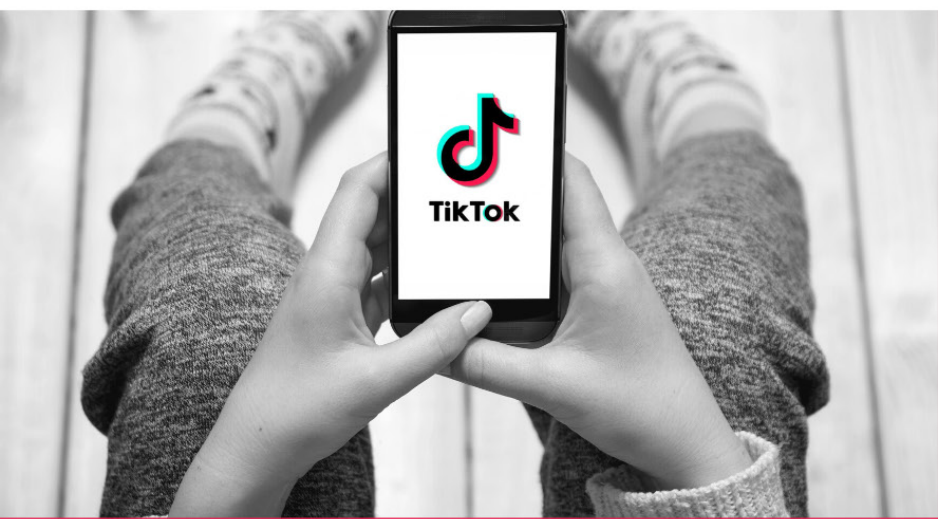# CC DH Center for Countering Digital Hate

DEADLY BY DESIGN

TikTok pushes harmful content promoting eating disorders and self-harm into users' feeds.

The Center for Countering Digital Hate is a US-headquartered international nonprofit NGO that disrupts the architecture of online hate and misinformation.

Digital technology has forever changed how we communicate, build relationships, share knowledge, set social standards, and negotiate and assert our societies' values.

Digital spaces have been colonized, and their unique dynamics exploited by malignant actors that instrumentalize hate and misinformation. These movements are opportunistic, agile, and confident in influencing and persuading people.

Over time these malignant actors, advocating diverse causes – from hatred of women to racial and religious intolerance to denial of science – have formed a digital Counter-Enlightenment. The disinformation they spread to bolster their causes has socialized the offline world for the worse.

The Center's work combines both analysis and disruption of these networks. CCDH's solutions seek to increase the economic, political, and social costs of all parts of the infrastructure – the actors, systems, and culture – that support and profit from hate and misinformation.

If you appreciate this report, you can donate to CCDH at counterhate.com/donate. In the United States, Center for Countering Digital Hate Inc is a 501(c)(3) charity. In the United Kingdom, Center for Countering Digital Hate Ltd is a non-profit company limited by guarantee.

Contents

Please Read

This report contains content on the following themes which may be distressing to readers:

• Eating disorders
• Suicide
• Self-harm
• Mental health
• Sexual assault
• Fat-shaming

If you are affected by the content of this report, you may find the following resources helpful.

NEDA — US eating disorder support and advice.
To reach a helpline call 800 931 2237 from 11am – 9pm ET Monday to Thursday, and from 11am – 5pm ET on Friday. To access web chat support use this link between 9am – 9pm ET on Monday to Thursday, and 9am – 5pm on Friday.

BEAT — UK eating disorder support and advice.
To reach a helpline use this link to find phone numbers for England, Scotland, Wales and Northern Ireland, 365 days a year, 9am – midnight during the week and 4pm – midnight on weekends. For 24 hour web chat support use this link.

Butterfly — Australian eating disorder support and advice.
To reach a helpline call 1800 33 4673, available from 8am – midnight seven days a week.

Nedic — Canadian eating disorder support and advice.
For their helpline call 1866 633 4220 from 9am – 9pm Monday to Thursday, and 9am – 5pm on Friday. For web chat support use this link at the same times above, and between 1pm – 7pm on weekends.

This report examines TikTok's role in recommending eating disorder and mental health content to vulnerable users. Our findings should not be interpreted as a criticism of those TikTok users who use the app to discuss their own mental health.

1 Introduction

Two-thirds of American teenagers use TikTok, and the average viewer spends 80 minutes a day on the application.[1] The app, which is owned by the Chinese company, Bytedance, rapidly delivers a series of short videos to users and has overtaken Instagram, Facebook, and YouTube in the bid for young people's hearts, minds, and screen time.

And yet most people understand very little about how TikTok works or the potential dangers of the platform. Journalists love to talk about Twitter, their platform of choice. Facebook remains the most used platform worldwide, giving politicians, brands, and bad actors an unparalleled pool of potential users to target, and it has received proportionate scrutiny. But TikTok reveals a generational gap in usage and understanding. This report seeks to break down those barriers and give parents and policymakers insight into the content and algorithms shaping young lives today.

For our study, Center for Countering Digital Hate researchers set up new accounts in the United States, United Kingdom, Canada, and Australia at the minimum age TikTok allows, 13 years old. These accounts paused briefly on videos about body image and mental health, and liked them. What we found was deeply disturbing. Within 2.6 minutes, TikTok recommended suicide content. Within 8 minutes, TikTok served content related to eating disorders. Every 39 seconds, TikTok recommended videos about body image and mental health to teens. The results are every parent's nightmare: young people's feeds are bombarded with harmful, harrowing content that can have a significant cumulative impact on their understanding of the world around them, and their physical and mental health.

TikTok operates through a recommendation algorithm that constructs a personalized endless-scroll 'For You' feed, ostensibly based on the likes, follows, watch-time, and interests of a user. CCDH researchers created "standard" and "vulnerable" accounts in the geographies covered. Research has indicated that users who seek out content about eating disorders will often choose usernames with related language; thus, our "vulnerable" accounts contained the term "loseweight" in their usernames. TikTok identifies the user's vulnerability and capitalizes on it. The vulnerable accounts in our study received 12 times more recommendations for self-harm and suicide videos than the standard accounts. Young people who engage with this content are left to cope with a staggering onslaught of more and more recommended videos in their feeds.

This year, for the first time, a Coroner's inquest in the United Kingdom ruled that social media platforms contributed to the suicide of 14-year-old Molly Russell. Molly had liked, shared, or saved 2,100 posts related to suicide, self-harm, or depression on Instagram in the 6 months before her death.[2] Molly's inquest has shown that Big Tech's negligence has real, life-altering consequences – and that comprehensive regulation is needed to protect children online.

This year, Tiktok's Chief Operating Officer, Vanessa Pappas, testified before the Senate Homeland Security and Government Affairs Committee.[3] She stated that safety was a "priority" for her company and that the mission of TikTok was "to inspire creativity and bring joy." Her assurances of transparency and accountability are buzzword-laden empty promises that legislators, governments, and the public have all heard before. CCDH researchers found a community for eating disorder content on the platform, amassing 13.2 billion views across 56 hashtags often designed to evade moderation. Rather than entertainment and safety, our findings reveal a toxic environment for TikTok's youngest users, intensified for its most vulnerable.

This report underscores the urgent need for reform of online spaces. CCDH's STAR Framework argues that legislators must demand platforms embed safety by design, transparency of their algorithms and economic incentives, and accountability and responsibility for failure to enforce their terms of service and the harms their platforms perpetuate. Without oversight, TikTok's opaque algorithm will continue to profit by serving its users – children as young as 13, remember – increasingly intense and distressing content without checks, resources, or support.

It should be clear that this report aims to examine TikTok's role in recommending harmful content to vulnerable users – by no means should users who post about their mental health and experiences be shamed for sharing their experiences. For those affected by any of the issues discussed, please refer to this report's content warning for additional resources and support.

Imran Ahmed
CEO, CCDH

2 Executive Summary

Researchers investigated TikTok's algorithm by establishing two new accounts for users aged 13 in each of the USA, UK, Australia, and Canada. One of these accounts was given a username that indicates a concern about body image.

For each account, we recorded the first 30 minutes of algorithmically recommended content on each account's 'For You' feed, watching and liking any videos about body image, mental health or eating disorders. The resulting recordings were analyzed to examine the frequency of recommendations for body image, mental health, self-harm and eating disorder content.

TikTok hosts eating disorder content with over 13.2 billion views

- TikTok hashtags hosting eating disorder content have over 13.2 billion views.
- Some pro-eating disorder content is evading moderation by using coded hashtags, in some cases co-opting the name of singer Ed Sheeran.

TikTok recommended eating disorder and self-harm content to new teen accounts in minutes

- New TikTok accounts in our study were recommended self-harm and eating disorder content within minutes of scrolling the app's For You feed.
  - Suicide content was recommended within 2.6 minutes
  - Eating disorder content was recommended within 8 minutes

TikTok showed teens body image and mental health content every 39 seconds

- A new TikTok account set up by a 13-year-old user that views and likes content about body image and mental health will be recommended that content every 39 seconds.
- Experts have warned that such content can have a damaging effect on teens' mental health, even where it does not explicitly promote eating disorders.

TikTok targeted vulnerable teens with much more harmful content

- TikTok accounts established with the phrase "loseweight" in their name received significantly more recommendations for eating disorder and self-harm content.
- These Vulnerable Teen Accounts were shown:
  - 3 times as many harmful videos as standard teen accounts
  - 12 times as many self-harm videos as standard teen accounts
- Vulnerable Teen Accounts were also shown more videos relating to mental health.

Recommendations

- TikTok must provide full transparency of its algorithms and rules enforcement, or regulators should step in and compel the platform to do so.

- Dealing with eating disorder content within videos or using coded hashtags requires proactive, informed enforcement that leverages the experience of public health, civil society and academic bodies with expertise in this issue.

- Legislators can change the incentives that shape TikTok's business model by implementing our STAR Framework for social media regulation, which includes: Safety by Design, Transparency, Accountability, and Responsibility.[4]

- Until social media companies are liable for negligence in the coding of their algorithms, instead of hiding behind the Section 230 liability shield, they will continue to behave in a negligent manner that puts children and adults at risk.[5]

3 Methodology

This report is based on recordings of the content served by TikTok's recommendation algorithm to new accounts in the first 30 minutes that they spend browsing the platform's 'For You' feed.

TikTok's For You feed is an endlessly scrollable stream of algorithmically recommended video content. TikTok automatically presents users with the For You feed when they open the app, and the platform has stated that For You is "central to the TikTok experience and [is] where most of our users spend their time."[6]

Each TikTok user is shown different, personalized content on their For You feed as TikTok's algorithm seeks to recommend content catering to each user's interests. TikTok's algorithm bases these recommendations on user actions, for example which videos they like and how long they watch each video.[7]

Our study was designed to investigate how algorithmic recommendations on TikTok's For You feed respond to users that express an interest in body image, mental health and eating disorders.

How we set up accounts for the study

Researchers created eight new TikTok accounts, in each case establishing the accounts using a new email address, a clear phone cache and a fresh installation of the Android TikTok app. All eight accounts were set up with a stated birth date that informed TikTok that the user is 13 years old.

The table below shows how each of the eight accounts was assigned a location and username, giving each a name that is used to refer to that account throughout the report. Further details on how locations and usernames were established are outlined below.

Location	Female Username[8]	'loseweight' Username
United States	Standard US Teen	Vulnerable US Teen
United Kingdom	Standard UK Teen	Vulnerable UK Teen
Australia	Standard Australia Teen	Vulnerable Australia Teen
Canada	Standard Canada Teen	Vulnerable Canada Teen

How we set usernames for accounts used in the study

We decided to test if naming accounts on TikTok influenced the content shown on the For You feed. To do this, we created two types of accounts.

1. **Standard Teen Accounts:** Four accounts created in line with the methodology outlined above with female usernames created by a random name generator.
2. **Vulnerable Teen Accounts:** Four accounts created in line with the methodology outlined above but with usernames including the phrase 'loseweight'.

The design of the "Vulnerable Teen Accounts" used in our study is based on research from the tech reform initiative Reset that has shown that pro-eating disorder users on Instagram will choose usernames with related words, such as "anorexia".[9] Pro-eating disorder users are more vulnerable to eating disorder content on social media as they have existing eating disorders and actively seek out harmful content.

Similarly, there is evidence that users that are vulnerable to content about depression, self-harm and suicide will choose usernames with related words. Data retrieved from the mobile phone of Molly Russell, whose death has been linked to harmful social media content on these topics, shows that she created a new Twitter account to consume this content with the handle "Idfc_nomore", believed to have meant "I don't f***ing care no more".[10]

How we set locations for accounts used in the study

Researchers assigned two accounts each to four English-speaking locations: the United States, the United Kingdom, Australia and Canada.

Researchers experimented with TikTok to understand how the app uses location data. TikTok uses IP address, GPS and phone language settings to localize its content recommendations. Researchers were able to change the location of each TikTok account by adjusting the language settings and using Surfshark's VPN and GPS override function.[11] Researchers then examined videos recommended on the For You feed to confirm that accounts were being recommended localized content.

How we determined the behavior of accounts used in the study

Accounts used in the study expressed a preference for videos about body image, mental health and eating disorders by pausing on relevant videos and pressing the 'like' button.

