

Self-Harm Policies and Internet Platforms

Shelby Perkins, Elena Cryst, and Shelby Grossman

We thank David Thiel for technical assistance.

Stanford Internet Observatory
April 13, 2021



Contents

1	Introduction	2
1.1	Key Findings	2
2	Why Self-Harm Policies are Complicated	4
3	Platform Policies	4
3.1	Search Engines	5
3.2	Social Networks	6
3.3	Creator Platforms	9
3.4	Gaming	11
3.5	Dating	12
3.6	Chat	13
3.7	Other	14
4	Conclusion	16

A Note:

This post discusses suicide and other forms of self-harm in the context of online platform policies. In the United States, the National Suicide Prevention Lifeline is available 24/7 in English at 1-800-273-8255 and in Spanish at 1-888-628-9454. It offers Tele-Interpreter services in over 150 additional languages.

1 Introduction

Online platforms, such as social media and search engines, can be both positive and negative forces for an individual experiencing thoughts about suicide, self-injury, eating disorders, or other forms of self-harm.¹ The platform may point individuals to helpful resources such as emergency help lines; it may also alert friends and family that an individual is at risk and provide space to reach out and help. But platforms can also direct individuals to content that encourages and glorifies self-harm.

What are platforms' publicly known policies related to self-harm content? In this report we collect information on these policies for 39 online platforms, including search engines, social media networks, **creator platforms**, gaming platforms, dating apps, and chat apps. We code the policies and rank platforms based on policy comprehensiveness across various categories. We believe we are the first to collect this information systematically. It should be noted that a platform simply having a policy does not mean the policy cannot be improved. Our framework initially looks at whether policies exist and whether policies comprehensively address the spectrum of harm from content about suicide, self-injury, and eating disorders.

We want to address two possible critiques head on. First, a critic may argue that it's not important for a platform to have a public policy around self-harm content, as long as it has an internal one. We argue that having a public policy is important for platform users and civil society. When an individual's content is removed, labeled, or actioned in some other way, platforms with a public-facing policy can link to the policy to help explain to the individual why the platform took action. This helps ensure the individual is not confused about the reason, provides the opportunity for the individual to better educate themselves about the platforms' policies, and increases the likelihood that they abide by these policies going forward. Additionally, policies that link to support services can reduce the barriers an individual may face in their search for assistance. Public-facing policies also allow civil society groups such as self-harm prevention organizations to understand how a platform addresses self-harm content. This, in turn, allows the prevention organization to know if the platform policies reflect the latest research in this field and whether advocacy may be appropriate. While many platforms appear to actively work with prevention organizations in developing their policies, this is likely not the case for all platforms. Public-facing policies also empower civil society groups to hold platforms to their own stated policies.

Second, a critic may argue that a platform could have an exceptionally thoughtful policy that is never implemented. What, then, is the point of focusing on the policy? We agree that implementation also matters. We are working on a longer-term project to systematically collect data on search results and when "knowledge panels" with hotline numbers do and do not appear.

We realize that different platforms have different risks and policy needs: comparing platforms like Bing's search engine to platforms like Hinge's dating app does not make sense. As such, we adjust our metrics slightly across platforms, and provide only apples-to-apples comparisons in our rankings.

1.1 Key Findings

Our key findings are as follows:

¹For the purposes of this paper, we use the term "self-harm" to refer to suicide; suicide attempts or ideation; self-injury, such as cutting; eating disorders; and other forms of physical harm to oneself. The platforms themselves may use "self-harm" as a general term or to indicate "self-injury" specifically.

- **There is vast unevenness in the comprehensiveness of public-facing policies.** For example, Facebook policies address not only suicide but also euthanasia, suicide notes, and livestreaming suicide attempts. In contrast, Instagram and Reddit have no policies related to suicide in their primary policy documents.
 - Search engines, including Google and Bing, lack public-facing policies for how they handle suicide content. While Google and Bing certainly have policies on these topics, for example frequently linking to hotlines at the top of self-harm search results, these policies are not public.
 - What we term **creator platforms** (TikTok, Twitch and YouTube) perform relatively well on our ratings.
- **Policy information can be difficult to locate.** Platforms sometimes provide detailed updates on how and when they enforce community guidelines in blog posts, but fail to change the official community guidelines. This leaves users searching for relevant information that should be in one location. This point has been made by the **Facebook Oversight Board** as well.
- **Only a handful of platforms' public-facing policies include information about which self-harm prevention resources they direct users to.** These resources can lower barriers for users' seeking help.
- **Only a few platforms allude to downranking self-harm content, and those that do are not specific.**
- **Policies around suicide, self-injury, and eating disorders are complicated.** First, there is a fine line between content used by support communities and content encouraging further harm, making automatic moderation difficult. Second, new technologies, particularly livestreams, require unique forms of moderation. Third, it can be difficult to take into account local norms across countries.

