



# DRAGGED INTO THE RABBIT HOLE

NEW EVIDENCE OF TIKTOK'S RISKS TO CHILDREN'S MENTAL HEALTH

AMNESTY  
INTERNATIONAL



**Amnesty International is a movement of 10 million people which mobilizes the humanity in everyone and campaigns for change so we can all enjoy our human rights. Our vision is of a world where those in power keep their promises, respect international law and are held to account. We are independent of any government, political ideology, economic interest or religion and are funded mainly by our membership and individual donations. We believe that acting in solidarity and compassion with people everywhere can change our societies for the better.**

© Amnesty International 2025

Except where otherwise noted, content in this document is licensed under a Creative Commons (attribution, non-commercial, no derivatives, international 4.0) licence.

<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

For more information please visit the permissions page on our website: [www.amnesty.org](http://www.amnesty.org)

Where material is attributed to a copyright owner other than Amnesty International this material is not subject to the Creative Commons licence.

First published in 2025

by Amnesty International Ltd

Peter Benenson House, 1 Easton Street,

London WC1X 0DW, UK



*Cover illustration: © Marine Fustec*

Index: POL 40/0360/2025

Original language: English

[amnesty.org](http://amnesty.org)

**AMNESTY  
INTERNATIONAL**



**CONTENT WARNING** This report covers sensitive issues and contains examples of content which include references to self-harm and suicide. Contacts for helplines and organizations providing free emotional support around the world can be found in Amnesty International's guide *Staying Resilient While Trying to Save the World (Volume 2): A Well-being Workbook for Youth Activists*.<sup>1</sup>

### **Acknowledgements**

Amnesty International would like to express its gratitude to members of the Algos Victima collective who participated in this research as well as our technical partners from the Algorithmic Transparency Institute.

### **The Algos Victima collective**

On 12 September 2023, a first French complaint against TikTok was filed for provocation to suicide by Marie Le Tiec's family, represented by Laure Boutron-Marmion. In 2024, the collective "Algos Victima" was created to support minors, and their families, who were exposed to social media harms through legal proceedings before French institutions. In November 2024, seven families filed a class action lawsuit against TikTok for exposing their children to videos that trivialize or even idealize depression, self-harm, or suicide. Today, the collective brings together eleven French families mobilized to establish TikTok's responsibility for the deterioration of their children's mental and physical health.

### **Algorithmic Transparency Institute**

The Algorithmic Transparency Institute (ATI) is a project of the National Conference on Citizenship. The Algorithmic Transparency Institute brings greater transparency to the digital platforms that impact civic discourse. ATI develops tools, collects data, produces research, and supports analysis that furthers a shared understanding of the relationship between online conversation and our democracy. ATI enables researchers, journalists, and civil society advocates to track and understand our digital civic discourse across social media.

Levi Kaplan and Piotr Sapiezynski of Northeastern University assisted with the automated testing and data analysis.

---

<sup>1</sup> Amnesty International, *Staying Resilient While Trying to Save the World (Volume 2): A Well-Being Workbook for Youth Activists* (Index: ACT 10/3231/2020), January 2021, [amnesty.org/en/documents/act10/3231/2020/en/](https://www.amnesty.org/en/documents/act10/3231/2020/en/), p. 99.

# CONTENTS

<b>1. EXECUTIVE SUMMARY</b>	<b>5</b>
<b>2. METHODOLOGY</b>	<b>7</b>
<b>3. MIXED AWARENESS OF SOCIAL MEDIA RISKS AMONGST YOUNG PEOPLE IN FRANCE</b>	<b>9</b>
<b>4. TESTIMONIES OF YOUNG SURVIVORS AND AFFECTED AND BEREAVED PARENTS</b>	<b>11</b>
4.1 HOW TIKTOK'S 'FOR YOU' FEED PUSHES STRUGGLING CHILDREN AND YOUNG PEOPLE DEEPER INTO DEPRESSION, SELF-HARM AND SUICIDAL IDEATION	11
4.2 THE DEADLY CONSEQUENCES OF TIKTOK'S SUICIDAL IDEATION RABBIT HOLES	14
4.3 UNDERSTANDING TIKTOK'S HARMFUL DESIGN AND FAILING CONTENT MODERATION IN THE CONTEXT OF YOUNG PEOPLE'S AND PARENTS' LIVED EXPERIENCES	15
<b>5. TECHNICAL RESEARCH EVIDENCE</b>	<b>18</b>
5.1 MANUAL TESTS	18
5.2 AUTOMATED TESTS	20
<b>6. HUMAN RIGHTS FRAMEWORK</b>	<b>24</b>
<b>7. EU AND NATIONAL LEGAL FRAMEWORKS</b>	<b>27</b>
7.1 EUROPEAN UNION REGULATION: THE DIGITAL SERVICES ACT	27
7.2 THE ONLINE PROTECTION OF MINORS IN FRANCE	29
<b>8. TIKTOK'S FAILURE TO MAKE THE PLATFORM SAFE FOR CHILDREN AS A BREACH OF THE DIGITAL SERVICES ACT AND INTERNATIONAL HUMAN RIGHTS STANDARDS</b>	<b>31</b>
<b>9. CONCLUSION AND RECOMMENDATIONS</b>	<b>34</b>

# 1. EXECUTIVE SUMMARY

- TikTok's 'For You' feed drags teen accounts into deeply harmful "rabbit holes" two years after Amnesty International first raised the alarm: In a new experiment, French teen accounts signalling an interest in content related to sadness or mental health struggles were drawn into depressive content within less than an hour of scrolling through TikTok's 'For You' feed.
- Within three to four hours, accounts were shown videos romanticizing suicide or showing young people sharing their intention to end their lives, including information on suicide methods.
- Testimonies from young survivors of self-harm and bereaved parents reveal the catastrophic impact on struggling adolescents' mental and physical health.

In November 2023, Amnesty International published a damning report highlighting how easily children and young people expressing an interest in mental health could be drawn into "rabbit holes" of depressive and suicidal content in TikTok's 'For You' feed. The research revealed how TikTok employs addictive platform design to maximize engagement, undermining the right to privacy, freedom of thought and mental and physical health.

Since then, Amnesty International, together with young people from Argentina to Ireland, Kenya and Pakistan, has been campaigning for change. More than 125,000 people have signed a petition calling on TikTok to make the platform safer. And yet, whilst regulators are all too slowly beginning to tackle the harms of the Big Tech industry, TikTok is still failing to address its addictive design and its harmful "rabbit holes", prioritizing profits over children's safety.

Nowhere did Amnesty International's global campaign attract as much support and public debate as in France. For all too many young people and their families, these harms were not abstract, they were - and are - their lived experience. Between 2024 and 2025, eleven French families joined a class action lawsuit against TikTok to establish, and seek accountability for, the company's responsibility for the deterioration of their children's mental and physical health, including a contributing role in the deaths of 15-year-old Marie Le Tiec and Charlize Dapui-Parkiet. This follow-up report brings together testimony from some of these families and young survivors and renewed qualitative and quantitative evidence of TikTok's continued failure to address the risks and harms of its business model.

The evidence presented in this briefing documents TikTok's failure to address its systemic design risks for children and young people both under international business and human rights standards as well as under the company's binding obligations under the EU's Digital Services Act. It is an urgent appeal to the company itself, but also to EU and French regulators to take decisive action to force the company to respect children's and human rights to privacy, freedom of thought and health.

Amnesty International shared its key findings with TikTok and offered it the opportunity to comment, but the company did not provide a response ahead of publication.

To address the risks identified in this briefing, Amnesty International calls on the European Commission to:

- Include these findings in its ongoing investigation into TikTok's potential infringements of its obligations under the DSA and ensure a timely and effective investigation and enforcement process,

and on TikTok to:

- Transition to a rights-respecting business model that is not based on invasive data tracking. As a first step, TikTok must ensure that its human rights due diligence policies and processes address the systemic and widespread human rights impacts of its business model, in particular the right to privacy, the right to freedom of thought and the right to health.
- Stop maximizing “user engagement” at the expense of its users’ health and other human rights. As part of its human rights due diligence process, TikTok must identify design elements in cooperation with users, including children and young people, and independent experts, which encourage addictive platform use and social comparison, and replace these with a user experience that is focused on ‘safety by design’ and the best interests of the child.
- Cease collecting intimate personal data and drawing inferences from a user’s watch time and engagement about their interests, emotional state or well-being for the purposes of ‘personalizing’ content recommendations and ad targeting. Rather than using pervasive surveillance to adapt feeds to a user’s interests, TikTok should enable users to communicate their interests through deliberate prompts (for example, users could be asked to enter specific interests if they would like to be served personalized recommendations) and only when based on users’ freely given, specific and informed consent.

# 2. METHODOLOGY

This research is a follow-up project to Amnesty International's 2023 report, *Driven into the Darkness*, which documented the serious human rights risks and harms caused and contributed to by TikTok's engagement-based business model for children and young people.<sup>2</sup> The report found that TikTok had created an addictive platform, despite mounting evidence of the serious health risks associated with children's compulsive use of social media. Examining TikTok's hyper-personalized 'For You' feed more closely, the research revealed how children expressing an interest in mental health-related content could easily be drawn into "rabbit holes" of harmful content, risking exacerbating children and young people's struggles with depression and self-harm. Nowhere did Amnesty International's subsequent global campaign to make TikTok safer attract as much support and public debate as in France, where TikTok is regulated under the European Union's Digital Services Act (DSA), requiring the company to identify and mitigate systemic risks to children's rights, since 2023. Responding to the growing evidence base of TikTok's risks for children and young people, the European Commission launched an investigation into TikTok's potential infringements of the DSA in February 2024. Amidst this ongoing investigation, *Dragged into the rabbit hole*, a renewed collaboration with the National Conference on Citizenship's Algorithmic Transparency Institute, builds on the human rights analysis and technical research methodology of the 2023 report to present new evidence of TikTok's continued failure to make the platform safe for children and young people.

## Interviews

Through a collaboration with the Algos Victima collective, Amnesty International researchers interviewed three young survivors of depression and/or self-harm about the role of TikTok in exacerbating their mental health struggles as well as eight affected mothers and fathers (including Stéphanie Mistre, mother to Marie Le Tiec, and the parents of Charlize Dapui-Parkiet, who lost their daughters to suicide) and two close friends of Marie's who shared perspectives on Marie's experiences in relation to her use of TikTok as well as their own. Amnesty International France also interviewed the neurologist Servane Mouton who spoke about the effects of TikTok on adolescents' cognitive development.

To assess and mitigate the risk of retraumatization and discuss psychological support, researchers had preparatory conversations with the parents of the young survivors interviewed for this research. Parental consent was sought and granted for all participants under the age of 18 and informed consent was taken from each adolescent. Participants over the age of 18 gave their informed consent in advance of the interviews. To protect the identity of our interviewees, this report uses pseudonyms for those research participants who chose to remain anonymous, these are marked with an asterisk.

---

<sup>2</sup> Amnesty International, *Driven into the Darkness: How TikTok's 'For You' Feed Encourages Self-Harm and Suicidal Ideation* (Index: POL 40/7350/2023), 7 November 2023, <https://www.amnesty.org/en/documents/POL40/7350/2023/en/>

## Survey data

To examine public attitudes towards social media and risk awareness amongst young people in France, Amnesty International France collaborated with the Harris Institute in late 2024 polling 1,000 young people between the ages of 13 to 25. The results can be found in the following chapter.

## Manual tests

Amnesty International set up three fake teen accounts, two female, one male, registered as 13-year-olds on TikTok in France to manually track the “rabbit hole” effect. The researcher scrolled through the ‘For You’ feed and watched posts twice if they related to sadness or mental health issues, skipping other content, for a total of three to four hours. This process was screen-recorded and subsequently manually analysed to better understand the types and scale of algorithmic amplification of mental health-related content recommended to children on TikTok.

## Automated tests

Watch history data from one of the manual accounts was then used to create automated TikTok accounts in addition to automated baseline accounts (watching only benign content), all registered as 13-year-olds, to study the personalization or “rabbit hole” effect at greater scale. Researchers used partitions of a watch history containing up to 387 videos in total to test the response of TikTok’s algorithmic recommender system to different ratios of previously watched sad/depressive and self-harm-related videos.

Before the automated testing began, researchers used the TikTok data export functionality to collect the full watch history of one of the manual accounts. The downloaded file contains the IDs of all videos that were recommended to the account. Researchers assume that the difference in time between recommendations of up to 5 minutes indicates the user watched a video for that time. Differences above five minutes indicate separate sessions (as the account was manually run across multiple sessions), and the last video from each session was ignored as its watch time cannot be inferred.

