DRIVEN INTO THE DARKNESS

HOW TIKTOK'S 'FOR YOU' FEED ENCOURAGES SELF-HARM AND SUICIDAL IDEATION

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.

First published in 2021
by Amnesty International Ltd
Peter Benenson House, 1 Easton Street,
London WC1X 0DW, UK

Index: POL 40/7350/2023
Original language: English

amnesty.org

Cover photo: *© Luisa Balaban*

AMNESTY INTERNATIONAL

CONTENT WARNING

This report covers sensitive issues including self-harm and suicide and contains graphic imagery of TikTok content related to these issues. 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.***[1]

ACKNOWLEDGEMENTS

Amnesty International would like to express its gratitude to its technical partners, the Algorithmic Transparency Institute (National Conference on Citizenship) and AI Forensics, as well as the individuals, organizations, activists and experts who facilitated and guided this research. These include among others: Dr S. Bryn Austin, Zeyna Awan, Dr Jerome Cleofas, Julien Cornebise, Kevin Gachee, In Touch Community Services, Lillian Kariuki, KERI: Caring for Activists, Dr Amanda Raffoul, Dr Marc Reyes, Aliya Shah, Talang Dalisay, Ian Wafula, and Youth for Mental Health Coalition.

Amnesty International is also indebted to the children and young people in Kenya and the Philippines who shared their experiences and their ideas during the research for this report.

1. 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/, p. 99.

CONTENTS

GLOSSARY

ADHD	Attention deficit hyperactivity disorder
AI	Artificial intelligence: There is no widely accepted definition of the term "artificial intelligence" or "AI". The United Nations Office of the High Commissioner for Human Rights uses the term to refer to a constellation of processes and technologies enabling computers to complement or replace specific tasks otherwise performed by humans, such as making decisions and solving problems, including machine learning and deep learning.[2]
Big Tech	A popular shorthand for the leading information technology companies shaping and dominating the internet.
CDC	Centers for Disease Control and Prevention (USA)
CoE	Council of Europe
CRC	Convention on the Rights of the Child
Data scraping	Automated collection of web data at scale.
DSA	The European Union's Digital Services Act
ECtHR (in footnotes)	European Court of Human Rights
HRC (in footnotes)	UN Human Rights Committee
ICCPR	International Covenant on Civil and Political Rights
ICESCR	International Covenant on Economic, Social and Cultural Rights
LGBTI	Lesbian, Gay, Bisexual, Transgender, Intersex
Metadata	Data about digital files and communications, for example timestamps, location data.
OECD Guidelines	OECD Guidelines for Multinational Enterprises.
OHCHR	Office of the United Nations High Commissioner for Human Rights
"Rabbit hole" effect	A commonly used term in the study of social media platforms; it is used in this report to refer to the narrowing down of content recommendations in TikTok's 'For You' feed to one subject, based on the assumption that this content will elicit strong emotional reactions and that it will keep the user engaged.
UDHR	Universal Declaration of Human Rights
UN Guiding Principles	UN Guiding Principles on Business and Human Rights
VPN	Virtual private network

2. UN Office of the High Commissioner for Human Rights (OHCHR), "The right to privacy in the digital age", 15 September 2021, A/HRC/48/31, fn 2.

1. EXECUTIVE SUMMARY

"Have you ever looked at a bottle of pills and thought of overdosing? Have you ever just wanted for it all to end? Have you ever thought about the release it would give, all the pain that would go away?" Set against the background of small, pixelated figures moving through a maze in an arcade game, an AI-generated voice utters these thoughts in a half-minute-long clip. Watching, rewatching, then moving on to the next clip, the 13-year-old TikTok user scrolling through these videos doesn't feel anything. That is because it is an automated account, set up and programmed by Amnesty International and the Algorithmic Transparency Institute to simulate and explore the digital reality of children and young people living with mental health concerns such as anxiety or depression. With each hour spent on TikTok's 'For You' feed, more of the video clips "recommended" to the teenage account show children and young people crying, or alone in the dark overlaid with text expressing depressive thoughts or faceless voices describing their suffering, self-harm and suicidal thoughts.

In the last three years, marked by the Covid-19 pandemic, TikTok has emerged as a global platform, attracting hundreds of millions of children and young people. TikTok's 'For You' page and the algorithmic recommender system behind it have played a crucial role in catapulting the platform to its current ubiquity in children's and young people's lives. *I feel exposed: Caught in TikTok's Surveillance Web*, published as a companion report to this piece, exposes the deeply discriminatory and invasive model of surveillance for profit that sustains TikTok's business model. This complementary report focuses on the user surveillance, design decisions and personalized content recommendations employed by TikTok in the pursuit of "user engagement" — and their detrimental real-world impact on children and young people.

TikTok's recommender system, the set of algorithms which analyse users' interests and engagement patterns, then match these with new video clips to keep users emotionally engaged and their eyes fixed on the feed, was not built to produce a "rabbit hole" of depressive content. But if a young user shows a clear interest in mental health-related content, that is what the system matches them with in the pursuit of optimal user engagement. If the automated research account set up for this investigation were a human being, better able to distinguish between relevant and irrelevant content, the system would produce such recommendations much faster still.

"Luis" (name changed), a 21-year-old undergraduate student in Manila diagnosed with bipolar disorder, described his experience of TikTok's 'For You' feed to Amnesty International:

> **"When I felt low, I think 80% [of the content] related to mental health. It's like a spiral. It's a rabbit hole because it starts with just one video. If one video is able to catch your attention, even if you don't like it, it gets bumped to you the next time you open TikTok and because it seems familiar to you, you watch it again and then you watch it again and then the frequency of it in your feed rises exponentially."**

This report is based on desk research, a scoping survey, focus group discussions and interviews with more than 300 children and young people in two of the countries with the highest social media usage

worldwide (Kenya and the Philippines) as well as a technical investigation, conducted together with the Algorithmic Transparency Institute and AI Forensics as technical partners, involving more than 30 automated accounts comparing results across Kenya and the United States of America (USA), in addition to manual experiments in Kenya, the Philippines and the USA.

For the automated audit, researchers set up 40 automated accounts with four different pre-defined personas displaying different levels of interest in mental health-related content to mimic different children's behaviours on TikTok. Each account was set up to run for just under 60 minutes in a single session each day for 10 days. The accounts were divided into sub-groups following four different scrolling behaviours, each subgroup rewatching videos associated with a different set of terms and hashtags and skipping unrelated content. 20 accounts were set up to simulate 13-year-olds in the USA and another 20 accounts simulated child users in Kenya (of which 11 were included in the analysis). The age of the accounts (13 years) was chosen to examine the recommendations served to the youngest permitted age group of TikTok users, to whose accounts TikTok applies teen safety measures. The additional manual experiment consisted of hour-long screen recordings covering a researcher's interaction with the 'For You' feed of newly set-up accounts, again representing 13-year-olds.

Amnesty International's research shows that TikTok has maximized the addictive qualities of design choices and engagement strategies employed by competing social media companies, incentivising a race to the bottom between a small number of leading social media companies vying for the highest user numbers and engagement rates. TikTok has done this in spite of mounting scientific evidence of the serious risks associated with addictive use of social media especially for children and young people's health, including sleep and attention problems and even changes in brain structure similar to those observed in people experiencing drug addiction.

Beyond its addictive nature, TikTok's 'For You' feed poses additional risks for children and young people with pre-existing mental health concerns. A technical investigation conducted by Amnesty International, the Algorithmic Transparency Institute (National Conference on Citizenship) and AI Forensics shows that children and young people who watch mental health-related content on TikTok's 'For You' page can easily be drawn into "rabbit holes" of potentially harmful content, including videos that romanticize and encourage depressive thinking, self-harm and suicide.

After 5-6 hours on the platform, almost one in every two videos served to automated accounts programmed to simulate 13-year-old children in Kenya and the USA with an interest in mental health were mental health-related and potentially harmful, roughly 10 times the volume served to accounts with no interest in mental health. A manual review of 540 videos recommended to a sample of these bot accounts showed a steady progression from 17% of the videos served in the first hour being categorized as potentially harmful to 44% of content in the tenth hour (based on hour-long sessions spread out across ten days).

Amnesty International observed an even faster "rabbit hole" effect with even higher rates of potentially harmful content when researchers manually rewatched mental health-related videos suggested to supposed 13-year-olds in Kenya, the Philippines and the USA. Amongst the recommendations served to an account located in the Philippines, the first video tagged with #depresionanxiety [sic] showing a young boy in distress was suggested within the first 67 seconds of scrolling through recommended content on the 'For You' page. From minute 12 onwards, 58% of the recommended posts related to anxiety, depression, self-harm and/or suicide and was categorized as potentially harmful for children and young people with pre-existing mental health concerns.

In the US-based manual experiment, the fourth video shown was tagged #paintok and focused on text reading "when you realize you've never been put first your entire life but instead are just that person that fills a void in other people's lives until they don't need you anymore". From the 20th video onwards (less than three minutes in), 57% of the videos are related to mental health issues, with at least nine posts romanticizing, normalizing or encouraging suicide in a single hour.

The Kenyan account in the manual experiments saw the slowest progression towards a feed filled with depressive content. However, once that point was reached (20 minutes into the experiment), 72% of the videos recommended in the next 40 minutes related to mental health struggles, with at least five references to suicidal thinking or the content creator's death wish. Not a single mental health-related video was produced by a mental health care professional or recognized mental health organization.

The case of Molly Russell, the 14-year-old British girl who died from an act of self-harm after having viewed depressive content on Instagram, shows in the most tragic way how exposing a young person experiencing depressive symptoms to a social media feed consisting of a high volume of posts that discuss, normalize or even romanticize depressive thinking, self-harm and suicide has the potential to exacerbate young users' pre-existing mental health issues and can potentially contribute to harmful and even devastating real-world actions. Such interference with a person's thoughts and emotions constitutes an abuse to the right to freedom of thought and the right to health.

Given the well-documented emotional vulnerabilities of children and young adults and the extensive evidence based on previous civil society and media reports, TikTok should know and have identified that its algorithmic recommender systems risk exposing young users to "rabbit holes" of potentially harmful posts.

To fulfil its responsibilities as laid out in the UN Guiding Principles on Business and Human Rights, TikTok should be conducting appropriate human rights due diligence to identify, prevent, mitigate and account for how it is addressing its potential and actual harms. As part of this human rights due diligence, the company should have identified the risks to children and young people inherent in the design of its platform given the growing evidence of systemic risks associated with algorithmic recommender systems compiled by media, civil society organizations and international institutions and should have taken steps to mitigate them. The company's response to Amnesty International's questions points to a patchwork of individual measures such as redirecting certain searches to help resources, an option to "refresh" the feed and partnerships with mental health support organizations. TikTok's response demonstrates a lack of adequate procedures to address the systemic nature of these risks and the measures and safety tools implemented by the company fall short of addressing the magnitude of the systemic risks identified in this report.

Individual actions by a single company are however insufficient to rein in a business model that is fundamentally incompatible with human rights, in particular the right to privacy, the right to freedom of thought and the right to health. States must therefore effectively regulate "Big Tech" companies like TikTok in line with international human rights law and standards to protect and fulfil children and young people's rights.

Amnesty International calls on TikTok to urgently implement the following recommendations:

- 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.
- TikTok must 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.
- 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.

- TikTok must 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.

To fulfil children and young people's rights, states must:

- 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 in line with international human rights law and standards 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 sensitive personal data (for example, recommendations based on watch time and likes which allow for inferences of sensitive information) to personalize ads and content recommendations must be banned.

2. METHODOLOGY

2.1 PURPOSE AND SCOPE OF THE REPORT

This report explores the risks to human rights inherent in leading social media platforms' use of targeted content recommendations and their competition for young users' attention, with a specific focus on TikTok's addictive platform design and its amplification of depressive and self-harm related content.

It forms one of two reports by Amnesty International investigating the human rights impacts of the surveillance-based business model used by TikTok and other social media platforms – a model involving invasive tracking of users which Amnesty International has previously described as a "Faustian bargain, whereby [people] are only able to enjoy their human rights online by submitting to a system predicated on human rights abuse"[3] – on children (people under the age of 18) and young people (15-24-year-olds, also referred to as "young users" in this report).

The companion to this report, *I feel exposed: Caught in TikTok's Surveillance Web*, highlights abuses by tech companies of children's right to privacy, which are inherent in the business model of "Big Tech" platforms.[4] This complimentary report focuses on the detrimental real-world impact on children and young people of the user surveillance, design decisions and personalized content recommendations employed by social media companies in the pursuit of "user engagement". It explores the risks that TikTok poses to children and young people, particularly for those experiencing mental health issues such as depression, anxiety and self-harm.

It builds on previous work by Amnesty International, including its 2022 report, *The Social Atrocity – Meta and the Right to Remedy for the Rohingya*, which documented Facebook's role in the amplification of posts inciting violence, hatred and discrimination against the Rohingya people, ultimately contributing to the ethnic cleansing against them by Myanmar security forces in 2017.[5]

Together, these reports contribute to the growing evidence base of Amnesty International's global campaign for corporate accountability and redress for human rights abuses associated with the surveillance-based business model of Meta, Google and TikTok and other "Big Tech" platforms.

2.2 RESEARCH FOCUS

All leading social media companies vie for ever-increasing user numbers and the greatest share of their screen time. The focus of this report is on TikTok, because, in the last three years, marked

3. 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, p. 5.
4. Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web* (Index: POL 40/7349/2023), 7 November 2023.
5. 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

by the Covid-19 pandemic, it has emerged as a global platform, attracting hundreds of millions of children and young people, many of which now call TikTok their favourite platform according to market research.[6] Its rapid growth saw the number of active users on TikTok roughly doubling between December 2019 and September 2021,[7] when the company announced it had surpassed 1 billion monthly users.[8] The majority of TikTok users are thought to be children and young people.[9] In August 2020, it was reported that TikTok had "classified more than a third of its 49 million daily users in the USA as being 14 years or younger", and that many were believed to be below 13 years old, the minimum age at which access to the adult version of the platform is permitted by TikTok in the United States of America (USA) and most other countries, in which it operates.[10]

TikTok's rapid rise to market dominance, along with the particular design elements it uses to maximize the amount of time users spend on its platform, is giving rise to growing concern that the company is exposing children and young people to an addictive and potentially unsafe platform.[11]

Although TikTok is used almost worldwide, the available literature on its impact (and that of other social media platforms) on children and young people, including potential harms, is overwhelmingly focused on Australia, Europe and the USA.[12] Almost 90% of the world's youth population live in developing countries[13] but their experiences are vastly underrepresented in studies and public debates about social media. In order to broaden the geographic evidence base, research for this report focused on TikTok use by children and young people in Kenya and the Philippines, which were selected because they have exceptionally high rates of time spent daily by users aged 16 to 64 on social media (just over three hours per day in Kenya and just over four hours in the Philippines which are some of the highest rates globally compared with a two-and-a-half-hour worldwide average).[14] The USA was included in the technical investigation element of the research (see research methodology below) in order to explore possible differences in platform design choices made by TikTok in a country where there is a high-level of public debate about the risks of social media for teenage users, and which can also be considered a key market for the company.

---

6. Qustodio, *Social Media Annual Data Report 2021:* Living and Learning in a Digital World, Chapter 2, Social Media, https://qustodio.com/en/social-media-qustodio-annual-data-report-2021/
7. We Are Social and Hootsuite, *Digital 2022 Global Overview Report*, 26 January 2022, https://wearesocial.com/hk/blog/2022/01/digital-2022-another-year-of-bumper-growth
8. Reuters, "TikTok hits 1 billion monthly active users globally – company", 27 September 2021, https://www.reuters.com/technology/tiktok-hits-1-billion-monthly-active-users-globally-company-2021-09-27
9. Meltwater, "54 TikTok stats you need to know [2023]", 30 December 2022, https://www.meltwater.com/en/blog/tiktok-statistics; Guardian, "What TikTok does to your mental health: 'It's embarrassing we know so little'", 30 October 2022, https://www.theguardian.com/technology/2022/oct/30/tiktok-mental-health-social-media; Reuters Institute, *Digital News Report 2023*, June 2023, https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/dnr-executive-summary
10. New York Times, "A third of TikTok's U.S. users may be 14 or under, raising safety questions", 14 August 2020, https://www.nytimes.com/2020/08/14/technology/tiktok-underage-users-ftc.html; TikTok offers a "limited app experience" with "additional safety and privacy protections" for users under the age of 13, see TikTok, "TikTok for younger users", 13 December 2019, https://www.newsroom.tiktok.com/en-us/tiktok-for-younger-users. The New York Times article refers to allegations by a former TikTok employee that TikTok allowed video content from children younger than 13 to "remain online for weeks". As the under-13s version of the app does not permit child users to share videos, this implies that the under-13s were using the adult version of the app, likely having submitted false age information. In 2019, TikTok agreed to pay a fine of $5.7 million to settle allegations by the FTC that the company had illegally collected information from children under the age of 13. Wire, "FTC Hits TikTok With Record $5.7 Million Fine Over Children's Privacy", 27 February 2019, https://www.wired.com/story/tiktok-ftc-record-fine-childrens-privacy/; for more information on TikTok's age policy see TikTok, "Guardian's Guide", https://www.tiktok.com/safety/en/guardians-guide/ (accessed on 26 September 2023).
11. See for example: Wall Street Journal, "'The corpse bride diet': How TikTok inundates teens with eating-disorder videos", 17 December 2021, https://www.wsj.com/articles/how-tiktok-inundates-teens-with-eating-disorder-videos-11639754848; Sophia Petrillo, "What makes TikTok so addictive?: An analysis of the mechanisms underlying the world's latest social media craze", 13 December 2021, Brown Undergraduate Journal of Public Health, Issue 2021-2022, https://www.sites.brown.edu/publichealthjournal/2021/12/13/tiktok/; BBC, "Inside Tiktok's real-life frenzies - from riots to false murder accusations", 21 September 2023, https://www.bbc.co.uk/news/technology-66719572; Guardian, "What TikTok does to your mental health: It's embarrassing we know so little", 30 October 2022, https://www.theguardian.com/technology/2022/oct/30/tiktok-mental-health-social-media
12. Sakshi Ghai, Lucía Magis-Weinberg and others, "Social media and adolescent well-being in the Global South", August 2022, Current Opinion in Psychology, Volume 46, https://pubmed.ncbi.nlm.nih.gov/35439684/
13. UNESCO, "Thematic factsheet: youth and empowerment", 31 January 2023, https://www.unesco.org/en/youth-and-empowerment
14. We are social, Digital 2022 Global Overview Report, 26 January 2022, https://www.wearesocial.com/hk/blog/2022/01/digital-2022-another-year-of-bumper-growth/; Amnesty researchers were unable to find comparable global statistics for children and young people specifically.

2.3 RESEARCH METHODOLOGY

The research for this report took place between October 2022 and July 2023 and comprised the following elements:

1. In-depth desk research, including a review of relevant academic, UN, NGO, media and other secondary sources.

2. A scoping exercise in the form of an online questionnaire, distributed via Amnesty International and partner organizations' social media channels and responded to by 550 children and young adults between the ages of 13 and 24 in 45 countries in October and November 2022. The questionnaire asked about use of leading social media platforms, experiences on these platforms, likes and dislikes, reactions to negative experiences and visions for change, with a view to better understand lived experiences, concerns and attitudes towards social media.[15]

3. Five semi-structured, in-person interviews and 20 focus group discussions (FGDs) with more than 300 children and young people aged 14 to 24 (of which about 180 were TikTok users) in Kenya (conducted in March 2023) and the Philippines (conducted in April and May 2023). In Kenya, in-person interviews and FGDs with school children, young activists, university students and recent graduates took place in the capital Nairobi, in the cities of Kisumu and Mombasa, and in Machakos County. In the Philippines, in-person interviews were conducted in the capital Manila and Batangas Province. Nine additional online interviews were held with children and young people in regions of the Philippines that were not accessible to the researchers for practical reasons or due to security concerns. Amnesty International researchers spoke to 37 children and young people in the Philippines, of which 34 spoke about their experiences on TikTok in addition to other social media platforms.

   In Kenya, Amnesty International relied on Amnesty International Kenya's network of 'human rights friendly schools' and Amnesty International chapters to approach young people to interview. In the Philippines, a call for participation was shared through Amnesty International Philippines' youth network and with networks of mental health support services and youth-led mental health campaigning organizations.

   This referral pathway to children and young people, the focus on interviews or very small focus groups and the greater number of young adults with some access to mental health support services in our cohort in the Philippines likely contributed to a much higher rate of participants' self-reporting of past or current mental health issues compared with the mostly teenage group in Kenya. Learnings from the FGDs in Kenya conducted in March 2023 led to a more focused approach to inviting research participants, with the assistance of mental health organizations and Amnesty International's young activist network, who felt that their mental health had been impacted by their use of social media.

   All interviews and FGDs were conducted by Amnesty International researchers in English, with occasional translation support when interviewees expressed thoughts and experiences in Kiswahili and Dholuo (Kenya) or Tagalog (Philippines). Questions were asked about their use of different social media platforms, their experiences on these platforms, their opinions about regulatory efforts with regards to social media use by children, the impact of their social media use on their lives and the perceived impact of their platform experiences on their emotional state and mental health.

15. Amnesty International. "'We are totally exposed': Young people share concerns about social media's impact on privacy and mental health in global survey", 7 February 2023, https://www.amnesty.org/en/latest/news/2023/02/children-young-people-social-media-survey-2

Guardian or parental consent was sought and granted for all participants under the age of 18 and informed consent was taken from each child. Participants over the age of 18 gave their informed consent in advance of interviews or participation in FGDs. To protect the identity of our interviewees, this report uses pseudonyms throughout for all research participants other than the expert interviewees who consented to the publication of their names.

4. 14 expert interviews with specialist adolescent psychologists, (youth) mental health campaigners, public health experts and teachers in Kenya, the Philippines and the USA were conducted between February and May 2023.

5. The focus on children and young people necessarily set ethical limits on how we could explore the means through which algorithmic systems interact with users. To avoid exposing young users to harmful content or compromising their privacy, while at the same time producing the most representative results possible, the research also involved a two-part technical investigation conducted in June and July 2023:

    - An automated algorithmic audit of TikTok's 'For You' page recommender system using research accounts simulating 13-year-olds' engagement with the platform in Kenya and the USA.
    - Screen recordings of three manually run research accounts, set up to represent 13-year-old users in the Kenya, the Philippines and the USA, tracking the amplification of mental health-related content in TikTok's 'For You' page.

Further details about the methodology of our technical investigation are discussed in Chapter 6.

6. A research letter asking for detailed information about TikTok's human rights due diligence processes, data protection policies and measures to promote child users' and young adults' well-being and mental health as well as a Right of Response letter outlining the key findings of this report. TikTok's responses from 12 July 2023 and 29 October 2023 can be found in Annex 3 and 4.

3. BACKGROUND

3.1 MISSING THE POINT: RESEARCH GAPS AND CORPORATE OBSTACLES TO STUDYING THE HUMAN RIGHTS RISKS OF SOCIAL MEDIA PLATFORMS FOR CHILDREN AND YOUNG PEOPLE

In October 2022, a Coroner's inquest into the death of British teenager Molly Russell was the first to formally identify social media as having "likely contributed" to a person's death.[16] A considerable proportion of Molly Russell's Instagram feed in the months before she died at the age of 14 consisted of content depicting and promoting self-harm, depression and suicide.[17] The Coroner ruled that her death had resulted from "an act of self-harm while suffering from depression and the negative effects of online content", some of which "tend[ed] to portray self-harm and suicide as an inevitable consequence of a condition that could not be recovered from" with recommender systems presenting Molly with a "limited and irrational view without any counterbalance of normality".[18] The Coroner recommended that the UK government and social media companies "review the use of algorithms to provide content".[19] It is one of the most strongly worded indictments to date by a public authority in the debate around social media's impact on teenagers' mental health.

SOCIAL MEDIA FEEDS AND TIKTOK'S 'FOR YOU' FEED

Contrary to traditional mass media, social media platforms create a personalized feed of user-generated content. Originally largely populated by content created by the 'friends' with whom the user connected and the pages they followed, social media platforms including TikTok increasingly recommend content created by individuals outside of the user's own network, based on the user's declared or inferred interests. TikTok's 'For You' feed, the default user experience, is often portrayed

16. North London Coroner's Service, "Regulation 28 report to prevent future deaths", 13 October 2022, https://www.judiciary.uk/wp-content/uploads/2022/10/Molly-Russell-Prevention-of-future-deaths-report-2022-0315_Published.pdf
17. 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
18. North London Coroner's Service, "Regulation 28 report to prevent future deaths", 13 October 2022 (previously cited).
19. BBC News, "Molly Russell: Coroner's report urges social media changes", 14 October 2022, https://www.bbc.co.uk/news/uk-england-london-63254635

by the media as well as marketing and technology experts as the most elaborate result of this shift towards a hyper-personalized feed, where users only need to watch recommended content for different amounts of time for the system of algorithms to gauge the user's interests, match these with the available video content and to serve up more personalized recommendations and advertisements.*

* Though the EU's Digital Services Act bans targeted advertising towards minors. For a detailed discussion of TikTok's data collection and advertising practices, see Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, 7 November 2023 (previously cited).

Yet, social media companies make it notoriously difficult for independent researchers to audit their algorithmic systems.[20] Amnesty International and other civil society organizations have called for regulatory action to prevent companies from hindering independent, public interest research into the functioning and impact of social media platforms, warning that:

> **"Many [social media] platforms actively hinder research through their terms of service, through technical measures, or through intimidation and threats of legal action. In particular, platform efforts to prevent scraping [the automated collection of data at scale] have had a chilling effect on the researchers trying to hold them accountable."[21]**

Nevertheless, as the number of social media users and the amount of time spent by children on social media has increased year-on-year, news coverage and public discourse on its role in children's lives has become increasingly critical of the potentially harmful role of these platforms on young people. This is most notable in Europe and the USA, where public debate was fuelled by the publication of leaked internal documents, known as the "Facebook Papers" in 2021, which included survey findings shared within the company in 2019, according to which Instagram made "body image issues worse for one in three teen girls."[22]

Prior to 2021, there was already much debate among psychologists and public health experts about whether social media use has contributed to a rise in depression and self-harm among young people.[23] A key limitation of many of the available studies is that they examine the effect of social media in general, or, even more non-specifically, the effect of screen time on American and European teenagers in general. There is a notable absence of similar evidence from other parts of the world.[24] In reality, social media platforms differ in important ways and change over time, as do children and young

20. Algorithm Watch, "AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook", 13 August 2021, https://www.algorithmwatch.org/en/instagram-research-shut-down-by-facebook; Algorithm Watch, "A guide to the EU's new rules for researcher access to platform data", 7 December 2022, https://www.algorithmwatch.org/en/instagram-research-shut-down-by-facebook
21. Scraping can pose considerable privacy risks depending on the context, the type of data collected and its use, and thus should be limited. However, Amnesty International and others have criticized the blanket ban on scraping imposed by social media companies as an undue impediment to research into systemic risks. In the absence research data access frameworks, such as that soon to be implemented in Europe thanks to new obligations on very large online platforms under the European Union's (EU) Digital Services Act, researchers have at times had to conduct public interest research in contravention of platforms' Terms of Services. Mozilla Foundation, Amnesty International and others, "Response to the European Commission's call for evidence for a Delegated Regulation on data access provided for in the Digital Services Act", May 2023, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3423646_en * Though the EU's Digital Services Act bans targeted advertising towards minors. For a detailed discussion of TikTok's data collection and advertising practices, see Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, 7 November 2023 (previously cited).
22. Guardian, "Facebook aware of Instagram's harmful effect on teenage girls, leak reveals", 14 September 2021, https://www.theguardian.com/technology/2021/sep/14/facebook-aware-instagram-harmful-effect-teenage-girls-leak-reveals; Meta dismissed the findings contained in the "Facebook Papers", arguing that they were "based on selected documents that are mischaracterized and devoid of any context.", see Washington Post, "A whistleblower's power: Key takeaways from the Facebook Papers", 26 October 2021, https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/
23. For an extensive ongoing overview, see Jonathan Haidt and others, "Social Media and Mental Health: A Collaborative Review", New York University, https://docs.google.com/document/d/1w-HOfseF2wF9YIpXwUUtP65-olnkPyWcgF5BiAtBEy0/edit (accessed on 12 September 2023).
24. Sakshi Ghai and others, "Social media and adolescent well-being in the Global South" (previously cited); Sakshi Ghai, Luisa Fassi, Faisal Awadh & Amy Orben, "Lack of sample diversity in research on adolescent depression and social media use: A scoping review and meta-analysis", 7 February 2023, Clinical Psychological Science, https://www.doi.org/10.1177/21677026221114859

people. In thinking about risks and regulation, the question social media platforms should be asked is not whether all platforms harm all children or even just the 'average' child, if such a proxy can ever exist, but *which* specific aspects of social media platforms could pose risks to *any* of their users or specific groups of users.

