

## X Corp.

### **California Terms of Service Report**

Report Date: Apr. 1, 2024

Reporting Period: Oct. 1—Dec. 31, 2023

X Corp. maintains that AB 587 is unlawful and submits this report – and all Terms of Service Reports under the law – under protest.<sup>1</sup>

(1) The current version of the terms of service of the social media platform.

For the purposes of this report, and consistent with the definition in Cal. Bus. & Prof. Code § 22675(f), X Corp.'s "Terms of Service" consist of its Legal Terms of Service, Rules, and Policies.

The current version of X Corp.'s Legal Terms of Service can be found at <a href="https://x.com/en/tos">https://x.com/en/tos</a> and are attached to this report as **Appendix A**.

The current version of X Corp.'s Rules can be found at <a href="https://help.x.com/en/rules-and-policies/x-rules">https://help.x.com/en/rules-and-policies/x-rules</a> and is attached to this report as **Appendix B**. The X Rules page contains hyperlinks to X Corp.'s relevant Policies, which are attached to this report as **Appendix C**.

(2) If a social media company has filed its first report, a complete and detailed description of any changes to the terms of service since the previous report.

**Legal Terms of Service:** X has not made any changes to its Legal Terms of Service since its previous report on January 1, 2024.

**Rules & Policies:** X has made the following changes to its Rules and Policies since its previous report on January 1, 2024:

- X deprecated its Crisis Misinformation Policy on February 5, 2024.
- X updated its Abuse and Harassment policy on March 3, 2024 to align with local laws. Among other changes, the policy was updated to include the following:

**Use of Prior Names and Pronouns.** Where required by local laws, we will reduce the visibility of posts that purposefully use different pronouns to address someone other than what that person uses for themselves, or that use a previous name that someone no longer goes by as part of their transition. Given the complexity of determining whether

<sup>&</sup>lt;sup>1</sup> See X Corp. v. Bonta, 2:23-cv-01939-WBS-AC (E.D. Cal.), appeal pending, 24-271 (9th Cir.).

such a violation has occurred, we must always hear from the target to determine if a violation has occurred.

A comparison of the current version of the Abuse and Harassment policy against the prior version of the policy is attached to this report as **Appendix D**.

(3) A statement of whether the current version of the terms of service defines each of the following categories of content, and, if so, the definitions of those categories, including any subcategories:

#### (A) Hate speech or racism.

The current version of the X Terms of Service does not define "hate speech" or "racism." X has policies that address "Hateful Conduct" and "Abuse and Harassment," among other policies.

#### (B) Extremism or radicalization.

The current version of the X Terms of Service does not define "extremism" or "radicalization." X has policies that address "Violent & Hateful Entities" and "Hateful Conduct," among other policies.

#### (C) Disinformation or misinformation.

The current version of the X Terms of Service does not define "disinformation" or "misinformation."

X addresses misinformation through its "Civic Integrity" and "Synthetic and Manipulated Media" policies. Further, X's Community Notes feature provides relevant additional context to posts.

#### (D) Harassment.

The current version of the X Terms of Service includes a category defining "Targeted Harassment" as follows:

"[M]alicious, unreciprocated targeting (such as mentioning or tagging) of individual(s), particularly when shared to humiliate or degrade someone. This can mean:

- Sharing multiple Posts, over a short period of time, or continuously posting replies with malicious content, to target an individual. This includes accounts dedicated to harassing an individual or multiple individuals.
- Mentioning or tagging users with malicious content."

X enforces Targeted Harassment through its Abuse and Harassment policy.

#### (E) Foreign political interference.

<sup>2</sup> The definition of "misinformation" included in the January 1, 2024 X California Terms of Service came from a Help Center article that was deprecated along with the Crisis Misinformation Policy.

The current version of the X Terms of Service does not define "Foreign political interference." X does have a policy that addresses "Civic Integrity" and "Synthetic and Manipulated Media," among other policies.

#### (F) Controlled substance distribution.

The current version of the X Terms of Service does not define "Controlled substance distribution." X does have a policy that addresses "Illegal or Certain Regulated Goods or Services."

- (4) A detailed description of content moderation practices used by the social media company for that platform, including, but not limited to, all of the following:
  - (A) Any existing policies intended to address the categories of content described in paragraph (3).

X Corp. has one policy that is intended to address the category of content described in paragraph (3)(A) ("Harassment"):

<u>Abuse and Harassment:</u> the Abuse and Harassment Policy can be found at <a href="https://help.x.com/en/rules-and-policies/abusive-behavior">https://help.x.com/en/rules-and-policies/abusive-behavior</a>, and is attached here in **Appendix C**.

X Corp. has two policies that are intended to address the category of content described in paragraph (3)(C) ("Disinformation or misinformation"):

<u>Synthetic and Manipulated Media Policy:</u> the Synthetic and Manipulated Media Policy can be found at <a href="https://help.x.com/en/rules-and-policies/manipulated-media">https://help.x.com/en/rules-and-policies/manipulated-media</a> and is attached here in **Appendix C.** 

<u>Civic Integrity Policy:</u> the Civic Integrity Policy can be found at <a href="https://help.x.com/en/rules-and-policies/election-integrity-policy">https://help.x.com/en/rules-and-policies/election-integrity-policy</a> and is attached here in **Appendix C**.

X Corp. has one policy that is intended to address the category of content described in paragraph (3)(F) ("Controlled Substance Distribution"):

<u>Illegal or Certain Regulated Goods or Services Policy:</u> the Illegal or Certain Regulated Goods or Services Policy can be found at

https://help.x.com/en/rules-and-policies/regulated-goods-services and is attached here in **Appendix C**.

X Corp. does not have policies intended to address the other categories of content as they are described in Cal. Bus. & Prof. Code § 22677(a)(3). X Corp. does have other policies and rules that are intended to address a broad range of safety, privacy, and legal issues. These policies and rules can be found at <a href="https://help.x.com/en/rules-and-policies/">https://help.x.com/en/rules-and-policies/</a>.

(B) How automated content moderation systems enforce terms of service of the social media platform and when these systems involve human review.

X employs a combination of heuristics and machine learning algorithms to automatically detect content that violates the X Rules and policies enforced on our platform.

#### MACHINE LEARNING MODELS

We use combinations of natural language processing models, image processing models and other sophisticated machine learning methods to detect potentially violative content. These models vary in complexity and in the outputs they produce. For example, the model used to detect abuse on the platform is trained on abuse violations detected in the past. Content flagged by these machine learning models is either reviewed by human content reviewers before an action is taken or, in some cases, automatically actioned based on predicted accuracy of the model's output.

#### **HEURISTIC MODELS**

Heuristics are typically utilized to enable X to react quickly to new forms of violations that emerge on the platform. Heuristics are common patterns of text or keywords that may be typical of a certain category of violations. Pieces of content detected by heuristics may also get reviewed by human content reviewers before an action is taken on the content. These heuristics are used to flag content for review by human agents and to prioritize the order in which such content is reviewed.

## (C) How the social media company responds to user reports of violations of the terms of service.

How X moderates content can be broadly categorized in three buckets: content and accounts that remain, those that are restricted, and those that are removed.

**Remain**: The overwhelming majority of content on X is healthy—meaning it does not violate our Terms of Service and Rules or our content-specific policies such as Hateful Conduct, Abuse & Harassment, and others.

Restrict: This is where X's Freedom of Speech, Not Reach enforcement philosophy is used. For content that X may interpret as potentially violating its policies—meaning it's awful, but lawful—X restricts the reach of posts by making the content less discoverable. When X decides to restrict a piece of content, a restricted reach label is applied, the ability to engage with the content is taken away, and its reach is restricted to views occurring directly on the author's profile. Restricted reach labels are not in use for all policies; X's restricted reach labels were initially only applied to Hateful Conduct, but their application has since been expanded to X's Abuse & Harassment, Civic Integrity, and Violent Speech Policies. That said, X has a range of enforcement options for the variety of use cases that it faces every day. For example, X may also place an account in read-only mode, temporarily limiting its ability to post, Repost, or Like.

**Remove**: Certain types of content, such as targeted violent threats, targeted harassment, or privacy violations, can be extremely harmful if not removed and X either suspends users outright for such content or requires that the content be deleted before the user is permitted to return to the platform. We also suspend accounts that share child sexual exploitative material as well as accounts that belong to terrorist organizations, perpetrators of violent attacks, and/or individuals who affiliate with and promote their illicit activities.

(D) How the social media company would remove individual pieces of content, users, or groups that violate the terms of service, or take broader action against individual users or against groups of users that violate the terms of service.

To enforce its Rules, X uses a combination of machine learning and human review. X's systems are able to surface content to human moderators who use important context to make decisions about potential rule violations. This work is led by an international, cross-functional team with 24-hour coverage and the ability to cover multiple languages. X also has a complaints process for any potential errors that may occur.

#### Examples of actions X may take:

- Placing a post behind a notice: X may place some forms of sensitive media like adult content or graphic violence behind an interstitial advising viewers to be aware that they will see sensitive media if they click through. X also gives users the option to control whether they see sensitive media.
- Withholding a post based on age: X restricts views of specific forms of sensitive media such as adult content for viewers who are under 18, or who do not include a birth date on their profile, with interstitials.
- Withholding a post or account in a country: X may withhold access to certain content in a particular country if it receives a valid and properly scoped request from an authorized entity in that country.
- (E) The languages in which the social media platform does not make terms of service available, but does offer product features, including, but not limited to, menus and prompts.
  - Bangla বাংলা
  - Basque (beta) euskara
  - Galician (beta) galego
  - Gujarati ગુજરાતી
  - Hebrew עברית
  - Hindi हिन्दी
  - Kannada ಕನ್ನ
  - Malay Melayu
  - Marathi मराठी
  - Norwegian norsk

- Serbian српски
- Tamil தமிழ்
- Thai ไทย
- Turkish Türkçe
- Ukrainian українська
- Urdu (beta) اردو

(5) (A) Information on content that was flagged by the social media company as content belonging to any of the categories described in paragraph (3), including all of the following:

#### (i) The total number of flagged items of content.

Policy	Total Number of Reports
Abuse & Harassment	25,467,017
Civic Integrity	N/A <sup>3</sup>
Synthetic & Manipulated Media	N/A <sup>4</sup>

#### (ii) The total number of actioned items of content.

	Enforcement Action Taken		
Policy	Restricted Reach of Post(s)	Post Removal	Labels (Post-Level)
Abuse & Harassment	126,154	616,226	N/A
Civic Integrity	18	8	N/A
Synthetic & Manipulated Media	N/A	4	2,659

(iii) The total number of actioned items of content that resulted in action taken by the social media company against the user or group of users responsible for the content.

Policy	Account Suspensions
Abuse & Harassment	295,860
Civic Integrity	3
Synthetic & Manipulated Media	N/A

<sup>&</sup>lt;sup>3</sup> This policy is enforced proactively; users do not have the ability to report violations of it.

<sup>&</sup>lt;sup>4</sup> This policy is enforced proactively; users do not have the ability to report violations of it.

Policy	Enforcement Type	Account suspensions
Abuse & Harassment	Auto Enforced	0
Abuse & Harassment	Manual Enforced	295,859
Abuse & Harassment	Manual Enforced - Proactive Detection	1
Civic Integrity	Manual Enforced	3
Synthetic & Manipulated Media	N/A	N/A

## (iv) The total number of actioned items of content that were removed, demonetized, or deprioritized by the social media company.

	Enforcement Action Taken		
Policy	Restricted Reach of Post Removal		
Abuse & Harassment	126,154	619,662	
Civic Integrity	0	8	
Synthetic & Manipulated Media	N/A	N/A	

Policy	Enforcement	Total Number of Posts Removed
Abuse & Harassment	Auto Enforced	4,647
Abuse & Harassment	Manual Enforced	601,541
Abuse & Harassment	Manual Enforced - Proactive Detection	10,037
Civic Integrity	Auto Enforced	2
Civic Integrity	Manual Enforced	5
Synthetic & Manipulated Media	Manual Enforced	4

Policy	Enforcement	Total Number of Posts Restricted in Reach
Abuse & Harassment	Manual	126,154
Civic Integrity	Manual	18
Synthetic & Manipulated Media	N/A	N/A

#### (v) The number of times actioned items of content were viewed by users.

Policy	Total Impressions	Total Linger Impressions before Post Removal
Abuse & Harassment	617,309,135	600,935,906
Civic Integrity	-	-
Synthetic & Manipulated Media	N/A	N/A

Policy	Total Impressions	Total Linger Impressions before Restricting Reach of Posts
Abuse & Harassment	264,079,198	226,264,440
Civic Integrity	462	462
Synthetic & Manipulated Media	N/A	N/A

(vi) The number of times actioned items of content were shared, and the number of users that viewed the content before it was actioned.

This data is not available due to technical limitations.

(vii) The number of times users appealed social media company actions taken on that platform and the number of reversals of social media company actions on appeal disaggregated by each type of action.

1. For posts that were restricted in reach:

Policy	Restricted Reach of Posts	Total Appeals	Overturn Count
Abuse & Harassment	126,154	2,697	1,851
Civic Integrity	N/A	N/A	N/A
Synthetic & Manipulated Media	N/A	N/A	N/A

#### 2. For posts that users were asked to remove:

Policy	Content Removal	Total Appeals	Overturn Count
Abuse & Harassment	616,226	9,546	1,472
Civic Integrity	8	0	0
Synthetic & Manipulated Media	N/A	N/A	N/A

#### 3. For account level suspensions:

Policy	Account Suspensions	Total Appeals	Overturn Count
Abuse & Harassment	295,860	19,822	2,742
Civic Integrity	3	2	0
Synthetic & Manipulated Media	N/A	N/A	N/A

- (B) All information required by subparagraph (A) shall be disaggregated into the following categories:
  - (i) The category of content, including any relevant categories described in paragraph (3).

Categories of content are indicated by the "Policy" column in the charts provided above.

(ii) The type of content, including, but not limited to, posts, comments, messages, profiles of users, or groups of users.

The tables provided in response to Question 5(A) include enforcement data at the post, profile, and account-level. Please note that the X platform does not have a separate "comments" feature, and posts that reference or respond to other posts are themselves considered posts. The enforcement categorisation is the same for posts and "comments."

X enforces its policies differently in the context of direct messages. When a direct message is reported, X immediately hides the content of the reported post and replaces it with a notice saying that the content has been reported. The reported message and conversation are also deleted from the reporter's inbox. For direct messages, the enforcement is different as the original content of reported posts is replaced with a notice stating that the content has been reported and the reported messages and conversations disappear from the reporters' inbox. In a group direct message conversation, the violating direct message may be placed behind a notice to ensure no one else in the group can see it again.

In Q4 2023, X took action on 204,761 direct messages for violations of its Abuse & Harassment Policy. Data is not available for Civic Integrity and Synthetic & Manipulated Media enforcement.

X does not segregate its users into defined groups.

(iii) The type of media of the content, including, but not limited to, text, images, and videos.

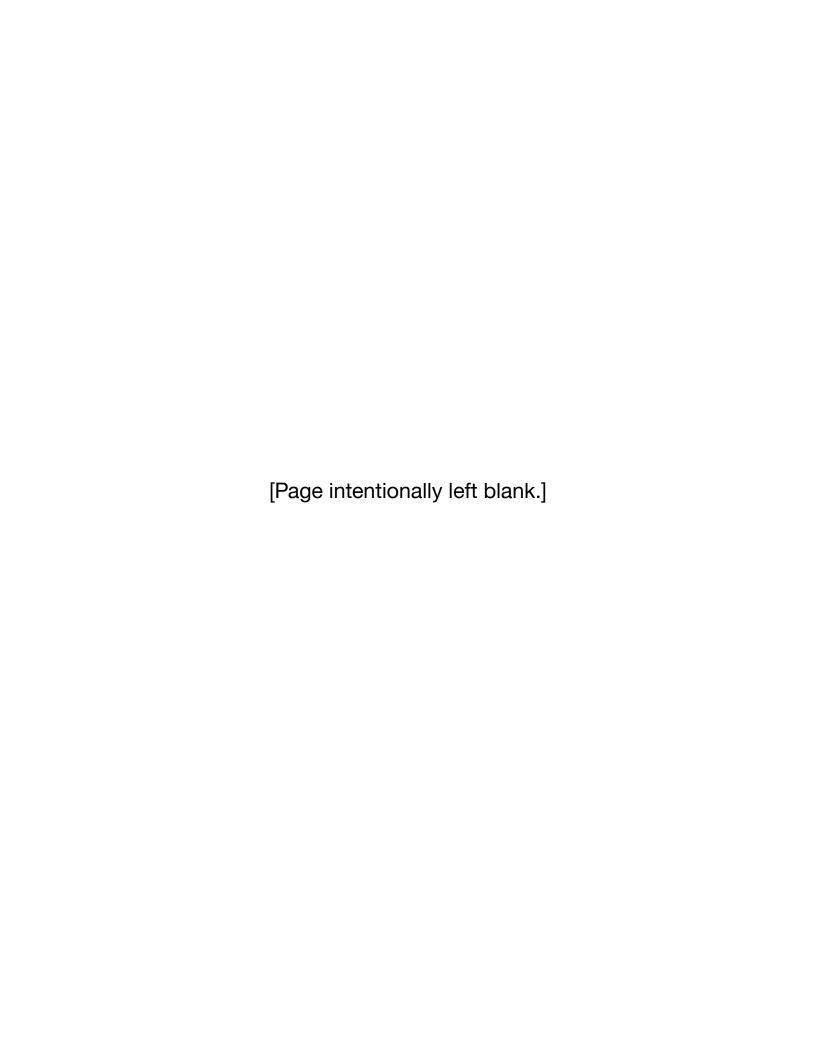
Our data includes post-level enforcement where a piece of media falls within the scope of the Synthetic and Manipulated Media policy. We do not have a breakdown by type of media.

(iv) How the content was flagged, including, but not limited to, flagged by company employees or contractors, flagged by artificial intelligence software, flagged by community moderators, flagged by civil society partners, and flagged by users.

Our data includes reports made by our users through our in-app reporting option.

(v) How the content was actioned, including, but not limited to, actioned by company employees or contractors, actioned by artificial intelligence software, actioned by community moderators, actioned by civil society partners, and actioned by users.

The tables provided in response to Question 5(A) include enforcement data breakdowns by method of enforcement (i.e., manual enforcement, proactive enforcement, and auto-actioned content).



## Appendix A: X Terms of Service as of 3/31/2024

## X Terms of Service

### **Summary of our Terms**

These Terms of Service ("Terms") are part of the User Agreement– a legally binding contract governing your use of X. You should read these Terms of Service ("Terms") in full, but here are a few key things you should take away:

- You will see advertising on the platform: In exchange for accessing the Services, X and our third-party providers and partners may display advertising to you.
- When posting Content and otherwise using the Services, you must comply
  with this User Agreement and Applicable Law: You are responsible for your use
  of the Services and your Content. You must comply with this User Agreement, its
  incorporated policies, and all applicable laws.
- You must abide by the Services' acceptable use terms: You may not access the
  Services in any way other than through the currently available, published interfaces
  that we provide. For example, this means that you cannot scrape the Services, try
  to work around any technical limitations we impose, or otherwise attempt to disrupt
  the operation of the Services.
- We have broad enforcement rights: X reserves the right to take enforcement
  actions against you if you do violate these terms, such as, for example, removing
  your Content, limiting visibility, discontinuing your access to X, or taking legal
  action. We may also suspend or terminate your account for other reasons, such as
  prolonged inactivity, risk of legal exposure, or commercial inviability.
- There are Intellectual Property Licenses in these Terms: You retain ownership and rights to any of your Content you post or share, and you provide us with a broad, royalty-free license to make your Content available to the rest of the world and to let others do the same. Conversely, we provide you a license to use the software we provide as part of the Services, such as the X mobile application, solely for the purpose of enabling you to use and enjoy the benefit of the Services.
- Your use of the Services is at your own risk: We provide the Services on an "AS IS" and "AS AVAILABLE" basis, and we disclaim all warranties, responsibility, and

liability to you or others to the extent permitted by law. You may be exposed to offensive or harmful content posted by other users. The Services may change from time to time, and we may limit or terminate availability of the Services or particular features to you or other users at any time.

You have remedies and redress mechanisms, but our liability is limited: You have a right to terminate this agreement at any time by deactivating your account and discontinuing use of the Services. Note that we will not be liable for certain types of damages as described in the agreement, and in any event, our aggregate liability shall not exceed the greater of \$100 USD or the amount you paid us, if any, in the past six months for the Services giving rise to the claim. Further, if you believe that your Content has been copied in a way that constitutes copyright infringement, the reporting process is detailed in these Terms.

