなどです。Metaは17の質問に完全に回答し、2つについては部分的に回答しました。部分的に回答したのは、審査バックログと、日本政府によるコンテンツの削除要請に関する質問です。

4. 一般の意見

監督委員会には、[提出の条件](#)を満たす一般の意見が20件寄せられました（アジア太平洋およびオセアニアから13件、米国およびカナダから3件、ヨーロッパから3件、中央および南アジアから1件）。これらの一般の意見は、公開されることに同意のうえでお寄せいただいています。[こちら](#)をクリックしてお読みください。

意見の内容としては、日本における社会政治的背景、政治家に対するオンラインでの暴力の脅迫および表現の自由に関する制限事項、日本の政治的議論における誇張された脅迫の行使または暴力の呼びかけ、「死ね」という言葉の言語的背景、ユーザーにフォローされていないThreadsのページでは政治的コンテンツを推奨しないとするMetaの選択、などがありました。

5. 監督委員会による分析

委員会は、Metaのコンテンツに関するポリシー、人権に関する責任、および価値観を分析し、当該コンテンツを削除すべきかどうか検討しました。さらに、この事例がMetaのコンテンツガバナンスに関する幅広いアプローチにどう影響するかも評価しました。



5.1 Metaのコンテンツに関するポリシーの遵守

1. コンテンツに関するルール

暴力と扇動に関するコミュニティ規定

委員会では、この事例のコンテンツは「死につながる可能性のある暴力（または他の形態による深刻度の高い暴力）の脅迫」を禁止する、[暴力と扇動](#)に関するポリシーの違反にはあたらないと考えます。「死ね」という言葉は文字どおりの意図で使われておらず、脅迫に信憑性はありません。

委員会の依頼を受けた言語の専門家は、この言葉は文字どおり誰かの命を脅かすために使われることもあるものの、嫌悪感や非難を示すために比喩的な意味合いで広く用いられていると説明しています。この専門家によると、当該コンテンツにおけるこの言葉の使われ方は、比喩的なカテゴリに当てはまるということです。Threadsやその他のプラットフォームにおけるこの言葉の出現を調査したデータの専門家は、この言葉は比喩的にまたは皮肉を込めてよく使われるとの結論を下しました。たとえば、「（痛くて）死にそう」、「（誰かのコメントに対して）死ねばいいのに」というように使われます。

返信自体もまた、この言葉が比喩的に使われていることを示しています。このユーザーはこのThreads投稿に対する返信の中で、国税庁長官に対し、国会に出席して不正疑惑について説明するよう求めています。また、返信には「へえ」という間投詞が含まれています。委員会の見解では、「へえ」という語は通常興味や皮肉を表すため、ここでの「死ね」という言葉には文字どおりの意味はないといえます。Metaは、このユーザーが対策案（岸田首相が国会で説明責任を果たすこと）を示していることから、当該コンテンツは文字どおり死を求めているわけではなく、政治批判が目的だと評価しており、委員会もこれに同意しています。



委員会は、日本での最近の出来事から、人々が政治指導者の死を求める呼びかけに敏感になっていることは認識しています。2022年の安倍元首相暗殺や、2023年の岸田首相を狙ったパイプ爆弾の使用を受け、信憑性のある死の脅迫を重く受け止めることは極めて重要になっています。ただ、この事例における死の呼びかけはあくまで信憑性の低いものでした。

いじめと嫌がらせに関するコミュニティ規定

委員会は、この事例のコンテンツはいじめと嫌がらせに関するポリシーに違反するものではないと考えます。また、岸田首相はポリシーにおける公人・著名人の基準を満たしているが、当該コンテンツにより「意図的にさらされて」はないとするMetaの見解に同意します。このユーザーは岸田首相のページに直接返信しておらず、タグ付けもしていないため、当該コンテンツは岸田首相に直接向けられたものではありません。

5.2 Metaの人権に関する責任の遵守

委員会は、当該コンテンツをプラットフォームから削除したことは、Metaの人権に対する責任に反すると判断します。

表現の自由 (ICCPR第19条)

