

EMPUJADOS A LA OSCURIDAD

EL FEED “PARA TI” DE TIKTOK FOMENTA LA AUTOLESIÓN Y LA IDEACIÓN SUICIDA

AMNISTÍA
INTERNACIONAL



Amnistía Internacional es un movimiento integrado por 10 millones de personas que activa el sentido de humanidad dentro de cada una de ellas y que hace campaña en favor de cambios que permitan que todo el mundo disfrute de sus derechos humanos. Nuestra visión es la de un mundo donde quienes están en el poder cumplen sus promesas, respetan el derecho internacional y rinden cuentas. Somos independientes de todo gobierno, ideología política, interés económico y credo religioso, y nuestro trabajo se financia principalmente con las contribuciones de nuestra membresía y con donativos. Creemos que actuar movidos por la solidaridad y la compasión hacia nuestros semejantes en todo el mundo puede hacer mejorar nuestras sociedades.

© Amnesty International 2023

Salvo cuando se indique lo contrario, el contenido de este documento está protegido por una licencia 4.0 de Creative Commons (atribución, no comercial, sin obra derivada, internacional) (véase <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.es>)

El material atribuido a titulares de derechos de autor distintos de Amnistía Internacional no está protegido por la licencia Creative Commons.

Para más información, visiten la página Permisos del sitio web de Amnistía Internacional.

Publicado por primera vez en 2023

por Amnesty International Ltd

Peter Benenson House, 1 Easton Street

London WC1X 0DW, Reino Unido



Foto de portada: © Luisa Balaban

Índice: POL 40/7350/2023 SPANISH

Idioma original: Inglés

amnesty.org

**AMNISTÍA
INTERNACIONAL**



ADVERTENCIA SOBRE EL CONTENIDO

Este informe aborda algunas cuestiones delicadas, como las autolesiones y el suicidio, y contiene imágenes gráficas de contenidos de TikTok relacionados con esos temas. En la guía de Amnistía Internacional ***Resiliencia y autocuidado: Salvar al mundo y no morir en el intento (Volumen 2). Cuaderno de trabajo para activistas jóvenes.***¹

AGRADECIMIENTOS

Amnistía Internacional desea dar las gracias a sus socios técnicos, el Instituto de Transparencia Algorítmica (Conferencia Nacional de Ciudadanía) [Algorithmic Transparency Institute (National Conference on Citizenship)] e Investigación Forense de la Inteligencia Artificial (AI Forensics), así como a las organizaciones y las personas particulares, activistas y especialistas que facilitaron y orientaron esta investigación. Éstas son, entre otras, S. Bryn Austin, Zeyna Awan, Jerome Cleofas, Julien Cornebise, Kevin Gachee, Servicios Comunitarios En Contacto, Lillian Kariuki, KERI, Amanda Raffoul, Marc Reyes, Aliya Shah, Talang Dalisay, Ian Wafula y Coalición de Jóvenes por la Salud Mental.

Amnistía Internacional está también muy agradecida a los niños, niñas y jóvenes de Kenia y Filipinas que relataron sus experiencias y expresaron sus ideas durante la investigación realizada para este informe.

1. Amnistía Internacional, *Resiliencia y autocuidado: Salvar al mundo y no morir en el intento (Volumen 2). Cuaderno de trabajo para activistas jóvenes* (Índice: ACT 10/3231/2020), enero de 2021, [amnesty.org/es/documents/act10/3231/2020/es/](https://www.amnesty.org/es/documents/act10/3231/2020/es/), p. 99.

ÍNDICE

Glosario	5
1. RESUMEN EJECUTIVO	6
2. METODOLOGÍA	11
2.1 Propósito y alcance del informe	11
2.2 Eje de investigación	12
2.3 Metodología de investigación	13
3. INFORMACIÓN GENERAL	16
3.1 No entender nada: Lagunas de investigación y obstáculos de las empresas al estudio de los riesgos en materia de derechos humanos que las plataformas de redes sociales suponen para los niños, niñas y jóvenes	16
3.2 Entre la política y el populismo: escasos esfuerzos para regular las plataformas de redes sociales	19
3.3 Intensificación de los problemas de salud mental entre los niños, niñas y jóvenes	22
4. MARCO DE DERECHOS HUMANOS	24
4.1 El derecho a la privacidad en la era digital	25
4.2 El derecho a no sufrir intromisiones	27
4.3 El derecho a controlar nuestra información personal	29
4.4 El interés superior de la infancia, y el derecho de los niños y niñas a ser escuchados	31
4.5 La responsabilidad empresarial de respetar los derechos humanos	31
5. ADICTIVO POR DISEÑO	34
5.1 Estrategias de TikTok para la interacción	36
5.2 Cómo puede afectar el diseño adictivo de las redes sociales a la salud de las personas jóvenes	38
6. ESPIRALES DESCENDENTES	42
6.1 El feed 'Para ti' de TikTok lleva al extremo un sistema inherentemente peligroso	42
6.2 Mayores riesgos para las personas jóvenes con problemas de salud mental	46
6.3 Explorar el sistema de recomendación de TikTok sistemáticamente	47
6.4 Medición y categorización de los contenidos nocivos a efectos de la investigación	49
6.5 Conclusiones destacadas	52
6.6 El interés por la salud mental desencadena el efecto espiral	53
6.7 Examples of recommended posts	62
6.8 Limitaciones del proyecto	66
7. FALLOS EMPRESARIALES	68
7.1 Falta de la diligencia debida adecuada	68
7.2 Riesgos graves combinados con respuestas inadecuadas	71
8. CONCLUSIÓN Y RECOMENDACIONES	74
Anexo	80

GLOSARIO

TDAH	Trastorno por déficit de atención con hiperactividad
IA	Inteligencia artificial. No existe una definición generalmente aceptada del término “inteligencia artificial” o “IA”. La Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos emplea este término para hacer referencia a una constelación de procesos y tecnologías que permite que los ordenadores complementen o reemplacen tareas específicas que de otro modo serían ejecutadas por seres humanos como tomar decisiones y resolver problemas, lo que comprende, entre otras cosas, el aprendizaje automático y el aprendizaje profundo. ²
Big Tech	Grandes tecnológicas (<i>big tech</i>): Forma popular de referirse a las principales empresas de tecnologías de la información que determinan y dominan la Internet.
CDC	Centros para el Control y la Prevención de Enfermedades (Estados Unidos)
CdE	Consejo de Europa
CDN	Convención sobre los Derechos del Niño
Data scraping	Extracción de datos (<i>scrapping</i>): Recopilación automatizada de datos de sitios web a gran escala
LSD	Ley de Servicios Digitales de la Unión Europea
TEDH (en las notas al pie)	Tribunal Europeo de Derechos Humanos
CDH (en las notas al pie)	Comité de Derechos Humanos de la ONU
PIDCP	Pacto Internacional de Derechos Civiles y Políticos
PIDESC	Pacto Internacional de Derechos Económicos, Sociales y Culturales
LGBTI	Personas lesbianas, gays, bisexuales, transgénero e intersexuales
Metadatos	Información sobre los archivos y las comunicaciones digitales como, por ejemplo las marcas de tiempo y los datos de localización
Líneas Directrices de la OCDE	Líneas Directrices de la OCDE para Empresas Multinacionales
OACNUDH	Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos
“Rabbit hole” effect	Efecto espiral (del inglés, “ <i>rabbit hole</i> ” effect o efecto “madriguera de conejo”): Expresión que suele utilizarse en el estudio de las plataformas de redes sociales; en este informe se utiliza para hacer referencia a la reducción de las recomendaciones de contenidos del feed “Para ti” de TikTok a un solo tema, partiendo del supuesto de que ese contenido provocará intensas reacciones emocionales y eso mantendrá interesada a la persona usuaria.
DUDH	Declaración Universal de Derechos Humanos
Principios rectores de la ONU	Principios rectores de la ONU sobre las empresas y los derechos humanos
VPN	Red privada virtual

2. Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (OACNUDH), “El derecho a la privacidad en la era digital. Informe del Alto Comisionado de las Naciones Unidas para los Derechos Humanos”, 15 de septiembre de 2021, doc. ONU A/HRC/48/31, nota al pie 2.

1. RESUMEN EJECUTIVO

“¿Alguna vez has mirado un bote de pastillas y pensado en sufrir una sobredosis? ¿Alguna vez has deseado que todo llegara a su fin? ¿Alguna vez has pensado en la liberación que sería, en todo el dolor que desaparecería?”. Con unas figuras pequeñas pixeladas que se desplazan por el laberinto de un juego de arcade de fondo, una voz generada por inteligencia artificial pronuncia estos pensamientos en un fragmento de vídeo de medio minuto de duración. Cuando mira, mira de nuevo y luego pasa al siguiente fragmento, la persona de 13 años usuaria de TikTok que navega por estos vídeos no siente nada. Eso se debe a que es una cuenta automatizada, configurada y programada por Amnistía Internacional y el Instituto de Transparencia Algorítmica para simular y estudiar la realidad digital de los niños, niñas y jóvenes que sufren problemas de salud mental tales como ansiedad o depresión. A cada hora transcurrida en el feed “Para ti” de TikTok aumenta el número de fragmentos de vídeo “recomendados” a la cuenta adolescente que muestran niños, niñas y jóvenes llorando, o solos en la oscuridad con un texto superpuesto que expresa pensamientos depresivos o voces sin rostro que describen sufrimiento, autolesiones y pensamientos suicidas.

En los últimos tres años, marcados por la pandemia de COVID-19, TikTok se ha convertido en una plataforma global y ha atraído a centenares de millones de niños, niñas y jóvenes. La página “Para ti” de TikTok y el sistema algorítmico de recomendación que hay detrás de ella han sido decisivos para catapultar la plataforma a su ubicuidad actual en la vida de los niños, niñas y jóvenes. En la publicación *“Siento que no se me protege”: Emboscada en la red de vigilancia de TikTok*, complementaria a este trabajo, se pone al descubierto el modelo profundamente invasivo y discriminatorio de vigilancia con fines lucrativos que sustenta el modelo empresarial de TikTok. El presente estudio se centra en la vigilancia de las personas usuarias, las decisiones de diseño y las recomendaciones de contenido personalizado empleadas por TikTok en aras de la “interacción de las personas usuarias”, así como en los efectos negativos que esas medidas tienen en la vida real de los niños, niñas y jóvenes.

El sistema de recomendación de TikTok —el conjunto de algoritmos que analizan los intereses y las constantes de interacción de las personas usuarias, y luego relaciona esa información con nuevos fragmentos de vídeo para mantener a la gente conectada desde el punto de vista emocional y con la vista fija en el feed— no se creó para generar espirales de contenidos depresivos. Sin embargo, si una persona joven muestra un interés claro por un contenido relacionado con la salud, eso es con lo que el sistema la relaciona en aras de la interacción óptima de las personas usuarias. Si la cuenta automatizada de investigación configurada para este estudio fuera un ser humano, más capaz de distinguir entre contenidos relevantes y contenidos irrelevantes, el sistema generaría esas recomendaciones de forma aún mucho más rápida.

“Luis” (nombre ficticio), estudiante universitario de 21 años de Manila diagnosticado con trastorno bipolar, describió a Amnistía Internacional su experiencia con el feed “Para ti” de TikTok de este modo:

“Cuando me sentía decaído, creo que el 80% [del contenido] tenía que ver con la salud mental. Es como una espiral. Es una madriguera en la que te adentras a partir de un solo vídeo. Si un vídeo logra captar tu atención, aunque no te guste, te lo presentan cuando vuelves a abrir TikTok y, como te resulta familiar, lo vuelves a ver. Al verlo de nuevo, su frecuencia en tu feed aumenta exponencialmente”.

El presente informe se basa en investigación documental, una encuesta de evaluación del alcance, y debates con grupos-muestra y entrevistas realizados con más de 300 niños, niñas y jóvenes de 2 de los países que presentan las cifras promedio más elevadas del mundo de uso de redes sociales (Kenia y Filipinas), así como una indagación técnica —realizada junto con el Instituto de Transparencia Algorítmica e Investigación Forense de la Inteligencia Artificial como socios técnicos— con más de 30 cuentas automatizadas que comparaban los resultados obtenidos en Kenya y Estados Unidos de América (Estados Unidos), y varios experimentos manuales llevados a cabo en Kenya, Filipinas y Estados Unidos.

Para la auditoría automatizada, el equipo investigador creó 40 cuentas automatizadas con 4 personalidades distintas predefinidas, que mostraban distintos niveles de interés por los contenidos relacionados con la salud mental a fin de imitar el comportamiento de diferentes niños y niñas en TikTok. Cada cuenta se configuró para estar activa durante algo menos de 60 minutos en una única sesión diaria a lo largo de 10 días. Las cuentas se dividieron en subgrupos que mostraban cuatro formas diferentes de desplazar la pantalla, y cada subgrupo miraba de nuevo los vídeos asociados con un conjunto distinto de términos y etiquetas, y se saltaba los contenidos no relacionados. Se configuraron 20 cuentas en Estados Unidos y otras 20 en Kenya (de las cuales 11 se incluyeron en el análisis) para simular a niños y niñas de 13 años. La edad indicada en las cuentas (13 años) se eligió con vistas a examinar las recomendaciones ofrecidas al grupo etario más joven permitido entre quienes usan TikTok, cuentas a las que esta plataforma aplica medidas de seguridad para adolescentes. El experimento manual adicional consistió en grabar la pantalla durante una hora de interacción del personal investigador con el feed “Para ti” en cuentas recién creadas, de nuevo en representación de personas de 13 años.

La investigación de Amnistía Internacional pone en evidencia que TikTok ha aumentado al máximo las cualidades adictivas de las opciones de diseño y las estrategias de interacción empleadas por empresas rivales del sector, e incentivado la armonización a la baja entre un pequeño número de las principales empresas de redes sociales que compiten por conseguir las cifras y tasas de interacción de personas usuarias más altas. TikTok ha hecho esto a pesar de la creciente evidencia científica de los graves riesgos asociados con el uso adictivo de las redes sociales sobre todo para la salud de los niños, niñas y jóvenes, que incluyen problemas de sueño y de atención, e incluso cambios en la estructura cerebral similares a los observados en personas adictas a las drogas.

Más allá de su carácter adictivo, el feed “Para ti” de TikTok entraña riesgos adicionales para los niños, niñas y jóvenes con problemas previos de salud mental. La indagación técnica realizada por Amnistía Internacional, el Instituto de Transparencia Algorítmica (Conferencia Nacional de Ciudadanía) e Investigación Forense de la Inteligencia muestra que los niños, niñas y jóvenes que ven contenidos relacionados con la salud mental en la página “Para ti” de TikTok pueden verse arrastrados fácilmente hacia espirales de contenidos potencialmente nocivos, incluidos vídeos que idealizan y fomentan el pensamiento depresivo, las autolesiones y el suicidio.

Al cabo de 5 o 6 horas en la plataforma, prácticamente la mitad de los vídeos mostrados a las cuentas automatizadas que se habían programado para que simularan a niños y niñas de 13 años de Kenya y Estados Unidos con interés por la salud mental guardaban relación con ésta y eran potencialmente nocivos, cifra que multiplicaba aproximadamente por 10 el volumen mostrado a las cuentas sin interés por

la salud mental. La revisión manual de 540 vídeos recomendados a una muestra de estas cuentas bots mostraron una progresión constante desde el 17% de los vídeos mostrados en la primera hora y categorizados como potencialmente nocivos hasta el 44% del contenido de la décima hora (datos basado en sesiones de una hora de duración repartidas a lo largo de 10 días).

Amnistía Internacional observó un efecto espiral aún más rápido y con tasas aún mayores de contenidos potencialmente nocivos cuando el equipo investigador miraba de nuevo los vídeos relacionados con la salud mental sugeridos a supuestas personas de 13 años de Kenia, Filipinas y Estados Unidos. Entre las recomendaciones mostradas a una cuenta localizada en Filipinas, el primer vídeo etiquetado con #depresionanxiety [sic] que mostraba a un niño angustiado fue sugerido en los primeros 67 segundos de exploración del contenido recomendado en la página “Para ti”. A partir del minuto 12, el 58% de las publicaciones recomendadas estaban relacionadas con la ansiedad, la depresión, las autolesiones o el suicidio, y se clasificaron como potencialmente nocivas para niños, niñas y jóvenes con problemas de salud mental previos.

En el experimento manual llevado a cabo en Estados Unidos, el cuarto vídeo mostrado llevaba la etiqueta #paintok y se centraba en un texto con el mensaje siguiente: “Cuando te das cuenta de que nunca en tu vida has ocupado el primer lugar, sino que eres sólo esa persona que llena un vacío en la vida de otras hasta que dejas de hacerles falta”. A partir del vídeo 20 (en menos de 3 minutos), el 57% de los vídeos están relacionados con problemas de salud mental y al menos 9 de las publicaciones idealizaban, normalizaban o fomentaban el suicidio en una sola hora.

La cuenta de Kenia en los experimentos manuales fue la que más lentamente avanzó hacia un feed de contenidos depresivos. Sin embargo, una vez alcanzado ese punto (a los 20 minutos de comenzar el experimento), el 72% de los vídeos recomendados en los 40 minutos siguientes estaban relacionados con problemas de salud mental, incluidas al menos 5 referencias a pensamientos suicidas o al deseo de morir de la persona autora del contenido. Ni un solo vídeo relativo a la salud mental había sido publicado por especialistas en salud mental u organizaciones reconocidas en la materia.

El caso de Molly Russell, la niña británica de 14 años que murió por autolesiones tras haber visto contenidos depresivos en Instagram, muestra de la forma más trágica que exponer a una persona joven con síntomas depresivos a un feed de redes sociales compuesto por un alto volumen de publicaciones que comentan, normalizan o incluso idealizan el pensamiento depresivo, las autolesiones y el suicidio tiene el potencial de intensificar problemas de salud mental preexistentes en personas usuarias jóvenes y podría contribuir a la comisión de actos nocivos e incluso devastadores en el mundo real. Este tipo de injerencia en los pensamientos y emociones de las personas constituye un abuso contra el derecho a la libertad de pensamiento y el derecho a la salud.

Dadas las bien documentadas vulnerabilidades emocionales de niños, niñas y jóvenes, y la amplia evidencia basada en informes previos de la sociedad civil y los medios de comunicación, TikTok debería saber y haber identificado que sus sistemas de recomendación algorítmicos crean el riesgo de exponer a las personas usuarias jóvenes a espirales de publicaciones potencialmente nocivas.