Researchers then replicated the manual account as follows. First, researchers created a new TikTok account located in France (using a VPN). They logged into that account on a physical Android device, then programmatically ensured that it watched the first pre-defined number of videos from the source account history, each for the same amount of time as the source account did, a process we refer to as ‘cloning’. Researchers waited 30 minutes to give the recommendation algorithm time to update its user model. They then repeatedly collected the next eight recommendations until they collected at least 201 video recommendations without sending any further watch (or skip) decisions back to TikTok, a process researchers refer to as ‘fetching’; this way they avoided further influencing the recommendation algorithm.

After the recommendations for a given account were collected, researchers cleared the TikTok app data, uninstalled the app, changed the Android Advertising ID (which is a semi-persistent device ID that allows app developers to identify users between installations of different apps), and repeated the process with another new account.

## Communications with TikTok

Amnesty International wrote to TikTok in March 2025 to ask what measures the company had put in place since the publication of *Driven into the Darkness* in late 2023. TikTok’s response listed familiar ‘well-being’ measures, most of them already in place when the research was conducted and failed to acknowledge the app’s “rabbit hole” problem. It also failed to produce evidence of any new targeted measures to address it.<sup>3</sup> Ahead of publishing this follow-up report, Amnesty International shared its key findings with TikTok. *Amnesty International shared its key findings with TikTok and offered it the opportunity to comment, but the company did not provide a response ahead of publication.*

---

<sup>3</sup> Amnesty International, “TikTok fails to address risks to children and young people’s mental health despite past warnings”, 12 May 2025, <https://www.amnesty.org/en/documents/act10/9354/2025/en/>

# 3. MIXED AWARENESS OF SOCIAL MEDIA RISKS AMONGST YOUNG PEOPLE IN FRANCE

A study commissioned by Amnesty International France and carried out at the end of 2024 by the Harris Institute polling young people aged between 13 and 25 shows that many young people are aware of the addictive effect of social media platforms and some of the dangers they present. More than 8 out of 10 young people reported spending too much or far too much time on social media, with an even higher rate amongst young women (88%). In research conducted in 2023, Amnesty International had already highlighted the particularly addictive design of TikTok, a social media platform that is particularly popular with young people. Amnesty International's interviews with children and young people in 2023 and 2025 show however that awareness of social media's addictive effects often does not translate into an ability to stop the excessive use thereof.

Moreover, more than half of the young people surveyed, especially women aged between 16 and 25, say they are often exposed to disturbing content on social networks, including content showing idealized bodies or promoting eating disorders. The study reveals that 58% of young people surveyed say they are affected by this content. Specifically, they report feeling uncomfortable, sad, demoralized or affected in their self-esteem. For women aged 16 to 21, this figure rises to 75%.

While a large majority of young people surveyed (69%) have already sought to avoid being exposed to toxic content, only one in five young people report having succeeded in doing so. For vulnerable people, these "rabbit holes" of disturbing content can have serious consequences in real life, as Amnesty International's interviews with young people struggling with mental health issues in 2023 and 2025 highlight.

Finally, the study shows that almost half of the young people surveyed are unaware that the content offered to them on social media platforms is personalized based on their personal data and interests. This lack of knowledge is particularly marked among young people aged 13 to 15 (64%). Research conducted by Amnesty International in 2023 showed that TikTok intrudes on the privacy of its users, tracking everything they do on the platform, in order to infer their interests, emotional state and level of well-being to serve them "personalized" content in the 'For You' feed.<sup>4</sup> TikTok, like all major social

---

<sup>4</sup> Amnesty International, I feel exposed: Caught in TikTok's Surveillance Web (Index: POL 40/7349/2023), 7 November 2023, <https://www.amnesty.org/en/documents/POL40/7349/2023/en/>

media companies, aims to maximize engagement and time spent on the platform in order to collect ever more personal data. This data allows it to display targeted advertisements and earn money.<sup>5</sup> Amnesty International has repeatedly documented serious human rights harms associated with this surveillance-based business model, leading Big Tech companies including Meta, X and TikTok to prioritize engagement and profits over platform safety.<sup>6</sup>

---

5 The EU's Digital Services Act bans targeted advertisements towards minors. The vast majority of TikTok's young user base around the world can however still be shown targeted ads, leaving the basis of TikTok's business model largely unaffected.

6 Amnesty International, Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights (Index: POL 30/1404/2019), 21 November 2019, <https://www.amnesty.org/en/documents/pol30/1404/2019/en>, Amnesty International, The Social Atrocity: Meta and the Right to Remedy for the Rohingya (Index: ASA 16/5933/2022), 29 September 2022, <https://www.amnesty.org/en/documents/asa16/5933/2022/en>, Amnesty International, I feel exposed: Caught in TikTok's Surveillance Web (Index: POL 40/7349/2023), 7 November 2023 (previously cited), Amnesty International, A thousand cuts: technology-facilitated gender-based violence against Poland's LGBTI community on X (Index: EUR 37/0098/2025), 1 September 2025, <https://www.amnesty.org/en/documents/eur37/0098/2025/en/>

# 4. TESTIMONIES OF YOUNG SURVIVORS AND AFFECTED AND BEREAVED PARENTS

## 4.1 HOW TIKTOK'S 'FOR YOU' FEED PUSHES STRUGGLING CHILDREN AND YOUNG PEOPLE DEEPER INTO DEPRESSION, SELF-HARM AND SUICIDAL IDEATION

To gain a better understanding of the risks of TikTok's hyper-personalized 'For You' feed for adolescents and young adults who struggle with common mental health issues, often linked to experiences of bullying and social exclusion, Amnesty International spoke with affected young people and their parents in France. Researchers also spoke to parents and friends of Marie who died by suicide at the age of 15 in 2021 and Charlize who died at the same age in November 2023. They shared their stories to raise awareness of risks which they felt were still poorly understood by parents, schools, politicians and even mental health professionals, and to demand accountability from TikTok and decisive action from regulators.

**Maëlle**, 18 years old, told Amnesty International researchers how she got drawn into a "rabbit hole" of depressive and self-harm content in TikTok's 'For You' feed shortly after signing up in early 2021.<sup>7</sup> A friend had encouraged her to install the app so they could share funny and entertaining videos with each other at a time when Maëlle was struggling with her mental health after being bullied at school.

*"At first, I used to go on there to have fun. I got these videos that were really the image we had of the app [...] not very informative, but entertaining. And then there was a song that came back [into the feed] lots of times because it had just come out. And gradually, I became interested in the lyrics, and I felt something inside me, I found that it touched me. The singer talks about his struggles and dark moments. I had already started to self-harm. And so, I watched more and more videos with this music. [...] And bit by bit, it became something darker and darker, like death maybe isn't such a bad idea."*

Within weeks of signing up, her TikTok feed turned into a stream of videos that normalized and even encouraged self-harm and suicide. She estimates that up to three quarters of the recommended videos related to mental health struggles in ways that dragged her further into self-harming behaviours. In September 2022, Maëlle was hospitalized for the first time. Yet, neither her parents nor her psychologists knew about what she was seeing online.

---

7 Interview, Paris, 8 July 2025

*"There is a part of the content that is glamorous. Young people flirting with the forbidden. And there are some really trashy things, people in the hospital or not doing well. It feels like they're on a tightrope. It's not the same goal but in the end, it still has the same effect on the user. [...] It attracted me even more; it kept my head underwater. And it made me feel guilty about not doing as badly as the people on the platform."*

*"On the one hand, I felt a sense of belonging. I felt less alone because I told myself that other people were going through these things. But for all that, I knew it was unhealthy and very vicious."*

*"There was never anything encouraging. Like go see someone if you feel bad, things like that. It was always something very dark."*

In the winter of late 2023/early 2024, Maëlle's mental health declined further.

*"There was a time when I didn't have the strength to do anything. I didn't play sports anymore. I was barely going to class anymore. To keep me busy, I used to go on TikTok, and it kept my head underwater."*

*"Seeing people who cut themselves live [on TikTok], people who say what to take as medication to end it, it influences and encourages you to harm yourself. There are videos [...] that are still burnt into my retina. There are things, behaviours that remain."*

Maëlle's mother, Morgane, only learned about this type of content in 2024 when a radio programme reporting on the beginnings of *Algos Victima* provoked a conversation with her daughter about her own experiences.<sup>8</sup> She told Amnesty International:

*"When I discovered this content on TikTok, it was a bit of a shock. But I quickly understood what she had locked herself into. The shame and guilt she felt for not being able to tell us about it. [...] But I couldn't even imagine that it could exist [...] She became addicted to deadly content. Addiction impacts your social life, health, work and schooling. Clearly, it had impacted all these environments [...] No matter what we did, no matter what we said, it's as if she was caught up in this parallel world and the real world no longer had an impact on her emotions or who she was."*

17-year-old **Édouard** described the same captivating effect of harmful TikTok content. He was hospitalized after suffering from severe depression. He regularly watched TikTok videos reposted on Instagram.<sup>9</sup>

*"I was in a deep depression. Little by little, I felt even worse, and I wanted to go further, stronger, deeper. And little by little, I descended into the abyss [...] I came across videos of people cutting themselves. Not of the act, but of the scars. I was thinking "I have to reach this level of scars or I'm not happy". I really wanted to be marked, deeper and deeper [...] I wanted to see myself destroyed, I no longer loved myself. So, I might as well destroy myself completely, mentally and physically."*

Édouard's mother, Géraldine, believes that these videos played a crucial role in exacerbating his condition:

*"Édouard wasn't feeling very well but the online videos were a catalyst. I'm not saying this is the origin of his discomfort, but it has increased it tenfold, like a machine that sucked him in."<sup>10</sup>*

Though he mostly watched TikTok content that was reposted to Instagram, he told Amnesty International that comparing the availability of self-harm and suicide-related content on the two platforms recently, on TikTok "it's really much easier to find. You directly hit the flag of Switzerland or the zebra and you come across these videos." (The Swiss flag is popular "algospeak" for suicide, the zebra stands for self-harm.)

---

8 Interview, Paris, 8 July 2025

9 Interview, near Toulon, 29 July 2025

10 Interview, near Toulon, 29 July 2025

17-year-old **Maëlllys** similarly described how she was sucked into a TikTok community completely separated from her real life, which reinforced her suicidal thoughts:

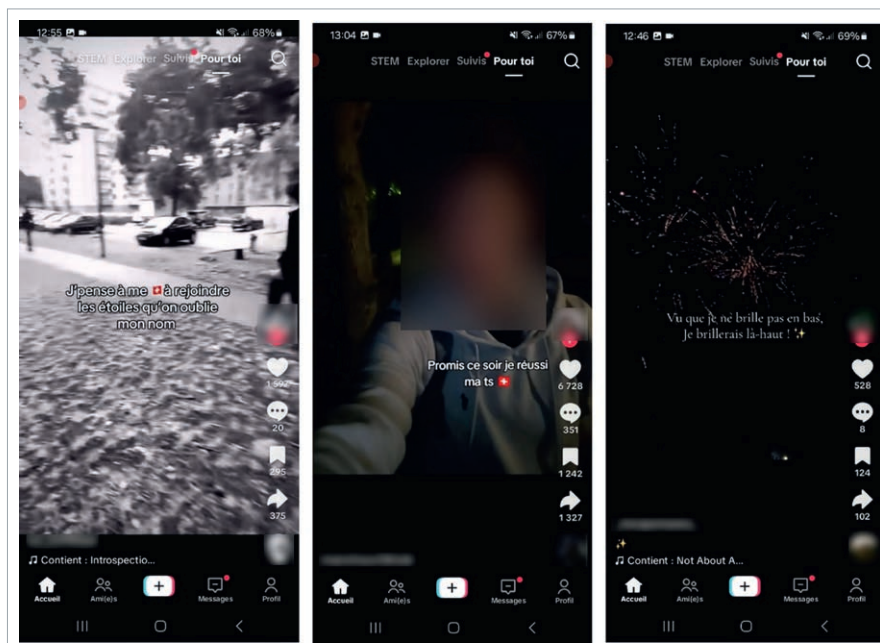
*“Initially, I watched dances, Disney playbacks, reality TV, movie scenes. [...] In 7th grade [secondary school in France, aged 12-13], I started to get sad things, I liked these videos very quickly because I related to them. Then TikTok recommended them to me again and all the contents of my feed were sad, spoke of depression [...]. It was really catastrophic, but at that moment, we don't realize it because we tell ourselves that we are not alone, that it happens to other people. We feel listened to. But the problem is that we close in on ourselves because we tell ourselves that only the people on TikTok can understand us, listen to us and experience the same things as us.”*

*“The more I watched, the more 'solutions' I found, for example hiding places for razor blades. [...] The more you watch, the deeper you sink, the more you indulge in your discomfort. You have no memory of before, when you were fine, when you felt happy. You tell yourself that you have always been unhappy. [...] We stay in there, we continue and it's a vicious circle.”*

*“After I was discharged from the hospital, I continued to go on TikTok. Always the same things, but worse [...]. There were challenges like “Labello” [a popular lip balm]: You wear Labello [lip balm] every day and the day the lip balm is empty, you kill yourself.”*

The “Labello challenge” or “lipbalm trend” became subject to widespread media reporting in 2022; TikTok denied the existence of related TikTok videos at the time.<sup>11</sup> Yet Amnesty International researchers came across two such videos in the feed of a manually administered test account in July 2025 (see section 5.1).