Table 1 summarizes our ratings of platforms' self-harm policies. We expand on our findings for each platform in the report.

Platform	Score	Platform	Score
Search Engines		Dating	
Google	1/3	Grindr	2/3
Bing	0/3	Tinder	2/3
Yandex	0/3	SCRUFF	0/3
Baidu	0/3	Her	0/3
DuckDuckGo	0/3	Hinge	0/3
Social Networks		Chat	
Facebook	5/5	Kik	3/3
Twitter	3/5	Discord	3/3
Instagram	3/5	LINE	2/3
Reddit	0/5	WeChat	1/3
Parler	0/5	iMessage	0/3
Gab	0/5	Signal	0/3
Creator Platforms		WhatsApp	0/3
TikTok	5/6	Telegram	0/3
YouTube	4/6	Other	
Twitch	3/6	Pinterest	3/3
Gaming		Tumblr	3/3
Roblox	1/3	Snapchat	2/3
Epic Games	1/3	Nextdoor	0/3
PlayStation Network	1/3	Venmo	0/3
Xbox	0/3	4chan	0/3
Minecraft	0/3	Clubhouse	0/3

Table 1: Self-harm policy ratings across platforms.

In this paper, we first outline some reasons why designing policies around self-harm is complicated, and then proceed to assess the comprehensiveness of platform policies.

2 Why Self-Harm Policies are Complicated

Like most content moderation policies, reasonable people can disagree about what types of policies about self-harm content platforms should have. In this section we discuss four topics that make self-harm policies difficult to develop: support groups, stated intent to self-harm, norms across countries, and enforcement actions.

First, many individuals who are in recovery from self-harm activities, such as eating disorders, participate in online support communities. An individual may for example share a photo montage showing their prior lowest weight and their new healthy weight. On the one hand, such a post may provide solace to others struggling with eating disorders, showing them that recovery is possible. On the other hand, research shows that seeing low weight numbers can lead to a “**race to the bottom**.” How should platforms balance letting people share their story with limiting content that could inadvertently encourage self-harm?

Second, there is the question of whether content showing an individual making a credible threat to kill themselves should be removed. This can be particularly challenging in the context of livestreams. Facebook currently has the **most detailed policy** on livestreaming and suicide, a policy that reflects thoughtful consideration of tradeoffs:

With respect to live content, experts have told us that if someone is saying they intend to attempt suicide on a livestream, we should leave the content up for as long as possible, because the longer someone is talking to a camera, the more opportunity there is for a friend or family member to call emergency services.

However, to minimize the risk of others being negatively impacted by viewing this content, we will stop the livestream at the point at which the threat turns into an attempt. As mentioned above, in any case, we will contact emergency services if we identify someone is at immediate risk of harming themselves.

Third, many of the platforms discussed in this report are available in almost all countries, and norms related to what is considered promoting self-harm vary greatly. Should platform policies vary by country? While implementation is not the focus of this report, we note that prior Stanford Internet Observatory **research** has shown gaps in policy implementation across languages. For example, knowledge panels linking to suicide-prevention hotlines appear on Google for suicide search terms in English, but not Japanese.

Last, self-harm content is different from other types of online abuse in that many users posting the content are not malicious actors. In this light, platforms can provide helpful resources to these users. Resources such as the National Suicide Prevention Lifeline or an attempt to contact local emergency services could save an individual’s life. However, many platforms simply moderate the content as they would any other type of harmful material, such as harassment, spam, or threats of harm to another. Providing helpful resources instead of or in addition to removing content can make moderation more complicated, but is an important step.

3 Platform Policies

We proceed to evaluate platform policies for the following platform categories: search engines, social networks, creator platforms, gaming platforms, dating apps, chat apps, and “other” platforms. Because each platform category operates differently, we have defined separate assessment rubrics for each category.