Para cumplir con sus responsabilidades según lo establecido en los Principios rectores de la ONU sobre las empresas y los derechos humanos, TikTok debería llevar a cabo la adecuada diligencia debida en materia de derechos humanos para identificar, prevenir y mitigar los daños potenciales y reales, y explicar cómo los aborda. Como parte de la diligencia debida en este ámbito, esta empresa debería haber identificado los riesgos para niños, niñas y jóvenes inherentes al diseño de su plataforma teniendo en cuenta la creciente evidencia de los riesgos sistémicos asociados con los sistemas algorítmicos de recomendación reunida por los medios de comunicación, las organizaciones de la sociedad civil y las instituciones internacionales, y haber tomado medidas para mitigarlos. La respuesta de TikTok a las preguntas de Amnistía Internacional apunta a un conjunto heterogéneo de

medidas individuales tales como redirigir determinadas búsquedas hacia recursos de ayuda, ofrecer la posibilidad de “actualizar” el feed y establecer alianzas con organizaciones de apoyo a la salud mental. Esta respuesta demuestra la carencia de procedimientos adecuados para abordar el carácter sistémico de estos riesgos, y las medidas y herramientas de seguridad aplicadas por la empresa distan mucho de abordar la magnitud de los riesgos sistémicos identificados en este informe.

No obstante, las medidas particulares de tan sólo una empresa no bastan para refrenar un modelo empresarial que es intrínsecamente incompatible con los derechos humanos, concretamente el derecho a la privacidad, el derecho a la libertad de pensamiento y el derecho a la salud. Por consiguiente, los Estados deben controlar las “grandes empresas tecnológicas” como TikTok de conformidad con las leyes y normas internacionales de derechos humanos a fin de proteger y hacer efectivos los derechos de los niños, niñas y jóvenes.

Amnistía Internacional pide a TikTok que aplique con urgencia las siguientes recomendaciones:

- Un cambio progresivo a un modelo de empresa respetuoso con los derechos humanos que no se base en un seguimiento de datos invasivo. Como primera medida, TikTok debe garantizar que sus políticas y procesos de diligencia debida en materia de derechos humanos abordan los efectos generalizados y sistémicos que su modelo empresarial tiene en los derechos humanos, concretamente el derecho a la privacidad, el derecho a la libertad de pensamiento y el derecho a la salud.
- TikTok debe dejar de aumentar al máximo la “interacción de las personas usuarias” en detrimento de la salud y otros derechos humanos de quienes las utilizan. En el marco de su proceso de diligencia debida en materia de derechos humanos, TikTok, en colaboración con las personas usuarias —incluidos niños, niñas y jóvenes— y especialistas independientes, debe identificar los elementos de diseño que estimulan el uso adictivo de las plataformas y la comparación social, y sustituirlos por una utilización que se centre en la “la seguridad desde el diseño” y el interés superior de la infancia.
- Para respetar la privacidad y proporcionar a las personas usuarias la posibilidad real de elección y control, la existencia de un ecosistema de redes sociales exento del uso de perfiles no debería simplemente ser una opción sino lo normal. Por consiguiente, los algoritmos determinantes de contenidos que usan TikTok y otras plataformas online no deberían basarse por defecto en la elaboración de perfiles (por ejemplo, a partir del tiempo de visión o la interacción); y deben solicitar la inclusión voluntaria en lugar de la exclusión voluntaria, de manera tal que el consentimiento de inclusión voluntaria se otorgue libremente y sea específico, informado (lo que incluye utilizar un lenguaje adaptado a la infancia) e inequívoco.
- TikTok debe dejar de recopilar datos personales privados y hacer inferencias sobre los intereses, el estado emocional o el bienestar de las personas usuarias a partir del tiempo de visión y su interacción con la finalidad de “personalizar” las recomendaciones de contenido y segmentar la publicidad. En lugar de utilizar la vigilancia generalizada para adaptar los feeds a los intereses de las personas usuarias, TikTok debería permitir a éstas expresar sus intereses mediante avisos explícitos (por ejemplo, se les podría pedir que indicasen intereses concretos si quisieran recibir recomendaciones personalizadas) y únicamente si ello fuera acompañado de su consentimiento informado, específico y libremente otorgado.

Para proteger y hacer efectivos los derechos de los niños, niñas y jóvenes, los Estados deben:

- Impedir que las empresas condicionen el acceso a sus servicios a que las personas usuarias “den su consentimiento” a la recopilación, tratamiento y difusión de sus datos personales para la segmentación de contenidos o con fines comerciales o publicitarios.

- Regular las empresas de redes sociales, de conformidad con las leyes y normas internacionales de derechos humanos, para que garanticen que los algoritmos que usan las plataformas online a fin de determinar los contenidos no se basan por defecto en la elaboración de perfiles; y que esos algoritmos solicitan la inclusión voluntaria en lugar de la exclusión voluntaria, de manera tal que el consentimiento de inclusión voluntaria se otorga libremente y es específico, informado e inequívoco. La recopilación y el uso de los datos personales sensibles inferidos (por ejemplo, las recomendaciones basadas en el tiempo de visión y los “me gusta” que permiten inferir información sensible) para personalizar los anuncios y los contenidos deberían estar prohibidos.

2. METODOLOGÍA

2.1 PROPÓSITO Y ALCANCE DEL INFORME

En este informe se analizan los riesgos para los derechos humanos inherentes al uso de las recomendaciones de contenido dirigido que ofrecen las principales plataformas de redes sociales y su disputa por la atención de las personas usuarias jóvenes, prestando especial atención al diseño adictivo de la plataforma TikTok y la amplificación que éste hace de los contenidos relacionados con la depresión y las autolesiones.

Éste es uno de los dos informes de Amnistía Internacional que estudian los efectos sobre los derechos humanos del modelo empresarial basado en la vigilancia utilizado por TikTok y otras plataformas de redes sociales —un modelo que supone el seguimiento invasivo de las personas usuarias que Amnistía Internacional ya había descrito anteriormente como “pacto faustiano, por el cual [las personas] sólo pueden disfrutar del pleno ejercicio de sus derechos humanos en línea sometiéndose a un sistema fundado sobre la base del abuso de los derechos humanos”³ — tiene sobre los niños y niñas (personas menores de 18 años) y jóvenes (personas de entre 15 y 24 años, también denominadas “personas usuarias jóvenes” en este informe).

El documento “*Siento que no se me protege*”: *Emboscada en la red de vigilancia de TikTok*, que complementa el presente informe, pone de relieve los abusos que las empresas tecnológicas cometen contra el derecho a la privacidad de los niños y niñas, y que son inherentes al modelo empresarial de las “grandes plataformas tecnológicas”.⁴ Este estudio se centra en los efectos negativos en el mundo real que tienen la vigilancia de las personas usuarias, las decisiones de diseño y las recomendaciones de contenido personalizado que las empresas de las redes sociales emplean en aras de la “interacción de las personas usuarias”. En concreto, en él se analizan los riesgos que TikTok supone para los niños, niñas y jóvenes, sobre todo para quienes sufren problemas de salud mental tales como depresión, ansiedad y autolesiones.

El informe se inspira en otros trabajos de Amnistía Internacional, como la publicación de 2022 titulada *La atrocidad social. Meta y el derecho a un recurso para la comunidad rohinyá*, en la que se documentó el papel desempeñado por Facebook en la amplificación de publicaciones que incitaban a la violencia contra la población rohinyá, lo que en última instancia contribuyó a la limpieza étnica que las fuerzas de seguridad de Myanmar llevaron a cabo contra esa comunidad en 2017.⁵

3. Amnistía Internacional, *Gigantes de la vigilancia: La amenaza que el modelo de negocios de Google y Facebook representa para los derechos humanos* (Índice: POL 30/1404/2019), 21 de noviembre de 2019, <https://www.amnesty.org/en/documents/pol30/1404/2019/en>, p. 5.

4. Amnistía Internacional, “*Siento que no se me protege*”: *Emboscada en la red de vigilancia de TikTok* (Índice: POL 40/7349/2023), 7 de noviembre de 2023.

5. Amnistía Internacional, *La atrocidad social: Meta y el derecho a un recurso para la comunidad rohinyá* (Índice: ASA 16/5933/2022), 29 de septiembre de 2022, <https://www.amnesty.org/es/documents/asa16/5933/2022/es/> (resumen ejecutivo).

2.2 EJE DE INVESTIGACIÓN

Todas las principales empresas de redes sociales se disputan las cifras cada vez más altas de personas usuarias y la cuota más importante de su tiempo en pantalla. El presente informe se centra en TikTok porque, en los últimos tres años —marcados por la pandemia de COVID-19—, esta empresa se ha convertido en una plataforma global y ha atraído a centenares de millones de niños, niñas y jóvenes, muchos de los cuales la consideran ya su plataforma preferida según los estudios de mercado.⁶ El rápido crecimiento de TikTok posibilitó que el número de personas activas en esta plataforma prácticamente se duplicara entre diciembre de 2019 y septiembre de 2021,⁷ cuando la empresa anunció que había superado los 1.000 millones de personas usuarias mensuales.⁸ Se considera que la mayoría de las personas que utilizan TikTok son niños, niñas y jóvenes.⁹ En agosto de 2020 se publicó que TikTok había incluido a más de un tercio de sus 49 millones de personas usuarias diarias en Estados Unidos en la categoría de 14 años o menos, y que se estimaba que muchas de ellas tenían menos de 13 años, la edad mínima con la que la empresa permite acceder a la versión para adultos de la plataforma en Estados Unidos de América (Estados Unidos) y la mayoría de los demás países en los que opera.¹⁰

El rápido ascenso de TikTok a las posiciones dominantes del mercado, junto con los elementos concretos de diseño que utiliza para aumentar al máximo la cantidad de tiempo que las personas usuarias pasan en su plataforma, está generando una preocupación creciente sobre el hecho de que esta empresa expone a niños, niñas y jóvenes a una plataforma adictiva y potencialmente peligrosa.¹¹

Aunque se utiliza prácticamente en todo el mundo, la bibliografía disponible sobre los efectos de TikTok (y otras plataformas de redes sociales) sobre los niños, niñas y jóvenes, incluidos daños potenciales, se centra mayoritariamente en Australia, Europa y Estados Unidos.¹² Casi el 90% de la población joven del mundo vive en países en desarrollo,¹³ pero sus experiencias están claramente infrarrepresentadas en los estudios y los debates públicos sobre las redes sociales. Con el fin de ampliar la base de pruebas geográfica, la investigación realizada para este informe se centró en el uso

6. Qustodio, *Social Media Annual Data Report 2021: Living and Learning in a Digital World*, capítulo 2 (redes sociales), <https://qustodio.com/en/social-media-qustodio-annual-data-report-2021/>
7. We Are Social y Hootsuite, *Digital 2022 Global Overview Report*, 26 de enero de 2022, <https://wearesocial.com/hk/blog/2022/01/digital-2022-another-year-of-bumper-growth>
8. Reuters, "TikTok hits 1 billion monthly active users globally – company", 27 de septiembre de 2021, <https://www.reuters.com/technology/tiktok-hits-1-billion-monthly-active-users-globally-company-2021-09-27>
9. Meltwater, "54 TikTok stats you need to know [2023]", 30 de diciembre de 2022, <https://www.meltwater.com/en/blog/tiktok-statistics>; Guardian, "What TikTok does to your mental health: 'It's embarrassing we know so little'", 30 de octubre de 2022, <https://www.theguardian.com/technology/2022/oct/30/tiktok-mental-health-social-media>; e Instituto Reuters, *Digital News Report 2023*, junio de 2023, <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/dnr-executive-summary>
10. New York Times, "A third of TikTok's U. S. users may be 14 or under, raising safety questions", 14 de agosto de 2020, <https://www.nytimes.com/2020/08/14/technology/tiktok-underage-users-ftc.html>; TikTok ofrece una experiencia limitada de la aplicación con salvaguardias y protecciones de privacidad adicionales para personas menores de 13 años, véase TikTok, "TikTok for younger users", 13 de diciembre de 2019, <https://www.newsroom.tiktok.com/en-us/tiktok-for-younger-users>. El artículo del New York Times menciona que una persona que había trabajado en TikTok denunció que la plataforma permitía que los contenidos de video de niños y niñas menores de 13 años permanecieran online durante semanas. La versión para menores de 13 años no permite a este grupo de población publicar videos, lo que significa que las personas menores de 13 años utilizaban la versión para adultos de la aplicación, probablemente tras haber facilitado información falsa sobre su edad. En 2019, TikTok aceptó pagar una multa de 5,7 millones de dólares estadounidenses para que se retiraran las acusaciones de la Comisión Federal de Comercio de Estados Unidos de que la empresa había recopilado de forma ilegal información de niños y niñas menores de 13 años. Wire, "FTC Hits TikTok With Record \$5.7 Million Fine Over Children's Privacy", 27 de febrero de 2019, <https://www.wired.com/story/tiktok-ftc-record-fine-childrens-privacy/>; para más información sobre la política de edad de TikTok, véase TikTok, "Guía para padres y tutores", <https://www.tiktok.com/safety/es-es/guardians-guide/> (visitada el 26 de septiembre de 2023).
11. Véanse, por ejemplo: Wall Street Journal, "The corpse bride diet: How TikTok inundates teens with eating-disorder videos", 17 de diciembre de 2021, <https://www.wsj.com/articles/how-tiktok-inundates-teens-with-eating-disorder-videos-11639754848>; Sophia Petrillo, "What makes TikTok so addictive?: An analysis of the mechanisms underlying the world's latest social media craze", 13 de diciembre de 2021, Brown Undergraduate Journal of Public Health, número 2021-2022, <https://www.sites.brown.edu/publichealthjournal/2021/12/13/tiktok/>; BBC, "Inside TikTok's real-life frenzies - from riots to false murder accusations", 21 de septiembre de 2023, <https://www.bbc.co.uk/news/technology-66719572>; y Guardian, "What TikTok does to your mental health: It's embarrassing we know so little", 30 de octubre de 2022, <https://www.theguardian.com/technology/2022/oct/30/tiktok-mental-health-social-media>
12. Sakshi Ghai, Lucia Magis-Weinberg y otros, "Social media and adolescent well-being in the Global South", agosto de 2022, *Current Opinion in Psychology*, volumen 46, <https://pubmed.ncbi.nlm.nih.gov/35439684/>
13. UNESCO, "Thematic factsheet: youth and empowerment", 31 de enero de 2023, <https://www.unesco.org/en/youth-and-empowerment>

de TikTok que hacen los niños, niñas y jóvenes de Kenia y Filipinas, países seleccionados por el tiempo excepcionalmente elevado que las personas de entre 16 y 64 años dedican diariamente a las redes sociales (algo más de tres horas al día en Kenia y algo más de cuatro horas en Filipinas, cifras que figuran entre las más elevadas del planeta frente al promedio de dos horas y media a nivel mundial).¹⁴ Estados Unidos se incluyó en el componente de indagación técnica (véase la metodología de investigación *infra*) para analizar las posibles diferencias en las decisiones de diseño de la plataforma tomadas por TikTok en un país en el que hay un debate público general sobre los riesgos que las redes sociales suponen para las personas adolescentes que las usan, y que también puede considerarse un mercado clave de la empresa.

Aunque la investigación se centró especialmente en los efectos que la plataforma tiene sobre la salud mental, en el alcance del proyecto no se contempló demostrar que TikTok causaba o agravaba los problemas de salud mental de quienes participaran en el estudio. Para hacerlo habría sido necesario tener un conocimiento profundo de cada circunstancia que afectaba a la salud y bienestar de esas personas y una opinión profesional experta en materia de salud mental, lo que no sería posible en el caso de la mayoría de los y las jóvenes participantes en esta investigación debido a las grandes limitaciones de acceso a servicios de salud mental gratuitos o asequibles (más información en el apartado 3.3). No obstante, aunando los métodos cualitativos y la indagación técnica sobre el sistema de recomendaciones de TikTok, la investigación realizada por Amnistía Internacional para este informe ofrece una imagen lo suficientemente clara como para justificar que se actúe con urgencia para mitigar ese riesgo.

2.3 METODOLOGÍA DE INVESTIGACIÓN

La investigación realizada para este informe tuvo lugar entre octubre de 2022 y julio de 2023, y constaba de los siguientes elementos:

1. Investigación documental detallada, incluida una revisión de las fuentes secundarias pertinentes académicas y de la ONU, las ONG y los medios de comunicación, entre otros actores.
2. Un estudio exploratorio en forma de cuestionario online, distribuido a través de los canales de redes sociales de Amnistía y las organizaciones asociadas, y al que 550 niños, niñas y jóvenes de 13 a 24 años de 45 países respondieron entre octubre y noviembre de 2022. En el cuestionario se preguntaba sobre el uso de las principales plataformas de redes sociales, las experiencias en estas plataformas, lo que gustaba y disgustaba, las reacciones ante las experiencias negativas y las visiones de cambio, con miras a comprender mejor las vivencias, las preocupaciones y las actitudes hacia las redes sociales.¹⁵
3. Un total de 5 entrevistas presenciales semiestructuradas y 20 debates con grupos-muestra en los que participaron más de 300 niños, niñas y jóvenes de entre 14 y 24 años (de los cuales aproximadamente 180 utilizaban TikTok) en Kenia (en marzo de 2023) y Filipinas (en abril y mayo de 2023). En Kenia, las entrevistas presenciales y los debates con escolares, activistas juveniles, estudiantes universitarios y jóvenes que habían terminado sus estudios recientemente

14. We are social, *Digital 2022 Global Overview Report*, 26 de enero de 2022, <https://www.wearesocial.com/hk/blog/2022/01/digital-2022-another-year-of-bumper-growth/>; El equipo de investigación de Amnistía no pudo encontrar estadísticas mundiales similares relativas específicamente a niños, niñas y jóvenes.

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

tuvieron lugar en la capital (Nairobi), en las ciudades de Kisumu y Mombasa y en el condado de Machakos. En Filipinas, las entrevistas presenciales se realizaron en la capital (Manila) y la provincia de Batangas. Se llevaron a cabo 9 entrevistas adicionales online con niños, niñas y jóvenes de regiones de Filipinas a las que el equipo investigador no podía acceder por razones prácticas o debido a problemas de seguridad. El equipo de investigación de Amnistía Internacional habló con 37 niños, niñas y jóvenes, de los cuales 34 contaron sus experiencias en TikTok además de en otras plataformas de redes sociales.

