Explorando lo que las y los adolescentes pueden ver en las redes sociales en México.

Informe de investigación

Junio 2026

REVEALING REALITY

Sobre Revealing Reality
Revealing Reality es una agencia independiente de investigación social y análisis. Disfrutamos trabajar en proyectos desafiantes con propósitos sociales para fundamentar políticas, diseño y cambios de comportamiento.

Esto incluye colaborar con organismos reguladores, gobiernos y organizaciones benéficas, para brindar un análisis riguroso de los comportamientos y experiencias en línea de las personas jóvenes. Estudiar cómo el mundo digital está moldeando la vida de las personas es algo que hacemos a diario. Esto incluye explorar cómo los servicios y plataformas digitales están moldeando el comportamiento de las personas, en las relaciones sociales, las apuestas, los productos financieros, la salud y más.

Frecuentemente realizamos investigaciones cualitativas y cuantitativas detalladas para comprender a fondo los comportamientos digitales y observar cómo las personas realmente experimentan la tecnología y el mundo en línea. Visite www.revealingreality.co.uk para obtener más información sobre nuestro trabajo o ponerse en contacto con nosotros.

Queremos agradecer a El hilo de Ariadne su apoyo en la realización de esta investigación.

Glosario

- **Feed:** página principal o flujo de contenido (actualizado de forma permanentes) que muestra publicaciones y actividades de las cuentas que una persona sigue o a las que está suscrita.

- **Input:** En el presente trabajo y en el contexto de redes sociales, input se refiere a las distintas interacciones que se llevan a cabo con los avatares (ejemplo: dar 'me gusta', hacer 'scroll', etc).

- **Scroll:** se refiere a la acción de deslizar el dedo o el cursor hacia abajo o hacia arriba en la pantalla para ver más contenido en un 'feed', página web o aplicación.

- **Clickbait:** práctica de usar titulares o miniaturas llamativas, a menudo sensacionalistas o engañosas, para generar clics en un enlace o contenido, incluso si este no cumple con las expectativas creadas.

Resumen ejecutivo

¿A qué pueden estar expuestos las y los adolescentes de México en las redes sociales?

Las plataformas de redes sociales, incluyendo TikTok, Instagram y X, afirman estar comprometidas con mantener a las y los adolescentes seguros en línea. Para su uso se requiere que las personas tengan al menos 13 años, y algunas plataformas mencionan usar tecnología para detectar usuarios(as) menores de edad o para aplicar protecciones adicionales para ellos(as). Estas protecciones incluyen configuraciones de privacidad, políticas de moderación de contenido y herramientas que buscan limitar la exposición a material dañino.

Esta investigación se propuso explorar cómo estas promesas se traducen en experiencias reales para adolescentes de México. Utilizando perfiles de redes sociales simulados - avatares - basados en los comportamientos en línea de adolescentes reales, la investigación examinó qué contenido podía aparecer al realizar acciones cotidianas como hacer 'scroll', dar 'me gusta' o 'seguir' cuentas así como buscar términos específicos.

Los hallazgos sugieren una brecha significativa entre las políticas de las plataformas y lo que las y los adolescentes pueden encontrar en la práctica.

- Los avatares creados para el proyecto pudieron acceder a todas las plataformas fácilmente, sin encontrar mecanismos importantes de verificación de edad.
- En algunos casos, los avatares fueron expuestos a contenido potencialmente dañino - incluyendo material relacionado con autolesión, violencia de cárteles e imágenes sexualizadas- en un periodo de uso relativamente corto.
- Las funciones de búsqueda de términos específicos parecen desempeñar un papel clave en la facilitación del acceso a ciertos tipos de contenido, incluso cuando algunos términos estaban bloqueados o generaron advertencias.
- En TikTok, a un pequeño número de avatares se les mostraron monedas digitales que parecían recompensar la visualización prolongada del contenido digital, aunque el propósito y el impacto de estas monedas no son claros.

Resumen ejecutivo

¿A qué pueden estar expuestos las y los adolescentes de México en las redes sociales?

Se ejecutaron 5 avatares durante un total de 60 minutos a lo largo de un período de investigación de 12 días y se agregaron 5 avatares ejecutados en una ventana más corta de 2 días. A pesar de esta interacción limitada, el contenido que resultó podría considerarse angustiante o inapropiado para adolescentes.
Dado que muchos(as) adolescentes cada día pasan una gran cantidad de tiempo en estas plataformas, los hallazgos plantean preguntas importantes sobre lo que una o un adolescente real podría encontrar con el tiempo.

Aunque estas observaciones no indican con qué frecuencia ocurren tales experiencias a gran escala, sí señalan patrones de diseño que pueden exponer a adolescentes a ciertos riesgos. La investigación subraya cómo los sistemas de recomendación pueden cambiar rápidamente según la interacción del o de la usuaria y cómo los mecanismos de protección - cuando están presentes - pueden ser inconsistentes o limitados en su efectividad.

Resumen ejecutivo

¿A qué pueden estar expuestos las y los adolescentes de México en las redes sociales?

Para reducir el riesgo de daño y lograr una mejor alineación entre la práctica y las políticas declaradas, las plataformas tendrían que:

- Evaluar y mitigar el posible impacto de las funciones de diseño que dirigen a adolescentes hacia riesgos o que fomentan una visualización prolongada o repetitiva del contenido digital.
- Cumplir con sus propios términos, políticas y normas publicadas.
- Revisar cómo responden los sistemas de recomendación a la interacción con contenido emocionalmente intenso o material sensible.
- Fortalecer los procesos de verificación de edad para distinguir de mejor manera entre personas usuarias adultas y menores, con el fin de ofrecer experiencias apropiadas para su edad y respetuosas de sus derechos.
- Aplicar medidas de moderación de contenido de manera más consistente, especialmente en los algoritmos recomendados.

Este informe presenta evidencia de una investigación basada en avatares en México para apoyar la reflexión y la acción sobre cómo el diseño de las plataformas afecta las experiencias de adolescentes en línea.

INTRODUCCIÓN

Contexto de la investigación

Basado en una investigación previa llamada 'Pathways', la cual exploró las experiencias de adolescentes en las redes sociales en Gran Bretaña, este informe tiene como objetivo explorar lo que las y los adolescentes en México pueden experimentar en las redes sociales y cómo las opciones de diseño de plataformas moldean esas experiencias. Utilizando perfiles simulados – o avatares – basados en su comportamiento real, la investigación revela cómo los sistemas de recomendación responden a acciones cotidianas como hacer ‘scroll’, dar ‘me gusta’, ‘seguir’ cuentas y buscar términos específicos.

A pesar de los compromisos públicos de las plataformas sobre la seguridad, las y los adolescentes pueden acceder fácilmente a las redes sociales y pueden verse expuestos rápidamente a contenido angustiante, gráfico o sexualizado.

Las funciones diseñadas para fomentar la interacción parecen amplificar el daño en lugar de limitarlo, mientras que las medidas de protección son inconsistentes o inexistentes.

Al comparar lo que las plataformas dicen que hacen con lo que los avatares realmente vieron, este informe destaca una clara brecha entre la política y la práctica. Muestra cómo aportes simples pueden dirigir a contenido potencialmente dañino y refuerza por qué el diseño es importante cuando se trata de proteger a las y los adolescentes en línea.

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

METODOLOGÍA

Del comportamiento y experiencias de adolescentes reales a perfiles digitales

Para entender lo que las y los adolescentes en México pueden ver y experimentar en las redes sociales, utilizamos una metodología de 'avatar': creando perfiles de redes sociales que imitan el comportamiento de adolescentes reales en línea. Este enfoque nos permite explorar y probar 'los feeds' impulsados por algoritmos sin poner en riesgo a las y los adolescentes. El objetivo fue probar lo que el algoritmo mostraría a una o un adolescente promedio a través de acciones cotidianas como hacer ‘scroll’, dar ‘me gusta’, ‘seguir’ cuentas y buscar términos específicos.

Previo a la creación de 'avatares', se realizaron entrevistas a profundidad a 12 adolescentes de distintas partes de México para entender cómo experimentan el mundo en línea. Las entrevistas se realizaron en línea en idioma español por una investigadora hispanohablante junto con una investigadora de México.

La muestra incluyó a adolescentes de diferentes edades y géneros, residentes de diversas localidades del país, como se muestra en el diagrama de la derecha. Algunos(as) adolescentes fueron incluidos porque habían visto tipos particulares de contenido o habían tenido experiencias que se relacionaban directamente con los temas explorados a través de los avatares.

Se les preguntó a las y los participantes sobre sus vidas digitales cotidianas: qué plataformas utilizan, qué tipo de contenido ven, cómo se sienten al respecto y cómo se comportan en línea.

Tras la entrevista, cada participante compartió grabaciones de sus 'feeds' de las redes sociales que más usa. También compartieron la lista de cuentas que seguían en cada perfil.

Estas entrevistas proporcionaron la información conductual y el contexto del mundo real que moldearon el diseño de cada avatar.

Posteriormente, los avatares se basaron en los perfiles y las listas de seguimiento de aquellas y aquellos adolescentes cuyas experiencias se alineaban más claramente con los temas de la investigación.

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

METODOLOGÍA

Escalada de comportamientos de 'input'

Se configuraron avatares en TikTok, Instagram y X. Todos los perfiles utilizaron comportamientos comunes como hacer 'scroll', dar 'me gusta', seguir cuentas y buscar términos, para simular cómo interactúan típicamente las y los adolescentes en las redes sociales. Los 'inputs' de interacción se incrementaron gradualmente para poner a prueba cómo se adaptaba el algoritmo en respuesta a esas interacciones.