Social media companies promote the benefits of their platforms for the lives of children and young people and many of the young people and indeed, many of those who participated in research for this report agreed. Young research participants praised TikTok for its entertainment value, the way it encourages creativity and for creating a sense of community. Numerous studies have highlighted the benefits of online communities, brought together by social media, particularly for marginalized young people searching for peers who share their experiences, including LGBTI youths and young people experiencing mental health issues.[25] But these benefits do not absolve social media companies of their responsibility to identify, assess, prevent and mitigate potential harms, and to remedy, where necessary and appropriate, actual harms resulting from the use of their platforms.

However, while marketing the benefits of their products, social media companies have failed to provide users with transparent risk assessments. They have also made it extraordinarily difficult for independent researchers to investigate the way in which their platforms function and the potential risks they pose to users.[26] In the absence of necessary regulation that might enable researchers to do so in an effective and user privacy-respecting way, research projects have been stifled and researchers threatened with legal action.[27] Moreover, as a fast-paced video platform TikTok poses additional research challenges compared to other platforms because video clips are more difficult and resource-intensive to analyse at scale compared with written posts.

3.2 BETWEEN POLICY AND POPULISM: LIMITED EFFORTS TO REGULATE SOCIAL MEDIA PLATFORMS

It is perhaps unsurprising, then, that it was a social media company's own *internal* research that contributed to a more critical public discourse and policy debates on the impact of social media on children and young people. As noted above, the 2021 "Facebook Papers", included damning evidence of Instagram's role in exacerbating negative body image issues among teenagers.[28] These and other findings from the leaked documents made global headlines and intensified pressure on policymakers to overcome long-standing challenges to the modernization of platform regulatory frameworks that were by then at least two decades out of date.[29]

Nevertheless, progress towards addressing the systemic risks associated with large social media platforms is limited, and few states or regional organizations have adopted legislation. Neither the

25. John Naslund, Ameya Bondre and others, "Social media and mental health: Benefits, risks, and opportunities for research and practice", 20 April 2020, Journal of Technology in Behavioral Science, Volume 5, https://www.doi.org/10.1007/s41347-020-00134-x; Matthew N. Berger, Melody Taba and others, "Social media use and health and well-being of lesbian, gay, bisexual, transgender, and queer youth: Systematic review", 21 September 2022, Journal of Medical Internet Research, Volume 24, Issue 9, https://www.jmir.org/2022/9/e38449
26. For further information on recommendations by Amnesty International and other civil society organizations on how to identify and mitigate systemic risks, see People vs Big Tech, "Briefing: Fixing Recommender Systems: From identification of risk factors to meaningful transparency and mitigation", 25 August 2023, https://www.peoplevsbig.tech/briefing-fixing-recommender-systems-from-identification-of-risk-factors-to-meaningful (accessed 25 August 2023)
27. Algorithm Watch, "AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook", 13 August 2021, https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/; Algorithm Watch, "A guide to the EU's new rules for researcher access to platform data", December 2022, https://www.algorithmwatch.org/en/dsa-data-access-explained/
28. Guardian, "Facebook aware of Instagram's harmful effect on teenage girls, leak reveals", September 2021 (previously cited).
29. The main piece of US legislation governing the liability of online platforms, the Communications Decency Act, was enacted in 1996. The predecessor to the EU's 2022 Digital Security Act (DSA), was the E-Commerce Directive which was adopted in 2000.

governments of Kenya nor the Philippines, two of the focus countries of this report, have yet passed legislation to this effect.[30]

Among the few jurisdictions that have taken action is the European Union (EU), whose Digital Service Act (DSA), adopted in July 2022, became the first major regional level "Big Tech" regulation, aimed at limiting the harmful effects of social media platforms, including by banning intrusive "targeted advertising" towards children.[31] The DSA imposes obligations on the largest online platforms, including TikTok, to assess their systemic risks, including risks to public health and children, to take measures to mitigate these risks and to subject themselves to independent audits to assess their compliance with these obligations (albeit with many questions remaining concerning the precise nature and enforcement of these measures).[32] Under the DSA, TikTok and other very large platforms are also required to provide users with at least one feed which is not based on user profiling. Amnesty International has described this opt-in approach a "missed opportunity" because changing settings to more privacy-respecting options is often cumbersome and users tend to stick to the default setting.[33]

The effectiveness of the DSA is yet to be tested; some of the obligations on very large online platforms such as the submission of risk assessments took effect in August 2023, but the DSA does not fully enter into force until 2024. While the DSA seeks to curtail certain risks associated with the surveillance-based business model of social media companies, for example by banning targeted advertising towards children, it does not ban the business model as a whole and is thus necessarily limited. Nonetheless, the DSA represents the most ambitious piece of platform regulation to date. Even though the DSA is limited in scope and application to platforms' operations in the EU, it nonetheless raises hopes of positive ripple effects beyond the EU as changes implemented inside the EU might be more easily extended or copied over into legislative proposals elsewhere.

In the UK, the Online Safety Bill, passed in September 2023, imposes new obligations on platforms with the intention of creating a safer online environment, with special protections applied to under 18s.[34] The Online Safety Bill was drafted to add to the privacy protections and data minimization requirements contained in the UK Children's Code, which first imposed a duty on online services to only engage in the "profiling" of children if they have "appropriate measures in place to protect the child from any harmful effects (in particular, being fed content that is detrimental to their health or well-being)".[35] The Online Safety Bill criminalizes the posting of material that encourages "serious self-harm".[36] A requirement in the Bill for platforms to stop children from accessing material that "does not meet a criminal threshold but which promotes, encourages or provides instructions for suicide, self-harm or eating disorders"[37] has attracted significant controversy. The Samaritans and a number of

30. Both states do however regulate the collection and use of personal data, including by online services. The companion report to this study, *I feel exposed: Caught in TikTok's Surveillance Web*, includes an analysis of relevant legislation in Kenya. The Philippines passed data privacy regulation in the form of the Data Privacy Act in 2012, which includes principles also established in the EU's General Data Protection Regulation (GDPR), such as purpose limitation and data minimization. See International Association of Privacy Professionals, "GDPR matchup: The Philippines' Data Privacy Act and its Implementing Rules and Regulations", July 2017, https://iapp.org/news/a/gdpr-matchup-the-philippines-data-privacy-act-and-its-implementing-rules-and-regulations/
31. EU, Regulation 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/?toc=OJ%3AL%3A2022%3A277%3ATOC&uri=uriserv%3AOJ.L_.2022.277.01.0001.01.ENG
32. For a detailed analysis of the DSA see *Amnesty International, What the EU's Digital Services Act Means for Human Rights and Harmful Big Tech Business Models*, (Index: POL 30/5830/2022), 7 July 2022, https://www.amnesty.org/en/documents/pol30/5830/2022/en; Euractiv, "Europe enters patchy road to audit online platforms' algorithms", 7 July 2023, https://www.euractiv.com/section/platforms/news/europe-enters-patchy-road-to-audit-online-platforms-algorithms/
33. Alfred Ng, "Default settings for privacy – we need to talk", 21 December 2019, https://www.cnet.com/tech/tech-industry/default-settings-for-privacy-we-need-to-talk/
34. UK Department for Digital, Culture, Media and Sport, Online Safety Bill (as amended in Committee), https://bills.parliament.uk/bills/3137
35. Information Commissioner's Office (ICO), Age-appropriate design code, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/code-standards/ (accessed on 13 July 2023).
36. UK Department for Digital, Culture, Media and Sport, Online Safety Bill (as amended in Committee), (previously cited)
37. UK Department for Digital, Culture, Media and Sport, *A Guide to the Online Safety Bill*, 16 December 2022, https://www.gov.uk/guidance/a-guide-to-the-online-safety-bill

mental health organizations have argued that limiting the rules to children was too narrow, while other groups have warned against incentivizing platforms to unduly restrict users' freedom of expression in an effort to comply with such measures to restrict "legal but harmful" content.[38]

New regulations in the USA remain deadlocked at the federal level.[39] However, several states have passed laws regulating online spaces. Among these is California whose Age-Appropriate Design Code Act will require platforms that "are likely to be accessed by children" to offer children a greater level of privacy through their default settings and prohibit the use of a child's information "in a way that the business knows, or has reason to know, is materially detrimental to the physical health, mental health, or well-being of a child".[40] The Act was however blocked by a federal judge in September 2023 while a lawsuit lodged by a tech industry group proceeds, which argues that the law violates the First Amendment.[41] Similar to the DSA, companies will have to conduct a non-public data protection impact assessment to identify whether a platform's design elements or algorithms could harm children. Although many child rights advocates have welcomed the Act as a step towards a safer online space,[42] privacy and free speech advocates have raised privacy concerns about the possible negative implications of a wider use of age verification measures.[43]

In the meantime, legislation enacted in Utah in March 2023 gives parents unprecedented access to and control over their children's social media accounts, raising concerns about children's privacy. Children's rights campaigners have cast doubt on whether shifting responsibility onto parents will protect children from systemic harms.[44] Arkansas and Louisiana nonetheless followed suit adopting similar pieces of legislation in early 2023.[45] In May 2023, Montana became the first US state to ban TikTok over separate concerns regarding the platform's data security given its Chinese roots, although experts have questioned whether a state-level ban is technically feasible.[46]

China, in turn, has imposed controls on its leading tech companies, including ByteDance, the founder of TikTok and the domestic version, Douyin. Taking effect in March 2022, the Regulations on the Administration of Internet Information Service Recommendation Algorithms combines requirements for transparency, user control and risk assessment obligations – reminiscent of the EU's DSA – with China's repressive control of media, that require algorithmic recommender systems to "present information conforming to mainstream values" and to "prevent or reduce controversies or disputes".[47] In order to ensure that platforms do not "endanger national security or the social public interest", China's Cyberspace Administration has been given extensive access to information about platforms'

38. Guardian, "Online safety bill must protect adults from self-harm content, say charities", 14 October 2022, https://www.theguardian.com/technology/2022/oct/14/online-safety-bill-must-protect-adults-self-harm-content-charities-samaritans
39. Of the multiple proposed federal-level bills relating to platform regulation, only the Kids Online Safety Act (KOSA) looks likely to be passed in the near term. KOSA combines elements of the California Design Code and the EU's DSA, requiring any large platform likely to be accessed by under-17s to publish reports on compliance activities, to implement "measures in its design and operation of products and services to prevent and mitigate" harms and to offer young users ways to opt-out of personalized recommendation-based feeds. See Congress.gov, Text - S.1409 - 118th Congress (2023-2024): Kids Online Safety Act, 2 May 2023, https://www.congress.gov/bill/118th-congress/senate-bill/1409/text
40. Alexander Misakian and Tiffany Young, "California enacts the California Age-Appropriate Design Code Act", 20 September 2022, https://www.foley.com/en/insights/publications/2022/09/california-enacts-age-appropriate-design-code-act
41. The Verge, "Court blocks California's online child safety law", 18 September 2023, https://www.theverge.com/2023/9/18/23879489/california-age-appropriate-design-code-act-blocked-unconstitutional-first-amendment-injunction
42. See for example: Fairplay, "Josh Golin statement regarding California Age Appropriate Design Code", 23 May 2022, https://fairplayforkids.org/josh-golin-ca-aadc/
43. Iapp, "California Age Appropriate Design Code final passage brings mixed reviews", 31 August 2022, https://iapp.org/news/a/california-age-appropriate-design-code-final-passage-brings-mixed-reviews/
44. NPR, "Utah's new social media law means children will need approval from parents", 24 March 2023, https://www.npr.org/2023/03/24/1165764450/utahs-new-social-media-law-means-children-will-need-approval-from-parents
45. CNN, "Arkansas governor signs sweeping bill imposing a minimum age limit for social media usage", 12 April 2023, https://edition.cnn.com/2023/04/12/tech/arkansas-social-media-age-limit/index.html; CNN, "Louisiana lawmakers approve parental consent bill for kids' social media use and other online services", 8 June 2023, https://edition.cnn.com/2023/06/08/tech/louisiana-parental-consent-bill-social-media/index.html
46. NPR, "Montana banned TikTok. Whatever comes next could affect the app's fate in the U.S.", 18 May 2023, https://www.npr.org/2023/05/18/1176940592/montana-ban-tiktok-lawsuit-constitution
47. Friedrich Ebert Stiftung, *China's Regulations on Algorithms: Contexts, impacts and comparisons with the EU*, January 2023, https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf

algorithmic systems, although meetings between the Cyberspace Administration and ByteDance reportedly revealed a lack of technical competency within the administration to understand and interrogate the data.[48] Likely responding to a wider government campaign to limit the time spent by children on online entertainment, Douyin itself imposed a 40-minute limit on under-14s' daily use of the platform in 2021.[49]

Other states such as Brazil, India and Turkey have sought to curtail the power of "Big Tech" platforms with the stated aim of stopping the spread of illegal content and disinformation. However, civil society activists and journalists have raised serious concerns about government overreach and the impact of measures on freedom of expression online.[50]

Taking a global view, further inroads towards responding to the systemic risks covered in this report in human rights-respecting ways urgently need to be made.

3.3 ESCALATING MENTAL HEALTH CHALLENGES AMONG CHILDREN AND YOUNG PEOPLE

Underlying the debate about social media's effect on children and young people's health and well-being is a documented increase in levels of anxiety, depression and self-harm among adolescents in the USA and other English-speaking countries in the 2010s, which further accelerated during the Covid-19 pandemic.[51] Although evidence of similar trends in other parts of the world is limited, partly due to a lack of data, there is nevertheless growing awareness of the level of unaddressed mental health needs globally.[52] The Global Health Data Exchange estimates that roughly one in seven people aged between 10 and 19 years old worldwide experience mental health issues, and the World Health Organization (WHO) has highlighted anxiety and depression as being particularly common.[53] Public health experts believe that the proportion of young people experiencing mental health issues is likely to be higher in Sub-Saharan Africa compared with high-income countries, although a lack of representative data from parts of the region and probable under-reporting due to a lack of affordable services and social stigma hinder efforts to understand and address the issue.[54]

In Kenya and the Philippines, where Amnesty International conducted qualitative research for this report, access to free or affordable mental health services remains very limited and these providers are often based in and around major cities.[55] In the Philippines, some young research participants said that

48. Matt Sheehan and Sharon Du, "What China's Algorithm registry reveals about AI governance", 9 December 2022, https://carnegieendowment.org/2022/12/09/what-china-s-algorithm-registry-reveals-about-ai-governance-pub-88606
49. BBC, "China: Children given daily time limit on Douyin – its version of TikTok", 20 September 2021, https://www.bbc.co.uk/news/technology-58625934
50. Global Voices, "How India's new internet regulations will change social media, online news and video streaming", 16 March 2021, https://globalvoices.org/2021/03/16/how-indias-new-internet-regulations-will-change-social-media-online-news-and-video-streaming; CNN, "Critics fear Turkey's new social media law could hurt freedom of expression. Here's how", 29 July 2020, https://edition.cnn.com/2020/07/29/europe/turkey-social-media-law-intl/index.html; MIT Technology Review, "Brazil's 'fake news' bill won't solve its misinformation problem", 10 September 2020, https://www.technologyreview.com/2020/09/10/1008254/brazil-fake-news-bill-misinformation-opinion
51. For an extensive overview of the available evidence see: Jonathan Haidt, Zach Rausch & Jean Twenge, (ongoing). *Social Media and Mental Health: A Collaborative Review* (previously cited).
52. World Health Organization, *World mental health report: Transforming mental health for all*, June 2022, https://www.who.int/publications/i/item/9789240049338
53. World Health Organization, "Factsheet: Mental health of adolescents" (previously cited).
54. Elsabe Brits, "High mental health burden for Africa's youth", 20 October 2021, https://www.nature.com/articles/d44148-021-00097-y; Ismail Temitayo Gbadamosi, Isaac Tabiri Henneh and others"Depression in Sub-Saharan Africa", 17 March 2022, IBRO Neuroscience Reports, Volume 12, https://www.pubmed.ncbi.nlm.nih.gov/35746974/
55. Finding based on conversations with mental health organizations, psychologists and mental health campaign organizations in Kenya and the Philippines, including e.g. In Touch Community Services, Dr Jerome Cleofas, Dr Marc Reyes between March and May 2023.

although they were in some cases able to receive diagnoses for mental health conditions, they could not afford prescribed medications and/or struggled to find therapists who could provide long-term care.[56]

Although there is greater availability of mental health care in the USA, specialist adolescent psychiatrists and therapists are reportedly struggling to cope with the increased demand for professional help with mood disorders and suicidal thoughts.[57] According to a survey by the Centers for Disease Control and Prevention (CDC), 30% of female US high-school students "seriously considered attempting suicide" in 2021 (up from 19% in 2011), with 13% reporting actual suicide attempts.[58] Emergency departments in the USA reported twice the eating disorder caseload for female adolescents in 2022 compared to pre-pandemic numbers.[59] Throughout CDC's dataset, young women and LGBI[60] teens reported experiencing higher levels of sexual violence, sadness and suicide risk compared to young men and heterosexual teenagers.[61]

All this data underlines the urgent case for taking a closer look at how the large number of young people experiencing mental health issues worldwide engage with social media, and how this particularly vulnerable group of users is affected by social media's engagement-maximizing business strategies.

56. Interview with a 20-year-old woman in Manila, 5 May 2023; interview with a 23-year-old trans man in Manila, 6 May 2023; focus group conversation including statements from a 22-year-old woman in Manila, 6 May 2023.
57. New York Times Magazine, "How do you actually help a suicidal teen?", 17 March 2023, https://www.nytimes.com/2023/05/17/magazine/suicide-teens.html
58. CDC, "CDC report shows concerning increases in sadness and exposure to violence among teen girls and LGBQ+ youth", updated 9 March 2023, https://www.cdc.gov/nchhstp/newsroom/fact-sheets/healthy-youth/sadness-and-violence-among-teen-girls-and-LGBQ-youth-factsheet.html
59. CDC Morbidity and Mortality Weekly Report, "Pediatric emergency department visits associated with mental health conditions before and during the COVID-19 pandemic — United States, January 2019-January 2022", 25 February 2022, Morbidity and Mortality Weekly Report, Volume 21, Issue 8, https://dx.doi.org/10.15585/mmwr.mm7108e2
60. No data on transgender people was included in the survey.
61. CDC, "Pediatric emergency department visits associated with mental health conditions before and during the COVID-19 pandemic — United States, January 2019-January 2022", 25 February 2022, Morbidity and Mortality Weekly Report (previously cited).

4. HUMAN RIGHTS FRAMEWORK

In its 2019 report, Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights, Amnesty International documented how Meta (then called Facebook) and Google (owned by Alphabet) have built a business model predicated on the abuse of their users' right to privacy, turning the collection of intimate personal data of billions of people into a means to generate advertising revenue.[62] Since then, hundreds of millions more users have joined newer social media platforms that are using the same surveillance-based business model to target and exploit the most rapidly growing audience online – children and young people.[63]

Amnesty International's 2023 report, I feel exposed: Caught in TikTok's Surveillance Web, describes how TikTok collects and analyses the data of children and young people in order to know and grow a user base that can be targeted with personalized advertisements.[64] It examines in detail how this surveillance-based business model contravenes the right to privacy, undermining young people's ability to exercise control over their personal information, and violates the right to freedom of thought, in particular the right to keep thoughts and opinions private.

Many of these concerns relate to issues covered in this report, notably how TikTok uses the trove of personal data it collects and the inferences drawn from it to target specific content at users to keep them engaged, as well as how its algorithmic content recommendation system exposes children and young adults with pre-existing mental health concerns to serious risks of harm to their mental health.

The following summarizes the applicable human rights framework related to these concerns.

4.1 THE RIGHT TO PRIVACY IN THE AGE OF SOCIAL MEDIA

"Privacy is vital to children's agency, dignity and safety and for the exercise of their rights", Committee on the Rights of the Child General Comment 25 on Children's Rights in Relation to the Digital Environment.[65]

62. Amnesty International, *Surveillance Giants* (previously cited).
63. International Telecommunications Union (ITU), "Facts and Figures 2022: Latest on global connectivity amid economic downturn", 30 November 2022, https://www.itu.int/hub/2022/11/facts-and-figures-2022-global-connectivity-statistics; Forbes, "What The Rise Of TikTok Says About Generation Z", 7 July 2020, https://www.forbes.com/sites/forbestechcouncil/2020/07/07/what-the-rise-of-tiktok-says-about-generation-z/
64. TikTok allows advertisers to target ads at users as young as 13 in all but the most closely regulated region, the EU, and the two other states (Switzerland and the UK), which TikTok covers in the region's privacy policy. Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web,* 2023 (previously cited).
65. UN Committee on the Rights of the Child, General Comment 25: Children's Rights in Relation to the Digital Environment, 2 March 2021, UN Doc. CRC/C/GC/25, para. 67.

The right to privacy is recognized in numerous human rights instruments but it has likely never been as widely challenged as by the rise of now globally dominant social media. 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.[66] The leading children's rights instrument, the Convention on the Rights of the Child (CRC), echoes that "the child has the right to the protection of the law against such interference or attacks".[67] Recognizing that technological innovation has created many new challenges to the right to privacy since these human rights instruments were created, the UN Human Rights Committee has further clarified that such protection includes regulating "the gathering and holding of personal information on computers, data banks and other devices, whether by public authorities or private individuals or bodies."[68]

The UN Office of the High Commissioner for Human Rights (OHCHR) has likewise elaborated on the meaning of privacy in relation to digital environment:

> **"Privacy can be considered as the presumption that individuals should have an area of autonomous development, interaction and liberty, a 'private sphere' with or without interaction with others, free from State intervention and from excessive unsolicited intervention by other uninvited individuals. In the digital environment, informational privacy, covering information that exists or can be derived about a person and her or his life and the decisions based on that information, is of particular importance."[69]**

The right to privacy encompasses 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 surveillance-based nature of TikTok's business model undermines each of these three elements to such an extent that it has undermined the very essence of privacy.

FREEDOM FROM INTRUSION

The UN High Commissioner for Human Rights has recognized that "even the mere generation and collection of data relating to a person's identity, family or life already affects the right to privacy, as through those steps an individual loses some control over information that could put his or her privacy at risk."[70] The protection of the right to privacy extends not only to the content of communications "but equally to metadata [data about digital files, communications, etc., for example timestamps, location] as, when analysed and aggregated, such data "may give an insight into an individual's behaviour, social relationship, private preference and identity that go beyond even that conveyed by accessing the content of a communication"".[71]

Interference with an individual's right to privacy is only permissible under international human rights law if it is neither arbitrary nor unlawful. Human rights mechanisms have consistently interpreted those words as pointing to the overarching principles of legality, necessity and proportionality.[72] However, the Committee on the Rights of the Child has found that "[d]igital practices, such as automated data processing, profiling, behavioural targeting, mandatory identity verification, information filtering and mass surveillance are becoming routine" in children's lives.[73] As *I feel exposed: Caught in TikTok's*

66. Universal Declaration of Human Rights (UDHR), Article 12 and International Covenant on Civil and Political Rights (ICCPR), Article 17.
67. CRC, Article 16.
68. UN Human Rights Committee (HRC), General Comment 16: The Right to Respect of Privacy, Family, Home and Correspondence, and Protection of Honour and Reputation (Article 17), 8 April 1988, para. 10.
69. UN High Commissioner for Human Rights, Report: *The Right to Privacy in the Digital Age*, 3 August 2018, UN Doc. A/HRC/39/29, para. 5.
70. UN High Commissioner for Human Rights, Report: *The Right to Privacy in the Digital Age, 3 August 2018* (previously cited), para. 7.
71. UN High Commissioner for Human Rights, Report: *The Right to Privacy in the Digital Age, 3 August 2018* (previously cited), para. 6.
72. Office of the High Commissioner for Human Rights (OHCHR), Report: *The Right to Privacy in the Digital Age*, 30 June 2014, UN Doc. A/HRC/27/37, paras. 21-27.
73. Committee on the Rights of the Child, General Comment 25 (previously cited), para. 68.

Surveillance Web explores in greater detail, dominant social media platforms, including TikTok, are instituting forms of mass corporate surveillance that are inherently unnecessary, disproportionate and can never be a permissible interference with the right to privacy.[74]

THE RIGHT TO CONTROL OUR PERSONAL INFORMATION

The second component of privacy provides that people have the right to control their personal information, or the right to "informational self-determination", to be able to decide when and how our personal data can be shared with others ("informational self-determination").[75] This forms the foundation for data protection regulation. The European Court of Human Rights (ECtHR) has recognized that the protection of personal data is of fundamental importance to a person's enjoyment of his or her right to privacy,[76] and that privacy provides for the right to a form of informational self-determination.[77]

THE RIGHT TO A SPACE IN WHICH TO FREELY EXPRESS IDENTITY

Finally, there is a broad consensus that privacy is also fundamental in creating and protecting the space necessary to construct our own identities.[78] The UN Human Rights Committee has defined privacy as "a sphere of a person's life in which he or she can freely express his or her identity."[79] This reflects an understanding that one's sense of identity is both socially constructed and dynamic: that individuals explore and display different sides of themselves in different contexts, whether it is with friends, at school, at work or in public, and these identities are constantly shifting and adapting.

