Exploring what children can see on social media in Mexico

Research report

June 2026

REVEALING REALITY

About Revealing Reality
Revealing Reality is an independent social research and insight agency. We take on complex projects with social purpose, using evidence to inform policy, design and behaviour change.

We work with regulators, government and charities to provide rigorous insight into people's online behaviours and experiences. A large part of our work examines how digital services and platforms shape everyday life - influencing relationships, gambling, financial decisions, health and more.

Through detailed qualitative and quantitative research, we build deep understanding of digital behaviours and reveal how people truly experience technology. Visit www.revealingreality.co.uk to find out more or get in touch.

We would like to thank El hilo de Ariadne for its support in carrying out this research.

Glossary

- **Feed:** the main page or continuously updating content stream that displays posts and activity from accounts a person follows or is subscribed to.
- **Input:** in this paper, and within the context of social media, input refers to the different interactions carried out with avatars (for example: liking, scrolling, etc.).
- **Scroll:** refers to the action of moving a finger or cursor up or down the screen in order to view more content in a feed, website, or application.
- **Clickbait:** the practice of using eye-catching headlines or thumbnails, often sensationalist or misleading, to generate clicks on a link or piece of content, even when it does not meet the expectations created.

Executive summary:

What children in Mexico can be exposed to on social media

Social media platforms, including TikTok, Instagram, and X, state that they are committed to keeping children safe online. All require users to be at least 13 years old, and some describe using technology to detect underage users or apply additional protections for users. These protections include privacy settings, content moderation policies, and tools that aim to limit exposure to harmful material.

This research set out to explore how these commitments translate into real experiences for children in Mexico. Using simulated social media profiles - avatars - based on real children's online behaviours, the research examined what content could be surfaced through everyday actions such as scrolling, liking, following, and searching.

The findings suggest a significant gap between platform policies and what children may encounter in practice.

- The avatars were able to access all platforms easily, without encountering meaningful age verification measures.
- In some cases, avatars were shown potentially harmful content - including material related to self-harm, cartel violence, and sexualised imagery - within a relatively short period of engagement.
- Search functions appeared to play a key role in enabling access to specific types of content, even where some terms were blocked or triggered warnings.
- On TikTok, a small number of avatars were also shown coin-based features that appeared to reward extended viewing, although the purpose and impact of these features were unclear.

Executive summary:

What children in Mexico can be exposed to on social media

Five avatars were run for a total of just 60 minutes across a 12-day research period, with five further avatars run across a shorter 2-day window. Despite this limited engagement, the content surfaced included material that could be considered distressing or inappropriate for children. Given that many children spend far longer on these platforms each day, the findings raise important questions about what a real child might encounter over time.

While these observations do not indicate how frequently such experiences occur at scale, they do point to design patterns that may expose children to risk. The research highlights how recommender systems can quickly shift based on user interaction and how safeguards - where present - may be inconsistent or limited in effectiveness.

Executive summary:

What children in Mexico can be exposed to on social media

To reduce the risk of harm and better align practice with stated policy, platforms may need to:

- Assess and mitigate the potential impact of design features that direct children towards risk or encourage prolonged or repeated viewing by children.
- Uphold their own published terms, policies, and community standards.
- Review how recommender systems respond to engagement with emotionally intense or sensitive material.
- Strengthen age assurance processes to better distinguish between adult and child users in order to deliver age-appropriate and rights-respecting experiences.
- Apply content moderation measures more consistently, especially in algorithmically curated feeds.

This report presents evidence from avatar-based research in Mexico to support reflection and action on how platform design affects children’s experiences online.

INTRODUCTION

Research context

Building from Pathways, which explored the experiences of children on social media in Britain, this report aims to explore what children in Mexico can experience on social media and how platform design choices shape those experiences. Using simulated profiles – or avatars – based on real children's behaviours, the research reveals how recommender systems respond to everyday actions like scrolling, liking, following, and searching.

Despite platforms' public commitments to safety, children are easily able to access social media and can quickly be exposed to distressing, graphic, or sexualised content. Features designed to drive engagement appear to amplify rather than limit harm, while safeguards are inconsistent or absent.

By comparing what platforms say they do with what the avatars actually saw, this report highlights a clear gap between policy and practice. It shows how simple inputs can lead to potentially harmful content and reinforces why design matters when it comes to protecting children online.

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

METHODOLOGY

From real children's behaviours and experiences to digital profiles

To understand what children in Mexico can be shown and experience on social media, we used an 'avatar' methodology – creating social media profiles that mimic real children's behaviour online. This approach allows us to explore and test algorithmically driven feeds without putting children at risk. The aim was to test what the algorithm would show a typical child through everyday actions like scrolling, liking, following, and searching.

In-depth interviews were carried out with 12 children across Mexico to understand how they experience the online world.

Interviews were conducted remotely, in Spanish, by a Spanish-speaking researcher alongside a Mexico-based research consultant.

Children were asked about their everyday digital lives – what platforms they use, the kinds of content they see, how it makes them feel, and how they behave online.

The sample included children of different ages and genders, living in a range of locations across the country as shown by the diagram on the right. Some children were included because they had seen particular types of content, or had experiences that directly related to the themes explored through the avatars.

Following the interview, each child shared recordings of their feeds across their most used social media platforms. They also shared the list of accounts they followed on each profile.

These interviews provided the behavioural insight and real-world context that shaped the design of each avatar. Avatars were then based on the profiles and follow lists of those children whose experiences most clearly aligned with the social media themes of the research.

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

METHODOLOGY

Escalating ‘input’ behaviours

Avatars were set up on TikTok, Instagram, and X. All profiles used common, low-friction behaviours – scrolling, liking, following, and searching – to simulate how children typically engage. Interaction inputs were gradually increased to test how the algorithm adapted in response.

Although the avatars were operated by researchers in the UK, they were set up using Mexican SIM cards and accessed via VPNs routing through Mexico, enabling platforms to serve local content to the avatars.

This method revealed how potentially harmful content can be surfaced quickly and easily and how platform design reinforces continued exposure.

While the avatars reflect realistic user experiences, they don’t measure what users will see or are likely to see. They illustrate what can be shown – not what will be shown, how often, or to how many children.

Fieldwork for this project including interviews and the avatars were conducted between October 2024 and February 2025.

Set-up

- Age of child or age they entered on their social media account
- Registered with a ‘dummy’ email address
- Following a randomly selected sample of a child’s follow list

Escalating phases (interaction)

- Passive phase: 4 days of scrolling
- Engagement: 4 days ‘liking’ content and ‘following’ accounts that relate to the theme of the avatar
- Engagement: 4 days ‘searching’ terms related to the theme of the avatar

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

METHODOLOGY

Avatars informed by real children's ages and behaviours

Themes and avatars

Each avatar reflected the behaviours, interests, and engagement patterns of real children in Mexico and was assigned to one of five content themes:

- Mental health and depression – content ranging from low mood to self-harm
- Sexual content – suggestive videos and sexualised imagery
- Violence – including interpersonal violence and cartel-related content
- Eating disorders – extreme dieting tips and material promoting disordered eating
- 'Buchona' culture – an aesthetic that glorifies luxury, usually as a 'narco wife' and hyper-feminine beauty ideals, often sexualised and glamorised in style

During the interviews, each child reported having a social media profile that was older than their real age – often using a false or random date of birth to sign up. These accounts had aged up with them over time. As a result, the children were now using accounts aged 18+. To reflect this reality, two avatars were set up for each theme, with one account at the adult age children entered when setting up their profile and one at the child's real age.

METHODOLOGY

Avatars informed by real children's ages and behaviours

The fieldwork period ran for 14 days, with 12 days using avatars set up with the adult ages and 2 days using avatars set to the children's real ages. This was done to assess whether age accuracy affected safeguards or content recommendations and to reflect the real experiences of children on social media.