Where researchers identified a recommended video matching one of the below categories, they viewed the video for 10 seconds and liked it. For all other videos, researchers would immediately scroll the For You feed to view the next video recommended by TikTok.

How we recorded and categorized recommended videos

Researchers used Android's built-in screen recording feature to record the content presented to each account in the first 30 minutes they spent browsing the For You feed.

These recordings were then analyzed by researchers to identify videos that fell into one of the categories outlined in the table below. Where videos fell into more than one category, researchers assigned it the category that best represented its content. For example, a video about eating disorders that makes reference to self-harm would be assigned the "eating disorders" category. Researchers identified 595 such videos in total.

This report argues that the speed and frequency with which TikTok recommends mental health, body image, eating disorder and self-harm content to new users is harmful. At points we also refer to eating disorder and self-harm videos as "harmful content", because this content is widely believed to carry an inherent risk of harm to users, and for this reason most platforms including TikTok have set community standards that ban users from posting it.[12]

Within these categories, we have not distinguished content with a positive intent, for example educational or recovery content, from that with a clearer negative intent. This is because researchers are not able to definitively determine the intent of a video in many cases, and because content with a positive intent can still be distressing and may cause harm.

Category	Definition
Mental health	Videos about anxieties, insecurities and mental health conditions, excluding eating disorders, self-harm, and suicide.
Body image	Videos about weight loss, diets, cosmetic surgeries, and beauty products, excluding videos explicitly about eating disorders. All videos about weight loss were included regardless of if they contained healthy advice.
Harmful content	
Eating disorders	Videos explicitly about eating disorders, such as videos about thinspo or discussing eating disorders, including videos tagged with eating disorder hashtags.
Self-harm and suicide	Videos explicitly about self-harm or suicide, such as videos tagged #sh or containing references to self-harm and suicide, such as razors and overdoses.

4 TikTok's algorithm carries mental health risks

Since launching outside of China in 2017, TikTok has become the world's fastest growing social media platform, reaching a billion users faster than any other app. Today, two-thirds of US teens use the app and an average American user watches 80 minutes of TikTok videos a day.

Key to TikTok's success has been its use of algorithmic recommendations. People logging into TikTok for the first time are presented with the app's For You feed, an endlessly scrollable stream of videos selected based on the platform's assessment of your interests, based on analysis of the videos you watch and engage with.[13]

What makes TikTok unique is that the For You feed and its algorithmic recommendations are designed, in the words of its maker, to be "central to the TikTok experience".[14] But this is also what introduces unique dangers to TikTok, as that same algorithm can recommend harmful content as it seeks to keep users viewing its content and the ads that earn the platform money.

This report explores some of the risks posed by TikTok's algorithm by measuring the extent to which it recommends content about mental health, body image, eating disorders and self-harm to young users in their first hours using the app. This section sets out what we already know about the harms that such content can pose to social media users of all ages.

Content about mental health and body image can be harmful

We know that social media content about mental health and body image can contribute to real harm, particularly for younger users.

In September this year, a British court ruled that Instagram posts contributed to the suicide of 14-year-old Molly Russell.[15] Data obtained by Molly Russell's family show that she saved, shared or liked 2,100 posts about depression, self-harm or suicide in the six months before her death.[16]

During that case, Instagram's parent company Meta claimed that it had never studied the effects of suicidal and depressive content on young users.[17] However, we know from the Facebook Papers, internal documents leaked by former Meta employee Frances Haugen, that Instagram had evidence that teens who struggle with mental health say using the platform made it worse, in part because of "pressure to conform to social stereotypes… and body shapes of influencers".[18]

More recent research has shown that many popular TikTok videos relating to body image and diet could "lead to development of eating disorders", with young women particularly at risk of developing "internalized body image and disordered eating behaviors".[19]

Harms posed by eating disorders and poor mental health extend into adulthood

We know that the harms posed by mental health, body image, eating disorder and self–harm or suicide content are unlikely to be limited to just young users.

This report focuses on the experience of teen users with an interest in mental health and body image in their first hours using TikTok, an age group for which TikTok claims to set "higher default standards for user privacy and safety."[20] But we know that the potential harms identified by our report affect adult users too:

- Eating disorders are most prevalent at age 21[21]
- Suicide is more prevalent from age 25 onwards[22]
- About 5% of adults self–harm, although rates are highest for teens and college students[23]

So despite focusing on teen users in particular, our research raises questions about the impact of TikTok recommendations on adult users who are known to be susceptible to these problems and who do not benefit from any extra safety measures that TikTok may provide for minors.

Social media platforms serve large networks focused on mental health and eating disorders

We know that these are not small problems, as social media platforms are hosts to large networks posting mental health and eating disorder content, and must therefore take a principled and responsible approach to managing the impact of this content on their users.

Research from the tech reform initiative Reset has shown that Instagram is host to a network of eating disorder accounts reaching 20 million unique followers on the platform.[24] Other studies have shown that videos on TikTok's #mentalhealth hashtag have over 25.3 billion views, with almost half of these featuring "symptoms of mental distress".[25] Our own report reveals that TikTok videos posted with hashtags relating to eating disorders have 13.2 billion views.

Our aim is not to criticize the people that make up these communities, many of whom turn to social media for support or to express themselves, but to shine a light on TikTok's responsibilities in hosting this content and recommending it to vulnerable young users.

5 Teens are bombarded with body image and mental health content on TikTok every 39 seconds

In order to investigate the rate at which teens are recommended body image and mental health content, researchers created four TikTok accounts with a stated age of 13.

These accounts, which we term **Standard Teen Accounts**, expressed an interest in body image, mental health and eating disorders by watching and liking relevant videos. Later sections will contrast our findings for these accounts with those belonging to vulnerable teens.

Our analysis found that TikTok recommended videos about mental health or body image were served to Standard Teen Accounts every 39 seconds.

<div style="border:1px solid #00CED1;">

Every 39 seconds

Rate at which **Standard Teen Accounts** were shown body image or mental health content

</div>

Experts on the relationship between body image, mental health and social media have warned that the prevalence of content about body image and mental health can have damaging effects on teens' mental health, even where that content does not explicitly promote eating disorders.[26]

This table shows how many mental health and body image videos TikTok recommended to the four Standard Teen Accounts used in our study.

Content	Videos Shown to Standard Accounts	Video Frequency
Mental health	98 videos	Every 73 seconds
Body image	87 videos	Every 83 seconds
Mental health or body image	185 videos	Every 39 seconds

Categories of videos served to Standard Teen Accounts

Total videos shown on average to Standard Teens
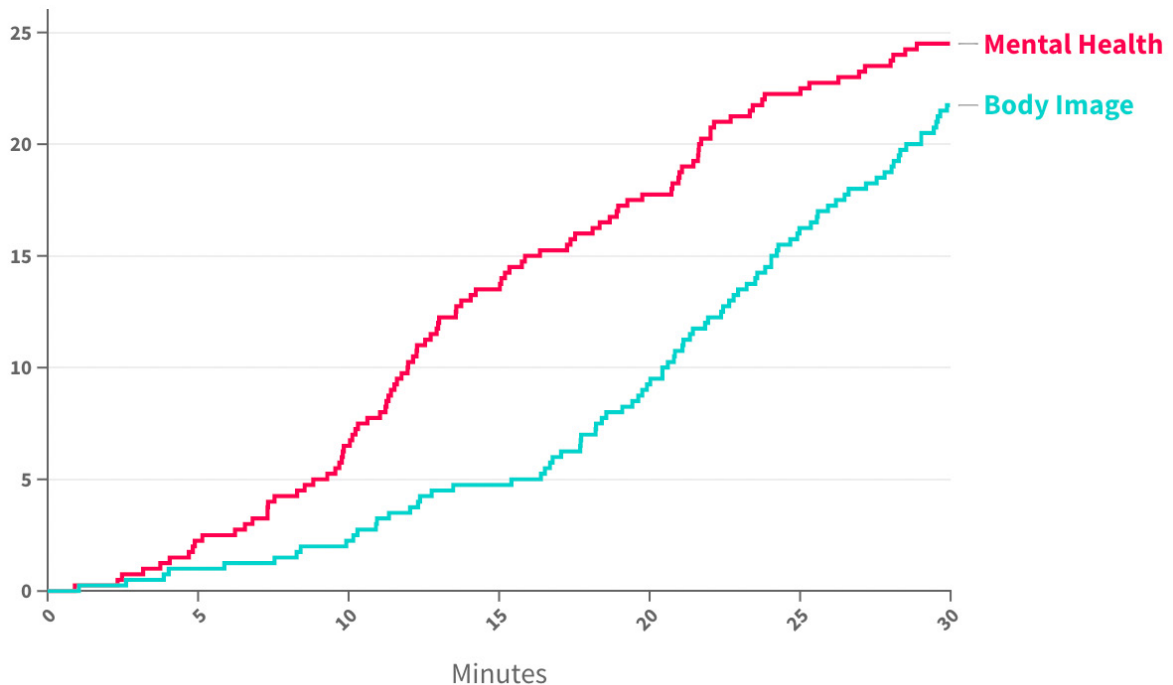
This graph displays the average number of mental health and body image videos shown to Standard Teen Accounts in the first 30 minutes of using the app

Examples of mental health and body image content recommended to Standard Teen Accounts

Studies of Instagram have shown that teens who struggle with mental health say using the platform made it worse, in part because of "pressure to conform to social stereotypes… and body shapes of influencers".[27]

While social media has been a useful tool for destigmatising mental health, studies have found that almost half of mental health content on TikTok "reported or expressed symptoms of mental distress" and that there is "limited insight into the impact of traumatic or 'triggering' events" posted on TikTok.[28]

Content recommended by TikTok in our study shows that teens with an interest in body image and mental health may face similar harms from a high rate of recommendations for body image and mental health content.

The majority of mental health videos recommended to Standard Teen Accounts in our study consisted of users sharing their anxieties and insecurities. In contrast, body image content appeared to be more harmful in nature, with videos advertising weight–loss drinks and 'tummy tuck' surgeries to accounts with a stated age of 13.

2,153 likes

"Boiled Egg Diet", weekly meal plan. Recommends eating almost exclusively boiled eggs to lose 24 lbs in two weeks.
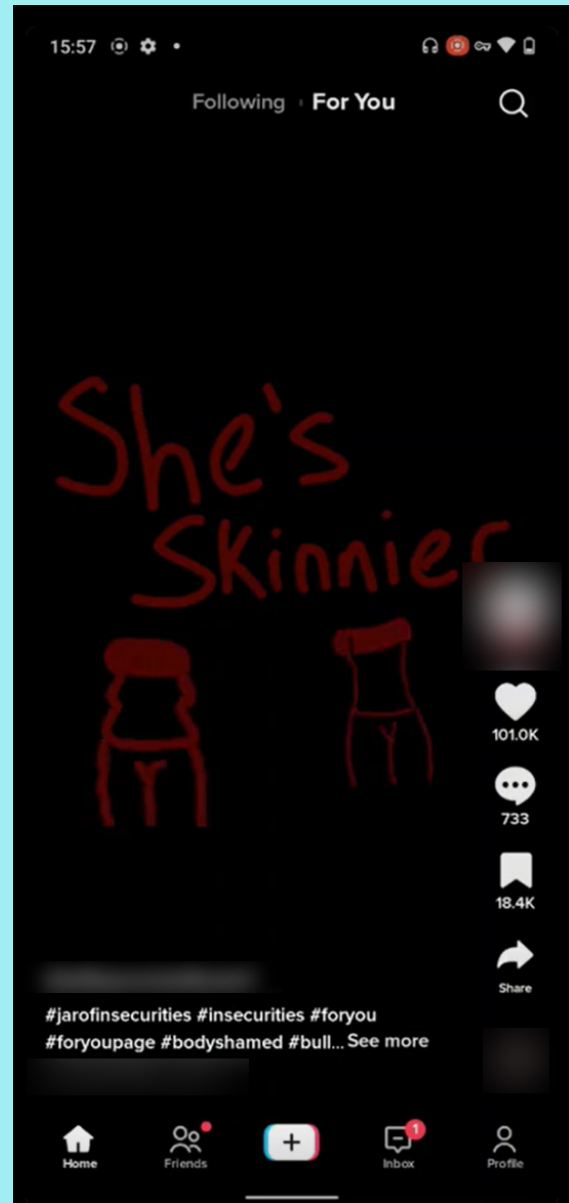
101,000 likes

An animation about wanting to be skinny. Text reads "She's skinnier". Audio says "I've been starving myself for you".
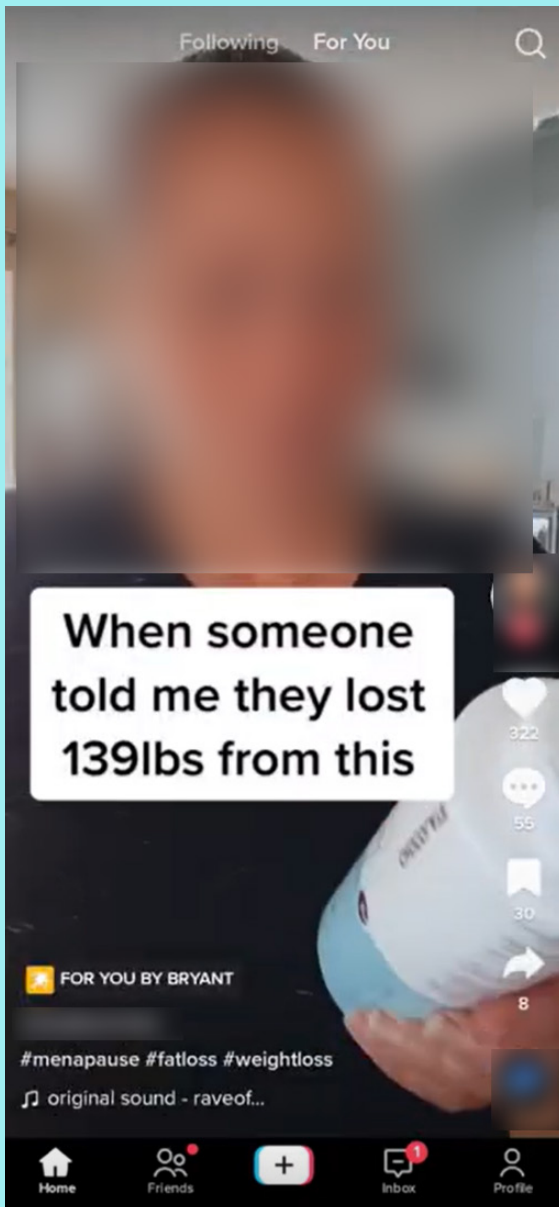
322 likes

Unmarked advert for Modere Trim, a weight-loss drink.