Maëlllys is now finally able to distance herself and break the vicious circle she described after years of counselling and hospitalizations linked to self-harm and suicide attempts. She still uses TikTok but says that aged 17, she now feels more capable of avoiding being dragged back into a spiral of harmful content when videos related to self-harm or suicide appear in her feed.



←  
Examples of recommended videos shown to manually run test accounts (registered as 13-year-olds) within the first four hours of signing up and signalling interest in sad and mental health-related content. Video caption one translates to “I’m thinking of [Swiss flag = killing myself] to join the stars so my name is forgotten”, video 2 translates to “Promised tonight I will succeed in my ts [suicide attempt]”, video 3 translates to “Since I don’t shine down here, I’ll shine up there”.

<sup>11</sup> Euronews, “A new self-harm Tik Tok ‘challenge’ is worrying France. But is the panic missing the point?”, 20 May 2022, <https://uk.news.yahoo.com/self-harm-tik-tok-challenge-135716162.html>

## 4.2 THE DEADLY CONSEQUENCES OF TIKTOK'S SUICIDAL IDEATION RABBIT HOLES

Marie and Charlize did not escape the rabbit holes of suicidal ideation that their TikTok feeds dragged them into, their parents told Amnesty International. Both families only discovered the hidden online lives of their children after their deaths by suicide in 2021 and 2023. Stéphanie Mistre, Marie's mother, and two of Marie's closest friends shared that **Marie** seemed to be doing well to all who knew her in her offline life. She was always described as a sunny person.<sup>12</sup> She had no history of self-harm. Yet unbeknownst to her mother, Marie had shared her mental health struggles in TikTok videos seemingly looking for support and validation from the online community in her feed.

*"I didn't know what TikTok was at all. Until Marie passed and, a month later, I went to check her phone. There, I fell into horror: songs that advocate suicide, the creepy, always dark contents, with a voice that tells you that you are not well, that you feel alone. Things that get you in the mood, not to mention tutorials on how to cut yourself or how to tie a knot to hang yourself."*

Stéphanie believes that Marie was dragged under by this stream of suicidal ideation content and the toxic reinforcement she was exposed to on the platform. In one instance, she received hundreds of likes in response to a video she posted of herself crying and talking about being unwell. Stéphanie was utterly shocked by the comments in her daughter's feed, which romanticized and openly encouraged suicide.

*"For these platforms, our children become products instead of human beings. They use our children as products with an algorithm and a filter bubble, using their emotions to capture them. The algorithm will capture your interests, which is not normal. They enter the child's private life. But children have rights. They take advantage of their fragility and weakness, of adolescence, which is recognized by all psychologists as a fragile period when young people are searching for themselves, and even more so for girls. This was the case for Marie."*

In the wake of Marie's death, her friends **Manon** and **Hugo** were themselves exposed to the harmful effects of TikTok's hyper-personalization:

*"When you're not well, you only have videos that aren't well. Because it keeps us in the mood. After Marie passed, for two months, I only had videos about death and suicide."<sup>13</sup> (Manon)*

*"At one point, I was not very well. And we find comfort in the videos, we like them, and they come like a tsunami."<sup>14</sup> (Hugo)*

**Charlize** was sensitive, she wanted to become a psychologist to help others.<sup>15</sup> The Covid lockdowns were a difficult time for her and her family, her paternal grandfather died, Charlize's mother was at the forefront of the fight against the pandemic working at the local hospital. When Charlize experienced bullying at school, she began to spend more time online, but her father says he'd see Charlize laugh at videos on her phone and was glad that she found something to make her smile rather than focus on dark thoughts. She was seeing a psychologist as she had begun to self-harm and even tried to end her life. But like in Marie's case, Charlize's parents Delphine and Jérémy did not learn about their daughter's online life and the feelings she was expressing online until it was too late. She ended her life in November 2023 at the age of 15.

---

12 Interviews, Paris, 16 July 2025

13 Interview, Paris, 16 July 2025

14 Interview, Paris, 16 July 2025

15 Interview, Nice, 29 July 2025

The following day, Delphine and Jérémy were told that Charlize had reposted a video romanticizing explicit suicidal thinking on TikTok shortly before her death. As they began to discover, she had been reposting suicidal ideation videos for months. “TikTok kept our daughter unwell”, Jérémy told Amnesty International.<sup>16</sup> “How can they deny the fact that content that says ‘you have problems? Kill yourself, it will be much better’ encourages you to do so? [...] Reporting videos doesn’t work. In addition, the problem is the functionality of the infinite scroll. You would have to spend your days reporting. It can’t be the responsibility of the users.”

These observations were echoed by Sophie\*,<sup>17</sup> mother of Emma\* who was exposed to harmful content on self-harm and eating disorders on TikTok, which Sophie\* believes to have exacerbated her daughter’s struggles with these issues. Despite putting a time limit through parental controls in place, she observed her daughter being dragged into a “spiral” of content that normalized mental health struggles to the extent that adolescents like her daughter “build an identity around” being unwell.<sup>18</sup> She felt that neither the available parental controls her family had used, nor TikTok’s content moderation were sufficient to address the systemic risks her daughter was exposed to.

*“Today, my daughter reports any content she deems dangerous. But she says they are never deleted, that nothing ever happens.”*

Her daughter continued to use TikTok under a strict time limit of 15 minutes per day until early 2025 when she decided to quit using the app. Up until she stopped using the app, she continued to be exposed to and report what she perceived to be harmful content.

## 4.3 UNDERSTANDING TIKTOK’S HARMFUL DESIGN AND FAILING CONTENT MODERATION IN THE CONTEXT OF YOUNG PEOPLE’S AND PARENTS’ LIVED EXPERIENCES

The testimonies of these affected young people and the families and friends of Marie and Charlize expand on the testimonies of children, young people and mental health professionals interviewed by Amnesty International in 2023 for *Driven into the Darkness*. As Amnesty International found then, TikTok’s design risks are systemic and inherently linked to its business model. TikTok, like its key competitors in the social media platform market, seeks to maximize the time spent by new and existing users of the platform to generate advertising revenue. To do so, it has built in addictive features, first and foremost amongst these the deeply privacy-invasive ‘For You’ feed recommendation system, which turns a user’s watch and engagement data into signals from which the system infers which content will keep the user’s eyes fixed on the screen for as long as possible, aided by the infinite scroll and autoplay features. Together, these addictive design features encourage unhealthy use patterns, which medical professionals have linked to serious health risks for children and adolescents.<sup>19</sup>

Neurologist Servane Mouton further explained the mechanisms and effects in a 2025 interview with Amnesty International:

---

16 Interview, Nice, 29 July 2025

17 Sophie\* and Emma\* are pseudonyms used at the participants’ request.

18 Interview, Paris, 15 July 2025

19 US Surgeon General, Social Media and Youth Mental Health, May 2023, <https://www.hhs.gov/surgeongeneral/reports-and-publications/youth-mental-health/social-media/index.html>; Qinghua He, Ofir Turel and others, “Brain anatomy alterations associated with Social Networking Site (SNS) addiction”, March 2017, Scientific reports, 7, 45064. <https://www.nature.com/articles/srep45064>; Yvonne Kelly, Afshin Zilanawala and others, “Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study”, December 2018, EClinicalMedicine, 6, 59–68. [https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370\(18\)30060-9/fulltext](https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(18)30060-9/fulltext)

*“Behaviours on social networks are similar to addiction, with this difficulty in getting off, the encroachment on other activities, the obsession with thoughts in a loop around having access to them. Missing [them] when you’re not there. The economic model of social networks is based on stimulating the short-term reward system so that we go there as often as possible, for as long as possible, etc. It is a bit like chasing after dopamine shots that will immediately produce the sensation of pleasure and make us want to go back.”*

As Amnesty International’s 2023 research and these new testimonies reveal, these design choices pose additional severe risks for children and young people whose interest in mental health predisposes them to be dragged into the kinds of harmful “rabbit holes” described by Maëlle, Maëllys and other research participants.

Psychologists interviewed in 2023 and prior independent research also pointed to negative impacts of design features encouraging social comparison between children and young people and their peers, such as the “heart” or “like” feature, on the mental health of adolescents.<sup>20</sup>

For instance, Maëlle and Édouard reflected on the risks of the “like” or “heart” button in relation to videos discussing self-harm and suicidal ideation:

*“When I liked the video, it was to encourage, to give strength, to let the person know that they can be well...But I later realized that it might actually mean to the person, “I’m being encouraged to self-harm.” (Maëlle)*

*“The online support can also be interpreted as ‘go ahead and kill yourself.’” (Édouard)*

Another common thread across the young people’s testimonies is the negative reinforcement they experienced from the online community they became immersed in after they were dragged into the “rabbit hole”. Similarly to Édouard’s observations on his thoughts of wanting to match the levels of self-harm displayed by others in their TikTok videos, Maëlle shared how commentators nudged others to go further in their self-harm or even encouraged others to end their lives.

*“There were comments like: If you make a suicide attempt but you don’t end up in intensive care or in the hospital, it doesn’t count. If you are not under-weight, you cannot be anorexic. If you don’t have stitches, it doesn’t count, things like that make you feel guilty. Cutting yourself, deep or not, still leaves scars, and the gesture is a call for help either way.”*

She added:

*“I rarely went to the comment section. But when I went there, I found it even darker than the video itself. Because there, it was: “Be brave my beauty, you’ll get there tomorrow, it’s the right day. You’re going to end your life. [...]” And it was snowballing.”*

These observations offer important psychological insights into online communities in relation to suicidal ideation; from a technological perspective they also highlight clear failures of TikTok’s content moderation process, which should ensure that content which promotes or encourages self-harm and suicide is swiftly removed. They also raise further questions about potential design risks in relation to the engagement-based ranking of comments and “algorithmic audiencing”, the automated targeting of specific user groups with particular types of content, potentially encouraging the development of “filter bubbles” or “echo chambers”.<sup>21</sup>

Whilst Amnesty International’s findings underline the importance of tackling not merely failures of content moderation but the “rabbit hole” effect (resulting from the algorithmic amplification of content

---

20 Amnesty International, *Driven into the Darkness: How TikTok’s ‘For You’ Feed Encourages Self-Harm and Suicidal Ideation*, 7 November 2023 (previously cited).

21 Owen Doyle, “Algorithms and the “Anti-Preference”: A Quantitative Investigation of “Reaching the Wrong Audience” on TikTok”, 2024, *Berkeley Undergraduate Journal*, 38(1), <https://escholarship.org/uc/item/2vq7291w>; Kai Riemer and Sandra Peter, “Algorithmic Audiencing: Why We Need to Rethink Free Speech on Social Media,” December 2021, *Journal of Information Technology* 36, no. 4: 409–26, <https://journals.sagepub.com/doi/10.1177/02683962211013358>

based on user profiling) to prevent both the privacy abuses that facilitate the development of such dangerous clusters of content and their harmful impact, the real-life examples included in these testimonies do also point to failures of TikTok's content moderation systems in preventing the spread of evidently violative content promoting self-harm and suicide. Laetitia, Maëllys's mother, shared that when she learnt about her daughter's online life, she spent months reporting harmful mental health content to TikTok but eventually gave up after she found that these videos were often reviewed and deemed permissible. She told Amnesty International:

*"For 6 months, I reported all the time. As soon as I had 5 minutes at work, my lunch breaks, my weekends, I spent hours and hours reporting videos. For nothing. In 90% of cases, I was told that it does not violate the rules of the community. After a while, I stopped because I was too unwell. I only lived to report TikTok videos."*

Section 8 will delve deeper into TikTok's failure to identify, mitigate and prevent risks and harms linked to these design choices and business decisions.