[市民的、政治的権利に関する国際規約](#) (ICCPR) 第19条では、「政治分野の公人・著名人および公的機関に関する公の場での議論」に対して「特に高い」保護を規定しています ([一般的意見34号](#) 38項)。国が表現を制限する場合は、適法性、正当な目的、ならびに、必要性および比例性の要件を満たす必要があります (ICCPR第19条 3項)。これらの要件は「3部テスト」と呼ばれることが多く、委員会ではこの3部テストのフレームワークを利用して、[国連のビジネスと人権に関する指導原則](#)に従いつつ、Metaの人権に関する責任を解釈しています (Meta自身には、[企業の人権に関するポリシー](#)があります)。委員会はこの解釈を、審査中の個別コンテンツに関する決定と、コンテンツガバナンスに対するMetaの幅広いアプローチに関する説明の両方について行っています。表現の自由に関



する国連特別報告者が述べたように、「企業に政府のような義務はないものの、自身が与える影響を鑑みれば、企業はユーザーの表現の自由に関する権利の保護について、同種の質問を評価する必要があります」(A/74/486、41項)。委員会は、国家元首に対する政治的言論については、攻撃的なものであろうとも、その重要性を認識しています。そのような指導者は、批判や政治的な反対意見を受けることが正当だからです(イランにおける抗議のスローガンおよびコロンビアにおける抗議に関する決定、一般的意見34号11項および38項を参照)。

1. 適法性 (ルールの明瞭さとわかりやすさ)

適法性の原則により、表現を制限するルールは、各個人がその内容に従って自らの行動を制御できるよう、アクセスしやすくかつ明確であり、十分な詳細を含めて策定される必要があります(一般的意見34号 25項)。さらに、そうしたルールは「履行の義務を負う者に対し、表現の自由を制限するための無制限の自由裁量を与えるものではなく」、「履行の義務を負う者に対し、適切に制限されるべき表現とそうでない表現とを区別できるよう、十分な指針を提供する」必要があります(同上)。表現の自由に関する国連特別報告者は、民間の主体によるオンラインでの発言のガバナンスを行う場合、適用されるルールは明確かつ具体的であるべきだと述べています(A/HRC/38/35 46項)。Metaのプラットフォームのユーザーは、ルールを確認し理解することができなければならず、コンテンツの審査担当者は、審査方法に関する明快なガイダンスを得ていなければなりません。

委員会は、高リスク者に対し「～に死を」という言葉を使って死を求める内容の投稿が禁止されていることについて、ユーザーにとって明快さに欠け、十分に周知もされていないとしました。

ポリシーの基本理念では、投稿されたコンテンツにおける脅迫の信憑性を評価する際には、文脈(非難、認知の向上、真剣でない、気軽など、そのコンテンツが投稿された状況)を考慮するとしています。しかし、文字どおりの意図はない



発言と、信憑性のある脅迫をどのように区別すべきなのかは明記されていません。委員会が[イランにおける抗議のスローガン](#)に関する事例で指摘したように、大規模に配置される人間の審査担当者は、標的に死を求めるなどのシグナルや基準に基づく特定のガイドラインに従っています。審査担当者には脅迫の意図や信憑性を評価する権限はなく、投稿に「～に死を」や「死ね」（今回の事例のように）などの脅迫的な文言および標的が含まれている場合、その投稿は削除されます。そのため委員会は、[イランにおける抗議のスローガン](#)に関する事例の勧告1を再提示し、Metaに対して、「～に死を」という語を使った誇張的な脅迫表現は、高リスク者に向けられた場合を除いて一般的に認められる旨の説明を、暴力と扇動に関する公開ポリシーの文言に追加すること、そして明らかに誇張的な政治的言論を保護するため、国家元首に向けられた脅迫的な発言が許可される場合の基準を設定することを求めています。