Please also note that these Terms incorporate our Privacy Policy (https://x.com/privacy) as well as other terms applicable to your use of the Services and your Content. Finally, these terms may vary depending on where you live, but in any case, you must be at least 13 years old to use X.

If you live outside the European Union, EFTA States, or the United Kingdom, including if you live in the United States, the X User Agreement comprises these Terms of Service, our Privacy Policy, our Rules and Policies, and all incorporated policies.

If you live in the European Union, EFTA States, or the United Kingdom, the X User Agreement comprises these Terms of Service, our Privacy Policy, our Rules and Policies, and all incorporated policies.

## X Terms of Service

If you live outside the European Union, EFTA States, or the United Kingdom, including if you live in the United States

These Terms of Service ("Terms") govern your access to and use of our services, including our various websites, SMS, APIs, email notifications, applications, buttons, widgets, ads,



commerce services, and our other covered services

(https://help.x.com/rules-and-policies/x-services-and-corporate-affiliates) that link to these Terms (collectively, the "Services"), and any information, text, links, graphics, photos, audio, videos, or other materials or arrangements of materials uploaded, downloaded or appearing on the Services (collectively referred to as "Content"). By using the Services you agree to be bound by these Terms.

These Terms are an agreement between you and X Corp., which provides X and the Services, 1355 Market Street, Suite 900, San Francisco, CA 94103 U.S.A. The words "we," "us," and "our" mean X Corp.

## 1. Who May Use the Services

You may use the Services only if you agree to form a binding contract with us and are not a person barred from receiving services under the laws of the applicable jurisdiction. In any case, you must be at least 13 years old to use the Services. If you are accepting these Terms and using the Services on behalf of a company, organization, government, or other legal entity, you represent and warrant that you are authorized to do so and have the authority to bind such entity to these Terms, in which case the words "you" and "your" as used in these Terms shall refer to such entity.

## 2. Privacy

Our Privacy Policy (https://www.x.com/privacy) describes how we handle the information you provide to us when you use the Services. You understand that through your use of the Services you consent to the collection and use (as set forth in the Privacy Policy) of this information, including the transfer of this information to the United States, Ireland, and/or other countries for storage, processing and use by us and our affiliates.

## 3. Content on the Services

You are responsible for your use of the Services and for any Content you provide, including compliance with applicable laws, rules, and regulations. You should only provide Content that

you are comfortable sharing with others.

Any use or reliance on any Content or materials posted via the Services or obtained by you through the Services is at your own risk. We do not endorse, support, represent or guarantee the completeness, truthfulness, accuracy, or reliability of any Content or communications posted via the Services or endorse any opinions expressed via the Services. You understand that by using the Services, you may be exposed to Content that might be offensive, harmful, inaccurate or otherwise inappropriate, or in some cases, postings that have been mislabeled or are otherwise deceptive. All Content is the sole responsibility of the person who originated such Content. We may not monitor or control the Content posted via the Services and, we cannot take responsibility for such Content.

We reserve the right to remove Content that violates the User Agreement, including for example, copyright or trademark violations or other intellectual property misappropriation, impersonation, unlawful conduct, or harassment. Information regarding specific policies and the process for reporting or appealing violations can be found in our Help Center (https://help.x.com/rules-and-policies/x-report-violation#specific-violations and https://help.x.com/managing-your-account/suspended-x-accounts).

If you believe that your Content has been copied in a way that constitutes copyright infringement, please report this by visiting our Copyright reporting form (https://help.x.com/forms/dmca) or contacting our designated copyright agent at:

X Corp.

Attn: Copyright Agent 1355 Market Street, Suite 900 San Francisco, CA 94103

Reports: https://help.x.com/forms/dmca

Email: copyright@x.com

### Your Rights and Grant of Rights in the Content

You retain your rights to any Content you submit, post or display on or through the Services. What's yours is yours — you own your Content (and your incorporated audio, photos and videos are considered part of the Content).

By submitting, posting or displaying Content on or through the Services, you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in

any and all media or distribution methods now known or later developed (for clarity, these rights include, for example, curating, transforming, and translating). This license authorizes us to make your Content available to the rest of the world and to let others do the same. You agree that this license includes the right for us to provide, promote, and improve the Services and to make Content submitted to or through the Services available to other companies, organizations or individuals for the syndication, broadcast, distribution, repost, promotion or publication of such Content on other media and services, subject to our terms and conditions for such Content use. Such additional uses by us, or other companies, organizations or individuals, is made with no compensation paid to you with respect to the Content that you submit, post, transmit or otherwise make available through the Services as the use of the Services by you is hereby agreed as being sufficient compensation for the Content and grant of rights herein.

We have an evolving set of rules for how ecosystem partners can interact with your Content on the Services. These rules exist to enable an open ecosystem with your rights in mind. You understand that we may modify or adapt your Content as it is distributed, syndicated, published, or broadcast by us and our partners and/or make changes to your Content in order to adapt the Content to different media.

You represent and warrant that you have, or have obtained, all rights, licenses, consents, permissions, power and/or authority necessary to grant the rights granted herein for any Content that you submit, post or display on or through the Services. You agree that such Content will not contain material subject to copyright or other proprietary rights, unless you have necessary permission or are otherwise legally entitled to post the material and to grant us the license described above.

## 4. Using the Services

Please review our Rules and Policies, which are part of the User Agreement and outline conduct that is prohibited on the Services. You may use the Services only in compliance with these Terms and all applicable laws, rules and regulations. X takes enforcement actions when Content or user behavior is in violation of our Rules and Policies or in relation to sensitive media. You can review X's enforcement options and how you can appeal our enforcement decision here.

The Services evolve constantly. As such, the Services may change from time to time, at our discretion. We may stop (permanently or temporarily) providing the Services or any features within the Services to you or to users generally. We also retain the right to create limits on use

and storage at our sole discretion at any time. We may also remove or refuse to distribute any Content on the Services, limit distribution or visibility of any Content on the service, suspend or terminate users, and reclaim usernames without liability to you.

In consideration for our granting you access to and use of the Services, you agree that we and our third-party providers and partners may place advertising on the Services or in connection with the display of Content or information from the Services whether submitted by you or others. We also reserve the right to access, read, preserve, and disclose any information as we reasonably believe is necessary to (i) satisfy any applicable law, regulation, legal process or governmental request, (ii) enforce the Terms, including investigation of potential violations hereof, (iii) detect, prevent, or otherwise address fraud, security or technical issues, (iv) respond to user support requests, or (v) protect the rights, property or safety of X, its users and the public. We do not disclose personally-identifying information to third parties except in accordance with our Privacy Policy.

Certain services or features may be offered on X for which additional terms and conditions may apply in connection with your use of those services. By using or paying for any of these additional services, you agree to any additional terms applicable to those services, and those additional terms become part of our agreement with you. If any of the applicable additional terms conflict with these Terms, the additional terms will prevail while you are using those services to which they apply.

If you use paid features of the Services, you agree to the applicable Terms for Paid Services (https://legal.x.com/purchaser-terms.html).

If you use developer features of the Services, including but not limited to X for Websites (https://developer.x.com/docs/twitter-for-websites), X Cards (https://developer.x.com/docs/twitter-for-websites/cards/overview/abouts-cards), Public API (https://developer.x.com/docs), or Sign in with X (https://developer.x.com/docs/authentication/guides/log-in-with-twitter), you agree to our Developer Agreement (https://developer.x.com/developer-terms/agreement) and Developer Policy (https://developer.x.com/developer-terms/policy). If you want to reproduce, modify, create derivative works, distribute, sell, transfer, publicly display, publicly perform, transmit, or otherwise use the Services or Content on the Services, you must use the interfaces and instructions we provide, except as permitted through the Services, these Terms, or the terms provided on https://developer.x.com/developer-terms. Otherwise, all such actions are strictly prohibited. If you are a security researcher, you are required to comply with the rules of our Vulnerability Reporting Program (https://hackerone.com/x). The requirements set out in the preceding paragraph may not apply to those participating in our Vulnerability Reporting

Program.

If you use advertising features of the Services, you agree to our Master Services Agreement (https://ads.x.com/terms).

#### **Your Account**

You may need to create an account to use the Services. You are responsible for safeguarding your account, so use a strong password and limit its use to this account. We cannot and will not be liable for any loss or damage arising from your failure to comply with the above.

You can control most communications from the Services. We may need to provide you with certain communications, such as service announcements and administrative messages. These communications are considered part of the Services and your account, and you may not be able to opt-out from receiving them. If you added your phone number to your account and you later change or deactivate that phone number, you must update your account information to help prevent us from communicating with anyone who acquires your old number.

#### Your License to Use the Services

We give you a personal, worldwide, royalty-free, non-assignable and non-exclusive license to use the software provided to you as part of the Services. This license has the sole purpose of enabling you to use and enjoy the benefit of the Services as provided on X, in the manner permitted by these Terms.

The Services are protected by copyright, trademark, and other laws of both the United States and other countries. Nothing in the Terms gives you a right to use the X name or any of the X trademarks, logos, domain names, other distinctive brand features, and other proprietary rights. All right, title, and interest in and to the Services (excluding Content provided by users) are and will remain our and our licensors' exclusive property. Any feedback, comments, or suggestions you may provide regarding X, or the Services is entirely voluntary and we will be free to use such feedback, comments or suggestions as we see fit and without any obligation to you.

#### Misuse of the Services

You also agree not to misuse the Services, for example, by interfering with them or accessing them using a method other than the interface and the instructions that we provide. You agree that you will not work around any technical limitations in the software provided to you as part of the Services, or reverse engineer, decompile or disassemble the software, except and only to the extent that applicable law expressly permits. You may not do any of the following while accessing or using the Services: (i) access, tamper with, or use non-public areas of the Services, our computer systems, or the technical delivery systems of our providers; (ii) probe, scan, or test the vulnerability of any system or network or breach or circumvent any security or authentication measures; (iii) access or search or attempt to access or search the Services by any means (automated or otherwise) other than through our currently available, published interfaces that are provided by us (and only pursuant to the applicable terms and conditions), unless you have been specifically allowed to do so in a separate agreement with us (NOTE: crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited); (iv) forge any TCP/IP packet header or any part of the header information in any email or posting, or in any way use the Services to send altered, deceptive or false source-identifying information; (v) engage in any conduct that violates our Platform Manipulation and Spam Policy or any other Rules and Policies; or (vi) interfere with, or disrupt, (or attempt to do so), the access of any user, host or network, including, without limitation, sending a virus, overloading, flooding, spamming, mail-bombing the Services, or by scripting the creation of Content in such a manner as to interfere with or create an undue burden on the Services. It is also a violation of these Terms to facilitate or assist others in violating these Terms, including by distributing products or services that enable or encourage violation of these Terms.

### **Ending These Terms**

You may end your legal agreement with us at any time by deactivating your accounts and discontinuing your use of the Services. See

https://help.x.com/managing-your-account/how-to-deactivate-x-account for instructions on how to deactivate your account and the Privacy Policy for more information on what happens to your information.

We may suspend or terminate your account or cease providing you with all or part of the Services at any time if we reasonably believe: (i) you have violated these Terms or our Rules and Policies, (ii) you create risk or possible legal exposure for us; (iii) your account should be removed due to unlawful conduct; (iv) your account should be removed due to prolonged inactivity; or (v) our provision of the Services to you is no longer commercially viable. We will

make reasonable efforts to notify you by the email address associated with your account or the next time you attempt to access your account, depending on the circumstances. To the extent permitted by law, we may also terminate your account or cease providing you with all or part of the Services for any other reason or no reason at our convenience. In all such cases, the Terms shall terminate, including, without limitation, your license to use the Services, except that the following sections shall continue to apply: 2, 3, 5, 6, and the misuse provisions of Section 4 ("Misuse of the Services"). If you believe your account was terminated in error you can file an appeal following the steps found in our Help Center (https://help.x.com/forms/account-access/appeals). For the avoidance of doubt, these Terms survive the deactivation or termination of your account.

## 5. Disclaimers and Limitations of Liability

#### The Services are Available "AS-IS"

Your access to and use of the Services or any Content are at your own risk. You understand and agree that the Services are provided to you on an "AS IS" and "AS AVAILABLE" basis. The "X Entities" refers to X Corp., its parents, affiliates, related companies, officers, directors, employees, agents, representatives, partners, and licensors. Without limiting the foregoing, to the maximum extent permitted under applicable law, THE X ENTITIES DISCLAIM ALL WARRANTIES AND CONDITIONS, WHETHER EXPRESS OR IMPLIED, OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT. The X Entities make no warranty or representation and disclaim all responsibility and liability for: (i) the completeness, accuracy, availability, timeliness, security or reliability of the Services or any Content; (ii) any harm to your computer system, loss of data, or other harm that results from your access to or use of the Services or any Content; (iii) the deletion of, or the failure to store or to transmit, any Content and other communications maintained by the Services; and (iv) whether the Services will meet your requirements or be available on an uninterrupted, secure, or error-free basis. No advice or information, whether oral or written, obtained from the X Entities or through the Services, will create any warranty or representation not expressly made herein.

## **Limitation of Liability**

TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, THE X ENTITIES SHALL NOT BE LIABLE FOR ANY INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR PUNITIVE

DAMAGES, OR ANY LOSS OF PROFITS OR REVENUES, WHETHER INCURRED DIRECTLY OR INDIRECTLY, OR ANY LOSS OF DATA, USE, GOODWILL, OR OTHER INTANGIBLE LOSSES. RESULTING FROM (i) YOUR ACCESS TO OR USE OF OR INABILITY TO ACCESS OR USE THE SERVICES; (ii) ANY CONDUCT OR CONTENT OF ANY THIRD PARTY ON THE SERVICES, INCLUDING WITHOUT LIMITATION, ANY DEFAMATORY, OFFENSIVE OR ILLEGAL CONDUCT OF OTHER USERS OR THIRD PARTIES; (iii) ANY CONTENT OBTAINED FROM THE SERVICES; OR (iv) UNAUTHORIZED ACCESS, USE OR ALTERATION OF YOUR TRANSMISSIONS OR CONTENT. IN NO EVENT SHALL THE AGGREGATE LIABILITY OF THE X ENTITIES EXCEED THE GREATER OF ONE HUNDRED U.S. DOLLARS (U.S. \$100.00) OR THE AMOUNT YOU PAID US, IF ANY, IN THE PAST SIX MONTHS FOR THE SERVICES GIVING RISE TO THE CLAIM. THE LIMITATIONS OF THIS SUBSECTION SHALL APPLY TO ANY THEORY OF LIABILITY, WHETHER BASED ON WARRANTY, CONTRACT, STATUTE, TORT (INCLUDING NEGLIGENCE) OR OTHERWISE, AND WHETHER OR NOT THE X ENTITIES HAVE BEEN INFORMED OF THE POSSIBILITY OF ANY SUCH DAMAGE, AND EVEN IF A REMEDY SET FORTH HEREIN IS FOUND TO HAVE FAILED OF ITS ESSENTIAL PURPOSE.

### 6. General

We may revise these Terms from time to time. The changes will not be retroactive, and the most current version of the Terms, which will always be at x.com/tos, will govern our relationship with you. We will try to notify you of material revisions, for example via a service notification or an email to the email associated with your account. By continuing to access or use the Services after those revisions become effective, you agree to be bound by the revised Terms. To the extent permitted by law, you also waive the right to participate as a plaintiff or class member in any purported class action, collective action or representative action proceeding.

The laws of the State of California, excluding its choice of law provisions, will govern these Terms and any dispute that arises between you and us. All disputes related to these Terms or the Services will be brought solely in the federal or state courts located in San Francisco County, California, United States, and you consent to personal jurisdiction and waive any objection as to inconvenient forum. To the extent permitted by law, you also waive the right to participate as a plaintiff or class member in any purported class action, collective action or representative action proceeding.

If you are a federal, state, or local government entity in the United States using the Services in your official capacity and legally unable to accept the controlling law, jurisdiction or venue



clauses above, then those clauses do not apply to you. For such U.S. federal government entities, these Terms and any action related thereto will be governed by the laws of the United States of America (without reference to conflict of laws) and, in the absence of federal law and to the extent permitted under federal law, the laws of the State of California (excluding choice of law).

The X User Agreement is written in English but is made available in multiple languages through translations. X strives to make the translations as accurate as possible to the original English version. However, in case of any discrepancies or inconsistencies, the English language version of the X User Agreement shall take precedence. You acknowledge that English shall be the language of reference for interpreting and constructing the terms of the X User Agreement.

In the event that any provision of these Terms is held to be invalid or unenforceable, then that provision will be limited or eliminated to the minimum extent necessary, and the remaining provisions of these Terms will remain in full force and effect. Our failure to enforce any right or provision of these Terms will not be deemed a waiver of such right or provision.

If you have any questions about these Terms, please contact us.

Effective: September 29, 2023

**Archive of Previous Terms** 

## X Terms of Service

# If you live in the European Union, EFTA States, or the United Kingdom

These Terms of Service ("Terms") govern your access to and use of the services, including our various websites, SMS, APIs, email notifications, applications, buttons, widgets, ads, commerce services, and our other covered services

(https://help.x.com/rules-and-policies/x-services-and-corporate-affiliates) that link to these Terms (collectively, the "Services"), and any information, text, links, graphics, photos, audio, videos, or other materials or arrangements of materials uploaded, downloaded or appearing on the Services (collectively referred to as "Content"). By using the Services you agree to be bound by these



Terms.

These Terms are an agreement between you and Twitter International Unlimited Company (Co. number 503351, VAT number IE9803175Q), an Irish company, which provides X and the Services, with its registered office at One Cumberland Place, Fenian Street Dublin 2, D02 AX07 Ireland. The words "we," "us," and "our," mean Twitter International Unlimited Company.

## 1. Who May Use the Services

You may use the Services only if you agree to form a binding contract with us and are not a person barred from receiving services under the laws of the applicable jurisdiction. In any case, you must be at least 13 years old to use the Services. If you are accepting these Terms and using the Services on behalf of a company, organization, government, or other legal entity, you represent and warrant that you are authorized to do so and have the authority to bind such entity to these Terms, in which case the words "you" and "your" as used in these Terms shall refer to such entity.

## 2. Privacy

Our Privacy Policy (https://www.x.com/privacy) describes how we handle the information you provide to us when you use the Services. You understand that through your use of the Services you consent to the collection and use (as set forth in the Privacy Policy) of this information, including the transfer of this information to the United States, Ireland, and/or other countries for storage, processing and use by us and our affiliates.

### 3. Content on the Services

You are responsible for your use of the Services and for any Content you provide, including compliance with applicable laws, rules, and regulations. You should only provide Content that you are comfortable sharing with others.

Any use or reliance on any Content or materials posted via the Services or obtained by you through the Services is at your own risk. We do not endorse, support, represent or guarantee the

completeness, truthfulness, accuracy, or reliability of any Content or communications posted via the Services or endorse any opinions expressed via the Services. You understand that by using the Services, you may be exposed to Content that might be offensive, harmful, inaccurate or otherwise inappropriate, or in some cases, postings that have been mislabeled or are otherwise deceptive. All Content is the sole responsibility of the person who originated such Content. We may not monitor or control the Content posted via the Services and, we cannot take responsibility for such Content.

We reserve the right to remove Content that violates the User Agreement, including for example, copyright or trademark violations or other intellectual property misappropriation, impersonation, unlawful conduct, or harassment. Information regarding specific policies and the process for reporting or appealing violations can be found in our Help Center

(https://help.x.com/rules-and-policies/x-report-violation and https://help.x.com/managing-your-account/suspended-x-accounts).

If you believe that your Content has been copied in a way that constitutes copyright infringement, please report this by visiting our Copyright reporting form (https://help.x.com/forms/dmca) or contacting our designated copyright agent at:

X Corp.

Attn: Copyright Agent 1355 Market Street, Suite 900

San Francisco, CA 94103

Reports: https://help.x.com/forms/dmca

Email: copyright@x.com

### Your Rights and Grant of Rights in the Content

You retain your rights to any Content you submit, post or display on or through the Services. What's yours is yours — you own your Content (and your incorporated audio, photos and videos are considered part of the Content).

By submitting, posting or displaying Content on or through the Services, you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods now known or later developed (for clarity, these rights include, for example, curating, transforming, and translating). This license authorizes us to make your Content available to the rest of the world and to let others do the same. However, if you have chosen via our features to limit the distribution of your Content to a restricted community, we will

respect that choice. You also agree that this license includes the right to analyze text and other information you provide with the view to improve the Services. You agree that this license includes the right for us to provide, promote, and improve the Services and to make Content submitted to or through the Services available to other companies, organizations or individuals for the syndication, broadcast, distribution, repost, promotion or publication of such Content on other media and services, subject to our terms and conditions for such Content use. Such additional uses by us, or other companies, organizations or individuals, is made with no compensation paid to you with respect to the Content that you submit, post, transmit or otherwise make available through the Services as the use of the Services by you is hereby agreed as being sufficient compensation for the Content and grant of rights herein.