# 5. TECHNICAL RESEARCH EVIDENCE

## 5.1 MANUAL TESTS

In July and August 2025, Amnesty International set up three teen accounts, two female, one male, registered as 13-year-olds based in France to manually examine the algorithmic amplification of content in TikTok's 'For You' feed. The researcher rewatched recommended content if it was related to feelings of sadness or mental health themes and skipped videos related to other topics. Amnesty International does not presume that this is fully representative of real user behaviour. However, as TikTok's recommender system is known to use watch time as a key signal for inferring users' preferences, these experiments offer crucial insights into TikTok's recommendation systems, specifically the impact of watch time on recommended content related to feelings of sadness or mental health themes and the effectiveness of measures taken to mitigate risks to children's rights under the EU's Digital Services Act, which the company has put in place with the stated aim of "prevent[ing] the display of concentrated content for topics such as dietary discussions, weight loss, sadness, grief, loneliness, and hopelessness".<sup>22</sup>

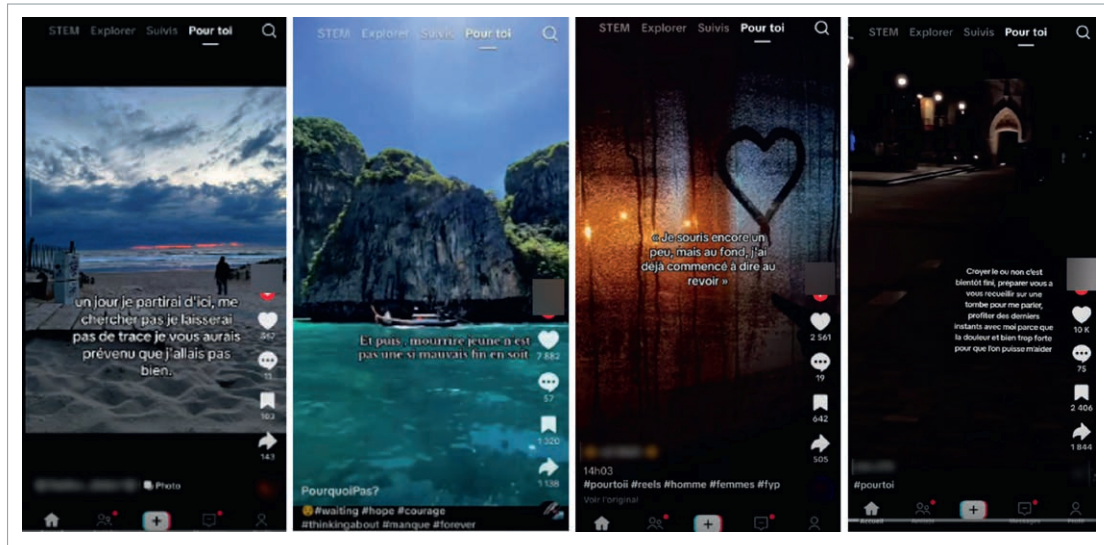
All three accounts encountered videos related to feelings of sadness or disillusionment within the first five minutes of scrolling through the 'For You' feed, before they had begun to signal any preferences. Watching these videos then led to a rapid increase in sadness and mental health-related content. Within fifteen to twenty minutes of starting the experiment, all three feeds were almost exclusively filled with videos related to mental health, with up to half of the videos at that point containing sad and depressive content. Two of the three accounts were shown videos expressing suicidal thoughts within 45 minutes. After less than three hours on the platform (based on hour-long sessions on different days), all accounts were shown content expressing suicidal thoughts. Examples include image captions such as, "I'm going through a period where I'm wondering whether I should really continue to exist or not". In some cases, videos referenced popular songs with lyrics that contain romanticized allusions to despair, to the idea of isolating oneself or going far away which are repurposed by users to evoke suicide ("One day I'll leave, don't look for me I won't leave any trace, I told you that I wasn't feeling well.").

Several videos mentioned users' plans to end their lives that day. Both female accounts received content recommendations stating or alluding to specific suicide methods within the third or fourth hour on the platform. One of the female accounts was shown two videos related to a known suicide challenge, which French media first reported on in 2022 (see section 4.1). The suicidal content often relies on romanticizing language and "algospeak" (using emojis or replacing letters with numbers to avoid detection by automated means of content moderation). Comparing the findings from these recent

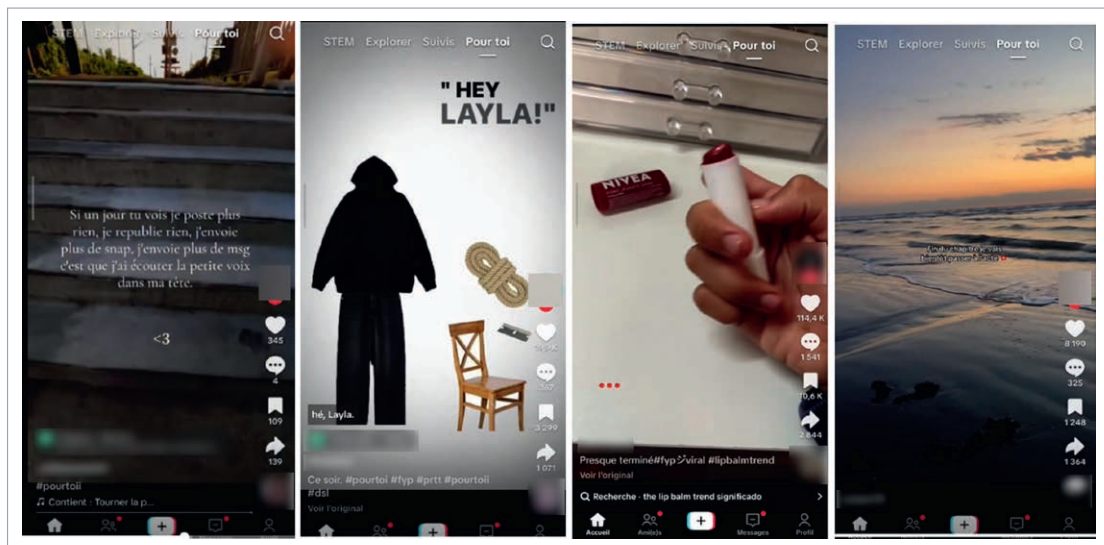
---

22 TikTok, "DSA Risk Assessment Report 2024, 28 August 2024, <https://www.tiktok.com/transparency/en/dsa-transparency/>

manual experiments with the young survivors' experiences reveals that at least one commonly used emoji used to disguise the word "suicide" seems to have been in use for multiple years and yet does not seem to trigger adequate measures to prevent the spread of harmful content.



Translations: 1: "One day I'll leave, don't look for me, I won't leave a trace, I would have told you that I wasn't doing well." 2: "And then dying young is not that bad an end in itself" 3: I'm still smiling a little, but deep down, I've already begun to say goodbye" 4: "[...] get ready to gather at a grave to talk to me, enjoy the last moments with me because the pain is too strong for anyone to help me"



5: "If one day you see that I don't post anything, I don't repost anything, I don't send anymore snaps, I don't send messages, that's because I've listened to the little voice in my head" 6: [caption below] "tonight" 7: [caption below] almost finished 8: "End of the chapter, I'll soon take action [Swiss flag as a popular suicide signifier]"

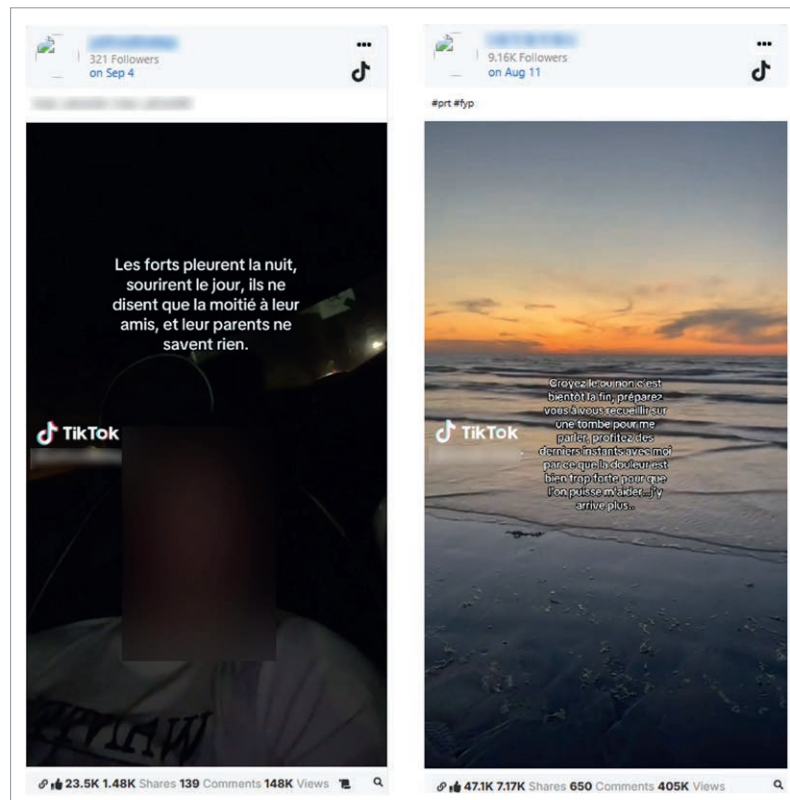
Videos recommended to a manual test account registered as a 13-year-old girl within the third and fourth hour of signalling interest in mental health-related content in TikTok's 'For You' feed. The account was drawn into a rabbit hole of melancholic and depressive content within less than hour and then saw a rapid increase in suicidal ideation content in the third and fourth hour. Contrary to TikTok's claims that it breaks up potentially harmful "clusters" of content, Amnesty International found no evidence of efforts to disrupt the steady stream of harmful content.

Crucially, these experiments revealed that TikTok’s stated DSA risk mitigation measures to break up “concentrated content” and proactively enforce its community guidelines are ineffective (see also section 8). Within mere hours, these accounts were drawn into rabbit holes of mental health-related content of potentially harmful nature with increasing numbers of videos romanticizing and normalizing suicide over time. None of the videos appeared to be produced by mental health professionals or support organizations.

## 5.2 AUTOMATED TESTS

After conducting the manual experiments, researchers then also ran automated tests to study TikTok’s algorithmic recommender system behind the ‘For You’ feed. The automated experiment builds upon the manual experiments, by using real user watch history (the result of the researcher’s methodical engagement with the ‘For You’ feed in the manual experiment) to build automated accounts and assess algorithmic amplification of harmful content at greater scale.

Automated experiments confirmed a potentially harmful personalization effect in relation to content categorized as sad or depressive. Just as in the manual experiments, each account was registered as a “fake” user aged 13. The collection methodology is described in chapter 2; the following analysis relates to the findings based on the manual categorization of the video recommendations collected from automated accounts whose preferences were expressed through watch histories derived from partitions of one of the manual account experiments.



The video on the **left** (“The strong ones cry at night, smile during the day, they only tell their friends half of it and their parents know nothing”) was categorized as **sad/depressive**.

The video on the **right** (“Believe it or not, soon it’s going to be the end, get ready to gather at a grave to talk to me, enjoy the last moments with me because the pain is too strong for anyone to help me”...I can’t do it any more”) was categorized as **self-harm/suicidal ideation content**. The video had already reached over 400,000 users when it was recommended to our research account.

Examples of categorized content:

The following scatter plots (graph 1 and 2) show the personalization rates in response to different ratios of sad/depressive and self-harm/suicidal ideation content contained in five partitions of the replicated

feed.<sup>23</sup> Researchers fetched 201 videos for two accounts per partition; for purposes of analysis, we combine the number of video recommendations for the two accounts for a total sample number of 402.