19,500 likes

How to get a cosmetic 'tummy tuck' surgery. Caption reads: "Find out more through the links in my bio".
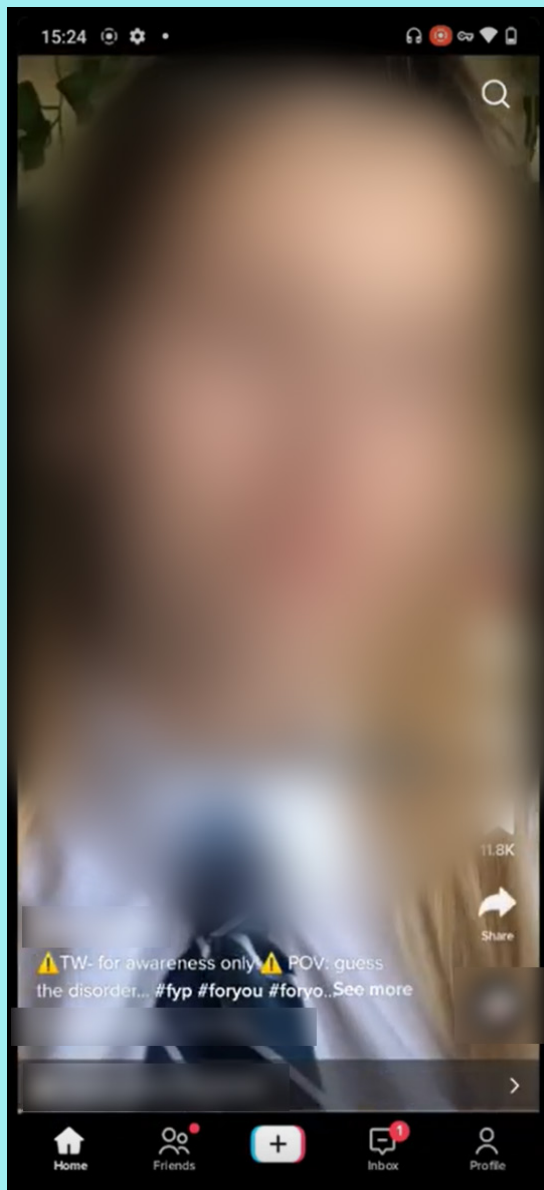
370,500 likes

This user acts out mental health conditions and asks the viewer to guess the name of the condition.

6 TikTok shows some teens eating disorder, self-harm and suicide content in minutes

Our study shows that TikTok recommended content about eating disorders, self-harm or suicide within minutes in some cases.

Analysis of recommended videos on the For You feed of the four **Standard Teen Accounts** used in our study shows that this content was sometimes recommended extremely quickly. At the earliest, we found that a video discussing suicide was recommended within 2 minutes and 38 seconds, while eating disorder content was recommended within 8 minutes.

2.6 minutes	8 minutes	Every 206 seconds
Fastest recommendation for suicide content to a Standard Teen Account	Fastest recommendation for eating disorder content to a Standard Teen Account	Rate at which Standard Teen Accounts were shown eating disorder or self-harm videos

TikTok's recommendation of self-harm, suicide or eating disorder content within 30 minutes in most tests raises concerns that teen users could quickly be exposed to such content and develop a greater interest in these topics, something explored in more detail in following sections.

While eating disorder and self-harm content was recommended to Standard Teen Accounts at a lower rate than the other mental health and body image content we identified, the data shows that TikTok's algorithm is nevertheless regularly recommending this content to new accounts.

This table shows how frequently Standard Teen Accounts were shown self-harm and eating disorder content. This is an average across the four Standard Teen Accounts in our study.

Content	Total Videos Shown to Standard Accounts	Frequency
Self-harm and suicide and suicide	6 videos	Every 20 minutes
Eating disorders	29 videos	Every 248 seconds
Self-harm and suicide or eating disorders	35 videos	Every 206 seconds

Standard Teen Accounts served up harmful content
Total videos shown on average to Standard Teens
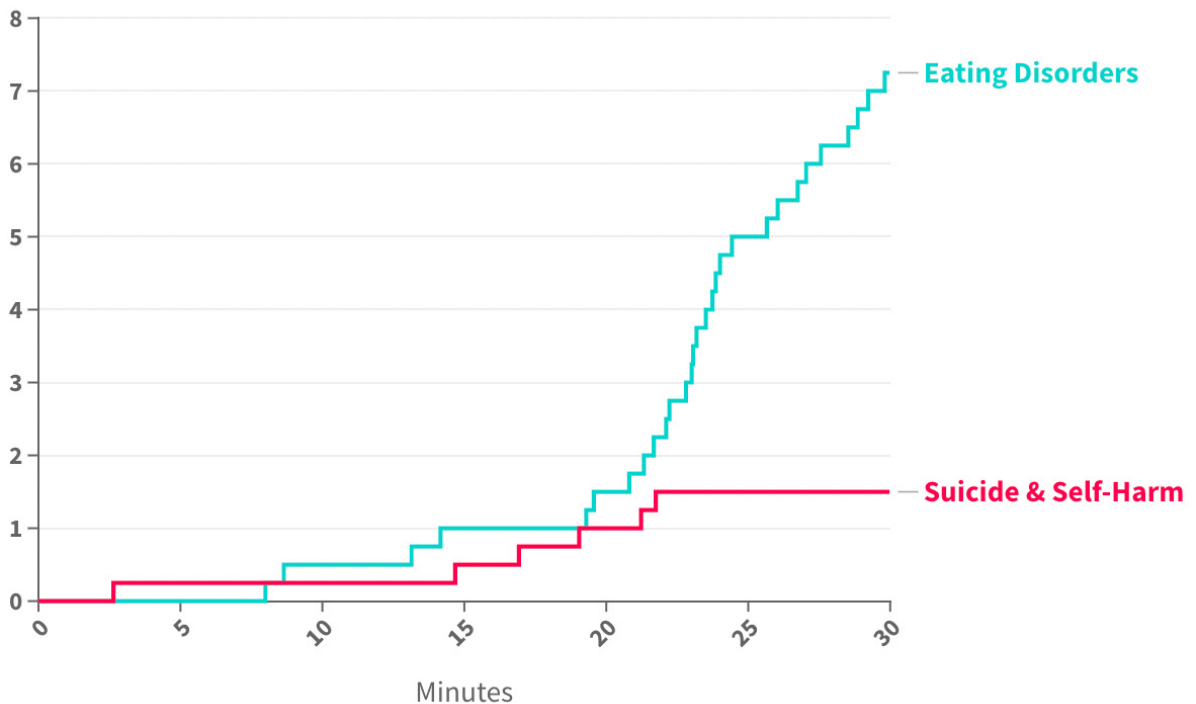
This graph displays an average of recommendations to the four Standard Teen Accounts, showing that the pace of eating disorder and self-harm recommendations increased over 30 minutes with no noticeable slowdown effects or other safeguarding measures.

Eating disorder and self-harm videos shown to teen accounts within 30 minutes

The severity of eating disorder and self-harm content recommended to Standard Teen Accounts in our study varied across videos and accounts.

The most extreme content shown to our accounts was videos about "junkorexia", going to school after attempting suicide, self-harming, insecurities about body weight, and disordered eating.
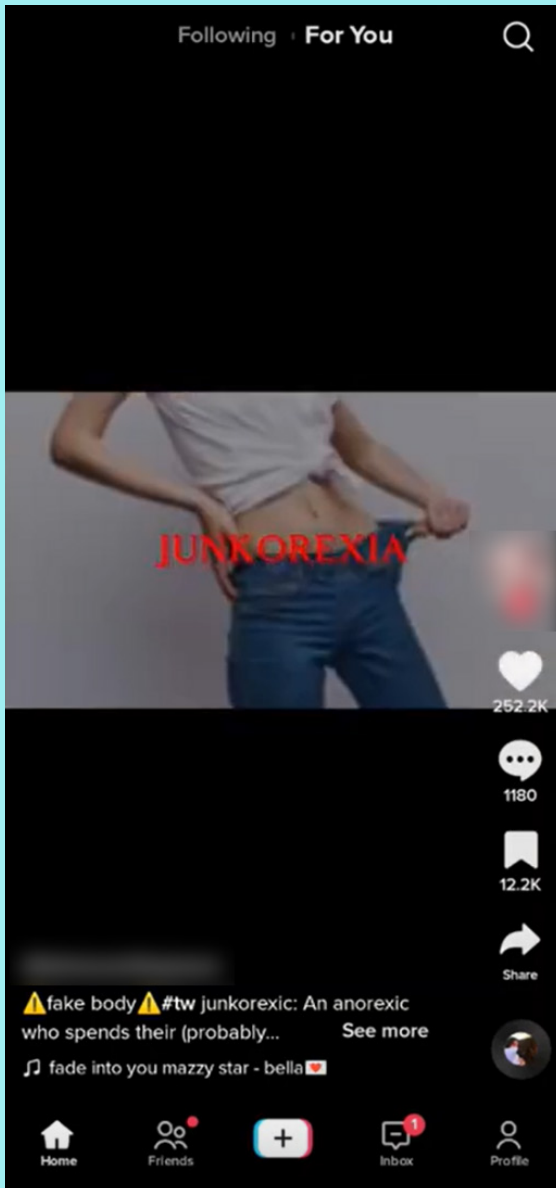
252,200 likes

"Junkorexia" is an unofficial term for people with anorexia who only eat junk food.

362,200 likes

'Jar Of Insecurities' is a meme where users fill in the jars to represent their insecurities. This meme appeared several times. This video uses a song with the lyrics "I've been starving myself for you".
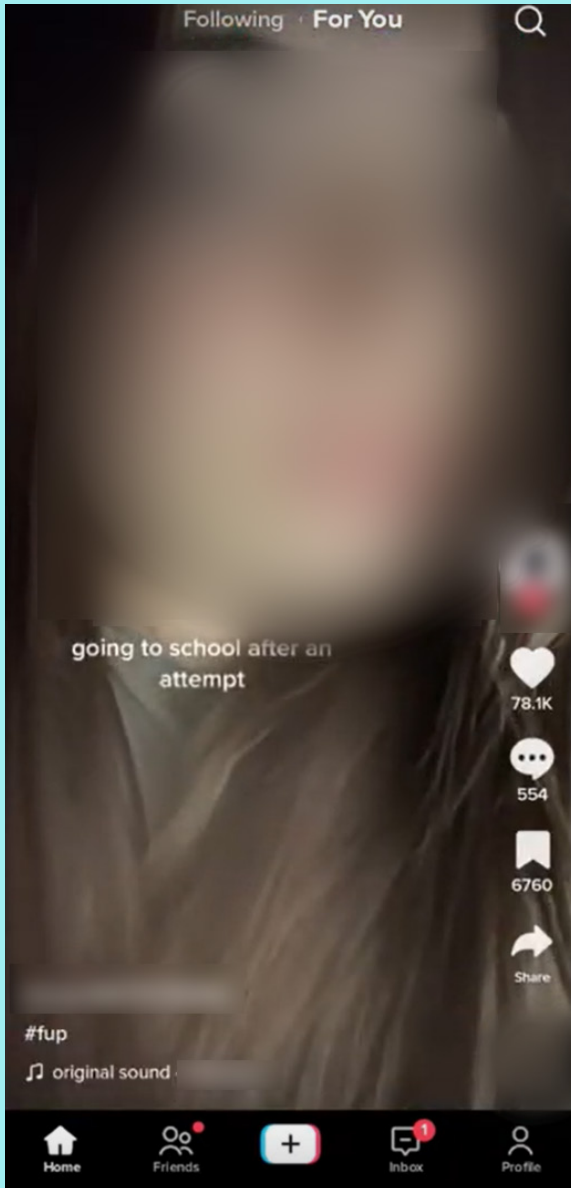
78,100 likes