We have an evolving set of rules for how ecosystem partners can interact with your Content on the Services. These rules exist to enable an open ecosystem with your rights in mind. You understand that we may modify or adapt your Content as it is distributed, syndicated, published, or broadcast by us and our partners and/or make changes to your Content in order to adapt the Content to different media.

You represent and warrant that you have, or have obtained, all rights, licenses, consents, permissions, power and/or authority necessary to grant the rights granted herein for any Content that you submit, post or display on or through the Services. You agree that such Content will not contain material subject to copyright or other proprietary rights, unless you have necessary permission or are otherwise legally entitled to post the material and to grant us the license described above.

## 4. Using the Services

Please review our Rules and Policies, which are part of the User Agreement and outline conduct that is prohibited on the Services. You may use the Services only in compliance with these Terms and all applicable laws, rules and regulations. X takes enforcement actions when Content or user behavior is in violation of our Rules and Policies or in relation to sensitive media. You can review X's enforcement options and how you can appeal our enforcement decision here.

The Services evolve constantly. As such, the Services may change from time to time, at our discretion. We may stop (permanently or temporarily) providing the Services or any features within the Services to you or to users generally. We also retain the right to create limits on use and storage at our sole discretion at any time. We may also remove or refuse to distribute any Content on the Services, limit distribution or visibility of any Content on the service, suspend or terminate users, and reclaim usernames if it is appropriate, including for the following reasons: (i) protecting the Services or our users; (ii) compliance with applicable laws or orders from competent authorities; (iii) breach of these Terms or our Rules and Policies or third parties' intellectual



property or other rights; (iv) if you or your Content exposes us, other users or any third party to legal or regulatory risk; and/or (v) your prolonged inactivity.

In consideration for our granting you access to and use of the Services, you agree that we and our third-party providers and partners may place advertising on the Services or in connection with the display of Content or information from the Services whether submitted by you or others. We also reserve the right to access, read, preserve, and disclose any information as we reasonably believe is necessary to (i) satisfy any applicable law, regulation, legal process or governmental request, (ii) enforce the Terms, including investigation of potential violations hereof, (iii) detect, prevent, or otherwise address fraud, security or technical issues, (iv) respond to user support requests, or (v) protect the rights, property or safety of X, its users and the public. We do not disclose personally-identifying information to third parties except in accordance with our Privacy Policy.

Certain services or features may be offered on X for which additional terms and conditions may apply in connection with your use of those services. These additional terms are accessible from our sites and applications dedicated to these services or features. By using or paying for any of these additional services, you will have to agree to any additional terms applicable to those services, and those additional terms will then also become part of our agreement with you. If any of the applicable additional terms conflict with these Terms, the additional terms will prevail while you are using those services to which they apply.

If you use paid features of the Services, you agree to the applicable Terms for Paid Services (https://legal.x.com/purchaser-terms.html).

If you use developer features of the Services, including but not limited to X for Websites (https://developer.x.com/docs/twitter-for-websites), X Cards (https://developer.x.com/docs/twitter-for-websites/cards/overview/abouts-cards), Public API (https://developer.x.com/docs), or Sign in with X (https://developer.x.com/docs/authentication/guides/log-in-with-twitter), you agree to our Developer Agreement (https://developer.x.com/developer-terms/agreement) and Developer Policy (https://developer.x.com/developer-terms/policy). If you want to reproduce, modify, create derivative works, distribute, sell, transfer, publicly display, publicly perform, transmit, or otherwise use the Services or Content on the Services, you must use the interfaces and instructions we provide, except as permitted through the Services, these Terms, or the terms provided on https://developer.x.com/developer-terms. Otherwise, all such actions are strictly prohibited. If you are a security researcher, you are required to comply with the rules of our Vulnerability Reporting Program (https://hackerone.com/x). The requirements set out in the preceding paragraph may not apply to those participating in our Vulnerability Reporting Program.



If you use advertising features of the Services, you agree to our Master Services Agreement (https://ads.x.com/terms).

#### **Your Account**

You may need to create an account to use the Services. You are responsible for safeguarding your account, so use a strong password and limit its use to this account. We cannot and will not be liable for any loss or damage arising from your failure to comply with the above.

You can control most communications from the Services. We may need to provide you with certain communications, such as service announcements and administrative messages. These communications are considered part of the Services and your account, and you may not be able to opt-out from receiving them. If you added your phone number to your account and you later change or deactivate that phone number, you must update your account information to help prevent us from communicating with anyone who acquires your old number.

#### Your License to Use the Services

We give you a personal, worldwide, royalty-free, non-assignable and non-exclusive license to use the software provided to you as part of the Services. This license has the sole purpose of enabling you to use and enjoy the benefit of the Services as provided on X, in the manner permitted by these Terms.

The Services are protected by copyright, trademark, and other laws of both the United States and other countries. Nothing in the Terms gives you a right to use the X name or any of the X trademarks, logos, domain names, other distinctive brand features, and other proprietary rights. All right, title, and interest in and to the Services (excluding Content provided by users) are and will remain our and our licensors' exclusive property. Any feedback, comments, or suggestions you may provide regarding X, or the Services is entirely voluntary and we will be free to use such feedback, comments or suggestions as we see fit and without any obligation to you.

#### Misuse of the Services

You also agree not to misuse the Services, for example, by interfering with them or accessing them using a method other than the interface and the instructions that we provide. You agree that you will not work around any technical limitations in the software provided to you as part of the

Services, or reverse engineer, decompile or disassemble the software, except and only to the extent that applicable law expressly permits. You may not do any of the following while accessing or using the Services: (i) access, tamper with, or use non-public areas of the Services, our computer systems, or the technical delivery systems of our providers; (ii) probe, scan, or test the vulnerability of any system or network or breach or circumvent any security or authentication measures; (iii) access or search or attempt to access or search the Services by any means (automated or otherwise) other than through our currently available, published interfaces that are provided by us (and only pursuant to the applicable terms and conditions), unless you have been specifically allowed to do so in a separate agreement with us (NOTE: crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited); (iv) forge any TCP/IP packet header or any part of the header information in any email or posting, or in any way use the Services to send altered, deceptive or false source-identifying information; (v) engage in any conduct that violates our Platform Manipulation and Spam Policy or any other Rules and Policies; or (vi) interfere with, or disrupt, (or attempt to do so), the access of any user, host or network, including, without limitation, sending a virus, overloading, flooding, spamming, mail-bombing the Services, or by scripting the creation of Content in such a manner as to interfere with or create an undue burden on the Services. It is also a violation of these Terms to facilitate or assist others in violating these Terms, including by distributing products or services that enable or encourage violation of these Terms.

## **Ending These Terms**

You may end your legal agreement with us at any time by deactivating your accounts and discontinuing your use of the Services. See

https://help.x.com/managing-your-account/how-to-deactivate-x-account for instructions on how to deactivate your account and the Privacy Policy for more information on what happens to your information.

We may suspend or terminate your account or cease providing you with all or part of the Services at any time if we reasonably believe: (i) you have violated these Terms or our Rules and Policies; (ii) you create risk or possible legal exposure for us; (iii) your account should be removed due to unlawful conduct; (iv) your account should be removed due to prolonged inactivity; or (v) our provision of the Services to you is no longer commercially viable. We will make reasonable efforts to notify you by the email address associated with your account or the next time you attempt to access your account, depending on the circumstances. In all such cases, the Terms shall terminate, including, without limitation, your license to use the Services, except that the following sections shall continue to apply: 2, 3, 5, 6, and the misuse provisions of Section 4 ("Misuse of the Services"). If you believe your account was terminated in error you can file an appeal following the steps found in our Help Center (https://help.x.com/forms/account-access/appeals). For the

avoidance of doubt, these Terms survive the deactivation or termination of your account.

## 5. Limitations of Liability

By using the Services you agree that X Corp., its parents, affiliates, related companies, officers, directors, employees, agents representatives, partners and licensors, liability is limited to the maximum extent permissible in your country of residence.

#### 6. General

We may revise these Terms from time to time. The changes will not be retroactive, and the most current version of the Terms, which will always be at x.com/tos, will govern our relationship with you. Other than for changes addressing new functions or made for legal reasons, we will notify you 30 days in advance of making effective changes to these Terms that impact the rights or obligations of any party to these Terms, for example via a service notification or an email to the email associated with your account. By continuing to access or use the Services after those revisions become effective, you agree to be bound by the revised Terms.

To the extent permitted by law, you waive the right to participate as a plaintiff or class member in any purported class action, collective action or representative action proceeding.

The X User Agreement is written in English but is made available in multiple languages through translations. X strives to make the translations as accurate as possible to the original English version. However, in case of any discrepancies or inconsistencies, the English language version of the X User Agreement shall take precedence. You acknowledge that English shall be the language of reference for interpreting and constructing the terms of the X User Agreement.

In the event that any provision of these Terms is held to be invalid or unenforceable, then that provision will be limited or eliminated to the minimum extent necessary, and the remaining provisions of these Terms will remain in full force and effect. Our failure to enforce any right or provision of these Terms will not be deemed a waiver of such right or provision.

If you have any questions about these Terms, please contact us.

Effective: September 29, 2023

#### Archive of Previous Terms

Appendix B: X Rules as of 3/31/2024

## The X Rules

X's purpose is to serve the public conversation. Violence, harassment and other similar types of behavior discourage people from expressing themselves, and ultimately diminish the value of global public conversation. Our rules are to ensure all people can participate in the public conversation freely and safely.

## **Safety**

<u>Violent Speech:</u> You may not threaten, incite, glorify, or express desire for violence or harm. <u>Learn more</u>.

<u>Violent & Hateful Entities:</u> You can't affiliate with or promote the activities of violent and hateful entities. <u>Learn more (https://help.x.com/en/rules-and-policies/violent-entities.html).</u>

<u>Child Sexual Exploitation:</u> We have zero tolerance for child sexual exploitation on X. Learn more.

<u>Abuse/Harassment:</u> You may not share abusive content, engage in the targeted harassment of someone, or incite other people to do so. <u>Learn more</u>.

<u>Hateful conduct:</u> You may not attack other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. <u>Learn more</u>.

<u>Perpetrators of Violent Attacks:</u> We will remove any accounts maintained by individual perpetrators of terrorist, violent extremist, or mass violent attacks, and may also remove posts disseminating manifestos or other content produced by perpetrators. <u>Learn more</u>.

<u>Suicide:</u> You may not promote or encourage suicide or self-harm. <u>Learn more.</u>

<u>Sensitive media:</u> You may not post media that is excessively gory or share violent or adult content within live video or in profile or header images. Media depicting sexual violence and/or assault is also not permitted. <u>Learn more</u>.

<u>Illegal or Certain Regulated Goods or Services:</u> You may not use our service for any unlawful purpose or in furtherance of illegal activities. This includes selling, buying, or facilitating transactions in illegal goods or services, as well as certain types of

## **Privacy**

<u>Private Information:</u> You may not publish or post other people's private information (such as home phone number and address) without their express authorization and permission. We also prohibit threatening to expose private information or incentivizing others to do so. <u>Learn more.</u>

<u>Non-Consensual Nudity:</u> You may not post or share intimate photos or videos of someone that were produced or distributed without their consent. <u>Learn more</u>.

<u>Account Compromise:</u> You may not use or attempt to use credentials, passwords, tokens, keys, cookies or other data to log into or otherwise access, add, delete or modify the private information or account features of any X account other than your own (or those you have been directly authorized to do so via X's Teams authorization, OAuth authorization or similar mechanism).

## **Authenticity**

<u>Platform Manipulation and Spam:</u> You may not use X's services in a manner intended to artificially amplify or suppress information or engage in behavior that manipulates or disrupts people's experience on X. <u>Learn more</u>.

<u>Civic Integrity</u>: You may not use X's services for the purpose of manipulating or interfering in elections or other civic processes. This includes posting or sharing content that may suppress participation or mislead people about when, where, or how to participate in a civic process. <u>Learn more</u>.

<u>Misleading and Deceptive Identities:</u> You may not impersonate individuals, groups, or organizations to mislead, confuse, or deceive others, nor use a fake identity in a manner that disrupts the experience of others on X. <u>Learn more</u>.

<u>Synthetic and Manipulated Media:</u> You may not deceptively share synthetic or manipulated media that are likely to cause harm. In addition, we may label posts containing synthetic and manipulated media to help people understand their

authenticity and to provide additional context. Learn more.

<u>Copyright and Trademark:</u> You may not violate others' intellectual property rights, including copyright and trademark. Learn more about our <u>trademark policy</u> and <u>copyright policy</u>.

## Third-party advertising in video content

You may not submit, post, or display any video content on or through our services that includes third-party advertising, such as pre-roll video ads or sponsorship graphics, without our prior consent.

## **Enforcement and Appeals**

Learn more about <u>our approach to enforcement</u>, including potential consequences for violating these rules or attempting to circumvent enforcement, as well as how to appeal.

#### Share this article



© 2024 X Corp.

Cookies (https://help.twitter.com/rules-and-policies/twitter-cookies)

Privacy (https://twitter.com/privacy)

Terms and conditions (https://twitter.com/tos)

English	
	$\neg$
	Help Center (https://help.twitter.com/en)

- English (https://help.twitter.com/en/rules-and-policies/x-rules)
- Español (https://help.twitter.com/es/rules-and-policies/x-rules)
- <u>日本語 (https://help.twitter.com/ja/rules-and-policies/x-rules)</u>
- 한국<u>어 (https://help.twitter.com/ko/rules-and-policies/x-rules)</u>
- <u>Português (https://help.twitter.com/pt/rules-and-policies/x-rules)</u>
- <u>Deutsch (https://help.twitter.com/de/rules-and-policies/x-rules)</u>
- <u>Türkçe (https://help.twitter.com/tr/rules-and-policies/x-rules)</u>
- <u>Français (https://help.twitter.com/fr/rules-and-policies/x-rules)</u>
- Italiano (https://help.twitter.com/it/rules-and-policies/x-rules)
- (https://help.twitter.com/ar/rules-and-policies/x-rules) العربية
- Nederlands (https://help.twitter.com/nl/rules-and-policies/x-rules)
- Bahasa Indonesia (https://help.twitter.com/id/rules-and-policies/x-rules)
- <u>Pyccкий (https://help.twitter.com/ru/rules-and-policies/x-rules)</u>
- हिंदी (https://help.twitter.com/hi/rules-and-policies/x-rules)
- <u>אברית (https://help.twitter.com/he/rules-and-policies/x-rules)</u>
- <u>简体中文 (https://help.twitter.com/zh-cn/rules-and-policies/x-rules)</u>
- <u>繁體中文 (https://help.twitter.com/zh-tw/rules-and-policies/x-rules)</u>
- <u>ภาษาไทย (https://help.twitter.com/th/rules-and-policies/x-rules)</u>
- <u>Tiếng Việt (https://help.twitter.com/vi/rules-and-policies/x-rules)</u>
- Melayu (https://help.twitter.com/ms/rules-and-policies/x-rules)
- <u>Filipino (https://help.twitter.com/fil/rules-and-policies/x-rules)</u>
- <u>المارسي (https://help.twitter.com/fa/rules-and-policies/x-rules)</u>
- <u>Dansk (https://help.twitter.com/da/rules-and-policies/x-rules)</u>
- Suomi (https://help.twitter.com/fi/rules-and-policies/x-rules)
- Svenska (https://help.twitter.com/sv/rules-and-policies/x-rules)
- Norsk (https://help.twitter.com/no/rules-and-policies/x-rules)
- Polski (https://help.twitter.com/pl/rules-and-policies/x-rules)
- Magyar (https://help.twitter.com/hu/rules-and-policies/x-rules)
- Română (https://help.twitter.com/ro/rules-and-policies/x-rules)
- Українська (https://help.twitter.com/uk/rules-and-policies/x-rules)
- Български (https://help.twitter.com/bg/rules-and-policies/x-rules)
- <u>Català (https://help.twitter.com/ca/rules-and-policies/x-rules)</u>
- Hrvatski (https://help.twitter.com/hr/rules-and-policies/x-rules)
- Српски (https://help.twitter.com/sr/rules-and-policies/x-rules)
- Slovenčina (https://help.twitter.com/sk/rules-and-policies/x-rules)
- <u> ಪಾಷ್ಟ್ರೋ (https://help.twitter.com/ps/rules-and-policies/x-rules)</u>
- <u>Dari (https://help.twitter.com/fa-af/rules-and-policies/x-rules)</u>
- Oromo (https://help.twitter.com/om/rules-and-policies/x-rules)
- <u>Tigrinya (https://help.twitter.com/ti/rules-and-policies/x-rules)</u>
- Kurdish (https://help.twitter.com/ckb)

- <u>Lietuvių (https://help.twitter.com/lt/rules-and-policies/x-rules)</u>
- Latviešu (https://help.twitter.com/lv/rules-and-policies/x-rules)
- Malti (https://help.twitter.com/mt/rules-and-policies/x-rules)
- <u>Slovenščina (https://help.twitter.com/sl/rules-and-policies/x-rules)</u>
- <u>Gaeilge (https://help.twitter.com/ga/rules-and-policies/x-rules)</u>
- <u>Lus Hmoob (https://help.twitter.com/hmn/rules-and-policies/x-rules)</u>
- <u>Հայերեն (https://help.twitter.com/hy/rules-and-policies/x-rules)</u>
- <u>\$\frac{1}{2}\$ (https://help.twitter.com/km/rules-and-policies/x-rules)</u>

## Appendix C: X Policies as of 3/31/2024

# **Violent Speech Policy**

#### **Overview**

#### October 2023

You may not threaten, incite, glorify, or express desire for violence or harm.

X is a place where people can express themselves, learn about what's happening, and debate global issues. However, healthy conversations can't thrive when violent speech is used to deliver a message. As a result, we may remove or reduce the visibility of violent speech in order to ensure the safety of our users and prevent the normalization of violent actions.

## What is in violation of this policy?

#### **Violent Threats**

You may not threaten to inflict physical harm on others, which includes (but is not limited to) threatening to kill, torture, sexually assault, or otherwise hurt someone. This also includes threatening to damage civilian homes and shelters, or infrastructure that is essential to daily, civic, or business activities.

#### Wish of Harm

You may not wish, hope, or express desire for harm. This includes (but is not limited to) hoping for others to die, suffer illnesses, tragic incidents, or experience other physically harmful consequences.

#### **Incitement of Violence**

You may not incite, promote, or encourage others to commit acts of violence or harm, which includes (but is not limited to) encouraging others to hurt themselves or inciting others to commit atrocity crimes including crimes against humanity, war crimes or genocide. This also includes using coded language (often referred to as "dog whistles") to indirectly incite violence.

#### **Glorification of Violence**

You may not glorify, praise, or celebrate acts of violence where harm occurred, which includes (but is not limited to) expressing gratitude that someone experienced physical harm or praising <u>Violent entities (https://help.twitter.com/rules-and-policies/violent-entities)</u> and <u>Perpetrators of Violent Attacks (https://help.twitter.com/rules-and-policies/perpetrators-of-violent-attacks)</u>. This also includes glorifying animal abuse or cruelty.

## What is not a violation of this policy?

We make sure to evaluate and understand the context behind the conversation before taking action, and we allow expressions of violent speech when there is no clear abusive or violent context, such as (but not limited to) hyperbolic and consensual speech between friends, or during discussion of video games and sporting events.

We also allow certain cases of figures of speech, satire, or artistic expression when the context is expressing a viewpoint rather than instigating actionable violence or harm.

## What happens if you violate this policy?

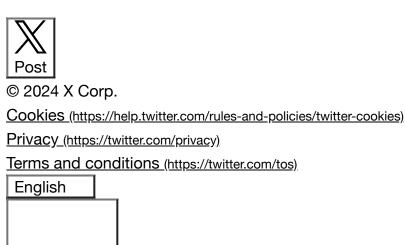
In certain cases, we will immediately and permanently suspend any account that violates this policy. However, for most violations, we may instead temporarily lock you out of your account before you can Post again. In rare cases, we may make the violative content less visible by <u>restricting its reach (https://help.twitter.com/rules-and-policies/enforcement-options)</u> on X. However, if you continue to violate this policy after receiving a warning, your account may be permanently suspended.

We also recognize that conversations regarding certain individuals credibly accused of severe violence may prompt outrage and associated violent speech. In these limited cases, we may take less punitive measures.

Learn more about <u>our range of enforcement options</u>. If someone believes we took an enforcement action in error, they can <u>submit an appeal</u>

(https://help.twitter.com/forms/account-access/appeals.html).