En Kenia, Amnistía Internacional se apoyó en la red de “colegios amigos de los derechos humanos” y las delegaciones de AI Kenia para dirigirse a personas jóvenes a las que entrevistar. En Filipinas, la invitación a la participación se difundió a través de la red de jóvenes de AI Filipinas y se envió a los servicios de apoyo a la salud mental y a organizaciones de activismo sobre salud mental dirigido por jóvenes.

Esta vía de referencias para llegar a los niños, niñas y jóvenes, el énfasis en las entrevistas o grupos-muestra muy reducidos, y el mayor número de personas adultas jóvenes con cierto acceso a servicios de apoyo a la salud mental en la cohorte de Filipinas contribuyeron seguramente a que se presentara una tasa mucho más alta de participantes que declararon tener o haber tenido problemas de salud mental en comparación con el grupo mayoritariamente de adolescentes de Kenia. Los aprendizajes adquiridos en los debates de Kenia que se habían realizado en marzo de 2023 llevaron a un planteamiento más enfocado consistente en invitar —con la ayuda de las organizaciones de salud mental y la red de activistas juveniles de Amnistía Internacional— a participantes en la investigación que consideraran que su salud mental se había visto afectada por el uso que hacían de las redes sociales.

El equipo de investigación de Amnistía Internacional realizó todas las entrevistas y debates en inglés, con apoyo puntual en la traducción cuando las personas entrevistadas expresaban las ideas y experiencias en kiswahili y luo (Kenia) o tagalo (Filipinas). Se plantearon preguntas acerca del uso que se hacía de las distintas plataformas de redes sociales, las experiencias en estas plataformas, las opiniones sobre los esfuerzos reguladores con respecto al uso de las redes sociales por parte de niños y niñas, el impacto del uso de las redes sociales en la vida, y los efectos percibidos que las experiencias en las redes sociales tenían en el estado emocional y la salud mental.

Se obtuvo el consentimiento de los progenitores o tutores de todas las personas participantes menores de 18 años, y el consentimiento informado de cada niño o niña. Quienes tenían más de 18 años dieron su consentimiento informado antes de tomar parte en las entrevistas o los debates. Con el fin de proteger la identidad de las personas entrevistadas, a lo largo de este informe se utilizan siempre seudónimos, excepto en el caso de especialistas que accedieron a que apareciera su nombre.

4. Entre febrero y mayo de 2023 se realizaron 14 entrevistas cualificadas a especialistas en psicología adolescente, activistas (juveniles) en materia de salud mental, especialistas en salud pública y docentes de Kenia, Filipinas y Estados Unidos.
5. El énfasis en los niños, niñas y jóvenes estableció forzosamente los límites éticos sobre la manera de analizar las formas en que los sistemas algorítmicos interactúan con las personas usuarias. Para evitar exponer a jóvenes a contenidos nocivos o poner en peligro su privacidad, y al mismo tiempo obtener resultados lo más representativos posible, en el estudio también se incluyó una indagación técnica en dos partes que se realizó en junio y julio de 2023:
 - Una auditoría algorítmica automatizada del sistema de recomendaciones de la página “Para ti” de TikTok por medio de cuentas de investigación que simulaban la interacción de personas de 13 años con la plataforma en Kenia y Estados Unidos,

- y grabaciones de pantallas de tres cuentas de investigación gestionadas manualmente, creadas para representar a las personas usuarias de 13 años de Kenia, Filipinas y Estados Unidos, y que monitoreaban la amplificación de contenidos relacionados con la salud mental en la página “Para ti” de TikTok.

La metodología de la indagación técnica llevada a cabo se analiza con más detalle en el capítulo 6.

6. Carta de presentación de la investigación para solicitar información detallada sobre los procesos de diligencia debida en materia de derechos humanos utilizados en TikTok, las políticas de protección de datos y las medidas para fomentar el bienestar y la salud mental de niños, niñas y jóvenes, además de una carta con derecho a respuesta resumiendo las principales conclusiones de este informe. Las respuestas de TikTok de 12 de julio y de 29 de octubre de 2023 se pueden consultar en los anexos 3 y 4, respectivamente.

3. INFORMACIÓN GENERAL

3.1 NO ENTENDER NADA: LAGUNAS DE INVESTIGACIÓN Y OBSTÁCULOS DE LAS EMPRESAS AL ESTUDIO DE LOS RIESGOS EN MATERIA DE DERECHOS HUMANOS QUE LAS PLATAFORMAS DE REDES SOCIALES SUPONEN PARA LOS NIÑOS, NIÑAS Y JÓVENES

En octubre de 2022, la investigación del pesquisidor de Reino Unido llevada a cabo sobre la muerte de la adolescente británica Molly Russell determinó oficialmente por primera vez que las redes sociales probablemente habían contribuido a la muerte de una persona.¹⁶ Una parte considerable del feed de Instagram de Molly Russell de los meses anteriores a que la niña muriera con 14 años consistía en contenidos que describían y fomentaban las autolesiones, la depresión y el suicidio.¹⁷ El pesquisidor determinó que la muerte de Molly había sido consecuencia de “un acto de autolesión al sufrir depresión y los efectos negativos de los contenidos online”, algunos de los cuales “tendían a presentar la autolesión y el suicidio como la consecuencia inevitable de una enfermedad sin posibilidad de recuperación” con sistemas de recomendación que ofrecían a Molly una “visión irracional y limitada sin ningún contrapeso de la realidad”.¹⁸ El pesquisidor recomendó que el gobierno de Reino Unido y las empresas de redes sociales “revisaran el uso de algoritmos para ofrecer contenidos”.¹⁹ Es una de las actas de acusación formal más contundentes realizadas hasta la fecha por una autoridad pública en el debate en torno a los efectos de las redes sociales en la salud mental de las personas adolescentes.

LOS FEEDS DE REDES SOCIALES Y EL FEED “PARA TI” DE TIKTOK

A diferencia de los medios de comunicación tradicionales, las plataformas de redes sociales crean un canal de contenido personalizado que genera la persona usuaria (“feed”). Inicialmente, las

16. Servicio del pesquisidor de Londres Norte, “Regulation 28 report to prevent future deaths”, 13 de octubre de 2022, https://www.judiciary.uk/wp-content/uploads/2022/10/Molly-Russell-Prevention-of-future-deaths-report-2022-0315_Published.pdf. [Traducción de Amnistía Internacional]
17. Politico, “Digital bridge: Platforms on the hook — Transatlantic AI rulebook — Let’s talk data transfers”, 6 de octubre de 2022, <https://www.politico.eu/newsletter/digital-bridge/platforms-on-the-hook-transatlantic-ai-rulebook-lets-talk-data-transfers/>; una portavoz de Meta que declaró en la investigación afirmó que los contenidos que Molly veía presentaban matices y eran complicados, y que era importante que las personas usuarias que tuvieran pensamientos suicidas pudieran expresarse online. BBC, “Molly Russell: Instagram posts seen by teen were safe, Meta says”. 26 de septiembre de 2022, <https://www.bbc.co.uk/news/uk-england-london-63034300>; para información detallada sobre la respuesta de Meta a la investigación, véase Meta, “Response from Meta”, 6 de diciembre de 2022, <https://www.judiciary.uk/wp-content/uploads/2022/10/2022-0315-Response-from-META.pdf>.
18. Servicio del pesquisidor de Londres Norte, “Regulation 28 report to prevent future deaths”, 13 de octubre de 2022 (antes citado).
19. BBC News, “Molly Russell: Coroner’s report urges social media changes”, 14 de octubre de 2022, <https://www.bbc.co.uk/news/uk-england-london-63254635>

plataformas de redes sociales se llenaron en gran medida de contenido creado por las “amistades” con las que un usuario o usuaria se conectaba y de las páginas que seguía. Sin embargo, plataformas de redes sociales como TikTok recomiendan cada vez más contenido creado por personas que no pertenecen a la red del usuario o usuaria, según los intereses que éste o ésta ha manifestado o que se han inferido. El feed “Para ti” de TikTok, que es el canal de contenido predeterminado de la plataforma, a menudo es presentado por los medios de comunicación y expertos/as de marketing y tecnología como el resultado más elaborado de este tipo de cambio hacia un feed hiperpersonalizado, en el que las personas usuarias sólo deben ver contenido recomendado durante cierto tiempo para que el sistema de algoritmos evalúe los intereses de éstas, los relacione con los contenidos de los vídeos disponibles y les ofrezca recomendaciones y anuncios más personalizados.*

* Pese a que la Ley de Servicios Digitales de la UE prohíbe la publicidad personalizada dirigida a menores de edad. Para un análisis detallado de las prácticas de recopilación de datos y publicidad de TikTok, véase Amnistía Internacional, “*Siento que no se me protege*”: Emboscada en la red de vigilancia de TikTok, 7 de noviembre de 2023 (antes citado).

Sin embargo, las empresas de redes sociales dificultan claramente la labor de auditar sus sistemas algorítmicos a quienes realizan investigaciones independientes.²⁰ Amnistía Internacional y otras organizaciones de la sociedad civil han pedido que se adopten medidas normativas para impedir que las empresas obstaculicen la investigación independiente y de interés público sobre el funcionamiento y los efectos de las plataformas de redes sociales, y han advertido:

“Muchas plataformas [de redes sociales] obstaculizan de forma activa la investigación mediante las condiciones de servicio, a través de medidas técnicas o mediante actos de intimidación y amenazas de emprender acciones judiciales. En concreto, los esfuerzos de las plataformas por impedir el *scraping* [la recopilación automatizada de datos a gran escala] han tenido un efecto inhibitorio en las personas investigadoras que tratan de hacerles rendir cuentas”.²¹

Sin embargo, al ir aumentando cada año el número de personas usuarias de redes sociales y la cantidad de tiempo que pasan los niños y niñas en ellas, la cobertura informativa y el debate público sobre el papel que estas plataformas desempeñan en la vida infantil son cada vez más críticos hacia sus efectos nocivos en jóvenes. Esto se aprecia claramente en Europa y Estados Unidos, donde el debate público se avivó por la publicación en 2021 de unos documentos internos filtrados, conocidos como los “Facebook Papers” (Papeles de Facebook), en los que figuraban las conclusiones de unas encuestas distribuidas en el interior de la empresa en 2019 y según las cuales Instagram dañaba la autoestima corporal de una de cada tres chicas adolescentes.²²

-
20. Observatorio de Algoritmos, “AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook”, 13 de agosto de 2021, <https://www.algorithmwatch.org/en/instagram-research-shut-down-by-facebook>; y Observatorio de Algoritmos, “A guide to the EU’s new rules for researcher access to platform data”, 7 de diciembre de 2022, <https://algorithmwatch.org/en/dsa-data-access-explained/>
21. La extracción de datos puede plantear riesgos considerables para la privacidad en función del contexto, el tipo de datos recopilados y el uso que se haga de ellos y, por tanto, debería estar limitada. No obstante, Amnistía Internacional y otros actores han criticado la prohibición general a la extracción de datos impuesta por las empresas de redes sociales por considerarla una traba indebida a la investigación sobre los riesgos sistémicos. Ante la falta de marcos de acceso a datos de investigación, como el que va a implementarse pronto en Europa gracias a las nuevas obligaciones impuestas a las plataformas online de gran tamaño en virtud de la Ley de Servicios Digitales de la Unión Europea (UE), las personas investigadoras en ocasiones se han visto obligadas a realizar su labor contraviniendo las condiciones de servicio de las plataformas. Fundación Mozilla, Amnistía Internacional y otros, “Response to the European Commission’s call for evidence for a Delegated Regulation on data access provided for in the Digital Services Act”, mayo de 2023, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3423646_en [Traducción de Amnistía Internacional] * Pese a que la Ley de Servicios Digitales de la UE prohíbe la publicidad personalizada dirigida a menores de edad. Para un análisis detallado de las prácticas de recopilación de datos y publicidad de TikTok, véase Amnistía Internacional, *I feel exposed: “Siento que no se me protege”*: Emboscada en la red de vigilancia de TikTok, 7 de noviembre de 2023 (antes citado).
22. Guardian, “Facebook aware of Instagram’s harmful effect on teenage girls, leak reveals”, 14 de septiembre de 2021, <https://www.theguardian.com/technology/2021/sep/14/facebook-aware-instagram-harmful-effect-teenage-girls-leak-reveals>; Meta descartó las conclusiones que aparecían en los “Papeles de Facebook” con el argumento de que se basaban en una selección de documentos mal interpretados y descontextualizados, véase Washington Post, “A whistleblower’s power: Key takeaways from the Facebook Papers”, 26 de octubre de 2021, <https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/>

Antes de 2021 ya había mucho debate entre profesionales de la psicología y especialistas en salud pública sobre si el uso de las redes sociales había contribuido al aumento de la depresión y las autolesiones entre las personas jóvenes.²³ Una limitación importante de muchos de los estudios disponibles es que examinan el efecto de las redes sociales en general o, de forma aún más inespecífica, el efecto del tiempo en pantalla sobre la población adolescente de Europa y América en general. Hay una carencia notoria de datos similares de otras partes del mundo.²⁴ En realidad, las plataformas de redes sociales difieren de manera significativa y cambian con el tiempo, lo mismo que los niños, niñas y jóvenes. Al reflexionar sobre los riesgos y la regulación, la cuestión que debe plantearse a las plataformas de redes sociales no es si todas las plataformas perjudican a todos los niños o simplemente al niño o niña “promedio” —si es que esa variable indirecta siquiera existe—, sino qué aspectos concretos de las plataformas de redes sociales podrían suponer un riesgo para cualquiera de las personas o grupos concretos de personas que las utilizan.

Las empresas de redes sociales publicitan las ventajas que sus plataformas aportan a la vida de los niños, niñas y jóvenes, y muchas de las personas jóvenes —y ciertamente muchas de las que participaron en la investigación realizada para este informe— están de acuerdo. Los jóvenes participantes en la investigación elogiaron a TikTok por su valor como entretenimiento, la manera en que estimula la creatividad y porque genera un espíritu de comunidad. Numerosos estudios han señalado los beneficios que aportan las comunidades online —reunidas por las redes sociales—, especialmente a las personas jóvenes marginadas que buscan a iguales que cuenten sus experiencias, como personas jóvenes LGBTI o con problemas de salud mental.²⁵ No obstante, estas ventajas no eximen a las empresas de redes sociales de su responsabilidad de identificar, evaluar, prevenir y mitigar posibles daños, y de reparar —cuando sea necesario y procedente— los daños reales derivados del uso de sus plataformas.

Sin embargo, aunque promocionan las ventajas de sus productos, las empresas de redes sociales no proporcionan a las personas usuarias evaluaciones transparentes de riesgos. Además, han dificultado enormemente las investigaciones independientes sobre el funcionamiento de sus plataformas y los posibles riesgos que suponen para quienes las utilizan.²⁶ En ausencia de una necesaria regulación que permita realizar estos estudios de forma eficaz y respetuosa con la privacidad, los proyectos de investigación han sido acallados y quienes llevan a cabo esta actividad han recibido amenazas de medidas judiciales.²⁷ Además, como plataforma de vídeos acelerados, TikTok plantea retos de investigación adicionales con respecto a otras porque el análisis a gran escala de fragmentos de vídeo es más difícil y requiere más recursos que el de publicaciones escritas.

-
23. Para una amplia revisión en curso sobre el tema, véase Jonathan Haidt y otros, “Social Media and Mental Health: A Collaborative Review”, Universidad de Nueva York, <https://docs.google.com/document/d/1w-HOfseF2wF9YlpXwUUtP65-olnkPyWcgF5BiAtBEy0/edit> (consultado el 12 de septiembre de 2023).
 24. Sakshi Ghai y otros, “Social media and adolescent well-being in the Global South” (antes citado); Sakshi Ghai, Luisa Fassi, Faisal Awadh y Amy Orben, “Lack of sample diversity in research on adolescent depression and social media use: A scoping review and meta-analysis”, 7 de febrero de 2023, *Clinical Psychological Science*, <https://www.doi.org/10.1177/21677026221114859>
 25. John Naslund, Armea Bondre y otros, “Social media and mental health: Benefits, risks, and opportunities for research and practice”, 20 de abril de 2020, *Journal of Technology in Behavioral Science*, volumen 5, <https://www.doi.org/10.1007/s41347-020-00134-x>; Matthew N. Berger, Melody Taba y otros, “Social media use and health and well-being of lesbian, gay, bisexual, transgender, and queer youth: Systematic review”, 21 de septiembre de 2022, *Journal of Medical Internet Research*, volumen 24, número 9, <https://www.jmir.org/2022/9/e38449>
 26. Para más información sobre las recomendaciones de Amnistía Internacional y otras organizaciones de la sociedad civil sobre la identificación y mitigación de los riesgos sistémicos, véase People vs Big Tech, “Briefing: Fixing Recommender Systems: From identification of risk factors to meaningful transparency and mitigation”, 25 de agosto de 2023, <https://www.peoplevsbig.tech/briefing-fixing-recommender-systems-from-identification-of-risk-factors-to-meaningful> (consultado el 25 de agosto de 2023).
 27. Observatorio de Algoritmos, “AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook”, 13 de agosto de 2021, <https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/>; y Observatorio de Algoritmos, “A guide to the EU’s new rules for researcher access to platform data”, diciembre de 2022, <https://www.algorithmwatch.org/en/dsa-data-access-explained/>

3.2 ENTRE LA POLÍTICA Y EL POPULISMO: ESCASOS ESFUERZOS PARA REGULAR LAS PLATAFORMAS DE REDES SOCIALES

Quizá no sorprenda, entonces, que fuera la propia investigación *interna* de una empresa de redes sociales la que favoreció un discurso público y unos debates políticos más críticos sobre los efectos de las redes sociales en los niños, niñas y jóvenes. Como se ha señalado *supra*, en los “Papeles de Facebook” de 2021 aparecía información incriminatoria del papel que Instagram desempeñaba en el empeoramiento de los problemas de autoestima corporal entre las adolescentes.²⁸ Ésta y otras conclusiones que se desprenden de los documentos filtrados fueron noticia en todo el mundo e intensificaron la presión sobre quienes elaboran las políticas para abordar los persistentes problemas de la modernización de los marcos normativos sobre las plataformas, que para entonces llevaban obsoletos al menos dos decenios.²⁹

Sin embargo, los avances encaminados a abordar los riesgos sistémicos asociados a las grandes plataformas de redes sociales son escasos, y pocos Estados u organizaciones regionales han adoptado legislación al respecto. Ni los gobiernos de Kenia ni los de Filipinas, dos de los países de interés del presente informe, han aprobado aún leyes en este sentido.³⁰

Entre las escasas jurisdicciones que han adoptado medidas figura la Unión Europea (UE), cuya Ley de Servicios Digitales (LSD) —adoptada en julio de 2022— se convirtió en la primera normativa importante de ámbito regional sobre las “grandes tecnológicas” encaminada a limitar los efectos nocivos de las plataformas de redes sociales, por ejemplo prohibiendo la intrusiva “publicidad personalizada” dirigida a menores de edad.³¹ La LSD impone a las grandes plataformas online, como TikTok, el deber de evaluar sus riesgos sistémicos, incluidos los que afectan a la salud pública y las personas menores de edad; de adoptar medidas para mitigar estos riesgos, y de someterse a auditorías independientes para valorar su grado de cumplimiento de estas obligaciones (si bien con muchas cuestiones pendientes con respecto al carácter y la aplicación concretos de esas medidas).³² En virtud de la LSD, TikTok y otras plataformas de gran tamaño también están obligadas a ofrecer a quienes las usan al menos un feed que no se base en la elaboración de perfiles de personas usuarias. Amnistía ha calificado este enfoque de inclusión voluntaria (*opt-in*) de “oportunidad perdida” porque cambiar la configuración a opciones más respetuosas con la privacidad suele ser engorroso y las personas usuarias tienden a mantener la configuración predeterminada.³³

28. Guardian, “Facebook aware of Instagram’s harmful effect on teenage girls, leak reveals”, septiembre de 2021 (antes citado).