Aunque los avatares fueron operados por investigadores en el Reino Unido, se configuraron utilizando tarjetas SIM mexicanas y se accedió a ellos a través de VPN que enrutaban a través de México, lo que permitió a las plataformas ofrecer contenido local a los avatares.

Este método mostró cómo se puede revelar contenido potencialmente dañino de forma rápida y sencilla, y cómo el diseño de la plataforma refuerza la exposición continua. Si bien los avatares reflejan experiencias de usuarios realistas, no miden lo que las y los usuarios verán o es probable que vean. Ilustran lo que se puede mostrar, no lo que se mostrará, con qué frecuencia ni a cuántas o cuántos adolescentes.

El trabajo de campo de este proyecto, que incluyó las entrevistas y los avatares, se realizó entre octubre de 2024 y febrero de 2025.

Configuración

- Edad de la o el adolescente o de su cuenta de redes sociales.
- Registro con una dirección de correo electrónico ficticia.
- Seguimiento de una muestra seleccionada al azar de la lista de seguimiento de una o un adolescente.

Fases de escalada (interacción)

- **Fase pasiva:** 4 días de solo hacer 'scroll'
- **Fase de interacción:** 4 días de dar 'me gusta' al contenido y seguir cuentas relacionadas con el tema del avatar.
- **Fase de búsqueda:** 4 días buscando términos relacionados con el tema del avatar.

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

METODOLOGÍA

Avatares basados en las edades y comportamientos de adolescentes reales

Temas y avatares
Cada avatar reflejaba los comportamientos, intereses y patrones de interacción de adolescentes reales en México y fue asignado a alguno de los cinco temas mencionados a continuación:

- **Salud mental y depresión** - contenido que abarca desde el bajo estado de ánimo hasta la autolesión.
- **Contenido sexual** – videos sugerentes e imágenes sexualizadas.
- **Violencia** – incluyendo violencia interpersonal y contenido relacionado con cárteles y narcotráfico.
- **Trastornos alimentarios** - consejos de dietas extremas y material que promueve la alimentación desordenada.
- **Cultura 'Buchona'** - una estética que glorifica el lujo, generalmente como una 'narcowife', y los ideales de belleza hiperfemeninos, a menudo sexualizados y glamorizados

Durante las entrevistas, cada adolescente reportó tener un perfil en redes sociales con una edad mayor a la suya real - a menudo utilizando una fecha de nacimiento falsa o aleatoria para registrarse. Con el tiempo, estas cuentas fueron 'envejeciendo' junto con ellos. Como resultado, las y los participantes estaban utilizando cuentas con una edad registrada de 18 años o más. Para reflejar esta realidad, se configuraron dos avatares por cada tema: uno con la edad adulta que la o el adolescente ingresó al crear su perfil, y otro con su edad real.

METODOLOGÍA

Avatares basados en las edades y comportamientos de las y los adolescentes reales

El período de trabajo de campo duró 14 días. Durante 12 días, se utilizaron avatares configurados con edades de personas adultas, y los 2 días restantes, se usaron avatares con las edades reales de las y los adolescentes . Esto se hizo para evaluar si la precisión de la edad impactaba las medidas de seguridad o las recomendaciones de contenido, y para reflejar las experiencias genuinas de las y los adolescentes en redes sociales.

Los perfiles ajustados a la edad real de las y los participantes se ejecutaron con un diseño de avatar acelerado de dos días. Estos avatares siguieron las mismas cuentas que los avatares adultos de 12 días, pero aumentaban la interacción con temas potencialmente dañinos mucho antes. Esto nos permitió explorar qué mostraban las plataformas a las y los usuarios con la edad de una o un adolescente.

Tema	Edad de adolescente	Edad de perfil	Plataforma
Contenido sexual	15	24	
“Buchona’	13	18	
Salud mental	15	24	
Trastornos Alimentarios	13	29	
Violencia	14	38	

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

ACCESO

Las y los adolescentes pueden acceder a plataformas a pesar de las restricciones de edad de las propias plataformas

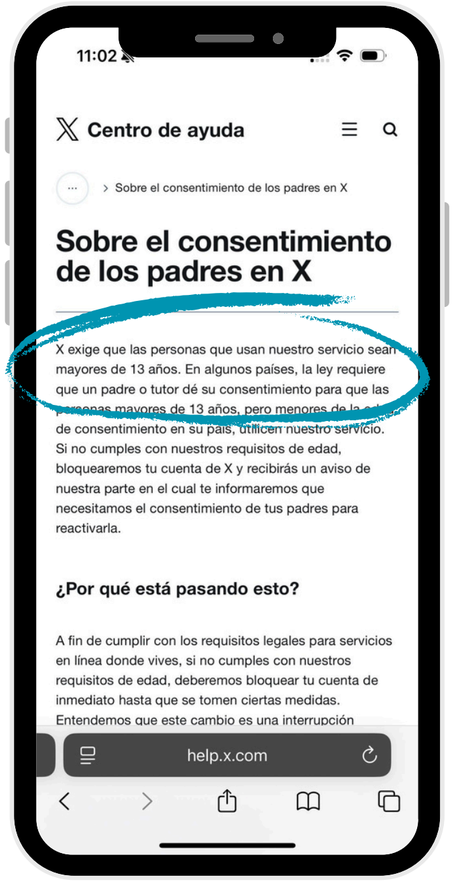
ACCESO

Las plataformas detallan sus enfoques para la verificación de edad

La verificación de edad puede ser un pilar fundamental para diseñar experiencias en línea seguras, apropiadas para la edad y que respeten los derechos de las y los adolescentes.
Sin saber quiénes son realmente sus usuarios, las plataformas no pueden ofrecer las protecciones adecuadas, aplicar las políticas pertinentes, ni diseñar espacios que respondan a las necesidades de las y los adolescentes. Esto obliga a las empresas a diseñar experiencias que protejan, empoderen y apoyen a las y los adolescentes, no solo que los limiten.

TikTok, Instagram, y X establecen que sus usuarios(as) deben tener al menos 13 años para registrarse.

TikTok e Instagram van más allá, afirmando que utilizan tecnología para detectar a usuarios(as) menores de edad y aplicar protecciones especiales para adolescentes.
Sin embargo, esto no parece reflejar las experiencias digitales de las y los adolescentes.



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

ACCESO

Experiencias de adolescentes en la creación de cuentas

La información recabada de las entrevistas mostró que todas las y los adolescentes usaron cuentas configuradas con una edad mayor de 18 años, lo que significa que estaban oficialmente registrados como personas adultas en las redes sociales. Para reflejar esta realidad, en una primera etapa del presente proyecto, los avatares se registraron inicialmente con la edad que las y los participantes ingresaron al crear sus perfiles.

Configurar los avatares en todas las plataformas fue rápido y sencillo. Ingresar una fecha de nacimiento - ya sea precisa o no - fue el único paso relacionado con la edad, y no se requirió ninguna verificación adicional. Las plataformas no parecieron cuestionar la edad ingresada, y las cuentas obtuvieron acceso inmediato a contenido y funciones.

Si una plataforma no conoce la edad de sus usuarios(as), no puede hacer cumplir sus propias reglas ni garantizar que todas las personas usuarias – especialmente las y los adolescentes – vean contenido que respete sus derechos (seguridad, privacidad y contenido apropiado para su edad). Es entonces que no se puede impedir que adolescentes accedan a contenido para personas adultas, ni garantizar que los sistemas de recomendación o las funciones de recompensa sean seguros y apropiados.

El resultado es un sistema en el que las restricciones de edad existen solo en el papel, pero no en la práctica - y en el que, por diseño, las y los menores son tratados como personas adultas.

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

EXPOSICIÓN

Los ‘feeds’ de recomendación pueden amplificar rápidamente contenido potencialmente dañino

EXPOSICIÓN

Los feeds de recomendación pueden amplificar rápidamente contenido potencialmente dañino

Las plataformas de redes sociales ofrecen a las y los usuarios un flujo de contenido personalizado, diseñado para captar y mantener su atención. Las plataformas afirman que sus sistemas de recomendación están moderados para evitar la repetición y promover el bienestar. Sin embargo, esto no siempre parece reflejar la experiencia real de las personas usuarias.

Aunque hay herramientas disponibles en las plataformas para alentar a las y los usuarios a señalar y reportar contenido que 'no les interesa', no parecen existir medidas de protección efectivas integradas en el diseño de la plataforma.

Como resultado, las y los usuarios aún pueden recibir cantidades cada vez mayores de contenido potencialmente dañino.

EXPOSICIÓN

Los feeds de recomendación pueden amplificar rápidamente contenido potencialmente dañino

Para los avatares de este estudio, interacciones simples como hacer ‘scroll’ en el feed, buscar términos específicos, darle 'me gusta' a contenido o seguir cuentas relacionadas con el tema de ese avatar fueron suficientes para que los feeds se orientaran hacia un contenido cada vez más específico y, en algunos casos, potencialmente dañino.

Interacciones sencillas aceleraron rápidamente la trayectoria de los avatares hacia contenido dañino. Dar ‘me gusta’ al contenido o ‘seguir’ perfiles - ambas interaccienes básicas en la plataforma - transformaron el algoritmo de genérico a potencialmente dañino en tan solo 30 minutos de interacciones.