The profiles set to the child's real age were run with a rapid, 2-day avatar design. These avatars followed the same accounts as the 12-day avatars but escalated engagement with potentially harmful themes much earlier. This enabled us to explore what platforms showed users with a child's age.

Theme	Age of child	Age of profile	Platform
Sexual content	15	24	
'Buchona'	13	18	
Mental health	15	24	
Eating disorders	13	29	
Violence	14	38	

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

ACCESS

Children can access platforms despite the platforms' own age restrictions

ACCESS

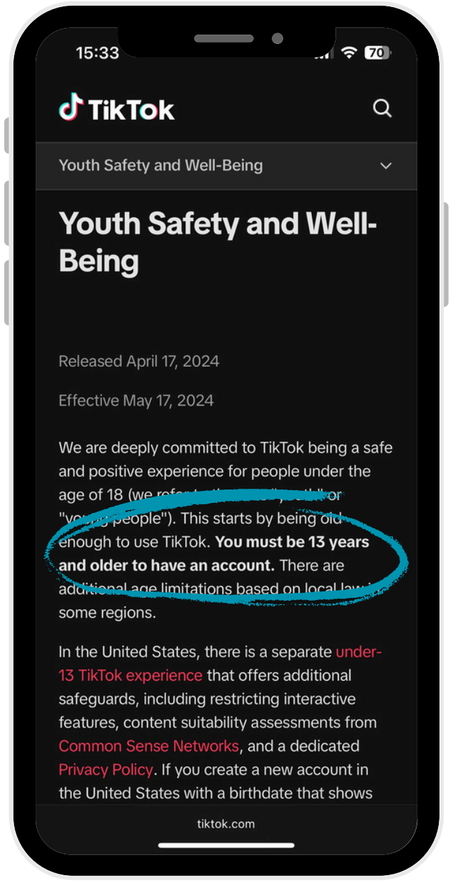
Platforms outline their approaches to age assurance

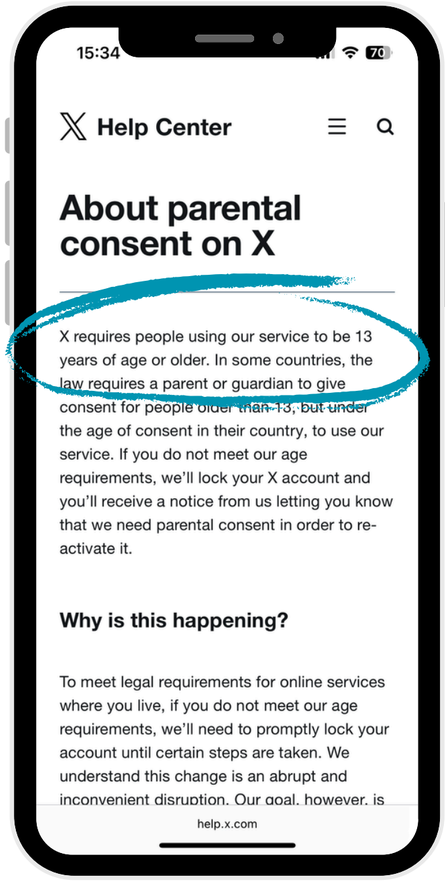
Age assurance can be a foundational part of designing safe, age-appropriate, and rights-respecting online experiences for children. Without knowing who their users are, platforms cannot offer the right protections, apply relevant policies, and design spaces that reflect children's needs. It obliges companies to design experiences that protect, empower, and support children - not just restrict them.

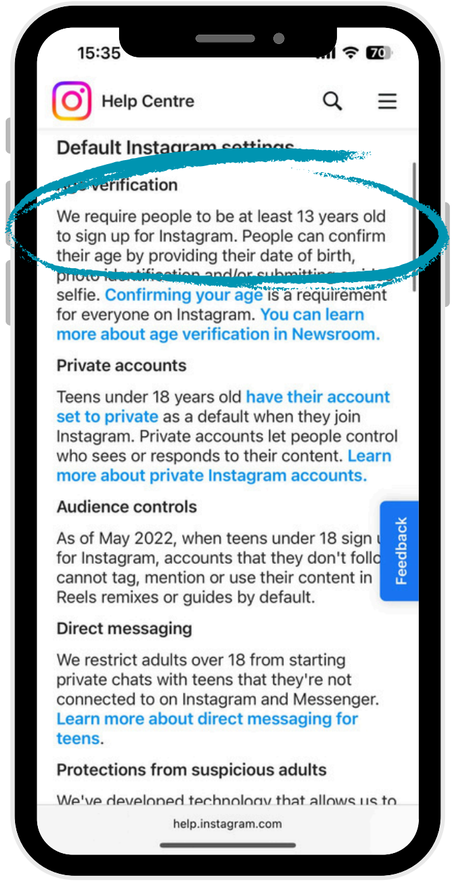
TikTok, Instagram, and X all state that users must be at least 13 years old to sign up.

TikTok and Instagram go further, claiming they use technology to detect underage users and apply special protections for teenagers.

However, this doesn't appear to reflect children's online experiences.







REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

ACCESS

Children’s experiences of account creation

As we learned during the interviews, all of the children used accounts set to an age older than 18, meaning they were officially registered as adults on social media. To reflect this reality, the avatars were first registered at the age the children entered when registering their profile.

Setting up the avatars on all platforms was quick and straightforward. Entering a date of birth – whether accurate or not – was the only step related to age, and no additional verification was required. The platforms did not appear to challenge the age entered, and accounts were granted immediate access to content and features.

If a platform does not know users’ ages, it cannot enforce its own rules or ensure that all users - especially children - receive experiences that respect their rights, such as safety, privacy, and age-appropriate content, by default. It cannot stop children from accessing adult content, nor can it ensure that recommender systems or reward features are safe and appropriate. The result is a system where age restrictions exist on paper, but not in practice – and where children are treated, by design, as adults.

EXPOSURE

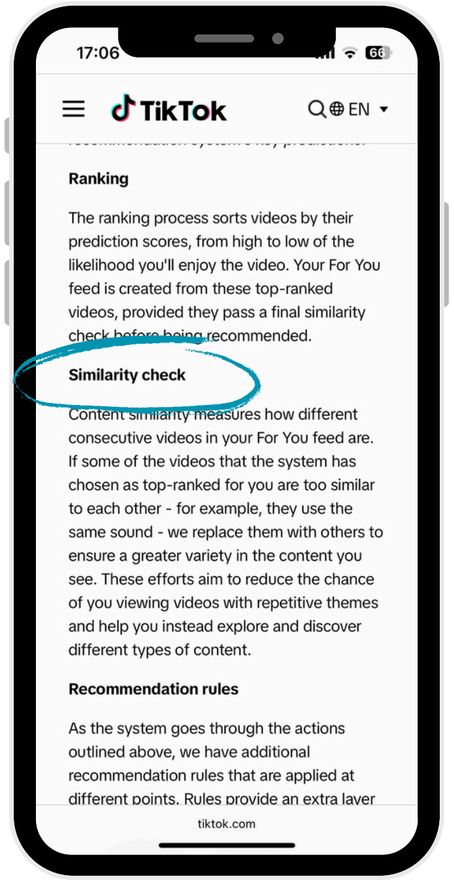
Recommender feeds can rapidly amplify potentially harmful content

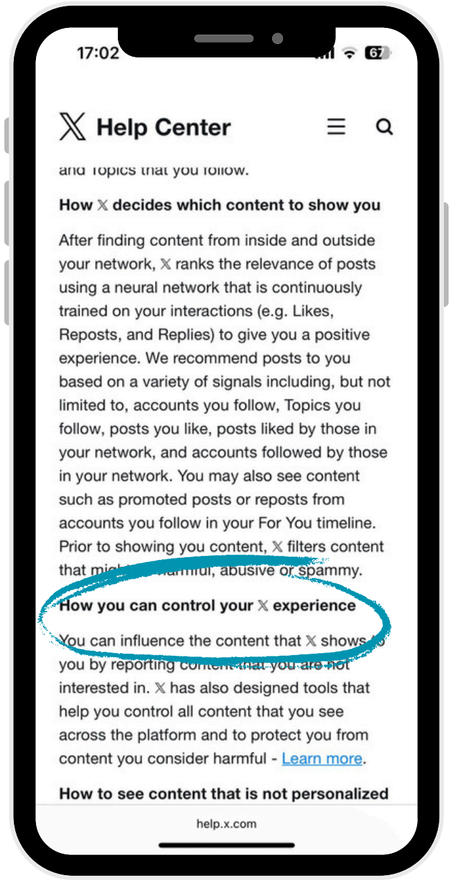
EXPOSURE

Recommender feeds can rapidly amplify potentially harmful content

Social media platforms serve users a personalised stream of content designed to hold attention. Platforms claim recommender systems are moderated to avoid repetition and reinforce wellbeing – but this does not appear to always reflect users' experiences.