3.1 Search Engines

We assess self-harm content-related policies for five popular search engines: Google, Bing, Yandex (a popular search engine in Russia), Baidu (a popular search engine in China), and DuckDuckGo. We evaluate them based on three categories:

1. Is there a policy on suicide?
2. Is there a policy on self-injury?
3. Is there a policy on eating disorders?

Findings are shown in Table 2. We note that in this section, and throughout, we code a platform as having a policy on suicide only if the word suicide is used. If only the term “self-harm” is used, we put the policy under “self-injury.” We encourage platforms that use the term self-harm to define it, as it was often unclear whether its use encompassed suicide.

The only search engine for which we were able to find a self-harm content-related policy was Google. Google has general terms of service, and many of its products have service-specific terms of service. However, there are no **service-specific terms of service for Google Search**. The general terms of service state, “Don’t abuse or harm others or yourself,” but they do not speak to Google’s policy for what content, if any, it may remove from Search or prevent from appearing high in search rankings.

	Google	Bing	Yandex	Baidu	DuckDuckGo
Policy on Suicide	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.
Policy on Self-Injury	Don’t abuse or harm others or yourself (or threaten or encourage such abuse or harm) — for example, by misleading, defrauding, defaming, bullying, harassing, or stalking others.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.
Policy on Eating Disorders	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.
Location	Terms of Service	Microsoft Services Agreement	Terms of Service	No relevant terms of service found.	No relevant terms of service found.
Score	1/3	0/3	0/3	0/3	0/3

Table 2: Search engine policies.

3.2 Social Networks

In Tables 3, 4, and 5 we assess the presence of self-harm content-related policies for Facebook, Twitter, Instagram, Reddit, Parler, and Gab. We evaluate them based on five categories:

1. Is there a policy on suicide?
2. Is there a policy on suicide notes?
3. Is there a policy on self-injury?
4. Is there a policy on eating disorders?
5. Is it clear what resources the platform provides to a user when this content is detected?

Facebook has policies addressing all five categories, while Twitter and Instagram each address three categories. Reddit, Parler, and Gab lack policies that cover any of the categories. Facebook’s policies explicitly address the blurry line between support and promotion content, as discussed in Section 2, **stating**: “we do allow people to discuss [suicide and self-injury] because we want Facebook to be a space where people can share their experiences, raise awareness about these issues, and seek support from one another.”

One issue, which we have seen around policies on other types of abuses, is whether the platform policies are thoroughly listed in a platform’s community guidelines. For example, based on our interpretation, the **Instagram Community Guidelines** have much to improve upon. The **Facebook Oversight Board** has noted that Facebook informed them that Facebook’s Community Standards apply to Instagram, and indeed one word in Instagram’s Community Guidelines about self-harm hyperlinks to Facebook’s Community Standards. However, given the importance of a clear public-facing policy, as we outlined in Section 1, we do not believe this hyperlink is sufficient to convey to Instagram users that Facebook’s Community Standards apply. We strongly agree with the Facebook Oversight Board’s **recommendation** that Facebook “clarify that where there are inconsistencies between Instagram’s Community Guidelines and Facebook’s Community Standards, the latter take precedence.” Similarly, we believe Instagram users should be able to find all self-harm policies in one place. For example, Instagram could link to relevant blog posts from their Community Guidelines, such as Instagram’s thoughtful “**Changes We’re Making to Do More to Support and Protect the Most Vulnerable People who Use Instagram**” post. This would clarify policy purpose, scope, and enforcement. Reddit also has a blog post, “**Reddit’s Partnership with Crisis Text Line**,” that is not linked from their Content Policy.

We were not able to find any applicable policies for Reddit, Parler,² and Gab. Parler’s new **Community Guidelines** state, “Parler’s viewpoint-neutral policies foster a community of individuals who tolerate the expression of all non-violent ideas.” Suicide-related content is often violent content, and as such may fall outside of tolerated expression. However, we do not find this sufficient to merit being called a “policy.”

²We assessed Parler’s policies both in December 2020 before it went offline, and in February 2021 after it found new web hosting.