Para ilustrar cómo evolucionó el contenido con el tiempo, tres estudios de caso muestran cómo diferentes avatares experimentaron contenido cada vez más específico y potencialmente dañino en respuesta a patrones cotidianos de interacción como hacer ‘scroll’, dar ‘me gusta’, ‘seguir’ cuentas y buscar términos específicos.

Estos incluyen:

- **Avatar de salud mental:** El contenido rápidamente se desplazó hacia publicaciones que hacen referencia a la depresión, la desesperanza y, en algunos casos, temas relacionados con el suicidio.

- **Avatar de violencia:** Tras una primera exposición a vídeos de acción genéricos, su feed se transformó rápidamente en contenido gráfico. Esto incluyó peleas callejeras, accidentes violentos o extremos, y noticias sobre violencia relacionada con los cárteles.

- **Avatar de contenido sexual:** La interacción ligera con contenido de estilo de vida y relaciones llevó a un feed dominado por vídeos sexualizados y material con temática para personas adultas.

ESTUDIO DE CASO: Salud mental

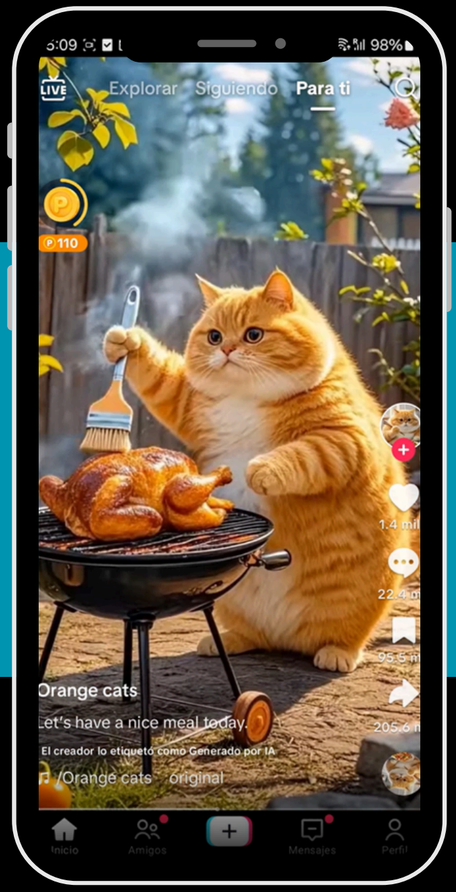
Contenido de salud mental en TikTok

Este avatar se basó en Helena, una adolescente de 16 años que describió ver contenido emocionalmente intenso en TikTok de forma regular, especialmente por la noche. Sus grabaciones de pantalla mostraban un feed lleno de temas tristes, incluyendo referencias a la autolesión.

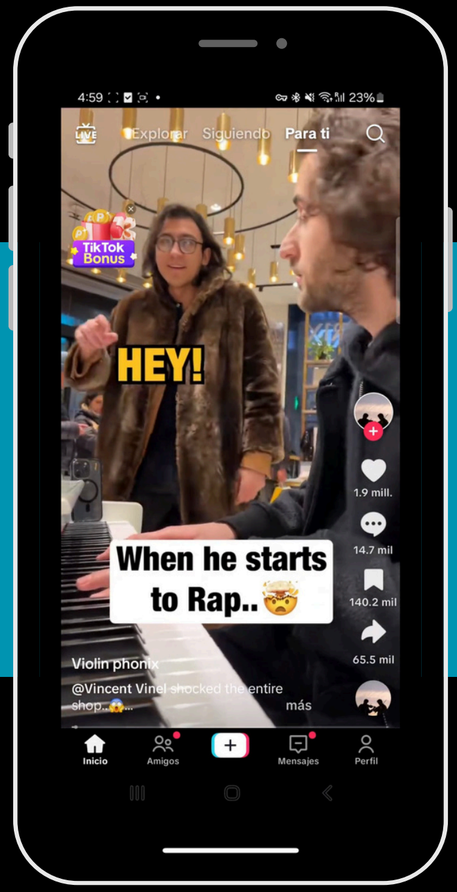
'Me meto a TikTok y me salen vídeos de puras cosas tristes..pero pues yo quiero otras cosas..O sea, yo me meto para distraerme, pero me salen con más cosas'

Para reflejar su experiencia, su avatar se configuró utilizando la misma edad que ella ingresó al registrarse en la plataforma y se empezó a seguir a una muestra de sus cuentas.

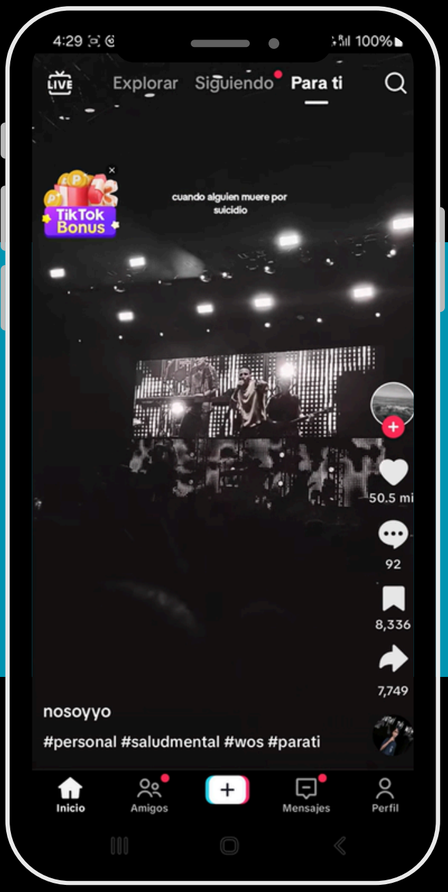
Tras la configuración inicial y la fase de 'scroll', el feed del avatar mostró una mezcla de vídeos virales genéricos, incluyendo memes, vídeos de canciones y publicaciones sobre relaciones.
Sin embargo, tras una mínima interacción con contenido de temática emocional, el 'feed' empezó a cambiar. Para el décimo día, casi todos los vídeos mostrados se relacionaban con la tristeza, la ansiedad y/o la desesperanza.



Día 1



Día 2



Día 12

ESTUDIO DE CASO: Salud mental

Fase uno: dar me ‘gusta’ y ‘seguir’ cuentas

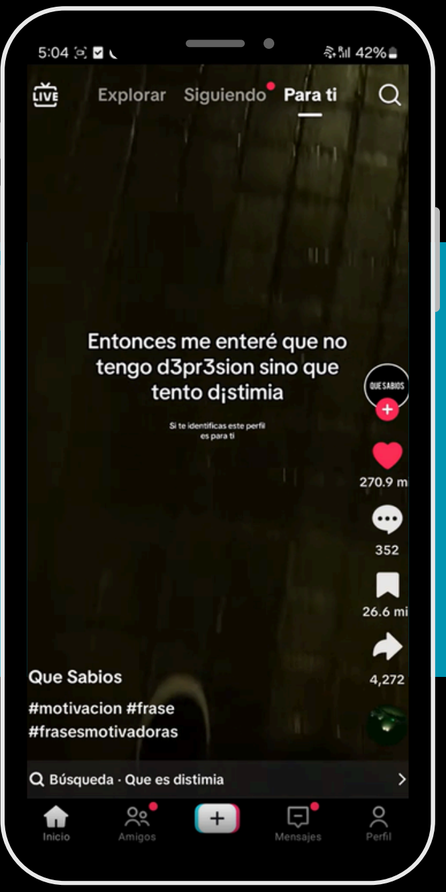
Cada día, los investigadores hicieron ‘scroll’ durante cinco minutos en el feed del avatar. Dieron ‘me gusta’ hasta a cinco contenidos relacionados con la salud mental o la depresión, y siguieron hasta tres de las cuentas que publicaban ese tipo de contenido.

Los primeros ejemplos de contenido emocional incluían reflexiones nostálgicas o de arrepentimiento sobre relaciones pasadas. Después de unos días de desplazamiento, estos parecieron volverse más intensos, con publicaciones centradas en la soledad, la baja autoestima y el malestar mental.

El volumen y la consistencia del contenido emocionalmente intenso se escalaron rápidamente. En pocos días, el feed quedó dominado por contenido relacionado con la depresión. En contraste, los avatares que exploraban otros temas en TikTok típicamente recibían de uno a tres videos relevantes por día.