While there are tools available across the platforms to encourage users to flag content they 'are not interested in', there do not appear to be effective safeguards built into the platform design. As a result, users can still be fed increasing amounts of potentially harmful content.





EXPOSURE

Recommender feeds can rapidly amplify potentially harmful content

For the avatars in this study, simple interactions like scrolling on the feed, searching, liking pieces of content, or following accounts related to that avatar's theme were enough to shift feeds towards increasingly specific and, in some cases, potentially harmful content.

Simple inputs rapidly accelerated avatars' trajectory to harmful content. 'Liking' content or 'following' profiles – both basic types of engagement on the platform – transformed the algorithm from generic to potentially harmful in just 30 minutes of scrolling and interactions.

To illustrate how content evolved over time, three case studies show how different avatars experienced increasingly specific and potentially harmful content in response to everyday engagement patterns such as scrolling, liking, and searching.

These include:

- **Mental health avatar:** content quickly shifted toward posts referencing depression, hopelessness, and, in some cases, suicide-related themes.
- **Violence avatar:** After early exposure to generic action videos, this avatar's feed shifted rapidly toward graphic content. This included street fights, violent or extreme accidents, and cartel-related violence news.
- **Sexual content avatar:** Light engagement with lifestyle and relationship content led to a feed dominated by sexualised videos and adult-themed material.

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

CASE STUDY: Mental health

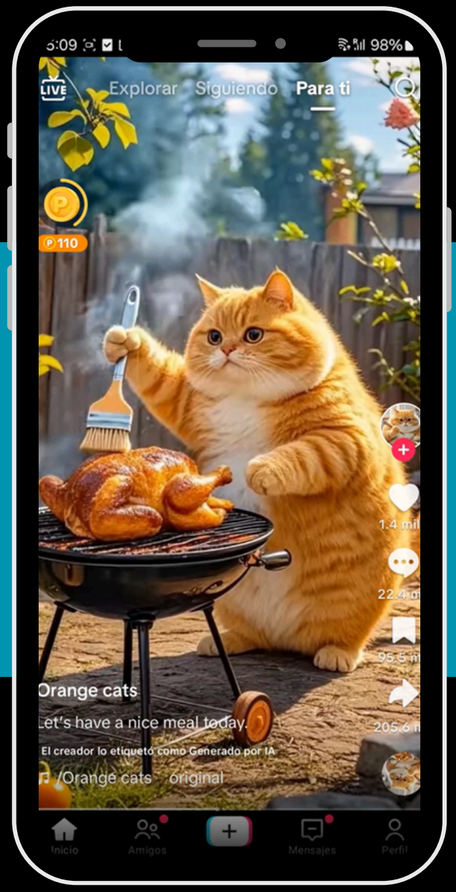
Mental health content on TikTok

This avatar was based on Helena, a 16-year-old girl who described regularly seeing emotionally intense content on TikTok, particularly at night. Her screen recordings showed a feed filled with sad themes, including references to self-harm.

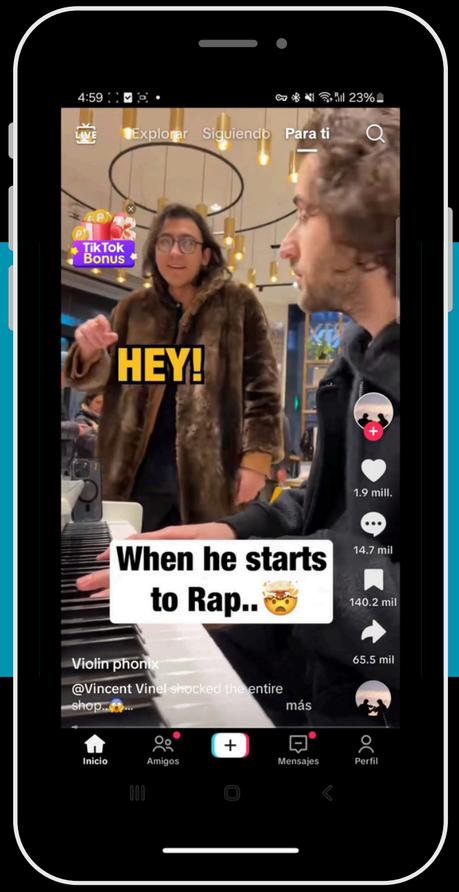
'I go on TikTok and I get videos of nothing but sad things... but I want other things... I mean, I go on TikTok to distract myself, but they come up with more things'.

To reflect her experience, the avatar was set up using the same age she entered when registering for the platform and populated with a sample of the accounts she followed.

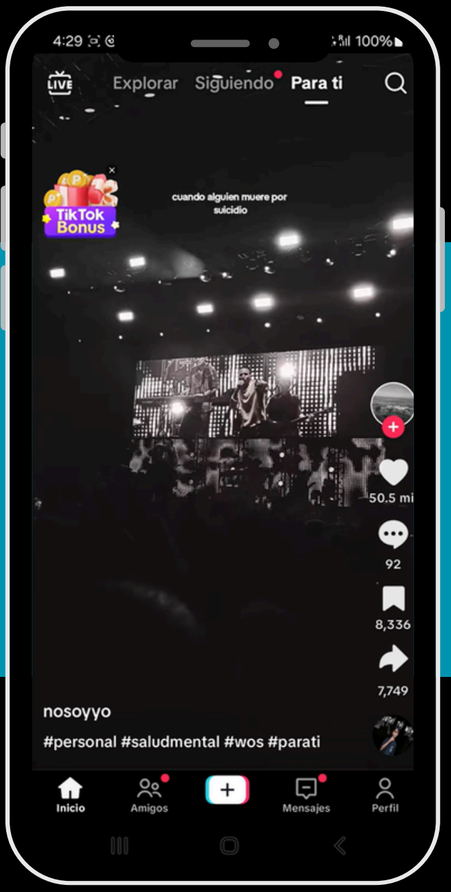
After initial setup and scrolling, the avatar's feed was served a mixture of generic viral videos – including memes, singing clips, and relatable posts about relationships. However, after minimal interaction with emotionally themed content, the feed began to shift. By day 10, almost every video shown related to sadness, anxiety, or hopelessness.