	Facebook	Twitter	Instagram	Reddit	Parler	Gab
Policy on Suicide	<p>We remove any content that encourages suicide or self-injury, including fictional content such as memes or illustrations and any self-injury content which is graphic, regardless of context. We also remove content that identifies and negatively targets victims or survivors of suicide or self-injury seriously, humorously or rhetorically, as well as real time depictions of suicide or self-injury. Content about recovery of suicide or self-harm that is allowed, but may contain imagery that could be upsetting, such as a healed scar, is placed behind a sensitivity screen.</p> <p>With respect to live content, experts have told us that if someone is saying they intend to attempt suicide on a livestream, we should leave the content up for as long as possible, because the longer someone is talking to a camera, the more opportunity there is for a friend or family member to call emergency services.</p> <p>However, to minimize the risk of others being negatively impacted by viewing this content, we will stop the livestream at the point at which the threat turns into an attempt. As mentioned above, in any case, we will contact emergency services if we identify someone is at immediate risk of harming themselves.</p> <p>Do not post: Content that promotes, encourages, coordinates, or provides instructions for: Suicide; Content that mocks victims or survivors of suicide, self-injury or eating disorders who are either publicly known or implied to have experienced suicide or self-injury</p> <p>Except in limited situations of newsworthiness, it is against our policies to post content depicting a person who engaged in a suicide attempt or death by suicide.</p>	<p>You may not promote or encourage suicide or self-harm.</p> <p>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 Tweets, images or videos, including live video.</p> <p>We define suicide to be the act of taking one's own life.</p> <p>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.</p>	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.
Policy on Suicide Notes	<p>For the following Community Standards, we require additional information and/or context to enforce: We may remove suicide notes when we have confirmation of a suicide or suicide attempt. We try to identify suicide notes using several factors, including but not limited to, family or legal representative requests, media reports, law enforcement reports or other third party sources (e.g., government agencies, NGOs).</p>	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.

Table 3: Social media policies (1 of 3).

	Facebook	Twitter	Instagram	Reddit	Parler	Gab
Policy on Self-Injury	<p>We define self-injury as the intentional and direct injuring of the body, including self-mutilation and eating disorders. We remove any content that encourages suicide or self-injury, including fictional content such as memes or illustrations and any self-injury content which is graphic, regardless of context. We also remove content that identifies and negatively targets victims or survivors of suicide or self-injury seriously, humorously or rhetorically, as well as real time depictions of suicide or self-injury. Content about recovery of suicide or self-harm that is allowed, but may contain imagery that could be upsetting, such as a healed scar, is placed behind a sensitivity screen.</p> <p>Do not post: Content that promotes, encourages, coordinates, or provides instructions for: Self-injury; Content that depicts graphic self-injury imagery; Content that mocks victims or survivors of suicide, self-injury or eating disorders who are either publicly known or implied to have experienced suicide or self-injury</p>	<p>You may not promote or encourage suicide or self-harm.</p> <p>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".</p> <p>Violations of this policy can occur via Tweets, images or videos, including live video.</p> <p>We define self-harm to include: self-inflicted physical injuries e.g., cutting; and eating disorders e.g., bulimia, anorexia.</p> <p>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.</p>	<p>Encouraging or urging people to embrace self-injury is counter to this environment of support, and we'll remove it or disable accounts if it's reported to us. We may also remove content identifying victims or survivors of self-injury if the content targets them for attack or humor.</p>	No policy could be found.	No policy could be found.	No policy could be found.
Policy on Eating Disorders	<p>Do not post:</p> <p>Content that focuses on depiction of ribs, collar bones, thigh gaps, hips, concave stomach, or protruding spine or scapula when shared together with terms associated with eating disorders.</p> <p>Content that contains instructions for drastic and unhealthy weight loss when shared together with terms associated with eating disorders. Content that mocks victims or survivors of suicide, self-injury or eating disorders who are either publicly known or implied to have experienced suicide or self-injury</p> <p>Content that mocks victims or survivors of suicide, self-injury or eating disorders who are either publicly known or implied to have experienced suicide or self-injury</p>	See cell above.	<p>The Instagram community cares for each other, and is often a place where people facing difficult issues such as eating disorders, cutting, or other kinds of self-injury come together to create awareness or find support [...]</p> <p>Encouraging or urging people to embrace self-injury is counter to this environment of support, and we'll remove it or disable accounts if it's reported to us. We may also remove content identifying victims or survivors of self-injury if the content targets them for attack or humor.</p>	No policy could be found.	No policy could be found.	No policy could be found.

Table 4: Social media policies (2 of 3).