29. La principal disposición legislativa estadounidense que rige la responsabilidad civil de las plataformas online, la Ley de Decencia en las Comunicaciones, se promulgó en 1996. La Ley de Servicios Digitales (LSD) de 2022 de la UE estuvo precedida de la Directiva sobre el comercio electrónico, que se había adoptado en 2000.

30. No obstante, ambos Estados regulan la recopilación y el uso de los datos personales, por parte también de los servicios online. El informe “*Siento que no se me protege*”: *Emboscada en la red de vigilancia de TikTok*, que complementa este estudio, incluye un análisis de la legislación pertinente de Kenia. Filipinas aprobó la regulación de la privacidad de datos mediante la Ley de Privacidad de Datos, de 2012, que incluye principios también contemplados en el Reglamento general de protección de datos de la UE tales como la limitación de la finalidad y la reducción al mínimo de la recopilación de datos. Véase Asociación Internacional de Profesionales de la Privacidad, “GDPR matchup: The Philippines’ Data Privacy Act and its Implementing Rules and Regulations”, julio de 2017, <https://iapp.org/news/a/gdpr-matchup-the-philippines-data-privacy-act-and-its-implementing-rules-and-regulations/>.
UE, Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales), <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32022R2065>

32. Para un análisis detallada de la LSD, véase Amnistía Internacional, *What the EU’s Digital Services Act Means for Human Rights and Harmful Big Tech Business Models*, (Índice: POL 30/5830/2022), 7 de julio de 2022, <https://www.amnesty.org/es/documents/pol30/5830/2022/en>; y Euractiv, “Europe enters patchy road to audit online platforms’ algorithms”, 7 de julio de 2023, <https://www.euractiv.com/section/platforms/news/europe-enters-patchy-road-to-audit-online-platforms-algorithms/>

33. Alfred Ng, “Default settings for privacy – we need to talk”, 21 de diciembre de 2019, <https://www.cnet.com/tech/tech-industry/default-settings-for-privacy-we-need-to-talk/>

La eficacia de la LSD aún no se ha probado: algunas de las obligaciones que afectan a las plataformas online de gran tamaño, como la presentación de evaluaciones de riesgos empezaron a aplicarse en agosto de 2023, pero la ley no entra plenamente en vigor hasta 2024. Aunque pretende reducir determinados riesgos asociados al modelo empresarial basado en la vigilancia que utilizan las empresas de redes sociales, por ejemplo prohibiendo la publicidad personalizada dirigida a menores de edad, la LSD no prohíbe el modelo empresarial en su totalidad y, por tanto, es obviamente limitada. No obstante, esta ley representa la regulación de plataformas más ambiciosa hasta la fecha. Además, pese a que presenta limitaciones en cuanto a su alcance y su aplicación a la actividad de las plataformas en la UE, la LSD aumenta las esperanzas de que se produzca un efecto dominó más allá de la UE porque los cambios implementados en la UE podrían extenderse con mayor facilidad a las propuestas legislativas de otros lugares, o copiarse en ellas.

En Reino Unido, el proyecto de ley sobre la seguridad online, aprobado en septiembre de 2023, impone nuevas obligaciones a las plataformas con el propósito de crear un entorno digital más seguro y aplica protecciones especiales a las personas menores de 18 años.³⁴ Este texto legislativo se elaboró para aumentar las protecciones en materia de privacidad y los requisitos de reducción al mínimo de la recopilación de datos contemplados en el Código de la Infancia de Reino Unido, que impuso por primera vez a los servicios online la responsabilidad de únicamente aplicar la “perfilación” de menores si “tienen en marcha las medidas adecuadas para proteger al niño o niña contra cualquier efecto nocivo (en concreto, recibir recomendaciones de contenido que sea perjudicial para su salud o bienestar)”.³⁵ El proyecto de ley sobre seguridad online tipifica como delito la publicación de materiales que fomenten las autolesiones graves.³⁶ El requerimiento que el proyecto de ley impone a las plataformas de que impidan a las personas menores de edad acceder a materiales “que no llegan a ser delito, pero promueven o proporcionan instrucciones para el suicidio, las autolesiones o los trastornos de la conducta alimentaria”³⁷ ha generado mucha polémica. La línea de atención telefónica Samaritanos y varias organizaciones de salud mental han argumentado que limitar las normas a las personas menores de edad es una medida muy estrecha de miras, y otros grupos se han pronunciado contra el hecho de incentivar a las plataformas a restringir indebidamente la libertad de expresión de quienes las usan en el empeño de cumplir con esas medidas para restringir los contenidos legales pero nocivos.³⁸

La nueva normativa de Estados Unidos permanece bloqueada a nivel federal,³⁹ pero varios estados del país han aprobado leyes que regulan los espacios digitales. Uno de ellos es California, cuya Ley de Código de Diseño Adecuado a la Edad exigirá a las plataformas “a las que es posible que accedan personas menores de edad” que ofrezcan a éstas un mayor nivel de privacidad mediante la configuración predeterminada, y prohibirá el uso de la información de niños y niñas “de una manera

34. Ministerio de Asuntos Digitales, Cultura, Medios de Comunicación y Deporte, proyecto de ley sobre seguridad online (modificado por la Comisión), <https://bills.parliament.uk/bills/3137>

35. Autoridad de Protección de Datos de Reino Unido, código de diseño adecuado a la edad, <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/code-standards/> (consultado el 13 de julio de 2023). [Traducción de Amnistía Internacional]

36. Ministerio de Asuntos Digitales, Cultura, Medios de Comunicación y Deporte, proyecto de ley sobre seguridad online (modificado por la Comisión), (antes citado)

37. Ministerio de Asuntos Digitales, Cultura, Medios de Comunicación y Deporte, *A Guide to the Online Safety Bill*, 16 de diciembre de 2022, <https://www.gov.uk/guidance/a-guide-to-the-online-safety-bill> [Traducción de Amnistía Internacional]

38. Guardian, “Online safety bill must protect adults from self-harm content, say charities”, 14 de octubre de 2022, <https://www.theguardian.com/technology/2022/oct/14/online-safety-bill-must-protect-adults-self-harm-content-charities-samaritans>

39. De las múltiples propuestas de proyecto de ley a nivel federal relacionadas con la regulación de las plataformas, sólo la Ley de Seguridad Infantil en Internet (KOSA, por sus siglas en inglés) se aprobará probablemente a corto plazo. Esta ley combina aspectos del Código de Diseño de California y la LSD de la UE, y exige a las plataformas de gran tamaño a las que puedan acceder menores de 17 años que publiquen informes sobre las actividades de cumplimiento normativo, apliquen medidas en el diseño y funcionamiento de sus productos y servicios para prevenir y mitigar el daño, y ofrezcan a las personas usuarias jóvenes formas de exclusión voluntaria (*opt-out*) de los feeds personalizados basados en recomendaciones. Véase Congress.gov, Texto: S. 1409 - 118 Congreso (2023-2024): Ley de Seguridad Infantil en Internet, 2 de mayo de 2023, <https://www.congress.gov/bills/118th-congress/senate-bill/1409/text>

que la empresa sepa, o tenga motivos para saber, que es sustancialmente perjudicial para la salud física, la salud mental o el bienestar de una persona menor de edad”.⁴⁰ Sin embargo, esa Ley fue bloqueada por una jueza federal en septiembre de 2023 mientras sigue adelante una demanda presentada por un grupo de la industria tecnológica en la que se argumenta que esa ley viola la Primera Enmienda.⁴¹ Como en el caso de la LSD, las empresas tendrán que realizar una evaluación del impacto sobre la protección de los datos no públicos para determinar si los elementos de diseño o los algoritmos de las plataformas podrían ser nocivos para los niños y niñas. Aunque muchas personas defensoras de los derechos de la infancia han valorado favorablemente la Ley como un paso hacia un espacio online más seguro,⁴² quienes defienden la privacidad y la libertad de expresión han manifestado su preocupación en materia de privacidad por las posibles implicaciones negativas del uso más generalizado de medidas de verificación de la edad.⁴³

Por su parte, la legislación promulgada en Utah en marzo de 2023 otorga a los progenitores un acceso sin precedente a las cuentas de redes sociales de los niños y niñas y el control de esas cuentas, lo que suscita inquietudes sobre la privacidad de las personas menores de edad. Quienes llevan a cabo trabajo de campaña sobre los derechos de la infancia han cuestionado que pasar la responsabilidad a los progenitores proteja a los niños y niñas de los daños sistémicos.⁴⁴ No obstante, Arkansas y Luisiana siguieron su ejemplo y adoptaron leyes similares a principios de 2023.⁴⁵ Y, en mayo de 2023, Montana se convirtió en el primer estado del país en prohibir TikTok por distintos motivos de preocupación relacionados con la seguridad de la información debido al origen chino de la plataforma, aunque hay especialistas que han puesto en duda que la prohibición a nivel de estado sea factible desde el punto de vista técnico.⁴⁶

A su vez, China ha impuesto controles sobre sus principales empresas tecnológicas, incluidas la fundadora de TikTok (ByteDance) y su versión nacional (Douyin). Tras su entrada en vigor en marzo de 2022, las Normas sobre la Administración de los Servicios Informativos en Internet añaden los requisitos de transparencia, control por parte de las personas usuarias y deber de realizar evaluaciones de riesgos —que recuerdan a la LSD de la UE— al control represivo de los medios de comunicación de China, que exige que los sistemas algorítmicos de recomendación “ofrezcan información que se ajuste a los valores dominantes” y “prevengan o reduzcan las disputas o controversias”.⁴⁷ Para garantizar que las plataformas no “ponen en peligro la seguridad nacional ni el interés público social”, la Administración del Ciberespacio de China ha recibido un amplio acceso a información sobre los sistemas algorítmicos de las plataformas, aunque, según informes, las reuniones celebradas entre la Administración del Ciberespacio y ByteDance pusieron de manifiesto la carencia de competencia técnica en la agencia para comprender y examinar los datos.⁴⁸ Posiblemente en respuesta a

-
40. Alexander Misakian y Tiffany Young, “California enacts the California Age-Appropriate Design Code Act”, 20 de septiembre de 2022, <https://www.foley.com/en/insights/publications/2022/09/california-enacts-age-appropriate-design-code-act> [Traducción de AI]
 41. The Verge, “Court blocks California’s online child safety law”, 18 de septiembre de 2023, <https://www.theverge.com/2023/9/18/23879489/california-age-appropriate-design-code-act-blocked-unconstitutional-first-amendment-injunction>
 42. Véase, por ejemplo: Fairplay, “Josh Golin statement regarding California Age Appropriate Design Code”, 23 de mayo de 2022, <https://fairplayforkids.org/josh-golin-ca-aadc/>
 43. Asociación Internacional de Profesionales de la Privacidad, “California Age Appropriate Design Code final passage brings mixed reviews”, 31 de agosto de 2022, <https://iapp.org/news/a/california-age-appropriate-design-code-final-passage-brings-mixed-reviews/>
 44. NPR, “Utah’s new social media law means children will need approval from parents”, 24 de marzo de 2023, <https://www.npr.org/2023/03/24/1165764450/utahs-new-social-media-law-means-children-will-need-approval-from-parents>
 45. CNN, “Arkansas governor signs sweeping bill imposing a minimum age limit for social media usage”, 12 de abril de 2023, <https://edition.cnn.com/2023/04/12/tech/arkansas-social-media-age-limit/index.html>; CNN, “Louisiana lawmakers approve parental consent bill for kids’ social media use and other online services”, 8 de junio de 2023, <https://edition.cnn.com/2023/06/08/tech/louisiana-parental-consent-bill-social-media/index.html>
 46. NPR, “Montana banned TikTok. Whatever comes next could affect the app’s fate in the U. S.”, 18 de mayo de 2023, <https://www.npr.org/2023/05/18/1176940592/montana-ban-tiktok-lawsuit-constitution>
 47. Friedrich Ebert Stiftung, *China’s Regulations on Algorithms: Contexts, impacts and comparisons with the EU*, enero de 2023, <https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf> [Traducción de Amnistía Internacional]
 48. Matt Sheehan y Sharon Du, “What China’s Algorithm registry reveals about AI governance”, 9 de diciembre de 2022, <https://carnegieendowment.org/2022/12/09/what-china-s-algorithm-registry-reveals-about-ai-governance-pub-88606> [Traducción de Amnistía Internacional]

una campaña gubernamental más amplia para limitar el tiempo que los niños y niñas dedican al entretenimiento online, la propia Douyin impuso en 2021 un límite de 40 minutos al uso diario de la plataforma para las personas menores de 14 años.⁴⁹

Otros Estados tales como Brasil, India y Turquía han intentado restringir el poder de las “grandes plataformas tecnológicas” con el objetivo expreso de detener la difusión de contenidos ilegales y la desinformación. Sin embargo, hay activistas de la sociedad civil y periodistas que han expresado su gran preocupación por las extralimitaciones de los gobiernos y los efectos de esas medidas sobre la libertad de expresión online.⁵⁰

Desde un punto de vista global, urge realizar más avances para responder a los riesgos sistémicos tratados en este informe de manera respetuosa con los derechos humanos.

3.3 INTENSIFICACIÓN DE LOS PROBLEMAS DE SALUD MENTAL ENTRE LOS NIÑOS, NIÑAS Y JÓVENES

Lo que subyace al debate sobre los efectos de las redes sociales sobre la salud y el bienestar de los niños, niñas y jóvenes es el aumento documentado de los niveles de ansiedad, depresión y autolesión entre la población adolescente de Estados Unidos y otros países anglófonos en la década de 2010, que se aceleró aún más durante la pandemia de COVID-19.⁵¹ Aunque la evidencia de tendencias parecidas en otras partes del mundo es escasa, en parte por la falta de datos, la sensibilización sobre el nivel de necesidades insatisfechas en materia de salud mental en todo el mundo va en aumento.⁵² Según la iniciativa Intercambio Mundial de Datos de Salud, aproximadamente una de cada siete personas de entre 10 y 19 años de todo el mundo sufre problemas de salud mental, y la Organización Mundial de la Salud ha señalado que la ansiedad y la depresión figuran entre los más comunes.⁵³ Las personas especialistas en salud pública consideran que el porcentaje de jóvenes que sufren problemas de salud mental en África subsahariana es posiblemente superior al de los países de ingresos elevados, aunque la falta de datos representativos de algunas partes de la región y el probable subregistro debido a la falta de servicios asequibles y el estigma social dificultan los intentos de comprender y abordar esta cuestión.⁵⁴

En Kenia y Filipinas, donde Amnistía Internacional realizó la investigación cualitativa para este informe, el acceso a servicios de salud mental gratuitos o asequibles sigue siendo muy escaso y estos servicios suelen estar ubicados en las principales ciudades y sus alrededores.⁵⁵ En Filipinas, algunas personas jóvenes participantes en la investigación afirmaron que, aunque en algunos casos podían

49. BBC, “China: Children given daily time limit on Douyin – its version of TikTok”, 20 de septiembre de 2021, <https://www.bbc.co.uk/news/technology-58625934>.

50. Global Voices, “How India's new internet regulations will change social media, online news and video streaming”, 16 de marzo de 2021, <https://globalvoices.org/2021/03/16/how-indias-new-internet-regulations-will-change-social-media-online-news-and-video-streaming/>; CNN, “Critics fear Turkey's new social media law could hurt freedom of expression. Here's how”, 29 de julio de 2020, <https://edition.cnn.com/2020/07/29/europe/turkey-social-media-law-intl/index.html>; y MIT Technology Review, “Brazil's 'fake news' bill won't solve its misinformation problem”, 10 de septiembre de 2020, <https://www.technologyreview.com/2020/09/10/1008254/brazil-fake-news-bill-misinformation-opinion>.

51. Para una información detallada de la evidencia disponible, véase: Jonathan Haidt, Zach Rausch y Jean Twenge, (en curso). *Social Media and Mental Health: A Collaborative Review* (antes citado).

52. Organización Mundial de la Salud, *Informe mundial sobre la salud mental: Transformar la salud mental para todos*, junio de 2022, https://iris.paho.org/bitstream/handle/10665.2/57878/9789275327715_spa.pdf?sequence=5.

53. Organización Mundial de la Salud, “Nota descriptiva Salud mental del adolescente” (antes citado).

54. Elsabe Brits, “High mental health burden for Africa's youth”, 20 de octubre de 2021, <https://www.nature.com/articles/d44148-021-00097-y>; e Ismail Temitayo Gbadamosi, Isaac Tabiri Henneh y otros, “Depression in Sub-Saharan Africa”, 17 de marzo de 2022, IBRO Neuroscience Reports, volumen 12, <https://www.ncbi.nlm.nih.gov/35746974/>.