このポリシーではまた、「公人・著名人」と「高リスク者」の扱いがあまり明確ではありません。現在、公人・著名人に対する保護の度合いは比較的弱く、深刻度が中程度の暴力の脅迫は「信憑性が高い」場合にのみ削除されますが、それがその他の人物に対するものである場合は、「信憑性の有無にかかわらず」削除されます。一方で、高リスク者に対する保護の度合いは強く、このポリシーには高リスク者に対して「追加的保護」を提供するという文言が存在します。上記のとおり、内部ガイドラインでは高リスク者に対する保護の度合いが強く、高リスク者が標的の場合は、「～に死を」を削除するよう求めています。委員会からの質問に対しMetaは、成人の公人・著名人に対する誇張表現を用いた不満や不同意の表明はよく行われているが、多くの場合暴力をあおる意図はないため、このポリシーでは公人・著名人に対する深刻度が中程度の脅迫をめぐるユーザーの言論に幅広い保護を提供していると回答しました。一方、高リスク者に死を求めるなど、深刻度が高い暴力の脅迫は大きなリスクを伴い、オフラインでの暴力行為につながる恐れがあります。この事例では、Metaは岸田首相が両方のカテゴリに該当するとみなしました。委員会は、暴力と扇動に関するポリシーにおける「公人・著名人」と「高リスク者」の定義が十分に明確でないため、ユーザーが各カテゴリの内容を理解できず、さらに両者の境界が曖昧な場合は一層わかりにくくなることを懸念しています。



[イランにおける抗議のスローガン](#)に関する事例で、委員会はMetaに勧告を行い、暴力と扇動に関するコミュニティ規定を修正し、高リスク者を例示して、国家元首がこのカテゴリに含まれることを説明するよう求めました。この決定の公開後、Metaは暴力的な発言と政治表現のバランスを整えるべく、ポリシー開発プロセスに着手しました。しかしMetaは、まだどのような人物が高リスク者に該当するのかを開示していません。委員会がこの事例への対処中にMetaとのブリーフィングセッションを行ったところ、Metaは、高リスク者の内部的な定義を公開することで、一部のユーザーが既存のポリシー実施ガイドラインを回避できてしまう可能性があるとして説明しました。

詳細なガイドラインを公開することで、特定のユーザーが既定の実施ルールを回避してしまうのではないかとするMetaの懸念には、委員会も同意します。しかし委員会では、Metaはイチかゼロかのアプローチではなく、高リスク者の基本的な定義を公開し、実例を添えて説明するというアプローチを取るべきだと考えます。そうすることで、どのような職業、政治活動、公務、その他のリスク関連の行動に従事する人物が保護されるのかを、ユーザーが理解できるようになります。委員会は、このようなアプローチがポリシー実施の効率低下につながることはないと考えます。実際、委員会はイランにおける抗議スローガンに関する事例において、Metaの同意を得てそうしたリストを提示し、次のように述べました。「高リスク者には、国家元首のほか、元国家元首、国家元首の立候補者および元立候補者、国政選挙および超国家組織の選挙での立候補者（落選した場合、選挙後最大30日間）、暗殺未遂に遭った人物、活動家、ジャーナリストなどが含まれます」。この例にあげられた人物はすでに公知となっているため、コミュニティ規定に反映させるべきです。

イランの抗議スローガンに関する事例での委員会の所見と、Metaがすでに暴力と扇動に関するポリシーに加えた更新事項をふまえ、委員会はMetaに対し、高リスク者の基本的な定義を示して、どのような人物が該当するか（暗殺またはその他の暴力に遭う危険性の高い政治指導者など）を明らかにしたうえで、実例（イラ



ンの抗議スローガンに関する事例で議論された人物など) を示すよう勧告します。

II. 正当な目的

表現の自由に対するあらゆる制限は、ICCPRに記載された正当な目的のいずれかを追求するものでなければなりません。暴力と扇動に関するコミュニティ規定は、「人身に実際の危害を及ぼす、または公共の安全を直接脅かす」おそれのあるコンテンツを削除し、「起こりうるオフラインでの暴力行為を防止する」ことが狙いです。このポリシーは、生命に対する権利や身体の安全に対する権利（ICCPR第6条および第9条）を守るという正当な目的を果たすものです。