	Facebook	Twitter	Instagram	Reddit	Parler	Gab
Resource Provided when Content Detected	When people post or search for suicide or self-injury-related content, we will direct them to local organizations that can provide support and if someone is at immediate risk of harming themselves, we will contact local emergency services to get them help.	No policy could be found.	We try to do our part by providing education in the app and adding information in the Help Center so people can get the help they need.	No policy could be found.	No policy could be found.	No policy could be found.
Location	Community Standards	Rules and Policies	Community Guidelines	Content Policy	No applicable terms of service found.	Terms of Service
Score	5/5	3/5	3/5	0/5	0/5	0/5
Topic Specific Enforcement Actions	<p>For the following content, we restrict content to adults over the age of 18, and include a sensitivity screen so that people are aware the content may be upsetting:</p> <ul style="list-style-type: none"> • Photos or videos depicting a person's death by suicide that are determined to be newsworthy • Photos or videos depicting a person who engaged in euthanasia/assisted suicide in a medical setting <p>For the following content, we include a sensitivity screen so that people are aware the content may be upsetting to some:</p> <ul style="list-style-type: none"> • Content that depicts older instances of self-harm such as healed cuts or other non-graphic self-injury imagery in a context of recovery • Content that depicts ribs, collar bones, thigh gaps, hips, concave stomach, or protruding spine or scapula in a recovery context. <p>We provide resources to people who post written or verbal admissions of engagement in self injury, including:</p> <ul style="list-style-type: none"> • Suicide • Euthanasia/assisted suicide • Self-harm • Eating disorders 	<p>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.</p> <p>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 Tweet 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. If you believe that your account was suspended in error, you can submit an appeal.</p> <p>We may also take steps to prevent the spread of instructional material hosted on third-party websites by marking such links as unsafe.</p>	<p>Encouraging or urging people to embrace self-injury is counter to this environment of support, and we'll remove it or disable accounts if it's reported to us. We may also remove content identifying victims or survivors of self-injury if the content targets them for attack or humor.</p>	No policy could be found.	No policy could be found.	No policy could be found.

Table 5: Social media policies (3 of 3).

3.3 Creator Platforms

We term our third category “creator platforms.” These include YouTube, TikTok, and Twitch. We evaluate these platforms based on six criteria:

1. Is there a policy on suicide that applies to content creators?
2. Is there a policy on suicide notes that applied to content creators?
3. Is there a policy on self-injury that applies to content creators?
4. Is there a policy on eating disorders that applies to content creators?
5. Do the same policies apply to non-creators on the platform?
6. Is it clear what resources the platform provides to a user when this content is detected?

TikTok has policies that cover five categories, the most among creator platforms, while YouTube and Twitch cover four and three categories, respectively. We observed many noteworthy trends: All platforms in this category are clear that policies apply to both creators and non-creators on the platform. Additionally, both YouTube and TikTok are explicit in allowing creators to share their stories about self-harm to raise awareness and find community support. We were impressed that YouTube's

Community Guidelines on **suicide and self-injury** provide resources, including hotlines and websites, for those having thoughts of suicide or self-harm, for 27 countries.

At the same time, none of the creator platforms are clear on whether they provide resources in addition to or instead of standard enforcement actions such as removal or suspension. Detailing this policy is especially important for platforms that offer livestreaming capabilities, as all of the creator platforms do. Additionally, although all platforms in this category have policies on content about suicide, none have policies about suicide notes. It is not clear, for example, whether a suicide note would fall under “promoting or glorifying suicide” in YouTube’s current policy.

