

YouTube がヘイトやハラスメントからコミュニティを保護する方法

YouTube はあらゆる人に表現の場を提供したいと考えています。悪意のある表現にさらされたり嫌がらせを受けたりすると、人々は自分のストーリーや経験を世界と共有することに消極的になってしまいます。そのため、YouTube はそうした行為を許可しないポリシーを策定しました。他の有害なコンテンツに対するアプローチと同様、ヘイトやハラスメントに関するポリシーに違反するコンテンツをすみやかに削除します。



YouTube のポリシーで保護されるのは誰ですか？

YouTube は、悪意のあるコンテンツのターゲットに応じてヘイトスピーチとハラスメントを区別しています。

ヘイトスピーチに関するポリシーでグループを保護

Google のヘイトスピーチに関するポリシーでは、特定の保護対象となる属性を持つグループに対して、暴力を扇動したり、憎悪を引き起こしたりするコンテンツを制限しています。

ハラスメントに関するポリシーは個人を保護します

YouTube のハラスメントに関するポリシーは、保護の対象となる属性や身体的特徴などの生まれ持った性質から特定できる個人を標的にした有害な脅威や侮辱を制限します。

保護の対象となる属性

YouTube は、既存の法律に照らし合わせ、さらに人権やヘイトスピーチなどの分野に精通した専門家と協議して、こうしたグループを定義しています。

- | | | | |
|------|-------|--------|------------------------------|
| ✓ 人種 | ✓ 宗教 | ✓ 性的指向 | ✓ カースト |
| ✓ 民族 | ✓ 障がい | ✓ 在留資格 | ✓ 従軍経験 |
| ✓ 国籍 | ✓ 性別 | ✓ 年齢 | ✓ 重大な暴力事件の被害者としての立場(またはその家族) |

YouTube のポリシーで制限されるのはどのようなコンテンツですか？

YouTube では次の行為は許可されません。[ヘイトスピーチとハラスメントに関するポリシー全文](#)は、オンラインでご確認いただけます。

ヘイトスピーチ

- | | |
|--------------------------|--------------|
| × 実際の暴力を正当化するために使用される陰謀論 | × 劣等感を与える言動 |
| × 人間性を失わせる行為 | × 至上主義 |
| × 暴力の扇動 | × 脅迫 |
| | × 暴力的な出来事の否定 |

ハラスメント

- | | |
|----------------------|-------------------|
| × 晒し行為 | × 人間性を失わせる行為 |
| × 暴力を連想させる言動 | × 悪質な侮辱 |
| × ストーキング | × 集中的な侮辱 |
| × 脅迫 | × 弱い立場の個人を標的にしたもの |
| × 本人の意に反した性的対象化 | |
| × 暴力事件の被害者に対するハラスメント | |
| × 死を望むこと | |

ヘイトスピーチやハラスメントの動向に YouTube はどのように対応していますか？

他の表現方法と同じように、ヘイトスピーチやハラスメントの表現も絶えず変化しています。YouTube はそのポリシーを、新しい傾向や符丁、差別の対象に先んじて進化させるべく措置を講じています。

Intelligence Desk は、ニュースソース、ソーシャルメディア、ユーザーレポートをモニタリングする社内チームです





暴力的な過激主義、至上主義、人権、ネットいじめ、言論の自由などの分野で外部の専門家と協力し、

定期的にさまざまなクリエイターからそれぞれの経験に基づいた助言を受けポリシーの策定と更新に反映しています

YouTube はどのようにポリシーを施行していますか？





YouTube はコミュニティと機械学習システムを活用してヘイトスピーチやハラスメントを検出しています

YouTube には毎分 500 時間以上の動画がアップロードされているため、視聴者やクリエイター各々が問題提起できるようにするとともに、問題のあるコンテンツをプラットフォーム全体から検出できる機械学習システムにも投資しています。