Day 1



Day 2



Day 12: 'When somebody dies from suicide, everyone feels bad, wishes they saw the signs and saw it coming'

CASE STUDY: Mental health

Phase one: 'liking' and 'following'

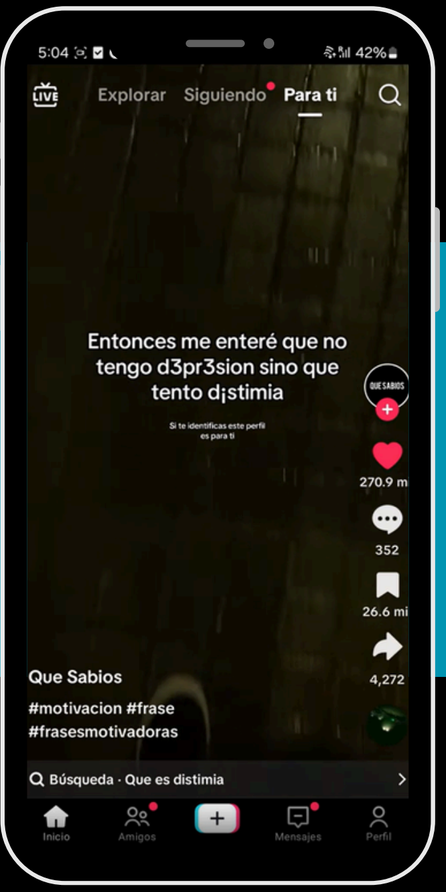
Each day, researchers scrolled for five minutes on the avatar's feed. They liked up to five pieces of content related to mental health or depression, following up to three of the accounts posting that content.

Early examples of emotional content included nostalgic or regretful reflections on past relationships. After a few days of scrolling, these seemed to become more intense, with posts focused on loneliness, low self-worth, and mental distress.

The volume and consistency of emotionally intense content escalated rapidly. Within a few days, the feed had become dominated by content related to depression and rumination.

Seen on Day 5 from an account already followed: 'And he went to sleep calm, while I was drowning in tears'.

Liked on Day 6: 'Me explaining to my nervous system that I just communicated something that made me feel bad and not that I'm being chased by a lion'.



Followed and liked on Day 7: 'So I realised that I don't have depression but dysthymia [persistent depressive disorder]'.

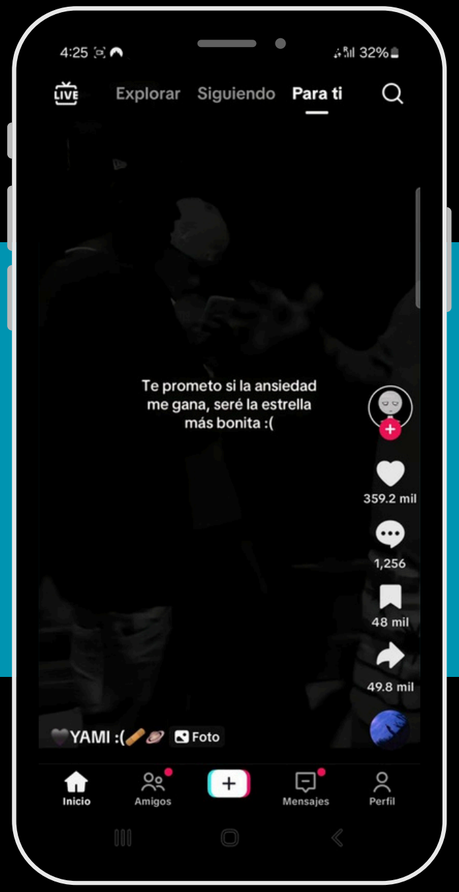
CASE STUDY: Mental health

Phase two: ‘searching’

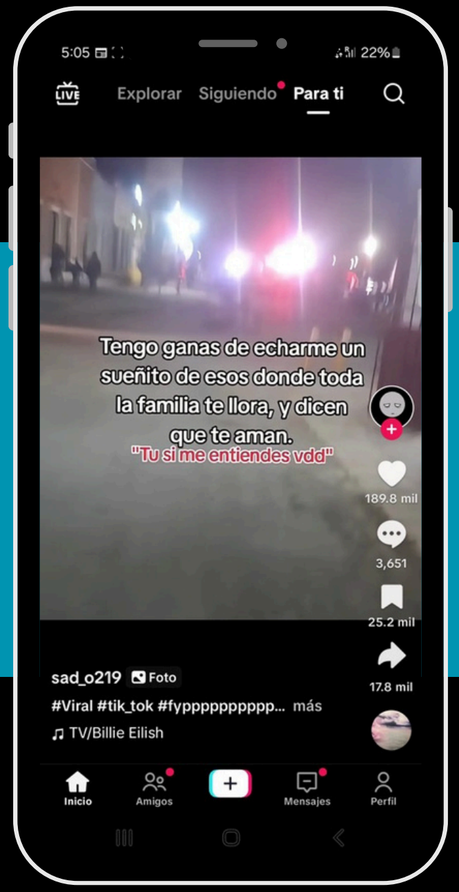
On day eight– with less than 40 minutes of total account activity – the avatar was first served a video explicitly referencing suicide. No warnings, restrictions, or support resources were shown in response.

Subsequent searches for mental health-related terms surfaced further emotionally intense and potentially harmful content. In some cases, content appeared to bypass TikTok’s moderation by using coded language, misspellings, or text overlays. Content included references to self-harm, hopelessness, and ideation, presented without contextual support or intervention.

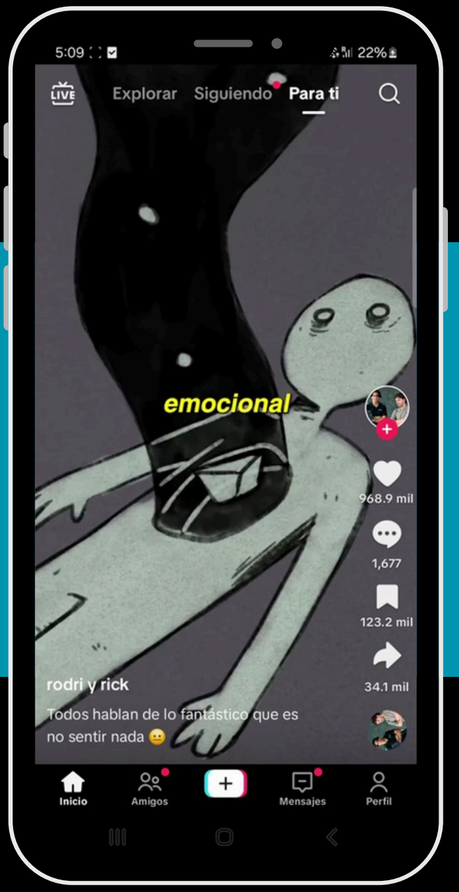
By the final days of the testing period, the avatar’s feed was almost entirely dominated by depressive content, including posts that normalised self-harm and suicidal thinking.



Day 8: 'I promise you that if anxiety beats me, I’ll be the most beautiful star'.



Day 9: 'I have the urge to have one of those naps where your family cries and tell you that they love you “you truly understand me”'.



Day 11: 'Everyone talks about how wonderful it is to feel nothing, but no one talks about how frustrating it is to be in an emotional block where you can't connect with anyone at all'.

CASE STUDY: Sexual content

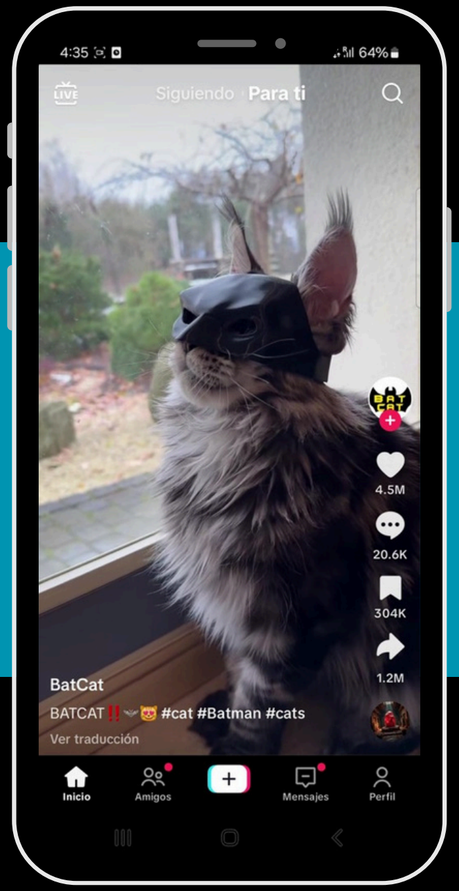
Sexual content on TikTok

Javier is a 15-year-old boy who described regularly seeing sexual content on his social media feeds.