Social media platforms such as Instagram and TikTok exert huge influence on young people as dominant spaces of identity exploration and representation. To participate in this space, children and young people willingly share, but are also nudged to publish personal information through posts, stories, reels and comments as well as to divulge personal preferences and interests through 'likes' or by sharing other people's content. And that only covers what other users of these platforms can see. The companies in question gain much wider insights into the private lives of young people who engage with their platforms through data on what children and young people watch or engage with, how often and for how long, their location data, data about the times they log on, the devices they use and the purchases they make, to name but a few.[80]

While research shows that younger teenagers may not necessarily yet be fully aware of the corporate surveillance they are subjected to, they are usually keenly aware of the social pressures inherent in platforms that rely on users to see others and be seen and the boundaries this creates within which young people feel able to explore their identity.[81] Amnesty International's interviews with young people offered extensive insights into the complex ways in which many of them control who in their social environment gets access to what kinds of posts, often through multiple accounts on one platform with various levels of visibility and public representation or anonymity.[82]

---

74. Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web* (previously cited).
75. The term "informational self-determination" was first used in the context of a German Constitutional Court ruling relating to personal information collected during the 1983 census. The Court ruled that it is the authority of the individual to decide when and within what restrictions information about their private life should be communicated to others. BVerfG, Urteil des Ersten Senats vom 15. Dezember 1983 – 1 BvR – Rn. 1-215 https://www.bverfg.de/e/rs19831215_1bvr020983.html
76. ECtHR, S *and Marper v. UK*, Applications 30562/04 and 30566/04, Grand Chamber judgement, 4 December 2008, https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-90051%22]}
77. ECtHR, *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland*, Application 931/13, Grand Chamber judgement, 27 June 2017, https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-175121%22]}, para.137.
78. See for example Philip Agre and Marc Rotenburg (editors), *Technology and Privacy: The New Landscape*, 1998.
79. HRC, Decision, *Coeriel and Aurik v. the Netherlands, adopted on 31 December 1994*, UN Doc. CCPR/C/52/D/453/1991, para. 10.2.
80. Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, 7 November 2023 (previously cited).
81. Mariya Stoilova, Sonia Livingstone and others, "Digital by default: children's capacity to understand and manage online data and privacy", 3 November 2020, Media and Communication, Volume 8, Issue 4, https://eprints.lse.ac.uk/107114/; Kathryn C. Montgomery, Jeff Chester and others, "Children's privacy in the big data era: Research opportunities", 1 November 2017, Pediatrics, Volume 140, Issue Supplement 2, https://doi.org/10.1542/peds.2016-17580
82. Many young adults in the Philippines in particular spoke about the distinct uses of "OG" (original) public accounts and their "dump" or "rant" accounts (accounts with a smaller number of trusted followers, used to give a more "unfiltered", honest version of the young people's lives, feelings and opinions; often on, but not limited to, Instagram.

4.2 THE RIGHT TO FREEDOM OF THOUGHT

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. Unlike freedom of expression, the speaking of one's mind, which can be limited by law when necessary and proportionate and in order to achieve a legitimate aim,[83] – for example to protect the dignity and rights of others – the thoughts, beliefs and opinions inside our head are entirely our own, not to be drawn out against our will or manipulated. The CRC protects children's freedom of thought in Article 14, which requires states parties to "respect the right of the child to freedom of thought, conscience and religion."

Intimately connected to this, the right to freedom of opinion, which is also an absolute right, is protected by Article 19(1) of the ICCPR and holds that "[e]veryone shall have the right to hold opinions without interference".

The rights to freedom of thought and freedom of opinion comprise three elements:

- The right to keep your thoughts and opinions private;
- The right not to have your thoughts and opinions manipulated; and
- The right not to be penalized for your thoughts and opinions.[84]

When the ICCPR and other human rights treaties were written, no one could have foreseen a challenge to the right to freedom of thought as large and pervasive as the one posed by modern technology today. The surveillance-based business model of social media companies, which involves the massive collection of intimate personal data on all its users (and the collection of certain pieces of non-intimate data even from people who have not registered for an account[85]) poses a direct threat to these rights in a way never seen before. After collecting users' personal data, social media companies use it to analyse people, aggregate them into groups and make predictions about their interests, characteristics and ultimately their behaviour – primarily so that they can use these insights to generate advertising revenue.[86]

I feel exposed: Caught in TikTok's Surveillance Web highlights how companies' profit-driven aim to collect and aggregate as much intimate personal data on social media users as possible – drawing inferences from posts, likes, reactions, follows, connections, watch time, comments, location, device data and more to piece together a person's interests and sexual and political orientation, to name but a few characteristics – violates the first element of this right: the right to keep your thoughts and opinions private.[87]

This report will examine more closely how TikTok uses this trove of personal data and the inferences drawn from it to target specific content at users to keep them engaged, thereby potentially infringing upon the right not to have your thoughts and opinions manipulated. The Committee on the Rights of the Child explores this argument when stating that it:

> **"[E]ncourages States parties to introduce or update data protection regulation and design standards that identify, define and prohibit practices that manipulate or interfere with children's right to freedom of thought and belief in the digital environment, for example by**

83. ICCPR, Article 19; CRC, Article 13.
84. Vermeulen, "Freedom of thought, conscience and religion (article 9)," in *Theory and Practice of the European Convention on Human Rights, 4th Edn.*, eds P. van Dijk, F. van Hoof, A. van Rijn and L. Zwaak (Cambridge: Intersentia Press), 751–772 cited in McCarth-Jones, "The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century", Front. Artif. Intell., 26 September 2019, Sec. Technology and Law, https://doi.org/10.3389/frai.2019.00019; Evelyn Aswad, "Losing the freedom to be human", 8 July 2020, Columbia Human Rights Law Review, Volume 52, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635701
85. Vice, "TikTok is watching you – even if you don't have an account", 21 January 2021, https://www.vice.com/en/article/jgqbmk/tiktok-data-collection; Wired, "All the ways TikTok tracks you and how to stop it", 23 October 2021, https://www.wired.co.uk/article/tiktok-data-privacy
86. Amnesty International, *Surveillance Giants* (previously cited).
87. Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, 7 November 2023 (previously cited).

emotional analytics or inference. Automated systems may be used to make inferences about a child's inner state. They should ensure that automated systems or information filtering systems are not used to affect or influence children's behaviour or emotions or to limit their opportunities or development".[88]

The Council of Europe (CoE) has also warned that "[f]ine grained, 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".[89]

Amnesty International's findings, detailed in the following sections, not only show that TikTok employs manipulative and addictive design practices, but also that the platform's algorithmic content recommendation systems, credited with enabling the rapid global rise of the platform,[90] exposes children and young adults with pre-existing mental health concerns to serious risks of harm to their mental health.

ALGORITHMIC RECOMMENDER SYSTEMS

Systems enabling the curation and personalization of social media feeds are often simply referred to as a single "algorithm". In reality, TikTok and other social media platforms rely on a multitude of automated tools to collect, combine, sort and infer data and to make decisions about which content to display to a particular user at a particular time. The complexity of these interacting tools – and researchers' severely limited access to the relevant systems and data – poses particular challenges in 'auditing' (that is, examining) recommender systems such as the one which powers TikTok's 'For You' feed.

4.3 THE 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. Several human rights instruments recognize the right to health.[91] The Committee on Economic, Social and Cultural Rights has stated that sufficient health care facilities and services must be available, within reach and affordable to all sections of the population.[92]

Yet mental health care remains widely under-resourced and neglected, leading the former UN Special Rapporteur on the Right to Everyone to the Enjoyment the Highest Attainable Standard of Physical and Mental Health to remark in his 2020 report to the UN Human Rights Council:

"There is no health without mental health. The rich links between mind, body and the environment have been well-documented for decades. As the third decade of the millennium begins, nowhere in the world has achieved parity between mental and physical health and this remains a significant human development challenge. An important message within that collective failure is that without addressing human rights seriously, any investment in mental health will not be effective."[93]

88. Committee on the Rights of the Child, General Comment 25 (previously cited), para. 62.
89. 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
90. Alex Hern, "How TikTok's algorithm made it a success: 'It pushes the boundaries'", 24 October 2022, https://www.theguardian.com/technology/2022/oct/23/tiktok-rise-algorithm-popularity
91. 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; the African Charter on Human and People's Rights, Article 16; the European Social Charter (Revised).
92. Committee on Economic, Social and Cultural Rights, General Comment 14, 11 August 2000, UN Doc. E/C.12/2000/4.
93. UN Special Rapporteur on the Right of Everyone to the Enjoyment of the Highest Attainable Standard of Physical and Mental Health (UN Special Rapporteur on the right to health), Report, 15 April 2020, UN Doc. A/HRC/44/48, para. 1.

He also warned in June 2020 that the Covid-19 pandemic had aggravated the "historical neglect of dignified mental health care" and that "the combination of school closures and lockdown conditions have a particular impact on children's stress, anxiety and mental health issues. This is especially worrying given the overall lack of recognition and awareness of the nature and scale of mental health problems among adolescents in many countries."[94]

Digital technologies have undoubtedly played a positive role in reducing access barriers to health care information and services, for example by enabling therapists to speak to patients remotely during lockdowns, by keeping young people connected to support and by allowing marginalized groups such as LGBTI adolescents to explore their identities and express themselves.[95] But there is a growing recognition that in the absence of effective state regulation, social media platforms in particular have also exposed children and young people to systemic risks such as those described in this report.[96] As the Committee on the Rights of the Child noted in General Comment 25:

> **"The digital environment can include gender-stereotyped, discriminatory, racist, violent, pornographic and exploitative information, as well as false narratives, misinformation and disinformation and *information encouraging children to engage* in unlawful or *harmful activities*. Such information may come from multiple sources, including other users, commercial content creators, sexual offenders or armed groups designated as terrorist or violent extremist. *States parties should protect children from harmful and untrustworthy content and ensure that relevant businesses and other providers of digital content develop and implement guidelines to enable children to safely access diverse content*, recognizing children's rights to information and freedom of expression, while protecting them from such harmful material in accordance with their rights and evolving capacities."[97]** [Emphasis added]

Focusing more specifically on risks to children and young people's health, the US Surgeon General warned in May 2023:

> **"Nearly every teenager in America uses social media, and yet we do not have enough evidence to conclude that it is sufficiently safe for them. Our children have become unknowing participants in a decades-long experiment… 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."[98]**

This urgent action requires both regulatory efforts from governments in line with their obligations under international human rights law to progressively fulfil people's right to health and take steps to protect people's human rights in the context of corporate activities, as well as immediate steps by social media companies to respect human rights and to fulfil their human rights due diligence responsibilities (discussed in greater detail in the following section).

---

94. OHCHR, "COVID-19 has exacerbated the historical neglect of dignified mental health care, especially for those in institutions: UN expert," 23 June 2020, https://www.ohchr.org/en/news/2020/06/covid-19-has-exacerbated-historical-neglect-dignified-mental-health-care-especially
95. Benjamin Hanckel and Shiva Chandra, "How young LGBTQIA+ people used social media to thrive during COVID lockdowns", The Conversation, 15 March 2021, https://theconversation.com/how-young-lgbtqia-people-used-social-media-to-thrive-during-covid-lockdowns-156130
96. 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; Committee on the Rights of the Child, General Comment 25, 2 March 2021, UN Doc. CRC/C/GC/25
97. Committee on the Rights of the Child, General Comment 25, 2 March 2021, UN Doc. CRC/C/GC/25, para. 54.
98. US Surgeon General, *Social Media and Youth Mental Health*, May 2023, https://www.hhs.gov/surgeongeneral/priorities/youth-mental-health/social-media/index.html

4.4 BEST INTERESTS OF THE CHILD AND THE RIGHT TO BE HEARD

An essential principle enshrined in the CRC is that of the 'best interests of the child'. Article 3(1) establishes that: "In all actions concerning children, whether undertaken by public or private social welfare institutions, courts of law, administrative authorities or legislative bodies, the best interests of the child shall be a primary consideration."[99]

Article 12 of the CRC states that children have the right to express their views "freely in all matters affecting the child, the views of the child being given due weight in accordance with the age and maturity of the child."[100] The evolving capacities of the child must be taken into consideration along with the best interests principle and children's right to have their views heard.

Applying these principles to the digital environment, the Committee on the Rights of the Child recommends that States ensure that, "in all actions regarding the provision, regulation, design, management and use of the digital environment, the best interests of every child is a primary consideration",[101] and that "digital service providers offer services that are appropriate for children's evolving capacities."[102]

4.5 CORPORATE RESPONSIBILITY TO RESPECT HUMAN RIGHTS

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).[103] This corporate responsibility to respect human rights is independent of a state's own human rights obligations and exists over and above compliance with national laws and regulations protecting human rights.[104]

It requires companies to avoid causing or contributing to human rights abuses through their own business activities and to address impacts in which they are involved, including by remediating any actual abuses. 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 how they address their 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.[105]

---

99. CRC, Article 3(1).
100. CRC, Article 12.
101. Committee on the Rights of the Child, General Comment 25 (previously cited), para. 12.
102. Committee on the Rights of the Child, General Comment 25 (previously cited), para. 20.
103. *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.
104. OHCHR, *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, 2011*, UN Doc. HR/PUB/11/04, ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf, Principle 11 including Commentary.
105. UN Guiding Principles (previously cited), Principle 19 and Commentary.

States in turn have an obligation to respect and protect human rights in the context of corporate activities. They must "protect against human rights abuse within their territory and/or jurisdiction by third parties, including business enterprises. This requires taking appropriate steps to prevent, investigate, punish and redress such abuse through effective policies, legislation, regulations and adjudication."[106]

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"[107] 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". The Committee has emphasized that States are "responsible for infringements of children's rights caused or contributed to by business enterprises where it has failed to undertake necessary, appropriate and reasonable measures to prevent and remedy such infringements or otherwise collaborated with or tolerated the infringements."[108]

Key to effective due diligence is transparency and publicly accounting for how a company has identified, prevented or mitigated potential or actual adverse impacts on human rights. The UN Guiding Principles state that companies "need to know and show that they respect human rights",[109] where "showing involves communication, providing a measure of transparency and accountability to individuals or groups who may be impacted and to other relevant stakeholders."[110] The Committee on the Rights of the Child has similarly stressed that States should encourage, and where appropriate require, companies to publish details of what measures they have taken to address impacts on children's human rights caused by their activities.[111]

This transparency requirement stands in clear contrast to the realities of the ever greater influence of privately developed algorithms, empowered to determine which posts are amplified and subsequently consumed by masses of people on social media or which kinds of posts violate a platform's guidelines. Social media platforms have long sought to protect what they perceive to be the business secrets behind their business models,[112] and have sought to limit independent researchers' ability to track the outcomes of these so-called 'black box' algorithms.[113] In light of the growing acknowledgement of the real-world harms associated with platforms that favour and amplify extreme content,[114] various UN and other international bodies have sought to develop responses to this significant regulatory gap.

In 2021, the OHCHR set out recommendations to States and companies for addressing human rights risks related to the use of artificial intelligence (AI), which plays an ever-greater role in refining content recommender systems, including:

1. "Systematically conduct human rights due diligence throughout the life cycle of the AI systems they design, develop, deploy, sell, obtain or operate. A key element of their human rights due diligence should be regular, comprehensive human rights impact assessments.
2. Dramatically increase the transparency of their use of AI, including by adequately informing the public and affected individuals and enabling independent and external auditing of automated

106. UN Guiding Principles (previously cited), Principle 1.
107. Committee on the Rights of the Child, General Comment 16, 17 April 2013, UN Doc. CRC/C/GC/16, para. 62.
108. Committee on the Rights of the Child, General Comment 16 (previously cited), para. 28.
109. UN Guiding Principles (previously cited), Commentary to Principle 15.
110. UN Guiding Principles (previously cited), Commentary to Principle 21.
111. Committee on the Rights of the Child, General Comment 16 (previously cited), para. 65.
112. Alex Hern, "How TikTok's algorithm made it a success: 'It pushes the boundaries'" (previously cited).
113. Mozilla Foundation, Amnesty International and others, "Response to the European Commission's call for evidence for a Delegated Regulation on data access provided for in the Digital Services Act", (previously cited); Frederick Mostert, "Social media platforms must abandon algorithmic secrecy", 17 June 2021, https://www.ft.com/content/39d69f80-5266-4e22-965f-efbc19d2e776
114. Amnesty International, *Myanmar: The Social Atrocity – Meta and the Right to Remedy for the Rohingya* (previously cited); *Surveillance Giants* (previously cited).

systems. The more likely and serious the potential or actual human rights impacts linked to the use of AI are, the more transparency is needed.

3. Ensure participation of all relevant stakeholders in decisions on the development, deployment and use of AI, in particular affected individuals and groups.

4. Advance the explainability of AI-based decisions, including by funding and conducting research towards that goal."[115]

Following calls by the CoE[116] and other bodies for state regulation of the use of algorithmic systems, initiatives such as the EU's DSA and the draft Artificial Intelligence Act[117] are seeking to increase transparency, implement oversight mechanisms and prevent potentially abusive uses of AI.

Chapter 7 of this report sets out how TikTok is failing to fulfil its responsibility to respect human rights, failing to mitigate risks to the right to privacy, freedom of thought and the right to health, through its hyper-personalized feed and other engagement maximizing design elements that sustain TikTok's surveillance-based business model.

115. OHCHR, The right to privacy in the digital age, 15 September 2021, UN Doc. A/HRC/48/31.
116. CoE, "Recommendation CM/Rec (2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems", 8 April 2020.
117. The EU's draft AI Act seeks to regulate AI systems more broadly, whereas the Digital Services Act targets risks associated with the algorithmic systems of large social media platforms. Amnesty International, "EU: AI Act at risk as European Parliament may legitimize abusive technologies", 13 June 2023, https://www.amnesty.org/en/latest/news/2023/06/eu-ai-act-at-risk-as-european-parliament-may-legitimize-abusive-technologies/

5. ADDICTIVE BY DESIGN

Contrary to other mass media, social media companies personalize their users' experience to maximize the amount of time users spend on the platform and to drive engagement with user content and – most critically for the advancement of their business – advertisements. Amnesty International's report 2019 *Surveillance Giants* documented how Facebook (now called Meta) and Google (owned by Alphabet) have optimized the use of big data analytics and addictive design so effectively as to dominate the world's advertising market.[118] In 2021, the two companies shared between them an estimated 43% of all global ad spending online and offline.[119] TikTok entered this market determined to rapidly attract hundreds of millions of young users and usurp other media's share of users' time and attention.[120] Its marketing materials underline that the company sees itself not just in competition with other social media but also aims to take up space occupied by other daily activities too, advertising that "41% of Gen Z TikTok users globally say they spent less time listening to podcasts since starting to use TikTok".[121]

The effects of extensive TikTok use that children and young people interviewed by Amnesty International shared with researchers is not limited to spending less time on other media. Several young people we interviewed felt that their TikTok use resulted in them failing to submit school and university assignments, spending less time with friends offline and, most detrimental to their health, scrolling through their feeds late at night instead of catching enough sleep.[122]

For example, "Mary", who attends an all-girls school in Western Kenya said:

> **"When I wake up, the first thing I do is look at my phone, all I can think about is my phone".[123]**

"Amy", another adolescent in the focus group, said:

> **"If I use TikTok, my intention is usually to watch it for 5 minutes, I try to watch two (videos) but it lets me see and continue watching the whole night. I want to spend 5 minutes but the videos are so interesting I end up [watching for] at least six hours".**

Asked about which social media app was the most difficult to stop using, there was almost unanimous consensus amongst the 29 participants in the group that they felt this was TikTok.[124] "Joyce", a young woman in the Philippines, explained:

---

118. Amnesty International, *Surveillance Giants* (previously cited).
119. Ebiquity, "Google, Meta and Amazon are on track to absorb more than 50% of all ad money in 2022", 4 February 2022, https://ebiquity.com/news-insights/press/google-meta-and-amazon-are-on-track-to-absorb-more-than-50-of-all-ad-money-in-2022/
120. Yahoo Finance, "Forget Facebook, Snapchat, Twitter: TikTok Is The Breakout COVID-19 Social Media Platform", 24 April 2020, https://finance.yahoo.com/news/forget-facebook-snapchat-twitter-tiktok-210717993.html; Business Insider, "Nearly half of Gen Z is using TikTok and Instagram for search instead of Google, according to Google's own data", 13 July 2022, https://www.businessinsider.com/nearly-half-genz-use-tiktok-instagram-over-google-search-2022-7; TikTok, "TikTok Marketing Science Global Time Well Spent Study (Global Results) conducted by Kantar March 2021", https://tiktok.com/business/en-US/insights (accessed 1 June 2023).
121. TikTok, "TikTok Marketing Science Global Time Well Spent Study (Global Results) conducted by Kantar March 2021" (previously cited).
122. Mentioned in separate focus group discussions in Kisumu, Kenya, 13 and 14 March 2023 as well as FGDs in schools in Mombasa, on 15, 16, 17 March 2023.
123. Focus group discussion, Kisumu, Kenya, 14 March 2023.
124. Focus group discussion, Kisumu, Kenya, 14 March 2023.

"I deleted it [TikTok] for a while but that was because I was very addicted to it… I would spend so many hours on TikTok just scrolling through videos and it's because you can't help but wonder what goes up next when you scroll down".[125]

These and other testimonies from children and young people were corroborated by specialist adolescent psychologists consulted as part of the research, who said that they had observed a trend of TikTok aggravating existing issues with addictive patterns in children's and young adults' use of social media. "I'll hear a lot of them telling me they need a break from social media, they're getting addicted to social media. They just want to sit and scroll the whole day", said Zeyna Awan, a clinical psychologist in Nairobi.[126] Kevin Gachee, another psychologist in Nairobi, observed a generational change in that young people he works with in "Gen Z" appear to often struggle with an unhealthy excessive use of TikTok, whereas adults from the previous generation would be more likely to raise issues around unhealthy social comparison reportedly prompted by their use of Instagram.[127]

Available evidence from market research is skewed towards data on children's social media use in the USA and Europe. A 2021 study found TikTok to be the platform on which children spend the most time across the USA, UK and Spain, with children in the USA spending an average of 99 minutes per day on TikTok alone.[128] A recent representative survey of teenagers in the USA found that 45% of girl users of TikTok "say they feel 'addicted' to the platform or use it more than intended at least weekly."[129]

Our qualitative research in Kenya and the Philippines suggests similar trends. Out of all social media platforms, TikTok specifically was described by young research participants as "addictive" and "time-wasting". At least 25 participants in 10 out of 11 focus group discussions in Kenya made such comments and many more showed their agreement through nods. Young TikTok users described patterns of extreme night-time use[130] and a sizeable minority of school-age child research participants in Kenya reported that, because they share phones with their parents or older siblings they only have access to phones in the late evening or at night and therefore often use social media without parental supervision late into the night or in the early morning hours, with the result that they miss out on sleep before a school day.[131] This finding may point to additional risks to the health of economically disadvantaged children and adolescents.

The emerging literature on social media use and teen mental health further suggests a pattern of children with mental health concerns being more susceptible to the excessive use of social media. A 2021 report by the UK's National Health Service (NHS) found that, "children aged 11 to 16 years with a probable mental disorder were particularly likely to spend more time on social media than they intended; almost two-thirds reported this (63.8%) compared with less than half of those unlikely to have a mental disorder (45.5%)."[132] A systematic review of studies on the use and impact of social media by adolescents with pre-existing mental health issues also found that adolescents with a diagnosis of depression and those with "self-reported symptoms of anxiety, depression, self-injurious behaviour" are more likely to experience issues with regulating their internet use, including social media use.[133]

---

125. Online interview with a Manila-based participant, 22 May 2023.
126. Expert interview, conducted online, 9 March 2023.
127. Expert interview conducted in person in Nairobi, 9 March 2023.
128. Qustodio, *Annual Data Report 2021: Living and Learning in a Digital World*, Chapter 2, Social Media, https://www.qustodio.com/en/social-media-qustodio-annual-data-report-2021 (accessed on 20 July 2023).
129. Jacqueline Nesi, Supreet Mann & Michael B. Robb, *Teens and mental health: How girls really feel about social media*, 2023, https://www.commonsensemedia.org/sites/default/files/research/report/how-girls-really-feel-about-social-media-researchreport_web_final_2.pdf
130. FGDs conducted in Kisumu, Kenya, 13 and 14 March 2023, Mombasa, Kenya, 16 and 17 March, Machakos, Kenya, 22 March 2023.
131. At least nine young Kenyan participants mentioned losing out on sleep because of TikTok, including during nights in school terms, either amidst discussions about when and how they accessed social media or prompted to reflect on which social media platform they spent the most time on. The number of children who share phones varied greatly between groups from less than 10% to close to 50% of the students in a particular group.
132. NHS, *Mental Health of Children and Young People in England 2021 - Wave 2 follow up to 2017 survey,* 30 September 2021, https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2021-follow-up-to-the-2017-survey
133. Katarzyna Kostyrka-Allchorne, Mariya Stoilova and others, "Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and nonsuicidal self-injury conditions – a systematic review", February 2023, Child and Adolescent Mental Health, Volume 28, issue 1, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10108198/

5.1 TIKTOK'S ENGAGEMENT STRATEGIES

Individuals' abilities to regulate their social media use and contextual factors may vary, but like other social media platforms, TikTok has made design choices intended to maximize users' time spent on the platform.[134] In addition to embedding design elements that other social media platforms have successfully employed to keep users' eyes locked on the screen[135] – including the ability to 'like' posts, frequent notifications and a never-ending feed of content – TikTok stands out due to its 'For You' feed, a highly personalized and infinitely scrollable page of algorithmically recommended content, selected to reflect what the system has inferred to be the user's interests within minutes of first using the feed.[136]

Whereas social media platforms previously relied on users' actively declared social connections to predict interests, TikTok's content recommendations largely rely on analysing which videos a user engages with from among an initial selection of popular videos. This is then used to progressively narrow down the user's interests by recommending videos, which the system associates with the previously engaged with content.[137] The result is a much more personalized feed of short video clips, predicted to elicit the greatest entertainment value for the individual user.[138]

By combining these elements, TikTok taps into what psychologists describe as "the intermittent reward pattern of winning or losing on a slot machine... capitaliz[ing] on classical conditioning and reward-based learning processes to facilitate the formation of habit loops and encourage addictive use".[139] In short, TikTok has maximized the addictive qualities of different design choices, which Facebook's founding president warned back in 2017, could have unintended negative consequences, especially for children's health and development.[140] Young interview participants who reported compulsive TikTok use patterns often described the resulting behaviour as "mindless scrolling"[141], where users "don't really connect or feel anything about the post because you really don't have any time to process them when you're just scrolling."[142]

Adolescent research participants also commented on the perceived role of the 'like' function, common to TikTok and other social media platforms, in exacerbating insecurities and keeping their eyes fixed on the screen.[143] They explained that they felt compelled to keep checking their account to monitor how their posts fare in comparison with other's posts. School-age children in rural Kenya even reported using photo and video editing tools to create false posts about experiences their families could not financially afford to gain followers.[144]

Older research participants, including university students, referred to the perceived role of their social media profiles in defining their social status and self-worth: "When you're in college as a fresh graduate, you are your social media account", said Nikki, aged 24, from Manila. Victor, a young

134. Sophia Petrillo, "What makes TikTok so addictive?: An analysis of the mechanisms underlying the world's latest social media craze", 13 December 2021, Brown Undergraduate Journal of Public Health, Issue 2021-2022, https://sites.brown.edu/publichealthjournal/2021/12/13/tiktok/
135. ABC News, "Book excerpt: Jaron Lanier's 'Ten Arguments for Deleting Your Social Media Accounts Right Now'", 19 June 2018, https://abcnews.go.com/Technology/book-excerpt-jaron-laniers-ten-arguments-deleting-social/story?id=56009512; BBC, "Facebook founding president sounds alarm", 9 November 2017, https://www.bbc.co.uk/news/technology-41936791
136. TikTok, "How TikTok recommends videos #ForYou", 18 June 2020, https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you
137. Alex Hern, "How TikTok's algorithm made it a success: 'It pushes the boundaries'" (previously cited).
138. Catherine Wang, "Why TikTok made its user so obsessive? The AI Algorithm that got you hooked", 7 June 2020, https://towardsdatascience.com/why-tiktok-made-its-user-so-obsessive-the-ai-algorithm-that-got-you-hooked-7895bb1ab423
139. Sophia Petrillo, "What makes TikTok so addictive?: An analysis of the mechanisms underlying the world's latest social media craze", December 2021, Brown Undergraduate Journal of Public Health, Issue 2021-2022, (previously cited).
140. Facebook's founding president, Sean Parker, stated in 2017: "[Facebook] probably interferes with productivity in weird ways. God only knows what it's doing to our children's brains." See BBC, "Facebook founding president sounds alarm" (previously cited).
141. Manuel (pseudonym), 23, university student based in Luzon, remote interview on 4 May 2023.
142. Nikki (pseudonym), 24, Manila, interviewed on 8 May 2023.
143. Focus group discussions with secondary school pupils in Kisumu County, 14 March 2023, Mombasa, 17 March, and Machakos County, 22 March 2023.
144. Teenage focus group discussion participants at a secondary school in Machakos County, 22 March 2023.

adult focus group participant in Mombasa echoed the idea that social media requires curated self-representation, saying "the me you see on social media is not the me sitting here today".[145]

Numerous psychological studies and – as the "Facebook Papers" revealed – the company's own research[146] have documented the toxic effect of social media driven social comparison on the mental health of adolescents.[147] Instagram ran a two-year experiment in which it hid the number of likes a post received from all users except the content originator. It eventually discarded the idea, opting instead to allow users to choose for themselves whether likes would be shown, explaining that the experiment had received a mixed response.[148] The value of de-emphasizing popularity and social comparison in boosting young people's self-worth and healthier forms of engagement is however clear to experts such as Nairobi-based psychologist Kevin Gachee: "The way they design these apps, they're designed to keep you engaged. I commended Instagram when they [temporarily] stopped showing the number of likes because it did help significantly in terms of people saying, 'I don't feel the pressure to keep on posting all the time because, you know, my friend got 100 likes and I'd like to have 100 likes'."[149]

More worryingly still in terms of mental health impact is the effect of social media's push towards affirmation-seeking behaviour on young people who self-harm. Manila-based psychologist Dr. Marc Reyes explained his concerns to Amnesty International: "Most people who go through mental health issues feel they're invisible, but on their social media they're seen. So the likes become very detrimental; a lot of people are liking my post [about self-harm], so it's ok", adding that "if you post, you share, reshare [such content], it might trigger suicide contagion or self-harm contagion."[150]

Despite such concerns, TikTok – as well as Instagram – continue to show 'likes' as the default setting.