"Going to school after an [suicide] attempt".
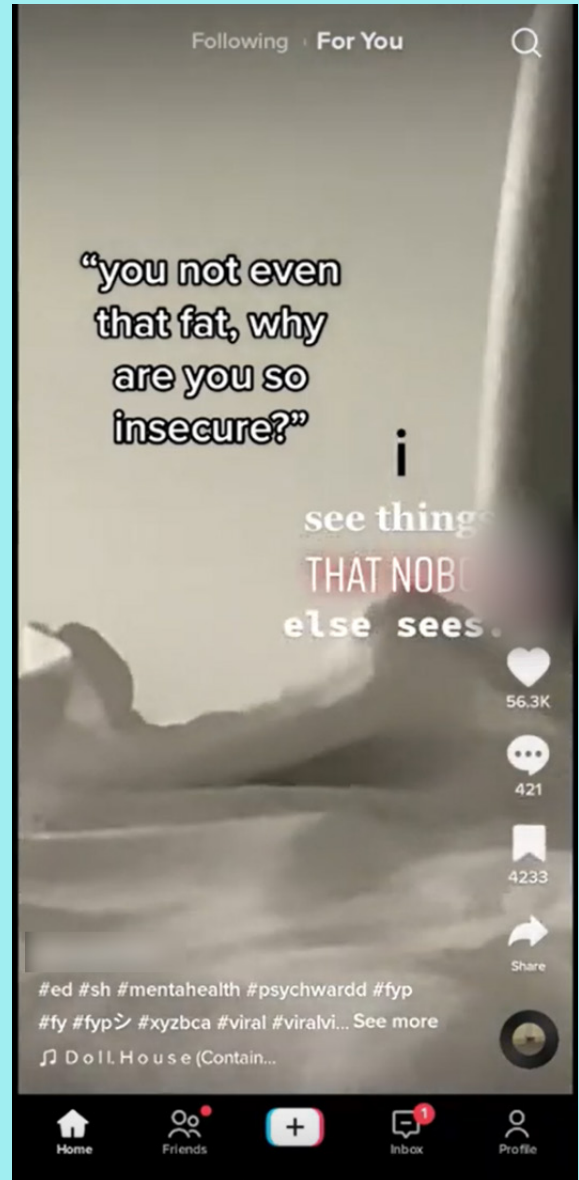
56,300 likes

Video taken from a hospital bed. "You not even that fat, why are you so insecure?" "I see things THAT NOBODY else sees." Tagged #ed and #sh.

76,300 likes

Video about disordered eating. Video cuts between an image of a woman eating nothing with the text "you have eating problems right?" to the second image of a woman eating donuts and the text "no I have eating problems".

7 Vulnerable teens with are exposed to more harmful content, not less

Researchers created four **Vulnerable Teen Accounts** with usernames containing the phrase "loseweight" and a stated age of 13 to investigate the content recommended to teen users with existing concerns about mental health or body image.

The design of these accounts is based on research from the tech reform initiative Reset that has shown that "pro-eating disorder" users on Instagram will choose usernames with related words, such as "anorexia".[29] These users are more vulnerable to eating disorder content on social media as they are vulnerable to eating disorders and actively seek out harmful content.

Similarly, there is evidence that users that are vulnerable to content about depression, self-harm and suicide will choose usernames with related words. Data retrieved from the mobile phone of Molly Russell, whose death has been linked to harmful social media content on these topics, shows that she created a new Twitter account to consume this content with the handle "Idfc_nomore", believed to have meant "I don't f***ing care no more".[30]

Following the same methodology used throughout our study, researchers displayed an interest in body image, mental health and eating disorder content by watching and liking relevant videos on TikTok's For You feed, recording and analyzing the first 30 minutes of recommendations.

Every 66 seconds	**3 times more**	**12 times more**
Rate at which Vulnerable Teen Accounts were shown harmful videos	Harmful recommendations to Vulnerable Teen Accounts	Self-harm and suicide videos shown to Vulnerable Teen Accounts

Data from our study shows that videos about eating disorders, self-harm and suicide were shown to Vulnerable Teen Accounts three times more often than to Standard Teen Accounts.

The rate of recommendations for self-harm and suicide videos increased particularly sharply, with Vulnerable Teen Accounts receiving 12 times as many recommendations as Standard Teen Accounts in our study.

This table shows how many eating disorder, self-harm and suicide videos TikTok recommended to the four Vulnerable Teen Accounts used in our study.

Content	Total Videos Shown to Vulnerable Teen Accounts	Frequency
Self-harm and suicide and suicide	74 videos	Every 97 seconds
Eating disorders	35 videos	Every 206 seconds
Self-harm and suicide or eating disorders	109 videos	Every 66 seconds

Vulnerable Teens were served up more harmful videos

Total videos shown to average Standard and Vulnerable Teens
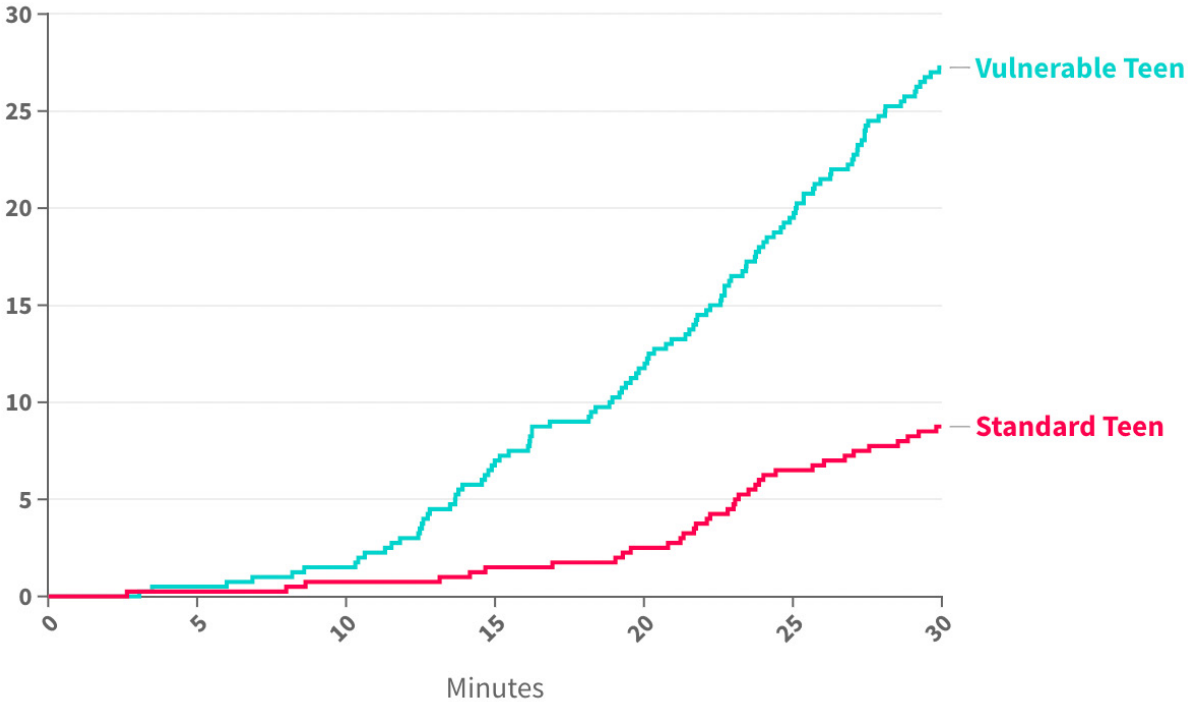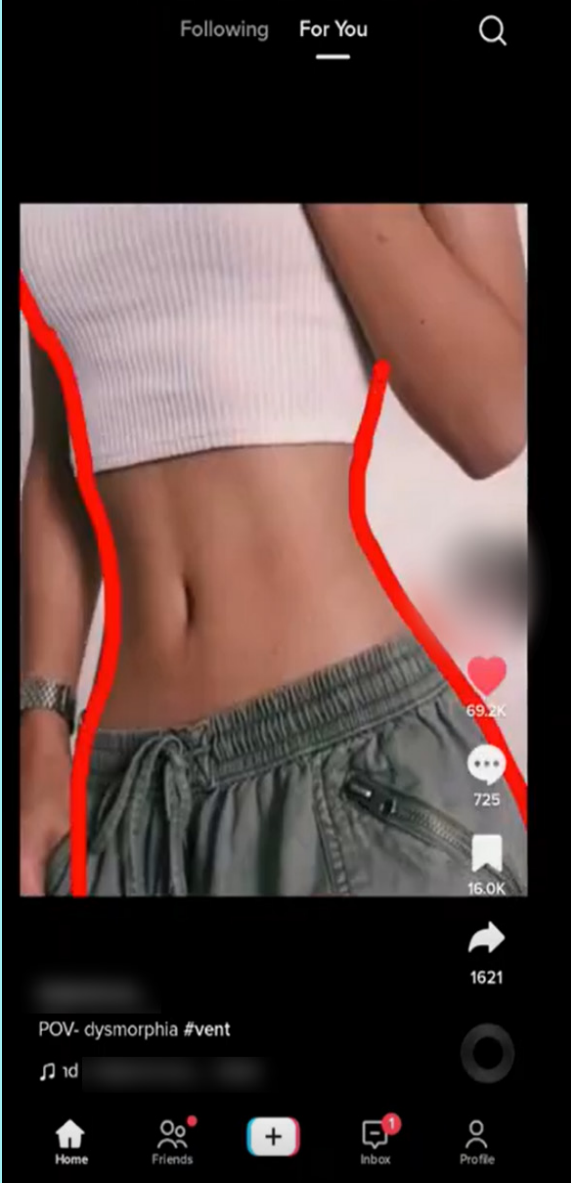
This graph displays the average harmful videos recommended to our Vulnerable and Standard Teen Accounts, illustrating the disparity between the two groups. Vulnerable teens were served three times the amount of harmful videos as standard teens.

Extreme content recommended to Vulnerable Teen Accounts in 30 minutes

The harmful content TikTok recommended to the Vulnerable Teen Accounts was more extreme than content shown to the Standard Teen Accounts. An increase in videos about eating disorders was accompanied by an increase in self-harm and suicide content, which included thinspo, methods to self-harm, and videos of teens discussing plans to commit suicide.
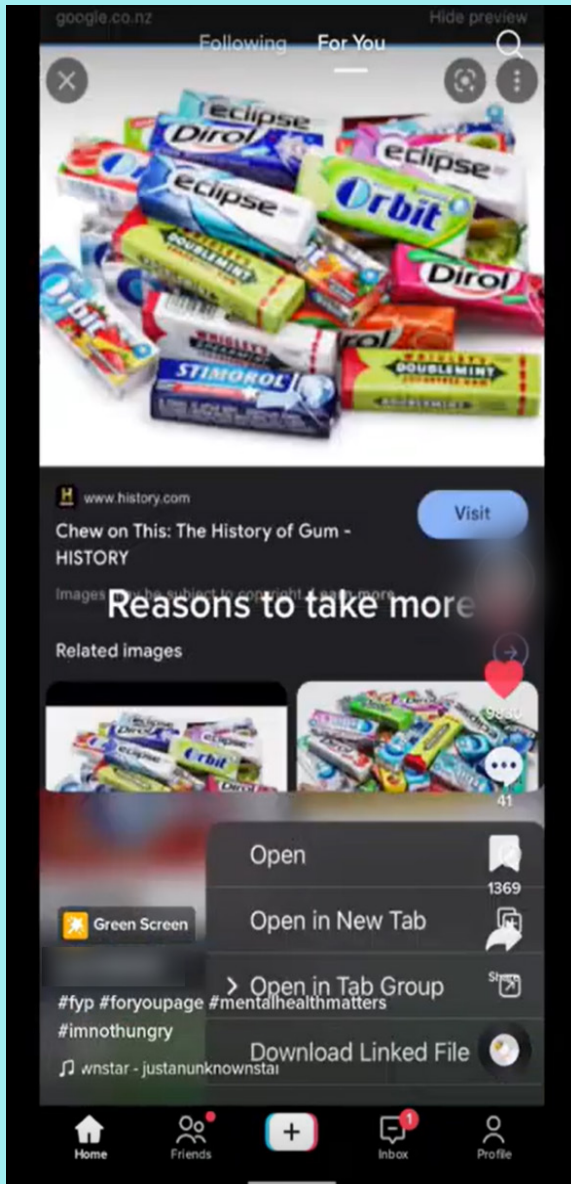
Below are examples of the eating disorder content recommended to our Vulnerable Teen Accounts.

69,200 likes

Thinspo, short for 'thinspiration'. Images aimed at motivating weight loss. This video contains several thinspo images with flashing red-outlines set to fast drum beats. Text reads "MAN I WISH I WAS PRETTY" "I'M SO FUCKING UGLY".

9,829 likes

'Yk [you know] eating gum can Mack u less hungry".