#### Share this article



Help Center (https://help.twitter.com/en)

# Violent and hateful entities policy

### **Overview**

#### **April 2023**

There is no place on X for violent and hateful entities, including (but not limited to) terrorist organizations, violent extremist groups, <u>perpetrators of violent attacks</u> (<a href="https://help.twitter.com/rules-and-policies/perpetrators-of-violent-attacks">https://help.twitter.com/rules-and-policies/perpetrators-of-violent-attacks</a>), or individuals who affiliate with and promote their illicit activities. The violence and hate these entities engage in and/or promote jeopardizes the physical safety of those targeted.

You may not threaten terrorism and/or violent extremism, nor promote violent and hateful entities.

Violent entities are those that deliberately target humans or essential infrastructure with physical violence and/or violent rhetoric as a means to further their cause. These include, but are not limited to, terrorist organizations, violent extremist groups, and <u>perpetrators of violent attacks (https://help.twitter.com/rules-and-policies/perpetrators-of-violent-attacks)</u>.

Hateful entities are those that have systematically and intentionally promoted, supported and/or advocated for <a href="https://help.twitter.com/rules-and-policies/hateful-conduct-policy">hateful-conduct-policy</a>), which includes promoting violence or engaging in targeted harassment towards a protected category.

## What is in violation of this policy?

Under this policy, you can't affiliate with or promote the activities of violent and hateful entities. Examples of the types of content that violate this policy include, but are not limited to, doing the following on behalf of, indirectly, or directly for a violent or hateful entity:

- Engaging in or promoting violent acts
- Recruiting, or providing or distributing services (such as media/propaganda) to further stated goals

## What is not a violation of this policy?

We may make limited exceptions for violent and hateful entities if we can determine the following:

- They have reformed or denounced their violence and/or hate-based purpose.
- They are currently engaged in a peaceful resolution process.
- They are state or governmental entities, including those that have representatives elected to public office.

Additionally, any discussions of violent and hateful entities for clearly educational, documentary, and/or newsworthy purposes is not a violation of this policy.

## Who can report violations of this policy?

Anyone can report potential violations of this policy, whether they have an X account or not.

## How can I report violations of this policy?

#### In-app

You can report this content for review in-app on iOS as follows:

- 1. Select **Report Account** from the ••• icon.
- Select Violent & hateful entities.
- 3. Submit your report.

#### Desktop

You can report this content for review via desktop as follows:

- 1. Select **Report Account** from the ••• icon.
- 2. Select Violent & hateful entities.
- 3. Submit your report.

## What happens if you violate this policy?

We will immediately and permanently suspend any account that we determine to be in violation of this policy. If you believe that your account was suspended in error, you can submit an appeal (https://help.twitter.com/forms/general?subtopic=suspended).

### **Additional resources**

Learn more about <u>our range of enforcement options (https://help.twitter.com/rules-and-policies/enforcement-options)</u> and our approach to <u>policy development and enforcement (https://help.twitter.com/rules-and-policies/enforcement-philosophy</u>).

To learn more about violations related to <u>specific threats of violence</u> (<a href="https://help.twitter.com/rules-and-policies/violent-threats-glorification">https://help.twitter.com/rules-and-policies/violent-threats-glorification</a>) or incidents involving wishing for the serious physical harm, death, or disease of an individual or group of people, please see our <u>abusive behavior</u> (<a href="https://help.twitter.com/rules-and-policies/abusive-behavior">https://help.twitter.com/rules-and-policies/abusive-behavior</a>) policy.

#### Share this article

X	
Post	
<u> </u>	-

© 2024 X Corp.

Cookies (https://help.twitter.com/rules-and-policies/twitter-cookies)

Privacy (https://twitter.com/privacy)

Terms and conditions (https://twitter.com/tos)

English	
	Help Center (https://help.twitter.com/en)

# Child sexual exploitation policy

#### **Overview**

October 2020

We have a zero-tolerance child sexual exploitation policy on X.

X has zero tolerance towards any material that features or promotes child sexual exploitation, one of the most serious violations of the <u>Our Rules</u> (<a href="https://help.twitter.com/rules-and-policies/twitter-rules">https://help.twitter.com/rules-and-policies/twitter-rules</a>). This may include media, text, illustrated, or computer-generated images. Regardless of the intent, viewing, sharing, or linking to child sexual exploitation material contributes to the revictimization of the depicted children. This also applies to content that may further contribute to victimization of children through the promotion or glorification of child sexual exploitation. For the purposes of this policy, a minor is any person under the age of 18.

## What is in violation of this policy?

Any content that depicts or promotes child sexual exploitation including, but not limited to:

- visual depictions of a child engaging in sexually explicit or sexually suggestive acts;
- illustrated, computer-generated or other forms of realistic depictions of a human child in a sexually explicit context, or engaging in sexually explicit acts;
- · sexualized commentaries about or directed at a known or unknown minor; and
- links to third-party sites that host child sexual exploitation material.

The following behaviors are also not permitted:

- sharing fantasies about or promoting engagement in child sexual exploitation;
- expressing a desire to obtain materials that feature child sexual exploitation;
- recruiting, advertising or expressing an interest in a commercial sex act involving a child, or in harboring and/or transporting a child for sexual purposes;
- sending sexually explicit media to a child;
- engaging or trying to engage a child in a sexually explicit conversation;
- trying to obtain sexually explicit media from a child or trying to engage a child in sexual activity through blackmail or other incentives;
- identifying alleged victims of childhood sexual exploitation by name or image;
   and
- promoting or normalizing sexual attraction to minors as a form of identity or sexual orientation.

## What is not a violation of this policy?

- Discussions related to child sexual exploitation are permitted, provided they don't normalise, promote or glorify child sexual exploitation **in any way**;
- Conversations about help-seeking behaviour of individuals who may struggle with an attraction to minors;
- Advocacy against illegal or harmful activity involving minors provided there is no sharing of, or linking to any material featuring child sexual exploitation; and
- Depictions of nude minors in a non-sexualized context or setting may be permitted in a limited number of scenarios (e.g., art works by internationally renowned artists that feature minors, news media reportage, media created and shared for scientific or educational purposes).

## Who can report violations of this policy?

Anyone can report potential violations of this policy, whether they have an X account or not.

## How can I report violations of this policy?

If you think you've found an X account distributing or promoting child sexual exploitation, you can report it via our child sexual exploitation form <a href="https://help.twitter.com/forms/cse">(https://help.twitter.com/forms/cse</a>). Provide the username and links to all relevant Posts that led you to believe the account should be reviewed.

## What happens if you violate this policy?

In the majority of cases, the consequence for violating our child sexual exploitation policy is **immediate and permanent suspension**. In addition, violators will be prohibited from creating any new accounts in the future. Note: when we're made aware of content depicting or promoting child sexual exploitation, including links to third party sites where this content can be accessed, they will be removed without further notice and reported to the National Center for Missing & Exploited Children (NCMEC).

In a limited number of situations, where we haven't identified any malicious intent, we will require you to remove this content. We will also temporarily lock you out of your account before you can Post again. Further violations will lead to your account being permanently suspended. If you believe that your account was suspended in error, you can submit an appeal. (https://help.twitter.com/forms/general?subtopic=suspended)

### **Additional resources**

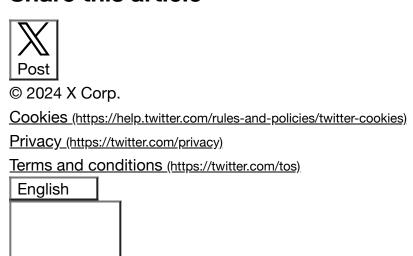
Learn more about our <u>range of enforcement options</u> (https://help.twitter.com/rules-and-policies/enforcement-options) and our approach to <u>policy development and enforcement</u> (https://help.twitter.com/rules-and-policies/enforcement-philosophy).

If you've found content elsewhere on the internet that is distributing or promoting child sexual exploitation, you should <u>file a report with NCMEC</u> (<a href="http://www.missingkids.org/gethelpnow/cybertipline">http://www.missingkids.org/gethelpnow/cybertipline</a>), or with your local law enforcement. If you believe the content is hosted on a website outside of the United States, you can report it on the <a href="https://www.inhope.org/EN#hotlineReferral">International Association of Internet Hotlines website</a> (<a href="https://www.inhope.org/EN#hotlineReferral">https://www.inhope.org/EN#hotlineReferral</a>).

We partner with multiple organisations whose work is dedicated to fighting child sexual exploitation around the world. Please see the full list of <u>our child protection partners (https://about.twitter.com/safety/safety-partners.html#child-protection-partners)</u>.

Help Center (https://help.twitter.com/en)

#### Share this article



## **Abuse and Harassment**

# Overview March 2024

You may not target others with abuse or harassment, or encourage other people to do so.

X's mission is to give everyone the power to create and share ideas and information, as well as express their opinions and beliefs without barriers. Free expression is a human right – we believe that everyone has a voice, and the right to use it. Our role is to serve the public conversation, which requires representation of a diverse range of perspectives.

We recognize that if anyone, regardless of background, experiences harassment on X, it can jeopardize their ability to express themselves and cause harm. To facilitate healthy dialogue on the platform, and empower individuals to express diverse opinions and beliefs, we prohibit behavior and content that harasses, shames, or degrades others. In addition to posing risks to people's safety, these types of behavior may also lead to physical and emotional hardship for those affected.

## What is in violation of this policy?

We prohibit behavior and content that harasses, shames, or degrades others, as defined below. Please note that, in order to help our teams understand the context, we sometimes need to hear directly from the person being targeted to ensure that we have the necessary information prior to taking appropriate and proportionate enforcement action.

## **Targeted Harassment**

We prohibit the malicious, unreciprocated targeting (such as mentioning or tagging) of individual(s), particularly when shared to humiliate or degrade someone. This can mean:

- Sharing multiple Posts, over a short period of time, or continuously posting replies with malicious content, to target an individual. This includes accounts dedicated to harassing an individual or multiple individuals.
- Mentioning or tagging users with malicious content.

#### **Violent Event Denial**

We prohibit content that denies that mass murder or other mass casualty events took place, where we can verify that the event occurred, and when the content is shared with abusive context. This may include references to such an event as a "hoax" or claims that victims or survivors are fake or "actors." It includes, but is not limited to, events like the Holocaust, school shootings, terrorist attacks, and natural disasters.

#### Incitement of Harassment

We prohibit behavior that encourages others to harass or target specific individuals or groups of people with abuse. This includes, but is not limited to: calls to target people with abuse or harassment online and behavior that urges offline action, such as physical harassment.

# Unwanted Sexual Content & Graphic Objectification

While some <u>consensual nudity and adult content is permitted</u>
(<a href="https://help.twitter.com/rules-and-policies/media-policy">https://help.twitter.com/rules-and-policies/media-policy</a>) on X, we prohibit unwanted sexual conduct and graphic objectification that sexually objectifies an individual without their consent. This includes, but is not limited to:

- sending someone unsolicited and/or unwanted adult media (images, videos, and GIFs)
- · unwanted sexual discussion of someone's body
- solicitation of sexual acts
- any other content that otherwise sexualizes an individual without their consent.

### **Insults**

We take action against the use of insults or profanity to target others. However, while some individuals may find certain terms to be offensive, we will not take action against every instance where insulting terms are used.

### **Use of Prior Names and Pronouns**

Where required by local laws, we will reduce the visibility of posts that purposefully use different pronouns to address someone other than what that person uses for themselves, or that use a previous name that someone no longer goes by as part of their transition. Given the complexity of determining whether such a violation has occurred, we must always hear from the target to determine if a violation has occurred.

## What is not in violation of this policy?

Some posts may appear to be harmful when viewed in isolation, but may not be when viewed in the context of a larger conversation. For example, friends may consensually use certain terms or phrases to engage with each other that could appear abusive without this context. We also recognise that our platform can be used to call attention to, condemn, or highlight others for harmful behaviors. In such cases, we do not take action where the context is clearly non-abusive and intended to counter these kinds of rhetoric.

We also believe that criticism of institutions, practices and ideas is a fundamental part of the freedom of expression and thus we will not take action on such critical commentary.

# Who can report violations of this policy?

Anyone can report violations of this policy using our dedicated reporting flow. However, we sometimes need to hear directly from the person being targeted to ensure that we have the information needed prior to taking any enforcement action.

## What happens if you violate this policy?

When determining the penalty for violating this policy, we consider a number of factors including, but not limited to the severity of the violation, whether someone has been targeted (such as being mentioned, referred to by full name, referenced with a photo, etc), and an individual's previous record of rule violations. The following is a list of potential enforcement options for content that violates this policy:

- Making content less visible on X by:
  - Removing the post from search results, in-product recommendations, trends, notifications, and home timelines
  - Restricting the post discoverability to the author's profile
  - Downranking the post in replies
  - Restricting likes, replies, reposts, quote, bookmarks, share, pin to profile, or engagement counts
  - Excluding the post from having ads adjacent to it
- Excluding posts and/or accounts in email or in-product recommendations.
- · Requiring post removal.
  - For example, we may ask someone to remove the violating content and serve a period of time in read-only mode before they can post again.
     Subsequent violations may lead to account suspension.
- Suspending accounts whose sole purpose is to violate our Unwanted Sexual Content & Graphic Objectification policy, or accounts that are dedicated to harassing individuals.

To learn more, please see our <u>range of enforcement options (https://help.twitter.com/rules-and-policies/enforcement-options)</u>, and if someone believes their account was enforced against in error, they can <u>submit an appeal (https://help.twitter.com/forms/account-access/appeals)</u>.

#### **Share this article**



© 2024 X Corp.

Cookies (https://help.twitter.com/rules-and-policies/twitter-cookies)

Privacy (https://twitter.com/privacy)

Terms and conditions (https://twitter.com/tos)

English

## **Hateful Conduct**

#### **Overview**

#### **April 2023**

You may not directly attack other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease.

X's mission is to give everyone the power to create and share ideas and information, and to express their opinions and beliefs without barriers. Free expression is a human right – we believe that everyone has a voice, and the right to use it. Our role is to serve the public conversation, which requires representation of a diverse range of perspectives.

We recognize that if people experience abuse on X, it can jeopardize their ability to express themselves. Research has shown that some groups of people are disproportionately targeted with abuse online. For those who identify with multiple underrepresented groups, abuse may be more common, more severe in nature, and more harmful.

We are committed to combating abuse motivated by hatred, prejudice or intolerance, particularly abuse that seeks to silence the voices of those who have been historically marginalized. For this reason, we prohibit behavior that targets individuals or groups with abuse based on their perceived membership in a protected category.

If you see something on X that you believe violates this policy, please report it to us.

## What is in violation of this policy?

We will review and take action against reports of accounts targeting an individual or group of people with any of the following behavior, whether within Posts or Direct Messages.

#### Hateful references

We prohibit targeting individuals or groups with content that references forms of violence or violent events where a protected category was the primary target or victims, where the intent is to harass. This includes, but is not limited to media or text that refers to or depicts:

- genocides, (e.g., the Holocaust);
- lynchings.

#### Incitement

We prohibit inciting behavior that targets individuals or groups of people belonging to protected categories. This includes:

- inciting fear or spreading fearful stereotypes about a protected category, including asserting that members of a protected category are more likely to take part in dangerous or illegal activities, e.g., "all [religious group] are terrorists."
- inciting others to harass members of a protected category on or off platform,
   e.g., "I'm sick of these [religious group] thinking they are better than us, if any of you see someone wearing a [religious symbol of the religious group], grab it off them and post pics!"
- inciting others to discriminate in the form of denial of support to the economic
  enterprise of an individual or group because of their perceived membership in
  a protected category, e.g., "If you go to a [religious group] store, you are
  supporting those [slur], let's stop giving our money to these [religious slur]."
  This may not include content intended as political in nature, such as political
  commentary or content relating to boycotts or protests.

Note: content intended to incite violence against a protected category is prohibited under <u>Violent Speech</u>.

#### **Slurs and Tropes**

We prohibit targeting others with repeated slurs, tropes or other content that intends to degrade or reinforce negative or harmful stereotypes about a protected category. In some cases, such as (but not limited to) severe, repetitive usage of slurs, or racist/sexist tropes where the context is to harass or intimidate others, we may

require Post removal. In other cases, such as (but not limited to) moderate, isolated usage where the context is to harass or intimidate others, we may limit Post visibility as further described below.

#### **Dehumanization**

We prohibit the dehumanization of a group of people based on their religion, caste, age, disability, serious disease, national origin, race, ethnicity, gender, gender identity, or sexual orientation.

#### **Hateful Imagery**

We consider hateful imagery to be logos, symbols, or images whose purpose is to promote hostility and malice against others based on their race, religion, disability, sexual orientation, gender identity or ethnicity/national origin. Some examples of hateful imagery include, but are not limited to:

- symbols historically associated with hate groups, e.g., the Nazi swastika;
- images depicting others as less than human, or altered to include hateful symbols, e.g., altering images of individuals to include animalistic features; or
- images altered to include hateful symbols or references to a mass murder that targeted a protected category, e.g., manipulating images of individuals to include yellow Star of David badges, in reference to the Holocaust.

Media depicting hateful imagery is not permitted within live video, account bio, profile or header images. All other instances must be marked as sensitive media. Additionally, sending an individual unsolicited hateful imagery is a violation of this policy.

#### **Hateful Profile**

You may not use hateful images or symbols in your profile image or profile header. You also may not use your username, display name, or profile bio to engage in abusive behavior, such as targeted harassment or expressing hate towards a person, group, or protected category.

# Do I need to be the target of this content for it to be a violation of the X Rules?

Some Posts may appear to be hateful when viewed in isolation, but may not be when viewed in the context of a larger conversation. For example, members of a protected category may refer to each other using terms that are typically considered as slurs. When used consensually, the context behind these terms is not abusive, but a means to reclaim terms that were historically used to demean individuals.

When we review this type of content, it may not be clear whether the context is to abuse an individual on the basis of their protected status, or if it is part of a consensual conversation. To help our teams understand the context, we sometimes need to hear directly from the person being targeted to ensure that we have the information needed prior to taking any enforcement action.

**Note:** individuals do not need to be a member of a specific protected category for us to take action. We will never ask people to prove or disprove membership in any protected category and we will not investigate this information.

## What happens if you violate this policy?

Under this policy, we take action against behavior that targets individuals or an entire protected category with hateful conduct, as described above. Targeting can happen in a number of ways, for example, mentions, including a photo of an individual, referring to someone by their full name, etc.

When determining the penalty for violating this policy, we consider a number of factors including, but not limited to the severity of the violation and an individual's previous record of rule violations. The following is a list of potential enforcement options for content that violates this policy:

- Making content less visible on X by:
  - Removing the Post from search results, in-product recommendations, trends, notifications, and home timelines
  - Restricting the Post discoverability to the author's profile
  - Downranking the Post in replies
  - Restricting Likes, replies, Reposts, Quote, bookmarks, share, pin to profile, or engagement counts
  - Excluding the Post from having ads adjacent to it
- Excluding Posts and/or accounts in email or in-product recommendations.
- Requiring Post removal.
  - For example, we may ask someone to remove the violating content and serve a period of time in read-only mode before they can Post again.
- Suspending accounts that violate our Hateful Profile policy.

Learn more about our range of enforcement options.

If someone believes their account was suspended in error, they can <u>submit an appeal</u>.

#### **Share this article**

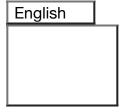


© 2024 X Corp.

Cookies (https://help.twitter.com/rules-and-policies/twitter-cookies)

Privacy (https://twitter.com/privacy)

Terms and conditions (https://twitter.com/tos)



Help Center (https://help.twitter.com/en)

# **Perpetrators of Violent Attacks**

Overview

#### February 2023

We will remove any accounts maintained by individual perpetrators of terrorist, violent extremist, or mass violent attacks, as well as any accounts glorifying the perpetrator(s), or dedicated to sharing manifestos and/or third party links where related content is hosted. We may also remove Posts disseminating manifestos or other content produced by perpetrators.

We want X to be a place where people can find reliable information and express themselves freely and safely without feeling burdened by unhealthy content. In the aftermath of terrorist, violent extremist and mass violent attacks, we know many want to express compassion for victims, condemn the attacks and/or the perpetrators, and discuss how these incidents impact people and their communities. Some might also wish to share manifestos or other similar content produced by the attack's apparent perpetrator or an accomplice, either to express outrage or condemnation of the perpetrator's possible motives.

The violence that these perpetrators engage in, and the rationale they outline in manifestos or other means that normalizes hatred and discrimination, may further jeopardize the physical safety and well-being of those who are targeted, and has the potential to inspire future attacks. Exposure to these materials may also cause harm to those that view them.