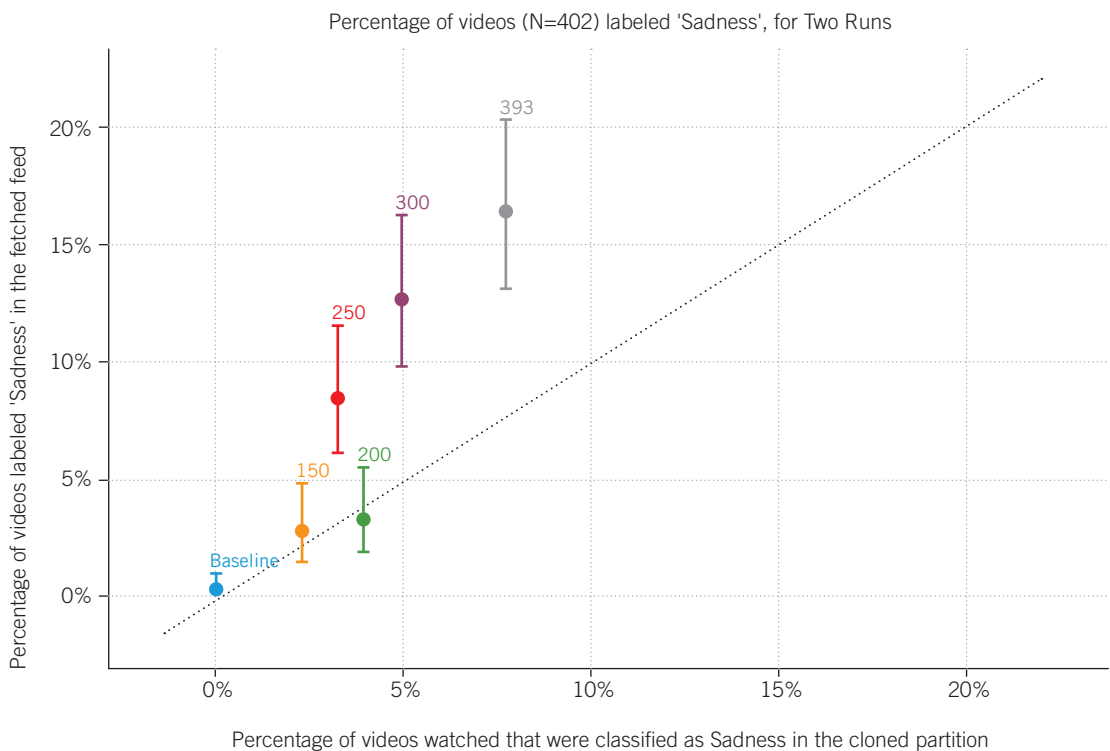
The x-axis represents the percentage of potentially harmful videos (sad/depressive in graph 1, self-harm/suicidal ideation in graph 2) in the *input* (the cloned feed). Videos were included if they were both categorized as on-topic (either related to sadness or self-harm) and watched by the source manual account, excluding videos which were either relevant but erroneously skipped in the manual experiment or erroneously watched but later categorized as on-topic. A video is defined as ‘watched’ if the watch time of the video is greater than 1.75 times the length of the video.

The y-axis represents the percentage of on-topic videos (sad/depressive in graph 1, self-harm/suicidal ideation in graph 2) in the output (the 402 fetched recommendations in response to the cloned feed).

In the graphs, if a dot falls above the  $y=x$  line (the dotted line down the diagonal), it suggests *over-emphasis*, where the resulting output had a higher percentage of harmful videos than the corresponding input. Therefore, graph 1 shows a clear over-emphasis on sad or depressive content in the recommendations surfaced to accounts as part of the ‘For You’ feed’s personalization process. If a dot falls below the  $y=x$  line, it suggests an *under-emphasis*, where the resulting output was lower in percentage than the corresponding input. Note that any upwards trend – even with the points falling below the dotted line – indicates that the more a user watches problematic content, the more such content TikTok recommends.

The bars either side of each dot/data point represent confidence intervals,<sup>24</sup> which are a statistical representation of the uncertainty and reliability of our results.

Graph 1



23 To confirm that TikTok correctly registered the automated accounts’ watch behaviours, the researchers exported each account’s watch history. They found that 10-15% of the videos the automated accounts watched, as per the replicated watch histories, were not present in the history exports. While it is not clear what caused this discrepancy, researchers accounted for this by only counting a video as watched if it appeared in the data download of the accounts. Researchers found similar numbers of watched on-topic videos in the data downloads between the two accounts per partition. Overall, researchers found that instead of the intended 150, 200, 250, 300, and 393 videos, the data download included 132, 181, 217, 266, and 355 videos.

24 estimated using a statistical approach known as the Agresti-Coull method

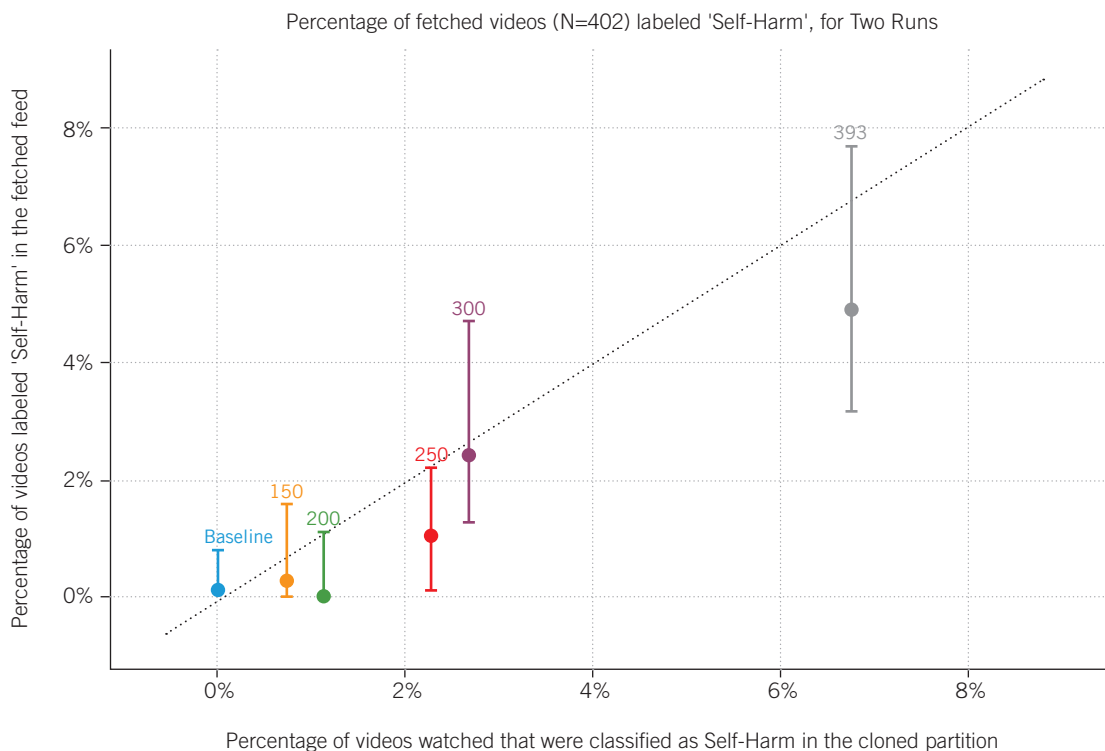
As the number of sad/depressive videos included in the partitions of the cloned feed increased, researchers observed an increasing number of video recommendations related to sadness and feelings of despair and inadequacy in response. The findings indicate that when engaging with sad or depressive content at rates as low as 4.9%, the rate of sad or depressive content in the subsequent video recommendations based on this set of signals jumps to almost 13%, more than doubling the original ratio of potentially harmful content. At 7.6% sad/depressive videos in the original watch history, the rate of sad or depressive content in the subsequent video recommendations jumps to 16%.

This indicates that TikTok’s recommender system continues to respond with, and multiply, the ratio of sad and depressive content in the ‘For You’ feed when a child user expresses interest in such content.

The calculated ratios are based on a conservative approach to disagreements between reviewers in the categorization of recommended videos. In cases where reviewers disagreed on whether content should be categorized as sad/depressive, the content was not included in the “sad/depressive” category.

Note also that the behaviour of these automated accounts is significantly pared back in comparison with a real user who would likely also signal content preferences by saving, reposting, liking and commenting on videos as well as following accounts. For ethical and practical reasons, the methodology was restricted to the signalling of content preferences through the cloned watch history. We must also note that these findings are still drawn from a sample size of 12 automated accounts, which whilst improving the generalizability of the findings, is still a relatively small sample size and would be complemented by further comparable research.

Graph 2



The automated experiments show less personalization in response to watching videos related to self-harm and suicidal ideation than in response to sadness content. However, engaging with this content nonetheless led to further recommendations of self-harm/suicidal ideation content in the feeds.

Compared to benign (not problematic) content, the self-harm videos have statistically fewer likes, views, and comments, and sadness videos have statistically fewer views and comments, with likes not being statistically significant. This may indicate that, due to personalization, potentially harmful content does not need to be as popular as benign content to be recommended to our test users.

The highest observed ratio, in the automated experiments, of 21.4% potentially harmful content (depressive and self-harm content) is lower than the “rabbit hole” effect observed in the manually run accounts. This is at least partly explained by a lower ratio of potentially harmful content in the input, as some videos from the original manual feed were no longer available (possibly because they were removed by the user or TikTok). Further, it is possible that a real user’s phone sends more information to TikTok than the automated setup does, resulting in a different set of recommendations.

***Nevertheless, the core principle holds – the more a user signals interest in potentially harmful content, the more such recommendations they receive.***

These findings are deeply concerning as even lower rates of continuous exposure to harmful content of this kind and extremity have been shown to contribute to real-world harm. Individual susceptibilities to depressive and self-harm content may differ, but real-world evidence from young people’s testimonies in this briefing and the widely reported case of British teenager Molly Russell give insights into the manipulative effect of continuous exposure to depressive and self-harm related content for adolescents with pre-existing mental health issues. For context, a UK Coroner found social media platforms to have played a contributing role in 14-year-old Molly Russell’s suicide; the investigation found that “roughly 13 per cent” of the 16,000 Instagram posts she viewed in the six months prior to her death “related to suicide, depression or self-harm”.<sup>25</sup>

---

25 Politico, “Digital bridge: Platforms on the hook — Transatlantic AI rulebook — Let’s talk data transfers”, 6 October 2022, <https://www.politico.eu/newsletter/digital-bridge/platforms-on-the-hook-transatlantic-ai-rulebook-lets-talk-data-transfers/>; a Meta spokesperson who spoke at the inquest said that the content Molly saw was “nuanced and complicated”, and that it was important to allow users experiencing suicidal thoughts to express themselves online. BBC, “Molly Russell: Instagram posts seen by teen were safe, Meta says”. 26 September 2022, <https://www.bbc.co.uk/news/uk-england-london-63034300>; for a more extensive response from Meta to the inquiry and key concerns raised in the process, see Meta, “Response from Meta”, 6 December 2022, <https://www.judiciary.uk/wp-content/uploads/2022/10/2022-0315-Response-from-META.pdf>

# 6. HUMAN RIGHTS FRAMEWORK

Amnesty International's 2023 report *Driven into the Darkness* documented how TikTok's business model, addictive design and failure to address and mitigate the risk of harmful "rabbit holes" undermine human rights, including the right to privacy, freedom of thought and health. Readers may want to refer to the report for a more detailed discussion of the international human rights instruments and standards relevant to this discussion. To briefly summarize the key obligations on States to respect, promote and fulfil children's and human rights as well as the responsibilities of businesses to respect human rights:

The **right to privacy** provides that no one should be subject to "arbitrary or unlawful interference" with their privacy, family, home or correspondence, and that this should be protected by law.<sup>26</sup> It comprises three interrelated concepts: freedom from intrusion into our private lives, the right to control information about ourselves, and the right to a space in which we can freely express our identities. It is Amnesty International's assessment that the user profiling inherent to TikTok's business model undermines each of these three elements. That is because social media platforms, including TikTok, are instituting forms of mass corporate surveillance to personalize content recommendations and ads that are inherently unnecessary, disproportionate and can never be a permissible interference with the right to privacy.<sup>27</sup> Children are exposed to enormous social pressure to participate in these spaces and willingly share information but are also nudged to publish personal information through posts and comments as well as to divulge personal preferences and interests through 'likes' or by sharing other people's content. Social media companies gain much wider insights still through data on what children and young people watch, how often and for how long, their location data, data about the times they log on and the devices they use, to name but a few. Children's access to these spaces is thus made conditional on deeply invasive profiling to which they cannot meaningfully consent at such a young age, and which cannot be said to be in their best interests. Yet the European Union's legal frameworks, including the Digital Services Act (DSA) and General Data Protection Regulation (GDPR), have thus far failed to rein in this rights abuse, through a lack of adequate safeguards contained in the law and through inadequate enforcement at EU and member state level, respectively.<sup>28</sup>

The **right to freedom of thought**, protected by Article 18(1) of the International Covenant on Civil and Political Rights (ICCPR), is an absolute right, meaning no one may interfere with our private thoughts and beliefs under any circumstances. The Council of Europe (CoE) has warned that "[f]ine grained,

---

26 Universal Declaration of Human Rights (UDHR), Article 12 and International Covenant on Civil and Political Rights (ICCPR), Article 17

27 Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web* (Index: POL 40/7349/2023), 7 November 2023 (previously cited); the EU's Digital Services Act bans targeted advertisements towards minors. The vast majority of TikTok's young user base around the world can however still be shown targeted ads, leaving the basis of TikTok's business model largely unaffected.