5.2 HOW ADDICTIVE SOCIAL MEDIA DESIGN CAN AFFECT YOUNG PEOPLE'S HEALTH

Digital technologies have undeniably played a positive role in reducing access barriers to health care information and services, for example by enabling therapists to speak to patients remotely during Covid-19 lockdowns, by keeping young people connected to friends and support networks, and by enabling marginalized groups such as LGBTI adolescents to explore their identities and express themselves.[151] But there is also growing recognition that, in the absence of effective state regulation, social media platforms also expose children and young people to systemic risks.[152]

Among the responses to questions in the scoping survey carried out for this research, there was praise for the diversity of ideas and the opportunities that social media offers, but many respondents raised concerns about the impact of what they described as "addictive" social media on their mental health. Young people reported feeling "anxious" and "self-conscious" about "unrealistic [body] images" viewed in their feeds. A female respondent from Kenya said that she believed the algorithmic content recommendations of social media platforms (referred to by the survey participant as "the algorithm"), "picked up" on her existing mental health issues. Others said that they had seen posts, which they

145. Focus group discussion with university students and young activists in Mombasa, 17 March 2023.
146. Wall Street Journal, "Facebook knows Instagram is toxic for teen girls company documents show" (previously cited).
147. Elia Abi-Jaoude, Karline Treurnicht Naylor and others, "Smartphones, social media use and youth mental health", 10 February 2020, Canadian Medical Association Journal, Volume 192, Issue 6, https://www.cmaj.ca/content/192/6/E136; For a summary of further research on social media-driven social comparison and its impact on mental health, see US Surgeon General, *Social Media and Youth Mental Health* (previously cited).
148. Instagram, "Giving people more control on Instagram and Facebook", 26 May 2021, https://about.instagram.com/blog/announcements/giving-people-more-control
149. Interview conducted in Nairobi on 9 March 2023.
150. Remote interview conducted on 4 May 2023.
151. Benjamin Hanckel and Shiva Chandra, "How young LGBTQIA+ people used social media to thrive during COVID lockdowns", The Conversation, 15 March 2021, https://theconversation.com/how-young-lgbtqia-people-used-social-media-to-thrive-during-covid-lockdowns-156130
152. See, for example, UN Special Rapporteur on the right to health, *Digital Innovation, Technologies and the Right to Health*, 21 April 2023, UN Doc. A/HRC/53/65; Committee on the Rights of the Child, General Comment 25 (previously cited).

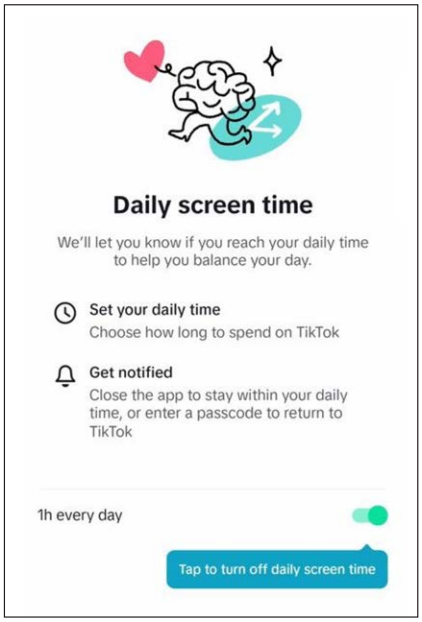
believed exacerbated their self-reported anorexia, described as "triggering", having "struggled" with self-harm or left them feeling "scared".[153]

In a 2023 advisory on social media's impact on teen mental health, the US Surgeon General issued clear words of warning about the effects of addictive social media design:

> **"Excessive and problematic social media use, such as compulsive or uncontrollable use, has been linked to sleep problems, attention problems, and feelings of exclusion among adolescents… Small studies have shown that people with frequent and problematic social media use can experience changes in brain structure similar to changes seen in individuals with substance use or gambling addictions".[154]**

The advisory went on to note that poor sleep does not just affect concentration and performance in school, it has also "been linked to altered neurological development in adolescent brains, depressive symptoms, and suicidal thoughts and behaviors".

The research into how excessive social media use may alter brain structures is still in its early stages but small studies point to some similarities in observed changes to the brain between substance addictions (e.g. cocaine) and addictive social media use.[155] Both trigger the release of high levels of dopamine, leading to a temporary emotional "high", making the user in turn feel low once they stop using the platform.[156] These findings point to a need for government regulation, in line with the regulation of other addictive substances, in order to fulfil children and young people's right to health.



A single click takes young users from the daily screen time limit notification to a page which lets them permanently disable the feature.

TikTok should clearly be aware of the platform's potential for harm to children and young adults' mental and physical health as a result of compulsive or harmful patterns of use. Based on the research evidence outlined above, TikTok risks potentially exposing children and young people to sleep and attention problems and may contribute to changes in brain structure similar to those observed in people experiencing drug addiction.

And yet its response has not gone far enough. In March 2023, TikTok announced the introduction of a new screen-time management tool, which requires under-18s to actively extend their time on the app once they have reached a 60-minute daily limit.[157] The effectiveness of the tool is yet to be independently assessed, but given that it shifts the responsibility to set limits on the use of TikTok onto teenagers who in large numbers describe themselves as "addicted", it is likely to be limited.[158] TikTok's prompt is only a suggestion to observe the time limit; a prompt that can be easily dismissed and does not include a health warning.

153. Responses to Amnesty International online questionnaire, October-November 2022.
154. US Surgeon General, *Social Media and Youth Mental Health* (previously cited).
155. 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
156. Scope, "Addictive potential of social media, explained", 29 October 2021, https://scopeblog.stanford.edu/2021/10/29/addictive-potential-of-social-media-explained/
157. TikTok, "New features for teens and families on TikTok", 1 March 2023, https://newsroom.tiktok.com/en-us/new-features-for-teens-and-families-on-tiktok-us
158. US Surgeon General, *Social Media and Youth Mental Health* (previously cited), p. 9; Jacqueline Nesi, Supreet Mann & Michael B. Robb, *Teens and mental health: How girls really feel about social media*, 2023 (previously cited), p.6.

Adding to the limitations of the measure, the change only applies to users whom the system identifies as being a child, yet the efficacy of TikTok's age verification has been called into question.[159] Indeed, the UK's media regulator Ofcom has found that 16% of British three- and four-year-olds have access to TikTok.[160] In April 2023, the UK's Information Commissioner's Office (ICO) fined TikTok £12.7 million for allowing children under the age of 13 to use the platform in 2020.[161] Interviews with children and young people for this report suggest that many under-18s either lie about their age to access the platform and circumvent child protective measures, or access TikTok through adults' accounts. None of these children would be shown the screen time prompt.

More effective age verification processes would arguably ensure that children under the age of 13 do not access the platform and that children above the age of 13 benefit from protective measures. However, currently available age verification tools, which offer a higher degree of certainty than a simple self-declaration by the user, inevitably require more extensive data collection, raising additional concerns for children's right to privacy. They also do not address the structural issues of the surveillance-based business model discussed in this report.

Some interviewees said that they had tried to limit their use of TikTok and other social media platforms including through screen-time control apps but found them ineffective in stopping the temptation. Nikki, a 24-year-old woman in the Philippines explained, "I've tried notifying myself by using alarm clocks and it doesn't really work. I find myself bargaining for five more minutes with this app when it's just a notification." Others explained how they circumvented efforts by their parents to restrict their social media use. For example, 13 out of 15 pupils at a secondary school in Mombasa, Kenya, admitted to fooling their parents by pretending to do schoolwork on their phones while actually using them to access social media accounts.[162]

A recurrent theme in the FGDs with children and young people was the idea that it is a matter of self-discipline and responsible use whether social media plays a constructive or destructive role in one's life. The agency of children and young people should be respected, and awareness-raising measures and education are needed to inform their decisions and better equip them to navigate online environments and identify risks and harms. However, it is deeply troubling how the lack of a safe online environment, which can impact the fulfilment of their right to health, has been normalized. Children and young people must not be expected to avoid the harms and mitigate the risks of a platform that is designed to keep them engaged at all costs. The responsibility to create a safe environment free from harms rests with the social media companies who have designed, manage and profit from these platforms.

Chapter 7 discusses these corporate responsibilities and TikTok's failure to fulfil its responsibilities in greater detail.

159. New York Times, "A third of TikTok's U.S. users may be 14 or under, raising safety questions", 14 August 2020, https://nytimes.com/2020/08/14/technology/tiktok-underage-users-ftc.html; BBC, "TikTok fined £12.7m for misusing children's data", 4 April 2023, https://bbc.co.uk/news/uk-65175902
160. Guardian, "TikTok being used by 16% of British toddlers, Ofcom finds", 29 March 2022, https://theguardian.com/technology/2022/mar/29/tiktok-being-used-by-16-of-british-toddlers-ofcom-finds
161. Information Commissioner's Office, "ICO fines TikTok £12.7 million for misusing children's data", 4 April 2023, https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/04/ico-fines-tiktok-127-million-for-misusing-children-s-data/
162. Focus group discussion, Mombasa, Kenya, 17 March 2023.

6. DOWN THE "RABBIT HOLE"

> ***"When I felt low, I think 80% [of the content] related to mental health. It's a rabbit hole because it starts with just one video. If one video is able to catch your attention, even if you don't like it, it gets bumped to you the next time you open TikTok and because it seems familiar to you, you watch it again and then you watch it again and then the frequency of it in your feed rises exponentially."***
>
> Luis, 21-year-old university student from Manila[163]

6.1 HOW TIKTOK'S 'FOR YOU' FEED PUSHES AN INHERENTLY DANGEROUS SYSTEM TO THE MAXIMUM

The core element of TikTok's engagement strategy is its 'For You' page, the default feed through which users are shown a fast-paced, personalized stream of short videos. Beyond its addictive nature, the feature is also at the centre of concerns about the platform's role in exposing children and young people to potentially harmful content.

Early debates on platform regulation focused on how to require companies to rapidly remove illegal and abusive posts amid ever increasing masses of user-generated content. With floods of disinformation and inciting posts sweeping across platforms in polarized political contexts and conflict settings, public scrutiny has shifted towards the ways in which social media platforms actively *spread* content.[164] As one unnamed Facebook employee explained:

> **"We… have compelling evidence that our core product mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish on the platform… The mechanics of our platform are not neutral."[165]**

163. Luis (pseudonym), a 21-year-old undergraduate student in Manila was interviewed in person on 6 May 2023.
164. Amnesty International, Myanmar: *The Social Atrocity* (previously cited); UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression (UN Special Rapporteur on freedom of expression), *Disinformation and Freedom of Opinion and Expression*, 13 April 2021, UN Doc. A/HRC/47/25; *Disinformation and Freedom of Opinion and Expression During Armed Conflicts*, 12 August 2022, UN Doc. A/77/288; Election Integrity Partnership, *The Long Fuse: Misinformation and the 2020 Election*, 2 March 2021, https://www.atlanticcouncil.org/in-depth-research-reports/the-long-fuse-eip-report-read
165. The Facebook Papers, "What is Collateral Damage?", 12 August 2019, cited in Amnesty International, *Myanmar: The Social Atrocity* (previously cited).

UN experts, Amnesty International and many other civil society organizations have documented how leading social media content recommender systems have contributed to spreading and amplifying "propaganda, extremist content and disinformation" in recent armed conflicts and situations of political violence.[166]

Although this report focuses on a different type of amplified content (depressive and self-harm-related content) and a different set of risks (potential harm to children and young users with pre-existing mental health concerns who are exposed to such content), the source of these risks is fundamentally the same: a social media platform's algorithmic recommender system, which is geared towards creating a personalized feed of viral content that will keep the user engaged for the greatest amount of time.

Despite evidence that the core mechanics of personalized recommendations, virality and "optimizing" to maximize user engagement not only create risks of human rights abuses and violations, but have contributed to concrete harms, social media platforms such as TikTok, Instagram and Facebook continue to employ them to fulfil their goal. Indeed, TikTok's success in doing this has pushed its competitor Meta (Instagram and Facebook's parent company) to emulate TikTok's reliance on personalized recommendations within its engagement strategy and to invest further in refining its algorithmic systems.[167]

And yet governments have thus far failed to put forward legislation to stop platforms' continued 'optimization' towards inferences-based personalization, virality and maximizing user engagement. TikTok openly markets its ability to shape an ever more personalized feed by analysing a user's interaction with the platform, effectively profiling the user, from the outset. In its own words:

"From the moment you start your TikTok journey, your first set of likes, comments, and replays will initiate recommendations as the system begins to learn more about your video tastes… These recommendations should become increasingly more personalized over time. Following new accounts, exploring hashtags, sounds, effects, and trending topics will all help to refine your feed."[168]

Previous reports by Amnesty International have detailed why the massive collection of intimate data that enables this kind of 'optimization' based on intimate profiling rather than deliberately given signals constitutes a corporate form of mass surveillance that is incompatible with the right to privacy and risks causing or contributing to knock-on harms involving further human rights abuses.[169] A leaked Facebook (now Meta) document revealed in 2017 that the company claimed to be able to identify "moments when young people need a confidence boost" and was apparently willing to effectively sell those moments of sadness or anxiety as sales opportunities to advertisers.[170] TikTok has more recently marketed its ability to identify when users are "emotionally engaged" in advertising content.[171] Meta and TikTok's stated objectives represent an attack on children and young people's right to privacy and freedom of thought. They are evidence of both companies' intent to analyse and infer users' private thoughts and emotions and to exploit these for commercial profit.

---

166. UN Special Rapporteur on freedom of expression, *Disinformation and Freedom of Opinion and Expression During Armed Conflicts*, 12 August 2022, UN Doc. A/77/288 ; Amnesty International, *Myanmar: The Social Atrocity* (previously cited); The Atlantic, "Who's behind #IStandWithPutin?", 5 April 2022, https://www.theatlantic.com/ideas/archive/2022/04/russian-propaganda-zelensky-information-war/629475; Hannah Porter, "A conversation on fighting disinformation in Yemen", Yemen Policy Center, March 2022, https://www.yemenpolicy.org/a-conversation-on-fighting-disinformation-in-yemen/; Paul Barrett and Justin Hendrix, *A Platform 'Weaponized': How YouTube Spreads Harmful Content – And What Can Be Done About It,* NYU Stern Center for Business and Human Rights, June 2022, https://www.stern.nyu.edu/experience-stern/faculty-research/platform-weaponized-how-youtube-spreads-harmful-content-and-what-can-be-done-about-it; Sara Brown, "In Russia-Ukraine War, Social Media Stokes Ingenuity, Disinformation", MIT Sloan School of Management, 6 April 2022, https://mitsloan.mit.edu/ideas-made-to-matter/russia-ukraine-war-social-media-stokes-ingenuity-disinformation
167. TechCrunch, "Meta expects recommendation models 'orders of magnitude' bigger than GPT-4. Why?", 29 June 2023, https://techcrunch.com/2023/06/29/metas-behavior-analysis-model-is-orders-of-magnitude-bigger-than-gpt-4-why/?guccounter=1; Platformer, "Meta's Nick Clegg on how AI is reshaping the feed", 30 June 2023, https://www.platformer.news/p/metas-nick-clegg-on-how-ai-is-reshaping
168. TikTok, "What is the 'For You' feed?", https://www.tiktok.com/creators/creator-portal/how-tiktok-works/whats-the-for-you-page-and-how-do-i-get-there (accessed on June 2023).
169. Amnesty International, *Surveillance Giants* (previously cited); Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, (previously cited).
170. MIT Technology Review, "Is Facebook targeting ads at sad teens?", May 2017, technologyreview.com/2017/05/01/105987/is-facebook-targeting-ads-at-sad-teens/
171. Access Now, "Open letter: TikTok's 'Focused View' feature must respect human rights law", 7 February 2023, https://www.accessnow.org/open-letter-tiktoks-focused-view/

Teenagers are not just part of user growth strategies, they are also particularly vulnerable to manipulation. Researchers in the field of neuroscience are still trying to better understand how adolescents' social environment and modern technology affect the brain at this particularly malleable stage in its development. Adolescents' "heightened sensitivity to rewards" appears to both promote positive behaviours and make them more likely to engage in risky behaviours.[172] Rational thinking as opposed to emotional processing does not fully develop until someone reaches their mid-twenties.[173] Teenagers are therefore much more likely than adults to act impulsively and to experience strong emotional reactions to external stimuli.[174] This means that caution is required not just in relation to the negative effects of user surveillance and its use in targeting advertisements but also in platforms' targeted amplification of content more widely.

Social media platforms play a huge part in this key developmental stage in young people's lives and their striving for self-understanding. Internal company documents, most prominently the "Facebook Papers",[175] and independent research show that leading platforms exert enormous influence over children and young people's moods, perceptions and self-image through the content they recommend, and the platform designs they employ.[176]

As the Committee on the Rights of the Child has previously argued, social media platforms' use of inferences drawn from their massive data collection to profile users and 'personalize' recommendations poses a very real risk of interfering with a child's thoughts and emotions in such a way as to undermine the child's freedom of thought.[177]

There are many studies documenting these systemic risks, including recent research reports specifically highlighting TikTok's role in amplifying problematic and harmful content. In 2021, Reset Australia exposed how TikTok's 'For You' feed promoted content that portrays and promotes negative ethnic and gender stereotypes.[178] In the same year, the Wall Street Journal documented how TikTok sent teenage users down "rabbit holes" of eating disorder videos[179] and depressive content.[180] In 2022, a study by the Center for Countering Digital Hate of TikTok's amplification of content related to eating disorders and self-harm found that new accounts with the term 'loseweight' in their name were more likely to receive recommendations for both types of harmful content.[181] In March 2023, another small-scale study found that TikTok's 'For You' feed required only "low-level engagement" from young users for it to rapidly start recommending extreme posts related to suicide and to the misogynistic and violent so-called 'incel' ideology.[182]

These findings point to some of the serious potential knock-on harms associated with TikTok's recommender system, adding risks to children and young persons' health to the previously discussed risks to the rights to privacy and to freedom of thought.

---

172. Zara Abrams, "What neuroscience tells us about the teenage brain", 25 August 2022, Monitor on Psychology, Volume 53, Issue 5, https://www.apa.org/monitor/2022/07/feature-neuroscience-teen-brain
173. University of Rochester Medical Centre, Health Encyclopaedia, "Understanding the teen brain", 2023, https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=1&contentid=3051 (accessed on 14 September 2023).
174. BBC, "The biggest myths of the teenage brain", 7 September 2022, https://www.bbc.com/future/article/20220823-what-really-goes-on-in-teens-brains
175. Wall Street Journal, "Facebook knows Instagram is toxic for teen girls company documents show" (previously cited).
176. Reset Australia, *Surveilling Young People Online: An Investigation Into Tiktok's Data Processing Practices*, July 2021, https://au.reset.tech/uploads/resettechaustralia_policymemo_tiktok_final_online.pdf; Jacopo Pruccoli and others , "The use of TikTok among children and adolescents with eating disorders: Experience in a third-level public Italian center during the SARS-CoV-2 pandemic", 30 July 2022, Italian Journal of Pediatrics, Volume 48, https://doi.org/10.1186/s13052-022-01308-4; Reset Australia, *Designing for Disorder: Instagram's Pro-Eating Disorder Bubble in Australia*, April 2022, https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf; Tech Transparency Project, "'Thinstagram': Instagram's algorithm fuels eating disorder epidemic", 8 December 2021, https://www.techtransparencyproject.org/articles/thinstagram-instagrams-algorithm-fuels-eating-disorder-epidemic
177. Committee on the Rights of the Child, General Comment 25 (previously cited), para. 62.
178. Reset Australia, *Surveilling young people online* (previously cited).
179. Wall Street Journal, "'The corpse bride diet': How TikTok inundates teens with eating-disorder videos", 17 December 2021, (previously cited).
180. Wall Street Journal, "Investigation: How TikTok's algorithm figures out your deepest desires", 21 July 2021, https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477
181. Center for Countering Digital Hate, *Deadly by Design*, 15 December 2022, https://counterhate.com/research/deadly-by-design/
182. Ekō, *Suicide, Incels, and Drugs: How TikTok's Deadly Algorithm Harms Kids*, March 2023, https://s3.amazonaws.com/s3.sumofus.org/images/eko_Tiktok-Report_FINAL.pdf

A number of bereaved parents claim that TikTok's content curation has already turned deadly. A lawsuit filed in the USA in June 2022 argues that TikTok played a role in the deaths of at least seven children who took part in the so-called 'blackout challenge', a dare, in which participants choke themselves and share videos of the act online, videos which were, according to one of multiple lawsuits, not just searchable, but amplified through the 'For You' feed.[183] TikTok disputes the parents' claim, having stated that the challenge "predates [the] platform and has never been a TikTok trend."[184]

Many of the children and young people who participated in the research for this report were acutely aware of the risks associated with TikTok's user engagement practices relative to other leading platforms.[185] "All the platforms that I use, there's a degree of unsafeness. But I think out of all the platforms, I really feel unsafe with TikTok", explained "Nikki", 24, based in Manila.[186]

In response to questions about the kinds of content that they had come across in their feeds, which they did not want to see, the young participants frequently mentioned sexual content and content containing extreme violence. For example, 17-year-old "Maria", a pupil in Manila, explained that a video she had come across in her feed showed the killing of a woman, which she believed was a real crime.[187] In another FGD with young adults in Kisumu, Kenya, most of the participants reported having come across the same viral video, which they said showed a young woman being forcibly undressed by a group of men and then submitted to an act of female genital mutilation.[188]

6.2 HEIGHTENED RISKS FOR YOUNG PEOPLE WITH MENTAL HEALTH CONCERNS

Multiple young people with self-reported mental health concerns told Amnesty International researchers about their experiences of getting drawn into TikTok's "rabbit holes" of triggering and mental health-related content. They felt that TikTok's amplification of problematic mental health-related content contributed to exacerbating their symptoms. For example, "Luis", an undergraduate student in Manila diagnosed with bipolar disorder, described his experience with TikTok's 'For You' feed:

> **"When I felt low, I think 80% [of the content] related to mental health. It's like a spiral. It's a rabbit hole because it starts with just one video. If one video is able to catch your attention, even if you don't like it, it gets bumped to you the next time you open TikTok and because it seems familiar to you, you watch it again and then you watch it again and then the frequency of it in your feed rises exponentially."[189]**

Echoing the experience of others, "Luis" described how periods of deliberate engagement with TikTok videos reflecting his own depressive or anxious thoughts would keep him trapped in negative content loops even after the depressive phase subsided:

> **"TikTok assumes that you want this particular content for a particular period of time. But the thing for me is that my preferred content changes very much. For instance, one day I'm very anxious, I want this content. And one day I'm very happy and very ecstatic about something. But the thing is, I'm still viewing content that's very sad and depressing, so that messes me up."**

183. The Verge, "The TikTok 'blackout challenge' has now allegedly killed seven kids", 8 July 2022, https://www.theverge.com/2022/7/7/23199058/tiktok-lawsuits-blackout-challenge-children-death
184. The Verge, "The TikTok 'blackout challenge' has now allegedly killed seven kids", 8 July 2022 (previously cited).
185. Among the Kenyan focus groups alone, at least 40 participants gave examples of feeling unsafe on TikTok, for example in the context of what they felt were violent videos (including videos they believed to show a rape and a murder) or age-inappropriate (e.g. pornographic) content, which they came across in their feeds.
186. Online interview, Manila, Philippines, 6 May 2023.
187. Online interview, Manila, Philippines, 9 May 2023.
188. Focus group discussion, Kisumu, Kenya, 13 March 2023.
189. Luis (pseudonym), an undergraduate student in Manila in his early 20s was interviewed in person on 6 May 2023.

"Francis", a student in Batangas Province, Philippines, similarly observed:

> **"When you 'heart' a sad video that you could relate to, suddenly my whole 'For You' Page is sad and I'm in 'sadtok'. It affects how I'm feeling".[190]**

Another young focus group participant explained:

> **"As an overthinker, the content I see makes me overthink [even] more, like videos in which someone is sick or self-diagnosing. It affects my mentality and makes me feel like I have the same symptoms and worsens my anxiety. And I don't even look them up, they just appear in my feed."[191]**

Eight young adults in the Philippines said they had been shown clusters of triggering content that affected their health and well-being from. Among the eight were "Cris", a young teacher trying to manage his attention deficit hyperactivity disorder (ADHD),[192] 22-year-old "Daisy", struggling with body image issues,[193] and "Lance", 23, a young transgender man dealing with self-reported body dysphoria.[194] Other young people, in particular in the Philippines, also reported negative mental health impacts of the amplification of political propaganda and, what they believed to be misinformation, in the context of the already highly polarized political environment in which they live. While it is difficult to attribute a person's mental health condition to any one factor, these testimonies highlight that there is a risk that TikTok may exacerbate pre-existing mental health concerns among children and young people.

6.3 EXPLORING TIKTOK'S RECOMMENDER SYSTEM SYSTEMATICALLY

Amnesty International's investigation into TikTok's recommender system aimed to contribute to the available evidence in two ways. First, it incorporated two locations (Kenya and the Philippines), which have been underrepresented in studies into this subject to date. Second, the audit was conducted with a larger sample size than has often been the case with previous civil society organizations' investigations in this field.[195]

6.3.1 AUTOMATED (SOCK PUPPET) RESEARCH DESIGN

Amnesty International, together with its technical partners, the Algorithmic Transparency Initiative and AI Forensics, conducted a two-part audit of TikTok's 'For You' feed recommender system with regard to mental health-related content. The first was an automated – or so-called 'sock-puppet' (bot) based – audit of recommendations made to accounts in Kenya and the USA.[196] The second was a manually run experiment involving an account each in Kenya, the Philippines and the USA.