	YouTube	TikTok	Twitch
Policy on Suicide for Creators	<p>YouTube users should not be afraid to speak openly about the topics of mental health or self-harm. However please don't post content on YouTube if it fits any of the descriptions noted below:</p> <ul style="list-style-type: none"> • Promoting or glorifying suicide • Providing instructions on how to self-harm or die by suicide 	<p>We do not allow content depicting, promoting, normalizing, or glorifying activities that could lead to suicide, self-harm, or eating disorders. We also do not permit users to share content depicting them partaking in, or encouraging others to partake in, dangerous activities that may lead to serious injury or death.</p> <p>However, we do support members of our community sharing their personal experiences with these issues in a safe way to raise awareness and find community support. We also encourage individuals who are struggling with thoughts of self-harm or suicide, or who know someone is seriously considering suicide, to immediately contact local emergency services or a suicide prevention hotline. In the event that our intervention could help a user who may be at risk of harming themselves, the TikTok team may also alert local emergency services.</p> <p>Suicide</p> <p>We remove content that displays suicide, suicidal ideation, or content that might encourage participation in other self-injurious behavior. We also remove content that depicts attempted suicide or content featuring a person engaging in behavior or intending to engage in behavior that is likely to lead to self-inflicted death. We prohibit any form of content that promotes, normalizes, or glorifies suicide, provides instructions for suicide, or posts that portray a suicide as heroic or honorable.</p> <p>Do not post, upload, stream, or share:</p> <ul style="list-style-type: none"> • Content that provides instructions for suicide • Content that depicts, promotes, normalizes, or glorifies suicide • Suicide games, dares, pacts, or hoaxes 	<p>Any activity that may endanger your life or lead to your physical harm is prohibited. This includes, but is not limited to: suicide threats... We do not make exceptions for self destructive behavior performed as a stunt or gag made in jest, or meant to entertain, when the behavior could reasonably be expected to cause physical injury.</p>
Policy on Suicide Notes for Creators	No policy could be found.	No policy could be found.	No policy could be found.
Policy on Self-Injury for Creators	<p>YouTube users should not be afraid to speak openly about the topics of mental health or self-harm. However please don't post content on YouTube if it fits any of the descriptions noted below:</p> <ul style="list-style-type: none"> • Graphic images of self-harm posted to shock or disgust viewers 	<p>Self-harm and eating disorders</p> <p>To avoid normalizing, encouraging, or triggering self-harm behavior, we do not allow imagery that depicts such behavior, regardless of the user's intention of posting it. We remove content that may encourage or normalize acts that are likely to lead to physical self-inflicted injury. Content that promotes eating habits that are likely to cause adverse health outcomes is also not allowed on the platform.</p> <p>Do not post, upload, stream, or share:</p> <ul style="list-style-type: none"> • Content that depicts, promotes, normalizes, or glorifies self-harm or eating disorders • Content that provides instructions on how to engage in self-harm or eating disorders • Self-harm or eating disorder games, dares, pacts, or hoaxes • Content that depicts, promotes, normalizes, or glorifies eating disorders or other dangerous weight loss behaviors associated with eating disorders 	No policy could be found.

Table 6: Creator policies (1 of 2).

	YouTube	TikTok	Twitch
Policy on Eating Disorders for Creators	Don't post content on YouTube if it fits any of the descriptions noted below. Eating Disorders: Content that praises, glorifies, or encourages viewers to imitate anorexia or other eating disorders. Eating disorders are characterized by abnormal or disturbed eating habits which negatively affect a person's health (including eating non-food items). Examples: The following types of content are not allowed on YouTube. This is not a complete list. Using terms or phrases in metadata that promote or glorify eating disorders. Videos that promote or glorify eating non food items.	See cell above.	No policy could be found.
Do these policies apply to viewers / non-creators?	This Community Guidelines policy applies to videos, video descriptions, comments, live streams, and any other YouTube product or feature. Please note this is not a complete list.	We will remove any content—including video, audio, livestream, images, comments, and text—that violates our Community Guidelines.	These guidelines...apply to all user generated content and activity on our services)
Resource Provided when Content Detected	No policy could be found.	In the event that our intervention could help a user who may be at risk of harming themselves, the TikTok team may also alert local emergency services.	No policy could be found.
Location	YouTube Policies - Suicide & Self-Injury	Community Guidelines	Community Guidelines
Score	4/6	5/6	3/6
Enforcement Actions Specific to Suicide, Self-Harm, or Eating Disorders	If your content violates this policy, we'll remove the content and send you an email to let you know. If this is your first time violating our Community Guidelines, you'll get a warning with no penalty to your channel. If it's not, we'll issue a strike against your channel. If you get 3 strikes, your channel will be terminated.	No policy could be found.	No policy could be found.

Table 7: Creator policies (2 of 2).

3.4 Gaming

Table 8 assesses the policies on five popular gaming platforms: Xbox, PlayStation Network, Minecraft, Epic Games, and Roblox. We evaluate them on three categories:

1. Is there a policy on suicide?
2. Is there a policy on self-injury?
3. Is there a policy on eating disorders?

Gaming platforms are important for this study because of the time users spend on the platforms, and because they create bridges between users of different ages. This necessitates thorough policies

around self-harm. In 2020, **32.6 million people** logged into Roblox daily, a platform where many of the developers are teenagers and young adults. Given the widespread use of gaming platforms and the community aspect of play, platforms should prioritize safe environments for users.