Hashtag reads #imnothungry.

153,100 likes

Reads: "POV: [point of view] you're slowly losing yourself again". Montage of eating disorder and self–harm images, including gum which is used as a substitute for food by some with eating disorders.
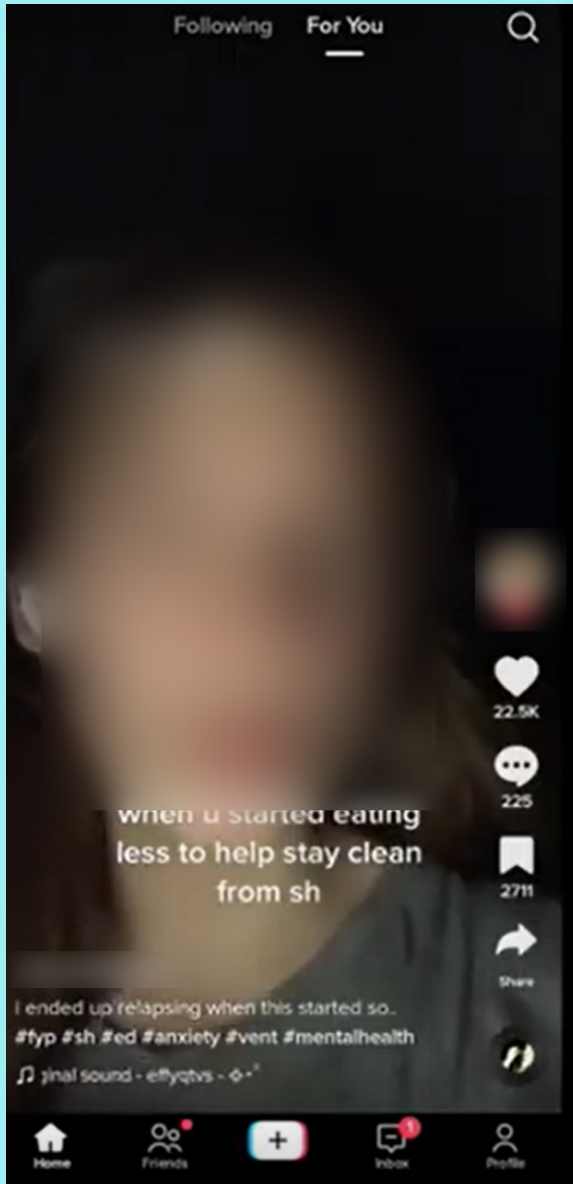
22,500 likes

"Tw: ed sh" "when u started eating less to help stay clean from sh".
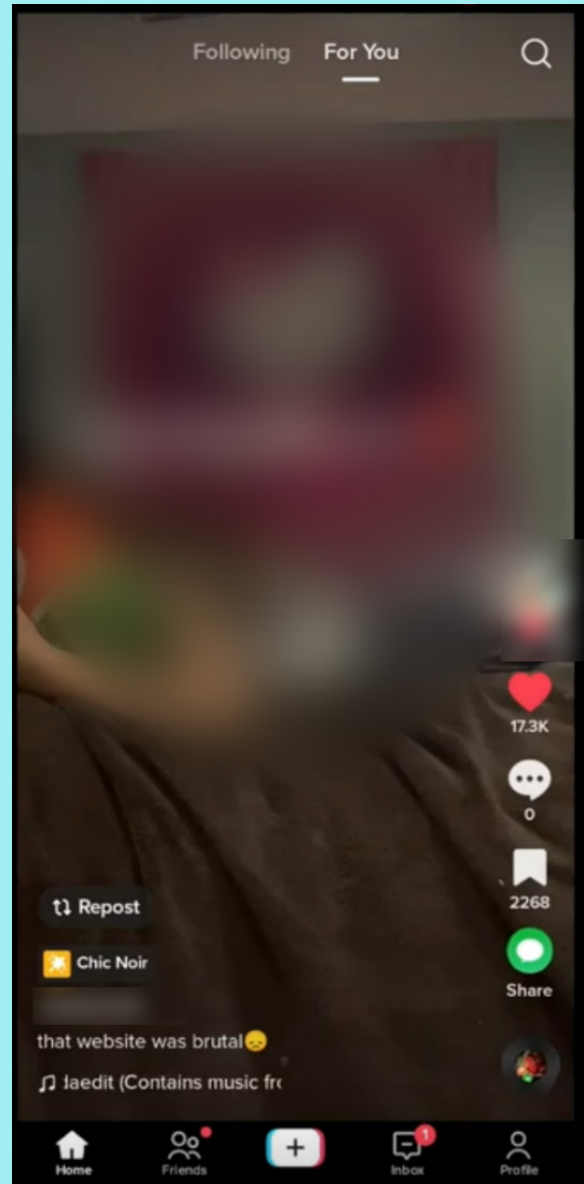
17,300 likes

Link to PrettyScale.com, a website where users upload pictures of the body and face to have their attractiveness ranked by an 'mathematical formula'.

Of the 39 videos about suicide shown to the Vulnerable Teen Accounts, six of them discuss desires or plans to attempt suicide. Some users disclosed their location and methods of hiding suicide attempts. In one instance, a Vulnerable Teen Account was shown three videos of users discussing suicide plans in one minute.
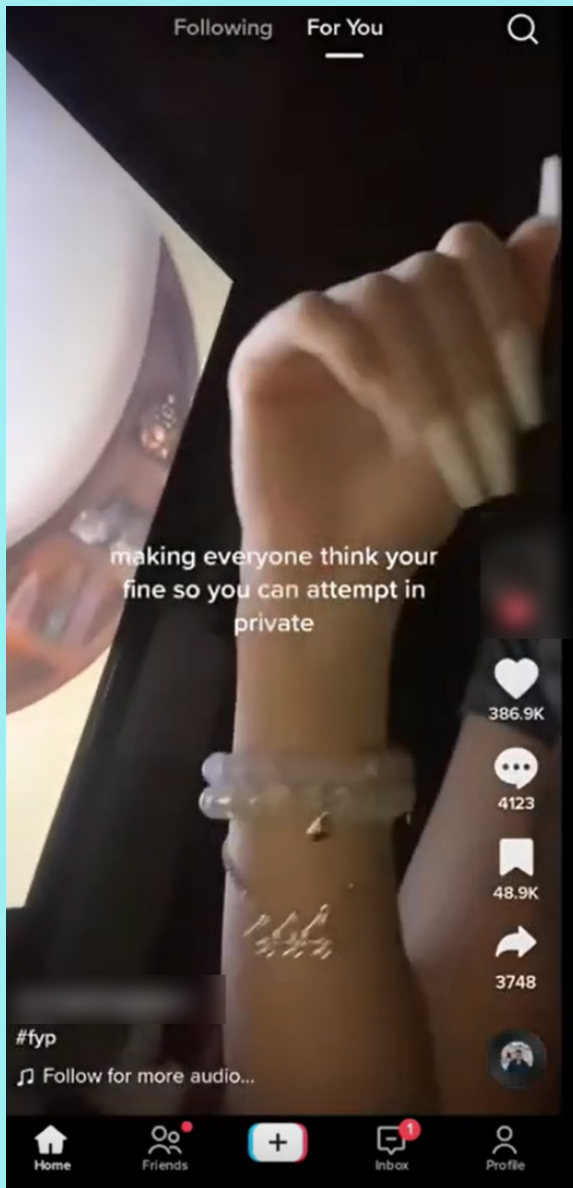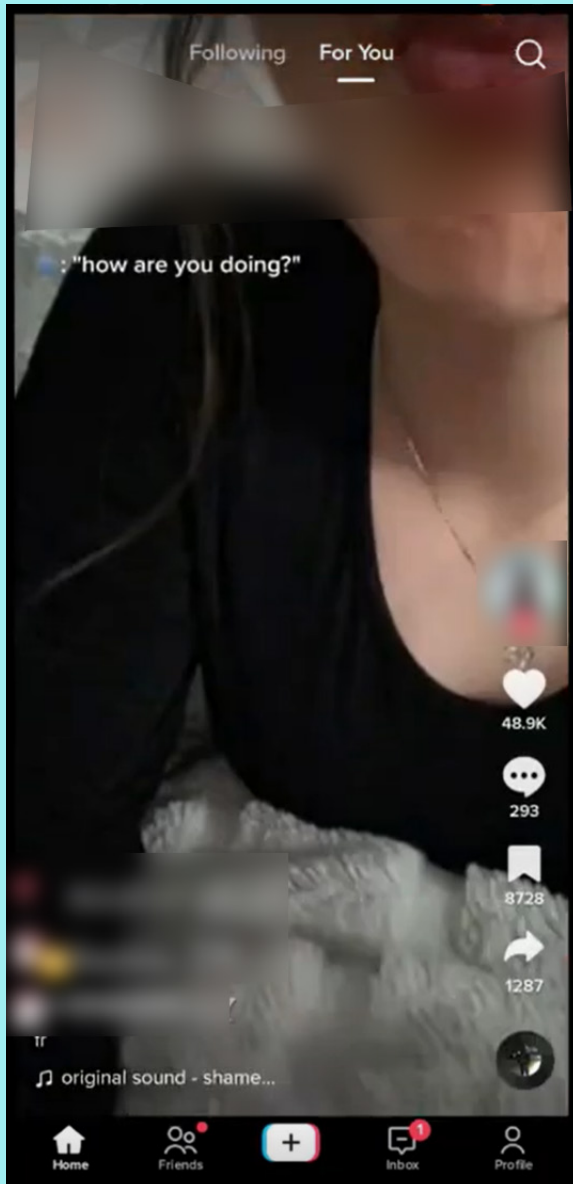

386,900 likes

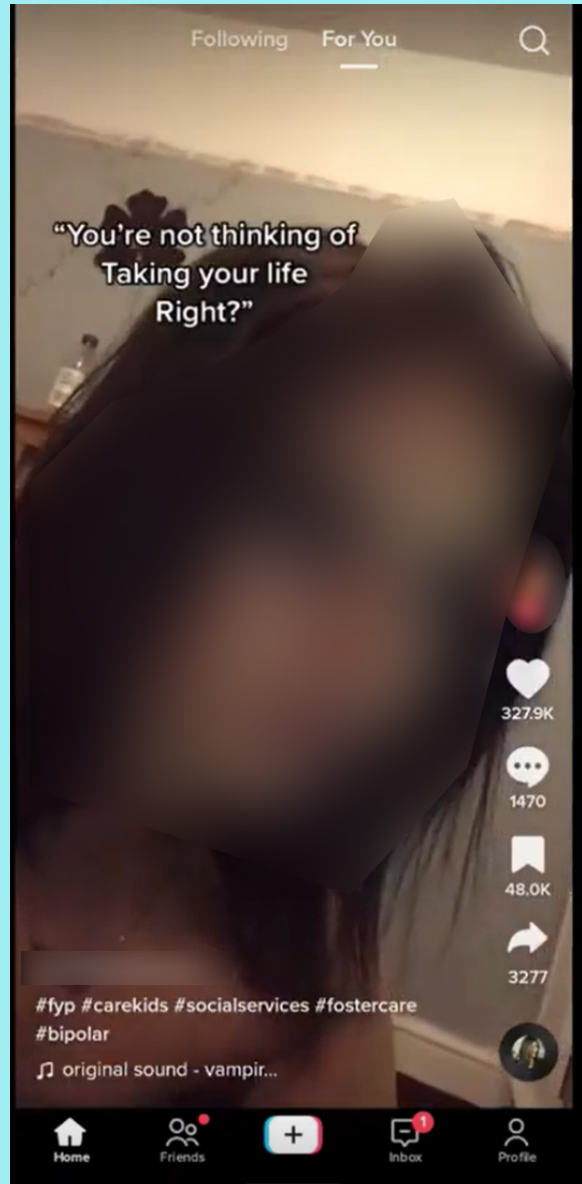"Making everyone think your fine so that you can attempt [suicide] in private".
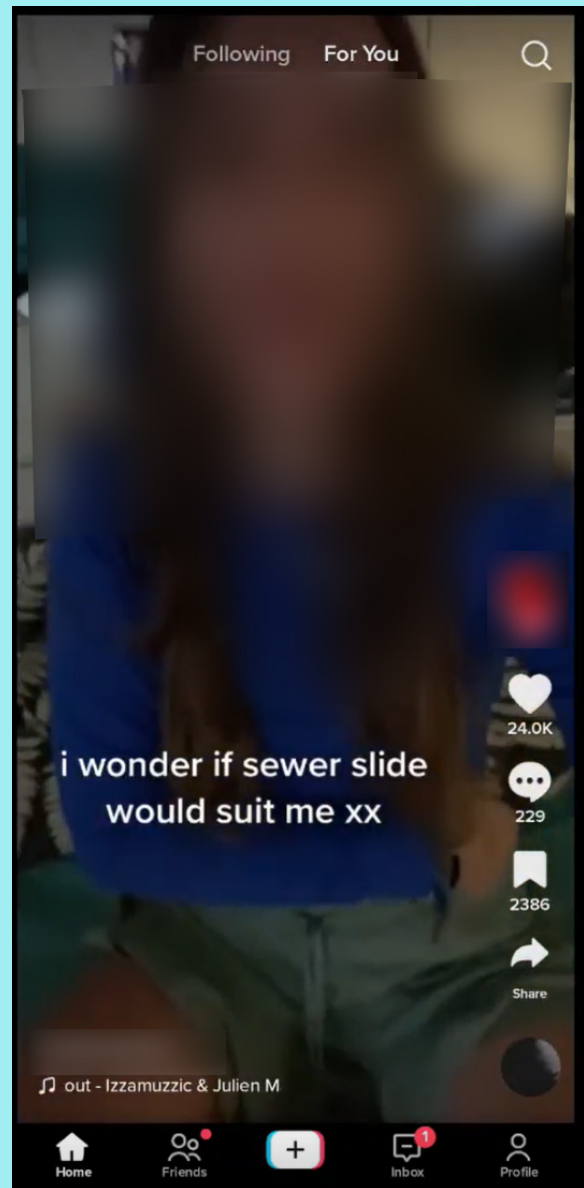
24,000 likes

"I wonder if sewer slide [suicide] would suit me xx".