55. Conclusión basada en las conversaciones mantenidas con organizaciones de salud mental, profesionales de la psicología y organizaciones de activismo sobre salud mental, como por ejemplo Servicios Comunitarios En Contacto, Jerome Cleofas y Marc Reyes, entre marzo y mayo de 2023.

conseguir que las diagnosticaran de enfermedades mentales, no podían pagar las medicinas que les recetaban o tenían dificultades para encontrar especialistas en psicología que pudieran ofrecer atención de largo plazo.⁵⁶

Aunque la disponibilidad de atención de la salud mental es mayor en Estados Unidos, el sector de la psiquiatría y psicología especializadas en adolescentes tiene dificultades para responder a la demanda creciente de ayuda profesional relacionada con trastornos del estado de ánimo y pensamientos suicidas.⁵⁷ Según una encuesta de los Centros para el Control y la Prevención de Enfermedades (CDC), el 30% de las alumnas de secundaria estadounidenses consideró seriamente la posibilidad de cometer suicidio en 2021 (porcentaje superior al 19% registrado en 2011), y el 13% manifestó haber cometido intentos reales de suicidio.⁵⁸ Los servicios de urgencias de Estados Unidos notificaron en 2022 el doble de número de casos de trastornos de la conducta alimentaria de mujeres adolescentes con respecto a las cifras prepandémicas.⁵⁹ En todo el conjunto de datos de los CDC, las mujeres jóvenes y las personas LGBTI⁶⁰ adolescentes declararon sufrir niveles de violencia sexual, tristeza y riesgo de suicidio superiores a los de los hombres jóvenes y adolescentes heterosexuales.⁶¹

Todos estos datos ponen de manifiesto la urgencia de examinar con mayor detenimiento la manera en que el gran número de jóvenes que sufren problemas de salud mental en todo el mundo se relacionan con las redes sociales, y la manera en que este grupo especialmente vulnerable de personas se ve afectado por las estrategias de negocio basadas en la maximización de la interacción que aplican las redes sociales.

56. Entrevista con una mujer de 20 años realizada en Manila el 5 de mayo de 2023; entrevista con un hombre trans de 23 años realizada en Manila el 6 de mayo de 2023; y conversación con grupo-muestra, que incluye las declaraciones de una mujer de 22 años, realizada en Manila el 6 de mayo de 2023.

57. New York Times Magazine, "How do you actually help a suicidal teen?", 17 de marzo de 2023, <https://www.nytimes.com/2023/05/17/magazine/suicide-teens.html>

58. CDC, "CDC report shows concerning increases in sadness and exposure to violence among teen girls and LGBTQ+ youth", actualizado el 9 de marzo de 2023, <https://www.cdc.gov/nchhstp/newsroom/fact-sheets/healthy-youth/sadness-and-violence-among-teen-girls-and-LGBTQ-youth-factsheet.html>

59. CDC, informe semanal de morbilidad y mortalidad, "Pediatric emergency department visits associated with mental health conditions before and during the COVID-19 pandemic — United States, January 2019-January 2022", 25 de febrero de 2022, Morbidity and Mortality Weekly Report, volumen 21, número 8, <https://dx.doi.org/10.15585/mmwr.mm7108e2>

60. En la encuesta no había información sobre las personas transgénero.

61. CDC, "Pediatric emergency department visits associated with mental health conditions before and during the COVID-19 pandemic — United States, January 2019-January 2022", 25 de febrero de 2022, Morbidity and Mortality Weekly Report (antes citado).

4. MARCO DE DERECHOS HUMANOS

En el informe de 2019 titulado *Gigantes de la vigilancia: La amenaza que el modelo de negocios de Google y Facebook representa para los derechos humanos*, Amnistía Internacional documentó la manera en que Meta (entonces llamada Facebook) y Google (propiedad de Alphabet) han creado un modelo empresarial fundamentado en la vulneración del derecho a la privacidad de las personas usuarias, y han transformado la recopilación de datos personales privados de miles de millones de personas en un medio para generar ingresos por publicidad.⁶² Desde entonces, centenares de millones de personas más se han unido a plataformas de redes sociales de más reciente aparición que utilizan el mismo modelo empresarial basado en la vigilancia para dirigirse a la audiencia online que crece a mayor velocidad: los niños, niñas y jóvenes.⁶³

En el informe de Amnistía Internacional de 2023 *“Siento que no se me protege”: Emboscada en la red de vigilancia de TikTok* se describe la forma en que TikTok recopila y analiza los datos de niños, niñas y jóvenes para tener al día y hacer crecer una base de personas usuarias a las que poder dirigir anuncios personalizados.⁶⁴ Además, se examina en detalle el hecho de que el modelo empresarial basado en la vigilancia vulnera el derecho a la privacidad, lo que menoscaba la capacidad de las personas jóvenes para controlar su información personal, y viola el derecho a libertad de pensamiento, concretamente el derecho a reservarse los pensamientos y opiniones.

Muchos de estos motivos de preocupación guardan relación con las cuestiones tratadas en el presente informe, especialmente el hecho de que TikTok utiliza el caudal de datos personales que recopila y las inferencias obtenidas de ellos para dirigir contenidos concretos a las personas usuarias y mantenerlas interesadas, así como que su sistema algorítmico de recomendación de contenidos expone a niños, niñas y jóvenes con problemas previos de salud mental a riesgos graves de sufrir daños psíquicos.

A continuación se presenta un resumen del marco de los derechos humanos de aplicación en relación con estos motivos de preocupación.

62. Amnistía Internacional, *Gigantes de la vigilancia* (antes citado).

63. Unión Internacional de Telecomunicaciones, “Hechos y cifras 2022: lo último sobre conectividad mundial en plena recesión económica”, 30 de noviembre de 2022, <https://www.itu.int/hub/2022/11/facts-and-figures-2022-global-connectivity-statistics/#/es>; y Forbes, “What The Rise Of TikTok Says About Generation Z”, 7 de julio de 2020, <https://www.forbes.com/sites/forbestechcouncil/2020/07/07/what-the-rise-of-tiktok-says-about-generation-z/>

64. TikTok permite a las empresas anunciantes dirigir la publicidad a personas usuarias de tan sólo 13 años en todas las regiones excepto en la más estrictamente regulada (la UE) y los otros dos Estados (Suiza y Reino Unido) que la plataforma incluye en la política de privacidad de esa región. Amnistía Internacional, *I feel exposed: Caught in TikTok's Surveillance Web*, 2023 (antes citado).

4.1 EL DERECHO A LA PRIVACIDAD EN LA ERA DIGITAL

“La privacidad es vital para la autonomía, la dignidad y la seguridad de los niños y para el ejercicio de sus derechos”, según se establece en la Observación general núm. 25 del Comité de los Derechos del Niño relativa a los derechos de los niños en relación con el entorno digital.⁶⁵

El derecho a la privacidad se reconoce en numerosos instrumentos de derechos humanos, pero probablemente nunca se ha visto cuestionado tan ampliamente como por el auge de las redes sociales actualmente dominantes a nivel mundial. El derecho a la privacidad establece que nadie debe ser objeto de “injerencias arbitrarias o ilegales” en su vida privada, su familia, su domicilio o su correspondencia, y que este principio debe estar protegido por la ley.⁶⁶ El principal instrumento de los derechos de la infancia, la Convención sobre los Derechos del Niño (CDN) se hace eco de que “[e]l niño tiene derecho a la protección de la ley contra esas injerencias o ataques”.⁶⁷ Reconociendo que la innovación tecnológica ha planteado muchos desafíos nuevos al derecho a la privacidad desde que se establecieron estos instrumentos de derechos humanos, el Comité de Derechos Humanos de la ONU ha precisado que esta protección incluye la regulación de “[l]a recopilación y el registro de información personal en computadoras, bancos de datos y otros dispositivos, tanto por las autoridades públicas como por las particulares o entidades privadas”.⁶⁸

La Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (OACNUDH) ha profundizado igualmente en el significado de la privacidad en relación con el entorno digital:

“La privacidad puede entenderse como la presunción de que el individuo debe tener una esfera de desarrollo autónomo, interacción y libertad, una ‘esfera privada’ con o sin relación con otras y libre de la intervención del Estado y de la intervención excesiva no solicitada de otros individuos no invitados [...]. En el entorno digital, la privacidad de la información, que abarca la información que existe o puede obtenerse acerca de una persona y de su vida y las decisiones basadas en esa información, tiene especial importancia”.⁶⁹

El derecho a la privacidad engloba tres conceptos interrelacionados: no sufrir intromisiones en nuestra vida privada, el derecho a controlar la información sobre nuestra persona y el derecho a un espacio donde podamos expresar libremente nuestra identidad. Amnistía Internacional considera que el rasgo característico de basarse en la vigilancia que presenta el modelo empresarial de TikTok menoscaba cada uno de estos tres aspectos hasta tal punto que ha menoscabado la esencia misma de la privacidad.

EL DERECHO A NO SUFRIR INTROMISIONES

El alto comisionado de la ONU para los derechos humanos ha reconocido que “[e]l simple hecho que se generen y reúnan datos relativos a la identidad, la familia o la vida de una persona ya afecta a su derecho a la privacidad, pues a través de esas acciones, la persona pierde en cierta medida el control sobre una información que podría poner en riesgo su vida privada”.⁷⁰ La protección del derecho

65. Comité de los Derechos del Niño, de la ONU, Observación general núm. 25: Derechos de los niños en relación con el entorno digital, 2 de marzo de 2021, doc. ONU CRC/C/GC/25, párr. 67.

66. Declaración Universal de Derechos Humanos (DUDH, artículo 12; y Pacto Internacional de Derechos Civiles y Políticos (PIDCP), artículo 17.

67. CDN, artículo 16.

68. Comité de Derechos Humanos de la ONU (CDH), Observación general núm. 16 sobre el derecho al respeto de la vida privada, la familia, el domicilio y la correspondencia, así como la protección de la honra y la reputación (artículo 17), 8 de abril de 1988, párr. 10.

69. Alta Comisionada de las Naciones Unidas para los Derechos Humanos, *El derecho a la privacidad en la era digital*, 3 de agosto de 2018, doc. ONU A/HRC/39/29, párr. 5.

70. Alto comisionado de la ONU para los derechos humanos, informe: *El derecho a la privacidad en la era digital*, 3 de agosto de 2018 (antes citado), párr. 7.

a la privacidad abarca no sólo el contenido de las comunicaciones, “sino también los metadatos [información sobre los archivos y las comunicaciones, etc. digitales, por ejemplo las marcas de tiempo y la localización], puesto que, al analizarse y reunirse, éstos pueden ‘dar una mejor idea del comportamiento, las relaciones sociales, las preferencias privadas y la identidad de una persona que la información obtenida accediendo al contenido de una comunicación privada’”.⁷¹

El derecho internacional de los derechos humanos sólo permite injerencias en el derecho a la vida privada de la persona si éstas no son ni arbitrarias ni ilegítimas. Los mecanismos de derechos humanos han interpretado sistemáticamente que esas palabras apuntan a los principios generales de legalidad, necesidad y proporcionalidad.⁷² Sin embargo, el Comité de los Derechos del Niño ha concluido que “[l]as prácticas digitales, como el procesamiento automatizado de datos, la elaboración de perfiles, la selección de comportamientos, la verificación obligatoria de la identidad, el filtrado de información y la vigilancia masiva, se están convirtiendo en procedimientos de rutina” en la vida infantil.⁷³ Tal como se estudia en más detalle en el documento “*Siento que no se me protege*”: *Emboscada en la red de vigilancia de TikTok*, las plataformas dominantes de redes sociales, como TikTok, están instaurando formas de vigilancia empresarial masiva que son intrínsecamente innecesarias y desproporcionadas y en ningún caso pueden considerarse una injerencia permisible en el derecho a la privacidad.⁷⁴

EL DERECHO A CONTROLAR NUESTRA INFORMACIÓN PERSONAL

El segundo componente de la privacidad establece que las personas tienen el derecho a controlar su información personal, o derecho a la “autodeterminación informativa”, para poder decidir el momento y la manera en que sus datos personales pueden facilitarse a terceros.⁷⁵ Esta idea es la base de la regulación de la protección de datos. El Tribunal Europeo de Derechos Humanos (TEDH) ha reconocido que la protección de los datos personales tiene una importancia fundamental para que las personas disfruten de su derecho a la privacidad,⁷⁶ y que la privacidad contempla el derecho a una forma de autodeterminación informativa.⁷⁷

EL DERECHO A UN ESPACIO EN EL QUE PODER EXPRESAR LIBREMENTE LA IDENTIDAD

Por último, existe un consenso generalizado en cuanto a de que la privacidad también es fundamental para crear y proteger el espacio necesario para construir nuestras propias identidades.⁷⁸ El Comité de Derechos Humanos de la ONU ha definido la privacidad como “una esfera de la vida en la que las personas pueden expresar libremente su identidad”.⁷⁹ Esto refleja la noción de que el sentimiento personal de la identidad se construye socialmente y es dinámico: las personas exploran y muestran diferentes aspectos de sí mismas en diferentes contextos —con sus amistades, en la escuela, en el trabajo o en público—, y estas identidades cambian y se adaptan constantemente.

71. Alto comisionado de la ONU para los derechos humanos, informe: *El derecho a la privacidad en la era digital*, 3 de agosto de 2018 (antes citado), párr. 6.

72. Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (OACNUDH), informe: *El derecho a la privacidad en la era digital*, 30 de junio de 2014, doc. ONU A/HRC/27/37, párrs. 21-27.

73. Comité de los Derechos del Niño, Observación general núm. 25 (antes citado), párr. 68.

74. Amnistía Internacional, “*Siento que no se me protege*”: *Emboscada en la red de vigilancia de TikTok* (antes citado).

75. El término “autodeterminación informativa” se utilizó por primera vez en el marco de una sentencia del Tribunal Constitucional alemán relacionada con una información personal recopilada durante la elaboración del censo de 1983. El Tribunal resolvió que corresponde a cada persona decidir cuándo y con qué limitaciones debe comunicarse la información sobre su vida privada a terceros. BVerfG, Urteil des Ersten Senats vom 15. Dezember 1983 – 1 BvR – Rn. 1-215, https://www.bverfg.de/e/rs19831215_1bvr020983.html

76. TEDH, *S. and Marper v. UK*, demandas 30562/04 y 30566/04, sentencia de la Gran Sala, 4 de diciembre de 2008, [https://hudoc.echr.coe.int/eng#{%22itemid%22:\[%22001-90051%22\]}](https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-90051%22]})

77. TEDH, *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland*, demanda 931/13, sentencia de la Gran Sala, 27 de junio de 2017, [https://hudoc.echr.coe.int/eng#{%22itemid%22:\[%22001-175121%22\]}](https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-175121%22]}), párr. 137.

78. Véase, por ejemplo, Philip Agre y Marc Rotenburg (editores), *Technology and Privacy: The New Landscape*, 1998.

79. CDH, Decisión, *Coeriel and Aurik v. the Netherlands*, adoptada el 31 de octubre de 1994, doc. ONU CCPR/C/52/D/453/1991, párr. 10.2. [Traducción de Amnistía Internacional]

Las plataformas de redes sociales tales como Instagram y TikTok ejercen una enorme influencia sobre la gente joven porque son los espacios dominantes de exploración y representación de la identidad. Para participar en este espacio, los niños, niñas y jóvenes se expresan por propia voluntad, pero también se ven empujados a dar a conocer información personal mediante las publicaciones, las historias, los reels (vídeos cortos) y los comentarios, así como a revelar sus preferencias e intereses personales a través de los “me gusta” o difundiendo contenidos de otras personas. Y esto sólo es lo que las demás personas usuarias de estas plataformas pueden ver. Las empresas en cuestión conocen mucho más de la vida privada de la gente joven que interactúa con sus plataformas a través de la información sobre lo que los niños, niñas y jóvenes ven o lo que capta su atención, la frecuencia y duración de los accesos, los datos de localización, los datos sobre los horarios en los que se conectan, los dispositivos que utilizan y las compras que realizan, por mencionar sólo algunos aspectos.⁸⁰

Aunque los estudios muestran que los y las adolescentes más jóvenes quizá no son plenamente conscientes de la vigilancia empresarial a la que están sometidos, normalmente son muy conscientes de las presiones sociales inherentes a las plataformas que dependen de que las personas usuarias vean a otras y sean vistas, y los límites que se generan y en los cuales la gente joven considera que puede explorar su identidad.⁸¹ Las entrevistas que Amnistía Internacional realizó a los y las jóvenes brindaron mucha información sobre los complejos modos en los que muchos de ellos controlan qué personas de su entorno social acceden a determinados tipos de publicaciones, a menudo utilizando varias cuentas en una misma plataforma con distintos niveles de visibilidad y representación pública o anonimato.⁸²

4.2 EL DERECHO A LA LIBERTAD DE PENSAMIENTO

El derecho a la libertad de pensamiento, amparado por el artículo 18.1 del Pacto Internacional de Derechos Civiles y Políticos (PIDCP), es un derecho absoluto, lo que significa que nadie puede inmiscuirse en nuestros pensamientos y creencias privados en ninguna circunstancia. A diferencia de la libertad de expresión (la comunicación de las opiniones personales), que puede limitarse por ley cuando sea necesario y de manera proporcionada y para alcanzar un objetivo legítimo⁸³ —por ejemplo para proteger la dignidad y los derechos de otras personas—, los pensamientos, las creencias y las opiniones que guardamos en la cabeza nos pertenecen por completo, y no deben sonsacarse contra nuestra voluntad ni manipularse. La CDN protege la libertad de pensamiento de los niños y niñas en su artículo 14, que dispone que los Estados Partes “respetarán el derecho del niño a la libertad de pensamiento, de conciencia y de religión”.

Íntimamente ligado a lo anterior, el derecho a la libertad de opinión, que también es un derecho absoluto, está amparado por el artículo 19.1 del PIDCP y establece que “[n]adie podrá ser molestado a causa de sus opiniones”.