6.3.2 AUTOMATED (SOCK PUPPET) RESEARCH DESIGN

For the automated audit, researchers set up 40 automated accounts with four different pre-defined personas to mimic different young people's behaviours on TikTok. Each account was set up to run for just under 60 minutes in a single session each day for 10 days.[197] The accounts were divided into

190. Focus group discussion in Tanauan City, held on 13 May 2023.
191. Focus group discussion in Tanauan City, held on 13 May 2023.
192. Interview in Manila, conducted on 6 May 2023.
193. Interview in Manila, conducted on 6 May 2023.
194. Interview in Manila, conducted on 6 May 2023.
195. See for example Center for Countering Digital Hate, *Deadly by Design*, 15 December 2022 (previously cited); Reset Australia, *Surveilling Young People Online: An Investigation Into Tiktok's Data Processing Practices*, July 2021 (previously cited).
196. No accounts from the Philippines were included in the automated research due to project limitations. The research focused instead on a comparative analysis of a country that was underrepresented in previous research in this field (Kenya) and a key market and country where there is active public debate of social media risks (USA).
197. Where technical issues disrupted a session, an additional session was run so that the account was active for roughly 10 hours in total. Since TikTok sets the voluntary time limit at 60 minutes per day for under-18s, the user experience was tested within that same daily time limit, although this is lower than the average time spent by teens on TikTok according to US market research.

sub-groups following four different scrolling behaviours, of which 20 accounts were set up to simulate 13-year-olds in the USA and another 20 simulated 13-year-olds in Kenya (of which 11 were included in the analysis).[198] The age of the accounts (13 years) was chosen to examine the recommendations served to the youngest permitted age group of TikTok users, to whose accounts TikTok applies teen safety measures.[199] In order to limit the extent to which the study promoted potentially harmful content for other users, the accounts did not like, comment, message, search for anything or follow anyone; they only watched recommended posts in the 'For You' feed twice if the post's description included a term from pre-defined lists:

Sub-group 1 (Depressive behaviour (funnelling)): The accounts in this sub-group simulated children with an interest in mental health who engage with increasingly extreme content once presented with it. Accounts in this category initially rewatched videos associated with a broader set of 150 mental health-related terms. Halfway into the experiment, they were then set to rewatch only content tagged with a shorter list of 64 terms that we identified prior to the experiment as often associated with overtly depressive or self-harm-related content.

Sub-group 2 (Broad mental health interest): Accounts rewatched videos tagged with a list of 150 mental health-related terms.

Sub-group 3 (General interest (referred to as "benign" terms in graphs) and mental health interest): Accounts in this category rewatched videos containing a selection of 480 general interest terms as well as the broad list of 150 mental health-related terms.

Sub-group 4 (General interest (referred to as "benign terms in graphs, control group)): These accounts only rewatched videos associated with 480 general interest terms, such as animalsoftiktok, comedy or traveltiktok.[200]

The broad mental health-related terms list included terms such as: mentalhealth, mentalhealthmatters, sadtiktok, schooldraining, struggling, ventingaccount as well as prominent abbreviations, misspellings and code words for depression and self-harm.[201]

The narrow mental health-related terms list comprises a subset of terms from the broad terms list that were found to be mostly associated with problematic and harmful content found in trial experiments.[202]

The primary research question was to what extent (if any) children and young people are shown potentially harmful videos depicting and/or romanticizing or encouraging depression, self-harm and suicide through algorithmic recommendations. The working hypothesis was that, if the amplification effect was strong, researchers would observe a steady increase in the volume of potentially harmful mental-health related content shown to the accounts in subgroups 1 and 2 (depressive behaviour and broad mental-health interest) relative to those with wider interests.

Although the study involved a larger sample compared to previous civil society research in this field, the sample size of 40 accounts was still not large enough for any statistical hypothesis testing. Therefore, the primary analysis descriptively compared the volume and frequency of mental-health related content seen by each sub-group. The research aimed to build on prior studies by including two quasi-control groups in the form of sub-groups that solely rewatched general interest content (referred to as "benign terms" in graphs), or general interest alongside mental health content. The inclusion of these two groups allowed researchers to observe the magnitude of amplification by the recommender

198. The experiment started with 20 accounts set up to represent teenage users in each country but only 11 Kenyan accounts ran until the conclusion of the investigation. Only 11 of the 20 Kenyan accounts ran for all ten sessions, the others were disabled by TikTok before their final session and were therefore excluded from the analysis.
199. See Annex 3: Written Response to Amnesty International's Research Questions.
200. For the full lists of terms, please see Annex 2.
201. For the full lists of terms, please see Annex 2.
202. For the full lists of terms, please see Annex 2.

system by providing a baseline measure of the volume and frequency of mental-health related content that TikTok serves to an "average" user.

It is highly likely factors such as the date, time of day, and location are used within the recommender algorithm and therefore are influential in determining the content that TikTok serves its users. To ensure the comparisons between sub-groups were as robust as possible, researchers attempted to control these by setting up bot accounts to access the platform on specific dates and times from the same location in Kenya and the USA. Technical issues such as bots being shut down or malfunctioning presented challenges to this, however researchers ensured to the greatest extent possible, that all data was collected in parallel across each sub-group.

6.3.3 MANUAL EXPERIMENT RESEARCH DESIGN

To counterbalance some of the limitations of the automated approach such as the need to pre-define terms through which the account would recognize relevant content and the inability of automated accounts to identify relevant posts based on mood, music or in-video captions rather than hashtags, Amnesty International and AI Forensics also conducted a small-scale manual experiment. This consisted of hour-long screen recordings covering a researcher's interaction with the 'For You' feed of newly set-up accounts, again representing 13-year-olds. One account each was set up using VPNs for Kenya, the Philippines and the USA. The researcher again only scrolled through the 'For You' feed and watched posts two to three times if they related to sadness or mental health issues.

6.4 MEASURING AND CATEGORIZING HARMFUL CONTENT FOR THE PURPOSES OF THE RESEARCH

6.4.1 CONSTRUCTING A MEASURE OF AMPLIFICATION

In the available literature on risks and harms associated with social media the precise meaning of "harmful content" is rarely defined.[203]

This presents a substantial challenge given that TikTok content comprises of a mixture of video, text and audio data and, in this context, is collected at such a large scale that makes manual approaches to categorization unfeasible. Despite these challenges, it was vital and integral to the investigation to classify content to provide a measure of amplification. Researchers therefore employed the following approach:

1. The primary measure of amplification was the percentage of recommended videos that contain a term from the broad mental health terms list (either within a hashtag or associated text). The researchers also analysed and graphically depicted terms that co-appear to understand how video topics are related.
2. Drawing on the qualitative research with children and young people and feedback from interviewed mental health experts, researchers constructed a content categorization framework. A panel of reviewers was convened, who manually classified a sample of 540 TikTok videos using this framework, with each piece of content reviewed independently by three reviewers.
3. Researchers then analysed how the manual categorizations corresponded to content hashtags and text descriptions. This approach allowed researchers to validate our primary measure of

203. The US Surgeon General's 2023 advisory for instance describes harmful content as "e.g., content that encourages eating disorders, violence, substance abuse, sexual exploitation, and suicide or discusses suicide means". See: US Surgeon General, *Social Media and Youth Mental Health* (previously cited).

amplification by assessing how much content containing a term from the broad list was deemed harmful in the manual classification exercise.

A key limitation to the primary measure of amplification is that assigned video hashtags are used to both: train the bots and mimic user behaviour, and to measure of the volume of mental health content they were served. With additional capacity, further manual categorization could have provided a more in-depth picture and possibly helped to refine the quantitative analysis methods. Civil society research efforts in this field remain constrained by budgetary and time constraints in addition to hurdles created by the lack of appropriate independently vetted research access frameworks at the time of writing (with the DSA's research access framework not yet implemented). This further underlines the need for regulatory efforts outside the EU to implement similar data access obligations as well as to create enforceable obligations on leading social media platforms to publicly report on their risk assessments and mitigating actions.

6.4.2 CLASSIFYING CONTENT

While Amnesty International's investigation was concerned with the cumulative effect of TikTok's amplification of specific kinds of content to a child over time rather than the effect of individual failures of content moderation, coding the data nonetheless required researchers to categorize individual pieces of content. The categories used in this report reflect feedback from various experts and clinicians but can nonetheless only ever present an approximation of the content observed and its potential for harming a person's mental health.[204] The categorization scheme that was used sought to provide a transparent description of the dimensions of potentially harmful mental health-related content that was recommended to the research accounts rather than a binary distinction between 'harmful' and 'non harmful' content.

In reality, beyond blatant examples of traumatizing content, perceptions and emotional reactions to content are highly individual. Amnesty International's interviews with children and young people underlined that what a person perceives as triggering or harmful to their mental health is often dependent on personal experiences and circumstances. It can also vary over time according to changes in emotional and mental state.

Moreover, in relation to user-generated content, psychologists also disagree about the pros and cons of non-experts sharing information related to mental health on social media. As noted above, platforms such as TikTok do offer unprecedented opportunities for self-expression and a sense of community for many, including those who share common mental health concerns.[205] They can also connect individuals who otherwise have no access to mental health resources with information that is designed to be easily and widely accessible. But experts have raised serious concerns about bitesize video clips conveying information and potential misinformation, as well as content containing (supposed) medical advice without any knowledge of the viewer and their individual needs, or encouraging self-diagnosis on the basis of incomplete or false information.[206]

Crucially, this content is not consumed in isolation, but in the context of a fast-paced feed that, as Amnesty International's research shows, can rapidly turn into a flood of problematic mental health-related content.

In addition to examples of content that clearly fitted into the categories defined for the experiment, researchers also identified recommended posts that were difficult to categorize with certainty, for example because they were too short, making the meaning and intent unclear. To mitigate possible individual biases and to validate our findings, each post was assigned a category by three separate reviewers on the Junkipedia platform[207] and the final categorization was based on a majority ruling decision in cases where reviewers assigned different categories.

204. Feedback was provided by S. Bryn Austin, ScD (STRIPED/Harvard Medical School), Amanda Raffoul, PhD (Harvard Medical School), Zeyna Awan, Nairobi-based psychologist, Kevin Gachee, Nairobi-based psychologist, Aliya Shah, Nairobi-based psychologist.
205. John A. Naslund, Ameya Bondre and others, "Social media and mental health: Benefits, risks, and opportunities for research and practice" (previously cited).
206. Joe Martin, "The big issue: Mental health and the TikTok effect", April 2023, https://www.bacp.co.uk/bacp-journals/therapy-today/2023/april-2023/the-big-issue/
207. Further technical details in Annex 1.

Types of mental health-related content with potential harmful effects when amplified at scale

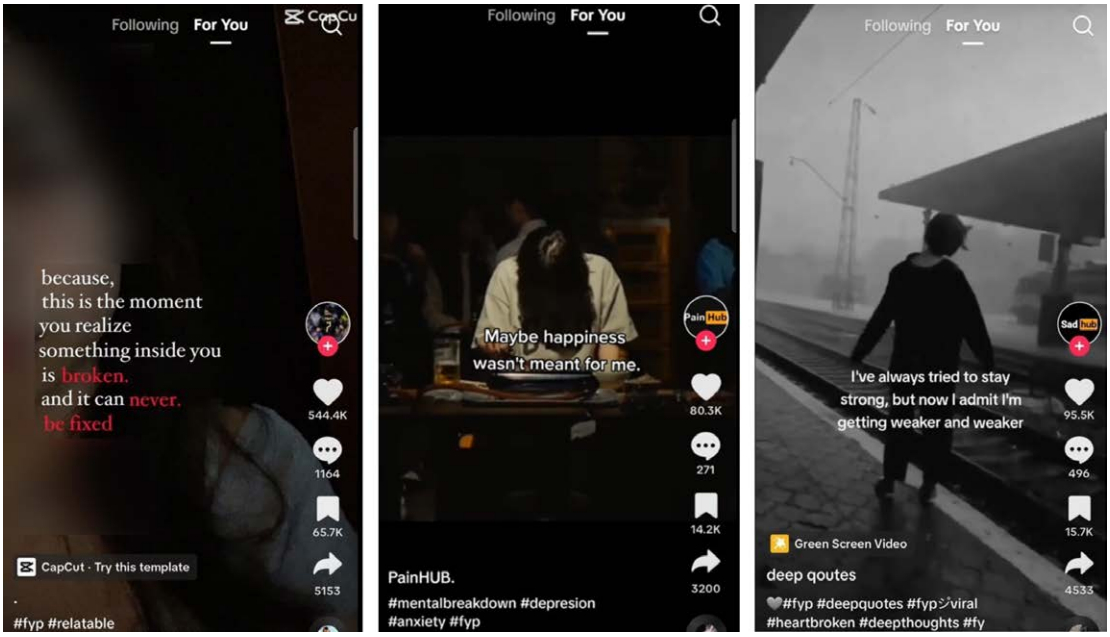
6.5 HEADLINE FINDINGS

The results of Amnesty International's research into TikTok show that children and young people who watch mental health-related content on TikTok's 'For You' page can easily be drawn into "rabbit holes" of potentially harmful content, including videos that romanticize and encourage depressive thinking, self-harm and suicide.

This was the case for automated accounts set to represent 13-year-olds in Kenya and the USA, with some of the same content being recommended to accounts in both countries. After 5-6 hours on the platform, almost one in every two videos served to accounts which had signalled an interest in mental health were mental health-related and potentially harmful, roughly 10 times the volume served to accounts with no interest in mental health.

This mirrored findings from the manual review of 540 videos recommended to US and Kenyan accounts, which showed a steady progression from 17% of the videos served on day one being categorized as potentially harmful to 44% of content on day 10.

The manual experiment conducted with one account each representing 13-year-olds in Kenya, the USA and the Philippines showed an even faster personalization process across all locations, reducing the user's feed to an even more pronounced "rabbit hole" of potentially harmful mental health content in under 21 minutes (compared to 4 to 5 hours for accounts within the automated study).



TikTok videos recommended to a manually-run teen account in its first hour.

6.6 MENTAL HEALTH INTEREST TRIGGERS "RABBIT HOLE" EFFECT

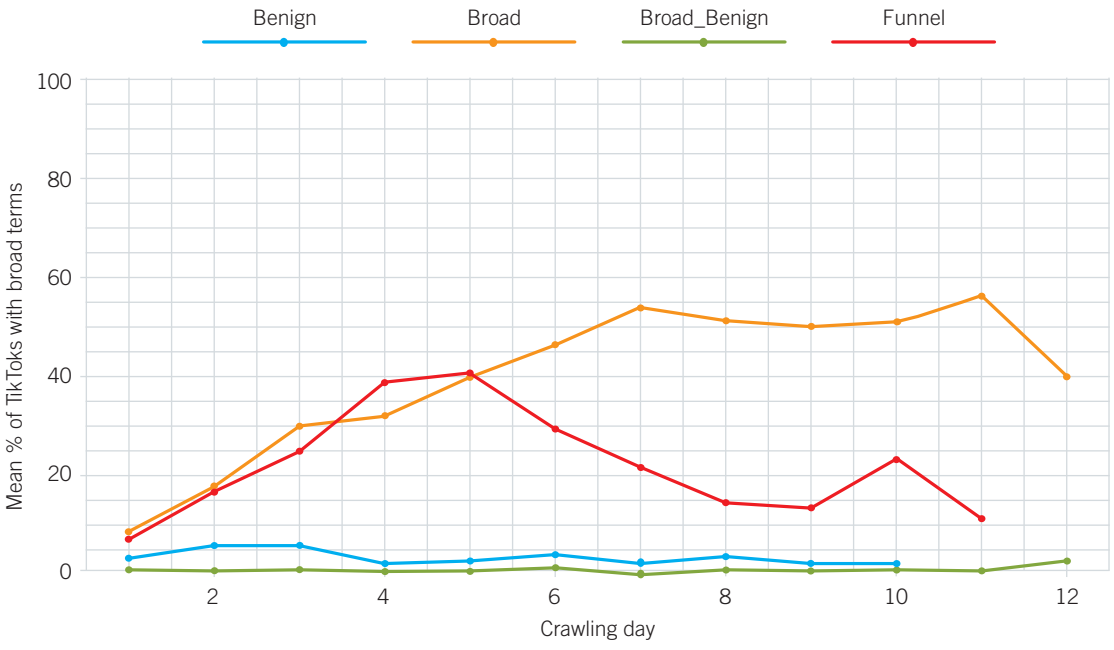
When TikTok's 'For You' page recommender system learns that a user is interested in videos related to mental health, it is highly likely to turn this signal into pronounced "rabbit holes" of content that discusses and encourages depressive thinking and self-harm.

Capacity limitations only allowed the manual review a sample of 540 of the thousands of recommended videos (details in section 6.6.4), meaning the validity of the sample results for the overall experiment is limited. Nevertheless, the sample suggests that the following percentages, based on a quantitative analysis of all video recommendations associated with the broad list of mental health terms across the whole experiment, likely underestimate the actual number of video recommendations containing potentially harmful content. The percentage of videos associated with a term from the broad mental health terms list in the sample was 6% lower than the actual percentage of videos manually classified as potentially harmful (categories 1-9) (the sample analysis will be further discussed in section 6.6.4).

Our quantitative analysis of the automated experiments showed that:

- The amplification effect appears to be strong. Researchers observed a steady increase during the first five to eight days in the volume of mental health content served to accounts signalling an interest in mental health content (subgroups 1 (depressive) and 2 (broad mental health interest)) compared to those with wider interests (figures 1 and 2).
- The volume of mental health-related content first peaked between day five and eight with it representing between 40-55% of all content served to our US-based depressive behaviour and broad mental-health interest sub-groups during this period (figure 1). Evidence from our manual categorization exercise suggests that 3 in every 5 videos that are identified as mental health-related based on terms and hashtags in the descriptions are potentially harmful.
- The volume of mental-health related content remained relatively low (under 10% of all content served) and constant over time to our sub-groups with wider interests throughout the study period regardless of location (figures 1 and 2).
- The US funnelling accounts saw a decline in mental-health related content after day 5, at which point the accounts were programmed to only rewatch content that contains a term from the narrow list (figure 1). Analysis suggests this is likely due to the fact the overall frequency of content containing a term from the narrow list amongst the total recommended content was very low (<1% both before and after funnelling) for four out of the five accounts. Therefore, the accounts were looking for specific content of which there appears to not have been much available within the recommendations at the time of the switch. In the absence of an effective signal, the recommender system likely used other factors to shape the accounts' 'For You' feeds (e.g. current trending topics, events).

Figure 1: Percentages (means per subgroup) of recommended videos served to US accounts (whose descriptions include terms from the broad list of mental health-related terms per day[208])



208. If due to technical issues an account ran for an insufficient amount of time in a planned one-hour-session, the incomplete session was merged with the next session, recorded on the same day. For table two: In one case, a Kenyan account did not collect a full session and no other session was recorded that day, so the incomplete session was excluded from the analysis.

- The Kenyan funnelling accounts, by contrast, retained a steadier volume of mental-health related content before and after day 5 (figure 2). Analysis shows this was due to the small number of Kenyan funnelling accounts that reached the end of the study (only two accounts were not suspended), of which one account experienced a substantial rabbit hole effect likely driving TikTok to continue recommending mental-health related content after the switch to the shorter list of terms that we identified prior to the experiment as often associated with overtly depressive or self-harm-related content.

Figure 2: Percentage of recommended videos served to Kenyan accounts (whose descriptions include terms from the broad list of mental health-related terms per day)

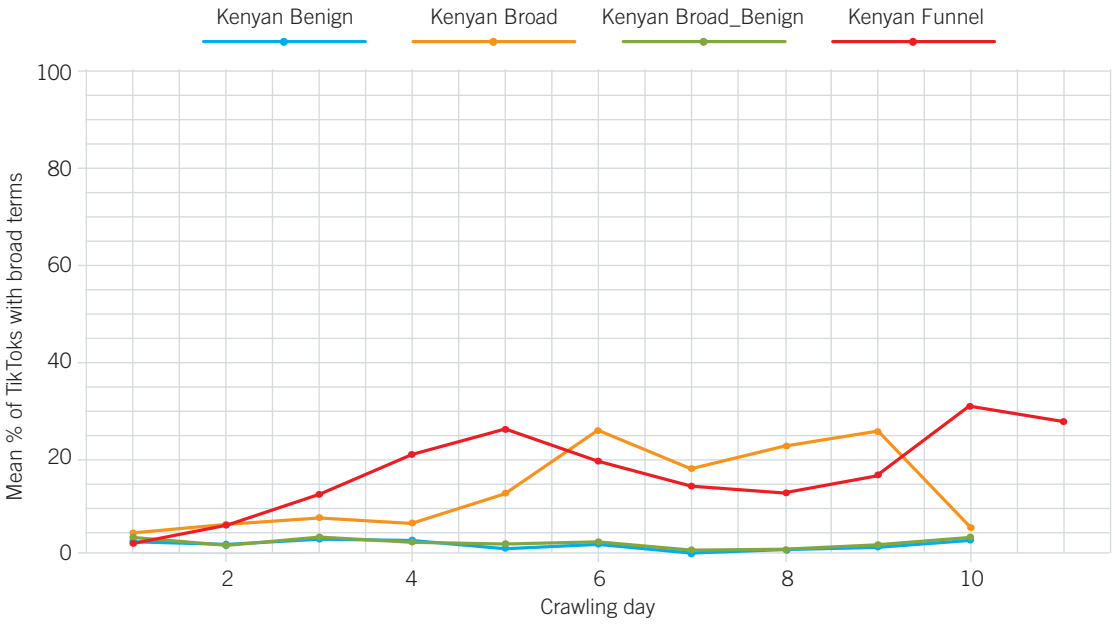
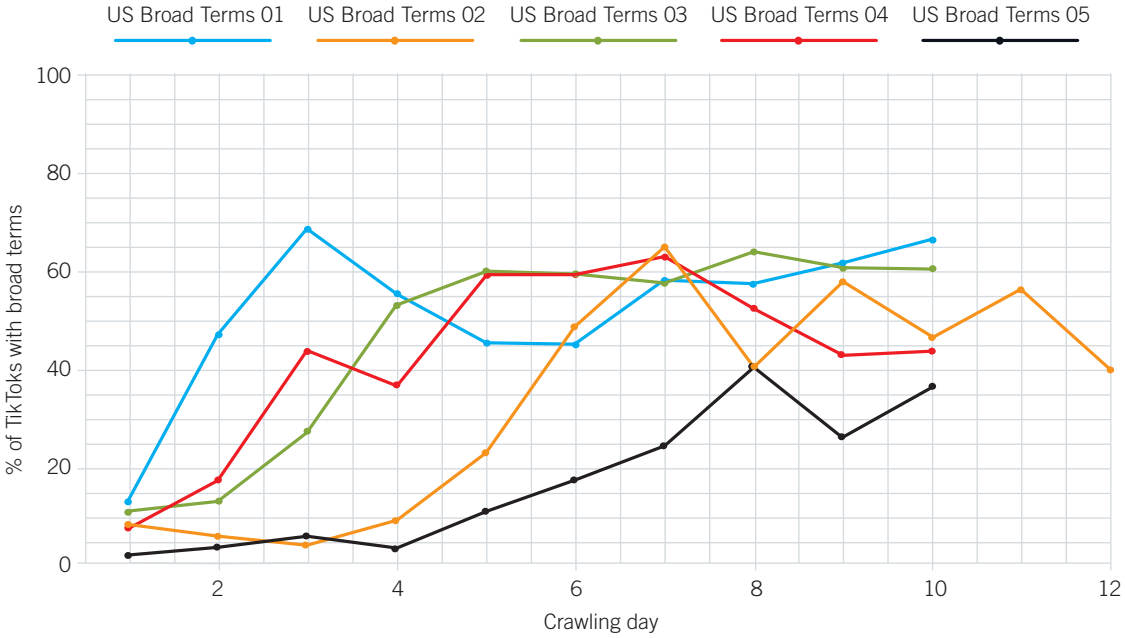


Figure 3: Percentage of recommended videos served to US accounts (group 2: broad mental health interest) whose descriptions include terms from the broad list of mental health-related terms per day



- There were considerable differences in amplification across accounts. Within the broad US-based mental-health interest sub-group, most accounts reached peaks of more than 60% of recommended videos relating to mental health issues (per day), whilst one account remained below 45% (figure 3).
- Observing accounts over multiple days provided valuable insights: For some accounts, it took 500 to 600 videos to reach the first peaks of the "rabbit hole" effect (during the second or third hour/day of watching recommended content). Others saw a slower personalization process with the first peak occurring after 1,500 to 2,000 videos (see figure 3, US-based account "USBroadTerms05" based on broad mental health terms list reaching its first peak on day eight).

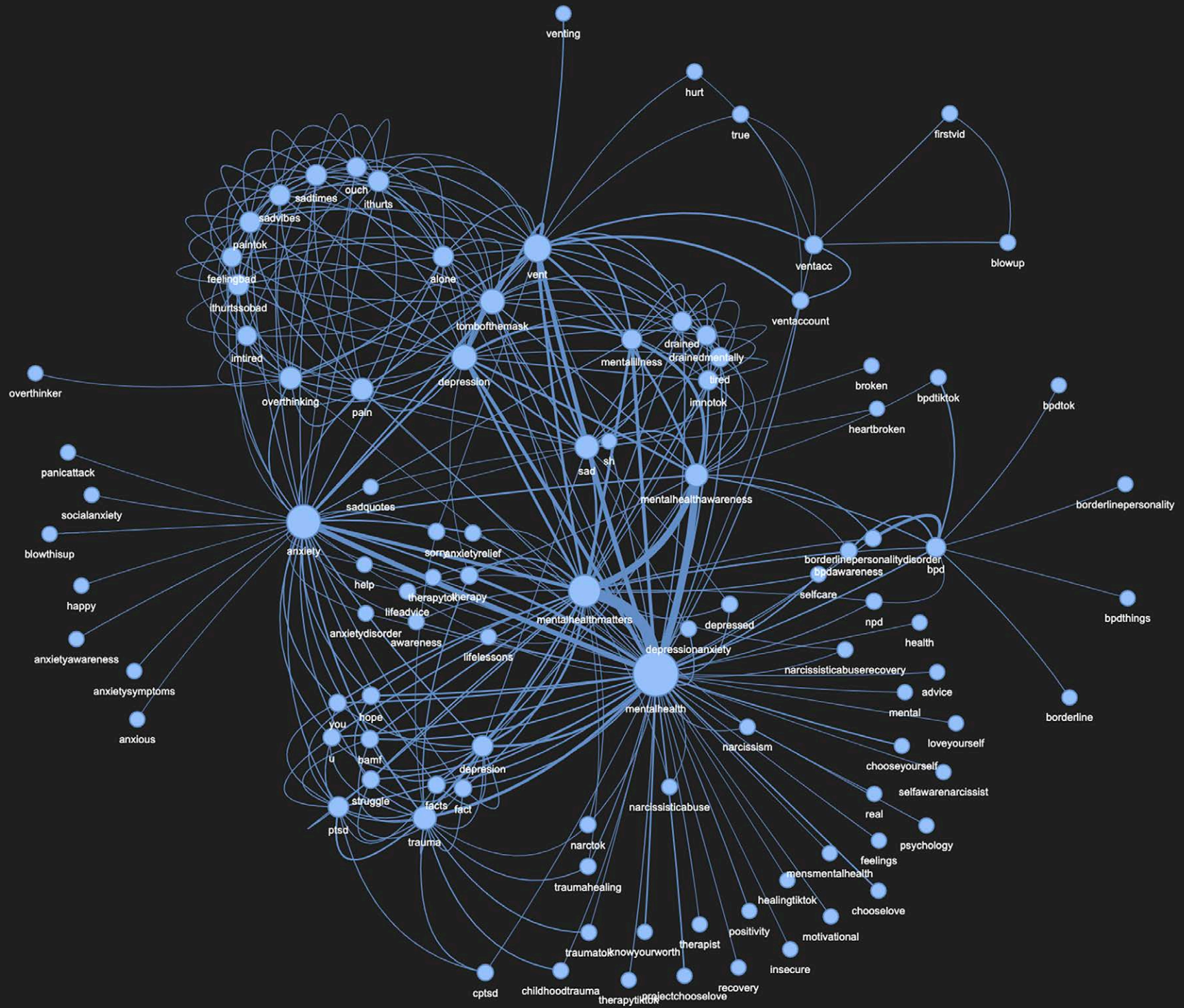
Four out of five accounts saw peaks of more than 60% related content. Across their entire feed, at least 20% of all videos recommended to US accounts in this category were mental health-related (associated with terms from the broad list).