48,900 likes

Expressing suicidal ideation. This user added a location, time and date.
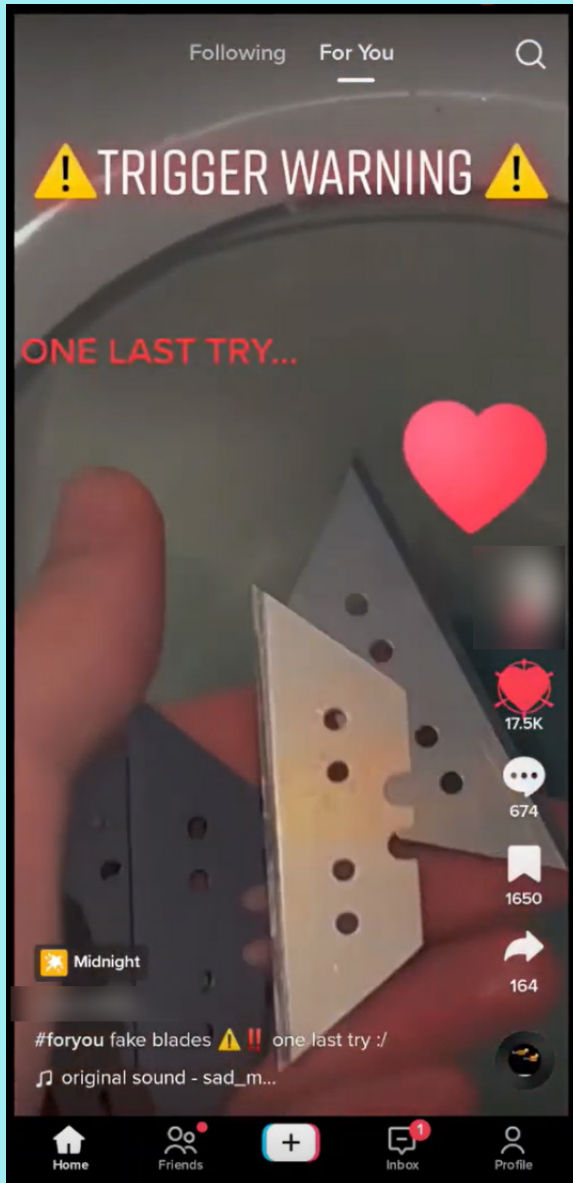
327,900 likes

Video of a teen girl crying. "You're not thinking of taking your life right?" "Right, Sarah Lynn?". Sarah Lynn is a TV character who dies from an overdose.

62,500 likes

The word "suicide" is hidden on the bush. The audio says "you know you're on my mind".

Our Vulnerable Teen Accounts were also shown 38 videos about self-harm. Razor blades were a common motif in these videos.

17,500 likes

Video of the user flushing razor blades down the toilet in an attempt to stay clean from self-harm. Reads: "TRIGGER WARNING" "ONE LAST TRY…".
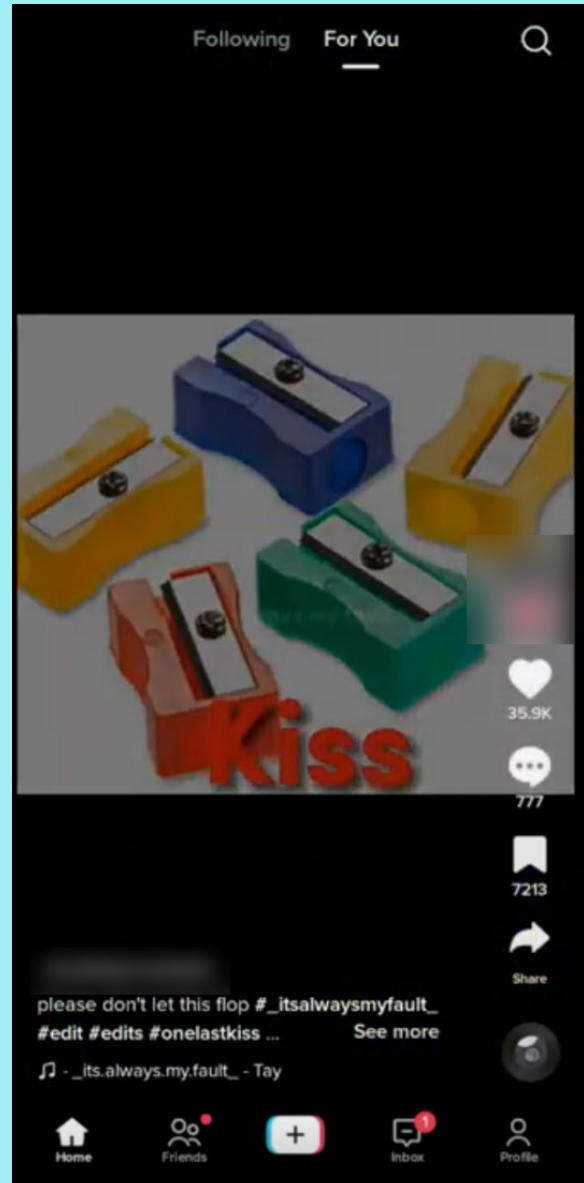
35,900 likes

Montage of pictures of sharpeners, razor blades and scissors. Set to a song with the lyrics "one last kiss I love you like an alcoholic".
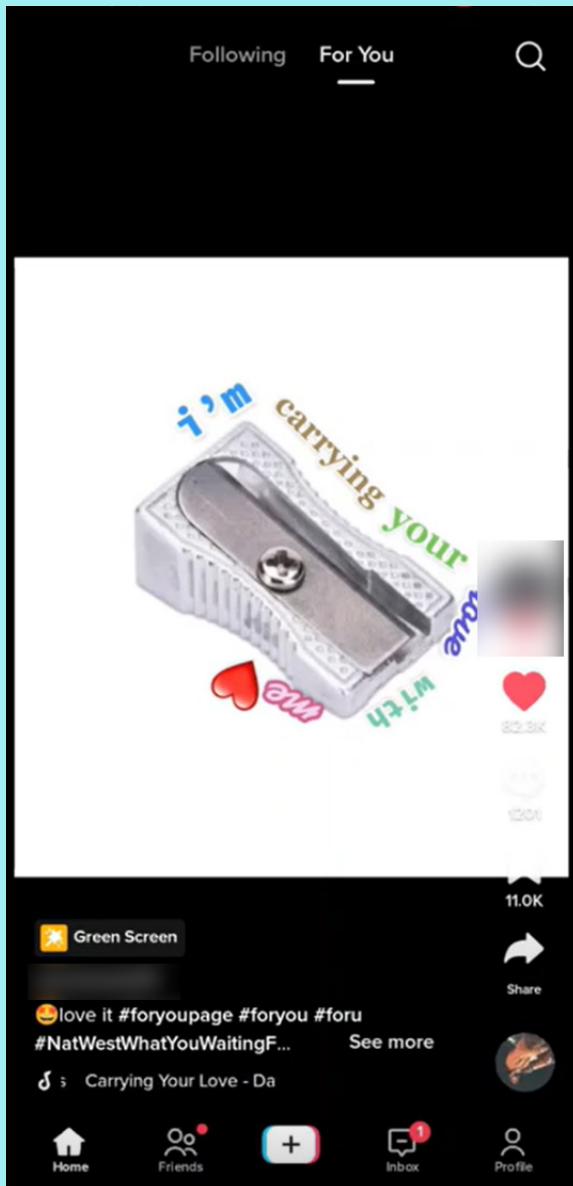
82,300 likes

"I'm carrying your love with me". Describing self-harm scars from razor blades taken from sharpeners.
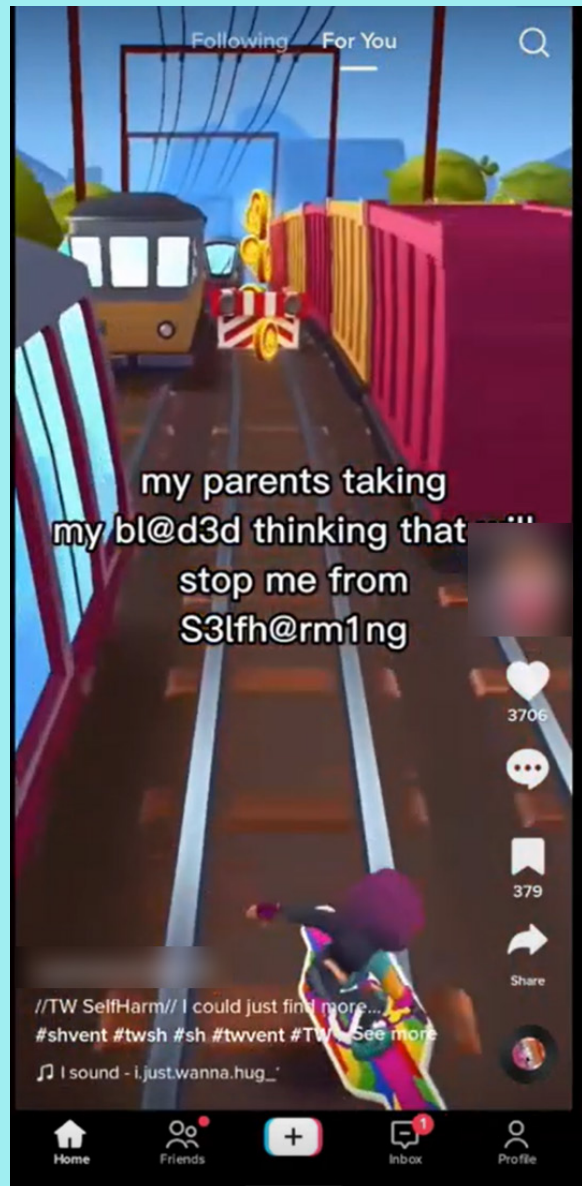
3,706 like

"my parents taking my bl@d3d [blades] thinking that will stop me from S3lfh@rm1ng [self-harming]".

Following For You

ya #mentalhealthmatters #foryoupage #sh

♫ ongs <3 je te lasserai des

Home Friends Inbox Profile

53,000 likes

"Pov: you have the same coping mechanism as me".

'POV 'means 'Point of View'. This video is a montage of coping mechanisms, including razor blades.

8 Mental health and body image videos shown to vulnerable teens every 27 seconds

Mental health and body image videos were recommended to Vulnerable Teen Accounts every 27 seconds, more frequently than the rate for Standard Teen Accounts.

This shows that recommendations for general mental health and body image content increased alongside a sharp increase in recommendations for self-harm and eating disorder videos.

Every 27 seconds	1.4 times	2.5 times
Rate at which Vulnerable Teen Accounts were shown body image or mental health content	As many mental health and body image videos recommended to Vulnerable Teen Accounts	As many mental health videos recommended to Vulnerable Teen Accounts

The table below shows that the increase in average frequency of mental health and body image videos was driven by an increase in the frequency of mental health videos.