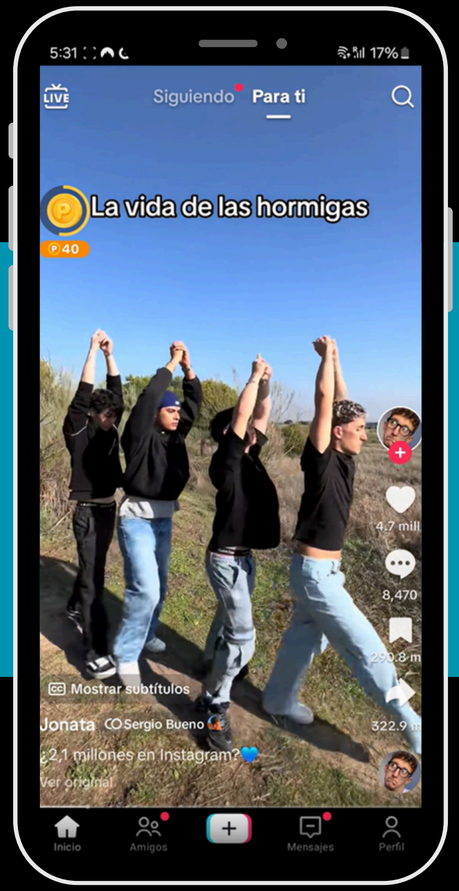
'I mean, no, there's no porn, but like women dancing with little clothing or things like that.'

This avatar was set up as the age the child had entered when setting up their own account and was populated with a combination of their follow list, as well as profiles seen on recordings of their TikTok feed.

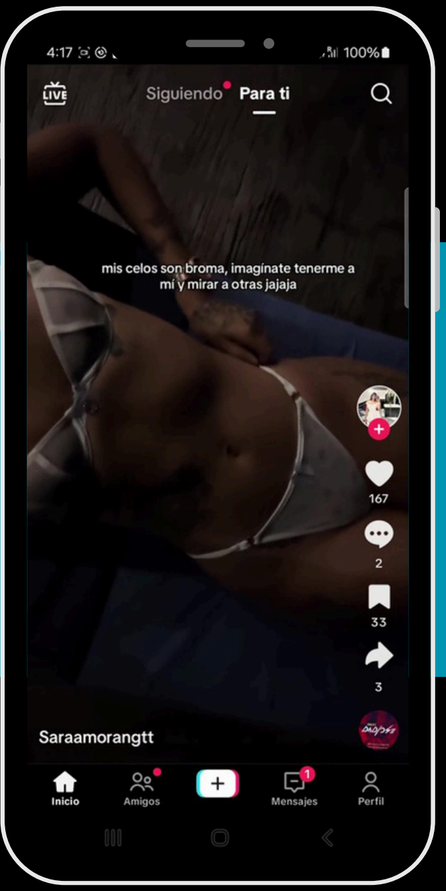
While this profile was initially served generic and viral content, like cats dressed as Batman, after minimal input, the avatar began regularly seeing references to sex acts and sexualised content. This includes instances of TikTok profiles directing users off the site towards private and explicit content on subscription sites.



Day 1



Day 3: 'The life of ants'.



Day 12: 'My jealousy is just a joke... Imagine having me and looking at other people hahaha'.

CASE STUDY: Sexual content

Phase one: 'liking' and 'following'

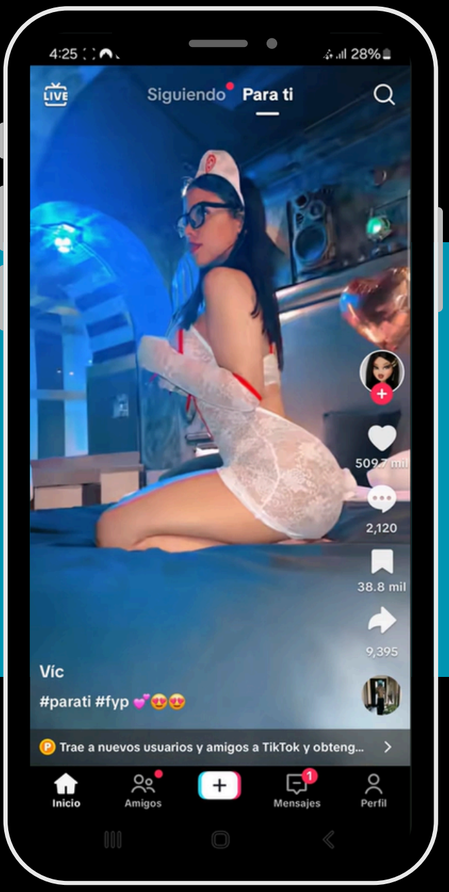
After the initial following of accounts, the avatar's feed began to be served content related to Javier's descriptions. This included 'women dancing with little clothing'.

On Day eight, the avatar watched 30 videos. Analysis of these pieces of content showed that 10 of the videos were posted by profiles promoting and directing users to explicit private members-only subscription platforms (e.g. OnlyFans).

While TikTok does not allow access to members-only websites through the in-app browser on the app, this can be bypassed by opening the link in a browser or by accessing it via another app, such as Instagram.

Liked from an account already followed on Day 6

Liked from an account followed on Day 7: 'They're saying you like them like me - naughty and likes to be spoiled'.

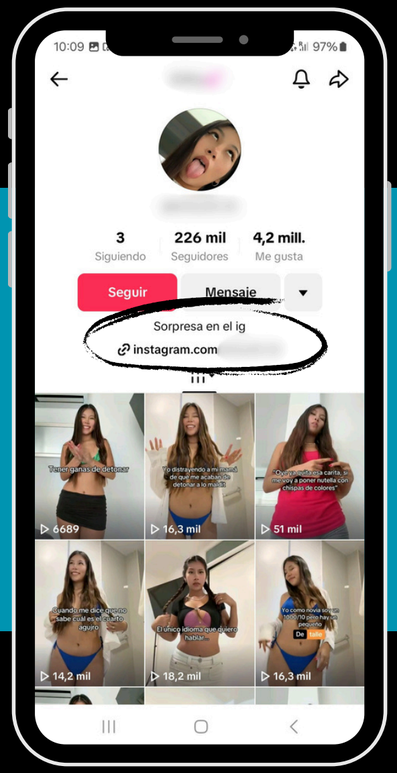


Seen on Day 8: 'Ok but me in my nurse uniform 💕'.

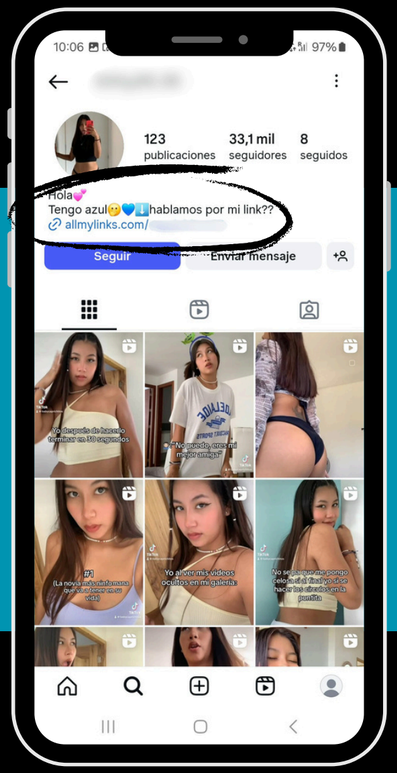
CASE STUDY: Sexual content

Journey from TikTok to pornography

*Liked and followed on Day 7: 'Girl, I'm on my way back, you'll stay with my dad right? Immediately me: *sexual moaning audio*'.*



Clicking on to this profile, users are directed to Instagram, with their bio reading 'surprise on my IG (Instagram)' with a link.