We believe the hateful and discriminatory views promoted in content produced by perpetrators are harmful for society and their dissemination should be limited in order to prevent perpetrators from publicizing their message. As a result we may remove Posts that include manifestos or other similar material produced by perpetrators, even if the context is not abusive. However, we may allow newsworthy content if it does not:

- Convey suggestions about how to arm oneself and choose targets;
- Share hateful slogans, symbols, memes, and/or hateful conspiracy theories;
- Outline the perpetrator's ideology, tactical choices, and/or plan of attack.

We define a manifesto as a statement by a perpetrator outlining their motivation, views, or intent to engage in a violent attack. A manifesto can be in the form of a written document, social media post, audio recording, video, external link, or letter or other forms of content. It may be shared in the aftermath, or at any period before a violent attack. A manifesto can be linked to the event through a statement of warning or intent.

# Who are the individual perpetrators of terrorist, violent extremist, and mass violent attacks?

Violent attacks claimed by a violent organization or by a member of such organizations are covered under our <u>Violent and Hateful Entitites policy</u>. We do not require that a person have been confirmed as members of terrorist organizations or other violent and hateful entities, nor that they have any official affiliation with any group, organization, or ideology, for us to enforce on content under this aspect of our policies.

What is in violation of this policy?

Under this policy, we will permanently suspend accounts belonging to apparent individual perpetrators of terrorist, violent extremist, and mass violent attacks, as well as accounts dedicated to the sharing of harmful and violent content associated with the perpetrators or the violent attack.

Additionally, you may not Post any of the following:

#### Manifestos and other content created by perpetrators

We may remove content containing manifestos and other content created by individual perpetrators or their accomplices. Violations can occur via Posts, Spaces, images, and videos, including live video. Examples of content that we may action under this policy include, but are not limited to:

- Sharing or linking full-length manifestos, regardless of context
- Any Posts containing excerpt(s) of a reproduced manifesto (edited or sliced) or
  of the original manifesto, except when shared in a newsworthy context.
  - Regardless of context, it is always prohibited to share any of the following:
    - Excerpt(s) that convey suggestions about how to arm oneself and choose targets.
    - Excerpt(s) that share hateful slogans, symbols, memes, and/or hateful conspiracy theories such as the Great Replacement Theory.
    - Excerpt(s) outlining the perpetrator's ideology, tactical choices, and/or plan of attack.
- Sharing perpetrator-generated media related to the attack in all instances. This
  includes, but is not limited to:
  - Media filmed by the perpetrator of the attack
  - Memes, stickers, or graffiti shared and/or produced by the perpetrator
  - Media of the weapons utilized in the attack
- Other perpetrator-generated content that is not associated with the violent attack but that is specific to the promotion or support of violence.
- Bystander-generated content of the attack as the attack is taking place, such as content that displays a moment of the assault or death, dead bodies, content that identifies victims, or content that depicts the perpetrator(s) conducting the attack.

#### Multimedia content, URLs, and hashtags

We may put a sensitive media interstitial over some media. The interstitial places images and videos behind a warning message that needs to be acknowledged before the media can be viewed. Using this feature means that people who don't want to see sensitive media can avoid it, or make an informed decision before they choose to view it. The types of media we may interstitial include, but are not limited to:

- Media depicting a perpetrator
- Excerpts of the attacker's manifesto that are shared as part of a news report

We may label URLs linking directly to documents believed to be a manifesto to prevent these URLs from being shared on X.

We may denylist hashtags identifying perpetrators on trends to reduce the visibility of perpetrators' identities, as well as all hashtags that are used uniquely to surface manifestos.

What is not a violation of this policy?

- · Accounts that belong to
  - Bystanders who happened to be close to the violent attack and/or managed to stop the attack, for example someone who shot the perpetrator(s)
  - o Perpetrators whose convictions were overturned after a not guilty verdict
- Content depicting any of the following:
  - The use of force by law enforcement and military personnel, as part of the scope of their official functions, that resulted in casualties;
  - Violence against military personnel and law enforcement;
  - Human rights abuses
  - Violent attacks that are part of an armed conflict
  - Violence used as a means of self defense
  - Violent attacks, where there is probable doubt about the intent behind the attack
  - Cases of vandalism and attacks resulting in damage to essential infrastructure
  - Sharing a few sentences from the manifesto(s), either in a newsworthy context or to bring more awareness to the rise of hate or violence towards a specific group. For example: "the shooter said he wants to eradicate x group" is not actionable under this policy.

Who can report violations of this policy?

Anyone can report potential violations of this policy, whether they have an X account or not.

What happens if I violate this policy?

The consequences for violating our violent events policy depends on the severity of the violation. Accounts maintained by perpetrators of terrorist, violent extremist, or mass violent attacks will be permanently suspended. As described above, we may also remove content containing manifestos and other content created by perpetrators or their accomplices.

In addition, we will also remove content that violates our policies regarding <u>Violent Speech</u> or <u>other parts of Our Rules</u>.

Additional resources

Learn more about our <u>range of enforcement options</u> and our approach to <u>policy</u> <u>development and enforcement</u>.

#### **Share this article**



© 2024 X Corp.

Cookies (https://help.twitter.com/rules-and-policies/twitter-cookies)

Privacy (https://twitter.com/privacy)

Terms and conditions (https://twitter.com/tos)

4	
English	
	Help Center (https://help.twitter.com/en)

# Suicide and Self-harm policy

Overview

#### You may not promote or encourage suicide or self-harm.

At X, we recognize that suicide and self-harm are significant social & public health challenges that require collaboration between all stakeholders – public, private, and civil society – and that we have a role and responsibility to help people access and receive support when they need it.

When developing this policy, we consulted extensively with experts to ensure that people who have engaged in self-harm or experienced suicidal thoughts can share their personal experiences. We also recognized the need to protect people from the potential harm caused by exposure to content that could promote or encourage self-harm – intentionally or inadvertently. That's why our policy prohibits content that promotes or encourages self-harming behaviors and provides support to those undergoing experiences with self-harm or suicidal thoughts.

## What is in violation of this policy?

Under this policy, you can't promote, or otherwise encourage, suicide or self-harm. We define promotion and encouragement to include statements such as "the most effective", "the easiest", "the best", "the most successful", "you should", "why don't you". Violations of this policy can occur via Posts, images or videos, including live video.

We define suicide to be the act of taking one's own life. We define self-harm to include:

- self-inflicted physical injuries e.g., cutting; and
- eating disorders e.g., bulimia, anorexia.

Violations of this policy include, but are not limited to:

- · encouraging someone to physically harm or kill themselves;
- asking others for encouragement to engage in self-harm or suicide, including seeking partners for group suicides or suicide games; and
- sharing information, strategies, methods or instructions that would assist people to engage in self-harm and suicide.

## What is not a violation of this policy?

Some examples of behavior that are not considered a violation of this policy include:

- sharing personal stories and experiences related to self-harm or suicide;
- sharing coping mechanisms and resources for addressing self-harm or suicidal thoughts; and
- discussions that are focused on research, advocacy, and education related to self-harm or suicide prevention.

**Note**: people can share their personal experiences, but should avoid sharing detailed information about specific strategies or methods related to self-harm, as this could inadvertently encourage this behavior.

## Who can report violations of this policy?

Anyone can report content that may encourage or promote suicide or self-harm via our in-app reporting or our <u>specialized reporting form</u>
<a href="mailto:(https://help.twitter.com/forms/suicide">(https://help.twitter.com/forms/suicide</a>). These reports are routed to a dedicated team who evaluate each case individually.

Note: if we receive a report that someone has expressed an intention to engage in self-harm or suicide, we will contact them directly, encourage them to seek support, and provide information about dedicated online and hotline resources. We may also

work with law enforcement officials where appropriate, for example, if we receive a valid emergency disclosure request as defined in our <u>Law Enforcement Guidelines</u> (https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support).

## How to report violations of this policy

To ensure that we handle reports sensitively, our in-app reporting provides separate options for people who may be expressing an intention to harm themselves, and content that is encouraging or promoting self-harm or suicide.

## **Expressing intentions of self-harm or suicide**

#### In-App

You can report content for review in-app as follows:

- 1. Select Report Post from drop-down menu
- 2. Select "It expresses intentions of self-harm or suicide"
- 3. Submit your report

#### Desktop

You can report this content for review on desktop as follows:

- Select Report Post from drop-down menu
- 2. Select "It expresses intentions of self-harm or suicide"
- 3. Submit your report

#### Reporting form

You can also report this content for review via <u>our dedicated report form</u> (<u>https://help.twitter.com/forms/report\_self\_harm</u>).

#### In-App

You can report content for review in-app as follows:

- 1. Select Report Post from drop-down menu
- 2. Select "It expresses intentions of self-harm or suicide"
- 3. Submit your report

## **Encouraging self-harm or suicide**

#### In-App

You can report content for review in-app as follows:

- 1. Select Report Post from drop-down menu
- 2. Select "It's abusive or harmful"
- 3. Select "They're encouraging self-harm or suicide"
- 4. Submit your report

#### **Desktop**

You can report this content for review on desktop as follows:

- 1. Select Report Post from drop-down menu
- 2. Select "It's abusive or harmful"
- 3. Select "They're encouraging self-harm or suicide"
- 4. Submit your report

## What happens if you violate this policy?

Our enforcement approach depends on the type of content being shared, whether or not the reported account is encouraging or promoting self-harm or suicide, and the account's previous history of violations.

If you violate this policy by sharing content that intentionally encourages others to harm themselves, ask others to encourage you to harm yourself, or share detailed information or instructions related to self-harm or suicide methods, we will require you to remove this content. We will also temporarily lock you out of your account before you can Post again. If you continue to violate this policy, or if your account is dedicated to promoting or encouraging self-harm or suicide, your account will be permanently suspended. If cases include images or videos related to self-harm or suicide, we will also evaluate this content under our sensitive media policy (<a href="https://help.twitter.com/rules-and-policies/media-policy">https://help.twitter.com/rules-and-policies/media-policy</a>). If you believe that your account was suspended in error, you can submit an appeal (<a href="https://help.twitter.com/forms/general?subtopic=suspended">https://help.twitter.com/forms/general?subtopic=suspended</a>).

We may also take steps to prevent the spread of instructional material hosted on third-party websites by <u>marking such links as unsafe (https://help.twitter.com/en/safety-and-security/phishing-spam-and-malware-links)</u>.

## Additional resources

Learn more about our <u>range of enforcement options (https://help.twitter.com/rules-and-policies/enforcement-options)</u> and our approach to <u>policy development and enforcement (https://help.twitter.com/rules-and-policies/enforcement-philosophy)</u>.

<u>Learn more (https://help.twitter.com/en/safety-and-security/self-harm-and-suicide)</u> about how you can support someone experiencing thoughts of self-harm or suicide.

Visit our Safety Center for a list of local <u>mental health resources</u>
<a href="mailto:(https://about.twitter.com/en\_us/safety/safety-partners.html#mental-health">(https://about.twitter.com/en\_us/safety/safety-partners.html#mental-health</a>) and read <u>our blog on suicide prevention (https://blog.twitter.com/en\_sea/topics/events/2019/Supporting-suicide-prevention-across-the-world.html)</u> to learn more about our work.

#### Share this article



© 2024 X Corp.

Cookies (https://help.twitter.com/rules-and-policies/twitter-cookies)

Privacy (https://twitter.com/privacy)

Terms and conditions (https://twitter.com/tos)

	·
English	
	Help Center (https://help.twitter.com/en)

## **Sensitive Media**

#### **Overview**

#### January 2024

You may post content that falls under our definitions of Graphic Content, Adult Nudity, and Sexual Behavior with a content warning, but you may not share this media in a live video, or in your profile picture or header, List banner, or Community cover photo. Media depicting Gratutitous Gore, Violent Sexual Conduct, or Bestiality and Necrophilia is not permitted.

People use X to show what's happening in the world, often sharing images and videos as part of the conversation. Sometimes, this media can depict sensitive topics, including graphic content, adult nudity, and sexual behavior. We recognize that some people may not want to be exposed to sensitive content, which is why we balance allowing people to share this type of media with helping people who want to avoid it to do so.

For this reason, you can't include graphic content, adult nudity, or sexual behavior within areas that are highly visible on X, including in live video, your profile picture or header, List banners, or Community cover photos. If you share this content on X, you need to mark (https://help.twitter.com/en/rules-and-policies/media-settings.html) your media or your account as sensitive. Doing so places images and videos behind a content warning that needs to be acknowledged before your media can be viewed. Using this feature means that people who don't want to see sensitive media can avoid it, or make an informed decision before they choose to view it. We also restrict graphic media, adult nudity, and sexual behavior for viewers who are under 18 or viewers who do not include a birth date on their profile. Learn more about age restricted content <u>here (https://help.twitter.com/en/rules-and-policies/notices-on-twitter.html)</u>. Beginning January 2024, you may begin to see new media content warnings on posts that X has designated as Graphic (containing violent or hateful imagery) or containing Adult media (adult nudity and sexual behavior). When these new content warnings are available for you to use, please be sure to continue marking your sensitive media accordingly.

Under this policy, there are also some types of sensitive media that we don't allow at all, because they have the potential to normalize violence and cause distress to those who view them.

We group sensitive media content into the following categories:

#### **Graphic Content**

Graphic Content is any media that depicts death, violence, medical procedures, or serious physical injury in graphic detail. Some examples include, but are not limited to, depictions of:

- · violent crimes or accidents;
- · physical fights;
- physical child abuse;
- bodily fluids including blood, feces, semen etc.;
- serious physical harm, including visible wounds; and
- severely injured or mutilated animals.

Note: exceptions may be made for documentary or educational content.

#### **Adult Nudity and Sexual Behavior**

Adult Nudity and Sexual Behavior is any consensually produced and distributed media that is pornographic or intended to cause sexual arousal. Some examples include, but are not limited to, depictions of:

- full or partial nudity, including close-ups of genitals, buttocks, or breasts (excluding content related to breastfeeding);
- · simulated sexual acts; and
- sexual intercourse or other sexual acts this also applies to cartoons, hentai, or anime involving humans or depictions of animals with human-like features.

Note: exceptions may be made for artistic, medical, health, or educational content.

For content that was created or distributed without the consent of those featured, please refer to our <u>non-consensual nudity policy</u>.

#### **Violent Sexual Conduct**

Violent Sexual Conduct is any media that depicts violence, whether real or simulated, in association with sexual acts. Some examples include, but are not limited to, depictions of:

- rape and other forms of violent sexual assault, or sexual acts that occur without the consent of participants, including a simulated lack of consent; and
- sexualized violence inflicting physical harm on an individual within an intimate setting, where it is not immediately obvious if those involved have consented to take part.

#### **Bestiality and Necrophilia**

Bestiality is any media that depicts sexual acts between a human and an animal. Necrophilia is any media that depicts sexuals acts between a living human and a human corpse.

#### **Gratuitous Gore**

Gratuitous Gore is any media that depicts excessively graphic or gruesome content related to death, violence or severe physical harm, or graphic content that is shared for sadistic purposes. Some examples include, but are not limited to, depictions of:

- dismembered or mutilated humans;
- charred or burned human remains;
- exposed internal organs or bones; and
- animal torture or killing.

**Note:** exceptions may be made for religious sacrifice, food preparation or processing, and hunting.

# What is in violation of this policy?

Our aim is to limit exposure to sensitive images and videos and to prevent the sharing of potentially disturbing types of sensitive media. For this reason, we differentiate our enforcement approach depending on the type of media that has been shared and where it has been shared.

#### **Graphic Content, Adult Nudity, and Sexual Behavior**

- you can't target people with unsolicited images or videos that contain Graphic Content, Adult Nudity, and Sexual Behavior
- you can't include Graphic Content, Adult Nudity, or Sexual Behavior within live video, or in your profile picture or header, List banners, or Community cover photos.

#### Violent Sexual Conduct, Bestiality and Necrophilia, and Gratuitous Gore

We strictly prohibit sharing Violent Sexual Conduct and Bestiality and Necrophilia to prevent the normalization of sexual assault and non-consensual violence associated with sexual acts. We also prohibit Gratuitous Gore content because research has shown that repeated exposure to excessively graphic content online may negatively impact an individual's wellbeing.

**Note:** very limited exceptions may be made for gory media associated with newsworthy events.

# What is not a violation of this policy?

You can share Graphic Content, Adult Nudity, and Sexual Behavior within your posts, provided that you mark this media with a content warning.

You can adjust your media settings if you intend to regularly post this content, or add one-time content warnings to individual posts. Learn more <a href="https://help.twitter.com/en/rules-and-policies/media-settings">hete</a>. If you don't mark your media as sensitive, we will do so manually if your content is reported for review. We may also limit exposure of posts depicting Graphic Content, Adult Nudity, and Sexual Behavior.

# Who can report violations of this policy?

Anyone can report potential violations of this policy via our dedicated <u>reporting flows</u> (<a href="https://help.twitter.com/en/safety-and-security/report-a-post">https://help.twitter.com/en/safety-and-security/report-a-post</a>).

# What happens if you violate this policy?

The enforcement action we take depends on the type of media you have shared, and where you have shared it.

**Graphic Content, Adult Nudity, and Sexual Behavior** 

- if shared in a live video, or in your profile picture or header, List banner, or Community cover photo we will require you to remove this content
- if you target someone with this content, we'll require you to remove it under our <u>Abuse & Harassment policy (https://help.twitter.com/en/rules-and-policies/abusive-behavior.html)</u>
- we may limit exposure of posts depicting this content

#### Violent Sexual Conduct, Bestiality and Necrophilia

- sharing these kinds of media anywhere on X is strictly prohibited. For the first time you violate this policy, we will require you to remove this content. If you violate this policy again after your first warning, your account may be permanently suspended
- your account may be permanently suspended if the majority of your activity on X is sharing this type of media.

#### **Gratuitous Gore**

- if shared in a post, we will require you to remove this content
- if shared in a live video, or in your profile picture or header, List banner, or Community cover photo – the first time you violate this policy, we will require you to remove this content. If you violate this policy again after your first warning, your account may be permanently suspended
- your account may be permanently suspended if the majority of your activity on X is sharing this type of media.

If you believe that your account was suspended in error, you can <u>submit an appeal</u> (https://help.twitter.com/forms/general?subtopic=suspended).

Learn more about <u>our range of enforcement options</u> and our approach to <u>policy</u> <u>development and enforcement</u>.

© 2024 X Corp.

Cookies (https://help.twitter.com/rules-and-policies/twitter-cookies)

Privacy (https://twitter.com/privacy)

Terms and conditions (https://twitter.com/tos)

English	
	Help Center (https://help.twitter.com/en)

- 1. Help Center (https://help.twitter.com/en)
- 2. <u>Safety and cybercrime</u> (https://help.twitter.com/en/rules-and-policies#safety-and-cybercrime)

# Illegal or certain regulated goods or services

## **Overview**

#### March 2023

You may not use our service for any unlawful purpose or in furtherance of illegal activities. This includes selling, buying, or facilitating transactions in illegal goods or services, as well as certain types of regulated goods or services.

X takes the safety of our users seriously. In an effort to address the potential for real-world harm, we do not allow the use of X for any unlawful behavior or to further illegal activities. This includes selling, buying, or facilitating transactions in illegal goods or services, as well as certain types of regulated goods or services. In some cases, we may ask you to contact a law enforcement agency and have them contact us via our law enforcement request page to ensure we have enough context to enforce this policy or request additional information directly. In addition to reports received, we proactively surface activity that may violate this policy for human review.

Goods or services covered under this policy include, but are not limited to:

- counterfeit goods and services\*;
- drugs and controlled substances;
- human trafficking;
- products made from endangered or protected species;
- sexual services;
- · illicitly obtained materials; and
- weapons, including firearms, ammunition, and explosives, and instructions on making weapons (e.g. bombs, 3D printed guns, etc.)

\*Learn more about our <u>counterfeit</u> (https://help.twitter.com/rules-and-policies/counterfeit-goods-policy) policy.

# What happens if you violate this policy?

If we determine that you violated this policy, we may suspend your account, including upon first review.

Accounts that appear to be using misleading account information in order to engage in spamming, abusive, or disruptive behavior to promote the sale of illegal and regulated goods and/or services may be subject to suspension under our <u>platform</u> <u>manipulation and spam (https://help.twitter.com/rules-and-policies/platform-manipulation)</u> policy.

If you believe that your account was suspended in error, you can <u>submit an appeal</u> (<a href="https://help.twitter.com/forms/account-access/appeals">https://help.twitter.com/forms/account-access/appeals</a>).

# **Additional resources**

Learn more about <u>our range of enforcement options (https://help.twitter.com/rules-and-policies/enforcement-options)</u> and our approach to <u>policy development and enforcement (https://help.twitter.com/rules-and-policies/enforcement-philosophy</u>).