Xbox and Minecraft had no policies related to suicide, self-injury, or eating disorders. PlayStation Network, Epic Games, and Roblox only had policies addressing “self-harm” without specifying suicide, self-injury, or eating disorders. Roblox has the most comprehensive self-harm content-related policy, addressing both users and developers.

	Xbox	Playstation Network	Minecraft	Epic Games	Roblox
Policy on Suicide	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.
Policy on Self-Injury	No policy could be found.	Do not engage in, threaten, or promote illegal activity, terrorism, or acts of violence or self-harm.	No policy could be found.	Threats of harm to yourself or others are taken seriously—don't make them, especially as a joke.	Threats and self-harm. In the digital environment, it is hard to tell what is a joke or is serious. So, we take all threats seriously and do not allow any threats on Roblox. We also take reports of self-harm very seriously. We are in active communication with police departments across the globe and often report threats related to physical harm or public safety. Additional rules for developers Creating positive game experiences. When designing your game, focus on providing positive game experiences for our community (e.g., working in a pizza restaurant, building a theme park, etc.). You should encourage the proper use of your game and never directly or indirectly encourage inappropriate behavior. For example, experiences within a game should not have an inappropriate theme (sleepover/slumber parties), contain nude or partially clothed characters, or display suggestive animations. Below is a list of inappropriate behaviors (and please note this list is not comprehensive; you should think through if you are providing or promoting inappropriate behavior): [...] Threats and self-harm; [...]
Policy on Eating Disorders	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.
Location	Community Standards	Code of Conduct	Community Standards	Community Rules	Community Rules
Score	0/3	1/3	0/3	1/3	1/3

Table 8: Gaming policies.

3.5 Dating

Our fourth category is dating apps, including Grindr, SCRUFF, Tinder, Her, and Hinge. Within this category, it is important to include apps targeting the LGBTQ community specifically (Grindr, SCRUFF, Her) because of the **higher risks** of experiencing mental health disorders. **One in three** LGBTQ adults have experienced mental illness, compared to one in five non-LGBTQ adults. **40% of transgender adults** have attempted suicide, compared to less than 5% of the US population overall.

48% of 18 to 29 year olds in the US have used a **dating site or app**. That number is 55% for lesbian, gay, or bisexual Americans. For these reasons, we believe dating platforms should have suicide and other self-harm related policies.

We evaluate these platforms on the following three categories:

1. Is there a policy on suicide?
2. Is there a policy on self-injury?
3. Is there a policy on eating disorders?

Grindr and Tinder have policies on both suicide and self-harm, though neither have policies on eating disorders. Hinge, Scruff and Her have no policies for any of the categories.

	Grindr	Scruff	Tinder	Her	Hinge
Policy on Suicide	Hate, Violence, and offensive content depictions of self-harm or suicide can be harmful for others, so we will remove them.	No policy could be found.	Violence and Physical Harm Content that advocates for or glorifies suicide or self-harm is also not allowed.	No policy could be found.	No policy could be found.
Policy on Self-Injury	See cell above.	No policy could be found.	See cell above.	No policy could be found.	No policy could be found.
Policy on Eating Disorders	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.
Location	Community Guidelines	Profile Guidelines	Community Guidelines	Community Guidelines	Terms of Service
Score	2/3	0/3	2/3	0/3	0/3

Table 9: Dating policies.

3.6 Chat

Our fourth category is chat platforms, including Kik, iMessage, Discord, Signal, WhatsApp, WeChat, LINE, and Telegram. We evaluate these platforms on the following three categories:

1. Is there a policy related to suicide?
2. Is there a policy on self-injury?
3. Is there a policy on eating disorders?

Kik and Discord have policies addressing all three categories while LINE addresses suicide and self-harm and WeChat addresses self-harm, without specifying what types of self-harm. Signal, iMessage, WhatsApp, and Telegram have no policies.