Visto el Día 5 de una cuenta ya seguida

Marcado como ‘me gusta’ el Día 6



Cuenta seguida y marcado como ‘me gusta’ el Día 7

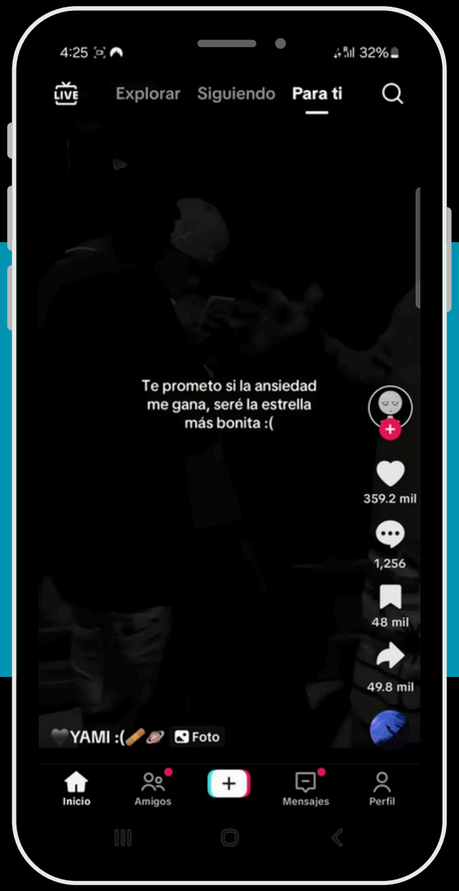
ESTUDIO DE CASO: Salud mental

Fase dos: ‘buscando’

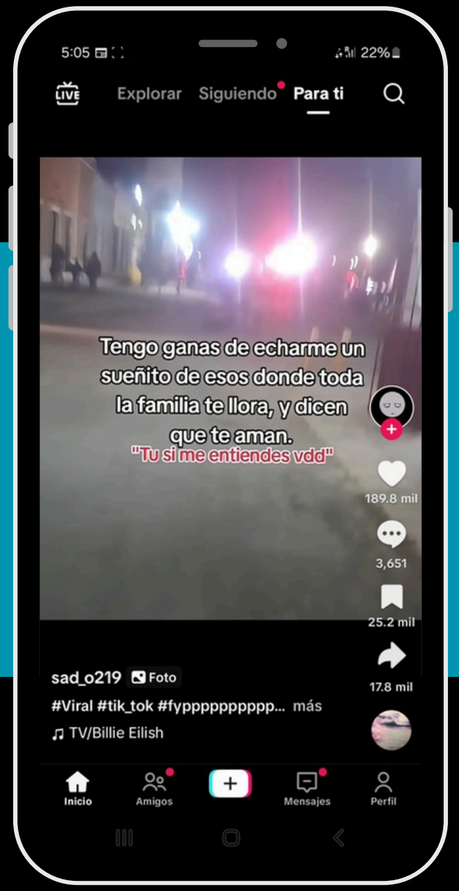
En el día ocho – con menos de 40 minutos de actividad total en la cuenta – al avatar se le mostró por primera vez un video que hacía referencia explícitamente al suicidio. No se mostraron advertencias, restricciones ni recursos de apoyo en respuesta.

Búsquedas posteriores de términos relacionados con la salud mental arrojaron más contenido emocionalmente intenso y potencialmente dañino. En algunos casos, el contenido parecía eludir la moderación de TikTok mediante el uso de lenguaje codificado, faltas de ortografía o superposiciones de texto. El contenido incluía referencias a autolesiones, desesperanza e ideación suicida, presentadas sin apoyo contextual o intervención.

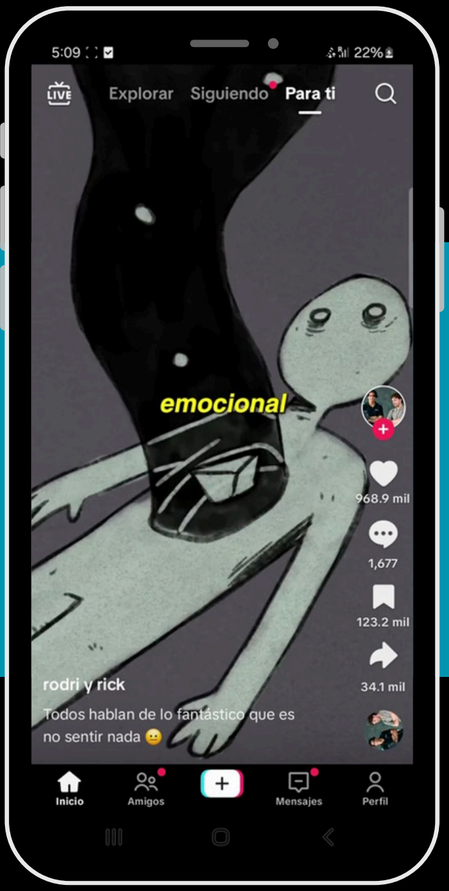
Para los últimos días del período de prueba, el feed del avatar estaba casi totalmente dominado por contenido depresivo, incluyendo publicaciones que normalizaban las autolesiones y el pensamiento suicida.



Día 8



Día 9: ‘Tengo ganas de echarme un sueñito de esos donde toda la familia te llora, y dicen que te aman. “Tu si me entiendes vdd”’.



Día 11: 'Todos hablan de lo maravilloso que es no sentir nada, pero nadie habla de lo frustrante que es estar en un bloqueo emocional donde no puedes conectar con nadie en absoluto'

ESTUDIO DE CASO: Contenido sexual

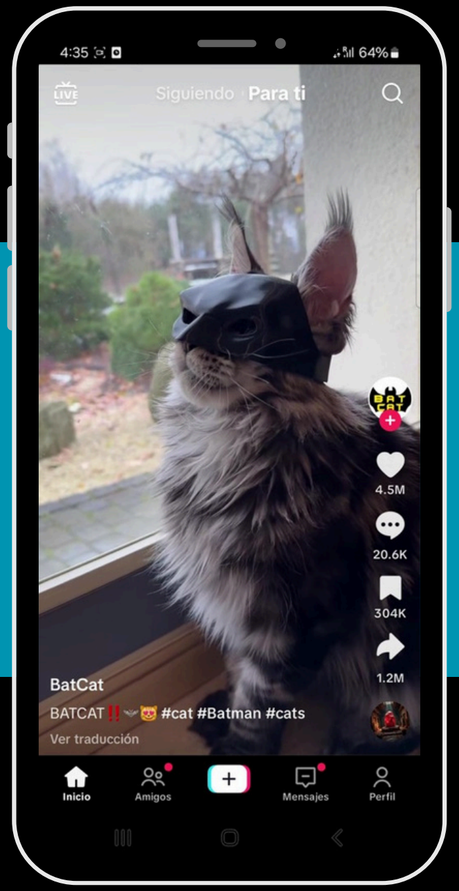
Contenido sexual en Tiktok

Javier es un chico de 15 años que describió ver contenido sexual regularmente en sus redes sociales.

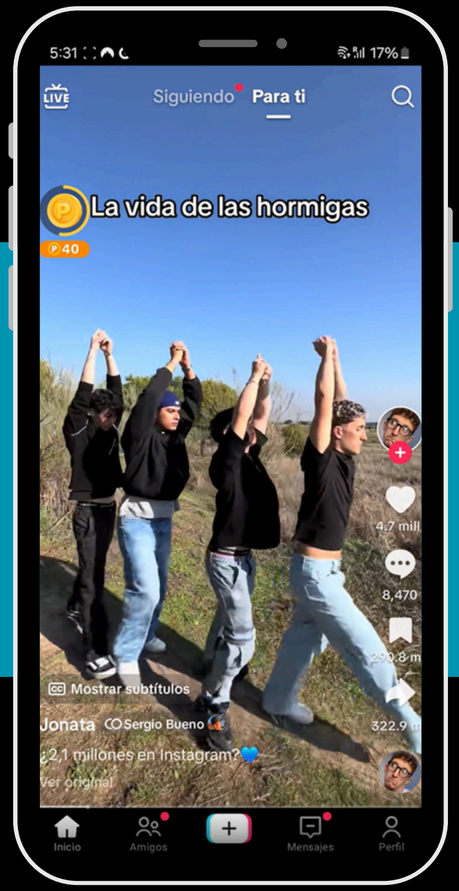
'*O sea, pues no hay porno porno, pero pues como mujeres bailando con poca ropa o cosas así.*'

Este avatar se configuró con la edad que el adolescente había ingresado al crear su propia cuenta y se configuró con una combinación de su lista de seguidos, así como perfiles vistos en las grabaciones de su 'feed' de TikTok.

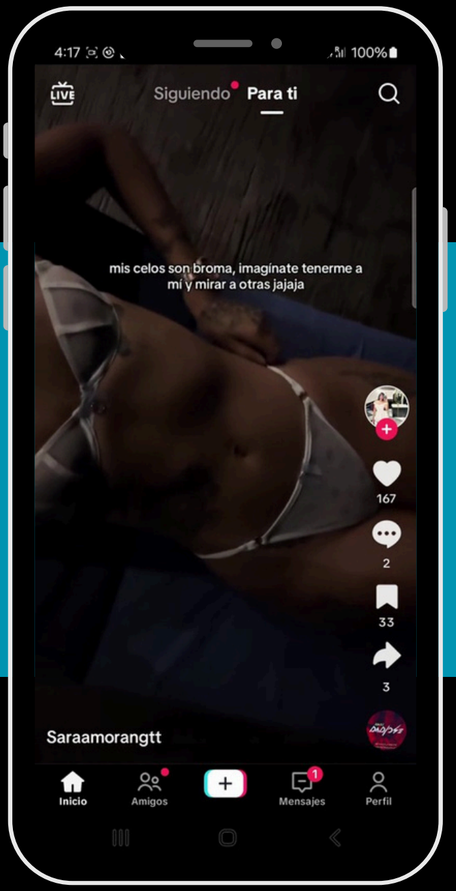
Aunque este perfil inicialmente recibió contenido genérico y viral, como gatos vestidos de Batman, después de una interacción mínima, el avatar comenzó a ver regularmente referencias a actos sexuales y contenido sexualizado. Esto incluye casos de perfiles de TikTok que dirigen al usuario fuera del sitio hacia contenido privado y explícito en sitios de suscripción.