28 EDRI and Amnesty International, "Civil society call and recommendations for concrete solutions to GDPR enforcement shortcomings", 16 March 2022, <https://www.amnesty.eu/wp-content/uploads/2022/03/EDRI-recommendations-for-better-GDPR-enforcement.pdf>

sub-conscious and personalized levels of algorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions”.<sup>29</sup> As our findings and the testimonies of affected young people underline, TikTok’s surveillance-based recommender system and its addictive design risk undermining children and young people’s ability to think freely and expose adolescents with pre-existing vulnerabilities to “rabbit holes” that can have severe knock-on impacts on their right to health.

States have an obligation to take steps to progressively achieve the full realization of the right of everyone to the highest attainable standard of **physical and mental health**.<sup>30</sup> Digital technologies have undoubtedly played a positive role in reducing access barriers to health care information and services. But there is growing recognition that in the absence of effective state regulation, social media platforms have also exposed children and young people to systemic risks to human rights.<sup>31</sup> Focusing on risks to children and young people’s health, the US Surgeon General warned in May 2023:

*“We must acknowledge the growing body of research about potential harms, increase our collective understanding of the risks associated with social media use, and urgently take action to create safe and healthy digital environments that minimize harm and safeguard children’s and adolescents’ mental health and well-being during critical stages of development.”<sup>32</sup>*

The UN Special Rapporteur on the right to health, framing the social media environment as an underlying determinant of health, also recommended that:

*“States should fulfil their obligations through the adoption of holistic strategies aimed at enhancing adolescents’ capacities to protect themselves from online harm, strengthening legislation and law enforcement mechanisms to tackle online abuse, including cross-border abuse, combating impunity and training parents and professionals who work with children.”<sup>33</sup>*

Importantly, pre-empting current political debates about bans on the use of social media by adolescents, the Special Rapporteur already noted then, that “it is neither appropriate nor possible to seek to restrict adolescents’ access to the digital environment.”<sup>34</sup>

**Companies have a responsibility to respect human rights** wherever they operate in the world and across all their business activities. This is a widely recognized standard of expected conduct as set out in international business and human rights standards including the UN Guiding Principles on Business and Human Rights (UN Guiding Principles) and the OECD Guidelines for Multinational Enterprises (OECD Guidelines).<sup>35</sup> The UN Guiding Principles establish that, to meet these corporate responsibilities, companies should have in place an ongoing and proactive human rights due diligence process to identify, prevent, mitigate and account for risks to human rights and how they address their

---

29 CoE, Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes, 13 February 2019, Decl(13/02/2019)1, <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168092dd4b>

30 UDHR, Article 25 ICESCR, Article 12, CRC, Article 24. See also regional human rights treaties including the African Charter on Human and People’s Rights and the European Social Charter

31 See for example OHCHR, Digital innovation, technologies and the right to health (Report of the Special Rapporteur on the Right to Health), April 2023, [ohchr.org/en/documents/thematic-reports/ahrc5365-digital-innovation-technologies-and-right-health](https://www.ohchr.org/en/documents/thematic-reports/ahrc5365-digital-innovation-technologies-and-right-health); Committee on the Rights of the Child, General Comment 25, 2 March 2021, UN Doc. CRC/C/GC/25

32 US Surgeon General, Social Media and Youth Mental Health, May 2023, <https://www.hhs.gov/surgeongeneral/reports-and-publications/youth-mental-health/social-media/index.html>

33 OHCHR, Report on the right to health of adolescents (Report of the Special Rapporteur on the Right to Health), 4 April 2016, <https://www.ohchr.org/en/documents/thematic-reports/ahrc3232-report-right-health-adolescents>, para. 45

34 OHCHR, Report on the right to health of adolescents (Report of the Special Rapporteur on the Right to Health), 4 April 2016 (previously cited), para. 45

35 Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework, 2011, endorsed by the UN Human Rights Council (UNHRC), UNHRC Resolution 17/4: Human rights and Transnational Corporations and other Business Enterprises, adopted on 16 June 2011, UN Doc. A/HRC/RES/17/4; and OECD Guidelines for Multinational Enterprises, 2011, <https://mneguidelines.oecd.org/mneguidelines>. In accordance with the UN Guiding Principles, corporate responsibility to respect human rights is independent of a State’s human rights obligations and exists over and above compliance with national laws and regulations protecting human rights. See UN Guiding Principles, Principle 11 and Commentary.

potential and actual impacts on human rights. If, in this process, a company finds that it is causing or contributing to abuses, it must cease or prevent the adverse human rights impacts, as well as provide remedy to those affected. Specific to children, the Committee on the Rights of the Child has stressed the duty of States to “require businesses to undertake child rights due diligence” and has set measures that should be taken by States to prevent businesses from causing or contributing to abuses of children’s rights and to “investigate, adjudicate and redress violations of children’s rights caused or contributed to by a business enterprise”.<sup>36</sup>

---

36 Committee on the Rights of the Child, General Comment 16, 17 April 2013, UN Doc. CRC/C/GC/16, paras. 62 and 28.

# 7. EU AND NATIONAL LEGAL FRAMEWORKS

## 7.1 EUROPEAN UNION REGULATION: THE DIGITAL SERVICES ACT

In July 2022, the EU passed a new landmark regulation, the Digital Services Act (DSA), defining the responsibilities and obligations of social media companies and other online platforms operating in the region.<sup>37</sup> Created to harmonize rules across the 27 EU member states for a safe online environment where fundamental rights, including children’s rights, are effectively protected, the DSA’s provisions are implemented and enforced by member states within their own legal systems through designated national authorities (Digital Services Coordinators), whilst the European Commission is responsible for enforcing compliance with the strictest rules applying to Very Large Online Platforms. The DSA tackles the spread of illegal content and disinformation as well as systemic risks arising from platforms’ “design, including algorithmic systems, functioning and use made of their services”.<sup>38</sup> As a legally binding framework, the DSA provides penalties for non-compliance, including fines of up to 6% of a company’s global annual turnover.

The DSA requires Very Large Online Platforms (designated platforms with more than 45 million users in the EU), including TikTok, to put in place “reasonable, proportionate and effective mitigation measures”, tailored to address systemic risks identified through a yearly risk assessment and with due consideration to the impacts of such measures on fundamental rights.<sup>39</sup> These measures may include adapting the design, features or functioning of their services, adapting content moderation processes, and testing and adapting a platform’s algorithmic recommender systems. The DSA also includes a ban on targeted advertising towards children and a requirement for large social media platforms to offer all users an optional feed which is not based on user profiling.<sup>40</sup> Very Large Online Platforms (VLOPs) are subject to yearly independent audits to assess compliance with the companies’ due diligence obligations.

The obligations contained in the DSA have applied to designated VLOPs since August 2023, and to all digital platforms since February 2024.<sup>41</sup>

---

37 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065>; Amnesty International, *What the EU’s Digital Services Act Means for Human Rights* (Index: POL 30/5830/2022), 7 July 2022, <https://www.amnesty.org/en/documents/pol30/5830/2022/en/>

38 Digital Services Act, Article 34

39 Digital Services Act, Article 35

40 Digital Services Act, Article 28

41 European Commission, “The Digital Services Act”, [https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act\\_en](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en) (accessed 8 August 2025)

On 19 February 2024, the European Commission announced that it had opened formal proceedings “to assess whether TikTok may have breached the Digital Services Act (DSA) in areas linked to the protection of minors, advertising transparency, data access for researchers, as well as the risk management of addictive design and harmful content”.<sup>42</sup> Echoing Amnesty International’s findings from November 2023, the Commission stated that it was looking into potential breaches of DSA obligations related to “actual or foreseeable **negative effects** stemming from the design of TikTok’s system, including algorithmic systems, **that may stimulate behavioural addictions and/ or create so-called ‘rabbit hole effects’**” (original emphasis). The investigation is ongoing at the time of writing.

A second probe, launched in April 2024, into TikTok Lite’s Rewards programme led to TikTok’s commitment to permanently withdraw the programme from the EU.<sup>43</sup> TikTok Lite is a data-light version of TikTok already widely used in Brazil, India and Indonesia that was quietly launched in France and Spain in April 2024. The “tasks and rewards” programme offers users financial incentives to increase their engagement with the app. The European Commission had questioned TikTok’s compliance with its risk assessment and mitigation obligations in relation to potential negative effects on (children’s) mental health “as a result of the new feature stimulating addictive behavior”.<sup>44</sup>

In July 2025, the European Commission published Guidelines for the Protection of Minors, providing recommendations to online platforms on how to mitigate risks to children’s rights in compliance with Article 28 the DSA, which requires providers of online platforms to “put in place appropriate and proportionate measures to ensure a high level of privacy, safety, and security of minors, on their service.”<sup>45</sup> Most relevant to the issue of addictive design, the Guidelines emphasize platforms’ responsibilities to ensure that

*“minors are not exposed to persuasive design features that are aimed predominantly at engagement and that may lead to extensive use or overuse of the platform or problematic or compulsive behavioural habits” including, among others, “the possibility to scroll indefinitely, the superfluous requirement to perform a specific action to receive updated information on an application, automatic triggering of video content, notifications artificially timed to regain minors’ attention.”<sup>46</sup>*

This recommendation evidently clashes with some of the core design features of TikTok’s “For You” feed.

With regards to recommender systems, the Guidelines ask platforms to

*“prioritise ‘explicit user-provided signals’ to determine the content displayed and recommended to minors” and to “implement measures to prevent minors’ exposure to content recommendations that could pose a risk to their safety and security, particularly when encountered repeatedly, such as content promoting unrealistic beauty standards or dieting, content that glorifies or trivialises mental health issues, such as anxiety or depression, discriminatory content, radicalisation content and distressing content depicting violence or encouraging minors to engage in dangerous activities.”<sup>47</sup>*

---

42 European Commission, “Commission opens formal proceedings against TikTok under the Digital Services Act”, 19 February 2024, [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_24\\_926](https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926)

43 European Commission, “TikTok commits to permanently withdraw TikTok Lite Rewards programme from the EU to comply with the Digital Services Act”, 5 August 2024, [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_24\\_4161](https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4161)

44 European Commission, “Commission opens proceedings against TikTok under the DSA regarding the launch of TikTok Lite in France and Spain, and communicates its intention to suspend the reward programme in the EU”, 22 April 2024, [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_24\\_2227](https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2227)

45 European Commission, Guidelines on the protection of minors, 14 July 2025, <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-protection-minors>; DSA, Article 28

46 European Commission, Guidelines on the protection of minors, 14 July 2025, [https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C\\_202505519](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C_202505519), section 6.4

47 European Commission, Guidelines on the protection of minors, 14 July 2025, section 6.5.1

The Guidelines thus recognize the risks associated with engagement-based recommender systems and aim to curb their harmful effects but fall short of banning these privacy-intrusive systems entirely. Whilst the recommendations are non-binding, they nonetheless offer an indication of how the Commission defines safety-by-design with regards to children’s rights and needs and what DSA enforcement ultimately aims towards in this area.