Content	Total Videos Shown to Vulnerable Teen Accounts	Average Frequency
Mental health	249 videos	Every 29 seconds
Body image	17 videos	Every 424 seconds
Mental health or body image	266 videos	Every 27 seconds

Videos served to Vulnerable and Standard Teen Accounts
Total videos shown to average user

Standard Teen
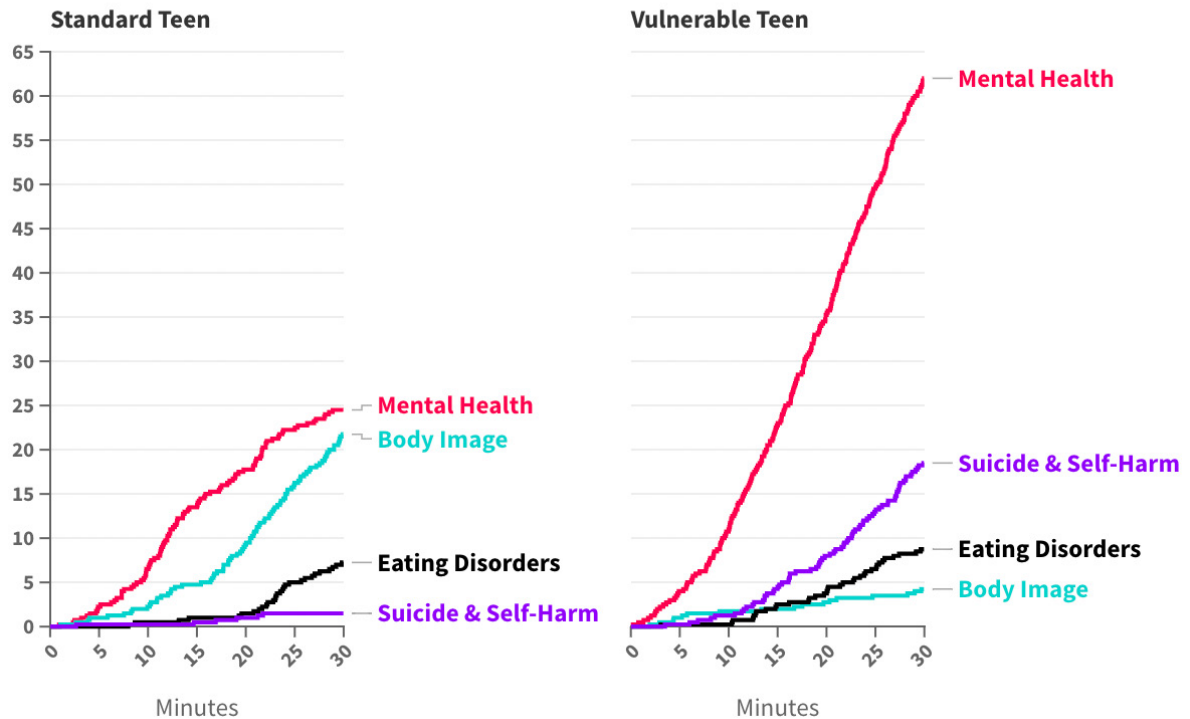
Vulnerable Teen

In the first 10 minutes of the recordings, mental health videos made up 6% of all content shown to the vulnerable accounts. In the final 10 minutes, mental health videos made up 27% of all content shown to the vulnerable accounts.

9 TikTok is host to an eating disorder community with over 13.2 billion views

Over the course of this study, researchers found 56 TikTok hashtags hosting eating disorder videos with over 13.2 billion views.

Researchers examined the videos carrying these hashtags to determine whether or not they promote eating disorders, identifying 35 that contain a high concentration of pro-eating disorder videos that promote harmful behaviors such as very low-calorie diets.[31] Videos posted using these hashtags received 59.9 million views in total.

Another 21 hashtags contained healthy discussion of eating disorders as well as harmful pro-eating disorder videos, with 13.2 billion views in total. TikTok does not appear to label or moderate content on these hashtags, making them a place where recovery content mingles freely with content that promotes eating disorders.

None of these hashtags carry warnings or links to help and advice for users struggling with eating disorders.

Users evade moderation by altering hashtags, for example by modifying #edtok to #edtøk. Another popular approach for avoiding content moderation is to co-opt singer Ed Sheeran's name, for instance #EdSheeranDisorder.

Korean pop music (K-pop) has a strong influence on the eating disorder community on TikTok, with many posts presenting the members of K-pop groups such as BlackPink and Aespa as inspirational body images or 'thinspo'. Pro-eating disorder users also employ K-pop references in their hashtags, for example #kpopwlacc which uses an abbreviation of the phrase "K-pop weight loss account" to promote eating disorder videos with 10 million views.

Previous research has shown that pro-eating disorder users are likely to connect through social media and encourage each other's eating disorders through 'accountability' posts. This involves sharing their current weight (cw), goal weight (gw) and ultimate goal weight (ugw), thinspo, and pairing-up as 'weight loss buddies".[32] Posts of this kind are easily found under TikTok's eating disorder hashtags.

Examples of content found on eating disorder hashtags
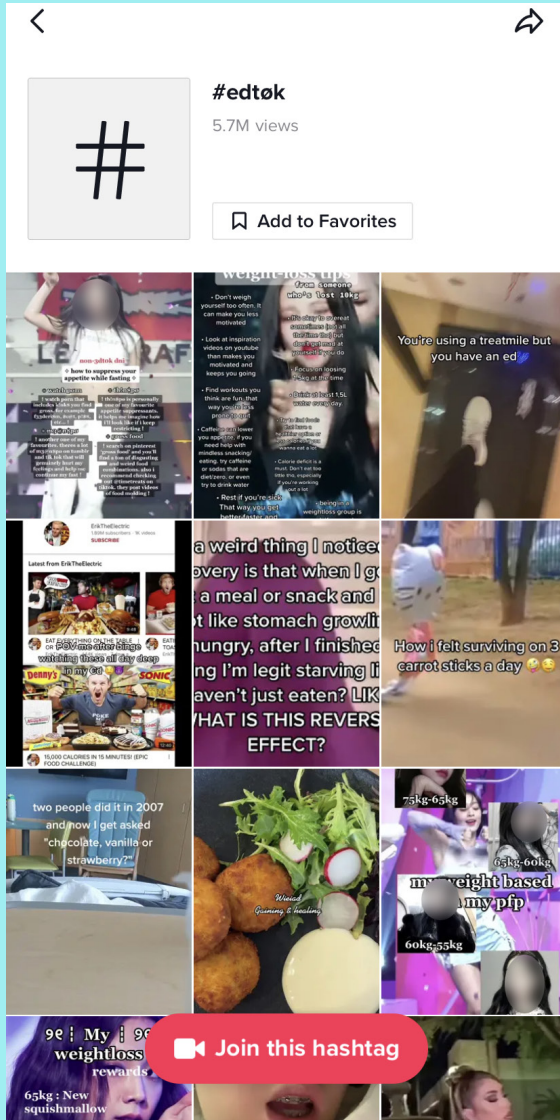
#edtøk
5.7 million views

First video reads: "non–3dtok dni" [non-eating disorder TikTok do not interact] "how to suppress your appetite while fasting", with a list of appetite suppressant tips.
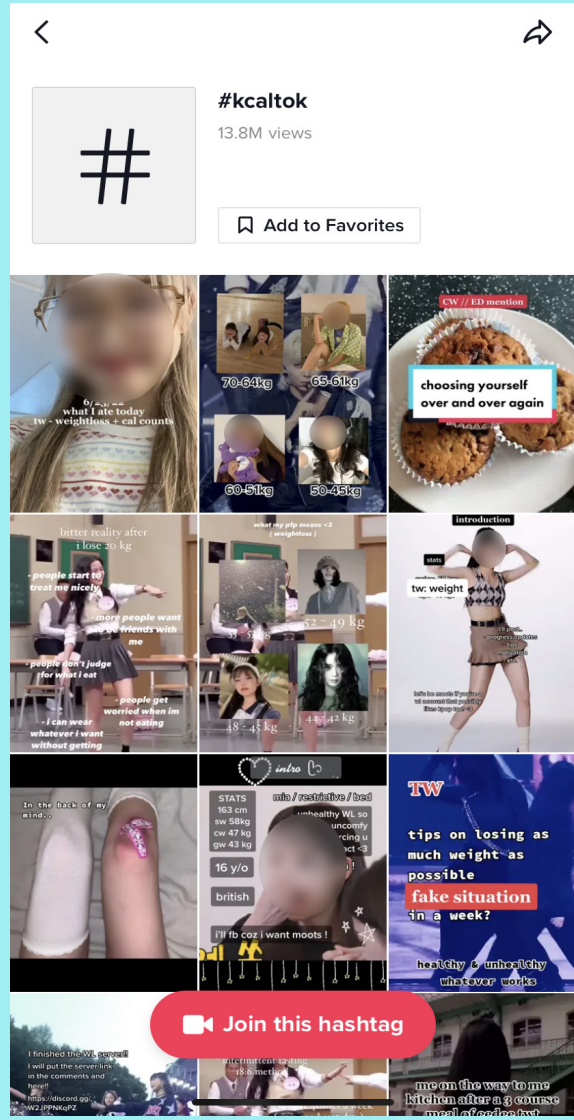
#kcaltok
13.8 million views

One video reads: "TW" "tips on losing as much weight as possible [fake situation] in a week" "healthy & unhealthy whatever works". 'Fake situation' is a disclaimer users put on videos to evade bans on TikTok.
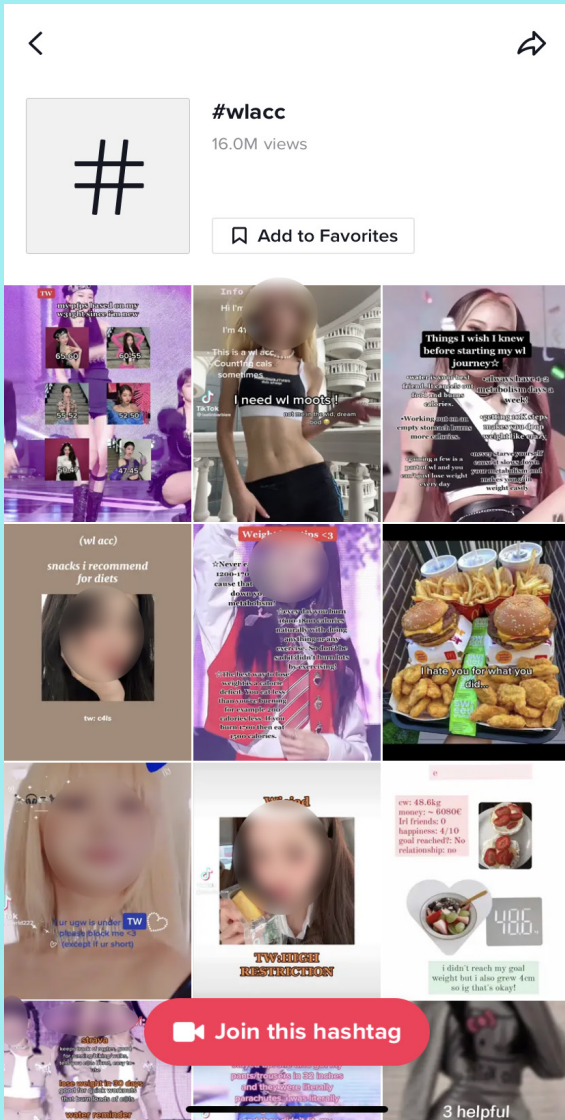
#wlacc
16 million views

First video reads: "my pfps [profile picture] based on my w31ght" with 6 profile picture options ranging from 65 kg to 45 kgs.
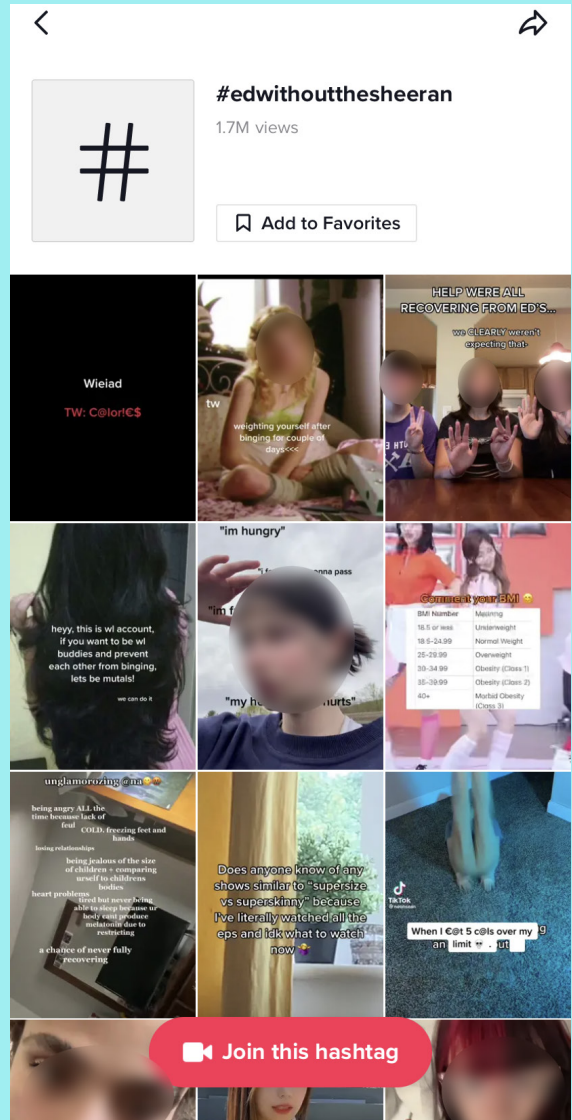
#edwithoutthe sheeran
1.7 million views

One video reads: "heyy, this is a wl account if you want to be wl buddies and prevent each other from binging lets me mutuals!" "we can do it".

10 Recommendations

Despite the promises made by Big Tech, this report shows that self-regulation has failed. Legislators, parents, and young people should be concerned that TikTok is far from being the friendly dance app it promotes itself to be to the public.

Without legislation, TikTok is behaving like other social media platforms and is primarily interested in driving engagement while escaping public scrutiny and accountability. The TikTok algorithm does not care whether it is pro-anorexia content or viral dances that drive the user attention they monetize. TikTok just wants to collect users' data and keep them on the platform, viewing advertisements.