El derecho a la libertad de pensamiento y la libertad de opinión comporta tres aspectos:

-
80. Amnistía Internacional, *I feel exposed: “Siento que no se me protege”: Emboscada en la red de vigilancia de TikTok*, 7 de noviembre de 2023 (antes citado).
 81. Mariya Stoilova, Sonia Livingstone y otros, “Digital by default: children’s capacity to understand and manage online data and privacy”, 3 de noviembre de 2020, *Media and Communication*, volumen 8, número 4, <https://eprints.lse.ac.uk/107114/>; y Kathryn C. Montgomery, Jeff Chester y otros, “Children’s privacy in the big data era: Research opportunities”, 1 de noviembre de 2017, *Pediatrics*, volumen 140, suplemento 2, <https://doi.org/10.1542/peds.2016-17580>.
 82. Muchas personas adultas jóvenes especialmente de Filipinas mencionaron los usos diferentes de las cuentas públicas “OG / Original Gangster” (originales) y sus cuentas “dump” (para volcar fotos) o “rant” (para despotricar) (cuentas con un menor número de personas seguidoras de confianza, que se utilizan para dar una versión más sincera y “sin filtros” de la vida, los sentimientos y las opiniones de los y las jóvenes; suelen usarse, de manera no exclusiva, en Instagram).
 83. PIDCP, artículo 19; y CDN, artículo 13.

- el derecho a reservarnos los pensamientos y opiniones;
- El derecho a que no se manipulen nuestros pensamientos y opiniones, y
- el derecho a que no se nos sancione por nuestros pensamientos y opiniones.⁸⁴

Cuando se redactaron el PIDCP y otros tratados de derechos humanos, nadie podía haber previsto un desafío al derecho a la libertad de pensamiento tan amplio y generalizado como el que plantea hoy la tecnología moderna. El modelo empresarial basado en la vigilancia propio de las empresas de redes sociales, que implica la recopilación masiva de datos personales privados sobre todas las personas usuarias (y la recopilación de ciertos datos no privados incluso de quienes no se han abierto una cuenta⁸⁵), plantea una amenaza directa a estos derechos de una forma completamente nueva. Tras recopilar esos datos, las empresas de redes sociales los utilizan para analizar a las personas, agruparlas y elaborar predicciones sobre sus intereses, características y, en última instancia, su comportamiento, principalmente para poder usar esta información y generar ingresos publicitarios.⁸⁶

En la publicación *“Siento que no se me protege”: Emboscada en la red de vigilancia de TikTok se destaca el hecho de que el propósito lucrativo de las empresas de recopilar y sumar la mayor cantidad posible de datos personales privados sobre las personas usuarias de las redes sociales —y obtener inferencias de las publicaciones, los “me gusta”, los contactos, el tiempo de visión, los comentarios, la ubicación, la información de los dispositivos, etc. para reconstruir los intereses y la orientación sexual y política de las personas, por mencionar sólo algunas características— viola el primer aspecto de este derecho: el derecho a reservarnos los pensamientos y opiniones.*⁸⁷

En el presente informe se examina más minuciosamente el hecho de que TikTok utiliza su caudal de datos personales y las inferencias obtenidas de ellos para dirigir contenidos concretos a las personas usuarias a fin de mantener su interés y, por tanto, podría vulnerar el derecho a que no se manipulen nuestros pensamientos y opiniones. “El Comité de los Derechos del Niño profundiza en este argumento al afirmar que

“alienta a los Estados partes a que establezcan una normativa de protección de datos y de diseño, o actualicen la existente, a fin de incluir la identificación, definición y prohibición de las prácticas que manipulen o inhiban el derecho de los niños a la libertad de pensamiento y de creencias en el entorno digital, por ejemplo mediante el análisis emocional o la inferencia. Los sistemas automatizados pueden utilizarse para hacer inferencias sobre el estado interior de un niño. Los Estados partes deben garantizar que los sistemas automatizados o los sistemas de filtrado de información no se utilicen para afectar o influenciar el comportamiento o las emociones de los niños ni para limitar sus oportunidades o su desarrollo”.⁸⁸

El Consejo de Europa (CdE) también ha advertido de que “[l]os niveles de persuasión meticulosos, subconscientes y personalizados pueden tener efectos considerables en la autonomía cognitiva de las personas y su derecho a formarse opiniones y tomar decisiones independientes.”⁸⁹

84. Vermeulen, “Freedom of thought, conscience and religion (article 9)”, en *Theory and Practice of the European Convention on Human Rights*, 4ª edición, P. van Dijk, F. van Hoof, A. van Rijn y L. Zwaak (editores) (Cambridge: Intersentia Press), 751–772 citado en McCarthy-Jones, “The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century”, *Front. Artif. Intell.*, 26 de septiembre de 2019, apartado “Tecnología y derecho”, <https://doi.org/10.3389/frai.2019.00019>; y Evelyn Aswad, “Losing the freedom to be human”, 8 de julio de 2020, *Columbia Human Rights Law Review*, volumen 52, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635701

85. Vice, “TikTok is watching you – even if you don’t have an account”, 21 de enero de 2021, <https://www.vice.com/en/article/jgqbm/tiktok-data-collection>; y Wired, “All the ways TikTok tracks you and how to stop it”, 23 de octubre de 2021, <https://www.wired.co.uk/article/tiktok-data-privacy>

86. Amnistía Internacional, *Gigantes de la vigilancia* (antes citado).

87. Amnistía Internacional, *“Siento que no se me protege”: Emboscada en la red de vigilancia de TikTok*, 7 de noviembre de 2023 (antes citado).

88. Comité de los Derechos del Niño, Observación general núm. 25 (antes citado), párr. 62.

89. CdE, Declaración sobre la capacidad de manipulación de los procesos algorítmicos, 13 de febrero de 2019, Decl(13/02/2019)1, <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168092dd4b> [Traducción de Amnistía Internacional]

SISTEMAS ALGORÍTMICOS DE RECOMENDACIÓN

Los sistemas que posibilitan la organización y personalización de los feeds de las redes sociales son más comúnmente conocidos como “el algoritmo”. En realidad, TikTok y otras plataformas de redes sociales se basan en una multitud de herramientas automatizadas para recopilar, combinar, clasificar e inferir datos y para tomar decisiones sobre los contenidos que deben mostrarse a una persona usuaria concreta en un momento concreto. La complejidad de estas herramientas interrelacionadas —y el muy limitado acceso de las personas investigadoras a los sistemas y datos pertinentes— plantea problemas específicos a la hora de “auditar” (es decir, examinar) los sistemas de recomendación como, por ejemplo, el que alimenta el feed “Para ti” de TikTok.

Las conclusiones de Amnistía Internacional, que se detallan en los apartados siguientes, indican no sólo que TikTok emplea prácticas de diseño adictivas y manipuladoras, sino también que los sistemas algorítmicos de recomendación de contenidos de la plataforma —a los que se atribuye su rápido auge mundial—⁹⁰ exponen a niños, niñas y jóvenes con problemas previos de salud mental a riesgos graves de sufrir daños psíquicos.

4.3 EL DERECHO A LA SALUD

Los Estados tienen la obligación de adoptar medidas para alcanzar progresivamente la realización plena del derecho de toda persona al nivel más alto posible de salud física y mental. El derecho a la salud está reconocido por varios instrumentos de derechos humanos.⁹¹ El Comité de Derechos Económicos, Sociales y Culturales ha manifestado que debe haber centros y servicios de atención a la salud disponibles en cantidad suficiente, al alcance y asequibles para todos los sectores de la población.⁹²

Sin embargo, la atención de la salud mental sigue estando ampliamente descuidada y desprovista de recursos suficientes, lo que llevó al relator especial de la ONU sobre el derecho de toda persona al disfrute del más alto nivel posible de salud física y mental a señalar en el informe que presentó en 2020 al Consejo de Derechos Humanos:

“No hay salud sin salud mental. Los abundantes vínculos existentes entre la mente, el cuerpo y el entorno están bien documentados desde hace décadas. Al iniciarse la tercera década del milenio, en ninguna parte del mundo se ha logrado la paridad entre la salud mental y la física y esto sigue siendo un problema considerable para el desarrollo humano. Un mensaje importante dentro de esa inacción colectiva es que, si no se presta la debida atención a los derechos humanos, cualquier inversión que se haga en salud mental será ineficaz”.⁹³

El relator también advirtió en junio de 2020 que la pandemia de COVID-19 había agravado el “descuido histórico de la atención de la salud digna” y que “[l]a combinación del cierre de los colegios y las condiciones de confinamiento t[enía] un efecto especial en el estrés, la ansiedad y las cuestiones

90. Alex Hern, “How TikTok’s algorithm made it a success: ‘It pushes the boundaries’”, 24 de octubre de 2022, <https://www.theguardian.com/technology/2022/oct/23/tiktok-rise-algorithm-popularity>

91. DUDH, artículo 25; PIDESC, artículo 12 y CDN, artículo 24. Véanse también los tratados regionales de derechos humanos tales como la Carta Africana de Derechos Humanos y de los Pueblos, artículo 16; y la Carta Social Europea (revisada).

92. Comité de Derechos Económicos, Sociales y Culturales, Observación general núm. 14, 11 de agosto de 2000, doc. ONU E/C.12/2000/4.

93. Relator especial de la ONU sobre el derecho de toda persona al disfrute del más alto nivel posible de salud física y mental (relator especial de la ONU sobre el derecho a la salud), informe, 15 de abril de 2020, doc. ONU A/HRC/44/48, párr. 1.

de salud mental de los niños” y “[esa] situación e[ra] especialmente preocupante dada la falta general de reconocimiento y conciencia sobre la índole y la envergadura de los problemas de salud mental entre los adolescentes de muchos países”.⁹⁴

Sin lugar a dudas, las tecnologías digitales han desempeñado un papel positivo en la reducción de las barreras de acceso a la información y los servicios de salud, por ejemplo posibilitando que las personas profesionales de la psicología hablaran con sus pacientes a distancia durante los confinamientos, manteniendo a la gente joven en contacto con los servicios de ayuda y permitiendo a grupos marginados tales como la población adolescente LGBTI explorar sus identidades y expresarse.⁹⁵ No obstante, se reconoce cada vez más que, en ausencia de una regulación estatal eficaz, las plataformas de redes sociales en concreto también han expuesto a niños, niñas y jóvenes a riesgos sistémicos tales como los que se describen en este informe.⁹⁶ En este sentido, el Comité de los Derechos del Niño señaló en su Observación general núm. 25:

“En el entorno digital puede haber información que propugne los estereotipos de género, la discriminación, el racismo, la violencia, la pornografía y la explotación, así como relatos falsos, información errónea y desinformación, e información que incite a los niños a participar en actividades ilícitas o perjudiciales. Esa información puede proceder de múltiples fuentes, tales como otros usuarios, creadores de contenidos comerciales, delincuentes sexuales o grupos armados designados como terroristas o extremistas violentos. Los Estados partes deben proteger a los niños contra los contenidos nocivos y poco fiables y garantizar que las empresas pertinentes y otros proveedores de contenidos digitales elaboren y apliquen directrices que permitan a los niños acceder de forma segura a contenidos diversos, reconociendo los derechos de los niños a la información y a la libertad de expresión, y protegiéndolos al mismo tiempo frente a ese material nocivo de conformidad con sus derechos y la evolución de sus facultades”.⁹⁷ [la cursiva es nuestra].

Centrándose de forma más específica en los riesgos para la salud de los niños, niñas y jóvenes, el cirujano general de Estados Unidos advirtió en mayo de 2023:

“Casi todas las personas adolescentes de América utilizan las redes sociales y, sin embargo, los datos disponibles no bastan para concluir que éstas les ofrecen suficiente seguridad. Nuestros niños y niñas se han convertido en participantes inconscientes de un experimento que dura decenios [...]. Debemos reconocer el corpus cada vez mayor de investigaciones sobre los daños potenciales, aumentar el conocimiento colectivo que tenemos de los riesgos asociados al uso de las redes sociales y adoptar con urgencia medidas para crear entornos digitales sanos y seguros que minimicen el daño y protejan la salud mental y el bienestar de los niños, niñas y adolescentes durante las etapas críticas del desarrollo”.⁹⁸

Esta actuación urgente requieren de una labor reguladora por parte de los gobiernos en consonancia con sus obligaciones en virtud del derecho internacional de los derechos humanos para satisfacer

94. OACNUDH, “La COVID-19 ha agravado el descuido histórico de la atención de la salud mental digna, en especial para los que se encuentran en instituciones, según experto de las Naciones Unidas”, 23 de junio de 2020, <https://www.ohchr.org/es/2020/06/covid-19-has-exacerbated-historical-neglect-dignified-mental-health-care-especially-those>

95. Benjamin Hanckel y Shiva Chandra, “How young LGBTQIA+ people used social media to thrive during COVID lockdowns”, The Conversation, 15 de marzo de 2021, <https://theconversation.com/how-young-lgbtqia-people-used-social-media-to-thrive-during-covid-lockdowns-156130>

96. Véanse por ejemplo OACNUDH, *Innovación digital, tecnologías y derecho a la salud. Informe de la relatora especial sobre el derecho de toda persona al disfrute del más alto nivel posible de salud física y mental*, abril de 2023, <https://www.ohchr.org/es/documents/thematic-reports/ahrc5365-digital-innovation-technologies-and-right-health>; y Comité de los Derechos del Niño, Observación general núm. 25, 2 de marzo de 2021, doc. ONU CRC/C/GC/25

97. Comité de los Derechos del Niño, Observación general núm. 25, 2 de marzo de 2021, doc. ONU CRC/C/GC/25, párr. 54.

98. Cirujano general de Estados Unidos *Social Media and Youth Mental Health*, mayo de 2023, <https://www.hhs.gov/surgeongeneral/priorities/youth-mental-health/social-media/index.html> [Traducción de Amnistía Internacional]

de manera progresiva el derecho de la población a la salud, y medidas inmediatas por parte de las empresas de redes sociales para respetar los derechos humanos y cumplir con sus responsabilidades de diligencia debida en materia de derechos humanos (esta cuestión se analiza con mayor detalle en el apartado siguiente).

4.4 EL INTERÉS SUPERIOR DE LA INFANCIA, Y EL DERECHO DE LOS NIÑOS Y NIÑAS A SER ESCUCHADOS

La Convención sobre los Derechos del Niño consagra, entre otros, el principio del “interés superior del niño o niña”. En el artículo 3.1 se establece que: “En todas las medidas concernientes a los niños que tomen las instituciones públicas o privadas de bienestar social, los tribunales, las autoridades administrativas o los órganos legislativos, una consideración primordial a que se atenderá será el interés superior del niño”.⁹⁹

Según el artículo 12 de la CDN, los niños y niñas tienen el derecho a expresar sus opiniones “libremente en todos los asuntos que afectan al niño, teniéndose debidamente en cuenta las opiniones del niño, en función de la edad y madurez del niño”.¹⁰⁰ Se debe tomar en cuenta la evolución de sus facultades, así como el principio del interés superior y el derecho que tienen a que se escuchen sus opiniones.

Al aplicar estos principios al entorno digital, el Comité de los Derechos del Niño recomienda a los Estados que se cercioren de que “en todas las actuaciones relativas al suministro, la regulación, el diseño, la gestión y la utilización del entorno digital, el interés superior de todos los niños sea una consideración primordial”,¹⁰¹ y que “los proveedores de servicios digitales ofrezcan servicios acordes con la evolución de las facultades de los niños”.¹⁰²

4.5 LA RESPONSABILIDAD EMPRESARIAL DE RESPETAR LOS DERECHOS HUMANOS

Las empresas tienen la responsabilidad de respetar los derechos humanos con independencia del lugar del mundo donde operen y en toda su actividad económica. Se trata de una norma de conducta ampliamente reconocida que figura en las normas internacionales sobre empresas y derechos humanos, como los Principios rectores de la ONU sobre las empresas y los derechos humanos (Principios rectores de la ONU) y las Líneas Directrices de la OCDE para Empresas Multinacionales (Directrices de la OCDE).¹⁰³ La responsabilidad de las empresas de respetar los derechos humanos es independiente de las propias obligaciones de un Estado en materia de derechos humanos, y se trata de una responsabilidad adicional a la de cumplir las leyes y normas nacionales de protección de los derechos humanos.¹⁰⁴

99. CDN, artículo 3.1.

100. CDN, artículo 12.

101. Comité de los Derechos del Niño, Observación general núm. 25 (antes citado), párr. 12.

102. Comité de los Derechos del Niño, Observación general núm. 25 (antes citado), párr. 20.

103. Principios rectores sobre las empresas y los derechos humanos: Puesta en práctica del marco de las Naciones Unidas para “proteger, respetar y remediar”, 2011, aprobado por el Consejo de Derechos Humanos de la ONU (CDH), Resolución del CDH 17/4: *Los derechos humanos y las empresas transnacionales y otras empresas*, aprobada el 16 de junio de 2011, doc. ONU A/HRC/RES/17/4; y Líneas Directrices de la OCDE para Empresas Multinacionales, 2011, https://www.oecd-ilibrary.org/finance-and-investment/lineas-directrices-de-la-ocde-para-empresas-multinacionales-sobre-conducta-empresarial-responsable_7abea681-es. Según los Principios rectores de la ONU, la responsabilidad de las empresas de respetar los derechos humanos es independiente de las obligaciones de un Estado en materia de derechos humanos, y se trata de una responsabilidad adicional a la de cumplir las leyes y normas nacionales de protección de los derechos humanos. Véase Principios rectores de la ONU, principio 11 y su comentario.

104. OACNUDH, Principios rectores sobre las empresas y los derechos humanos: Puesta en práctica del marco de las Naciones Unidas para “proteger, respetar y remediar”, 2011, doc. ONU HR/PUB/11/04, http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_SP.pdf, principio 11 y su comentario.