Concerningly amidst growing political support for new restrictions to adolescents’ access to social media from EU member states including the French President, the European Commission’s Guidelines also appear to open the possibility for member states to raise the minimum age for access to social networks, though the Guidelines include the caveat that this must be done “in compliance with Union law” (which has previously been invoked to stop such member state action, see section 7.2). This provision appears to contradict the stated aim of the DSA to harmonize rules across the European Union and risks unduly restricting adolescents’ right to freedom of expression and peaceful assembly online.<sup>48</sup>

## 7.2 THE ONLINE PROTECTION OF MINORS IN FRANCE

### 7.2.1 IMPLEMENTATION OF THE EUROPEAN DIGITAL SERVICES ACT IN FRANCE

The law to secure and regulate the digital space (SREN)<sup>49</sup>, promulgated on 21 May 2024, adapts the European regulation on digital services (DSA) into French law. It entrusts the Audiovisual and Digital Communication Regulatory Authority (Arcom) with the role of Digital Services Coordinator (DSC). Arcom carries out these tasks in conjunction with the Commission nationale de l’informatique et des libertés (CNIL) with regard to the provisions protecting privacy, and the Direction générale de la concurrence, de la consommation et de la répression des fraudes (DGCCRF), in relation to the provisions specific to marketplaces.<sup>50</sup> Arcom also participates with its European counterparts in the supervision of the obligations of actors regulated by the European Commission. The Commission has exclusive powers to monitor and enforce the provisions specific to Very Large Online Platforms.

The DSA also provides for the implementation of ‘trusted flaggers’ in each EU Member State to report suspected illegal content to online platforms.<sup>51</sup> Their reports are treated as a priority by all online platforms subject to the DSA. Trusted flaggers, designated by Arcom, are “organizations recognized for their expertise in the detection, identification and notification of illegal content, in one or more specific areas”.<sup>52</sup> Among them is the association e-Enfance, which fights against harassment and digital violence suffered by young people.<sup>53</sup>

Almost a year before the SREN law came into force, the Law of 7 July 2023 aimed at “establishing a digital majority and combating online hate”<sup>54</sup> established a digital majority at the age of 15 to register

---

48 For further analysis by Amnesty International of the human rights risks of “teen social media bans” see: Amnesty International, “Australia: Authorities must effectively regulate social media instead of banning children and young people”, 21 November 2024, <https://www.amnesty.org/en/latest/news/2024/11/australia-must-effectively-regulate-social-media-than-ban-children/>

49 LOI n° 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l’espace numérique, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000049563368/>

50 Convention de coopération pour la mise en œuvre du règlement sur les services numériques entre l’Arcom, la DGCCRF et la CNIL : [https://www.arcom.fr/sites/default/files/2024-06/Arcom%20-%20Convention%20coop%C3%A9ration\\_ARCOM\\_CNIL\\_DGCCRF.pdf](https://www.arcom.fr/sites/default/files/2024-06/Arcom%20-%20Convention%20coop%C3%A9ration_ARCOM_CNIL_DGCCRF.pdf)

51 Digital Services Act, Article 22

52 Arcom, “Trusted flaggers”, <https://www.arcom.fr/signaleurs-de-confiance>

53 <https://e-enfance.org/>. See the regularly updated list of trusted flaggers here: <https://www.arcom.fr/signaleurs-de-confiance>

54 LOI n° 2023-566 du 7 juillet 2023 visant à instaurer une majorité numérique et à lutter contre la haine en ligne, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000047799533>

on social networks. However, the European Commission found that the provisions of the law did not conform with European law, and thus the decree (the executive act) specifying the application and enforcement of the law was not published and the measure was not applied.

## 7.2.2 INITIATIVES BY THE FRENCH AUTHORITIES

In January 2024, President Emmanuel Macron launched “a Commission made up of experts from ‘civil society’ to assess the issues related to children’s exposure to screens and make recommendations”.<sup>55</sup> Among other things, the Commission focused on the problems of addiction to screens and in particular to social networks, their impact on the mental health of children and adolescents, as well as the insufficient regulation of content to which minors may be exposed. In its report published in April 2024, the Commission recommends in particular to “tackle, in order to prohibit, the addictive and trapping design features of certain digital services in order to give young people choice”.<sup>56</sup>

In March 2025, a parliamentary commission of inquiry into the psychological effects of TikTok on minors was established.<sup>57</sup> In particular, it examined the attention-grabbing design features used by TikTok as well as the risks associated with exposing young users to dangerous content. Several TikTok representatives were interviewed and asked about the platform’s responsibility for “rabbit holes” of potentially harmful content, including videos that idealize and encourage depressive thoughts, self-harm and suicide. According to TikTok representatives, the company makes every effort to protect its young users and guarantee them a “safe” online space: “We are committed to acting responsibly [...] by supporting useful and positive content and ensuring the safety and well-being of our community”.<sup>58</sup> The commission presented its conclusions in September 2025. In his foreword, the chairman of the commission, Arthur Delaporte stated: “The verdict is final: this platform knowingly exposes our children, our young people, to toxic, dangerous, addictive content.”<sup>59</sup>

In parallel with the work of the parliamentary commission of inquiry, the Minister Delegate for Artificial Intelligence and Digital Affairs has strongly mobilized at the European level to ban access to social networks for minors under the age of 15. The French President also expressed his support, reaffirming the will of France to increase the digital age of majority to 15 years old for social networks as provided for by the law of July 2023.

Finally, mental health was designated a Great National Cause in 2025.<sup>60</sup> As such, four priority objectives have been set: destigmatization, the development of prevention and early identification, improving access to care and support for those concerned.

---

55 “Enfant et écrans, A la recherche du temps perdu”, April 2024, <https://www.elysee.fr/admin/upload/default/0001/16/fbec6abe9d9cc1bff3043d87b9f7951e62779b09.pdf>

56 “Enfant et écrans, A la recherche du temps perdu”, April 2024 (previously cited).

57 Commission d’enquête sur les effets psychologiques de TikTok sur les mineurs, <https://www.assemblee-nationale.fr/dyn/17/organes/autres-commissions/commissions-enquete/tiktok>

58 Marlène Masure, Head of Content Europe, Middle East and Africa States, Executive Director of Business Development and Marketing. Hearing of 12/06/25 before the Parliamentary Committee of Inquiry into the psychological effects of TikTok on minors, <https://www.assemblee-nationale.fr/dyn/actualites-accueil-hub/effets-psychologiques-de-tiktok-sur-les-mineurs-auditions-de-representants-de-tiktok-et-tiktok-france>

59 Le Monde, “TikTok est « un des pires réseaux sociaux à l’assaut de notre jeunesse » : les mots très durs de la commission d’enquête parlementaire”, 11 September 2025, [https://www.lemonde.fr/pixels/article/2025/09/11/commission-d-enquete-tiktok-le-rapport-et-rapport-etrille-une-plateforme-hors-de-contrôle-a-l-assaut-de-la-jeunesse-et-formule-43-recommandations\\_6640383\\_4408996.html](https://www.lemonde.fr/pixels/article/2025/09/11/commission-d-enquete-tiktok-le-rapport-et-rapport-etrille-une-plateforme-hors-de-contrôle-a-l-assaut-de-la-jeunesse-et-formule-43-recommandations_6640383_4408996.html)

60 French Ministry of Health, “La santé mentale, Grande Cause nationale 2025”, 24 March 2025, <https://solidarites.gouv.fr/la-sante-mentale-grande-cause-nationale-2025>

# 8. TIKTOK'S FAILURE TO MAKE THE PLATFORM SAFE FOR CHILDREN AND ADOLESCENTS AS A BREACH OF THE DIGITAL SERVICES ACT AND INTERNATIONAL HUMAN RIGHTS STANDARDS

Amnesty International's 2023 report *Driven into the Darkness* showed that TikTok had failed to fulfil its responsibility to respect human rights as laid out in the UN Guiding Principles on Business and Human Rights by failing to conduct adequate human rights due diligence and to mitigate the serious risks that its platform poses to the mental and physical health of young users with existing mental health concerns.

The new findings, contained in this briefing, of TikTok's failures to respect children's rights and to fulfil its responsibilities under international human rights standards as well as its obligations under the EU's Digital Services Act must now urgently be considered as part of the European Commission's ongoing investigation into potential infringements as a designated Very Large Online Platform under the DSA in relation to the assessment and mitigation of systemic risks to children's rights under Articles 28(1), 34(1), 34(2) and 35(1).<sup>61</sup>

Ahead of the Global Mental Health Awareness Week in May 2025, Amnesty International asked TikTok what changes the company had implemented in the close to 18 months since the publication of Amnesty International's 2023 research. TikTok's response listed familiar 'well-being' measures, most of which were already in place when the previous research in 2023 was conducted and failed to acknowledge the app's "rabbit hole" problem.<sup>62</sup> It also failed to produce evidence of any new targeted measures to address this issue.

---

61 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), (previously cited).

62 Amnesty International, "TikTok fails to address risks to children and young people's mental health despite past warnings", 12 May 2025 (previously cited).

In TikTok’s 2024 risk assessment under the EU’s Digital Services Act, the company acknowledged that “certain types of concentrated content, though not violative of TikTok’s Community Guidelines, may cause harm by inadvertently reinforcing a negative personal experience for some viewers. For example, there may be an impact to mental well-being, particularly for Younger Users, associated with concentrated content relating to extreme dieting and body-image-related content.” In the same section of the risk assessment, TikTok listed mitigation measures such as the proactive enforcement of its community guidelines, maintaining content eligibility standards and applying “dispersion techniques” to the ‘For You’ feed as well as user tools, including filtering options and a refresh function to reset the feed. All these measures were already in place in 2023, when Amnesty International conducted its research and demonstrated that young users were exposed to systemic risks to human rights on the platform despite these measures. TikTok’s risk assessment did not include any metrics used to measure the prevalence of these risks, nor benchmarks defining what the company considers to be acceptable levels of exposure to “concentrated content”.

In May 2025, TikTok launched in-app meditation exercises “to encourage young people to switch off at night”.<sup>63</sup> In July 2025, TikTok launched “Well-being Missions”, gamifying pathways towards the platform’s well-being tools.<sup>64</sup> Whilst these updates indicate that the company is aware of some of the risks it exposes children and young people to, they are inadequate and insufficient mitigation measures in the wider context of the app’s addictive design features,<sup>65</sup> which remain unaddressed. Whilst non-binding, the EU’s Guidelines on the Protection of Minors now offer clear guidance to platforms on how to interpret their obligations under Article 28 of the DSA. TikTok’s core design features implemented in the platform’s default user experience for children, including autoplay, infinite scroll and engagement-based content recommendations in its ‘For You’ feed evidently contradict this guidance.

The technical research conducted for this briefing shows that TikTok’s stated measures under the Digital Services Act to mitigate the risk of exposing young users to “rabbit holes” or “concentrated content” that may be harmful to their mental and physical health are ineffective and inadequate. Amnesty International found no evidence that harmful clusters of “concentrated content” (TikTok’s term) were broken up. Instead, like in 2023, researchers found that the “rabbit hole” effect took effect within mere minutes of starting the manual experiments and remained firmly in place for the next hours with increasing numbers of videos openly romanticizing or normalizing suicide.

Both the testimonies of young survivors of self-harm and the technical research also confirmed that TikTok’s stated preemptive enforcement of its community guidelines has failed to prevent children’s exposure to harmful and violative content. This includes content that contains known “algospeak” references (deliberate efforts to circumvent content moderation, for example through emojis, abbreviations and replacing letters with numbers) to suicide and self-harm, whose meanings were obvious to researchers conducting research on the platform, as well as a publicly well-documented suicide challenge (discussed above in section 4.1). Had TikTok conducted adequate due diligence and addressed the human rights risks in its platform, the company would have known about this content and been able to identify it, as well as the risk that it poses. The testimonies further indicate that TikTok has failed to adequately moderate content upon users’ flagging of content, which the research participants had reason to believe was violating the company’s guidelines.

Importantly, these findings need to be considered against the backdrop of TikTok’s increased reliance on artificial intelligence to make critical decisions on permissible content. The company laid off hundreds of content moderators worldwide in 2024 and 2025, including reportedly “phasing out”

---

63 TikTok, “Expanding TikTok’s Mental Health Education Fund and well-being features”, 15 May 2025, <https://newsroom.tiktok.com/en-us/tiktok-expanding-mental-health-education-fund-and-wellbeing-features>

64 TikTok, “New ways we’re helping our community build balanced digital habits”, 30 July 2025, <https://newsroom.tiktok.com/en-us/new-ways-to-help-our-community-build-balanced-digital-habits>

65 Amnesty International, Driven into the Darkness: How TikTok’s ‘For You’ Feed Encourages Self-Harm and Suicidal Ideation (Index: POL 40/7350/2023), 7 November 2023 (previously cited).

its entire Trust and Safety department in Germany.<sup>66</sup> Contacted by the Associated Press in January 2025 about the case of Marie, Stéphanie Mistre's daughter who died by suicide, TikTok still pointed to its "40,000 trust and safety professionals worldwide" to argue that the company was ensuring that content which encourages suicide is swiftly removed, whilst it was already in the process of decreasing its trust and safety workforce.<sup>67</sup> A recent study comparing the performance of leading multimodal large language models with that of human content moderators found that "human reviewers remain superior in accuracy, particularly in complex or nuanced classifications", with an even more marked difference in relation to non-English content.<sup>68</sup> Given the complexities of moderating suicidal ideation content that often relies on slang and "algospeak" in a non-English language, the difference in quality between human and AI moderation is likely significant. Yet the study also found that content moderators cost 40 times more than using AI, offering clear short-term financial incentives to favour cost-saving AI models over platform safety performance. Whilst trust and safety professionals value AI as an aid that saves them from reviewing every harmful graphic piece of content,<sup>69</sup> replacing entire trust and safety teams with AI is an alarming step in the wrong direction as the numerous inadequacies of automated content moderation are well documented.