	Kik	iMessage	Discord	Signal	WhatsApp	WeChat	Line	Telegram
Policy on Suicide	Can I promote self-harm? Nope. We want you and others to be safe. Do not show or promote self-harm—including cutting, eating disorders, suicide, or sharing information on this topic in ways that promotes it's OK.	No policy could be found.	You may not share content that glorifies or promotes suicide or self-harm, including any encouragement to others to cut themselves, or embrace eating disorders such as anorexia or bulimia.	No policy could be found.	No policy could be found.	No policy could be found.	LINE prohibits Users from engaging in any of the following acts when using the Services: expressions that induce or encourage suicide, self-injurious behavior or drug abuse	No policy could be found.
Policy on Self-Injury	See above.	No policy could be found.	See above.	No policy could be found.	No policy could be found.	You agree not to engage in any of the following prohibited activities on or in relation to WeChat, or allow any person to use your account with us to do the same: promotes or encourages self-harming	LINE prohibits Users from engaging in any of the following acts when using the Services: expressions that induce or encourage suicide, self-injurious behavior or drug abuse	No policy could be found.
Policy on Eating Disorders	See above.	No policy could be found.	See above.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.
Location	Community Standards	No relevant terms of service.	Community Guidelines	Terms of Service	Terms of Service	Acceptable Use Policies	Terms and Conditions of Use	Terms of Service
Score	3/3	0/3	3/3	0/3	0/3	1/3	2/3	0/3

Table 10: Chat policies.

3.7 Other

The “Other” category contains platforms that did not fit neatly into the previous categories but which are still of interest for a variety of reasons. For example, Clubhouse is a new app that uses audio chat rooms to form communities, raising challenges for live content moderation. Venmo is a popular money-transfer platform that could be used to signal a pending suicide. We note that it is difficult to compare these services to each other as they offer different services. We evaluate these platforms on the following three categories:

1. Is there a policy on suicide?
2. Is there a policy on self-injury?

3. Is there a policy on eating disorders?

Pinterest and Tumblr have policies addressing all three categories and Snap Inc. has one policy addressing both self-injury and eating disorders. Nextdoor, Venmo, 4chan, and Clubhouse have no relevant policies.

	Pinterest	NextDoor	Snap Inc.	Venmo	4chan	Clubhouse	Tumblr
Policy on Suicide	<p>Pinterest isn't a place for content that displays, rationalizes or encourages suicide, self-injury, eating disorders or substance abuse. We'll limit the distribution of or remove such content, including:</p> <ul style="list-style-type: none"> • Self-harm instructions • Suicidal thinking and quotes • Graphic or otherwise triggering imagery or descriptions of self-harm • Promotion of self-harm • Mocking of people who self-harm or who have attempted or died by suicide • Images of accessories used to self-harm • Negative self-talk and insensitive humor about self-harming behavior • Suicide pacts, challenges and hoaxes <p>If you or someone you know is struggling with suicidal thoughts or is considering hurting themselves, we have resources in our Help Center where you can get free, confidential and immediate support.</p>	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.	No policy could be found.	<p>Don't post content that actively promotes or glorifies self-harm. This includes content that urges or encourages others to: cut or injure themselves; embrace anorexia, bulimia, or other eating disorders; or commit suicide rather than, e.g., seeking counseling or treatment, or joining together in supportive conversation with those suffering or recovering from depression or other conditions. Dialogue about these behaviors is incredibly important and online communities can be extraordinarily helpful to people struggling with these difficult conditions. We aim for Tumblr to be a place that facilitates awareness, support and recovery, and we will remove only those posts or blogs that cross the line into active promotion or glorification of self-harm.</p>
Policy on Self-Injury	See above	No policy could be found.	We don't allow the glorification of self-harm, including the promotion of self-injury or eating disorders.	No policy could be found.	No policy could be found.	No policy could be found.	See above.
Policy on Eating Disorders	See above.	No policy could be found.	See above.	No policy could be found.	No policy could be found.	No policy could be found.	See above.
Location	Community Guidelines	Community Guidelines	Community Guidelines	Acceptable Use	Rules	Community Guidelines	Community Guidelines
Score	3/3	0/3	2/3	0/3	0/3	0/3	3/3

Table 11: "Other" policies.

4 Conclusion

It is clear after analyzing self-harm content-related policies across dozens of online platforms that many of these platforms have policy gaps, at least in their public-facing policies. We encourage platforms to make as much of their policies public as possible, for the reasons outlined in Section 1. For platforms that do not yet have self-harm policies, we encourage them to develop policies, ideally in collaboration with self-harm prevention groups.

A note to platforms: If you believe we made an error describing your policies, or want to update us on new policies, please email us at internetobservatory@stanford.edu.

The Stanford Internet Observatory is a cross-disciplinary program of research, teaching and policy engagement for the study of abuse in current information technologies, with a focus on social media. The Observatory was created to learn about the abuse of the internet in real time, and to translate our research discoveries into training and policy innovations for the public good.

Stanford | Internet Observatory

Cyber Policy Center