Para ello, las empresas han de evitar causar o contribuir a causar abusos contra estos derechos a través de sus propias actividades económicas, y hacer frente a las consecuencias negativas en las que tengan alguna participación, lo que incluye remediar cualquier abuso cometido. Los Principios rectores de la ONU establecen que, para cumplir con estas responsabilidades, las empresas deben contar con un proceso continuo y proactivo de diligencia debida en materia de derechos humanos para identificar, prevenir y mitigar los efectos que tienen en este ámbito, y explicar cómo los abordan. Si en ese proceso se concluye que causan o contribuyen a causar abusos, las empresas deben detener su actividad o evitar sus efectos adversos para los derechos humanos.¹⁰⁵

A su vez, los Estados tienen la obligación de respetar y proteger los derechos humanos en el marco de las actividades empresariales. En este sentido, deben “proteger contra las violaciones de los derechos humanos cometidas en su territorio y/o su jurisdicción por terceros, incluidas las empresas. A tal efecto deben adoptar las medidas apropiadas para prevenir, investigar, castigar y reparar esos abusos mediante políticas adecuadas, actividades de reglamentación y sometimiento a la justicia”.¹⁰⁶

En lo que se refiere concretamente a los niños y niñas, el Comité de los Derechos del Niño ha destacado el deber que tienen los Estados de “exigir a las empresas que procedan con la diligencia debida en lo que respecta a los derechos del niño”,¹⁰⁷ y ha establecido medidas que los Estados deben adoptar para impedir a las empresas que causen abusos contra los derechos de la infancia o contribuyan a ellos y para “investigar, enjuiciar y reparar las violaciones de los derechos del niño causadas por una empresa o a las que una empresa haya contribuido”. El Comité ha hecho hincapié en que los Estados son “responsable[s] de dichas violaciones si no ha[n] adoptado las medidas apropiadas, razonables y necesarias para impedir las o repararlas, o si ha[n] tolerado o colaborado de alguna otra forma en su comisión”.¹⁰⁸

La transparencia y la rendición pública de cuentas son fundamentales para realizar con eficacia la diligencia debida respecto a la manera en que una empresa ha identificado, prevenido o mitigados los efectos negativos reales o potenciales en los derechos humanos. En los Principios rectores de la ONU se dispone que las empresas “deben saber y hacer saber que respetan los derechos humanos”,¹⁰⁹ y que “[h]acer saber implica comunicar, ofrecer transparencia y rendir cuentas a las personas o grupos que puedan verse afectados y a otros interesados, incluidos los inversores”.¹¹⁰ El Comité de los Derechos del Niño ha destacado igualmente que los Estados deben fomentar —y, cuando proceda, exigir— que las empresas hagan pública la información sobre las medidas que hayan adoptado para abordar las repercusiones que su actividad tenga sobre los derechos humanos de la infancia.¹¹¹

Este requisito de transparencia contrasta claramente con la realidad de la influencia cada vez mayor que ejercen los algoritmos desarrollados en privado, habilitados para determinar las publicaciones que se amplifican y posteriormente son consumidas por una multitud de personas en las redes sociales, o para establecer los tipos de publicaciones que violan las directrices de las plataformas. Las plataformas de redes sociales llevan tiempo tratando de proteger lo que perciben como secretos de sus modelos empresariales,¹¹² y han tratado de limitar la capacidad de la investigación independiente para realizar un seguimiento de los resultados de la denominada “caja negra” de los algoritmos.¹¹³

105. Principios rectores de la ONU (antes citado), principio 19 y su comentario.

106. Principios rectores de la ONU (antes citado), principio 1.

107. Comité de los Derechos del Niño, Observación general núm. 16, 17 de abril de 2013, doc. ONU CRC/C/GC/16, párr. 62.

108. Comité de los Derechos del Niño, Observación general núm. 16 (antes citado), párr. 28.

109. Principios rectores de la ONU (antes citado), comentario al principio 15.

110. Principios rectores de la ONU (antes citado), comentario al principio 21.

111. Comité de los Derechos del Niño, Observación general núm. 16 (antes citado), párr. 65.

112. Alex Hern, “How TikTok’s algorithm made it a success: ‘It pushes the boundaries’” (antes citado).

113. Fundación Mozilla, Amnistía Internacional y otros, “Response to the European Commission’s call for evidence for a Delegated Regulation on data access provided for in the Digital Services Act”, (antes citado); y Frederick Mostert, “Social media platforms must abandon algorithmic secrecy”, 17 de junio de 2021, <https://www.ft.com/content/39d69f80-5266-4e22-965f-efbc19d2e776>

Ante el creciente reconocimiento de los daños en el mundo real asociados con las plataformas que favorecen y amplifican contenidos extremos,¹¹⁴ varios organismos de la ONU y otras entidades internacionales han tratado de generar respuestas a esta importante laguna normativa.

En 2021, la OACNUDH formuló las recomendaciones siguientes dirigidas a los Estados y las empresas para abordar los riesgos en materia de derechos humanos relacionados con el uso de la inteligencia artificial (IA), que desempeña un papel cada vez mayor en el perfeccionamiento de los sistemas de recomendación:

1. “Observ[ar] sistemáticamente la diligencia debida en materia de derechos humanos a lo largo del ciclo de vida de los sistemas de IA que diseñan, desarrollan, implantan, venden, obtienen o explotan. Un elemento fundamental de la diligencia debida en materia de derechos humanos debería ser la realización de evaluaciones periódicas y exhaustivas de los efectos de estos sistemas sobre tales derechos.
2. Aument[ar] considerablemente la transparencia en su uso de la IA, entre otras cosas, informando adecuadamente al público y a las personas afectadas y permitiendo una supervisión independiente y externa de los sistemas automatizados. Cuanto más probables y graves sean los efectos potenciales o reales sobre los derechos humanos relacionados con el uso de la IA, más necesaria será la transparencia.
3. Garanti[zar] la participación de todas las partes interesadas en las decisiones sobre el desarrollo, la implantación y el uso de la IA, en particular de las personas y grupos afectados.
4. Promov[er] la explicabilidad de las decisiones basadas en la IA, entre otras cosas mediante la financiación y la realización de investigaciones con ese objetivo”.¹¹⁵

Tras los llamamientos del Consejo de Europa¹¹⁶ y otros organismos en favor de la regulación estatal de los sistemas algorítmicos, algunas iniciativas tales como la Ley de Servicios Digitales y la Ley de Inteligencia Artificial de la UE¹¹⁷ tratan de aumentar la transparencia, implementar mecanismos de supervisión e impedir los usos potencialmente abusivos de la AI.

En el capítulo 7 del presente informe se explica el hecho de que TikTok incumple su responsabilidad de respetar los derechos humanos —al no mitigar los riesgos para el derecho a la privacidad, la libertad de pensamiento y el derecho a la salud— por medio de su feed y otros elementos estructurales de maximización de la interacción que sustentan su modelo empresarial basado en la vigilancia.

114. Amnistía Internacional, *Myanmar: Meta y el derecho a un recurso para la comunidad rohinyá* (antes citado); y *Gigantes de la vigilancia* (antes citado).

115. OACNUDH, El derecho a la privacidad en la era digital, 15 de septiembre de 2021, doc. ONU A/HRC/48/31.

116. CdE, “Recommendation CM/Rec (2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems”, 8 de abril de 2020.

117. El proyecto de ley sobre inteligencia artificial de la UE pretende regular los sistemas de IA en un sentido general, mientras que la Ley de Servicios Digitales se orienta a los riesgos asociados a los sistemas algorítmicos de las grandes plataformas de redes sociales. Amnistía Internacional, “Unión Europea: En peligro la Ley de Inteligencia Artificial ante la posibilidad de que el Parlamento Europeo legitime las tecnologías abusivas”, 13 de junio de 2023, <https://www.amnesty.org/es/latest/news/2023/06/eu-ai-act-at-risk-as-european-parliament-may-legitimize-abusive-technologies/>

5. ADICTIVO POR DISEÑO

A diferencia de otros medios de comunicación de masas, las empresas de redes sociales personalizan la experiencia de quienes las usan para aumentar al máximo el tiempo que pasan en la plataforma e impulsar la interacción con los contenidos de las personas usuarias y —de mayor importancia para potenciar su negocio— la publicidad. En el informe *Gigantes de la vigilancia*, publicado por Amnistía Internacional en 2019, se documentó cómo Facebook (actualmente Meta) y Google (propiedad de Alphabet) habían optimizado el uso de los análisis de macrodatos y el diseño adictivo con la eficacia suficiente como para dominar el mercado publicitario mundial.¹¹⁸ Se estima que, en 2021, entre ambas empresas sumaban un 43% de todo el gasto publicitario mundial online y offline.¹¹⁹ TikTok entró en este mercado con la determinación de atraer rápidamente a cientos de millones de jóvenes y usurpar a otras redes la cuota de tiempo y atención de sus personas usuarias.¹²⁰ Sus materiales de marketing ponen de manifiesto que la empresa no se ve a sí misma únicamente como en competencia con otras redes sociales, sino que también aspira a ocupar el espacio de otras actividades cotidianas, al publicitar que “el 41% de los miembros de la Generación Z que usan TikTok a nivel mundial afirman haber dedicado menos tiempo a escuchar podcasts desde que empezaron a utilizar TikTok”.¹²¹

Los efectos del uso extensivo de TikTok que aportaron a la investigación los niños, las niñas y las personas jóvenes que entrevistó Amnistía Internacional no se limitaban a dedicar menos tiempo a otros medios de comunicación. Varias de las personas jóvenes entrevistadas consideraban que el uso de TikTok las llevaba a no presentar trabajos escolares y universitarios, a pasar menos tiempo con sus amistades en el mundo real y —lo más perjudicial para la salud— a explorar sus feeds a altas horas de la madrugada en vez de dormir las horas suficientes.¹²²

Por ejemplo, “Mary”, alumna de una escuela femenina en Kenia occidental, afirmó:

“Cuando me despierto, lo primero que hago es mirar el teléfono, tan sólo pienso en mi teléfono”.¹²³

“Amy”, otra adolescente del grupo-muestra, indicó:

118. Amnistía Internacional, *Gigantes de la vigilancia* (antes citado).

119. Ebiquty, “Google, Meta and Amazon are on track to absorb more than 50% of all ad money in 2022”, 4 de febrero de 2022, <https://ebiquity.com/news-insights/press/google-meta-and-amazon-are-on-track-to-absorb-more-than-50-of-all-ad-money-in-2022/>

120. Yahoo Finance, “Forget Facebook, Snapchat, Twitter: TikTok Is The Breakout COVID-19 Social Media Platform”, 24 de abril de 2020, <https://finance.yahoo.com/news/forget-facebook-snapchat-twitter-tiktok-210717993.html>; Business Insider, “Nearly half of Gen Z is using TikTok and Instagram for search instead of Google, according to Google’s own data”, 13 de julio de 2022, <https://www.businessinsider.com/nearly-half-genz-use-tiktok-instagram-over-google-search-2022-7>; TikTok, “TikTok Marketing Science Global Time Well Spent Study (Global Results) conducted by Kantar March 2021”, <https://tiktok.com/business/es/insights> (consultado el 1 de junio de 2023).

121. TikTok, “TikTok Marketing Science Global Time Well Spent Study (Global Results) conducted by Kantar March 2021” (antes citado).

122. Mencionado en distintos debates con grupos-muestra en Kisumu (Kenia), los días 13 y 14 de marzo de 2023, así como en debates con grupos-muestra en escuelas de Mombasa, los días 15, 16 y 17 de marzo de 2023.

123. Debate con grupos-muestra, Kisumu (Kenia), 14 de marzo de 2023.

“Si uso TikTok, mi intención suele ser verlo durante cinco minutos, intento ver dos (vídeos), pero me deja ver y seguir visualizando toda la noche. Quiero pasar cinco minutos, pero los vídeos son tan interesantes que acabo [viéndolos durante] al menos seis horas”.

Al preguntar qué aplicación de las redes sociales era más difícil dejar de usar, hubo un consenso casi unánime entre los 29 participantes del grupo con respecto a considerar que era TikTok.¹²⁴ “Joyce”, una joven de Filipinas, explicó:

“La borré [la aplicación de TikTok] durante un tiempo, pero fue porque estaba muy enganchada a ella... Pasaba muchísimas horas en TikTok, simplemente navegando por los vídeos, debido a que no puedes evitar preguntarte qué saldrá a continuación al desplazar la pantalla”.¹²⁵

Estos y otros testimonios de niños, niñas y jóvenes fueron corroborados por especialistas en psicología de la adolescencia a quienes se consultó en el marco de la investigación, que afirmaron haber observado una tendencia a que TikTok agravara problemas de conductas adictivas existentes con respecto al uso de las redes sociales en niños, niñas y jóvenes. “Muchas personas me dicen que necesitan tomarse un descanso de las redes sociales, que se están enganchando a ellas. No quieren más que pasarse el día sentadas pasando pantallas”, afirmó Zeyna Awan, psicóloga clínica en Nairobi.¹²⁶ Kevin Gachee, también psicólogo en Nairobi, observó un cambio generacional en el sentido de que las personas jóvenes de la Generación Z con las que trabaja parecen tener dificultades relativas a un uso excesivo y poco saludable de TikTok, mientras que las personas adultas de la generación anterior son más propensas a referir problemas relativos a una comparación social malsana, al parecer propiciada por el uso de Instagram.¹²⁷

Las pruebas disponibles a partir de estudios de mercado se basan mayoritariamente en datos sobre el uso de las redes sociales por parte de niños y niñas en Estados Unidos y Europa. Un estudio de 2021 reveló que TikTok era la plataforma en la que niños y niñas pasaban más tiempo en Estados Unidos, Reino Unido y España; los niños y las niñas de Estados Unidos dedicaban una media de 99 minutos al día sólo a TikTok.¹²⁸ Una reciente encuesta representativa de adolescentes en Estados Unidos halló que el 45% de las niñas que usaban TikTok afirmaban “sentirse ‘enganchadas’ a la plataforma o usarla más de lo previsto, al menos con carácter semanal”.¹²⁹

Nuestra investigación cualitativa en Kenia y Filipinas sugiere tendencias similares. De entre todas las plataformas de redes sociales, las personas jóvenes participantes en la investigación describieron TikTok en concreto como “adictiva” y “una pérdida de tiempo”. Al menos 25 participantes en 10 de los 11 grupos-muestra en Kenia efectuaron tales comentarios y muchas otras personas asintieron con la cabeza para expresar su acuerdo. Las personas jóvenes usuarias de TikTok describieron constantes de uso nocturno extremo¹³⁰ y una minoría considerable de los niños y las niñas en edad escolar que participaron en la investigación en Kenia informaron de que, debido a que comparten los teléfonos con sus padres o hermanos mayores, sólo tienen acceso a éstos a última hora de la tarde o por la noche y, por lo tanto, a menudo utilizan las redes sociales sin supervisión parental hasta altas horas de la noche

124. Debate con grupos-muestra, Kisumu (Kenia), 14 de marzo de 2023.

125. Entrevista online con un participante residente en Manila, 22 de mayo de 2023.

126. Entrevista a especialista, efectuada online el 9 de marzo de 2023.

127. Entrevista a especialista, efectuada en persona en Nairobi el 9 de marzo de 2023.

128. Custodio, *Annual Data Report 2021: Living and Learning in a Digital World*, Chapter 2, Social Media, <https://www.custodio.com/en/social-media-custodio-annual-data-report-2021> (consultado el 20 de julio de 2023).

129. Jacqueline Nesi, Supreet Mann & Michael B. Robb, *Teens and mental health: How girls really feel about social media*, 2023, https://www.common-sense-media.org/sites/default/files/research/report/how-girls-really-feel-about-social-media-research-report_web_final_2.pdf

130. Debates con grupos-muestra llevados a cabo en Kisumu (Kenia) el 13 y 14 de marzo de 2023, en Mombasa (Kenia) el 16 y 17 de marzo, y en Machakos (Kenia) el 22 de marzo de 2023.

o a primera hora de la mañana, por lo que pierden horas de sueño antes de la jornada escolar.¹³¹ Este hallazgo puede apuntar a riesgos adicionales para la salud de niños, niñas y adolescentes económicamente desfavorecidos.

Las publicaciones recientes sobre el uso de las redes sociales y la salud mental en adolescentes sugiere además una constante en la que los niños y las niñas con problemas de salud mental son más susceptibles al uso excesivo de las redes sociales. Un informe de 2021 del Servicio Nacional de Salud (NHS) británico reveló que “los niños y las niñas de 11 a 16 años con una probable discapacidad psicosocial eran particularmente propensos a pasar más tiempo del que pretendían en las redes sociales; informaron de ello casi dos tercios (63,8%), en comparación con menos de la mitad de los que probablemente no padecieran ninguna discapacidad psicosocial (45,5%)”.¹³² Una revisión sistemática de estudios sobre el uso y el impacto de las redes sociales por parte de adolescentes con problemas de salud mental preexistentes también descubrió que aquéllos con un diagnóstico de depresión y aquéllos con “síntomas autodeclarados de ansiedad, depresión, comportamiento autolesivo” tenían más probabilidades de tener problemas para regular su uso de Internet, incluido el de las redes sociales.¹³³

5.1 ESTRATEGIAS DE TIKTOK PARA LA INTERACCIÓN

La capacidad de las personas para regular el uso de las redes sociales y los factores contextuales pueden variar, pero al igual que otras plataformas de redes sociales, TikTok ha tomado decisiones de diseño destinadas a maximizar el tiempo que pasan en la plataforma quienes la usan.¹³⁴

Además de incorporar elementos de diseño que otras plataformas de redes sociales han aplicado satisfactoriamente para mantener los ojos de las personas usuarias fijos en la pantalla¹³⁵ —como la posibilidad de votar “me gusta” en las publicaciones, notificaciones frecuentes y un feed de contenidos sin fin—, TikTok se diferencia por su feed “Para ti”, página altamente personalizada que puede desplazarse indefinidamente, compuesta por contenidos recomendados algorítmicamente y seleccionados en función de lo que el sistema deduce que interesa a la persona que lo usa, a los pocos minutos de abrir el feed por primera vez.¹³⁶

Mientras que anteriormente, para predecir intereses, las plataformas de redes sociales se basaban en las conexiones sociales que las personas usuarias declaraban de forma activa, las recomendaciones de contenidos de TikTok se basan en gran medida en el análisis de los vídeos con los que interactúa una persona usuaria de entre una selección inicial de vídeos populares. A continuación, esta interacción se utiliza para acotar progresivamente los intereses de la persona usuaria mediante

131. Al menos nueve personas jóvenes kenianas participantes mencionaron haber perdido horas de sueño a causa de TikTok, incluso durante las noches en periodos escolares, bien en las conversaciones acerca de cuándo y cómo accedían a las redes sociales, bien al pedirles que reflexionaran acerca de la plataforma de redes sociales a la que dedicaban más tiempo. El número de niños, niñas y jóvenes que comparten teléfonos variaba enormemente de un grupo a otro, desde menos del 10% hasta cerca del 50% de los alumnos de un grupo en particular.