III. 必要性と比例性

ICCPR第19条3項において、表現の自由の制限は、必要性と比例性の要件により、「対象を保護する機能を適切に果たし、そうした保護機能を果たす手段の中で最も非干渉的であり、保護対象の利害に応じた強さにする」必要があります（一般的意見34号 34項）。

Metaは当初、[暴力と扇動](#)に関するポリシーに基づいて当該コンテンツを削除するという決定を下しましたが、委員会はこれを不要と考えます。それは岸田首相の安全を守る手段として、最も非干渉的なものではないからです。この分析は今回の事例において最も重要な点であり、Metaが誇張された表現と実際の脅迫をどのように区別すべきかという問題を改めて投げかけています。委員会はこれまで、[イランにおける抗議スローガン](#)、[通りで阻まれたイラン人女性](#)、および[パキスタン議会での演説の報告](#)に関する事例で、比喩的な脅迫に対する過度な措置について、繰り返し懸念を表明してきました。これらの事例はスローガン、組織的な抗議運動、および間近に迫った選挙に関するものであるため、今回の事例とは異なるかもしれません。しかし根本的な問題は同じで、信憑性のない暴力の脅迫を理由に、政治的発言が制限されています。委員会はMetaがこのような議



論を認め、無用な壁をつくることなく、ユーザーが政治家の行動やふるまいに対する反感や反対を含む政治的見解を示せるようにすべきだと考えます。

ただし委員会は、Metaの暴力と扇動に関するポリシーで、実際の脅迫と比喩的な脅迫が明確に区別されていないことを懸念しています。この問題をさらに重要なものにしてしているのが、この事例のコンテンツを2名の人間のモデレーターが違反とみなしたという点です。Metaによると、彼らは日本語を話すことができ、日本の社会政治的背景にも詳しい人物とのことでした。

[ラバト行動計画](#)に記載されている6つの要素（文脈、発言者、意図、発言の内容、発言の範囲、差し迫った被害の可能性）は、脅迫の信憑性の評価に大変役立ちます。ラバトフレームワークは、国家、人種、宗教に対する憎悪による扇動を評価するために作成されたものですが、6つの要素のテストは幅広い暴力の扇動の評価に役立ちます（[イランにおける抗議のスローガン](#)や[キューバにおける女性への抗議参加の呼びかけ](#)などを参照）。Metaが当初、岸田首相の安全を守るにはコンテンツを削除する必要があるとみなしたことをふまえ、委員会はこの6つの要素を用いて、今回の事例における脅迫の信憑性を評価しました。

当該コンテンツが投稿されたのは、2023年の岸田派が関与した脱税疑惑のさなかでした。委員会の依頼を受けた専門家は、日本ではネット上の政治批判が増加しているものの、オンラインでの脅迫と近年の日本の政治家に対する暴力に明確なつながりはないとしています。今回のユーザーは公人・著名人ではなく、フォロワーは1,000人に届きません。当該コンテンツにはビューも「いいね」もなく、他のユーザーからの関心の低さがうかがえます。このユーザーの意図は、強い言葉を用いて、政治的な不正への注目を喚起することによる政治批判だと思われます。このような行為は日本のソーシャルメディアでは珍しいことではなく（一般の意見PC-29589およびPC-29594を参照）、被害が差し迫っていた可能性は低いといえます。