Día 1



Día 3



Día 12: 'Mis celos son broma, imaginate tenerme a mi y mirar a otras jajaja'

ESTUDIO DE CASO: Contenido sexual

Fase uno: dar me ‘gusta’ y ‘seguir’ cuentas

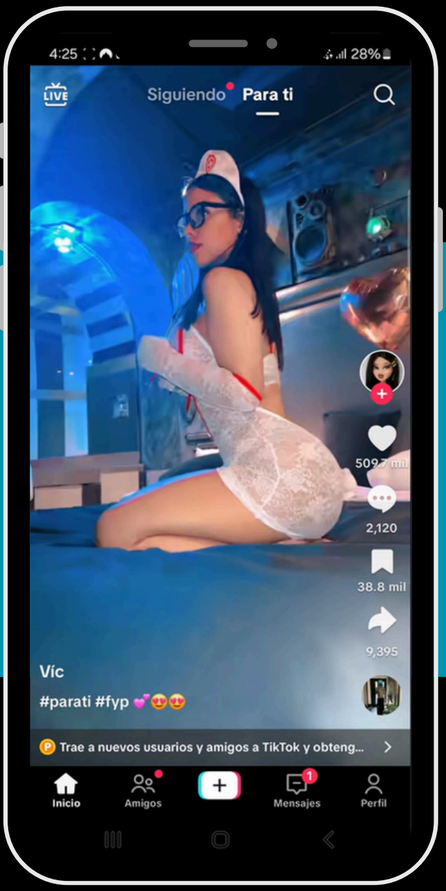
Después de seguir inicialmente las cuentas, el feed del avatar comenzó a mostrar contenido relacionado con las descripciones de Javier. Esto incluía '*mujeres bailando con poca ropa*'.

En el día ocho, el avatar vio 30 videos. El análisis de este contenido mostró que 10 de los videos fueron publicados por perfiles que promovían y dirigían a los usuarios a plataformas de suscripción privadas y explícitas (por ejemplo, OnlyFans).

Aunque TikTok no permite el acceso directo a través de su aplicación a páginas de suscripción (como OnlyFans), esto puede eludirse abriendo el enlace en un navegador externo o accediendo a través de otra aplicación, como Instagram.

Marcado como ‘me gusta’ de una cuenta ya seguida

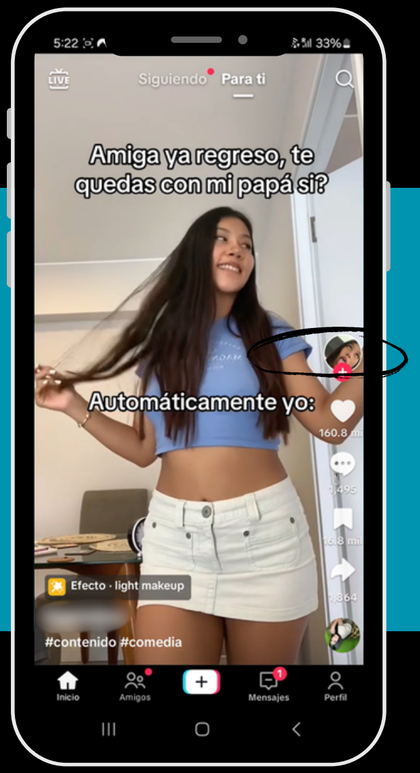
Marcado como ‘me gusta’ de una cuenta ya seguida el Día 7: 'Andan diciendo que te gustan como yo... Bien consentida y berrinchuda'



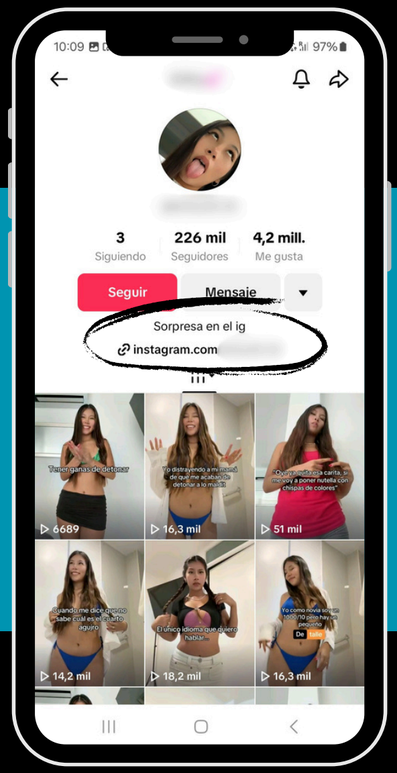
Visto el Día 8: 'Ok pero yo en mi uniforme de enfermera 💕'

ESTUDIO DE CASO: Contenido sexual

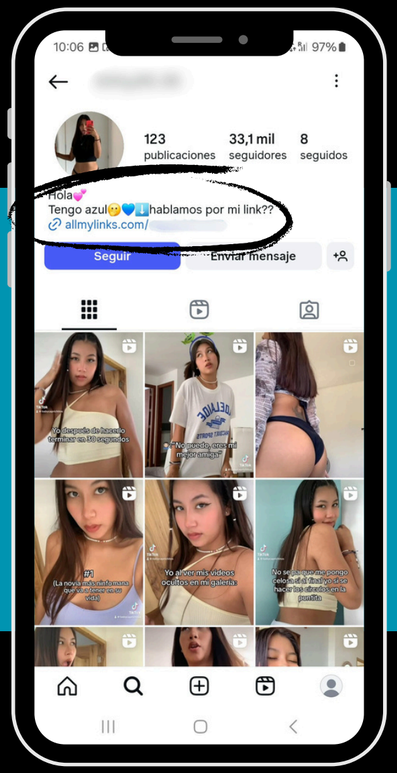
Camino de TikTok a la pornografía



*Cuenta seguida y marcado como 'me gusta' el Día 7: *audio de gemidos sexuales**



'Al hacer clic en este perfil, los usuarios son dirigidos a Instagram, con su biografía que dice 'sorpresa en mi IG (Instagram)' y un enlace'.



En su cuenta de Instagram, su biografía dice: 'Tengo un azul [referencia a OnlyFans]. Podemos hablar a través de mi enlace'.

Este enlace lleva a una de dos páginas privadas con contenido explícito para adultos, que promueven contenido de 'sexo duro', '(sexo oral) profundo', '(eyaculaciones) faciales' y '(sexo) anal'.

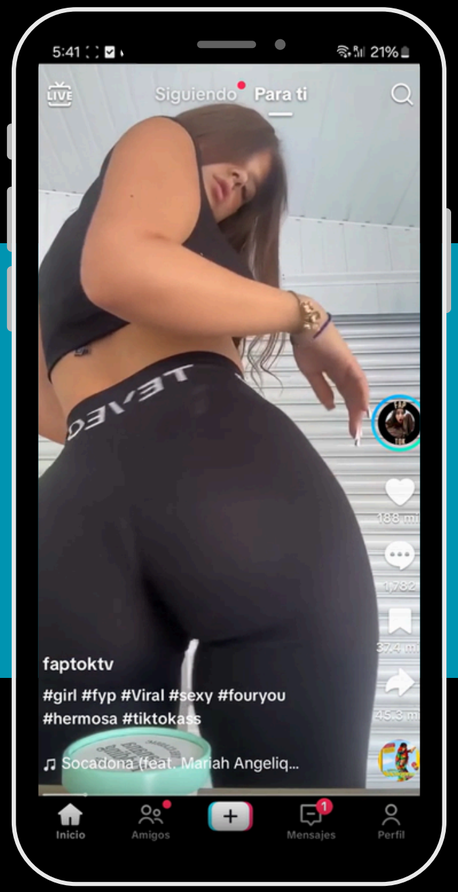
ESTUDIO DE CASO: Contenido sexual

Fase dos: ‘buscando’

Las búsquedas de términos como 'bikini' y 'belleza' resultaron en que se mostrara más contenido sexual en el feed del avatar. Esto incluyó videos de actos sexuales imitados y audio sexualizado, como gemidos.

Las Normas de la Comunidad de TikTok establecen que la plataforma restringe el contenido que contiene 'comportamiento sexual' y 'encuadre sexualizado'. También afirma que este contenido 'no es elegible para el FYF [For You Feed] si muestra besos íntimos, encuadre sexualizado o comportamiento sexualizado por parte de personas adultas'.

Con una mínima interacción en TikTok, los adolescentes pueden ser rápidamente expuestos a contenido sexualizado y a cuentas que promueven material de pago para personas adultas.



Día 9: Un video de la cuenta ‘faptoktv’ [refiriéndose a la masturbación] incluyendo emarcamiento sexualizado.

*Día 11: *audio de gemidos sexuales**

Día 12

ESTUDIO DE CASO: Contenido violento

Contenido violento en X

Este avatar se basó en la experiencia de Gabriel, un niño de 14 años que describió haber visto contenido violento en X.

'Son cuentas que encuentran vídeos exclusivos así que nadie puede encontrar y lo suben ellos para que lo vea otra gente... tienen mucha gente muerta, pues mucha sangre y mucha violencia'.

Se creó una cuenta avatar en X, registrada con la edad adulta que Gabriel había usado en su propia cuenta. No hubo un sistema sólido de verificación de edad.

Inicialmente, el 'feed' de la cuenta mostraba principalmente noticias mexicanas y estadounidenses. Al tercer día, con tan solo 15 minutos de actividad, se le mostró al avatar el cadáver de una celebridad ejecutada, sin censura. Durante el trabajo de campo, el 'feed' del avatar se vio dominado por temas violentos e imágenes explícitas.