To put these findings into context, "roughly 13 percent" of the videos in British teenager Molly Russell's Instagram feed in the six months before her death were reportedly "related to suicide, depression or self-harm".[209] The Coroner in this case found that this level of personalized amplification of mental health content was enough to play a significant contributing role in her death.

The significantly higher percentage of mental health, depression and self-harm related content found in Amnesty International's experiments[210] on TikTok resulting from just one of the signals usually given by users (such as rewatches, likes and follows), if found in a real child or young person's account, would expose them to an even more pronounced tunnel vision perspective. Particularly for vulnerable children and young people, this could risk exacerbating pre-existing mental health concerns with an accompanying risk of serious harm to their physical health. Whereas researchers working on this project did not like, actively seek out relevant content or follow any users to avoid contributing to the amplification of potentially harmful content, such signals typically sent by most users, would potentially have led to an even higher percentage of related content.

Section 6.6.2 below discusses the manual analysis of TikTok account feeds and takes a closer look at the types of content that were recommended based on the accounts' interest in videos associated with mental health terms and hashtags. In addition, quantitative methods offer insights into the subject and tone of the recommendations. The network map on the next page shows the most commonly amplified terms associated with mental health across all recommended videos and their connections with co-occurring hashtags. The most common terms associated with the broad list of mental health terms are firstly umbrella terms such as 'mentalhealth', 'mentalhealthmatters' and 'mentalhealthawareness' (15-20% of those videos in turn also counting 'vent' amongst their other hashtags and terms), followed by 'vent', 'anxiety' and 'sad'.

209. Politico, "Digital bridge: Platforms on the hook — Transatlantic AI rulebook — Let's talk data transfers", 6 October 2022 (previously cited).
210. Though a caveat applies in that Amnesty International researchers were unable to review the categorization applied to Molly Russell's feed and the categorization may not be directly comparable.



Network map of commonly co-occurring hashtags associated with the broad list of mental health terms tracked in Amnesty International's automated experiment. The size of the circle indicates how often the term appeared in our dataset; the thickness of the line indicates how often terms co-occur. Note: Researchers excluded all combinations where one of the two terms is in the 100 most common terms across videos without any connection to the broad mental health terms list, and where there are fewer than 15 total co-occurrences.

In contrast to the results observed in relation to the accounts only signalling an interest in mental health, researchers did not observe a "rabbit hole" effect for the accounts which combined the broad list of mental health-related terms and the list of benign terms: 0.4 to 1.4% of the Kenyan feeds included videos linked to the broad list of mental health terms and less than 1% of all US feeds in that group. This is likely explained by the weight that was given to interests not related to mental health (with three times as many general interest terms as mental health-related terms) and the likely much higher availability of content related to benign interests across all videos that could be recommended to a user.

Somewhat contradictorily to this, Amnesty International found that, despite showing no active interest in videos associated with any mental health-related terms, three out of five US-based accounts programmed to only rewatch content associated with pets, travel and similar general interest videos ("benign" terms) saw 2% to 4% of their feed relate to terms and hashtags linked to mental health (broad list of mental health terms), with peak periods of up to 12% (across 50 consecutive videos). The Kenyan accounts in that same subgroup viewed 0.6% to 0.7% mental health-related posts. Further analysis would be needed to contextualize this finding, but it could, for example, be due to one of the general interest terms appearing in the description of some of these videos, or high engagement rates of these videos leading to their amplification in the accounts' feeds.

6.6.1 FINDINGS FROM THE MANUAL EXPERIMENTS: "RABBIT HOLE" EFFECTS CAN BE TRIGGERED WITHIN 20 MINUTES

Compared to the automated accounts signalling their interest in mental health, Amnesty International observed an even faster "rabbit hole" effect in the manual experiments and an even higher rate of potentially harmful content. In the manual observation of the recommendations served to an account registered for a supposed 13-year-old in the Philippines, the first video tagged with #depresionanxiety [sic] showing a young boy in distress was suggested within the first 67 seconds of scrolling through recommended content on the 'For You' page. From minute 12 onwards, 58% of the recommended posts related to anxiety, depression, self-harm and/or suicide and fell into one of the aforementioned categories of mental health-related content with potentially harmful effects on children and young people with pre-existing mental health concerns.

In the US-based manual experiment, the fourth video shown was tagged #paintok and focused on text reading "when you realize you've never been put first your entire life but instead are just that person that fills a void in other people's lives until they don't need you anymore". From the 20th video onwards (less than three minutes in), 57% of the videos related to mental health issues, with at least nine posts romanticizing, normalizing or encouraging suicide in a single hour.

The Kenyan account in the manual experiments saw the slowest progression towards a feed filled with mental health-related content. However, once that point was reached (20 minutes into the experiment), 72% of the videos recommended in the next 40 minutes related to mental health struggles, with at least five references to suicidal thinking or the content creator's death wish. The first video referencing a content creator's mental health struggles was recommended 3 minutes into the experiment (25th video recommendation). Not a single mental health-related video was produced by a mental health care professional or recognized mental health organization.

6.6.2 FINDINGS FROM OUR MANUAL CATEGORISATION: A CLOSER LOOK AT THE TYPES OF MENTAL HEALTH CONTENT RECOMMENDED BY TIKTOK'S 'FOR YOU' FEED

To gain a better understanding of the kinds of videos that TikTok recommended to the accounts, researchers manually analysed 20-minute samples viewed by a US and a Kenyan account each from the two sub-groups of accounts that saw a significant "rabbit hole" effect (funnelling and broad mental

health interest) on the first, sixth and tenth day. Researchers manually coded 540 pieces of content, which were each separately coded by three reviewers and categorized according to a majority ruling decision across the reviewers.

Of these 21% contained a term from the broad list, and 27% were classified as potentially harmful (categories 1-9, table 1). This aligns with the above findings from the terms-based analysis as researchers only categorized content from accounts in the broad terms and funnelling sub-groups. Lived experience posts speaking about anxiety, depression and self-harm are the most common category of videos (13% of all content), which when amplified at this scale and in presenting a very narrow focus on this particular type of lived experience, could risk exacerbating existing mental health issues in children and young people.[211] This is followed by posts showing people crying or in emotional distress (4.3% of all videos) and dramatic descriptions of trauma, suffering and suicide (4%). Another 4% of all videos glamorize or trivialize depression, anxiety, self-harm and suicide or speak of plans to die by suicide, explain how to self-harm or encourage self-harm or suicide.

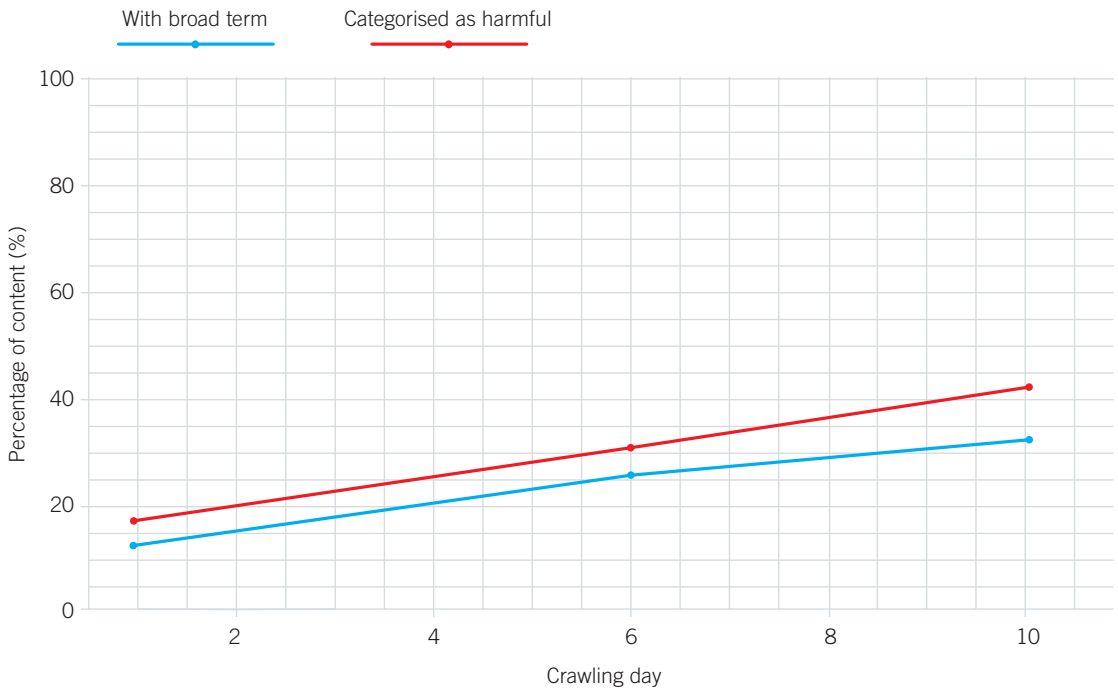
Table 1: Manual categorization of samples from the automated experiment

Code	Category	# Content manually categorized	% of content containing term from broad list
0	Not relevant	393 (73%)	11.2
1	**Lived experience** post speaking about anxiety, depression and self-harm (without romanticizing it)	68 (13%)	44.1
2	Post showing people in **emotional distress**	23 (4.3%)	39.1
3	Post portraying feelings of **loneliness or inadequacy** as **helpless and/or deserved**	8 (1.5%)	37.5
4	Post intended to portray self-harm, depression and suicidal thoughts as **inescapable**	4 (0.7%)	75
5	**Dramatic description** of trauma, suffering and suicide	22 (4%)	59.1
6	**Traumatizing** imagery	0	-
7	Post that **glamorizes, romanticizes or trivializes** depression, anxiety, self-harm and suicide (including through drawings, comics, use of music and captions)	15 (3%)	60
8	Post conveying **mental health-related misinformation**, e.g. posts dissuading users from seeking professional help, taking prescribed medications or encouraging self-diagnosis based on misleading information	2 (0.4%)	50
9	Post that mentions **plans to self-harm or die by suicide** or **explains how to** self-harm or die by suicide or otherwise **encourages self-harm or suicide**	5 (1%)	60

211. Katarzyna Kostyrka-Allchorne, Mariya Stoilova and others, "Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and nonsuicidal self-injury conditions – a systematic review", February 2023, Child and Adolescent Mental Health, Volume 28, issue 1, (previously cited); North London Coroner's Service, "Regulation 28 report to prevent future deaths", 13 October 2022 (previously cited).

Figure 5 shows the steady increase in the share of potentially harmful mental health-related content over time. Comparing the percentage of manually categorized content and the percentage of videos containing a term from the broad list of mental health related terms shows that the quantitative analysis (solely based on terms) underestimates the actual volume of potentially harmful content, which could pose risks to children and young people with existing mental health concerns.

Figure 5: Progression of manually categorized content on days 1, 6 and 10 of our experiments. We map both progression of content that was classified as harmful (according to our framework) alongside progression of content whose descriptions include terms from the broad list of mental health-related terms

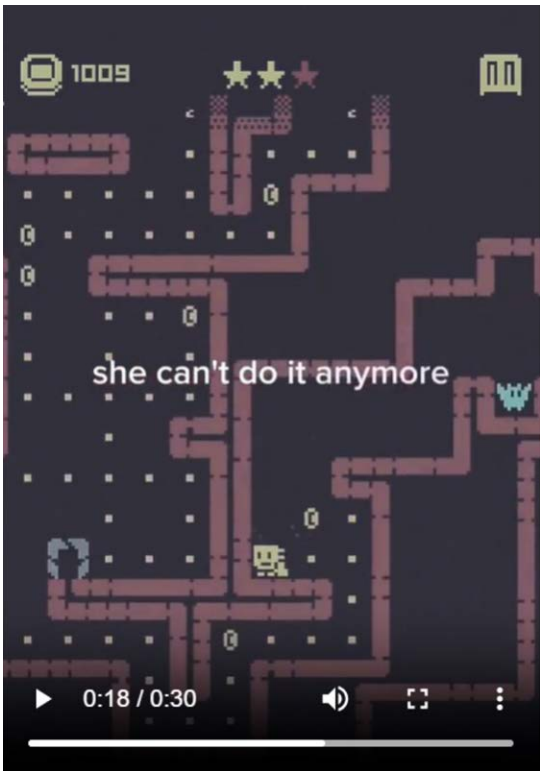


CONTEXTUALIZING THE TERMS-BASED ANALYSIS THROUGH THE MANUAL CLASSIFICATION EXERCISE

To help contextualize the findings from the primary analysis based on terms and hashtags, researchers aimed to understand where content that contained terms from the broad list of mental health terms fell within the manual classification framework. This simultaneously provides an understanding of how much content that is being amplified by the recommender system is potentially harmful, and how much potentially harmful content is missed by the bot accounts and not accounted for in the terms-based analysis because the descriptions do not include hashtags or terms that are included in the broad list.

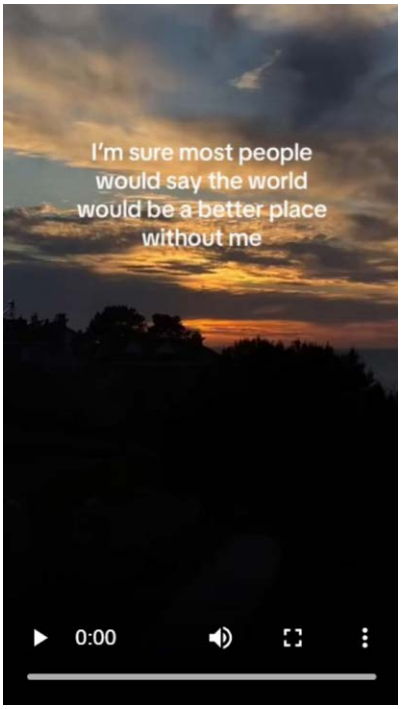
Of content that contained terms from the broad list, 62% was classified as potentially harmful. 18% of the videos that did not contain any of the terms were classified as potentially harmful. Taken together, this suggests that the broad list of mental health terms is relatively good predictor of content that is potentially harmful versus content that isn't, however there remains both: (1) a substantial volume of content (38%) that contains a term from the broad list that isn't categorized as potentially harmful, and (2) a substantial volume of content that is potentially harmful but that doesn't contain a term from the broad list.

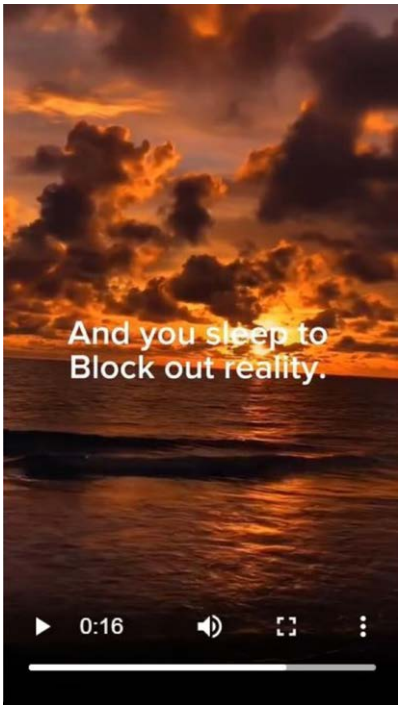
6.7 EXAMPLES OF RECOMMENDED POSTS



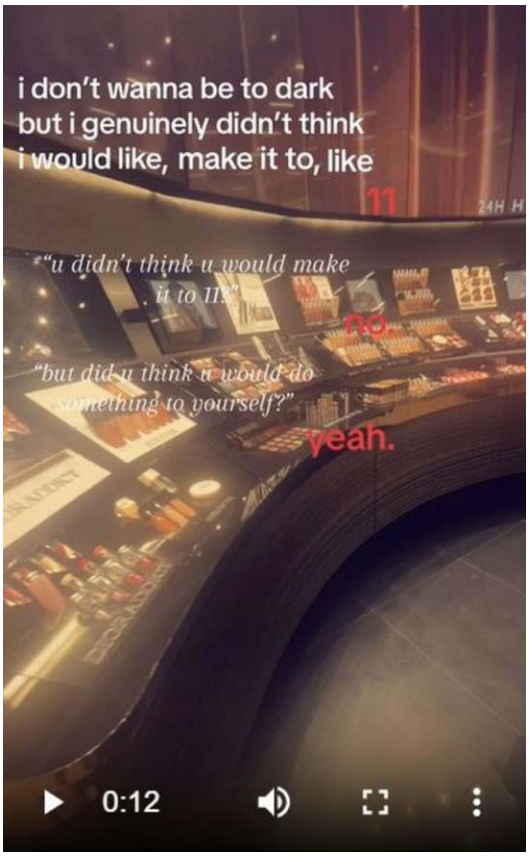
This post describing a girl's suffering and tagged #broken #vent #ventacc #overthinking #emotionallydrained #ventaccount #ventmentalhealth #yoursafeplace #sadtiktok and others was recommended to three US accounts and one Kenyan account.

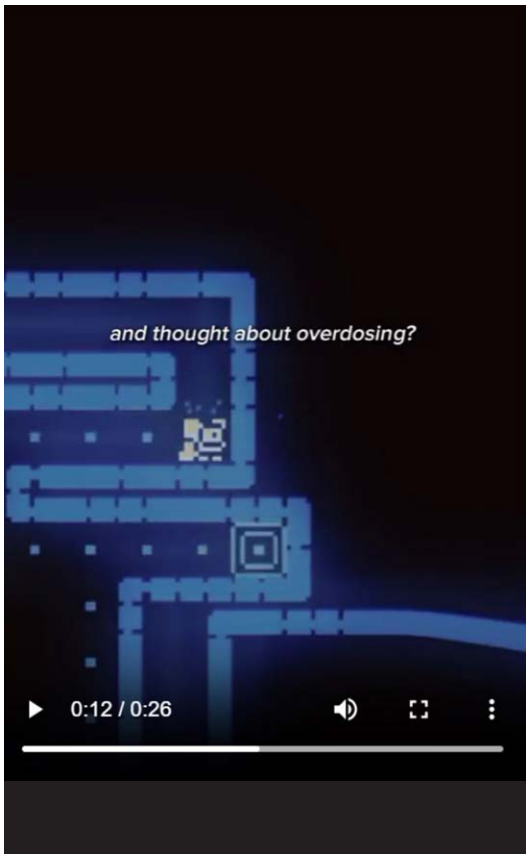
An AI-generated voiceover says: ***"You don't understand how drained she is, yet she still shows up to school every day. Yes, she always has a smile on her face but how do you know she's not struggling. She's so tired and with all the things you boys say to her makes it worse. She goes home crying silently so no one will hear and ask her what's wrong. She can't do it anymore. You don't see the pain she goes through and how hard she's trying to stay alive. You're making it worse with all the shit you say to her, she can't cope much longer. She really does try her hardest."***





These suicidal and depressive posts set against romantic backdrops were recommended to a US account in the depressive behaviour/funnelling sub-group.



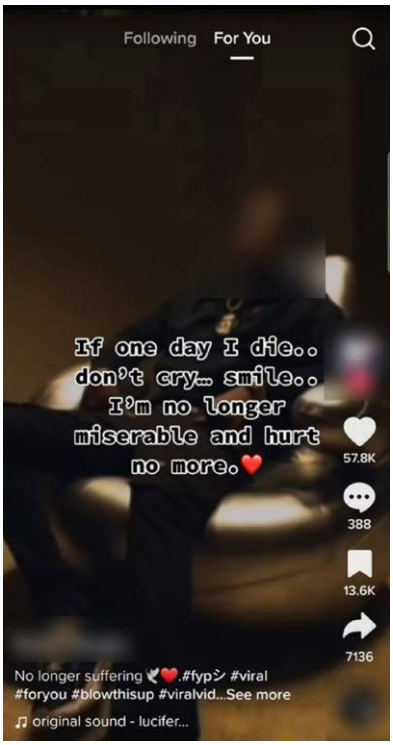


↑←

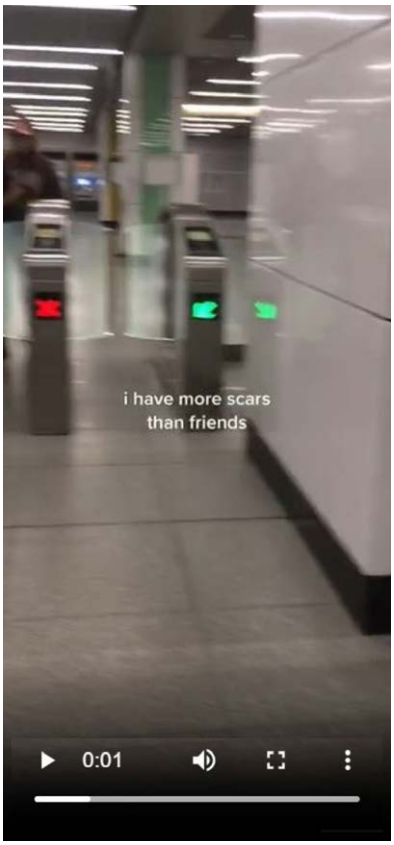
Examples of posts discussing suicide, including through coded language and abbreviations likely intended to avoid detection through automated content moderation ("kys" = kill yourself). Dark roads and arcade games are common backdrops to many of the depression- and suicide-related posts we tracked.

Quote from the video on the bottom left (tagged with #MentalHealthMatters):

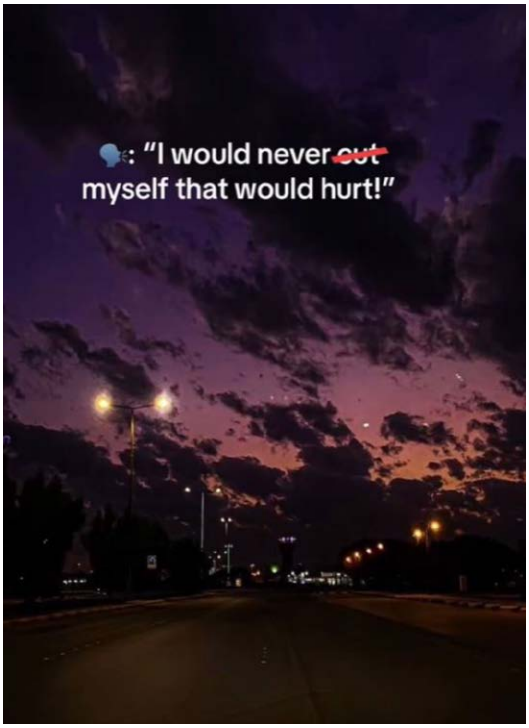
"Have you ever looked at a bottle of pills and thought of overdosing? Have you ever just wanted for it all to end? Have you ever just wanted to die? Have you ever thought about the release it would give, all the pain that would go away? but like I said, just a thought"

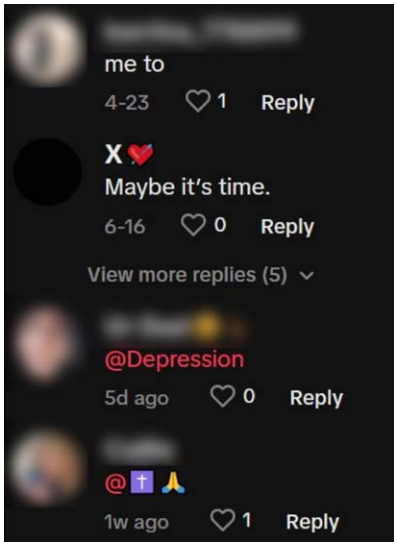


The posts (left) recommended to the teen accounts express suicidal thoughts in ways that romanticize and/or encourage suicide. The video on the right is a film scene overlaid with the creator's comment. It shows a girl in distress, walking into oncoming motorway traffic, shouting "come on!" It cuts out just before the assumed collision. The post was recommended to the manually run US account.



Posts discussing self-harm in response to anger or disappointment over social exclusion. The video on the right appears to have been filmed in London. It was recommended to a Kenyan account between videos from a professional UK mental health content creator. Kenyan accounts were often shown posts by UK- and US-based content creators.





A post trivializing self-harm. The video reuses a common sound, two voices singing the "la la lalalalalala" sound cut to in the second image. The video was recommended to a US account (depressive behaviour/ funnelling sub-group) and tagged with #ventaccount #mentalhealthmatters #ventacc #vent #deepvibes #sadtok #sadtiktok #sadvibes #MentalHealth #nightvibes #chilltok #feelinglonely #trustissues #chilltiktok #iwanttobealone #help. It was no longer available (either removed by the creator or TikTok) when we checked a few weeks after it was recommended to the account.

"I started thinking how everyone's lives would be better without me."

These examples illustrate the kinds of posts that children and young people are likely to be exposed to if they signal an interest in mental health related content. Posts which romanticize depressive and suicidal thinking or trivialize self-harm as well as the above post and comments, which appear to encourage self-harm and suicide, give a sense of the systemic risks found in this study. Many of these posts were amplified by TikTok's recommender system to the point that they attracted hundreds of thousands of views.

Altogether, these findings show that TikTok is not simply failing to appropriately moderate individual pieces of content but that its recommender system actively amplifies such content at scale, at the risk of exacerbating some children and young people's existing mental health issues.

6.8 PROJECT LIMITATIONS

A key limitation of the technical investigation was the exclusive focus on English-language content. Although many of the interviews with children and young adults in Kenya and the Philippines point to English as the main language of interaction with the platform, many young people also post, search for and consume content in other locally spoken languages. Mental health discourse, which is tied to more clinical language, seems to be shared predominantly in English in both countries,[212] but users' posts about their lived experience might also be produced and shared in other languages. Previous reports have flagged that TikTok and Meta dedicate far fewer resources to the moderation of non-English content.[213] It is therefore possible that researchers would have observed higher rates of problematic and harmful content being amplified in languages other than English.

LIMITATIONS OF AUTOMATED AUDITS

As noted earlier in this report, TikTok and other social media platforms have put in place obstacles to independent research, for example by banning automated data collection ('scraping') in their Terms of Service. During Amnesty International's technical investigation, many of the Kenyan accounts were suspended, possibly in connection with the use of a VPN to simulate the location, before they reached the end of our observation period and therefore had to be excluded from analysis.

Independent of these set-up challenges, the core issue of automating accounts to run an experiment like this at scale without exposing researchers to the strains of engaging with this distressing content over prolonged periods is that the extent to which bots can represent human behaviour is limited. The small-scale manual experiment highlighted such limitations; whereas a bot can only react to pre-identified traits of relevant videos (limited in this case to terms and hashtags in the video descriptions), a researcher or user would identify relevant content based on the interplay of descriptions, spoken content, visuals and displayed text, music, sounds and profile name. The automated accounts therefore likely failed to signal interest in many relevant posts, not least because the manual analysis and prior manual exploration showed that some of the most extreme content in relation to self-harm and suicide do not include descriptions or hashtags, often seemingly linking to other relevant content through common sounds instead.