Legislators need to act to change the incentives that shape social media companies' broken business models and allow greater transparency into their platforms. In consultation with lawmakers and legal experts worldwide, CCDH launched its STAR Framework in May 2022, a global standard for regulatory design that would help to create a safer environment for all users, including our children and grandchildren.

Our STAR Framework represents:

> **S** – Safety by Design
> **T** – Transparency of algorithms, rules enforcement, and economics
> **A** – Accountability
> **R** – Responsibility of Big Tech and senior executives

We explain the core elements of STAR in the Appendix, the full guidance can be accessed here. An explanation of how the STAR Framework relates to this report is outlined below.

Safety by Design

Safety by design is anchored in a duty of care approach – where regulation puts the public interest and public safety ahead of corporate commercial interests. Safety by design recognises that online harm leads to offline harm, frequently with tragic consequences. As a result, safety by design means taking a preventative approach and investing in the systems, processes, and teams that can identify and eliminate risks and harms to consumers. This is the basic consumer standard that we expect from other businesses in other sectors.

It is clear from the findings in this report that TikTok has not designed its product and service in a way that ensures the safety of its users. High-impact and harmful content is automatically directed to young and vulnerable users within seconds of setting up profiles on the platform and at a disturbing frequency. This is the inevitable consequence of an opaque platform and the absence of incentives and legislation focusing on safety.

Introducing a requirement to adopt safety-by-design would require companies like TikTok to conduct risk assessments of their products, policies and processes. It would ensure they design and implement transparency features within the platform (like the inputs and outputs and impacts of their algorithms). It would require built-in safety features, such as responsive reporting systems. Our approach would mean that any changes to the product or service are carefully monitored, and that risks are appropriately managed, with public safety rather than profit put first and foremost.

Transparency

Transparency is a core element of CCDH's STAR Framework and desperately lacking from social media platforms, especially TikTok. Crucially, it is clear from our research in this report that there is no transparency over the algorithm that is driving toxic and harmful content to young and vulnerable users and why, for example, there appear to be variations depending on factors such as a young person's chosen username. Greater public transparency of these elements will help everyone to understand what is happening and why, and put more pressure on TikTok to ensure that it embeds safety by design.

In the STAR Framework, we recommend that algorithmic transparency should include:

- Search algorithms and data – such as autocompleting a keyword and metadata used;
- Recommendation algorithms and data – which curate content that a user may be interested in;
- Ad-tech algorithms and data – that target users based on demographics and behavior to optimize advertising; and
- Moderation algorithms and data – that target content, users and groups that breach the law or the platform's or search engine's terms and conditions. This data should include internal metrics, such as the violative view rate.

To help assess the impact of algorithms and products and to identify emerging forms and trends of harm on platforms, the data above should be supported by public transparency on the most popular content on that platform, with clear data on the impact of algorithmic recommendations. For example, TikTok should declare which content on its platform falls into the following categories, ideally as live daily statistics:

- Most liked videos
- Most viewed videos
- Most recommended videos

A live public service has the benefits of being faster, giving broader access, providing a public record, and being harder to falsify or mislead. However, this transparency should be complemented by access to more detailed data via a public API. Within a legislative

framework, regulators and courts should have the right to access additional data to ensure legal duties are being complied with. Individuals should also have a clear right to access and share their own data.

Accountability

Accountability is necessary because Big Tech has shown that it cannot be trusted to act when it thinks no one is watching, and self-regulation has palpably failed. CCDH has released several studies that show that even when harmful content that breaches their policies is reported to platforms using their own reporting systems, they fail to act.[33] TikTok publicly makes claims about the safety of its platform that are not supported by the evidence presented in this report.

Any robust regulatory system needs accountability built in so that community standards, responsibilities and duties are upheld and problems are brought to light rather than concealed or minimized in corporate boardrooms. There are several ways to achieve this, from regulatory oversight to allowing assessment of liability through the courts system. The UK and the EU are in the process of adopting legislative models that rely on a regulator to oversee their new legislative schemes. Australia and New Zealand already have media regulators in place.

In this case, TikTok is now on notice that their algorithm is immersing young users in this harmful content, and that the situation appears to be worse if they are vulnerable. An accountability system would provide independent means to hold TikTok to account.

Responsibility

Responsibility means that there are consequences for social media companies and their senior executives for failing to fulfill their statutory duties and take reasonable steps to prevent harm. Currently, the public, including young and vulnerable users and their families, are bearing the costs of this harm. In the absence of a comprehensive statutory framework, these companies should be liable under general principles of negligence and civil law – which is the case for other companies in other sectors. Financial penalties and the risk of litigation will help to ensure that companies like TikTok are embedding safety by design and doing all that they can to ensure that their platforms are safe and harmful content is not amplified to young and vulnerable users.

Appendix: STAR Framework

Through CCDH's STAR Framework, we aim to establish key global standards for social media reform to ensure effectiveness, connectedness and consistency for a sector whose reach impacts people globally.

We need to reset our relationship with technology companies and collectively legislate to address the systems that amplify hate and dangerous misinformation around the globe. The STAR Framework draws on the most important elements for achieving this: Safety by Design, Transparency, Accountability and Responsibility. You can read the full version of CCDH's STAR Framework here.

S	**Safety** by Design: Safety by design means that technology companies need to be proactive at the front end to ensure that their products and services are safe for the public, particularly minors. Safety by design principles adopt a preventative systems approach to harm. This includes embedding safety considerations through risk assessments and decisions when designing, implementing, and amending products and services. Safety by design is the basic consumer standard that we expect from companies in other sectors.
T	**Transparency:** There are three key areas where transparency is desperately needed and should be prioritized: • Algorithms; • Rules enforcement; and • Economics, specifically related to advertising.
A	**Accountability** to democratic and independent bodies: Regulation is most effective where there are accountability systems in place for statutory duties and harm caused, particularly where there is a risk of inaction because of profit motives and commercial factors. Frequently, accountability systems include an enforcement and independent pathway for challenging decisions or omissions.
R	**Responsibility** for companies and their senior executives: The final element of the STAR Framework is responsibility – both social media and search engine companies and their senior executives that are responsible for implementing duties under a legislative framework. Responsibility means consequences for actions and omissions that lead to harm. A dual approach – targeting both companies and their senior executives – is a common intervention strategy for changing corporate behavior.

Endnotes

1   "How TikTok Ate the Internet", Washington Post, 14 October 2022. https://www.washingtonpost.com/technology/interactive/2022/tiktok-popularity/

2   "Social media firms 'monetising misery', says Molly Russell's father after inquest", The Guardian, 30 September 2022. https://www.theguardian.com/uk-news/2022/sep/30/molly-russell-died-while-suffering-negative-effects-of-online-content-rules-coroner

3   Senate Homeland Security and Governmental Affairs Committee hearing Social Media's Impact on Homeland Security, 14 September 2022, https://www.hsgac.senate.gov/imo/media/doc/Testimony-Pappas-2022-09-14-REVISED.pdf

4   "STAR Framework", Center for Countering Digital Hate, 21 September 2022, https://counterhate.com/research/star-framework/

5   Section 230 of the Communications Decency Act 1996 provides social media platforms with a shield against liability for any content produced by users on their website, even if that content is selected by the company's algorithms and published in the news feeds of billions of users.

6   "How TikTok recommends videos #ForYou", TikTok, 18 June 2020, https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

7   "How TikTok recommends videos #ForYou", TikTok, 18 July 2020, https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

8   Usernames were created by a random name generator set to provide traditionally female names

9   "Designing For Disorder: Instagram's Pro-Eating Disorder Bubble In Australia", Rys Farthing, Reset Australia, 22 April 2022, https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf

10   "In her own words - Molly Russell's secret Twitter account", BBC News, 30 September 2022, https://www.bbc.com/news/uk-62892636

11   "How to enable the GPS override function?", Surfshark, Accessed 10 November 2022, https://support.surfshark.com/hc/en-us/articles/360011517640-How-to-enable-the-GPS-override-function-

12   "Community Guidelines", TikTok, February 2022, https://www.tiktok.com/community-guidelines?lang=en#33

13   "How TikTok recommends videos #ForYou", TikTok, 18 July 2020, https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

14   "How TikTok recommends videos #ForYou", TikTok, 18 June 2020, https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

15   "British Ruling Pins Blame on Social Media for Teenager's Suicide", The New York Times, 1 October 2022, https://www.nytimes.com/2022/10/01/business/instagram-suicide-ruling-britain.html

16   "Molly Russell: Instagram posts seen by teen were safe, Meta says", BBC News, 26 September 2022, https://www.bbc.com/news/uk-england-london-63034300

17   "British Ruling Pins Blame on Social Media for Teenager's Suicide", The New York Times, 1 October 2022, https://www.nytimes.com/2022/10/01/business/instagram-suicide-ruling-britain.html

18   "Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show", Wall Street Journal, 14 September 2021, https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

19 "Nutrition content is huge on TikTok, but a new study shows trends like #WhatIEatInADay may influence disordered eating behaviors", CNBC, 9 November 2022, https://www.cnbc.com/2022/11/09/tiktok-nutrition-related-content-may-contribute-to-disordered-eating.html

20 "Strengthening privacy and safety for youth on TikTok", TikTok, 13 January 2021, https://newsroom.tiktok.com/en-us/strengthening-privacy-and-safety-for-youth

21 Ward ZJ, Rodriguez P, Wright DR, Austin SB, Long MW. Estimation of Eating Disorders Prevalence by Age and Associations With Mortality in a Simulated Nationally Representative US Cohort. JAMA Netw Open. 2019;2(10):e1912925. doi:10.1001/jamanetworkopen.2019.12925

22 "Suicide statistics", American Foundation for Suicide Prevention, retrieved 15 November 2022, https://afsp.org/suicide-statistics/

23 "Self-Harm Statistics and Facts", The Recovery Village, 2 May 2022, https://www.therecoveryvillage.com/mental-health/self-harm/self-harm-statistics/

24 "Designing For Disorder: Instagram's Pro-Eating Disorder Bubble In Australia", Rys Farthing, Reset Australia, 22 April 2022, https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf

25 Basch C, Donelle L, Fera J, Jaime C, Deconstructing TikTok Videos on Mental Health: Cross-sectional, Descriptive Content Analysis, JMIR Form Res 2022;6(5):e38340, https://formative.jmir.org/2022/5/e38340

26 "Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show", Wall Street Journal, 14 September 2021, https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

27 "Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show", Wall Street Journal, 14 September 2021, https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

28 Basch C, Donelle L, Fera J, Jaime C, Deconstructing TikTok Videos on Mental Health: Cross-sectional, Descriptive Content Analysis, JMIR Form Res 2022;6(5):e38340, https://formative.jmir.org/2022/5/e38340

29 "Designing For Disorder: Instagram's Pro-Eating Disorder Bubble In Australia", Rys Farthing, Reset Australia, 22 April 2022, https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf

30 "In her own words – Molly Russell's secret Twitter account", BBC News, 30 September 2022, https://www.bbc.com/news/uk-62892636

31 The Dangers of Pro-Ana and Pro-Mia, Beat, Accessed November 7 2022, https://www.beateatingdisorders.org.uk/get-information-and-support/about-eating-disorders/dangers-of-pro-ana-and-pro-mia/

32 Rys Farthing, Designing For Disorder: Instagram's Pro-Eating Disorder Bubble In Australia, Reset Tech, 22 April 2022, https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf

33 "Marketplace flagged over 800 social media posts with COVID-19 misinformation. Only a fraction were removed", CBC, 30 March 2021, https://www.cbc.ca/news/marketplace/marketplace-social-media-posts-1.5968539

    "Failure to Protect: Anti-Muslim Hate", Center for Countering Digital Hate, 28 April 2022, https://counterhate.com/research/anti-muslim-hate/

    "Failure to Protect", Center for Countering Digital Hate, 30 July 2021, https://www.counterhate.com/failuretoprotect

    "Instagram fails to take down more than 94% of racist abuse accounts targeting England players after Euros", iNews, 15 July 2021, https://inews.co.uk/news/technology/instagram-racist-abuse-posts-england-players-after-euros-1102896

"Twitter fails to delete 99% of racist tweets aimed at footballers in run-up to World Cup", The Guardian, 20 November 2022, https://www.theguardian.com/technology/2022/nov/20/twitter-fails-to-delete-99-of-racist-tweets-aimed-at-footballers-in-run-up-to-world-cup

"Hidden Hate", Center for Countering Digital Hate, 6 April 2022, https://www.counterhate.com/hiddenhate

"I get abuse and threats online – why can't it be stopped?", BBC, 18 October 2021, https://www.bbc.co.uk/news/uk-58924168

"New research shows Metaverse is not safe for kids", Center for Countering Digital Hate, 30 December 2021, https://counterhate.com/blog/new-research-shows-metaverse-is-not-safe-for-kids/

"Twitter fails to remove 100 abusive misogynists", The Times, 13 January 2022, https://www.thetimes.co.uk/article/twitter-fails-to-remove-100-abusive-misogynists-z7nwg6d9t

Deadly by Design

Published 15 December 2022