Help Center (https://help.twitter.com/en)

#### Share this article

Post	
© 2024 X Co	orp.
Cookies (http	s://help.twitter.com/rules-and-policies/twitter-cookies
Privacy (https	s://twitter.com/privacy)
Terms and c	onditions (https://twitter.com/tos)
English	

# **Private Content**

# **Overview**

#### March 2024

You may not threaten to expose, incentivize others to expose, or publish or post other people's private information without their express authorization and permission, or share private media of individuals without their consent.

Sharing someone's private information online without their permission, sometimes called "doxxing," is a breach of their privacy and can pose serious safety and security risks for those affected.

Additionally, posting images is an important part of our users' experience on X. However, where individuals have a reasonable expectation of privacy in an individual piece of media, we believe they should be able to determine whether or not it is shared. When we are notified by individuals depicted, or their authorized representative, that they did not consent to having media shared, we will remove the media. This policy is not applicable to public figures.

Lastly, having sexual, nude, or intimate media shared without express consent not only severely violates someone's privacy and psychological safety, but can lead to physical, emotional, and financial hardship and we work tirelessly to remove this content immediately.

When reviewing reports under our Private Content policies, we consider a number of things, including:

- What type of information is being shared? Certain types of private
  information carry higher risks than others, if they're shared without permission.
  Our primary goal is to protect individuals from potential physical harm as a
  result of their information being shared, so we consider information such as
  physical location and phone numbers to be a higher risk than other types of
  information.
- Who is sharing the information? It's important to consider who is sharing the
  reported information and whether or not they have the consent of the person it
  belongs to. We know there are times when people may want some forms of
  their personal information to be shared publicly, such as for professional
  networking, coordinating social events, or seeking help after a natural disaster.
- Is the information available elsewhere online? If the reported information was shared somewhere else before it was shared on X, such as someone sharing their personal phone number on their own publicly accessible website, we may not consider this information to be private, as the owner has made it publicly available elsewhere. However, we may take action against home addresses being shared, even if they are publicly available, due to the potential for physical harm.
- Why is the information being shared? The intent of the person sharing the
  information is also important to assess before taking action. For example, if we
  believe that someone is sharing information with abusive intent, or to harass
  another person, we will take action. On the other hand, if someone is sharing
  information in an effort to help someone involved in a crisis situation, we may
  not take action.

# What is in violation of this policy?

# **Posting Private Information**

You cannot share the following types of private information without the permission of the person it belongs to:

- home address or physical location information, such as street addresses, GPS coordinates, or other identifying information related to locations that are considered private
- identity documents, such as government-issued IDs or social security or other national identity numbers
- contact information, such as non-public personal phone numbers, email addresses, or passwords
- financial account information, such as bank account or credit card details
- health-related private information, such as biometric data or medical records
- · the identity of an anonymous user, such as their name or media depicting them

The following behaviors are also not permitted:

- threatening to publicly expose someone's private information
- sharing information that would enable individuals to hack or gain access to someone's private information without their consent, such as sharing login credentials for online banking services
- asking for or offering a bounty or financial reward in exchange for posting someone's private information
- asking for a bounty or financial reward in exchange for not posting someone's private information, sometimes referred to as blackmail.

# **Non-Consensual Nudity**

You cannot share sexual, nude, or intimate media (photos/videos) without the permission of the person involved, or that was taken or appears to have been taken without the consent of the people involved. This includes:

- hidden camera content featuring nudity, partial nudity, and/or sexual acts
- creepshots or upskirts images or videos taken of people's buttocks, up an individual's skirt/dress or other clothes that allows people to see the person's genitals, buttocks, or breasts
- images or videos that superimpose or otherwise digitally manipulate an individual's face onto another person's nude body
- images or videos that are taken in an intimate setting and not intended for public distribution

The following behaviors are also not permitted:

- threatening to publicly expose someone's sexual, nude, or intimate media
- sharing information that would enable individuals to hack or gain access to someone's sexual, nude, or intimate media without their consent, such as sharing login credentials for a private photo album
- asking for or offering a bounty or financial reward in exchange for posting someone's sexual, nude, or intimate media
- asking for a bounty or financial reward in exchange for not posting someone's sexual, nude, or intimate media, sometimes referred to as blackmail.

#### Right to Privacy

You cannot share media (photos/videos) of private individuals without the permission of the person(s) depicted.

# What is not a violation of this policy?

# **Posting Private Information**

The following are not in violation of this policy:

- people sharing their own private information
- sharing information that is publicly available elsewhere, in a non-abusive manner
- sharing identity documents or social security or other national identity numbers from regions where this information is not considered to be private
- sharing information that we don't consider to be private, including:
  - names
  - birthdates or ages
  - places of education or employment
  - location information related to commercial properties or places of business, where this information is publicly available
  - descriptions of physical appearance
  - gossip, rumors, accusations, and allegations
  - screenshots of text messages or messages from other platforms (unless they contain private information, such as someone's phone number).

# **Non-Consensual Nudity**

Pornography and other forms of consensually-produced adult content are allowed on X, provided you mark the post or <u>mark your account as sensitive</u> (<a href="https://help.twitter.com/rules-and-policies/media-settings.html">https://help.twitter.com/rules-and-policies/media-settings.html</a>). Doing so provides people who may not want to see this type of content with a warning that they will need to acknowledge before viewing your media. If you don't mark your media as sensitive, we may do so manually if your content is reported for review.

# **Right to Privacy**

The following are not in violation of this policy:

- the media is publicly available or is being covered by mainstream media
- the media and the accompanying tweet text add value to the public discourse or are shared in public interest
- the media contains eyewitness accounts or on the ground reports from developing events
- the subject of the media is a public figure.

# Who can report violations of this policy?

# **Posting Private Information**

Anyone can report violations of this policy using our dedicated reporting flow when the private information has been shared in a clearly abusive way. When this isn't the case, we may need to hear directly from the owner of the private information (or an authorized representative, such as a lawyer) before taking action.

# **Non-Consensual Nudity**

Due to X allowing some types of adult content on the platform, we may need additional context to determine if the content was created or shared without the consent of those involved. This means we may need to hear directly from the individual(s) featured (or an authorized representative, such as a lawyer) to ensure that we have sufficient context before taking any enforcement action.

However, to reduce the burden on those affected by non-consensual media, anyone can report the following types of content:

- · creepshots or upskirts
- content where a bounty or financial reward is offered in exchange for nonconsensual nudity media
- intimate images or videos that are accompanied by text that wishes/hopes for harm to come to those depicted or otherwise refers to revenge, such as "I hope you get what you deserve when people see this"
- intimate images or videos that are accompanied by information that could be used to contact those depicted, such as "you can tell my ex what you think by calling them at 123-456-7890."

# **Right to Privacy**

When reporting private media, we need a first person report in order to make the determination that the image or video has been shared without their permission. We accept reports from:

- · the individuals depicted in the reported media or
- their authorized representative, such as:
  - o parents/legal guardians
  - lawyers
  - legal representatives.

# What happens if you violate this policy?

When determining the penalty for violating this policy, we consider a number of factors including, but not limited to the severity of the violation and an individual's previous record of rule violations. The following is a list of potential enforcement options for content that violates this policy:

- · Requiring post removal.
  - For example, we may ask someone to remove the violating content and serve a period of time in read-only mode before they can post again.
     Subsequent violations may lead to account suspension.
- Suspending accounts whose sole purpose is to post others' private information or media of any kind. This includes accounts dedicated to sharing upskirt or creepshot media.

Learn more about our <u>range of enforcement options (https://help.twitter.com/rules-and-policies/enforcement-options.html)</u>. If someone believes their posts or account were enforced against in error, they can <u>submit an appeal (https://help.twitter.com/forms/account-access/appeals.html)</u>.

#### Share this article



# Non-consensual nudity policy

## **Overview**

#### December 2021

You may not post or share intimate photos or videos of someone that were produced or distributed without their consent.

Sharing explicit sexual images or videos of someone online without their consent is a severe violation of their privacy and the <u>X Rules</u>. Sometimes referred to as revenge porn, this content poses serious safety and security risks for people affected and can lead to physical, emotional, and financial hardship.

# What is a violation of this policy?

Under this policy, you can't post or share explicit images or videos that were taken, appear to have been taken or that were shared without the consent of the people involved.

Examples of the types of content that violate this policy include, but are not limited to:

- hidden camera content featuring nudity, partial nudity, and/or sexual acts;
- creepshots or upskirts images or videos taken of people's buttocks, up an individual's skirt/dress or other clothes that allows people to see the person's genitals, buttocks, or breasts;
- images or videos that superimpose or otherwise digitally manipulate an individual's face onto another person's nude body;
- images or videos that are taken in an intimate setting and not intended for public distribution; and
- offering a bounty or financial reward in exchange for intimate images or videos.

# What is not a violation of this policy?

Pornography and other forms of consensually produced adult content are allowed on X, provided that this media is marked as sensitive. Doing so provides people who may not want to see this type of content with a warning that they will need to acknowledge before viewing your media.

To mark your media as sensitive, navigate to your <u>safety settings</u>
<a href="mailto:chttps://twitter.com/settings/safety">chttps://twitter.com/settings/safety</a>) and select the **Mark media you post as containing**<a href="mailto:mark">material that may be sensitive</a> option. If you don't mark your media as sensitive,
we may do so manually if your content is reported for review.

# Who can report violations of this policy?

Because X allows some types of adult content, we need to evaluate context to assess if reported content has been created or shared without the consent of those involved.

#### Content reportable by anyone

We recognize that it can be difficult for those impacted to report this type of content for review. To reduce the burden on those affected, anyone can report the following types of content:

- creepshots or upskirts;
- content where a bounty or financial reward is offered in exchange for nonconsensual nudity media; and
- intimate images or videos that are accompanied by:
- text that wishes/hopes for harm to come to those depicted or otherwise refers to revenge e.g., "I hope you get what you deserve when people see this"; and
- information that could be used to contact those depicted e.g., "You can tell my ex what you think by calling them on 1234567".

#### Content reportable by featured individual(s)

For other types of content, we may need to hear directly from the individual(s) featured (or an authorized representative, such as a lawyer) to ensure that we have sufficient context before taking any enforcement action.

# How can I report violations of this policy?

#### In-app

You can report this content for review in-app as follows:

- 1. Select **Report post** from the \*\*oo\* icon.
- 2. Select It displays a sensitive photo or video.
- 3. Select An unauthorized photo or video.
- 4. Select It includes unauthorized, intimate content of me or someone else.
- 5. Select the relevant option depending on who you are reporting on behalf of.
- 6. Select up to 5 posts to report for review.
- 7. Submit your report.

#### **Desktop**

You can report this content for review via desktop as follows:

- 1. Select **Report post** from the \*\*o\*o icon.
- 2. Select It displays a sensitive photo or video.
- 3. Select **An unauthorized photo or video**.
- 4. Select It includes unauthorized, intimate content of me or someone else.
- 5. Select the relevant option depending on who you are reporting on behalf of.
- 6. Select up to 5 posts to report for review.
- 7. Submit your report.

#### Report form

You can also report this content via <u>our private information report form</u>, by selecting the **An unauthorized photo or video** option.

# What happens if you violate this policy?

We will **immediately and permanently suspend any account that we identify as the original poster of intimate media** that was created or shared without consent. We will do the same with any account that posts only this type of content, e.g., accounts dedicated to sharing upskirt images.

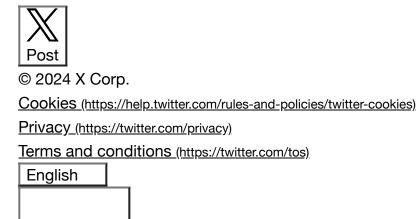
In other cases, we may not suspend an account immediately. This is because some people share this content inadvertently, to express shock, disbelief or to denounce this practice. In these cases, we will require you to remove this content. We will also temporarily lock you out of your account before you can post again. If you violate this policy again after your first warning, your account will be permanently suspended. If you believe that your account was suspended in error, you can <u>submit an appeal</u>.

## **Additional resources**

Learn more about <u>our range of enforcement options</u> and our approach to <u>policy</u> <u>development and enforcement</u>.

Help Center (https://help.twitter.com/en)

#### Share this article



# Platform manipulation and spam policy

# **Overview**

#### March 2023

You may not use X's services in a manner intended to artificially amplify or suppress information or engage in behavior that manipulates or disrupts people's experience or platform manipulation defenses on X.

We want X to be a place where people can make human connections, find reliable information, and express themselves freely and safely. To make that possible, we do not allow spam or other types of platform manipulation. We define platform manipulation as using X to engage in bulk, aggressive, or deceptive activity that misleads others and/or disrupts their experience.

Platform manipulation can take many forms and our rules are intended to address a wide range of prohibited behavior, including:

- commercially-motivated spam, that typically aims to drive traffic or attention from a conversation on X to accounts, websites, products, services, or initiatives;
- inauthentic engagements, that attempt to make accounts or content appear more popular or active than they are;
- coordinated activity, that attempts to artificially influence conversations through the use of multiple accounts, fake accounts, automation and/or scripting;
- coordinated harmful activity that encourages or promotes behavior which violates Our Rules; and
- leveraging X's open source code to circumvent remediations or platform defenses.

What is in violation of this policy?

Under this policy we prohibit a range of behaviors in the following areas:

# Multiple accounts and coordination

You can't mass-register X accounts or use automation to create X accounts.

X allows users to create and/or operate up to ten (10) accounts for different, non-duplicative purposes. Users who already own or operate more than ten accounts will not be impacted by these new limits unless a violation of the X Rules takes place. Some examples of what X considers to be non-duplicative uses include:

- Accounts tracking when objects in space travel over a specific location on Earth, e.g. New York, Hong Kong
- Accounts sharing news about different professional teams in the same sport or league, e.g. NFL, NBA, Premier League
- Accounts for personal projects, hobbies/interests, or for your business(es)
- Accounts for branded entities specific to unique locations or languages, e.g. brand accounts in English, Spanish and French

In addition, users are allowed to control multiple accounts on behalf of a third-party (e.g. "Social Media Managers"), provided that no violations of Our Rules (including this policy) occurs.

X may approve exceptions to multiple account limits at its discretion.

What is in violation of the Multiple Accounts policy?

You can't artificially amplify or disrupt conversations through the use of multiple accounts or by coordinating with others to violate Our Rules. This includes:

- operating multiple accounts that interact with the same or substantially similar content or X account(s) in order to inflate or manipulate the prominence of content and/or accounts; for example:
  - Do not create multiple accounts to "boost" trending topics or hashtags
  - Do not create multiple accounts to engage with the same posts, accounts, or polls
  - Do not create multiple accounts to misuse the mention/reply feature
  - Do not create multiple accounts to boost or amplify one of your own accounts by misusing engagement features (likes, reposts)
- operating multiple accounts that post substantially similar or identical content to one another; for example:
  - Do not cross-post content across multiple accounts; content that is cross-posted but localized into another language is permitted
  - Do not cross-post similar or duplicate content to the same trending topics or hashtags
- Employing workarounds to exceed the technical limits of account creation (e.g. account phone number limits)

Multiple accounts that exhibit a pattern or sole purpose of the preceding behaviors may be suspended by X. Users operating multiple accounts are still subject to all other X Rules, particularly our rules surrounding Platform Manipulation and Spam, and <u>Ban Evasion</u>. Any other violations of Our Rules may also result in all of your accounts being suspended.

**Engagement and metrics** 

You can't artificially inflate your own or others' followers or engagement. This includes:

- selling/purchasing Post or account metric inflation selling or purchasing followers or engagements (Reposts, Likes, mentions, X Poll votes);
- apps using or promoting third-party services or apps that claim to add followers or add engagements to Posts;
- reciprocal inflation trading or coordinating to exchange follows or Post engagements (including but not limited to participation in "follow trains," "decks," and "Repost for Repost" behavior); and
- account transfers or sales selling, purchasing, trading, or offering the sale, purchase, or trade of X accounts, usernames, or temporary access to X accounts.

Misuse of X product features

You can't misuse X product features to disrupt others' experience. This includes:

#### **Posts and Direct Messages**

- sending bulk, aggressive, high-volume unsolicited replies, mentions, or Direct Messages;
- posting and deleting the same content repeatedly;
- repeatedly posting identical or nearly identical Posts, or repeatedly sending identical Direct Messages;
- repeatedly posting content or sending Direct Messages consisting of links shared without commentary, so that this comprises the bulk of your Post/Direct Message activity; and
- Posting an existing phrase or content in a duplicative manner, whether individually or in concert with other accounts. Learn more in our <u>copypasta and</u> <u>duplicate content policy</u>.

#### **Editing Posts**

- editing Posts to artificially amplify content or deceive people. Examples include, but are not limited to:
  - using a Post's existing engagement to amplify substantially different content (e.g. editing a Post from "What's better? Pancakes or waffles?" to "Thousands of people trust my service. Like my post and subscribe to my channel to get tips on investing")
  - editing links (URLs) so that the final destination page has significantly changed, either in content or location (e.g. domain, URL path)
  - editing media (e.g. images, videos, audios, gifs, or URLs hosting relevant content) so the Post is meaningfully or entirely different or irrelevant to the previous versions.

#### **Following**

- "follow churn" following and then unfollowing large numbers of accounts in an effort to inflate one's own follower count:
- indiscriminate following following and/or unfollowing a large number of unrelated accounts in a short time period, particularly by automated means; and
- duplicating another account's followers, particularly using automation.

#### **Engagement**

- aggressively or automatically engaging with Posts to drive traffic or attention to accounts, websites, products, services, or initiatives.
- aggressively adding users to Lists or Moments.

#### **Hashtags**

- using a trending or popular hashtag with an intent to subvert or manipulate a conversation or to drive traffic or attention to accounts, websites, products, services, or initiatives; and
- Posting with excessive, unrelated hashtags in a single Post or across multiple Posts.

- publishing or linking to malicious content intended to damage or disrupt another person's browser (malware) or computer or to compromise a person's privacy (phishing); and
- posting misleading or deceptive links; e.g., affiliate links and clickjacking links.

#### **Open Source Code**

 exploiting X's open source code, including the logic behind visibility filtering or ML Model training, to circumvent policy enforcement for violations of <u>Our</u> Rules.

What is not a violation of this policy?

The following are not in violation of this policy:

- posting links without commentary occasionally;
- coordinating with others to express ideas, viewpoints, support, or opposition towards a cause, provided such behavior does not result in violations of Our Rules; and
- operating multiple accounts with distinct identities, purposes, or use cases.
   These accounts may interact with one another, provided they don't violate other rules. Some examples include:
  - organizations with related but separate chapters or branches, such as a business with multiple locations;
  - operating a personal account in addition to pseudonymous accounts or accounts associated with your hobbies or initiatives; and
  - hobby/artistic bots.

Who can report violations of this policy?

Anyone can report accounts or Posts via our dedicated reporting flow. These reports are used in aggregate to help refine our enforcement systems and identify new and emerging trends and patterns of behavior.

# How can I report violations of this policy?

#### In-app

You can report this content in-app as follows:

- 1. Select **Report Post** from the icon.
- 2. Select It's suspicious or spam.
- 3. Select the option that best tells us how the Post is suspicious or spreading spam.
- 4. Submit your report.

#### Web

You can report this content via web as follows:

- 1. Select **Report Post** from the icon.
- 2. Select It's suspicious or spam.
- Select the option that best tells us how the Post is suspicious or spreading spam.
- 4. Submit your report.

# What happens if you violate this policy?

The consequences for violating this policy depend on the severity of the violation as well as any previous history of violations. Our action is also informed by the type of spammy activity that we have identified. The actions we take may include the following:

#### **Anti-spam challenges**

When we detect suspicious levels of activity, accounts may be locked and prompted to provide additional information (e.g., a phone number) or to solve a reCAPTCHA.

#### **Denylisting URLs**

We denylist or provide warnings about URLs we believe to be unsafe. Read more about <u>unsafe links</u>, including how to appeal if we've falsely identified your URL as unsafe.

#### Limiting the visibility of Posts

Posts which we believe to be in violation of these policies may not appear in certain parts of the X product, and/or may not be recommended or amplified by X. <u>Learn more</u> about other instances when a Post's visibility may be limited.

#### Post deletion and temporary account locks

- If the platform manipulation or spam offense is an isolated incident or first offense, we may take a number of actions ranging from requiring deletion of one of more Posts to temporarily locking account(s). Any subsequent platform manipulation offenses will result in permanent suspension.
- In the case of a violation centering around the use of multiple accounts, you
  may be asked to choose one account to keep. The remaining accounts will be
  permanently suspended.

#### Temporary loss of access to X features or products

 We may temporarily limit or restrict access to X features or products, including (but not limited to) Posts, Edit Post, Direct Messages, Spaces, Communities, or Live.