Día 1

Día 2

Día 3

ESTUDIO DE CASO: Contenido violento

Fase uno: dar me ‘gusta’ y ‘seguir’ cuentas

Después de intensificar el ‘input’ del avatar al dar ‘me gusta’ a publicaciones y ‘seguir’ cuentas relacionadas, apareció contenido casi exclusivamente violento. Esto incluía vídeos de personas siendo heridas o asesinadas en accidentes, así como imágenes violentas asociadas con la actividad de pandillas o cárteles.

Después de dar ‘me gusta’ y ‘seguir’ el contenido mostrado, el ‘feed’ del avatar se llenó de contenido similar -noticias individuales o periodistas que publicaban sobre la actividad de los cárteles- a un volumen importante durante el resto del periodo de trabajo de campo.

Visto el Día 4: Video que muestra el interrogatorio a punta de pistola de un miembro de un cartel, con el video terminando cuando se le pone un cuchillo en la garganta antes de la decapitación.

Visto el Día 8

Visto el Día 8: 'Las ratas aprenden a los golpes' - video que muestra a hombres jóvenes posando con armas antes de un disparo que los muestra siendo golpeados por las autoridades.

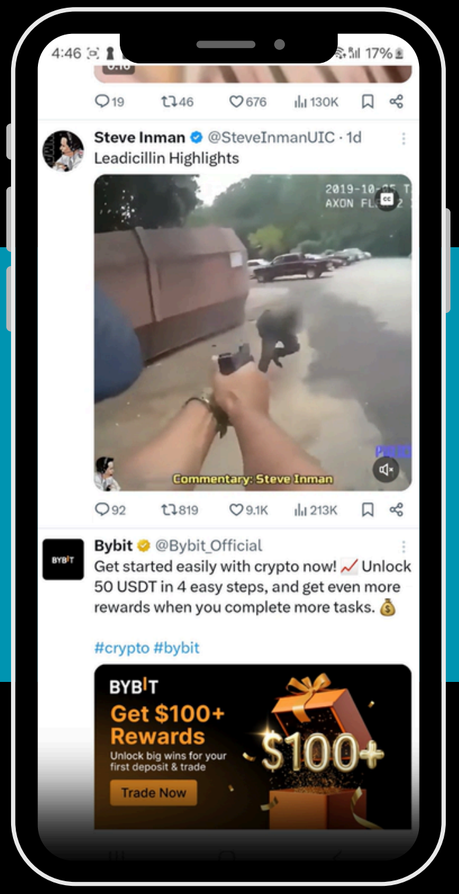
ESTUDIO DE CASO: Contenido violento

Fase dos: ‘buscando’

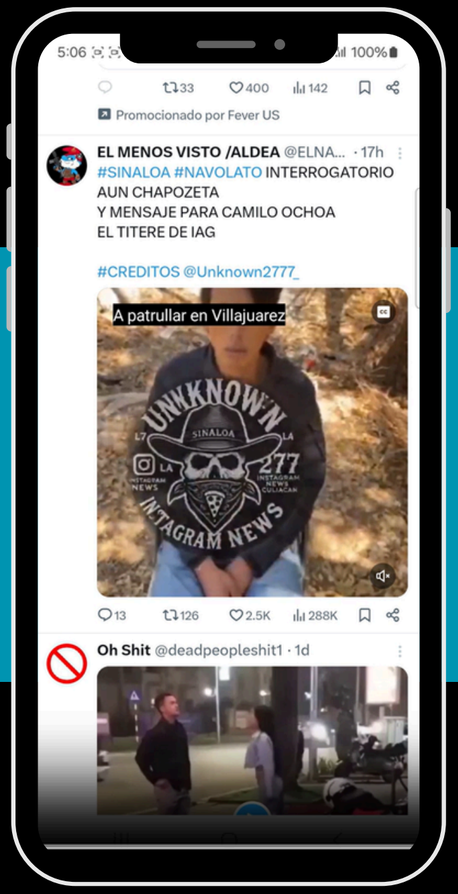
La búsqueda de términos como 'Sinaloa' o 'enfrentamientos' aumentó la exposición del avatar a contenido violento y continuó mostrándolo en el feed del avatar. Mientras que en una primera instancia el periodismo y la cobertura de noticias eran quienes mostraban la mayor parte de las imágenes gráficas, esto fue cambiando al ser mostrado posteriormente por cuentas dedicadas a la violencia o al ‘gore’.

Según las Reglas de X, el contenido violento está permitido en X, siempre que esté 'Puedes compartir contenido gráfico si está correctamente etiquetado, no se exhibe de forma prominente y no es excesivamente sangriento'. Las Políticas de X para menores de edad también indican que la plataforma tiene como objetivo restringir 'ciertos tipos de contenido multimedia delicado' a 'menores conocidos [de 13 a 17 años]' - aunque no se especifica claramente si esto incluye contenido violento de cualquier tipo.

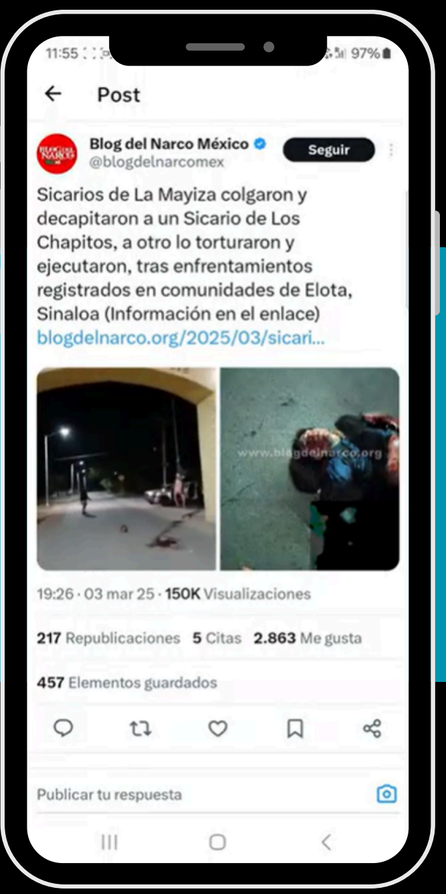
En el caso del avatar infantil, este mostró que las y los adolescentes pueden estar expuestos a este contenido con casi la misma intensidad que el avatar adulto.



Visto en el Día 11: Imágenes de la cámara corporal de un oficial de policía disparando a un individuo.



Visto el Día 12: Imágenes de un interrogatorio.



Visto por el avatar con edad de niño el Día 2: video e imagen de un ‘sicario’ decapitado colgando de un puente.

EXPOSICIÓN

Cómo las funciones de búsqueda revelaron contenido dañino

En algunos casos, el material dañino solo apareció después de buscar un término relevante. Estos términos de búsqueda se identificaron basándose en etiquetas y términos vistos en grabaciones de pantalla compartidas por las y los adolescentes durante las entrevistas, y se complementaron con el asesoramiento de una investigadora de México.

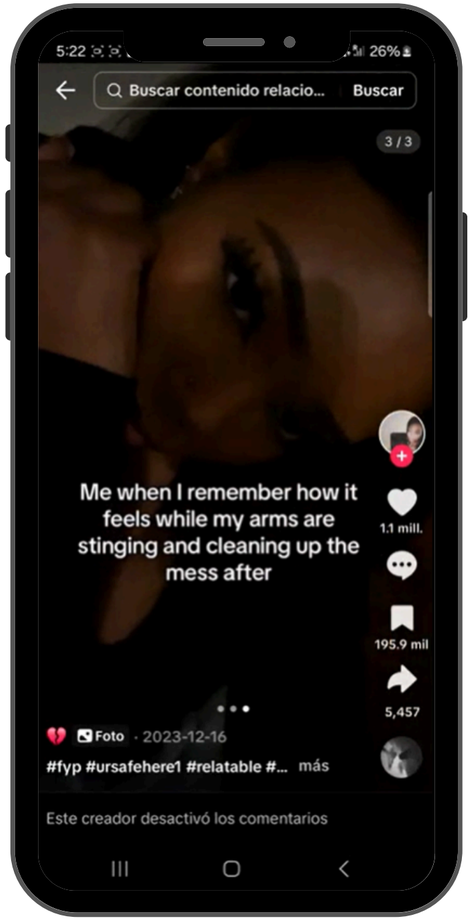
Por ejemplo, para el avatar de autolesiones y suicidio, en la fase inicial del 'input' se encontró contenido de profunda reflexión y depresión, sin embargo, fue hasta la etapa de búsqueda de palabras relacionadas con el tema que le apareció el contenido explícito sobre dichas temáticas. Es entonces que, dicho avatar pudo acceder a referencias explícitas de suicidio después de buscar términos como 'su1cui' y '$H' -identificados como términos comunes que evaden los filtros de moderación de la plataforma.

El avatar que exploraba la exposición a contenido relacionado con trastornos alimenticios solo fue expuesto a contenido relacionado a la temática a través de búsquedas directas, como 'bodygoals'.

En X, al avatar con edad de niño que exploraba la violencia, se le mostraba regularmente contenido violento. Esto incluía la posibilidad de ver un video sin censura de las consecuencias de una ejecución en Sinaloa después de buscar 'enfrentamientos'.