On their Instagram account, their bio reads, 'I have a blue [referencing OnlyFans]. We can talk through my link'.

This link takes you to one of two private and explicit adult content pages, promoting 'Hard Sex', 'Deep Throat', 'Facials', and 'Anal' content.

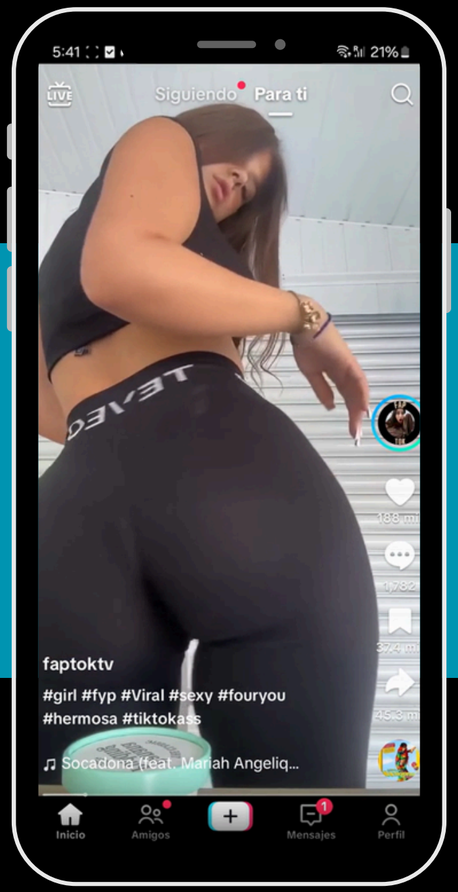
CASE STUDY: Sexual content

Phase two: ‘searching’

Searching for terms such as 'bikini' and 'belleza' led to more sexual content being shown on the avatar’s feed. This included videos of imitated sex acts and sexualised audio, such as moaning.

TikTok’s Community Guidelines state the platform restricts content containing ‘sexual behaviour’ and ‘sexual framing’. It also states that this content is 'ineligible for the FYF [for-you-feed] if it shows intimate kissing, sexualized framing, or sexualised behaviour by adults.'

With minimal engagement on TikTok, children can be quickly exposed to sexualised content and accounts promoting paid adult material.



Day nine: A video from ‘faptoktv’ [referring to masturbation] including sexual framing.

*Day 11: 'Girl, I’m on my way back, you’ll stay with my dad right? Immediately me: *sexual moaning audio*'.*

Day 12: 'Us girls who can do 'little circles' on the tip, are blessed - and I’m not talking about tongues'.

CASE STUDY: Violence

Violent content on X

This avatar was based on the experience of Gabriel, a 14-year-old boy who described seeing violent content on X.

'They're accounts that find exclusive videos that no one else can find, and they upload them for other people to see... they have a lot of dead people, a lot of blood, and a lot of violence'.

An avatar account on X was set up, registered as the adult age the child had used for their own account. There was no robust age verification in place.

Initially the account's feed was largely Mexican and American news stories. By day 3 - after just 15 minutes of activity on the account - the avatar was shown the uncensored corpse of an executed celebrity. During the fieldwork period, the avatar's feed became dominated by violent themes and graphic imagery.

Day 1

Day 2

Day 3: 'The Colombian singer Zair Guette and his manager Teddy vergara were murdered in rural Ginebra'.

CASE STUDY: Violence

Phase one: 'liking' and 'following'

After escalating the 'input' for the avatars, including 'liking' posts and 'following' related accounts, the avatar was served almost exclusively violent content. This included videos of people being hurt or killed in accidents, as well as violent imagery associated with gang-related or cartel activity.

For the remainder of the fieldwork period, the avatar's feed was filled with similar content at an intense volume. This was primarily delivered by news aggregator accounts or journalists posting about cartel activity.

Seen on Day 4: Video showing an interrogation of a cartel member by gunpoint, with the video ending as a knife is put to his throat prior to decapitation.

Seen on Day 8: Video of boys as young as 15 firing guns, having been recruited to fight in a cartel war.

Seen on Day 8: 'Rats learn from shock' - video showing young men posing with guns before footage of them being beaten by authorities.

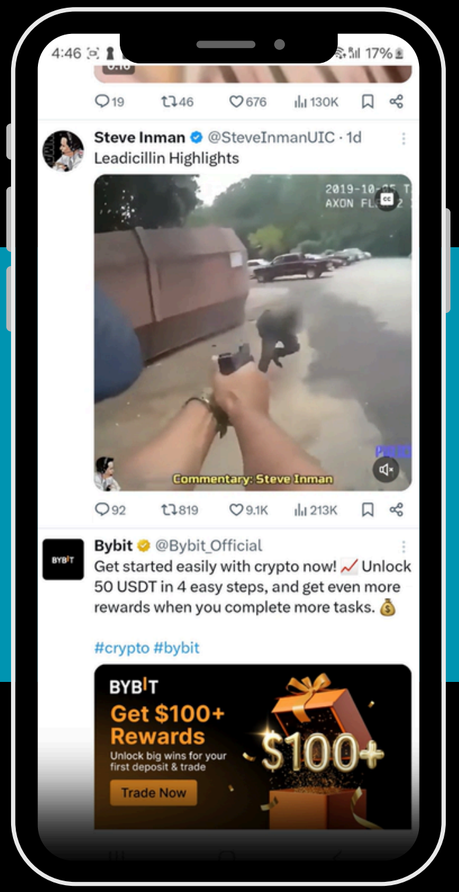
CASE STUDY: Violence

Phase two: ‘searching’

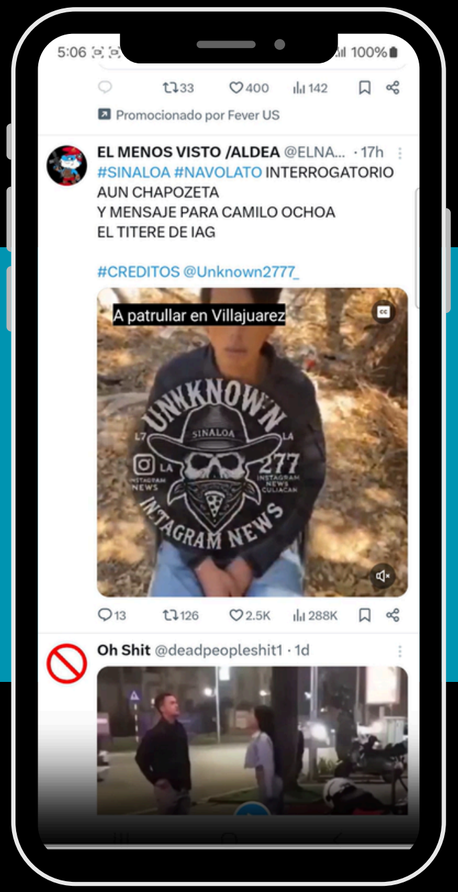
Searching terms such as 'Sinaloa' or 'enfrentamientos' increased the avatar’s exposure to violent content, and continued serving this content in the avatar’s feed. While journalism and news coverage served most of the graphic imagery before, this was later served by accounts dedicated to ‘violence’ or ‘gore.’

According to X’s Rules, violent content is allowed on X, as long as it is 'properly labeled, not prominently displayed and is not excessively gory.' X’s Guidelines for Minors also outline that the platform aims to restrict 'certain types of sensitive media' to 'known minors [aged 13-17]' - although it does not outline clearly if this includes violent content of any kind.

The day 2 child avatar showed that children are able to be exposed to this content at almost the same intensity as the adult avatar.



Seen on Day 11: Bodycam footage of a police officer shooting an individual.



Seen on Day 12: Footage of an interrogation.