#### **Permanent suspension**

For severe violations, accounts will be permanently suspended at first detection. Examples of severe violations include:

- operating accounts where the majority of behavior is in violation of the policies described above;
- using any of the tactics described on this page to undermine the integrity of elections;
- buying/selling accounts;
- creating accounts to replace or mimic a suspended account; and
- operating accounts that X is able to reliably attribute to entities known to violate Our Rules.

If you believe that your account was locked or suspended in error, you can <u>submit</u> an <u>appeal</u> (https://help.twitter.com/forms/general?subtopic=suspended).

# **Additional resources**

Learn more about our <u>automation rules</u> for developers, our <u>election integrity</u> (<a href="https://about.x.com/en/values/elections-integrity.html">https://about.x.com/en/values/elections-integrity.html</a>) efforts, our <u>financial scam policy</u>, our <u>hacked materials policy</u>, and our <u>guidelines for promotions and contests</u>.

Learn more about our <u>range of enforcement options</u> and our approach to <u>policy</u> <u>development and enforcement</u>.

## Share this article

Post	
© 2024 X Corp.	
Cookies (https://h	<u>elp.twitter.com/rules-and-policies/twitter-cookies)</u>
Privacy (https://tw	<u>itter.com/privacy)</u>
Terms and cond	itions (https://twitter.com/tos)
English	
	Help Center (https://help.twitter.com/en)

# **Civic integrity policy**

# **Overview**

# August 2023

You may not use X's services for the purpose of manipulating or interfering in elections or other civic processes, such as posting or sharing content that may suppress participation, mislead people about when, where, or how to participate in a civic process, or lead to offline violence during an election. Any attempt to undermine the integrity of civic participation undermines our core tenets of freedom of expression and as a result, we will apply labels to violative posts informing users that the content is misleading.

#### What is a civic process?

X considers civic processes to be events or procedures mandated, organized, and conducted by the governing and/or electoral body of a country, state, region, district, or municipality to address a matter of common concern through public participation. Some examples of civic processes may include:

- Political elections
- Censuses
- Major referenda and ballot initiatives

#### What is in violation of this policy?

This policy addresses 4 categories of misleading behavior and content:

#### Misleading information about how to participate

You may not advance verifiably false or misleading information about how to participate in an election or other civic process. This includes but is not limited to:

- misleading information about procedures to participate in a civic process (for example, that you can vote by Post, text message, email, or phone call in jurisdictions where these are not a possibility);
- misleading information about requirements for participation, including identification or citizenship requirements;
- misleading claims that cause confusion about the established laws,
   regulations, procedures, and methods of a civic process, or about the actions of officials or entities executing those civic processes; and
- misleading statements or information about the official, announced date or time of a civic process.

#### **Suppression**

You may not advance verifiably false or misleading information about the circumstances surrounding a civic process intended to intimidate or dissuade people from participating in an election or other civic process. This includes but is not limited to:

- misleading claims that polling places are closed, that polling has ended, or other misleading information relating to votes not being counted;
- misleading claims about police or law enforcement activity related to voting in an election, polling places, or collecting census information;
- misleading claims about long lines, equipment problems, or other disruptions at voting locations during election periods;

#### Intimidation

You may not engage in or promote behaviors that may coerce others to refrain from participating in a civic process. This includes, but is not limited to:

- inciting or promoting violent behaviors intentionally near a location where an electoral process is being conducted, including polling stations and vote counting locations;
- inciting the disruption or destruction of procedures, infrastructure, or election equipment that is necessary for someone to participate in a civic process;
- inciting others to harass voters or poll workers;
- promoting the brandishing of firearms near polling locations to intimidate voters and election workers;
- threats regarding voting locations or other key places or events (note that our <u>Violent Speech (https://help.twitter.com/en/rules-and-policies/violent-speech)</u> policy may also be relevant for threats not covered by this policy).

#### False or misleading affiliation

You may not create fake accounts which misrepresent their affiliation, or share content that falsely represents its affiliation, to a candidate, elected official, political party, electoral authority, or government entity. Read more about our <u>Misleading and deceptive identities policy</u>.

#### What is not a violation of this policy?

Not all false or untrue information about politics or civic processes constitutes manipulation or interference. In the absence of other policy violations, the following are generally not in violation of this policy:

- inaccurate statements about an elected or appointed official, candidate, or political party;
- organic content that is polarizing, biased, hyperpartisan, or contains controversial viewpoints expressed about elections or politics;
- discussion of public polling information; voting and audience participation for competitions, game shows, or other entertainment purposes;
- using X pseudonymously or as a <u>parody</u>, <u>commentary</u>, <u>or fan account</u> (<a href="https://help.twitter.com/en/rules-and-policies/twitter-impersonation-and-deceptive-identities-policy">https://help.twitter.com/en/rules-and-policies/twitter-impersonation-and-deceptive-identities-policy</a>) to discuss elections or politics.

#### What happens if you violate this policy?

Posts that are enforced under this policy will have their reach restricted on X by:

- Excluding the post from search results, trends, and recommended notifications
- Removing the post from the For you and Following timelines
- Restricting the post's discoverability to the author's profile
- Restricting Likes, replies, reposts, quotes, bookmarks, share, pin to profile, or Edit post
- Downranking the Post in replies

Posts enforced under this policy will receive a label informing both Post authors and viewers that we have limited the Post's visibility. Post authors are able to submit an appeal on the label if they think we incorrectly limited their Post's visibility. Learn more about our <u>range of enforcement options (https://help.twitter.com/en/rules-and-policies/enforcement-options.html)</u>.

#### Share this article



- 1. Help Center (https://help.twitter.com/en)
- 2. <u>Platform integrity and authenticity</u> (https://help.twitter.com/en/rules-and-policies#platform-integrity-and-authenticity)

# Misleading and deceptive identities policy

**Overview** 

**April 2023** 

# You may not misappropriate the identity of individuals, groups, or organizations or use a fake identity to deceive others.

We want X to be a place where people can find authentic voices. While you are not required to display your real name or image on your profile, your account should not use false profile information to represent itself as a person or entity that is not affiliated with the account owner, such that it may mislead others who use X.

# What is in violation of this policy?

We prohibit the following behaviors under this policy:

#### **Impersonation**

You may not pose as an existing person, group, or organization to mislead others about who you are or who you represent. Accounts that violate this policy will misrepresent their identity by using **at least two elements** of another identity, such

as the name, image, or false claims of affiliation with another individual or organization in their profile or posts.

#### **Deceptive Identities**

You may not pose as someone who doesn't exist to mislead others about who you are or who you represent. This includes using at least one element of someone else's identity on your profile or posts in a misleading way, such as using another individual's image or making a false statement of affiliation with an existing individual or entity. We also consider accounts to be deceptive if they use a computer generated image of a person to pose as someone who doesn't exist.

# What is not a violation of this policy?

We believe giving people choice in terms of how they represent themselves online enables them to express themselves and control their privacy. X allows the use of pseudonymous accounts, meaning an account's profile is not required to use the name or image of the account owner. Accounts that appear similar to others on X are not in violation of this policy, so long as their purpose is not to deceive or manipulate others. Accounts that share your name but have no other commonalities are not in violation of this policy.

#### Parody, Commentary, or Fan Accounts

Accounts that depict another person, group, or organization in their profile to discuss, satirize, or share information about that entity are not in violation of this policy. While these accounts may use elements of another's identity, they also include profile language or other indicators that inform people on X that the account is not affiliated with the subject of the profile.

To avoid confusing others about an account's affiliation, Parody, Commentary, and Fan accounts must distinguish themselves in their account name and in their bio. Accounts that fail to sufficiently distinguish themselves are considered non-compliant and in violation of this policy.

- Account name: The account name should clearly indicate that the account is
  not affiliated with the subject portrayed in the profile. Accounts can indicate
  this by incorporating words such as, but not limited to, "parody," "fake," "fan,"
  or "commentary." This language should be stated in a way that can be
  understood by any audience and should not be contradicted by other affiliation
  terminology such as "official." Please note, an account name is separate from
  the username (i.e. @handle).
- Bio: The bio should clearly state that the account is not affiliated with the subject portrayed in the profile. Non-affiliation can be indicated by incorporating words such as, but not limited to, "not affiliated with," "parody," "fake," "fan," or "commentary." This language should be stated in a way that can be understood by any audience.

# How can I report violations of this policy?

If you believe an account is posing as you or your brand, you can file a report <a href="https://help.twitter.com/forms/impersonation">here</a> (<a href="https://help.twitter.com/forms/impersonation">here</a>). If you believe an account is using a deceptive fake identity or misusing the identity of somebody else, you can flag it as a bystander by <a href="https://reporting.org/reporting">reporting</a> directly from the account's profile.

In cases where an account is suspected of misusing a specific individual or entity's identity, we may need more information to determine whether the account is run or authorized by the entity portrayed in the profile. To ensure we have enough context, we may need a report from the portrayed party in order to take action.

# What happens if you violate this policy?

The consequences for violating the policy depend on the severity and type of violation, as well as an account's history of previous violations. The actions we take may include the following:

#### **Profile modifications**

If your account is potentially confusing in terms of its affiliation, including as a non-compliant parody, commentary or fan account, we may require you to edit the content on your profile. If you violate this policy again after your first warning, your account may be suspended.

#### Suspension

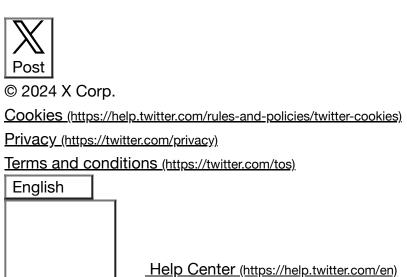
If you are engaged in impersonation or are using a deceptive identity, we may suspend your account. If we believe you may be in violation of this policy, we may require you to provide government issued identification (such as a driver's license or passport) in order to reinstate your account.

If you believe that your account was locked or suspended in error, you can <u>submit</u> an <u>appeal (https://help.twitter.com/forms/general?subtopic=suspended)</u>.

# **Additional resources**

To report an account that may be misusing your trademark, please see our <u>trademark policy</u>. For information about our spam rules, you can read more on our <u>platform manipulation and spam</u> policy. You may submit feedback using this <u>form (https://help.twitter.com/en/twitter-for-you)</u>.

#### Share this article



- 1. Help Center (https://help.twitter.com/en)
- 2. <u>Platform integrity and authenticity</u> (https://help.twitter.com/en/rules-and-policies#platform-integrity-and-authenticity)

# Synthetic and manipulated media policy

## **Overview**

# **April 2023**

You may not share synthetic, manipulated, or out-of-context media that may deceive or confuse people and lead to harm ("misleading media"). In addition, we may label posts containing misleading media to help people understand their authenticity and to provide additional context.

# What is in violation of this policy

In order for content with **misleading media** (including images, videos, audios, gifs, and URLs hosting relevant content) to be labeled or removed under this policy, it must:

- Include media that is significantly and deceptively altered, manipulated, or fabricated, or
- Include media that is shared in a deceptive manner or with false context, and
- Include media likely to result in widespread confusion on public issues, impact public safety, or cause serious harm

We use the following criteria as we consider posts and media for labeling or removal under this policy as part of our ongoing work to enforce our rules and ensure healthy and safe conversations on X:

1. Is the content significantly and deceptively altered, manipulated, or fabricated?

In order for content to be labeled or removed under this policy, we must have reason to believe that media are significantly and deceptively altered, manipulated, or fabricated. Synthetic and manipulated media take many different forms and people can employ a wide range of technologies to produce these media. Some of the factors we consider include:

- whether media have been substantially edited or post-processed in a manner that fundamentally alters their composition, sequence, timing, or framing and distorts their meaning;
- whether there are any visual or auditory information (such as new video frames, overdubbed audio, or modified subtitles) that has been added, edited, or removed that fundamentally changes the understanding, meaning, or context of the media;
- whether media have been created, edited, or post-processed with enhancements or use of filters that fundamentally changes the understanding, meaning, or context of the content; and
- whether media depicting a real person have been fabricated or simulated, especially through use of artificial intelligence algorithms

We will not take action to label or remove media that have been edited in ways that do not fundamentally alter their meaning, such as retouched photos or colorcorrected videos.

In order to determine if media have been significantly and deceptively altered or fabricated, we may use our own technology or receive reports through partnerships with third parties. In situations where we are unable to reliably determine if media have been altered or fabricated, we may not take action to label or remove them.

#### 2. Is the content shared in a deceptive manner or with false context?

We also consider whether the context in which media are shared could result in confusion or suggests a deliberate intent to deceive people about the nature or origin of the content, for example, by falsely claiming that it depicts reality. We assess the context provided alongside media to see whether it provides true and factual information. Some of the types of context we assess in order to make this determination include:

- whether inauthentic, fictional, or produced media are presented or being endorsed as fact or reality, including produced or staged works, reenactments, or exhibitions portrayed as actual events;
- whether media are presented with false or misleading context surrounding the source, location, time, or authenticity of the media;
- whether media are presented with false or misleading context surrounding the identity of the individuals or entities visually depicted in the media;
- whether media are presented with misstatements or misquotations of what is being said or presented with fabricated claims of fact of what is being depicted

We will not take action to label or remove media that have been shared with commentary or opinions that do not advance or present a misleading claim on the context of the media such as those listed above.

In order to determine if media have been shared in a deceptive manner or with false context, we may use our own technology or receive reports through partnerships with third parties. In situations where we are unable to reliably determine if media have been shared with false context, we will not label or remove the content.

## 3. Is the content likely to result in widespread confusion on public issues, impact public safety, or cause serious harm?

Posts that share misleading media are subject to removal under this policy if they are likely to cause serious harm. Some specific harms we consider include:

- Threats to physical safety of a person or group
- Incitement of abusive behavior to a person or group
- Risk of mass violence or widespread civil unrest
- Risk of impeding or complicating provision of public services, protection efforts, or emergency response
- Threats to the privacy or to the ability of a person or group to freely express themselves or participate in civic events, such as:
  - Stalking or unwanted and obsessive attention
  - Targeted content that aims to harass, intimidate, or silence someone else's voice
  - Voter suppression or intimidation

We also consider the time frame within which the content may be likely to impact public safety or cause serious harm, and are more likely to remove content under this policy if immediate harm is likely to result. Posts with misleading media that are not likely to result in immediate harm but still have a potential to impact public safety, result in harm, or cause widespread confusion towards a public issue (health, environment, safety, human rights and equality, immigration, and social and political stability) may be labeled to reduce their spread and to provide additional context.

While we have other rules also intended to address these forms of harm, including our policies on violent threats, civic integrity, and hateful conduct, we will err toward removal in borderline cases that might otherwise not violate existing rules for Posts that include misleading media.

#### What is not a violation of this policy

We seek to protect public conversation surrounding various issues. Media often accompany these conversations and encourage further discourse. In the absence of other policy violations, the following are generally not in violation of this policy:

- Memes or satire, provided these do not cause significant confusion about the authenticity of the media;
- Animations, illustrations, and cartoons, provided these do not cause significant confusion about the authenticity of the media.
- Commentary, reviews, opinions, and/or reactions. Sharing media with edits
  that only add commentary, reviews, opinions, or reactions allows for further
  debate and discourse relating to various issues and are not in violation of this
  policy.
- Counterspeech. We allow for direct responses to misleading information
  which seek to undermine its impact by correcting the record, amplifying
  credible information, and educating the wider community about the prevalence
  and dynamics of misleading information.

#### What happens if you violate this policy?

The consequences for violating our synthetic and manipulated media policy depends on the severity of the violation.

#### **Post Deletion**

For high-severity violations of the policy, including misleading media that have a serious risk of harm to individuals or communities, we will require you to remove this content.

#### Labeling

In circumstances where we do not remove content which violates this policy, we may provide additional context on posts sharing the misleading media where they appear on X. This means we may:

- Apply a label and/or warning message to the post
- Show a warning to people before they share or like the post;
- Reduce the visibility of the post on the platform and/or prevent it from being recommended;
- Turn off likes, replies, and Reposts; and/or
- Provide a link to additional explanations or clarifications, such as relevant X policies.

In most cases, we will take a combination of the above actions on posts we label.

#### **Account locks**

If we determine that an account has advanced or continuously shares harmful misleading narratives that violate the synthetic and manipulated media policy, we may temporarily reduce the visibility of the account or lock or suspend the account.

If you believe that your account was locked or suspended in error, you can <u>submit</u> an <u>appeal</u> (https://help.twitter.com/forms/general?subtopic=suspended).

#### **Additional resources**

Learn more about our range of enforcement options.

#### Share this article



© 2024 X Corp.

Cookies (https://help.twitter.com/rules-and-policies/twitter-cookies)

Privacy (https://twitter.com/privacy)

Terms and conditions (https://twitter.com/tos)

English	
	Help Center (https://help.twitter.com/en)

## **Copyright policy**

# What types of copyright complaints does X respond to?

X responds to copyright complaints submitted under the Digital Millennium Copyright Act ("DMCA"). Section 512 of the DMCA outlines the statutory requirements necessary for formally reporting copyright infringement, as well as providing instructions on how an affected party can appeal a removal by submitting a compliant counter-notice.

X will respond to reports of alleged copyright infringement, such as allegations concerning the unauthorized use of a copyrighted image as a profile or header photo, allegations concerning the unauthorized use of a copyrighted video or image uploaded through our media hosting services, or posts containing links to allegedly infringing materials. Note that not all unauthorized uses of copyrighted materials are infringements (see our <u>fair use</u> article for more information).

If you are concerned about the use of your brand or entity's name, please review <u>Twitter's trademark policy</u>. If you are concerned about a parody, newsfeed, commentary, or fan account, please see the relevant policy <u>here</u>. These are generally not copyright issues.

#### Am I a copyright holder? How do I know?

If you are unsure whether you hold rights to a particular work, please consult an attorney or another adviser as X cannot provide legal advice. There are plenty of resources to learn more about copyright law including <a href="http://copyright.gov/">http://copyright.gov/</a>) and <a href="http://www.eff.org/issues/bloggers/legal/liability/IP">http://copyright.gov/</a>) and <a href="http://www.eff.org/issues/bloggers/legal/liability/IP">http://www.eff.org/issues/bloggers/legal/liability/IP</a>), just to name a few.

# What to consider before submitting a copyright complaint

Before submitting a copyright complaint to us, please consider whether or not the use could be considered <u>fair use</u>.

If you have considered fair use, and you still wish to continue with a copyright complaint, you may want to first reach out to the user in question to see if you can resolve the matter directly with the user. You can reply to the user's post or send the

user a Direct Message and ask for them to remove your copyrighted content without having to contact X.

Prior to submitting a formal complaint with X, please be aware that under 17 U.S.C. § 512(f), you may be liable for any damages, including costs and attorneys' fees incurred by us or our users, if you knowingly materially misrepresent that material or activity is infringing. If you are unsure whether the material you are reporting is in fact infringing, you may wish to contact an attorney before filing a notification with us.

Note: In general, the photographer and NOT the subject of a photograph is the actual rights holder of the resulting photograph. If you're unsure whether or not you own the copyrights to a work, or if you're infringing upon someone else's work, please consult an attorney or another advisor.

# What information do you need to process a copyright complaint?

To submit a notice of claimed copyright infringement, you will need to provide us with the following information:

- 1. A physical or electronic signature (typing your full name will suffice) of the copyright owner or a person authorized to act on their behalf;
- Identification of the copyrighted work claimed to have been infringed (e.g., a link to your original work or clear description of the materials allegedly being infringed upon);
- 3. Identification of the infringing material and information reasonably sufficient to permit X to locate the material on our website or services;
- 4. Your contact information, including your address, telephone number, and an email address;
- 5. A statement that you have a good faith belief that the use of the material in the manner asserted is not authorized by the copyright owner, its agent, or the law; and
- A statement that the information in the complaint is accurate, and, under penalty of perjury, that you are authorized to act on behalf of the copyright owner.

If you are reporting the content of a post, please give us a <u>direct link</u> to that post. Or please specify if the alleged infringement is in the header, avatar, etc. A LINK TO A PROFILE PAGE IS INSUFFICIENT FOR X TO IDENTIFY INFRINGING MATERIALS.

#### How do I file a copyright complaint?

You can report alleged copyright infringement by visiting X Help Center and <u>filing a copyright complaint (https://help.twitter.com/forms/ipi)</u>. If you are logged in to x.com, you can visit the X Help Center directly from your X account by clicking the 'Help' link located in the sidebar.

Filing a DMCA complaint is the start of a pre-defined legal process. Your complaint will be reviewed for accuracy, validity, and completeness. If your complaint has satisfied these requirements, we will take action on your request - which includes forwarding a full copy of your notice (including your name, address, phone and email address) to the user(s) who posted the allegedly infringing material in question.