212. This appears to be related to socio-economic as well as to cultural factors. For instance, Tagalog, one of the most widely spoken languages in the Philippines, does not have an exact translation for the word 'depression'. The closest Tagalog term used is the word for 'sadness'. Therapists and their often more affluent patients are more likely to use English in the context of mental health conversations. The local mental health non-profits we spoke to for this project who are active on TikTok equally post in English.
213. Mozilla Foundation, *From Dance App to Political Mercenary: How Disinformation on Tiktok Gaslights Political Tensions in Kenya*, 8 June 2022, https://foundation.mozilla.org/en/blog/new-research-disinformation-on-tiktok-gaslights-political-tensions-ahead-of-kenyas-2022-elections; *Amnesty International, Myanmar: The Social Atrocity* (previously cited).

7. CORPORATE FAILURES

7.1 LACK OF ADEQUATE DUE DILIGENCE

The case of Molly Russell, the 14-year-old British girl who died from an act of self-harm after having viewed depressive content on Instagram, shows in the most tragic way how exposing a young person experiencing depressive symptoms to a social media feed consisting of a high volume of posts that normalize or even romanticize depressive thinking, self-harm and suicide has the potential to exacerbate young users' pre-existing mental health issues and can potentially contribute to harmful and even devastating real-world actions.[214] As previously discussed, the algorithmic recommender systems that TikTok and other social media companies employ to maximize user engagement have the potential to make inferences about a user's thoughts and emotions. The recommender system then responds to these inferred mental states by showing more highly personalized and individually appealing content that it predicts will engage a user with the potential to perpetuate the inferred mental or emotional state and trap users in a feedback loop. In this way, the recommender system has the potential to influence, interfere with or even manipulate a child and young person's thoughts. This represents a far more targeted and personal form of power over children and young people than was previously exerted by traditional mass media broadcasting their content to a broad audience.

Such interference or manipulation constitutes an abuse of the right to freedom of thought and has the potential to exacerbate young users' pre-existing mental health issues, presenting serious risks to physical health and even lives.[215] Given the well-documented emotional vulnerabilities of children and young adults,[216] cases such as that of Molly Russell, and the extensive evidence based on previous civil society and media reports, not to mention the revelations contained in the "Facebook Papers", TikTok should know and have identified that its algorithmic recommender systems risk exposing young users to "rabbit holes" of potentially harmful posts that could exacerbate pre-existing mental health issues.

As discussed in section 4.4, in order to fulfil its responsibilities as laid out in the UN Guiding Principles on Business and Human Rights (UN Guiding Principles), TikTok should be conducting appropriate human rights due diligence to identify, prevent, mitigate and account for how it is addressing its potential and actual harms.

214. BBC News, "Molly Russell: Coroner's report urges social media changes" (previously cited).
215. Katarzyna Kostyrka-Allchorne, Mariya Stoilova and others, "Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and nonsuicidal self-injury conditions – a systematic review", February 2023, Child and Adolescent Mental Health, Volume 28, issue 1, (previously cited); North London Coroner's Service, "Regulation 28 report to prevent future deaths", 13 October 2022 (previously cited).
216. University of Rochester Medical Centre, "Understanding the teen brain", 2023 (previously cited); US Surgeon General, Social Media and Youth Mental Health (previously cited).

In its written response to questions sent by Amnesty International on 12 July 2023 to TikTok, the company did not disclose what specific risks, including to children's rights, it had identified during the design and development stage of creating the app, nor the process that it undertook to identify risks and respective mitigation strategies put in place prior to its launch. Once launched, TikTok should have continued and should be continuing to undertake proactive due diligence.

The company claims in its letter that its approach to youth safety is informed by its commitment to human rights, in particular the UN Guiding Principles "and its call to conduct human rights due diligence". This commitment is included on its website, but TikTok does not have a publicly available human rights policy, beyond this high level commitment to human rights on its website, which is an essential part of human rights due diligence as outlined by the UN Guiding Principles and the OECD Due Diligence Guidance.[217]

Amnesty International followed up in writing to ask TikTok whether the company has such a policy, which covers a broad range of human rights including the rights to privacy, freedom of thought and health, and whether it is publicly available. Researchers also asked who within the senior management of the company is responsible for its implementation. TikTok responded by pointing Amnesty International to the commitment to human rights on its website, which the company wrote is informed by a number of international human rights frameworks including the UN Guiding Principles, which the company has pledged to uphold. However, TikTok did not share a human rights policy as requested by Amnesty International nor information on who at a senior level is responsible for its implementation and how it is embedded from the top of the company through all its functions.[218]

TikTok did state in its written response from 12 July 2023 that it has a centralized team that leads its human rights work, as well as other staff throughout the company who it says are "empowered by leadership to address any identified risks".[219] Amnesty International asked TikTok how large this team is and who on the senior management they report to. Amnesty International also asked how these staff are empowered to address any identified risks, what process is in place when staff identify a risk, and what action they can take. TikTok did not respond to this question.[220]

TikTok's Community Guidelines include a sub-section on TikTok's policies on youth safety and well-being which are overseen by a sub-team in the Trust and Safety Product Policy team. Additionally, according to TikTok, its "Platform Fairness team specializes in issues of human rights, fairness, and inclusion and applies a multifaceted approach to the review and revision of policies, product features, and algorithmic systems."

However, TikTok did not provide details in its letters on what this "multi-faceted approach" entails, nor what specific risks it is looking for or has identified, in particular in relation to its recommender system. It did say that it has embedded a human rights approach across its community guidelines, that it has global advisory councils, which include experts in children's rights, and that it is a member of the WeProtect Global Alliance and the Tech Coalition,[221] where it engages with its peers on current and upcoming risks. In June 2023, TikTok also announced the creation of a Youth Council,[222] which it says will provide "a more structured and regular opportunity for young people to provide their views" on the creation of policies, products and programmes.

---

217. OECD, Due Diligence Guidance for Responsible Business Conduct, https://mneguidelines.oecd.org/OECD-Due-Diligence-Guidance-for-Responsible-Business-Conduct.pdf (accessed 18 July 2023)
218. See Annex 4: Written response from TikTok, 29 October 2023.
219. See Annex 3 to this report: Written response from TikTok, 12 July 2023.
220. See Annex 4: Written response from TikTok, 29 October 2023.
221. The WeProtect Global Alliance and the Tech Coalition are two initiatives involving major tech platforms, which aim to protect children from online sexual abuse and exploitation.
222. TikTok Newsroom, "Updating family pairing and establishing TikTok's Youth Council", 27 July 2023, https://newsroom.tiktok.com/en-gb/keyword-filtering-and-youth-council-update

These initiatives have the potential for greater, ongoing engagement by TikTok with rightsholders, experts, civil society organizations and other stakeholders. However, it remains unclear exactly how this engagement will inform TikTok's due diligence and risk assessments and how frequently stakeholders will be consulted.

In its response to Amnesty, TikTok did not mention the specific risk to the rights to privacy and freedom of thought regarding its data collection practices, profiling and recommender algorithm systems, nor the risk to the right to health of highly personalized content feeds. The company stated in its 12 July letter that it will "in the future […] conduct periodic impact assessments in partnership with third parties, which may take the form of company-wide assessments, team assessments, and product or business line specific assessments." TikTok did not provide details, including when the assessments start, how regular they will be and whether the results will be made public. This contrasts with UN guidance, which calls for "regular comprehensive human rights impact assessments" as a "key element of […] human rights due diligence".[223]

Amnesty International requested further details from TikTok about the company's human rights due diligence policies and practices. TikTok responded that it "consults with a range of stakeholders to inform our human rights due diligence and are in the process of implementing a number of recommendations to our trust and safety operations that have resulted from our engagement with Article One Advisors on human rights. These recommendations are implemented by our platform fairness team in partnership with a human rights working group of colleagues on teams across the company. The assessment recommended that TikTok to conduct a child rights impact assessment, which we will be launching in partnership with Article One."[224]

While it is positive that the company has a human rights working group made up of staff on teams across the company and that TikTok will be conducting child rights impact assessments, the fact that the company has not to date implemented such impact assessments given the huge popularity of their platform with under-18s is a considerable oversight. It is questionable whether TikTok has been able to adequately assess the risks posed to children by its platform without such impact assessments and reflects a failure to carry out due diligence adequately.

Indeed, TikTok states in its 29 October 2023 letter to Amnesty International that another recommendation is to "develop a company-wide human rights due diligence process which will include conducting periodic human rights impact assessments", which the company is in the process of developing and "which will propose the triggers around when we need to conduct an assessment." That TikTok currently does not have a company-wide human rights due diligence process in place is a clear failure of the company's responsibility to respect human rights. It is also not clear what it means by triggers that will show when an assessment needs to be conducted. Given the serious risks to children's human rights posed by the platform, the huge numbers of children that use the platform globally and the importance that it has in many children's lives, it is imperative that the company conducts child rights impact assessments both regularly as part of its due diligence and whenever it implements new design features.

In its written response, TikTok did not include a list of specific risks identified. Given that TikTok is only now in the process of developing a human rights due diligence process that it says will be aligned with international human rights standards, including the UN Guiding Principles, it is unsurprising that the company has not shared a list of specific risks that they have identified or the steps taken to prevent and mitigate these specific risks, as to date the company has not been fulfilling its responsibility to respect human rights by conducting adequate human rights due diligence.

223. UN High Commissioner for Human Rights, Report: The Right to Privacy in the Digital Age, 15 September 2021 (previously cited).
224. See Annex 4: Written response from TikTok, 29 October 2023.

This lack of an adequate due diligence process is inconsistent with the UN Guiding Principles that requires businesses to have "in place policies and processes through which they can both know and show that they respect human rights in practice."[225] The UN Guiding Principles also clarify that showing "involves communication, providing a measure of transparency and accountability to individuals or groups who may be impacted".[226] Another essential element of due diligence as outlined in the OECD Due Diligence Guidance involves communicating how impacts have been addressed.[227] In its letters, TikTok pointed to various measures aimed at promoting children's safety, health and well-being, though some of these do not address the issues covered in this report. Furthermore, as discussed below, many of the measures that they have implemented and which may be intended to address any identified risk of children and young people being sent down "rabbit holes" of harmful mental health content are inadequate.

7.2 GRAVE RISKS MET WITH INADEQUATE RESPONSES

Responding to specific questions about what actions TikTok takes to prevent risks to young people's mental health, its written response to Amnesty International pointed to its Community Guidelines, which set out which types of content are banned and thus, if reported or otherwise identified, removed from the platform.[228] These include a ban on content "showing, promoting, or providing instructions on suicide and self-harm, and related challenges, dares, games, and pacts", "showing or promoting suicide and self-harm hoaxes" and "sharing plans for suicide and self-harm".[229]

However, Amnesty International's manual analysis of post samples in the 'For You' feeds of accounts signalling an interest in mental health found that approximately one in 100 of the recommended videos encouraged self-harm or suicide and a further 3% glamorized, romanticized or trivialized depression, self-harm or suicide. These videos would not just have been available to users searching for such content but were – as our research shows – also actively amplified by TikTok's recommender system.

Although TikTok has never publicly acknowledged the human rights risks associated with creating a highly personalized feed, it has nevertheless introduced tools that seem to be informed by the possibility that such risks exist. In 2021, the company announced that it was working to diversify content recommendations to avoid producing overly narrow feeds, stating that: "At TikTok, we recognize that too much of anything, whether it's animals, fitness tips, or personal well-being journeys, doesn't fit with the diverse discovery experience we aim to create."[230]

In July 2022, it was reported that TikTok was ready to roll out a tool working towards this diversification.[231] In its July letter to Amnesty International, TikTok also said that it "interrupt[s] repetitive patterns, so that content that may be fine if seen occasionally, like extreme fitness or dieting content, is not being viewed in clusters that may be more problematic."[232]

However, Amnesty International's investigation found no evidence that the "rabbit holes" the accounts were sent down were broken up by unrelated content to such an extent as to possibly mitigate the risks. On the contrary, researchers observed repeated peaks in the "rabbit hole" effect over time (see for example figure 3 in section 6.6.1).

225. UN Guiding Principles (previously cited), Principle 21 and commentary.
226. UN Guiding Principles (previously cited), Principle 21 and commentary.
227. OECD, Due Diligence Guidance for Responsible Business Conduct (previously cited).
228. See Appendix to this report: Written response from TikTok, 12 July 2023.
229. TikTok, "Community guidelines", updated in March 2023, tiktok.com/community-guidelines/en/mental-behavioral-health/
230. TikTok, "An update on our work to safeguard and diversify recommendations", 16 December 2021, https://newsroom.tiktok.com/en-us/an-update-on-our-work-to-safeguard-and-diversify-recommendations
231. The Verge, "TikTok is giving you new ways to control your For You page", 13 July 2022, https://www.theverge.com/2022/7/13/23205795/tiktok-algorithm-hashtag-filters-safety-content-levels
232. See Appendix to this report: Written response from TikTok, 12 July 2023.

Another new tool was introduced by TikTok in 2022 to allow users to specify words or hashtags to exclude videos tagged with these terms from their feed, adding to the pre-existing 'Not Interested' button, through which users can signal their wish not to be recommended similar videos. Users have cast doubt on the effectiveness of the 'Not Interested' tool,[233] an observation that was shared by several of the young research participants. They said that they had used the tool in the hope of avoiding further triggering content which they perceived to have a harmful effect on their mental health but found it had limited effects on subsequent recommendations. As "Nikki", the 24-year-old activist in Manila said, "I gave up. I don't click them [various platforms' user control tools] anymore, because although I click, I report, I hide them, it still comes up in my feed eventually".[234]

Amnesty International's investigation did not test the efficacy of the new keyword filter tool, but the large variety of hashtags and terms used by TikTok's 'creator community' for content related to anxiety, depression, types of self-harm and suicide, including deliberate misspellings and code words, would likely limit its effectiveness.

The company's letter of response to our findings from October 2023 also points to the "family pairing" function, which allows parents to customize their child's daily time limit and to make use of the keyword filter tool to stop their child from seeing content associated with specific words in their 'For You' and 'Following' feeds.[235] Such parental controls may add a protective layer, especially if children feel unable to take protective measures themselves, albeit limited by the same challenge as described above of filtering out all the terms that relate to any one topic or type of content. Parental controls do however also create new complexities and may pose risks for the fulfilment of the rights of the child in situations where the best interests of the child, for example to gain access to information about a specific topic, do not align with the beliefs and opinions of the parent.

In its written response to Amnesty International,[236] TikTok also highlighted its 'refresh' function, which allows users to reset their feed and retrain the recommender system.[237] Once again though, Amnesty International's research findings seem to suggest that this function would be of limited use to a child or young person signalling an interest in mental health, given that they would likely retrain the recommender system to amplify potentially harmful content within a short period of time unless they themselves actively skipped all mental health-related content. Moreover, even if they did skip the mental health content, findings from the technical investigation conducted as part of the research suggest that users who show no active interest in such content still receive some recommendations of related content, albeit lower than those who actively signal an interest.

Finally, researchers also tested the usefulness of TikTok's in-app 'why you are seeing this video' transparency tool, which is, according to TikTok, supposed to "bring more context to content recommended in For You feeds".[238] But the information panels researchers were shown as part of the manual experiments only included simplistic and unspecific explanations such as "this video is popular in your country" or "this video is longer, you seem to like longer videos", which do little to explain why a particular video was served to the account or how the personalization effect took hold for the account in question, which saw a pronounced "rabbit hole" effect.

It is clear that more effective user control tools are needed. However, even if these tools were effective,

233. Mashable, "TikTok doesn't seem to care if I'm 'not interested'", 28 April 2022, https://mashable.com/article/tiktok-not-interested
234. Nikki (pseudonym), 24, Manila, interviewed on 8 May 2023.
235. See Annex 4 to this report: Written response from TikTok, 29 October 2023, and TikTok, "User safety: What is Family Pairing?", accessed on 30 October 2023, https://support.tiktok.com/en/safety-hc/account-and-user-safety/user-safety#4
236. See Annex 3 to this report: Written response from TikTok, 12 July 2023.
237. TikTok, "Introducing a way to refresh your For You feed on TikTok", 16 March 2023, https://newsroom.tiktok.com/en-us/introducing-a-way-to-refresh-your-for-you-feed-on-tiktok-us
238. TikTok, "Learn why a video is recommended For You", 20 December 2022, https://newsroom.tiktok.com/en-us/introducing-a-way-to-refresh-your-for-you-feed-on-tiktok-us

they are isolated protective measures that shift responsibility for creating a safe environment onto the user, including young and vulnerable people who may not be able to actively distance themselves from triggering or harmful content. As such, they constitute an inadequate response to the systemic risks associated with TikTok's recommender system.

TikTok falls short of the expected measures involved in adequate human rights due diligence and, as Amnesty International's investigation shows, fails to prevent the serious risks its platform poses to the mental and physical health of young users with existing mental health concerns.

8 CONCLUSION AND RECOMMENDATIONS

CONCLUSION

TikTok's 'For You' page and the algorithmic recommender system behind it have helped to catapult the platform to its current ubiquity in children's and young people's lives. While marketers marvel at its potential to keep users hooked and TikTok's competitors scramble to emulate the design features driving its success, children and young people are being exposed to a system which turns their psychological vulnerabilities into a means to maximize "user engagement".

Through seamless personalization of users' feeds in record time, TikTok has created a platform that is highly addictive, and is exposing users to serious health risks. For children and young people across the world living with depression, anxiety and other mental health issues, TikTok poses a particular danger, because, as this report shows, its 'For You' feed can quickly send children and young people who signal an interest in mental health down "rabbit holes" of depressive content, including videos which romanticize, trivialize and encourage self-harm and suicide.

Despite the increased scrutiny of social media platforms' algorithmic systems in the wake of the publication of the "Facebook Papers" and the UK Coroner's inquest into Molly Russell's death, TikTok and other social media platforms continue to chase user engagement regardless of the potential harm to users. Instead, they prioritize virality and hours spent by users on the platform over the safety of their systems. In doing so, TikTok and other social media giants have normalized the surveillance and manipulation of users worldwide, undermining and abusing their rights to privacy and freedom of thought.

These abuses can seem abstract and intangible, but they have real life consequences for children and young people around the globe. By turning commercially-driven surveillance into a means to keep eyes on screens, in some circumstances through a flood of depressive and banned mental health content, TikTok is putting young users' right to mental and physical health at risk.

The full extent of these risks is still not fully understood, especially as research evidence and the companies' attention is unfairly skewed towards Europe and North America. However, as TikTok's use continues to grow, each day thousands more children and young people are signing up to a platform that could make their lives less safe.

TikTok must urgently overhaul its data collection and amplification processes and undertake comprehensive human rights due diligence. However, individual actions by a single company are not sufficient to rein in a business model that is fundamentally incompatible with human rights. States must therefore effectively regulate "Big Tech" companies like TikTok in line with international human rights law and standards to protect and fulfil children and young people's rights.

This requires an urgent shift in focus by States, which for too long have focused primarily on requiring platforms to remove illegal or, in the case of more repressive governments, unwanted content from their platforms. Instead, States must tackle the root cause of the proliferation of harmful content online, which is a business model predicated on maximizing engagement at the expense of users' human rights. As global discussions on the risks of artificial intelligence continue, it is critical that all governments adopt and implement regulations to prevent and mitigate the harms of some of the most pervasive algorithmic systems already dominating our digital public spaces today.

RECOMMENDATIONS

RECOMMENDATIONS FOR TIKTOK

- TikTok and other technology companies that depend on invasive data-driven operations amounting to mass corporate surveillance must transition to a rights-respecting business model. As a first step, they must ensure that their human rights due diligence policies and processes address the systemic and widespread human rights impacts of their business models, in particular the right to privacy, the rights to freedom of opinion and thought and the right to health. They must be transparent about the risks, including to human rights, that they have identified and how they have been addressed.
- TikTok must 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.
- TikTok must undertake 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.
- TikTok must 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.
- Human rights impact assessments should be published on a regular basis and should include detailed information on risks and mitigating measures taken with respect to specific countries (especially 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.

- TikTok should 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.
- TikTok must 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, 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 content moderation resourcing across all languages.
- TikTok should create public campaigns and awareness on its platform about the different safety features users can enable. Such campaigns could be promoted to users through various channels such as promoted posts on feeds and in-app notifications encouraging users to learn how to confidently use various safety tools.
- Enable independent researchers to access and review all relevant algorithmic systems and where necessary access data to conduct independent research on systemic risks and human rights harms. The EU's Digital Services Act has set out a research access framework that ought to serve as a model for TikTok and other large social media platforms' research access frameworks elsewhere. Where a lack of regulation does not provide clear guidelines on access criteria and vetting processes, TikTok and other social media companies should seek to establish independent advisory bodies to ensure an effective independent vetting process.

RECOMMENDATIONS FOR STATES

Recommendations for meaningful data protection and platform regulation.

States must:

- 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. This will require enacting and/or enforcing comprehensive data

protection laws in line with international human rights law and standards to prohibit targeted advertising on the basis of invasive tracking practices. These laws should restrict the amount and scope of personal data that can be collected, strictly limit the purpose for which companies process that data and ensure inferences about individuals drawn from the collection and processing of personal data are protected. They should further require that companies provide clear information to their users about the purpose of collecting their personal data from the start and that they do not further process it in a way that is incompatible with this purpose or their responsibility to respect human rights.

- 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 sensitive personal data (for example, recommendations based on watch time and likes which allow for inferences of sensitive information) to personalize ads and content recommendations must 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.
- Regulatory processes must involve meaningful consultation of affected groups, including children and young people, as well as independent experts and civil society organizations.
- Ensure that independent national data protection regulators are established, that their independence is guaranteed in law and that they have adequate resources, expertise and powers to meaningfully investigate and sanction abuses of regulations by social media companies in line with international human rights law and standards. They must be able to ensure independent and effective oversight over platform design as well as the design, development and implementation of algorithmic systems to ensure companies are held legally accountable for the identification, prevention and mitigation of human rights harms linked to such systems.
- Enact or 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.
- Require TikTok and other social media companies to provide age-appropriate explanations to children, or to parents and caregivers for very young children, of their terms of service. For children, these should use clear and simple language, provide transparent information throughout the user process and not only at the beginning, and provide clear explanations of user control choices and default settings. Social media companies should also be required to provide non-textual measures to aid understanding of their terms of service, such as images, videos and animations, and they should be easily contactable for any questions.
- Make the best interests of the child a primary consideration when regulating advertising and marketing addressed to and accessible to children. Sponsorship, product placement and all other forms of commercially driven content should be clearly distinguished from all other content and should not perpetuate discriminatory stereotypes (such as those based on gender, race, age, disability etc.)

Recommendations for human rights due diligence.

States must:

- Require in law that technology companies carry out ongoing and proactive human rights due diligence to identify and address human rights risks and impacts related to their global operations, including those linked to their algorithmic systems or arising from their business model as a whole. Where businesses target children or have children as end users, they should be required to integrate child rights into their due diligence processes, in particular to carry out and make publicly available child rights impact assessments, with special consideration given to the differentiated and at times severe impacts of the digital environment on children. They should take appropriate steps to prevent, monitor, investigate and punish child rights abuses by businesses.

Recommendations for effective remedies.

States must:

- Invest in, encourage and promote the implementation of effective digital educational programmes to ensure that individuals understand their rights, including their right to seek an effective remedy against any data protection, privacy or other human rights abuse, when accessing digital services.
- Guarantee access to effective remedy for human rights abuses linked to the impacts of technology companies, wherever the harms occur, including harms resulting from the operations of their subsidiaries (whether foreign or domestic). Redress mechanisms should be made easily accessible and understandable to enable individuals to file complaints when their rights have been infringed.

ANNEX

1. TECHNICAL RESEARCH SET-UP

For this investigation, we used a relatively recent version of the android TikTok mobile app operating on android devices. Our accounts were automatically operated, and we were intercepting the app's communication with the TikTok servers for our client-side data collection. We then sent the data to the Junkipedia servers, the auditing system built by our technical partner, the Algorithmic Transparency Institute, where we enriched our data collection with extra metadata including audio transcript text of the TikTok videos, engagement data and media data.

We created the US accounts directly on android devices based in the USA. For the Kenyan accounts, we created the accounts in a different country, but we operated them after their creation with a Kenyan virtual private network (VPN). After the creation of the accounts and the launch of the experiments, 9 of the 20 Kenyan accounts were disabled by TikTok before the experiments could conclude. We have excluded these experiments from our analysis.

The account creation process sometimes required solving a captcha, and occasionally required us to verify the email address we used for the account creation. As a matter of practice, we verified the emails for every account we created. All the user's birth dates were set to random days in May 2010. Thus, TikTok made all the accounts private. Occasionally, TikTok would show us a message that our account was set to private, and in many cases it would also include a pop-up named 'Privacy Highlights for Teens' with a set of videos related to TikTok's data collection and use.