委員会は、暴力の脅迫の信憑性を評価するにはコンテンツごとに異なるアプローチが必要であり、特に世界規模でのコンテンツの評価は非常に困難であることを



認識しています。また、エスカレーションされた脅威については、Metaがより正確に信憑性の評価を行えることも理解しています。委員会はMetaに対し、エスカレーションされた場合のみ、高リスク者に対して「～に死を」という言葉を使った脅迫を禁止するポリシーを施行するよう、勧告することを検討しました。この「エスカレーション限定」のポリシーでは、より詳しい文脈情報をコンテンツに適用する必要があるほか、大規模に配置された人間のモデレーターによるコンテンツの初期審査とは別に、対象分野の専門家による判断が求められることとなります。委員会は、Metaにおける対象分野の専門家の数が、大規模に配置された人間の審査担当者よりも大幅に少なく、前者の処理能力が限定的であることを理解しています。したがって、エスカレーション限定のポリシーを実施した場合は、専門家不足により、大量のコンテンツが審査しきれなくなる可能性があります。さらにエスカレーション限定のルールは、信頼できるパートナーや大々的な報道など、何らかの別の手段でMetaがその事例を認識した場合にのみ適用できます（[スーダンの即応支援部隊による捕虜の動画](#)に関する決定を参照）。つまり、「～に死を」という言葉を使った死の脅迫をMetaが審査できるのは、特定のチャンネルを通じて注意喚起された場合のみということになります。

最終的に委員会は、この方法ではポリシーの実施が抑制的になり、より多くの死の脅迫がMetaのプラットフォーム上に放置されることになると判断しました。さらに、Metaが自社プラットフォーム上でのそうしたコンテンツの拡散について、検証済みのデータを提示できなかったため、委員会もポリシーがどの程度抑制的に実施されているかを評価できませんでした。

そのため委員会は、政治的言論を適切に守るため、Metaが審査担当者に表現や現地の背景を評価するうえでの追加ガイダンスを提供すること、そしてモデレーター向けに発行するガイドラインは、その基盤となるポリシーの基本理念に合致したものにすること、が必要であると考えます。以前の事例でも同じような問題があり（イランにおける抗議スローガン、通りで阻まれたイラン人女性、パキスタン議会での演説の報告を参照）、時間に制約のあることが多い中、委員会は選挙、危機、紛争など特定の文脈に焦点を絞って、ポリシーとその実施に関する解決策を検討しました。その結果、Metaはポリシーの施行方法を調整し、危機管



理ポリシープロトコル（CPP）や完全性製品オペレーションセンター（IPOC）などの仕組みを利用して、特定の文脈を把握できるようになりました。

この事例において、Metaは委員会に、特別な対策は取っていないと報告しました。Metaによると、安倍晋三元首相の暗殺などの単発の事件は、悲劇ではあるものの、リスクや不安定な状況を拡大させるさらなるシグナルがない限り、このような仕組みを適用するには基本的に不十分だということです。その代わりにMetaは、暴力事件に関するコンテンツのみが対象となる、「違反的な暴力事件」プロトコルの適用範囲に、「暗殺」を指定しました。このような状況においては、Metaは全般的なポリシーと施行方法に頼るしかありません。そのため、信憑性のある脅迫と比喩的な脅迫を区別するための拡張可能な解決策を策定することが、政治表現を適切に守る唯一の方法になります。

さらに、Metaがこのポリシーを大規模に実施し続ける場合、自動システムの精度は人間のモデレーターが提供するトレーニングデータの質に左右され続けることとなります。委員会は、[通りで阻まれたイラン人女性](#)に関する決定で示した所見を再提示し、人間のモデレーターがルールを厳格に適用して比喩的な発言を削除した場合、そのミスは自動化によって再現、増幅され、過剰な措置につながる可能性があるとししました。

死の呼びかけについては、現実の危害が及ぶ可能性を文脈主体で評価すべきとする委員会の所見に基づき、大規模に配置される審査担当者には現行のものよりも細やかな実施ガイドラインが必要だといえます。Metaの審査担当者向けの内部ガイドラインでは、「～に死を」という言葉を使って死を求める内容のうち、高リスク者に向けられたものを削除するよう規定されています。こうしたガイドラインは、[暴力と扇動](#)に関するポリシーの基本理念を反映していません。基本理念では「文脈が重要」とされ、不満や不同意を示すために深く考えず思いつきで行われる脅迫や暴力の呼びかけは文脈によって説明されるとあります。そのため委員会は、Metaが審査担当者向けの内部ガイドラインや指示を更新して、現地の背景や言語への配慮を明確に認めるとともに、不満や不同意を示すために「深