If you are concerned about your contact information being forwarded, you may wish to use an agent to report for you.

Please be aware that under 17 U.S.C. § 512(f), you may be liable for any damages, including costs and attorneys' fees incurred by us or our users, if you knowingly materially misrepresent that material or activity is infringing. If you are unsure whether the material you are reporting is in fact infringing, you may wish to contact an attorney before filing a copyright complaint.

#### How are claims processed?

We process copyright complaints in the order in which they are received. Once you've submitted your ticket, we will email you a ticket confirmation. If you do not receive a ticket confirmation that means we did not receive your complaint and you should re-submit your complaint. However, please note, submitting duplicate copyright complaints will result in a delay in processing.

If we decide to remove or disable access to the material, we will notify the affected user(s) and provide them with a full copy of the reporter's complaint (including the provided contact information) along with instructions on how to file a counter-notice.

# What information gets forwarded to the reported user(s)?

If we remove or disable access to the materials reported in a copyright complaint, the reported user(s) will receive a copy of the complaint, including the reporter's full name, email, street address, and any other information included in the complaint.

If you are uncomfortable sharing your contact information with the reported user(s), you may wish to consider appointing an agent to submit your DMCA notice on your behalf. Your agent will be required to submit the DMCA notice with valid contact information, and identify you as the content owner that they are representing.

#### What happens next?

X's response to copyright complaints may include the removal or restriction of access to allegedly infringing material. If we remove or restrict access to user content in response to a copyright complaint, X will make a good faith effort to contact the affected account holder with information concerning the removal or restriction of access, including a full copy of the complaint, along with instructions for filing a counter-notice.

If you've not yet received a copy of the copyright complaint regarding the content removed from your account, please respond to the support ticket we sent you.

In an effort to be as transparent as possible regarding the removal or restriction of access to user-posted content, we clearly mark withheld posts and media to indicate to viewers when content has been withheld (examples below).

# My content was removed from X

#### Why did I receive a copyright complaint?

If you receive a copyright complaint, it means that access to the content described in the complaint has been restricted. Please take the time to read through our correspondence to you, which includes information on the complaint we received as well as instructions on how to file a counter-notice. Please ensure that you are monitoring the email address associated with your X account.

Tip: Removing the material reported in a copyright complaint will not resolve that complaint

#### What if I want to contest the takedown?

If you believe that the materials reported in the copyright complaints were misidentified or removed in error, you may send us a <u>counter-notification(s)</u> through our Help Center. A counter-notice is a request for X to

reinstate the removed material, and it has legal consequences. Alternatively, you may be able to seek a retraction of the copyright complaint from the reporter.

#### How do I seek a retraction?

The DMCA complaint you received includes the contact information of the reporter. You may want to reach out and ask them to retract their notice using our <u>retraction form (https://help.twitter.com/forms/ipi/dmca-retraction)</u>. This is the fastest and most efficient means of resolving an unresolved copyright complaint. Alternatively, the reporter can also send a retraction notice to <u>copyright@twitter.com</u>. Such notice should include: (1) identification of the material that was disabled, and (2) a statement that the reporter would like to retract their DMCA notice. A retraction is at the sole discretion of the original reporter, and only that reporter can issue a retraction. If a portion of your post has been withheld, the violation will not be resolved just by deleting the remaining portion.

#### When should I file a counter-notice?

A counter-notice is a request for X to reinstate the removed material, and is the start of a legal process that has legal consequences. For example, submitting a counter notice indicates that you consent to the jurisdiction of a U.S. Federal court and that you consent to the disclosure of your personal information to the reporter.

With these considerations in mind, you may file a counter-notice if you believe that this material was misidentified, or you have a good faith belief that the material should not have been removed. If you're unsure whether or not you should file a counter-notice, you may want to consult with an attorney.

Tip: Re-posting material removed in response to a copyright complaint may result in permanent account suspension. If you believe the content was removed in error, please file a counter-notice rather than re-posting the material.

#### What information do you need to process a counternotice?

To submit a counter-notice, you will need to provide us with the following information:

- A physical or electronic signature (typing your full name will suffice);
- 2. Your full name, address (including country), telephone number, and for verification purposes, your X username and associated email;
- 3. Identification of the material that has been removed or to which access has been disabled, and the location at which the material appeared before that removal or disabling (the description from the copyright notice will suffice);
- 4. A statement under penalty of perjury that you have a good faith belief that the material was removed or disabled as a result of mistake or misidentification of the material to be removed or disabled; and
- 5. One of the following jurisdictional consent statements:

(If your address is in the United States)

"I consent to the jurisdiction of the Federal District Court for the judicial district in which my address is located and I'll accept service of process from the person who provided notice under 17 U.S.C. 512 (c)(1)(C) or an agent of such person."

OR

(If your address is outside of the United States)

"I consent to any judicial district in which X may be found, and I'll accept service of process from the person who provided notice under 17 U.S.C. 512 (c)(1)(C) or an agent of such person."

To submit a counter-notice, please provide all of the above information using our web form, located <a href="https://help.twitter.com/forms/ipi/dmca-counternotice.html">https://help.twitter.com/forms/ipi/dmca-counternotice.html</a>).

#### What happens after I submit a counter-notice?

Upon receipt of a valid counter-notice, we will promptly forward a copy to the person who filed the original notice. This means that the contact information that is submitted in your counter-notice will be shared to the person who filed the original notice.

If the copyright owner disagrees that the content was removed in error or misidentification, they may pursue legal action against you. If we do not receive notice within 10 business days that the original reporter is seeking a court order to prevent further infringement of the material at issue, we may replace or cease disabling access to the material that was removed.

We cannot offer any legal advice. Should you have questions, please consult an attorney.

# Filing a copyright complaint or counter-notice is serious business!

Please think twice before submitting a claim or counter-notice, especially if you are unsure whether you are the actual rights holder or authorized to act on a rights holder's behalf. There are legal and financial consequences for fraudulent and/or bad faith submissions. Please be sure that you are the actual rights holder, or that you have a good faith belief that the material was removed in error, and that you understand the repercussions of submitting a false claim.

# What happens if my account receives one or more copyright complaints?

If multiple copyright complaints are received about an account, or other evidence suggests a pattern of repeat infringement, X may suspend that account in accordance with our Repeat Infringer Policy. Our Repeat Infringer Policy takes valid retractions and counter-notices into account. Users suspended under that policy may file a suspension appeal using <a href="mailto:third://thi

Additionally, re-posting material that was already removed in response to a copyright complaint, indicating where to access copyrighted material on third party websites, or certain other actions that demonstrate your account is intended solely or primarily to infringe the copyrighted material of others may result in permanent account suspension. If you are suspended on this basis, you may file an appeal using <a href="mailto:this://help.twitter.com/forms/general?subtopic=suspended">this form (https://help.twitter.com/forms/general?subtopic=suspended)</a>. Please be aware that appealing a suspension does not guarantee your account will be reinstated.

#### **Share this article**



### **Trademark policy**

#### **Overview**

#### **April 2023**

You may not violate others' intellectual property rights, including copyright and trademark.

A trademark is a word, logo, phrase, or device that distinguishes a trademark holder's good or service in the marketplace. Trademark law may prevent others from using a trademark in an unauthorized or confusing manner.

#### What is in violation of this policy?

Using another's trademark in a way that may mislead or confuse people about your affiliation may be a violation of our trademark policy.

#### What is not a violation of this policy?

Referencing another's trademark is not automatically a violation of X's trademark policy. Examples of non-violations include:

- using a trademark in a way that is outside the scope of the trademark registration e.g., in a different territory, or a different class of goods or services than that identified in the registration; and
- using a trademark in a nominative or other fair use manner. For more information, see our <u>Misleading and deceptive identities policy</u>.

#### Who can report violations of this policy?

X only investigates requests that are submitted by the trademark holder or their authorized representative e.g., a legal representative or other representative for a brand.

#### How can I report violations of this policy?

You can submit a trademark report through our <u>trademark report form</u> (<a href="https://help.twitter.com/forms/trademark">https://help.twitter.com/forms/trademark</a>). Please provide all the information requested in the form. If you submit an incomplete report, we'll need to follow up about the missing information. Please note that this will result in a delay in processing your report.

**Note:** We may provide the account holder with your name and other information included in the copy of the report.

#### What happens if you violate this policy?

If we determine that you violated our trademark policy, we may suspend your account. Depending on the type of violation, we may give you an opportunity to comply with our policies. In other instances, an account may be permanently suspended upon first review. If you believe that your account was suspended in error, you can <u>submit an appeal (https://help.twitter.com/forms/general?subtopic=suspended)</u>.

#### **Additional resources**

Learn more about <u>our range of enforcement options</u> (https://help.twitter.com/rules-and-policies/enforcement-options) and our approach to <u>policy development and enforcement</u> (https://help.twitter.com/rules-and-policies/enforcement-philosophy).

- 1. Help Center (https://help.twitter.com/en)
- 2. <u>Platform Use Guidelines</u> (https://help.twitter.com/en/rules-and-policies#platform-use-guidelines)

# Our approach to policy development and enforcement philosophy

X is reflective of real conversations happening in the world and that sometimes includes perspectives that may be offensive, controversial, and/or bigoted to others. While we welcome everyone to express themselves on our service, we will not tolerate behavior that harasses, threatens, or uses fear to silence the voices of others.

We have the <u>X Rules</u> in place to help ensure everyone feels safe expressing their beliefs and we strive to enforce them with uniform consistency. Learn more about <u>different enforcement actions</u> (https://support.twitter.com/articles/20175435).

#### Our policy development process

Creating a new policy or making a policy change requires in-depth research around trends in online behavior, developing clear external language that sets expectations around what's allowed, and creating enforcement guidance for reviewers that can be scaled across millions of posts.

While drafting policy language, we gather feedback from a variety of internal teams as well as our Trust & Safety Council. This is vital to ensure we are considering global perspectives around the changing nature of online speech, including how our rules are applied and interpreted in different cultural and social contexts. Finally, we train our global review teams, update the X Rules, and start enforcing the new policy.

#### Our enforcement philosophy

We empower people to understand different sides of an issue and encourage dissenting opinions and viewpoints to be discussed openly. This approach allows many forms of speech to exist on our platform and, in particular, promotes counterspeech: speech that presents facts to correct misstatements or misperceptions, points out hypocrisy or contradictions, warns of offline or online consequences, denounces hateful or dangerous speech, or helps change minds and disarm.

Thus, **context matters**. When determining whether to take enforcement action, we may consider a number of factors, including (but not limited to) whether:

- the behavior is directed at an individual, group, or protected category of people;
- the report has been filed by the target of the abuse or a bystander;
- the user has a history of violating our policies;
- the severity of the violation;
- the content may be a topic of legitimate public interest.

#### Is the behavior directed at an individual or group of people?

To strike a balance between allowing different opinions to be expressed on the platform, and protecting our users, we enforce policies when someone reports abusive behavior that targets a specific person or group of people. This targeting can happen in a number of ways (for example, @mentions, tagging a photo, mentioning them by name, and more).

#### Has the report been filed by the target of the potential abuse or a bystander?

Some posts may seem to be abusive when viewed in isolation, but may not be when viewed in the context of a larger conversation or historical relationship between people on the platform. For example, friendly banter between friends could appear offensive to bystanders, and certain remarks that are acceptable in one culture or country may not be acceptable in another. To help prevent our teams from making a

mistake and removing consensual interactions, in certain scenarios we require a <u>report</u> from the actual target (or their authorized representative) prior to taking any enforcement action.

#### Does the user have a history of violating our policies?

We start from a position of assuming that people do not intend to violate our Rules. Unless a violation is so egregious that we must immediately suspend an account, we first try to educate people about our Rules and give them a chance to correct their behavior. We show the violator the offending post(s), explain which Rule was broken, and require them to remove the content before they can post again. If someone repeatedly violates our Rules then our enforcement actions become stronger. This includes requiring violators to remove the post(s) and taking additional actions like verifying account ownership and/or temporarily limiting their ability to post for a set period of time. If someone continues to violate Rules beyond that point then their account may be permanently suspended.

#### What is the severity of the violation?

Certain types of behavior may pose serious safety and security risks and/or result in physical, emotional, and financial hardship for the people involved. These egregious violations of the X Rules — such as posting violent threats, non-consensual intimate media, or content that sexually exploits children — result in the immediate and permanent suspension of an account. Other violations could lead to a range of different steps, like requiring someone to remove the offending post(s) and/or temporarily limiting their ability to post new post(s).

#### Is the behavior newsworthy and in the legitimate public interest?

X moves at the speed of public consciousness and people come to the service to stay informed about what matters. Exposure to different viewpoints can help people learn from one another, become more tolerant, and make decisions about the type of society we want to live in.

To help ensure people have an opportunity to see every side of an issue, there may be the rare occasion when we allow controversial content or behavior which may otherwise violate our Rules to remain on our service because we believe there is a legitimate public interest in its availability. Each situation is evaluated on a case by case basis and ultimately decided upon by a cross-functional team.

Some of the factors that help inform our decision-making about content are the impact it may have on the public, the source of the content, and the availability of alternative coverage of an event.

<u>Public impact of the content:</u> A topic of legitimate public interest is different from a topic in which the public may be curious. We will consider what the impact is to citizens if they do **not** know about this content. If the post does have the potential to impact the lives of large numbers of people, the running of a country, and/or it speaks to an important societal issue then we may allow the the content to remain on the service. Likewise, if the impact on the public is minimal we will most likely remove content in violation of our policies.

<u>Source of the content:</u> Some people, groups, organizations and the content they post on X may be considered a topic of legitimate public interest by virtue of their being in the public consciousness. This does not mean that their posts will always remain on the service. Rather, we will consider if there is a legitimate public interest for a particular post to remain up so it can be openly discussed.

Availability of coverage: Everyday people play a crucial role in providing firsthand accounts of what's happening in the world, counterpoints to establishment views, and, in some cases, exposing the abuse of power by someone in a position of authority. As a situation unfolds, removing access to certain information could inadvertently hide context and/or prevent people from seeing every side of the issue. Thus, before actioning a potentially violating post, we will take into account the role it plays in showing the larger story and whether that content can be found elsewhere.

#### Share this article



© 2024 X Corp.

Cookies (https://help.twitter.com/rules-and-policies/twitter-cookies)

Privacy (https://twitter.com/privacy)

Terms and conditions (https://twitter.com/tos)

English

# Appendix D: Changes Made to Abuse & Harassment Policy

#### Abuse and harassment Harassment

#### Overview

June 2023

March 2024

You may not share abusive content, harass someone target others with abuse or harassment, or encourage other people to do so.

X's mission is to give everyone the power to create and share ideas and information, as well as express their opinions and beliefs without barriers. Free expression is a human right – we believe that everyone has a voice, and the right to use it. Our role is to serve the public conversation, which requires representation of a diverse range of perspectives.

On X, you should feel safe expressing your unique point of view. We believe in freedom of expression and open dialogue, and in order to We recognize that if anyone, regardless of background, experiences harassment on X, it can jeopardize their ability to express themselves and cause harm. To facilitate healthy dialogue on the platform, and empower individuals to express diverse opinions and beliefs, we prohibit behavior and content that harasses, shames, or degrades others. In addition to posing risks to people's safety, abusive these types of behavior may also lead to physical and emotional hardship for those affected.

#### What is in violation of this policy?

#### **Targeted harassment**

We consider targeted behavior as prohibit behavior and content that harasses, shames, or degrades others, as defined below. Please note that, in order to help our teams understand the context, we sometimes need to hear directly from the person being targeted to ensure that we have the necessary information prior to taking appropriate and proportionate enforcement action.

#### **Targeted Harassment**

<u>We prohibit the</u> malicious, unreciprocated, <u>and intended targeting (such as mentioning or tagging) of individual(s)</u>, <u>particularly when shared</u> to humiliate or degrade <del>an individual(s)</del>. We prohibit the following behavior on the platform someone. This can mean:

- Sharing multiple Posts, over a short period of time, or continuously posting replies with malicious content, to target an individual. This includes accounts dedicated to harassing an individual or multiple individuals.
- Mentioning or tagging users with malicious content.

#### Encouraging or calling for others to harass an individual or group of people

#### **Violent Event Denial**

We prohibit content that denies that mass murder or other mass casualty events took place, where we can verify that the event occurred, and when the content is shared with abusive context. This may include references to such an event as a "hoax" or claims that victims or survivors are fake or "actors." It includes, but is not limited to, events like the Holocaust, school shootings, terrorist attacks, and natural disasters.

#### **Incitement of Harassment**

We prohibit behavior that encourages others to harass or target specific individuals or groups of people with abuse. This includes, but is not limited to: calls to target people with abuse or harassment online and behavior that urges offline action, such as physical harassment.

#### **Unwanted Sexual Content & Graphic Objectification**

While some <u>consensual nudity and adult content is permitted</u> on X, we prohibit unwanted sexual conduct and graphic objectification that sexually objectifies an individual without their consent. This includes, but is not limited to:

- sending someone unsolicited and/or unwanted adult media, including (images, videos, and GIFs;)
- unwanted sexual discussion of someone's body;
- solicitation of sexual acts; and
- any other content that otherwise sexualizes an individual without their consent.

#### Insults

We take action against the use of insults or profanity to target others. In some cases, such as (but not limited to) severe, repetitive usage of insults or profanity where the context is to harass or intimidate others, we may require Post removal. In other cases, such as (but not limited to) moderate, isolated usage of insults and profanity where the context is to harass or intimidate others, we may limit Post visibility as further described below. Please also note that However, while some individuals may find certain terms to be offensive, we will not take action against every instance where insulting terms are used.

#### **Violent Event Denial**

We prohibit content that denies that mass murder or other mass casualty events took place, where we can verify that the event occurred, and when the content is shared with abusive context. This may include references to such an event as a "hoax" or claims that victims or survivors are fake or "actors." It includes, but is not limited to, events like the Holocaust, school shootings, terrorist attacks, and natural disasters.

# Do I need to be the target of this content for it to be reviewed for violating the X Rules?

To help our teams understand the context, we sometimes need to hear directly from the person being targeted to ensure that we have the necessary information prior to taking any enforcement action. Some Posts may appear to be abusive or harassing when viewed in isolation, but may not be when viewed in the context of a larger conversation. When we review cases, it may not be clear whether the context is to abuse an individual, or if it is Excluding the Posts from having ads adjacent to it

#### **Use of Prior Names and Pronouns**

Where required by local laws, we will reduce the visibility of posts that purposefully use different pronouns to address someone other than what that person uses for themselves, or that use a previous name that someone no longer goes by as part of their transition. Given the complexity of determining whether such a violation has occurred, we must always hear from the target to determine if a violation has occurred.

#### What is not in violation of this policy?

Some posts may appear to be harmful when viewed in isolation, but may not be when viewed in the context of a larger conversation. For example, friends may consensually use certain terms or phrases to engage with each other that could appear abusive without this context. We also recognise that our platform can be used to call attention to, condemn, or highlight others for harmful behaviors. In such cases, we do not take action where the context is clearly non-abusive and intended to counter these kinds of rhetoric.

We also believe that criticism of institutions, practices and ideas is a fundamental part of the freedom of expression and thus we will not take action on such critical commentary.

#### Who can report violations of this policy?

Anyone can report violations of this policy using our dedicated reporting flow. However, we sometimes need to hear directly from the person being targeted to ensure that we have the information needed prior to taking any enforcement action.

#### What happens if you violate this policy?

When determining the penalty for violating this policy, we consider a number of factors including, but not limited to the severity of the violation, whether someone has been targeted (such as being mentioned, referred to by full name, referenced with a photo, etc), and an individual's previous record of rule violations. The following is a list of potential enforcement options for content that violates this policy:

- Making content on X less visibe visible on X by:
  - Removing the Postpost from search results, in-product recommendations, trends, notifications, and home timelines
  - Restricting the Post's post discoverability to the author's profile
  - Downranking the Postpost in replies
  - Restricting <u>Likeslikes</u>, replies, <u>Reposts</u>, <u>Quotereposts</u>, <u>quote</u>, bookmarks, share, pin to profile, <u>or engagement counts</u>, <u>or Edit Post</u>
- <u> ← Excluding the Postpost</u> from having ads adjacent to it
- Excluding Postsposts and/or accounts in email or in-product recommendations.
- Requiring Postpost removal.

- For example, we may ask someone to remove the violating content and serve a period of time in read-only mode before they can <u>Postpost</u> again.
   <u>Subsequent violations may lead to account suspension.</u>
- Suspending accounts for those whose sole purpose is to violate our Unwanted Sexual Content & Graphic Objectification policy, or accounts that are dedicated to harassing individuals.

Learn To learn more about, please see our our range of enforcement options. If range of enforcement options, and if someone believes their account was suspended enforced against in error, they can submit an appeal.