Seen by the child-aged avatar on Day 2: video and image of a decapitated ‘sicario’ hanging from a bridge

EXPOSURE

How search functions surfaced harmful content

In some cases, harmful material only appeared after a relevant term was searched. These search terms were identified based on tags and terms seen in screen recordings shared by children during interviews, and further informed by advice from a Mexico-based consultant.

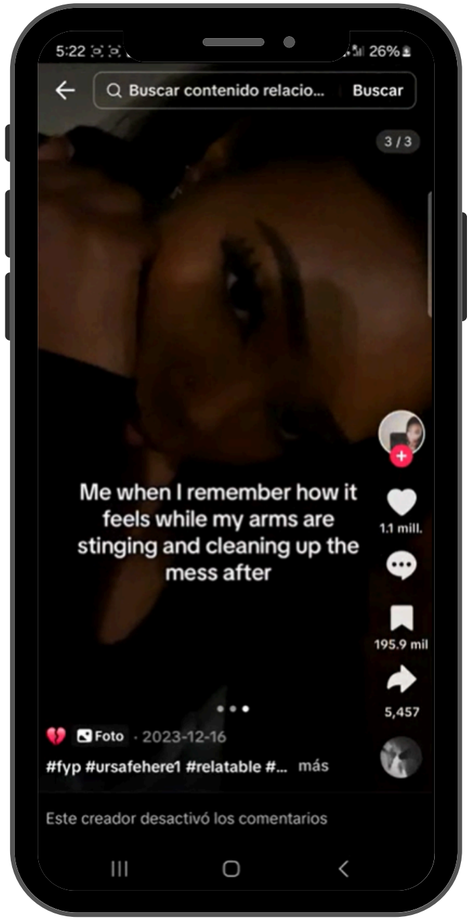
While the self-harm and suicide avatar was regularly served content about deep rumination and depression, content about self-harm and suicide was most clearly found through searching for the theme. The avatar based on a child who had previously seen self-harm content was able to access explicit references to suicide after searching terms like ‘su1cui’ and ‘$H’ – identified as common terms that bypass platform moderation filters.

The avatar exploring exposure to content relating to eating disorders was only exposed to related content through direct searches, such as ‘bodygoals’.

On X, the child-aged avatar exploring violence was regularly shown violent content. This included being able to see an uncensored video of the aftermath of an execution in Sinaloa after searching ‘enfrentamientos’.

When avatars searched for terms related to each theme, they were often shown potentially harmful content.

While some terms, such as ‘bodygoals’ and ‘chicas sexys’, triggered safety warnings or were blocked – particularly for child-aged accounts – the majority returned content with no restrictions or prompts.



Searching ‘SH’ on the adult-aged account.

EXPOSURE

How search functions surfaced harmful content

To reflect the real-world behaviour of children using adult-aged accounts, most avatars were initially set up using the ages children had inputted when signing up. Additional profiles were set up for each avatar, using the child's real age to observe what platforms show when the user is set up as a child.

In some cases, avatars with child-age profiles were blocked from searching specific high-risk terms – especially those linked to eating disorders or sexual content. However, self-harm content and emotionally intense material were still accessible and, in some cases, actively recommended.

This suggests that while some protections may be in place for underage users, they are inconsistent and limited in scope.

The child-aged avatar was not shown content when searching 'chicas sexys'.

Searching 'bodygoals' did not show reels relating to eating disorders to the adult or child-aged avatar, but did show non-reel content on the For You alongside a warning.

Searching 'mental health' on TikTok triggered a warning on both the adult and child aged avatars.

REVEALING REALITY · Instituto Nacional de Salud Pública · 5RIGHTS FOUNDATION

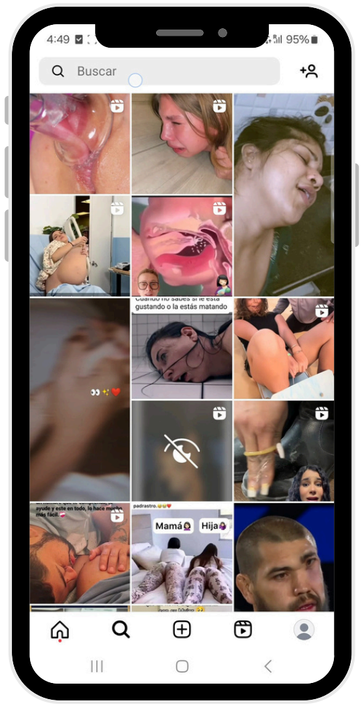
EXPOSURE

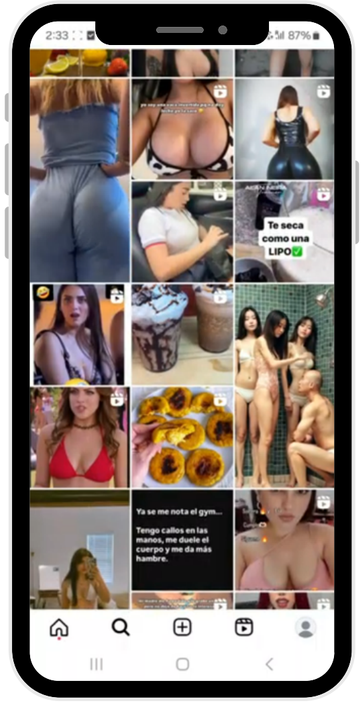
AI, clickbait and suggestive content on explore pages

Content on the explore pages in the later days on some avatars appeared to be designed to capture attention – particularly through thumbnails that gave the impression of more explicit or distressing content. After watching the videos, the content often turned out to be more benign.

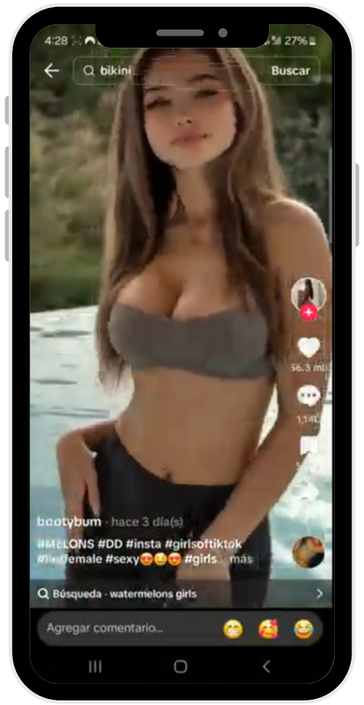
Alongside content displayed on the explore page, search terms such as 'belleza' led to sexually suggestive AI-generated videos being served to avatars. While not explicit, the videos included women in bikinis and some hyper-sexualised poses.

AI-generated content also appeared in adverts for the child-aged avatar. For example, a video of a woman losing half their body weight, going from ‘fat to skinny’ appeared as the top result after searching ‘version fit’.

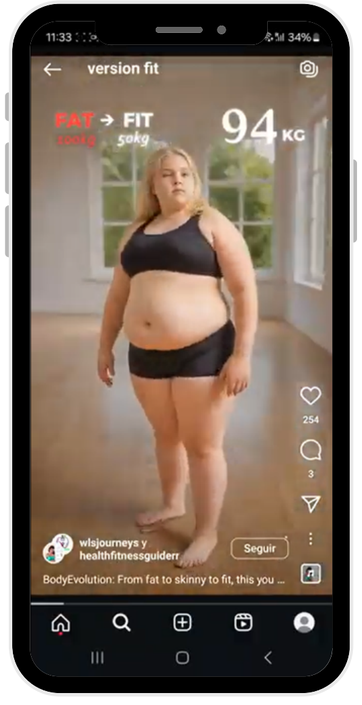




Clickbait and AI content seen by an avatar profile on Instagram.



An AI generated video shown to the sexual content avatar by searching ‘bikini’.



The child-aged avatar was exposed to an AI weight loss video searching ‘version fit’, where the generated girl loses half their weight.