く考えずに思いつきで」行われる脅迫や暴力の呼びかけを考慮するべきだと考えます。

最後に、委員会は、Threadsの文脈依存的なコンテンツを処理するMetaの能力にも懸念を抱いています。Metaは委員会に対し、未処理の事例がたまっていたため、この事例のコンテンツの審査が約3週間遅れたと報告しました。Metaによると、Threadsのコンテンツモデレーションでは当時、人間の審査担当者のみがThreadsレポートに対応していました。Metaは通常、レポートの自動クローズなど、複数の手法を用いて作業の遅れの累積を回避しています。レポートが48時間後に自動クローズされるということは、レポートをオープンのままにする仕組みがない限り、レポートは審査されないままクローズされ、ユーザーに効果的な救済策が与えられないこととなります。

6. 監督委員会の決定

監督委員会は、当該コンテンツを削除するとしてMetaの当初の決定を覆します。

7. 勧告

1. Metaは暴力と扇動に関するポリシーを更新し、「高リスク者」の基本的な定義を掲載して、暗殺などの暴力に見舞われるリスクの高い政治指導者などが含まれることを明確にすべきです。また、わかりやすいように例も添えること。

委員会は、暴力と扇動に関するポリシーの公開文言に変更案が反映された時点をもって、この勧告が実施されたものとみなします。

2. Metaは、大規模に配置される審査担当者向けの内部ガイドラインにおいて、「～に死を」という言葉を使って死を求める内容のうち、高リスク者に向けられたものに関する規定を更新すべきです。現地の背景や言語において、深く考えない思いつきによる暴力的脅迫を通じて不満や不同意を表明している投稿は、削除しない旨を盛り込んでください。



委員会はMetaが、高リスク者に対し「～に死を」という言葉を使って死を求める内容を含むコンテンツの誤認が削減されたという適切なデータを共有した時点をもって、この勧告が実施されたものとみなします。

3. Metaは、暴力と扇動に関するポリシー、および公人・著名人について言及されているその他のコミュニティ規定に、公人・著名人に対するいじめと嫌がらせの定義へのハイパーリンクを設置し、ユーザーがそれを高リスク者と区別できるようにすべきです。

委員会は、暴力と扇動に関するポリシーや、より幅広いMetaのコミュニティ規定の公開文言に、上記の変更案が反映された時点をもって、この勧告が実施されたものとみなします。

*手続きに関する注記：

- 監督委員会の決定は、5名のパネルにより行われ、委員会全員の多数決による承認を受けています。委員会の決定は、必ずしも全員の見解を示すものではありません。
- 監督委員会は[憲章](#)に基づき、Metaにコンテンツを削除されたユーザーや、通報したコンテンツがMetaにより削除されなかったユーザーによる申し立て、またそのように下されたMetaの決定（憲章第2条セクション1）を審査します。委員会は、コンテンツに関するMetaの決定を支持するまたは覆す権限を持ちます（憲章第3条セクション5および第4条）。委員会は、拘束力のない勧告を行うことができ、Metaはこれに対応する必要があります（憲章第3条セクション4および第4条）。Metaが勧告に従い行動することを約束した場合、委員会は実施状況を監視します。
- この事例において、委員会は独立した調査を委託しました。委員会は、地政学、信頼と安全、テクノロジーの重複領域を扱う顧問会社であるDuco Advisorsの支援を受けました。また、オンラインでの被害軽減に向けたリスクアドバイザーおよび脅威インテリジェンスサービスを提供するデジタル調査グループ、Memeticaにも調査協力を仰ぎました。言語に関する



る専門的事項については、350以上の言語に対応し世界5,000都市で活躍する人材を擁する、Lionbridge Technologies, LLCに助言を仰ぎました。