Cuando los avatares buscaban términos relacionados con cada tema, a menudo se les mostraba contenido potencialmente dañino.

Mientras que algunos términos, como 'bodygoals' y 'chicas sexys', activaron advertencias de seguridad o fueron bloqueados - particularmente para cuentas de menores de edad -, a la mayoría le arrojó contenido sin restricciones ni avisos.



Buscando 'SH' en la cuenta con edad adulta: 'Yo cuando recuerdo cómo me siento mientras me arden los brazos y limpio el desastre después.

EXPOSICIÓN

Cómo las funciones de búsqueda revelaron contenido dañino

Para reflejar el comportamiento en el mundo real de las y los adolescentes que usan cuentas con edad de adulto, la mayoría de los avatares se configuraron inicialmente usando las edades que las y los participantes habían ingresado al registrarse. Se configuraron perfiles adicionales para cada avatar, usando la edad real de las y los adolescentes para observar qué muestran las plataformas cuando el usuario se configura como una persona menor de edad.

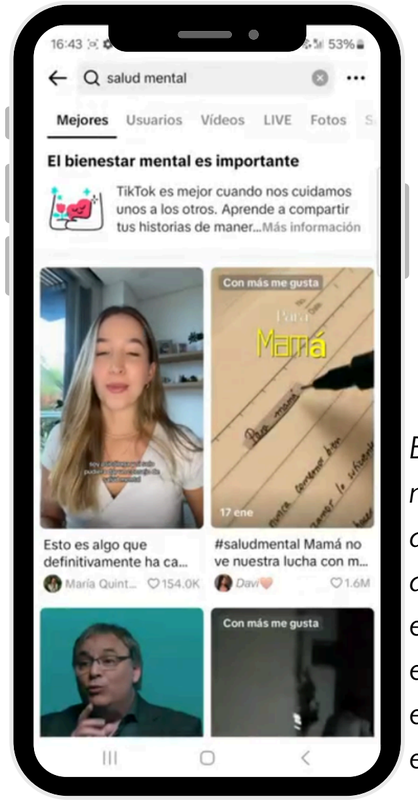
En algunos casos, los avatares con perfiles de edad infantil fueron bloqueados al buscar términos específicos de alto riesgo, especialmente aquellos vinculados a trastornos alimenticios o contenido sexual. Sin embargo, el contenido de autolesiones y el material emocionalmente intenso seguían siendo accesibles y, en algunos casos, activamente recomendados.

Esto sugiere que, si bien algunas protecciones pueden estar implementadas para usuarios menores de edad, son inconsistentes y de alcance limitado.

EXPOSICIÓN

Cómo las funciones de búsqueda revelaron contenido dañino

Al avatar con edad infantil no se le mostró contenido al buscar 'chicas sexys'.



Buscar 'salud mental' en TikTok activó una advertencia tanto en el avatar con edad adulta como en el avatar con edad de menor.

Buscar 'bodygoals' no expuso a ninguno de los avatares relacionados con trastornos alimenticios a 'reels' al hacer la búsqueda, pero sí los expuso a contenido en otras partes de la plataforma con una advertencia de salud.

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

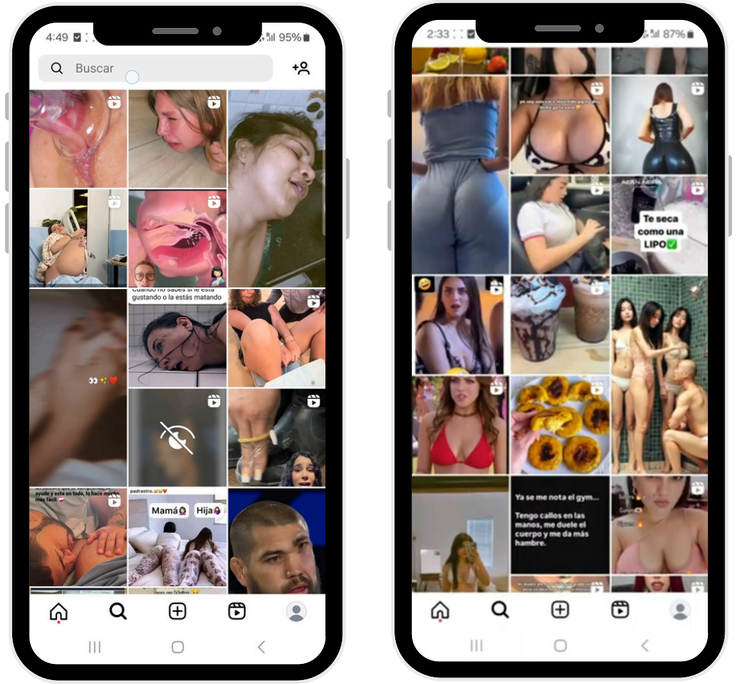
EXPOSICIÓN

IA, clickbait y contenido sugestivo en las páginas de exploración

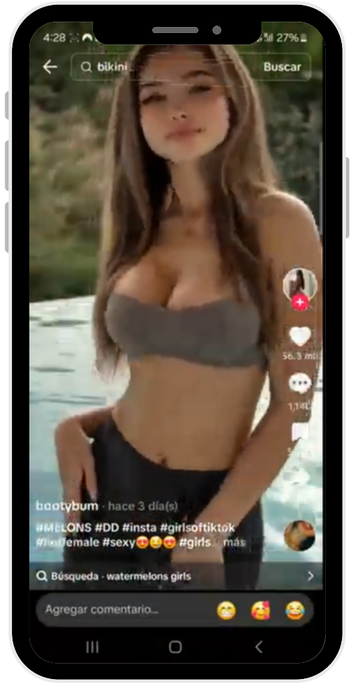
El contenido en las páginas de exploración, en los últimos días para algunos avatares, parecía estar diseñado para captar la atención, especialmente mediante miniaturas que daban la impresión de mostrar contenido más explícito o perturbador. Sin embargo, al hacer clic en los videos, el contenido a menudo resultó ser más inofensivo.

Junto con el contenido mostrado en la página de exploración, términos de búsqueda como 'belleza' llevaron a videos generados por inteligencia artificial (IA) de contenido sexualmente sugestivo a los avatares. Aunque no eran explícitos, los videos incluían mujeres en bikinis y algunas poses hipersexualizadas.

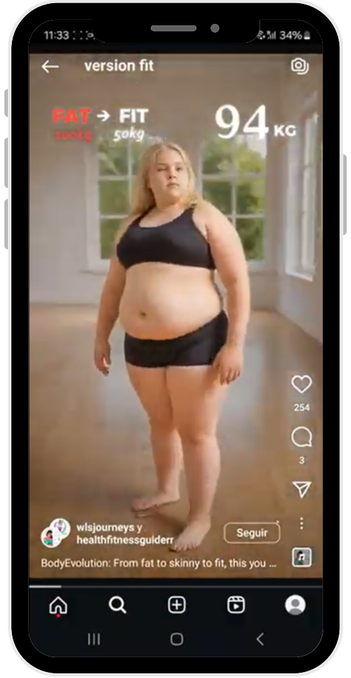
Contenido generado por IA también apareció en anuncios dirigidos a diversos avatares menores de edad. Por ejemplo, un video de una mujer que perdía la mitad de su peso corporal, pasando de 'gorda a delgada', apareció como el primer resultado tras buscar 'versión fit'.



Contenido de 'clickbait' y de IA visto por un perfil de avatar en Instagram



Un video generado por IA mostrado al avatar de contenido sexual al buscar 'bikini'.



El avatar de la edad infantil fue expuesto a un video de pérdida de peso de IA buscando 'versión fit', donde la chica generada pierde la mitad de su peso.

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

INCENTIVOS

Las características de diseño parecían incentivar a las y los usuarios a seguir viendo

INCENTIVOS

Características que parecieron incentivar a las y los usuarios a seguir viendo

Las plataformas de redes sociales están diseñadas principalmente para maximizar la participación de las y los usuarios. Sus modelos de negocio dependen de mantenerles activos(as) el mayor tiempo posible, fomentando el uso repetido, la atención sostenida y la conexión emocional. Esto a menudo significa priorizar características que impulsan el acceso y la participación.

En TikTok, este enfoque en la participación se manifiesta en las decisiones de diseño, como un feed impulsado por algoritmos o un flujo infinito de contenido, que promueven el desplazamiento constante, la retroalimentación instantánea y las recomendaciones de contenido algorítmicas. Si bien estas características pueden apoyar el descubrimiento o el entretenimiento, también pueden influir en lo que las y los usuarios ven y cuánto tiempo permanecen, especialmente las y los adolescentes cuyos comportamientos en línea y respuestas emocionales pueden ser diferentes a los de las personas adultas. Varios de los avatares de edad adulta en TikTok mostraron características (monedas) que parecían recompensar la cantidad del tiempo dedicado a ver videos. Estas incluían superposiciones y ventanas emergentes que ofrecían pequeñas cantidades de 'monedas' si la o el usuario continuaba viendo el video; por ejemplo, 'gana 10 monedas por cada 30 segundos vistos'.

Los anuncios entre videos también ofrecían monedas a cambio de invitar a amistades a unirse a TikTok, lo cual también apareció en los avatares menores de edad. Las monedas se acumulaban en un contador visible, pero su propósito y valor no se encontró. Esto parece ser parte del programa de recompensas de TikTok Lite, el cual está prohibido en la Unión Europea, pero disponible para usuarias y usuarios en México.