INCENTIVES

Design features appeared to incentivise users to keep watching

INCENTIVES

Features that appeared to incentivise users to keep watching

Social media platforms are primarily designed to maximise engagement. Their business models depend on keeping users active for as long as possible – encouraging repeated use, sustained attention, and emotional connection. This often means prioritising features that drive access and engagement.

On TikTok, this focus on engagement appears in design choices – such as an algorithmically driven feed or an infinite feed of content – that promote constant scrolling, instant feedback, and algorithmic content recommendations. While these features may support discovery or entertainment, they can also shape what users see and how long they stay – especially children, whose online behaviours and emotional responses may be different from those of adults.

Several of the adult-aged avatars on TikTok were shown features that appeared to reward time spent watching videos. These included overlays and pop-ups offering small amounts of 'coins' if the user continued watching the video; for example, 'earn 10 coins for every 30 seconds watched'.

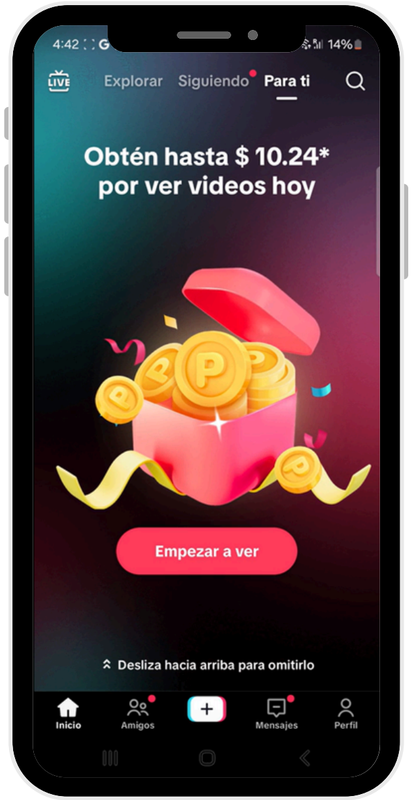
Ads between videos also offered coins in return for inviting friends to join TikTok, which also appeared on the child-aged avatars. The coins accumulated in a visible counter, but their purpose and value were not clearly explained to the user. This appears to be part of the TikTok Lite Rewards programme, which is banned in the EU, but available to users in Mexico.

INCENTIVES

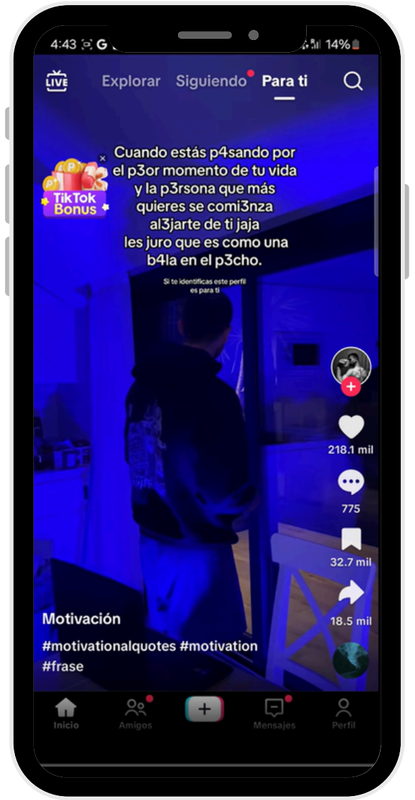
Features that appeared to incentivise users to keep watching

These features were only observed in a small number of cases and were not consistent across all avatars. However, where they did appear, they seemed to introduce a game-like element to scrolling – potentially encouraging users to watch for longer periods in exchange for perceived rewards.

In one case, the coin feature appeared, then served a series of emotionally charged videos, including content about loss, anxiety and sadness.



TikTok pop-up saying 'Win up to $10.24 for watching videos today'.



The next video: 'When you're going through the worst time of your life and the person you love the most starts to pull away from you, haha, I swear it's like a bullet to the chest'.

The video after: 'How do I explain to my immune system that I just have lots of stress and anxiety and that I don't need to have such bad headaches, nausea and appetite loss'.

INCENTIVES

Features that appeared to incentivise users to keep watching

Although the coin system's function was not clearly explained to the user, its presence suggests a mechanism that may encourage extended use. For children – who may not fully understand the nature or value of digital rewards – these features could contribute to prolonged time-on-platform.

TikTok's recommender system learns from user behaviour: the more time a user spends watching a particular kind of content, the more the algorithm shows them similar material. When combined with coin-based incentives, this system could contribute to patterns of viewing where emotionally intense content is shown more frequently than other types of material.

These incentives could encourage children to watch entire videos rather than swipe away, contributing to a feedback loop that rewards continued engagement, regardless of the topic or tone of the content being shown. While such systems are designed to optimise for interest, not harm, the result may still be an experience that pulls children deeper into content loops without meaningful opportunities for pause, reflection, or redirection.

CONCLUSION

Summary of key findings and recommendations

CONCLUSION

Key findings

Children in Mexico can be served potentially harmful content across a number of themes. A few minutes of scrolling each day on avatars mimicking real children's interests and behaviours can lead to feeds filled with content relating to depression, self-harm, violence, and sexual content.

Platforms are not only failing in their obligation to respect children's rights on policy – they are failing on enforcement, moderation, and design. In many cases, this is at odds with the standards and rules outlined in their own guidance and policies.

Platforms feed children potentially harmful content through recommendations and take no meaningful steps to prevent them from accessing and remaining on this trajectory.

Harmful pathways (e.g. from mental health interest to self-harm content) are fast and frictionless – and may be rewarded by the platform itself.

Platforms' current safeguards are insufficient to meet their own standards, let alone national or international children's rights responsibilities.

CONCLUSION

Recommendations

Technology companies

Respect children's rights and prevent violations, embedding safety, privacy and children's rights by design and by default across all digital products and services likely to be accessed by children, in line with international duties, and the Committee on the Rights of the Child General Comment No. 25.

- Undertake early, proactive, ongoing children's rights impact assessments throughout the lifecycle of products and services, and include detailed information on risks and mitigating measures.
- Introduce proportionate, effective, and privacy-preserving age verification at sign-up to direct children to age-appropriate experiences.
- End manipulative and exploitative business models, including engagement-driven algorithms, and deceptive reward systems that encourage excessive use, compulsive behaviour, or the commercial exploitation of children.
- Ensure robust, consistent, and transparent enforcement of terms of service, policies, and community standards.

CONCLUSION

Recommendations

Governments and legislators

Children have the right to protection from commercial exploitation, to safety, and to privacy, wherever they engage in the digital environment. These rights are guaranteed under international law, in particular the UN Convention on the Rights of the Child, as authoritatively interpreted by the Committee on the Rights of the Child, General Comment No. 25 on children's rights in relation to the digital environment.

In accordance with these obligations, States must ensure that children's rights are respected, protected, and fulfilled online, and that their best interests are a primary consideration in all actions concerning the design, regulation, governance, and operation of digital services.

To meet their international duties, governments and legislators must hold technology companies accountable, including by:

- Prioritising the adoption, implementation, and enforcement of legislation, regulation, policies, and technical standards specifically aimed at protecting children's rights in the digital environment.
- Legislating and enforcing companies' duties to respect children's rights, including obligations to prevent, mitigate, and remedy actual and potential adverse impacts on children.
- Including an overarching and legally enforceable duty of care for services likely to be accessed by children, and requiring the implementation of mandatory Safety-by-Design and Privacy-by-Design approaches across the full lifecycle of digital products and services.
- Mandating robust due diligence, including mandatory Child Rights Impact Assessments (CRIAs) that address risks relating to content, contact, conduct, and contract, as well as systemic and cross-cutting harms.
- Prohibiting manipulative, deceptive, and exploitative business models that undermine children's rights or exploit their vulnerabilities.
- Enforce platform terms of service, policies, and community standards.

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

El hilo de Ariadne