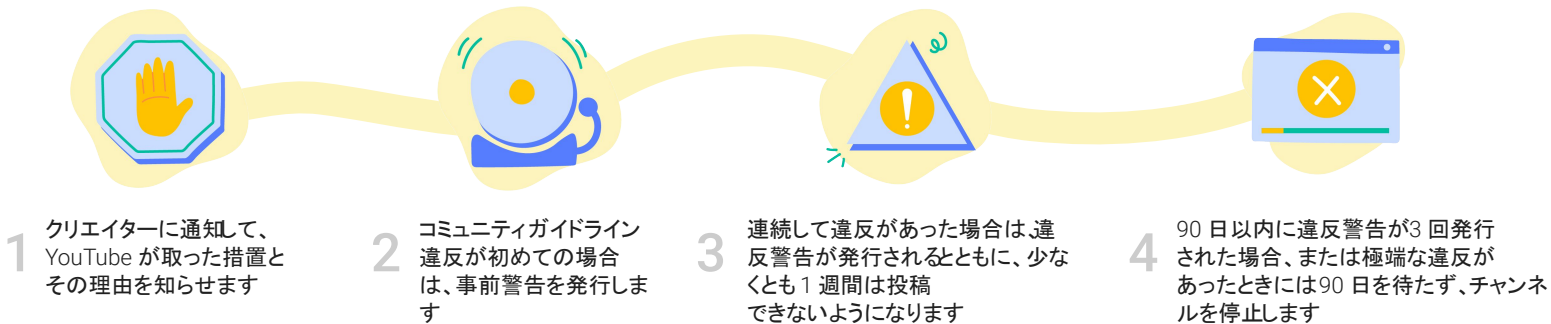
-  YouTube 公認報告者プログラムは、ヘイトスピーチやハラスメントの報告に関して特に精度の高い、政府機関パートナーと非政府組織 (NGO) 向けのプログラムです
-  クリエイターはコメント管理ツールを使用して、コメント セクションで特定の単語や個人をブロックできます
-  視聴者は動画やコメントの問題を報告できます
-  機械学習の活用より、違反コンテンツを大規模に検出しています。YouTube の機械学習システムは有害なコンテンツを特定するようにトレーニングされており、プラットフォーム全体で 24 時間稼働しています。人間の何倍もの速さで数十万本の動画を確認できます

コンテンツ審査で問題が報告されたコンテンツを評価します

問題が報告されたコンテンツは自動的に削除されません。世界中の専門家による YouTube 審査担当チームが、ヘイトスピーチやハラスメントに関するポリシーに基づいて、削除するコンテンツを決定します。この決定は、機械学習システムのトレーニングと改善に使用されます。

-  世界中の数十の言語に長けた担当者チームが、報告された問題を慎重に評価します
-  クリエイターの経歴、政治的な見解、立場、所属に関係なく、ガイドラインを一貫して適用します
-  文脈や言語の観点から、微妙な意味を理解するためのトレーニングと判断を行っています
-  ポリシーに違反しているコンテンツを削除します

ヘイトやハラスメントのコンテンツを投稿したクリエイターにはペナルティが課されます



ポリシーの例外

教育、ドキュメンタリー、科学、芸術を目的としたコンテンツ (EDSAJ コンテンツ) については、ポリシーに例外を設ける場合があります。たとえば、中傷やその他の攻撃的な表現を使ったコメディや風刺は、その背景が視聴者に明白である場合には許可される可能性があります。

ガイドラインのボーダーライン上のコンテンツの拡散を抑制

ヘイトスピーチやハラスメントに関するポリシー違反すれすれにありながら、完全にはボーダーラインを超えないコンテンツもあります。視聴者のトップページと [次の動画] リストにおすすめの動画として表示される割合を制限することで、こうしたコンテンツの拡散を抑制しています。

ヘイトスピーチやハラスメントの排除において、YouTube はどのような進歩を遂げていますか？

適切なテクノロジー、専門知識、ポリシーを導入し、ヘイトスピーチやハラスメントを含む動画、チャンネル、コメントを削除するスピードを上げています。