REVEALING REALITY

Instituto Nacional de Salud Pública

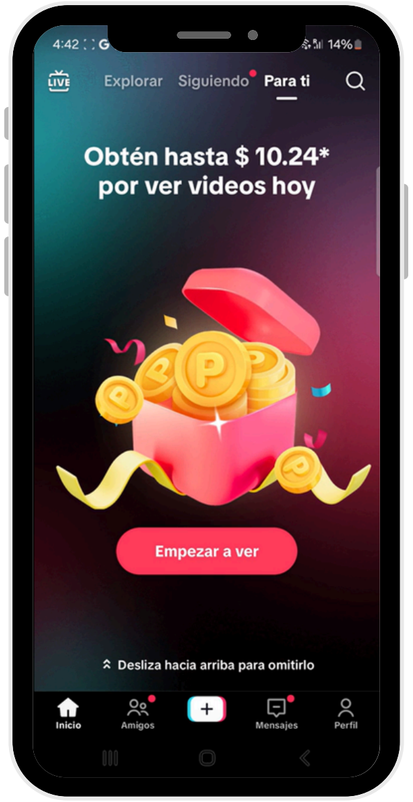
5RIGHTS FOUNDATION

INCENTIVOS

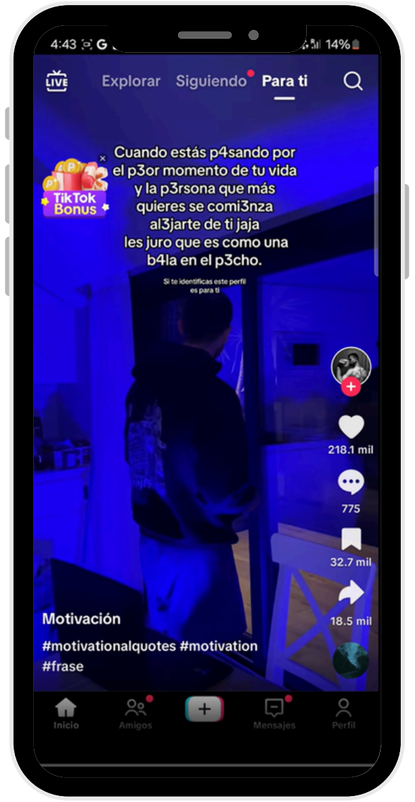
Características que parecieron incentivar a las y los usuarios a seguir viendo

Estas características solo se observaron en un número de casos y no fueron consistentes en todos los avatares. Sin embargo, donde sí aparecieron, parecían introducir un elemento similar a un juego, lo que potencialmente animaba a las y los usuarios a ver videos por períodos más largos a cambio de recompensas percibidas.

En un caso, la función de las monedas apareció y luego mostró una serie de videos con carga emocional, incluyendo contenido sobre pérdida, ansiedad y tristeza.



Pop-up de TikTok visto en el Avatar de salud mental.



El siguiente video: 'Cuando estás p4sando por el p3or momento de tu vida y la p3rsona que más quieres se comi3nza al3jar de ti jaja, les juro que es como una b4la en el p3cho".

El video después: '¿Cómo le explico a mi sistema inmunologico que solo estoy teniendo mucho estrés y ansiedad y no por eso me tiene que doler tando la cabeza, darme náuseas y perder el apetito?"

INCENTIVOS

Características que parecieron incentivar a las y los usuarios a seguir viendo

Aunque la función del sistema de monedas no está clara para la o el usuario, su presencia sugiere un mecanismo que puede fomentar un uso prolongado. Para las y los adolescentes, quienes quizás no comprendan completamente la naturaleza o el valor de las recompensas digitales, estas características podrían contribuir a un tiempo prolongado en la plataforma.

El sistema de recomendación de TikTok aprende del comportamiento de la o del usuario: cuanto más tiempo pasa una persona viendo un tipo particular de contenido, más material similar le muestra el algoritmo. Cuando se combina con los incentivos basados en monedas, este sistema podría contribuir a patrones de visualización donde el contenido emocionalmente intenso se muestra con mayor frecuencia que otros tipos de materiales.

Estos incentivos podrían animar a las y los adolescentes a ver videos completos en lugar de deslizarlos, lo que contribuye a un ciclo de retroalimentación que recompensa el compromiso continuo, sin importar el tema o el tono del contenido que se muestra. Si bien estos sistemas están diseñados para optimizar el interés y no el daño, el resultado aún podría ser una experiencia que atrae a las y los adolescentes en bucles de contenido sin oportunidades para pausar, reflexionar o redirigir.

CONCLUSIÓN

Resumen de hallazgos clave y recomendaciones

CONCLUSIÓN

Hallazgos clave

En México, las y los adolescentes pueden estar expuestos a contenido potencialmente dañino en diversas temáticas. Unos pocos minutos de desplazamiento diario en avatares que imitan los intereses y comportamientos de adolescentes reales, pueden llevar a feeds llenos de contenido relacionado con la depresión, la autolesión, la violencia y el contenido sexual.

Las plataformas no solo están fallando en su obligación de respetar los derechos de las y los adolescentes en cuanto a políticas, sino que también están fallando en la aplicación, la moderación y el diseño. En muchos casos, esto está en desacuerdo con los estándares y las reglas delineadas en sus propias guías y políticas.

Las plataformas alimentan a las y los adolescentes con contenido potencialmente dañino a través de recomendaciones y no toman medidas significativas para evitar que accedan y permanezcan en esta trayectoria.

Las trayectorias dañinas (por ejemplo, del interés en salud mental al contenido de autolesión) son rápidas y fáciles, y pueden ser incluso recompensadas por la propia plataforma.

Las medidas de protección actuales de las plataformas, son insuficientes para cumplir con sus propios estándares y con las responsabilidades nacionales o internacionales en cuanto a los derechos de las y los adolescentes.

CONCLUSIÓN

Recomendaciones clave

Empresas tecnológicas

Respetar los derechos de las niñas, niños y adolescentes (NNA) y prevenir violaciones, incorporando la seguridad, la privacidad y los derechos de la niñez desde el diseño y por defecto en todos los productos y servicios digitales que probablemente sean utilizados por NNA, en consonancia con las obligaciones internacionales y con la Observación General núm. 25 del Comité de los Derechos del Niño.

- Realizar evaluaciones tempranas, proactivas y continuas del impacto en los derechos de los NNA a lo largo de todo el ciclo de vida de los productos y servicios, e incluir información detallada sobre los riesgos y las medidas de mitigación.
- Introducir una verificación de la edad proporcionada, eficaz y que preserve la privacidad en el momento del registro para dirigir a las NNA hacia experiencias adecuadas a su edad.
- Poner fin a los modelos de negocio manipuladores y explotadores, incluidos los algoritmos basados en la participación y los sistemas de recompensas engañosos que fomentan el uso excesivo, el comportamiento compulsivo o la explotación comercial de las NNA.
- Garantizar una aplicación sólida, coherente y transparente de las condiciones de servicio, las políticas y las normas de la comunidad.

CONCLUSIÓN

Recomendaciones clave

Gobiernos y Legisladores

Las NNA tienen derecho a la protección contra la explotación comercial, a la seguridad y a la privacidad, independientemente del lugar en el que participen en el entorno digital. Estos derechos están garantizados por el derecho internacional, en particular la Convención de las Naciones Unidas sobre los Derechos del Niño, tal y como lo interpreta de manera autorizada el Comité de los Derechos del Niño en su Observación general n.º 25 sobre los derechos de las NNA s en relación con el entorno digital.

De acuerdo con estas obligaciones, los Estados deben velar por que los derechos de las niñas, niños y adolescentes sean respetados, protegidos y garantizados en el entorno digital, y que su interés superior constituya una consideración primordial en todas las acciones relativas al diseño, la regulación, la gobernanza y el funcionamiento de los servicios digitales

Para cumplir con sus obligaciones internacionales, los gobiernos y los legisladores deben exigir responsabilidades a las empresas tecnológicas, entre otras cosas:

- Priorizar la adopción, implementación y cumplimiento de leyes, regulaciones, políticas y normas técnicas específicamente dirigidas a proteger los derechos de las NNA en el entorno digital.
- Legislar y hacer cumplir las obligaciones de las empresas de respetar los derechos de las NNA, incluidas las obligaciones de prevenir, mitigar y remediar los efectos adversos reales y potenciales sobre las NNA.

CONCLUSIÓN

Recomendaciones clave

- Incluir una obligación general y legalmente exigible de diligencia para los servicios a los que puedan acceder las NNA, y exigir la aplicación de enfoques obligatorios de «seguridad desde el diseño» y «privacidad desde el diseño» a lo largo de todo el ciclo de vida de los productos y servicios digitales.
- Exigir una diligencia debida rigurosa, incluyendo evaluaciones obligatorias del impacto Evaluaciones de Impacto en los Derechos de la Niñez (Child Rights Impact Assessments (CRIAs)) que aborden los riesgos relacionados con el contenido, el contacto, la conducta y los contratos, así como los daños sistémicos y transversales.
- Prohibir los modelos de negocio manipuladores, engañosos y explotadores que socavan los derechos de las NNA o se aprovechan de sus vulnerabilidades.
- Hacer cumplir los términos de servicio, las políticas y las normas comunitarias de las plataformas.

REVEALING REALITY

Instituto Nacional de Salud Pública

5RIGHTS FOUNDATION

El hilo de Ariadne