Though not the focus of this research, the testimonies pertaining to the observed high prevalence of harmful comments and the formation of online communities collectively engaging in the normalization of self-harm point to potential further design risks in relation to the engagement-based amplification of harmful *comments* (as opposed to the amplification of video posts which was the focus of this research), which merit further investigation.

In light of these findings, Amnesty International urges the European Commission and national regulators to closely review the identified systemic risks and harms to children's rights and proceed in a timely manner with the Commission's ongoing investigation into potential DSA infringements.

---

66 Reuters, "ByteDance's TikTok cuts hundreds of jobs in shift towards AI content moderation", 11 October 2024, <https://ca.finance.yahoo.com/news/bytedance-cuts-over-700-jobs-022649989.html>; Euronews, "TikTok content moderators in Germany strike over AI taking their jobs", 23 July 2025, <https://uk.news.yahoo.com/tiktok-content-moderators-germany-strike-111824712.html>

67 AP, "Families sue TikTok in France over teen suicides they say are linked to harmful content", 25 January 2025, <https://apnews.com/article/tiktok-france-trial-suicide-lawsuit-fa8f979c3121a3c5712d52a300c9005f>

68 The Register, "Humans make better content cops than AI, but cost 40x more", 8 August 2025, <https://www.msn.com/en-us/health/medical/humans-make-better-content-cops-than-ai-but-cost-40x-more/ar-AA1KbhjQ>

69 Everything in Moderation, "Where are tomorrow's T&S experts coming from?", 18 August 2025, <https://www.everythinginmoderation.co/future-trust-safety-expert/>

# 9. CONCLUSION

Almost two years after Amnesty International published *Driven into the Darkness*, a damning report about TikTok's harmful "rabbit holes" and their risks to children and young people's mental health, this briefing has shown that TikTok continues to fail to address the systemic risks arising from its engagement-seeking business model. In 2024, the company first acknowledged the presence of systemic risks in relation to "concentrated content" - TikTok's euphemism for "rabbit holes" - for children when obliged to do so under new transparency requirements contained in the EU's Digital Services Act. Yet this renewed research has shown that the company's stated mitigation measures are failing to stop these same harmful "rabbit holes" from drawing child accounts into an online world in which self-harm and despair are the norm.

Survey evidence reveals that many children and young people are aware of some of the design risks associated with the profiling and engagement-based business model of leading social media companies but find themselves unable to shield themselves from these risks, even when they try. Amnesty International's interviews with young survivors of self-harm and affected and bereaved parents highlight the risks of exposing already struggling adolescents to an online platform that is designed to maximize engagement at all costs. Their testimonies reveal how drawing them into an artificial tunnel vision fixated on mental health struggles, TikTok normalized and exacerbated their self-harm and suicidal ideation up to the point of recommending content on suicide methods and challenges. Marie and Charlize's parents, who lost their daughters to suicide at the age of 15, believe that TikTok's 'For You' feed pushed their children over the edge, exacerbating their mental health challenges with the worst possible outcome. They now feel compelled to warn other parents and young people of the risks and to call for urgent change.

Amnesty International's new manual and automated research experiments have confirmed that TikTok's 'For You' feed can draw child accounts into "rabbit holes" of depressive and suicide-related content just as easily as in 2023. Mere hours on the platform suffice to expose child accounts who watch mental health-related content recommendations to an almost uninterrupted stream of potentially deeply harmful videos. This includes videos containing references to known signifiers of self-harm and suicide, revealing clear failures in TikTok's risk mitigation. TikTok is thus failing in its responsibility to respect human rights under the UN Guiding Principles, the OECD Guidelines, as well as failing to fulfil its obligations under the DSA.

Amidst an industry-wide trend to reduce the costs of trust and safety departments through increased reliance on artificial intelligence, regulators must urgently step up their enforcement action to prevent further risks and harms to children and young people in France and elsewhere. The evidence contained in this briefing points to clear infringements of the EU's Digital Services Act by TikTok and must urgently feed into the European Commission's ongoing investigation, resulting in timely and effective enforcement action to compel TikTok to finally make its popular app safe for children and young people in the EU and around the world.

# RECOMMENDATIONS

## RECOMMENDATIONS TO THE EUROPEAN COMMISSION

- Ensure that the Digital Services Act is robustly enforced. The European Commission must urgently include these findings in its ongoing investigation into TikTok’s potential infringements of its obligations under the DSA and ensure a timely and effective investigation and enforcement process.
- Introduce guidelines for Very Large Online Platforms conducting systemic risk assessments to ensure that the information contained in the assessments is appropriate and specific and can be used by civil society and regulators to scrutinize the platforms.
- Ensure that penalties issued as a result of non-compliance decisions are sufficient to encourage the non-repetition of harm.

## RECOMMENDATIONS TO THE FRENCH GOVERNMENT AND EU MEMBER STATES

- Ensure the fulfilment and protection of children’s rights by working with EU institutions and member states to create and enforce common regulatory frameworks, including the Digital Services Act, the AI Act, GDPR and forthcoming proposals to address addictive design under the Digital Fairness Act, to ensure “safety by design” rather than pursuing restrictions to adolescents’ access to social media platforms which would unfairly restrict their right to freedom of expression and freedom of peaceful assembly online.
- Ensure that access to and use of essential digital services and infrastructure such as TikTok and other social media platforms are not made conditional on ubiquitous surveillance of children, young people or adult users. As a first step, prevent companies from making access to their service conditional on individuals ‘consenting’ to the collection, processing or sharing of their users’ personal data for content targeting and marketing or advertising.
- Regulate social media companies to ensure that content-shaping algorithms used by online platforms are not based on profiling by default and that they require an opt-in rather than an opt-out, with the consent for opting in being freely given, specific, informed and unambiguous. The collection and use of inferred data (for example, recommendations based on watch time and likes) to personalize ads and content recommendations should be banned. Rather, users should be in control of which signals or declared interests they want the platform to factor into the shaping of their feed. For those who prefer a feed based on personalized recommendations, they must be given the option to communicate personal interests to the platform based on specific, freely given and informed consent and based on prompts made in child-friendly language.
- Enact/update and enforce regulatory frameworks to ensure people are able to exercise in practice their right to choose privacy-respecting alternatives to surveillance-based business models. This includes measures to ensure interoperability (the ability to communicate with existing contacts using another compatible platform) rather than just data portability so that people can move between services without social detriment, and to lessen network effects.

## RECOMMENDATIONS TO TIKTOK

- Transition to a rights-respecting business model that is not based on invasive data tracking. As a first step, ensure that human rights due diligence policies and processes address the systemic

and widespread negative human rights impacts of the current business model, in particular on the right to privacy, the rights to freedom of opinion and thought and the right to health. TikTok must be transparent about the risks and abuses, including to human rights, that it has identified and how they have been addressed.

- Stop maximizing “user engagement” at the expense of its users’ health and other human rights, given the available evidence of the negative impacts of compulsive platform use especially on young users’ health and well-being. As part of its human rights due diligence process, TikTok must identify design elements in cooperation with users, including children and young people, and independent experts, which encourage addictive platform use and social comparison, and replace these with a user experience that is focused on ‘safety by design’ and the best interests of the child.
- Undertake early, proactive, ongoing human rights due diligence throughout the lifecycle of algorithmic technologies, both before and after the roll-out and implementation of new systems and design features, in order that risks can be identified during the development stage and human rights abuses and other harms immediately picked up once the technologies have been launched.
- Engage children and young people, academic and civil society experts and other relevant stakeholders in its ongoing human rights due diligence processes. Children and young people should also play a core part in implementing “safety by design” by being involved in the development process of tools and features of social media platforms.
- Conduct and publish human rights impact assessments on a regular basis and include detailed information on risks and mitigating measures taken with respect to specific countries (especially when heightened due diligence processes are required, where systems may have a greater impact due to political conflicts or humanitarian emergencies), specific categories of users such as children and young people, and specific product changes.
- To respect privacy and to provide users with real choice and control, a profiling-free social media ecosystem should not just be an option but the norm. Content-shaping algorithms used by TikTok and other online platforms should therefore not be based on profiling (for example, based on watch time or engagement) by default and must require an opt-in instead of an opt-out, with the consent for opting in being freely given, specific, informed (including using child-friendly language) and unambiguous.
- Cease collecting intimate personal data and drawing inferences from a user’s watch time and engagement about their interests, emotional state or well-being for the purposes of ‘personalizing’ content recommendations and ad targeting. Rather than using pervasive surveillance to adapt feeds to a user’s interests, TikTok should enable users to communicate their interests through deliberate prompts (for example, users could be asked to enter specific interests if they would like to be served personalized recommendations) and only when based on users’ freely given, specific and informed consent.
- Introduce additional measures to prevent at-risk users from falling into compulsive use patterns and “rabbit holes” of potentially harmful content. These could include a mandatory daily limit on the number of personalized recommendations offered to children and a list of regularly updated terms related to borderline mental health content, which are deemed suitable to search for but not suitable for amplification in the ‘For You’ feed.
- Introduce “friction” measures as a mitigation strategy. As part of its human rights due diligence processes, TikTok should invest in research to identify and incorporate measures to limit the rapid and often disproportionate algorithmic amplification of borderline content.
- As an interim measure, urgently improve the effectiveness of measures aimed at diversifying the content recommendations in a user’s ‘For You’ feed to ensure content that risks encouraging harm does not populate a feed, including by introducing effective user control tools. These

should be easy to find and understand and offer users an effective way of suppressing future recommendations of content related to a specific topic, hashtag or user.

- Radically improve transparency in relation to the use of content-shaping and content-moderation algorithms, ensuring that their mechanics are publicly available and are also explained as part of the continued user experience and in clearly understandable and child-appropriate terms in all relevant languages.
- Ensure consistency in content moderation decision-making, ensure adequate human oversight of automated content moderation and appropriate investment in human and automated content moderation and trust and safety resourcing across all languages.
- Revamp TikTok's complaint mechanism to ensure it is accessible and effective in ensuring that TikTok can quickly and adequately address any concerns flagged by users or other concerned parties.



**AMNESTY INTERNATIONAL  
IS A GLOBAL MOVEMENT  
FOR HUMAN RIGHTS.  
WHEN INJUSTICE HAPPENS  
TO ONE PERSON, IT  
MATTERS TO US ALL.**

CONTACT US



[info@amnesty.org](mailto:info@amnesty.org)



+44 (0)20 7413 5500

JOIN THE CONVERSATION



[www.facebook.com/AmnestyGlobal](https://www.facebook.com/AmnestyGlobal)



[@amnesty](https://twitter.com/amnesty)

# DRAGGED INTO THE RABBIT HOLE

## NEW EVIDENCE OF TIKTOK'S RISKS TO CHILDREN'S MENTAL HEALTH

In November 2023, Amnesty International published *Driven into the Darkness*, a damning report highlighting how easily children and young people expressing an interest in mental health could be drawn into “rabbit holes” of depressive and suicidal content in TikTok’s ‘For You’ feed. The research revealed how TikTok employs addictive platform design to maximize engagement, undermining the right to privacy, freedom of thought and mental and physical health.

This follow-up research briefing presents new evidence that almost two years later, TikTok's ‘For You’ feed is still pushing children and young people engaging with mental health content into a cycle of depressive, self-harm and suicide content. Bringing together testimony from affected young people and parents in France as well as renewed technical research evidence, the briefing documents TikTok’s failure to address its systemic design risks for children and young people both under international business and human rights standards as well as under the company’s binding obligations under the EU’s Digital Services Act. It is an urgent appeal to the company itself, but also to EU and French regulators to take decisive action to force the company to respect children’s and human rights.