2. LISTS OF TERMS USED IN THE AUTOMATED EXPERIMENT

Broad Mental Health Interest:

alwayshereforyou❤❤
anxiety
bipolar
blade
blowup
bpd
broken
canileave
cantfightanylonger
childhoodtrama
cry
cryforhelp
crying
darkvibes
deppresesed
deppresiøn
depresion
depressed
depressedtiktok
depressing
depressingslideshow
depression
depressionanxiety
depressionawareness
depressions
dontknowwhattodo
drained
drainedmentally
emotionallydrained
endpoint
harm
heartbroken
hurting
icanttakethisanymore
idkanymore
ihatemylife
imgoinginsane
imnotok
insecurities
ithurts
ithurtssobad
itnevergetsbetter
itsalwaysmyfault
iwannacry
jarofinsecurities

justsadtbh
justventing
lifehurts
lifeisunfair
lifesucks
løsergrp
meantalhealthawareness
mental
mentalbreakdown
mentalhealth
mentalhealthawareness
mentalhealthawarness
mentalhealthmatters
mentalhealthtiktoks
mentalillness
mentallydrained
mentallyill
mentallyunstable
newsreportervoice
newsreportervoicesad
nobodycares
notok
notokay
overthinking
pain
painhub
paintok
panicattack
ptsd
readytoenditall
relatablevent
sad
sadhub
sadness
sadquotes
sadrelatabletiktok
sadslideshow
sadsound
sadstory
sadstuff
sadtiktok
sadtimes
sadtok
sadvibes
safeplace
safeplaceforeveryone
safespace
scars
schizoslideshows
schooldraining
sh
sharpener
shawareness
shizopost
shizoposting
shscars
shvent
socialanxiety
sotired
stressedout
struggling
suffering
suicide
tired
tiredoflife
tombofmask
tombofthemask
tombofthemaskgameplay
tombofthemaskvent
tompofthemask
totm
trauma
twvent
ursafeplace
vent
vent12677
ventacc
ventaccount
ventart
ventartist
ventdrawing
ventedit
venting
ventingaccount
ventingfeelings
ventingpost
ventingthoughts
ventmentalhealth
ventout
ventpost
ventpost
ventposting
ventpoststocryto
vents
venttiktok
venttok
venttome
venttomeifneeded❤
ventttttt
ventvideo
ventviral
ventwithme
webothinalotofpain
whendoesitend
yoursafeplace

Narrow Mental Health Interest:

blade
canileave
cantfightanylonger
crying
depressingslideshow
endpoint
icanttakethisanymore
idkanymore
ihatemylife
imgoinginsane
insecurities
itnevergetsbetter
itsalwaysmyfault
jarofinsecurities
justventing
løsergrp
mentalillness
mentallyill
mentallyunstable
readytoenditall
relatablevent
sadslideshow
scars
schizoslideshows
schooldraining
sh
sharpener
shizopost
shizoposting
shvent
sotired
suffering
suicide
tombofthemaskvent
twvent
vent
vent12677
ventacc
ventaccount
ventart
ventartist
ventdrawing
ventedit
venting
ventingaccount
ventingfeelings
ventingpost
ventingthoughts
ventmentalhealth
ventout
ventpost
ventpost
ventposting
ventpoststocryto
vents
venttiktok
venttok
venttome
venttomeifneeded❤
ventttttt
ventvideo
ventviral
ventwithme
webothinalotofpain

General Interest:

adorable
aesthetic
affirmations
agriculture
alaska
alberta
americanview
animal
animallover
animalsoftiktok
arizona
art
ascendingarts
australia
australianshepherd
awesome_earthpics
babiesoftiktok
babyanimals
babychicks
babycow
babycowsoftiktok
babydog
babyfever
babygoat
babygoats
babygoatsoftiktok
babylove
babytok
backpacking
backyardfarm
backyardflock
backyardgarden
backyardpoultry

beach
beautifuldestinations
beauty
beginnergardener
bestfriend
bird
birds
birdsoftiktok
birdtok
bonsai
bordercollie
bottlebaby
boymom
brain
braintraining
bull
bunnies
bunniesoftiktok
bunny
bunnylove
bushcraft
cactus
calf
california
calvesoftiktok
calvingseason
camping
canada
catlife
catlover
catlovers
catmom
cattle
cattledog
cattok
catvideo
chasingwaterfalls
chicken
chickencoop
chickenkeeping
chickenmath
chickenmom
chickensoftiktok
chicks
chicktok
chihuahua
chihuahuastiktok
chocolate
christmas
college
colorado
comedy
consciousness
conservation
construction
containergarden
containergardening
cooking
cottage
cottagecoreaesthetic
cottagegarden
countryboy
countrylife
countryliving
countrymusic
countrymusicug
countryside
cow
cowboy
cowgirl
cowtok
crazychickenlady
crystals
cuteanimals
cutebaby
cutecat
cutecow
cutecows
cutedog
cutedogs
cutepet
cutepets
cutepuppy
dachshund
dachshundpuppy
dachshundsoftiktok
dairy
dancewithturbotax
deer
doggo
doglife
doglove
doglover
doglovers
dogmom
dogmomlife
dogoftheday
dogsofinstagram
dogsofttiktok
dogtraining
dressage
duck
ducks
egg
eggs
energyhealing
england
englishcountryside
enlightenment
entrepreneur
equestrianlife
fairycore
family
fantasticearth

farmanimals
farmer
farmerlife
farmgirl
farmher
farmhouse
farming
farmkid
farmlifeisthebestlife
fish
fishing
fitness
florida
flower
flowerfarm
flowerfarmer
flowergarden
fluffy
fluffycow
fluffycows
fluffycowsoftiktok
foodie
foodies
foodlover
foodtiktok
foodtok
football
ford
forest
foryoupageofficiall
freerangechickens
frenchbulldog
frenchie
fresheggs
fruit
funnyanimals
funnycat
funnydog
funnydogs
funnydogvideos
funnypets
funnyvideo
furbaby
gardendesign
gardener
gardenersoftiktok
gardening101
gardeningforbeginners
gardeningtiktok
gardeningtips
gardeninspiration
gardenlife
gardenproject
gardentips
germanshepherdpuppy
germanshepherdsoftiktok
girlswhohike
glacier
glaciernationalpark
goat
goats
goatsoftiktok
goattok
godisgood
goldenretrieverlife
goldenretrieverpuppy
goodmorning
goodthing
gooutside
granola
granolagirl
granolatok
grasspuppies
grasspuppy
greenhouse
greenscreenvideo
growfood
growth
growyourown
growyourownfood
handmade
harvest
hawaii
health
healthylifestyle
healthyrecipes
hen
hens
herbalism
herbs
highlysensitiveperson
hike
hikersoftiktok
hikertok
hiketok
hikingszn
hikingtiktok
hilarious
history
hobbyfarm
homedecor
homegrown
homesteader
homesteadinglife
homesteadlife
horror
horsegirl
horselover
horseriding
horseshow
horsesontiktok
horsetok
horsetraining
housecow

pots
poultry
pregnant
puppies
puppiesoftiktok
puppydog
puppylife
puppytiktok
puppytok
pygmygoats
quakerpregrain
rabbits
rabbitsoftiktok
rain
ranch
ranching
ranchlife
reactivedog
reasonforbooking
relateable
relationshipadvice
rescue
rescueanimals
rescuedog
river
roadtrip
rodeo
rooster
roostersoftiktok
rural
sausagedog
scotland
scottishhighland
seeds
seedstarting
selfsufficient
seniordog
sheep
showcattle
silkiechicken
silkies
silly
slowliving
smallbusiness
smallfarm
snow
snowstorm
springgarden
springtime
springvibes
stallion
storytime
summer
sunrise
sunset
switzerland
teachersoftiktok
tennessee
texas
thruhike
tiktoktravel
toddlersoftiktok
tomato
tractor
travelbucketlist
traveltiktok
traveltok
tree
trees
treework
truck
uk
universe
upstateny
urbangarden
utah
vacation
vanlife
vegetablegarden
veggiegarden
visitengland
visitmontana
walk
wanderlust
washington
washingtonstate
water
waterfall
waterfallhike
waterfalls
weather
westcoast
wild
wildanimals
wildflowers
wildlife
wildlifephotography
winter
wisdom
womenwhohike
wood
woods
woodworking
work
workingdog
wyoming
yoga
yummy

3. TIKTOK'S WRITTEN RESPONSE OF 12 JULY 2023

TikTok response page 1

TikTok

July 12th, 2023

Dear Ms. Abdul Rahim,

Thank you once again for the outreach and for continuing the conversation started earlier this spring with the team at Amnesty Tech. We appreciate the time you've taken to develop this questionnaire related to children's rights in the digital environment and have collected and consolidated responses to the items highlighted in your letter. To that end, we would like to share more about TikTok's efforts across each of the areas highlighted in the questionnaire:

DUE DILIGENCE

Our approach to youth safety is informed by our larger commitment to human rights, including the United Nations Guiding Principles on Business and Human Rights and its call to conduct human rights due diligence. Our teams, which are composed of youth safety policy experts, proactively assess human rights risks related to young people. We also work to embed a human rights-based approach across all our Community Guidelines. We have Global Advisory Councils, which include experts in children's rights and we are members of the WeProtect Global Alliance and the Tech Coalition where we engage with our peers on current and upcoming risks. Furthermore, our recently announced Youth Council will serve as an additional sounding board to create policies, products, and programs in tune with the needs of our users. We also work with industry experts, non-governmental organizations, and industry associations around the world in our commitment to building a safe platform for our community. We collaborate with organizations in different regions to share best practices, create programs, and exchange ideas on safety-related topics. We regularly consult with external stakeholders and partners in this area and will continue to update you on our work in future.

We believe that human rights are integral to the work of all teams at TikTok. At TikTok, thousands of people are focused on helping to make our platform safe for our community to explore entertaining content and share their creativity. We have a centralized team that manages our efforts across the company, and we have champions across the organization, empowered by leadership to address any identified risks. TikTok's policies on youth safety and well-being, as articulated in this sub-section in the Community Guidelines, are developed by our Trust & Safety Product Policy team, and are overseen by a sub-team that specializes in youth safety and well-being issues. This sub-team closely collaborates with the Product Policy team responsible for account-level enforcement. The Platform Fairness team specializes in issues of human rights, fairness, and inclusion and applies a multifaceted approach to the review and revision of policies, product features, and algorithmic systems. In the future, we will conduct periodic impact assessments in partnership with third parties, which may take the form of company-wide assessments, team assessments, and product or business line specific assessments. In our human rights risk assessment process, gender is one of many considerations that has been identified to

have an impact on the risk level, and we design our mitigation measures to address such considerations.

TikTok embeds safety-by-design principles, which means that we implement a safety-first approach in the design of the platform and that user safety is embedded as a priority throughout the product and feature development decision-making processes. TikTok's goal is to provide young people with an experience that is developmentally appropriate and helps to ensure a safe space for self-exploration. TikTok is only for those aged at least 13, or 14 in certain jurisdictions. In the US, we offer a curated, view-only experience for those under age 13 that includes additional safeguards and privacy protections. We work to design tools and policies that promote a safe and age-appropriate experience for teens 13-17. In relation to product design, we take several steps including: (1) limiting access to certain product features, (2) utilizing Content Levels that sort content by levels of thematic comfort, (3) using restrictive default privacy settings, and (4) making content uploaded by accounts registered to users under 16 ineligible for the For You Feed (FYF).

Our policies prohibit content that may put young people at risk of exploitation, or psychological, physical, or developmental harm. We have many policies to promote youth safety on the platform. Below are the risks we have identified, and what is appropriate/inappropriate for our teen users.

NOT allowed

- Sexual exploitation of young people, including child sexual abuse material (CSAM), grooming, solicitation, and pedophilia
- Physical abuse, neglect, endangerment, and psychological abuse of young people
- Trafficking of young people, promotion or facilitation of underage marriage, and recruitment of child soldiers
- Sexual activity of young people
- Nudity or significant body exposure of young people
- Allusions to sexual activity by young people
- Seductive performances by young people
- Consumption of alcohol, tobacco products, and drugs by young people

Age-restricted (18 years and older)

- Cosmetic surgery that does not include risk warnings, including before-and-after images, videos of surgical procedures, and messages discussing elective cosmetic surgery
- Activities that are likely to be imitated and may lead to any physical harm
- Significant body exposure of adults
- Seductive performances by adults
- Sexualized posing by adults
- Allusions to sexual activity by adults
- Blood of humans and animals
- Consumption of excessive amounts of alcohol by adults
- Consumption of tobacco products by adults

For You Feed ineligible

- Any content created by an under-16 account
- Moderate body exposure of young people
- Intimate kissing or sexualized posing by young people

If we become aware of youth exploitation on our platform, we will ban the account, as well as any other accounts belonging to the person, and make reports to law enforcement/NCMEC as necessary.

We also offer several tools and controls to support our community's safety and well-being, such as the guides within our Safety Center, which focus on our approach to safety, privacy and security on TikTok. To specifically support our younger community on TikTok, we have developed a Youth Portal, which offers both in-app tools and educational content for our younger users to enjoy their best possible experience.

DATA COLLECTION

Our privacy policies set out, among other things, what data we collected as well as how we use or share such data from TikTok users. Please also refer to this article in our help center for information about our recommendation system. TikTok does not use "sensitive personal data," as the term is defined under GDPR, to personalize content. Nor does TikTok use data collected from users and machine learning to draw inferences about protected characteristics beyond gender and age-range. TikTok does not sell user's personal information or share user's personal information with third parties for purposes of cross-context behavioral advertising where restricted by applicable law.

At TikTok, we take special care when crafting the experiences teens have on the platform, including the ads they see. Currently, some teens may be shown ads based on their activities on and off TikTok, such as the accounts they follow, the videos they like, and their profile information. On June 28, 2023, we announced that we are restricting the types of data that can be used to show ads to teens by region. This means that people in the United States aged 13 to 15 will no longer see personalized ads on TikTok based on their activities off TikTok and people in the European Economic Area, United Kingdom, and Switzerland aged 13 to 17 will no longer see personalized ads on TikTok based on their activities on or off TikTok. We're continuing to work toward providing all people on TikTok with transparency and controls so they can choose the experience that's right for them.

TikTok's Anti-Discrimination Ad Policy prohibits advertisers from using our ads products to discriminate against people unlawfully. Accordingly, advertisers may not include any unlawfully discriminatory or harassing content in their advertising or any content that encourages unlawful discrimination or harassment. In addition, advertisers may not use audience selection tools to: (a) wrongfully target specific groups of people for advertising in a way that breaches applicable laws or regulations; or (b) wrongfully exclude specific groups of people from seeing their ads, in breach of applicable laws or regulations.

In order to prioritize our users and ensure a positive experience on our platform, we do not allow advertisers to unlawfully target or exclude users based on the following categories, including without limitation by using Custom Audiences or the tools and audiences made available on our platform:

Categories advertisers may not use for discriminatory ad targeting where unlawful:

- Legally protected classes based on the local laws of the region, such as race, ethnicity, age, familial status, and sexual orientation
- National identity, country of citizenship, country of origin, or veteran status or identity or beliefs in regards to political groups, religion or union affiliations
- Personal, financial, or legal hardships
- Individual health statuses or disabilities, including mental, physical, genetic, or emotional health and conditions

When advertisers use TikTok's advertiser tools (e.g., to display advertisements on their own websites and applications), TikTok prohibits those advertisers from transmitting certain types of sensitive information back to TikTok. Section 2.8 of TikTok's Business Products (Data) Terms prohibits advertisers from sharing with, or enabling TikTok to collect, "Business Products Data that you know or ought reasonably to know is from or about children or that includes health or financial information, or other categories of sensitive information (including any information defined as sensitive or special category data under applicable laws, regulations and applicable industry guidelines. TikTok restricts advertisers' use of lookalike audiences based on its Anti-Discrimination Ad Policy (referenced above) and we have policies that restrict specific ad categories to 18+.

AGE GATE AND MINIMUM AGE APPEALS

TikTok has a 12+ rating in the App Store, which lets parents use device-level controls to block people under the age of 12 from downloading the app. To help keep people from using TikTok if they're not yet old enough to do so, we've designed a neutral, industry-standard age gate that requires people to fill in their complete birthdate to discourage people from simply clicking a pre-populated minimum age. For accounts banned or restricted because we believe the account holder is under a particular minimum age, the account holder can appeal. See here and here for more information. TikTok continues to investigate industry standards and best practices when considering other options for users to submit age information.

ALGORITHMIC RECOMMENDER SYSTEMS, HELP FEATURES AND ACCESS TO MENTAL HEALTH-RELATED CONTENT

As a platform used by millions of people in the US and more than 1 billion people around the world, we're committed to protecting our community every day. TikTok removes content that violates our Community Guidelines and offers features that help people explore TikTok safely, including:

- Redirecting searches linked to terms like #eatingdisorders or #suicide to prompt people to view support resources, such as helplines along with information on how they can seek assistance.
- Enabling people to refresh their feed if they feel what they are seeing is no longer relevant to them
- Using keywords to tailor their feeds to avoid potentially triggering content

TikTok cares deeply about the well-being of our community members and wants to be a source of happiness, enrichment, and belonging. We welcome people coming together to find connections, participate in shared experiences, and feel part of a broader community. We work to make sure this occurs in a supportive space that does not negatively impact people's physical or psychological health. To accomplish this, we work with our internal experts, and external partners, such as Digital Wellness Lab, Crisis Text Line, Butterfly Foundation, and the International Association for Suicide Prevention, to shape our approach to mental health content. Research shows that content related to mental health impacts different people in different ways. Nonetheless, TikTok puts limits on certain types of mental health content that can appear on the platform

We want TikTok to be a place where people can discuss emotionally complex topics in a supportive way without increasing the risk of harm and we also want to ensure that TikTok encourages self-esteem and does not promote negative social comparisons. Our Community Guidelines have a range of policies that are designed to ensure that emotionally complex topics can be discussed in a supportive way without increasing the risk of harm. For example, we do not allow showing, promoting or sharing plans for suicide, nor do we allow showing or promoting of disordered eating or any dangerous weight loss behaviors. We also take steps to age restrict content related to certain topics, like cosmetic surgery, that may promote negative social comparison in younger users. We also interrupt repetitive patterns, so that content that may be fine if seen occasionally, like extreme fitness or dieting content, is not being viewed in clusters that may be more problematic.

TikTok uses a combination of machine and human moderation to identify suitable candidates for effective signposting of resources and evaluates these search terms in accordance with TikTok's policy frameworks. We also actively solicit feedback from external organizations, such as the International Association for Suicide Prevention, Samaritans, Comentzar de Nuevo, and others to identify emerging terms that could benefit from resource signposting.

As part of TikTok's commitment to safety, education, and uplifting our community and partners, we launched a Mental Health Media Education Fund and donated over $2 million in ad credits to organizations working on supporting mental well-being, including:

- **Alliance for Eating Disorders (**@allianceforeD**)** - National Alliance for Eating Disorders is a nonprofit organization providing education, referrals, and support
- **American Foundation for Suicide Prevention (**@afspnational**)** - American Foundation for Suicide Prevention, Saving lives + bringing hope

- **Crisis Text Line (**@crisistextline**)** - Crisis Text Line provides free, 24/7 mental health support. Text TIKTOK to 741741
- **Made of Millions (**@madeofmillions**)** - Made of Millions is a global advocacy nonprofit on a mission to change how the world perceives mental health
- **National Alliance on Mental Illness (**@nami**)** - National Alliance on Mental Illness helps Americans affected by mental illness
- **National Eating Disorders Association (**@neda**)** - NEDA supports those affected by eating disorders, and serves as a catalyst for prevention, cures and access to quality care
- **Peer Health Exchange (**@peerhealthexchange**)** - Peer Health Exchange provides youth with support, resources, and education to make healthy decisions

As part of this initiative, we're also hosting a series of TikTok training sessions to equip our partners with the tools they need to share information with their communities during critical moments, such as World Mental Health Day in October or back-to-school season. This collaboration represents just one part of our continued efforts to advocate for positive mental health and reach people in need of support, and we're grateful that nonprofits and advocacy groups choose TikTok as a platform to share their knowledge and to reach a wide audience.

Encouraging Supportive Conversations
To accompany our Media Education Fund, we also launched a #MentalHealthAwareness hub for our community to easily learn about well-being topics, connect with advocates, and support organizations that provide important resources.

#MentalHealthAwareness Creator Spotlight
TikTok is a vibrant and welcoming place, enabling creators and the wider community to share their personal stories. Whether they're advocating for more open discussion about depression and anxiety or sharing tips on how people can manage body or self-esteem issues, creators in the #MentalHealthAwareness community help foster open, honest, and authentic conversations. During Mental Health Awareness Month, we spotlighted 10 creators who use TikTok to educate the community on #MentalHeathAwareness and have made a significant impact both on and off the platform over the past year.

In closing, we want to reiterate TikTok's commitment to protecting all members of our community, especially our younger users. Our continued dialogue with Amnesty International is critical as TikTok works to build trust and improve our overall approach to providing young people with an experience that is developmentally appropriate and helps to ensure a safe space for self-exploration

We thank you for your questions and appreciate the opportunity to provide additional details as needed.

With warm regards,
TikTok Trust & Safety

4. TIKTOK'S WRITTEN RESPONSE OF 29 OCTOBER 2023

TikTok response page 1

October 29, 2023

Lauren Dean Armistead, Head of the Children's Digital Rights Team (Interim)
Rasha Abdul Rahim, Director of Amnesty Tech
Michael Kleinman, Director of Silicon Valley Initiative
Amnesty International
1 Easton Street
London, WC1X, ODW
United Kingdom

Dear Ms. Armistead, Ms. Abdul Rahim and Mr. Kleinman,

Thank you for letter dated October 12, 2023, in which you invite TikTok to respond to Amnesty International's research reports regarding TikTok's corporate responsibility to respect human rights in relation to children and young people's use of the TikTok platform. We appreciate the opportunity to address this important topic and reaffirm our deep commitment to protecting the human rights, safety and well-being of people under the age of 18 on the platform. To that end, we would like to share more about TikTok's efforts across the themes highlighted in your report findings:

Privacy and Advertising Policies

At TikTok, the privacy and security of our users is among our highest priorities. We take our responsibility to safeguard people's privacy and data security seriously. In line with industry practices, we collect information that users choose to provide to us and share with the broader TikTok community, as well as information that helps the app function, operate securely, and improve the user experience. We detail the information we collect in our privacy policies. There are a number of inaccuracies about our practices described in the reports that we would like to clarify.

Report 1's assertions centered on TikTok's data collection practices do not accurately describe TikTok's privacy practices, nor the platform's capabilities. The TikTok app has its own built-in search engine functionality and does not directly collect what people search for outside of the app or through other search engines. TikTok does not collect precise geolocation in the US, European Economic Area, UK and many other regions. In regions where we do collect precise geolocation, we obtain consent prior to collection and people can revoke this consent at any time.

In addition to the inaccuracies described above, the assertions about LGBTQ+ content made by the *The Wall Street Journal* and referenced in the report are incorrect, as we said at the time. TikTok does not identify individuals or infer sensitive information such as sexual orientation or race based on what they watch. Additionally, as we explained to the *The Wall Street Journal*, watching a video is not necessarily a sign of someone's

identity. There are many reasons someone may engage with content; there are allies who engage with LGBTQ+ content but may not identify as LGBTQ+ themselves. There are people who enjoy baking content but aren't bakers. There are people who watch sports content and aren't athletes. People come to TikTok to discover new, entertaining content.

When we build products and features for our platform, we do so by keeping privacy in mind and building in privacy principles throughout the product development lifecycle. We also believe it's important to ensure strong protections to help keep minors and young people safe, which is why we've introduced privacy features and tools to support age-appropriate experiences on our platform. In our previous letter and above, we provided an overview of the global default protections, tools and guidance we have implemented to protect teen privacy and keep them safer on the platform. We will continue this important work.

Report 1 also inaccurately describes our ads practices. First, TikTok prohibits advertisers from using our ads products to discriminate against people unlawfully. We provided our Anti-Discrimination Ads Policy in our prior letter, which describes our stance on discriminatory ads. The report inaccurately characterizes our enforcement of this policy. All advertisements on TikTok are subject to our Community Guidelines and Advertising Policies. As a result, advertisements are not permitted if they violate TikTok's policies. Second, we implemented restrictions regarding the types of data that can be used to show ads to teens by region in an announcement in July of 2023. This policy has been implemented and we will continue to move toward providing our community with transparency and controls so they can choose the experience that's right for them.

Human Rights

TikTok is committed to respecting the human rights of all people, especially community members between the ages of 13-17. Our commitment to human rights, available on our website, is informed by several international human rights frameworks which we have pledged to uphold. These include the International Bill of Human Rights (which includes the Universal Declaration of Human Rights [UDHR], the International Covenant on Civil and Political Rights [ICCPR], and the International Covenant on Economic, Social and Cultural Rights [ICESCR]), (2) the International Labour Organization [ILO] Declaration on Fundamental Principles and Rights at Work, (3) the Convention on the Rights of the Child [CRC], and (4) the United Nations Guiding Principles on Business and Human Rights [UNGPs].

TikTok consults with a range of stakeholders to inform our human rights due diligence. For instance, we are implementing a number of recommendations to our trust and safety operations that have resulted from our engagement with Article One Advisors on human rights. These recommendations are implemented by our platform fairness team in partnership with a human rights working group of colleagues on teams across the company. The assessment recommended that TikTok to conduct a child rights impact

assessment, which we will be launching in partnership with Article One. Another recommendations was to develop a company-wide human rights due diligence process which will include conducting periodic human rights impact assessments. In partnership with Business for Social Responsibility (BSR), we are developing this human rights due diligence toolkit which will propose the triggers around when we need to conduct an assessment. This toolkit and corresponding processes are aligned with international human rights standards, most notably the UNGPs.

Finally, as your report correctly indicates, we have embedded a human rights approach across our Community Guidelines and have advisory councils around the globe, which include experts in children's rights. We are members of the WeProtect Global Alliance and the Tech Coalition. TikTok recently announced the formation of our Youth Council, which will enable us to listen to the experiences of those who directly use our platform and be better positioned to make changes to create the safest possible experience for our community. We've been working to build the council with youth representing a diversity of backgrounds and geographies and will keep Amnesty Tech informed as work progresses.

Teen Safety and Mental Health

TikTok is committed to ensuring the safety and well-being of our teenage community members. We strive to navigate the complexity of supporting our community's well-being on our platform with nuance. We take a four-pronged approach that involves removing harmful content, age-restricting or dispersing content that may not suitable for younger members of the community, and empowering people by providing them with tools and connecting them to resources. Our Community Guidelines do not allow content that shows, promotes, or shares plans for suicide or self-harm or content that shows or promotes disordered eating or any dangerous weight loss behavior. We also age-restrict certain topics that may pose unique risks to children, such as videos showing or promoting cosmetic surgery that do not include risk warnings, including before-and-after images, videos of surgical procedures, and messages discussing elective cosmetic surgery.

As the report references, TikTok has developed and implemented systems that limit content related to certain topics that may be fine if seen occasionally, but potentially problematic when presented in aggregate. These systems include coverage for topics like misery, hopelessness, sadness, and diet and fitness. We continue to work on expanding and implementing these systems, including adding more mental health topics.

We regularly consult with health experts, remove content that violates our policies, and provide access to supportive resources for anyone in need, including children. We are mindful that triggering content is unique to each individual and remain focused on fostering a safe and comfortable space for everyone, including people who choose to share their recovery journeys or educate others on these important topics. We have

published a guide for creators with suggestions on how to talk about mental health while keeping themselves safe and being respectful to other community members.

TikTok also offers tools to help parents and younger members of our community manage their screen time. As referenced in your report findings, TikTok automatically sets a 60-minute screen time limit for every account belonging to a user below age 18. Before adopting this feature and choosing this limit, we consulted academic research and experts from the Digital Wellness Lab at Boston Children's Hospital. As your report mentions, teens are able to continue watching after the 60-minute limit is reached by entering a passcode. We have found that enacting more restrictive screen time controls may increase the risk of teens lying about their age, while the passcode feature requires teens to make an active decision to extend their screen time. The interruption introduces friction into the experience, which gives people an opportunity to pause and reflect on whether they wish to continue watching. Additionally, our Family Pairing features allow parents to customize the daily screen time limit for their teens and implement stricter standards if they feel it is needed. We are monitoring the efficacy of the time limit default and are continuing to innovate to make it more effective.

In addition to Family Pairing, TikTok offers a wealth of resources for parents and guardians to help safeguard their teen's safety, privacy, and well-being on the platform, which can be found directly in TikTok's Safety Center. Additionally, TikTok's Youth Portal offers both in-app tools and educational content that empower young users to keep their account secure and limit their online footprint to the degree they feel comfortable. The Youth Portal includes a "You're in Control" video series which provides safety and security tips for all users.

TikTok is committed to upholding human rights and maintaining a safe platform for all members of our community, especially our younger users. We appreciate the opportunity to respond to these reports and welcome a continued dialogue on these important issues.

Sincerely,

Lisa Hayes
Head of Safety Public Policy & Senior Counsel, Americas, TikTok

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

+44 (0)20 7413 5500

JOIN THE CONVERSATION

www.facebook.com/AmnestyGlobal

@amnesty

DRIVEN INTO THE DARKNESS

HOW TIKTOK'S 'FOR YOU' FEED ENCOURAGES SELF-HARM AND SUICIDAL IDEATION

During the Covid-19 pandemic, TikTok emerged as a global platform, attracting hundreds of millions of children and young people largely thanks to its 'For You' page, an infinitely scrollable feed of personalized video suggestions, and the algorithmic recommender system behind it.

Through its seamless hyper-personalization, TikTok has created an addictive platform, despite mounting evidence of the serious health risks associated with children's compulsive use of social media. Examining further risks of TikTok's content targeting, Amnesty International's research shows that TikTok's 'For You' feed can easily draw children and young people who signal an interest in mental health into "rabbit holes" of potentially harmful content, including videos that romanticize and encourage depressive thinking, self-harm and suicide. TikTok risks exacerbating children and young people's struggles with depression, anxiety and self-harm, putting young people's mental and physical health at risk.

TikTok must urgently overhaul its data collection and amplification processes and undertake comprehensive human rights due diligence. However, individual actions by a single company are not sufficient to rein in a business model that is fundamentally incompatible with human rights. States must regulate "Big Tech" companies in line with international human rights law and standards to protect and fulfil children and young people's rights.

INDEX: POL 40/7350/2023
OCTOBER 2023
LANGUAGE: ENGLISH

amnesty.org

AMNESTY INTERNATIONAL