132. NHS, *Mental Health of Children and Young People in England 2021 - Wave 2 follow up to 2017 survey*, 30 de septiembre de 2021, <https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2021-follow-up-to-the-2017-survey>

133. Katarzyna Kostyrka-Allchorne, Mariya Stoilova and others, “Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and nonsuicidal self-injury conditions – a systematic review”, febrero de 2023, *Child and Adolescent Mental Health*, volumen 28, número 1, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10108198/>

134. Sophia Petrillo, “What makes TikTok so addictive?: An analysis of the mechanisms underlying the world’s latest social media craze”, 13 de diciembre de 2021, *Brown Undergraduate Journal of Public Health*, edición 2021-2022, <https://sites.brown.edu/publichealthjournal/2021/12/13/tiktok/>

135. ABC News, “Book excerpt: Jaron Lanier’s ‘Ten Arguments for Deleting Your Social Media Accounts Right Now’”, 19 de junio de 2018, <https://abcnews.go.com/Technology/book-excerpt-jaron-laniers-ten-arguments-deleting-social/story?id=56009512>; BBC, “Facebook founding president sounds alarm”, 9 de noviembre de 2017, <https://www.bbc.co.uk/news/technology-41936791>

136. TikTok, “How TikTok recommends videos #ForYou”, 18 de junio de 2020, <https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you>

recomendaciones de vídeos, que el sistema asocia con el contenido con el que se ha interactuado previamente.¹³⁷ El resultado es un feed mucho más personalizado, compuesto por vídeos cortos, que presuntamente ofrecerán el máximo entretenimiento a la persona usuaria.¹³⁸

Al combinar estos elementos, TikTok recurre a lo que en psicología se describe como “el patrón de refuerzo intermitente al ganar o perder en una máquina recreativa... que se sustenta en el condicionamiento clásico y en los procesos de aprendizaje basados en el refuerzo para facilitar la formación de ciclos de costumbre y fomentar el uso adictivo”.¹³⁹ En resumen, TikTok ha maximizado las cualidades adictivas de diferentes opciones de diseño, algo que —como advirtió el presidente fundador de Facebook ya en 2017— podría tener consecuencias negativas no deseadas, especialmente para la salud y el desarrollo de niños y niñas.¹⁴⁰ Las personas jóvenes entrevistadas que comunicaron patrones de uso compulsivo de TikTok a menudo describieron la conducta resultante como un “desplazamiento de pantalla irreflexivo”,¹⁴¹ durante el cual las personas usuarias “en realidad no conectan con la publicación ni sienten nada al respecto porque no les da tiempo a procesarlas, al limitarse a pasar pantallas”.¹⁴²

Las personas adolescentes que participaron en la investigación comentaron también la supuesta contribución de la función “me gusta” —presente tanto en TikTok como en otras plataformas de redes sociales— a intensificar las inseguridades y a mantener la mirada fija en la pantalla.¹⁴³ Explicaron que se sentían obligadas a consultar su cuenta una y otra vez para ver cómo se posicionaban sus publicaciones en comparación con las demás. Niños y niñas en edad escolar de las zonas rurales de Kenia declararon haber utilizado incluso herramientas de edición de fotos y vídeos para crear publicaciones falsas acerca de experiencias que su familia no podía permitirse económicamente, con el fin de ganar seguidores.¹⁴⁴

Las personas de mayor edad que participaron en la investigación, población universitaria incluida, hicieron referencia al supuesto papel que desempeñaban sus perfiles en las redes sociales a la hora de definir su estatus social y su valor personal: “Cuando llegas a la universidad, después de graduarte, eres tu cuenta en las redes sociales”, explicó Nikki, 24 años, de Manila. Victor, un joven adulto que participó en un grupo-muestra en Mombasa se hizo eco de la idea de que las redes sociales requieren una autorrepresentación esmerada, al afirmar: “el yo que veis en las redes sociales no es el yo que está sentado aquí hoy”.¹⁴⁵

Numerosos estudios psicológicos y —como revelaron los “Papeles de Facebook”— la propia investigación de la empresa¹⁴⁶ han documentado el efecto tóxico que tiene la comparación social propiciada por las redes sociales sobre la salud mental en la adolescencia.¹⁴⁷ Instagram llevó a cabo durante dos años un experimento en el que ocultaba el número de “me gusta” que recibía una

137. Alex Hern, “How TikTok’s algorithm made it a success: ‘It pushes the boundaries’” (antes citado).

138. Catherine Wang, “Why TikTok made its user so obsessive? The AI Algorithm that got you hooked”, 7 de junio de 2020, <https://towardsdatascience.com/why-tiktok-made-its-user-so-obsessive-the-ai-algorithm-that-got-you-hooked-7895bb1ab423>

139. Sophia Petrillo, “What makes TikTok so addictive?: An analysis of the mechanisms underlying the world’s latest social media craze”, diciembre de 2021, *Brown Undergraduate Journal of Public Health*, número 2021-2022, (antes citado).

140. El presidente fundador de Facebook, Sean Parker, declaró en 2017: “[Facebook] probablemente interfiera con la productividad de formas extrañas. Sólo Dios sabe qué le está haciendo al cerebro de nuestros hijos”. Véase BBC, “Facebook founding president sounds alarm” (antes citado).

141. Manuel (seudónimo), 23 años, estudiante universitario residente en Luzón, entrevista a distancia el 4 de mayo de 2023.

142. Nikki (seudónimo), 24 años, Manila, entrevistada el 8 de mayo de 2023.

143. Debates con grupos-muestra del alumnado de secundaria en el condado de Kisumu el 14 de marzo de 2023, en Mombasa el 17 de marzo y en el condado de Machakos el 22 de marzo de 2023.

144. Adolescentes participantes en un debate con grupos-muestra en una escuela secundaria del condado de Machakos, 22 de marzo de 2023.

145. Debate con grupos-muestra de estudiantes de universidad y jóvenes activistas en Mombasa, 17 de marzo de 2023.

146. Wall Street Journal, “Facebook knows Instagram is toxic for teen girls company documents show” (antes citado).

147. Elia Abi-Jaoude, Karlene Treurnicht Naylor et al., “Smartphones, social media use and youth mental health”, 10 de febrero de 2020, *Canadian Medical Association Journal*, volumen 192, número 6, <https://www.cmaj.ca/content/192/6/E136>; Para consultar un resumen de otras investigaciones sobre la comparación social impulsada por las redes sociales y su impacto en la salud mental, véase Cirujano general de Estados Unidos, *Social Media and Youth Mental Health* (antes citado).

publicación a todas las personas usuarias, salvo a la que había creado el contenido. Terminó por desechar la idea, optando en su lugar por permitir a las personas usuarias elegir por sí mismas si se mostraban los “me gusta” o no, con la explicación de que el experimento había recibido una respuesta mixta.¹⁴⁸ Sin embargo, especialistas como el psicólogo Kevin Gachee, radicado en Nairobi, tienen claro el valor de restar importancia a la popularidad y la comparación social para fomentar la autoestima de las personas jóvenes y favorecer formas de interacción más sanas: “La forma en que diseñan estas aplicaciones está pensada para que interactúes con ellas continuamente. Alabé a Instagram cuando dejó [temporalmente] de mostrar el número de ‘me gusta’ porque ayudó significativamente a que la gente dijera: No siento la presión de seguir publicando continuamente porque, sabes, una amistad ha conseguido 100 ‘me gusta’ y yo también quiero llegar a 100 ‘me gusta’.”¹⁴⁹

Más preocupante aún en términos de impacto sobre la salud mental es el efecto de arrastre que ejercen las redes sociales hacia una conducta que busca la afirmación en personas jóvenes que se autolesionan. El Dr. Marc Reyes, psicólogo radicado en Manila, explicó sus motivos de preocupación a Amnistía Internacional: “La mayoría de las personas que se enfrentan a problemas de salud mental se sienten invisibles, pero en las redes sociales tienen visibilidad. Así pues, los ‘me gusta’ se vuelven muy perjudiciales; a mucha gente le gusta mi publicación [sobre autolesiones], así que está bien”, a lo que añadió: “al publicarse, compartirse y volverse a compartir [ese contenido], podría desencadenar una epidemia de suicidios o autolesiones”.¹⁵⁰

A pesar de estos motivos de preocupación, TikTok —al igual que Instagram— sigue mostrando los “me gusta” por defecto.

5.2 CÓMO PUEDE AFECTAR EL DISEÑO ADICTIVO DE LAS REDES SOCIALES A LA SALUD DE LAS PERSONAS JÓVENES

Es innegable que las tecnologías digitales han desempeñado un papel positivo en la reducción de las barreras de acceso a la información y los servicios sanitarios, por ejemplo, al permitir a los terapeutas hablar con los pacientes a distancia durante el confinamiento por la COVID-19, al mantener a las personas jóvenes en conexión con sus amistades y redes de apoyo, y al permitir a grupos marginados —como adolescentes LGBTI— explorar sus identidades y expresarse.¹⁵¹ Sin embargo, se reconoce cada vez más que, en ausencia de una normativa estatal eficaz, las plataformas de redes sociales también exponen a niños, niñas y jóvenes a riesgos sistémicos.¹⁵²

Entre las respuestas a las preguntas de la encuesta de evaluación del alcance llevada a cabo para esta investigación, se elogiaron la diversidad de ideas y las oportunidades que ofrecen las redes sociales, pero muchas personas encuestadas expresaron como motivo de preocupación el impacto de lo que describieron como redes sociales “adictivas” para la salud mental. Las personas jóvenes declararon sentirse “ansiosas” y “acomplejadas” por las “imágenes [corporales] poco realistas” mostradas en sus feeds. Una encuestada de Kenia afirmó que creía que las recomendaciones de contenidos algorítmicas de las plataformas de redes sociales (a las que la encuestada se refería como

148. Instagram, “Giving people more control on Instagram and Facebook”, 26 de mayo de 2021, <https://about.instagram.com/blog/announcements/giving-people-more-control>

149. Entrevista efectuada en Nairobi el 9 de marzo de 2023.

150. Entrevista a distancia efectuada el 4 de mayo de 2023.

151. Benjamin Hanckel and Shiva Chandra, “How young LGBTQIA+ people used social media to thrive during COVID lockdowns”, The Conversation, 15 de marzo de 2021, <https://theconversation.com/how-young-lgbtqia-people-used-social-media-to-thrive-during-covid-lockdowns-156130>

152. Véase, por ejemplo, el Informe de la Relatora Especial sobre el derecho de toda persona al disfrute del más alto nivel posible de salud física y mental, *Innovación digital, tecnologías y derecho a la salud*, 21 de abril de 2023, doc. ONU A/HRC/53/65; Comité de los Derechos del Niño, Observación general núm. 25 (antes citado).

“el algoritmo”) “detectaban” sus problemas de salud mental existentes. Otras personas dijeron haber visto publicaciones que consideraban que intensificaban su anorexia autodeclarada, calificadas como “provocadoras”, que habían “tenido problemas” contra las autolesiones o que las habían “asustado”.¹⁵³

En un aviso de 2023 relativa al impacto de las redes sociales sobre la salud mental de las personas adolescentes, el Cirujano general de Estados Unidos pronunció unas claras palabras de advertencia sobre los efectos del diseño adictivo de las redes sociales:

“El uso excesivo y problemático de las redes sociales, tal y como el uso compulsivo o incontrolable, se ha relacionado con problemas de sueño, problemas de atención y sentimientos de exclusión entre adolescentes... En estudios menores se ha demostrado que las personas que hacen un uso frecuente y problemático de las redes sociales pueden sufrir cambios en la estructura cerebral similares a los observados en personas adictas al consumo de sustancias o al juego”.¹⁵⁴

El aviso señalaba también que dormir mal no sólo afecta a la concentración y al rendimiento escolar, sino que también “se ha relacionado con alteraciones en el desarrollo neurológico del cerebro adolescente, síntomas depresivos y pensamientos y conductas suicidas”.

La investigación acerca del modo en que el uso excesivo de las redes sociales puede alterar las estructuras cerebrales está aún en fase inicial, pero estudios menores señalan varias similitudes entre los cambios cerebrales observados en casos de adicciones a sustancias (por ejemplo, cocaína) y de uso adictivo de las redes sociales.¹⁵⁵ Ambas situaciones desencadenan la liberación de altos niveles de dopamina, lo que provoca un “subidón” emocional temporal el cual, a su vez, hace que la persona usuaria decaiga en cuanto deja de usar la plataforma.¹⁵⁶ Estos hallazgos apuntan a la necesidad de una normativa gubernamental, en consonancia con la regulación de otras sustancias adictivas, a fin de observar el derecho a la salud de niños, niñas y jóvenes.

TikTok debe ser claramente consciente del potencial de la plataforma para dañar la salud mental y física de niños, niñas y jóvenes como resultado de patrones de uso compulsivos o nocivos. Sobre la base de las pruebas de investigación antes citadas, TikTok se arriesga a exponer a niños, niñas y jóvenes a problemas de sueño y de atención, y puede propiciar cambios en la estructura cerebral similares a los observados en personas adictas a las drogas.

Aun así, su respuesta no ha sido lo suficientemente profunda. En marzo de 2023, TikTok anunció la introducción de una nueva herramienta para administrar el tiempo frente a la pantalla, que requiere que las personas menores de 18 años prolonguen activamente su tiempo en la aplicación una vez alcanzado el límite de 60 minutos diarios.¹⁵⁷ Aún no se ha evaluado de forma independiente la eficacia de la herramienta, pero dado que traslada la responsabilidad de limitar el uso de TikTok a las personas adolescentes, que se autodescriben en grandes cifras como “adictas”, es probable que sea limitada.¹⁵⁸ El aviso de TikTok es sólo una sugerencia para que se respete el límite de tiempo; un mensaje que puede descartarse fácilmente y que no incluye ninguna advertencia sanitaria.

153. Respuestas al cuestionario online de Amnistía Internacional, octubre-noviembre de 2022.

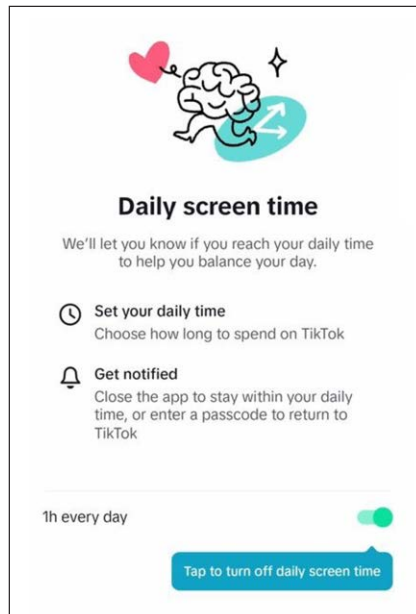
154. Cirujano general de Estados Unidos, *Social Media and Youth Mental Health* (citado anteriormente).

155. Qinghua He, Ofir Turel et al., “Brain anatomy alterations associated with Social Networking Site (SNS) addiction”, marzo de 2017, *Scientific reports*, 7, 45064. <https://www.nature.com/articles/srep45064>

156. Scope, “Addictive potential of social media, explained”, 29 octubre de 2021, <https://scopeblog.stanford.edu/2021/10/29/addictive-potential-of-social-media-explained/>

157. TikTok, “TikTok anuncia nuevas funcionalidades para contribuir al bienestar de los adolescentes y las familias”, 1 de marzo de 2023, <https://newsroom.tiktok.com/es-es/tiktok-anuncia-nuevas-funcionalidades-para-contribuir-al-bienestar-de-los-adolescentes-y-las-familias>

158. Cirujano general de Estados Unidos, *Social Media and Youth Mental Health* (antes citado), p. 9; Jacqueline Nesi, Supreet Mann & Michael B. Robb, *Teens and mental health: How girls really feel about social media*, 2023 (antes citado), p. 6.



Un solo clic redirige a las personas usuarias jóvenes desde la notificación del límite diario relativa al tiempo de pantalla hasta una página donde pueden desactivar permanentemente esta función.

Además de las limitaciones de la medida, el cambio sólo se aplica a las personas usuarias que el sistema identifica como niños o niñas, aunque se ha cuestionado la eficacia de la verificación de edad de TikTok.¹⁵⁹ De hecho, Ofcom, regulador de medios de comunicación de Reino Unido, ha descubierto que el 16% de los niños y las niñas de tres y cuatro años tienen acceso a TikTok en el país.¹⁶⁰ En abril de 2023, la Autoridad de Protección de Datos de Reino Unido (ICO) multó a TikTok con 12,7 millones de libras por haber permitido el uso de la plataforma a menores de 13 años en 2020.¹⁶¹ Las entrevistas con niños, niñas y jóvenes efectuadas para este informe sugieren que muchas personas menores de 18 años mienten acerca de su edad para obtener acceso a la plataforma y eludir las medidas de protección de la infancia, o que acceden a TikTok a través de cuentas de adultos. A estos niños y niñas no se les muestra la limitación del tiempo de pantalla.

Podría decirse que unos procesos de verificación de la edad más eficaces garantizarían que las personas menores de 13 años no accedieran a la plataforma y que las mayores contaran con medidas de protección. Sin embargo, las herramientas de verificación de la edad actualmente disponibles, que ofrecen un grado de certeza mayor que una simple autodeclaración de la persona usuaria, sin duda requieren una recopilación de datos más amplia, lo cual plantea problemas añadidos para el derecho a la privacidad de los niños y las niñas. Tampoco se abordan los problemas estructurales del modelo de negocio basado en la vigilancia, analizados en este informe.

Algunas personas entrevistadas afirmaron haber tratado de limitar su uso de TikTok y otras plataformas de redes sociales, incluso mediante aplicaciones de control del tiempo de pantalla, pero que no habían sido eficaces para frenar la tentación. Nikki, mujer de 24 años residente en Filipinas, explicó: “He probado a configurar notificaciones con alarmas y la verdad es que no funciona. Cuando sólo es una notificación, siempre quiero cinco minutos más en la aplicación”. Otras personas explicaron cómo habían sorteado los intentos de sus progenitores de restringir el uso de las redes sociales. Por ejemplo, 13 de 15 alumnos y alumnas de una escuela secundaria de Mombasa (Kenia) admitieron haber engañado a sus progenitores al fingir que hacían tareas escolares en el teléfono cuando en realidad lo utilizaban para conectarse a cuentas de redes sociales.¹⁶²

Un tema recurrente en los debates con grupos-muestra con niños, niñas y jóvenes fue la idea de que es una cuestión de autodisciplina y de uso responsable el hecho de que las redes sociales desempeñen un papel constructivo o destructivo en la propia vida. Debe respetarse la capacidad

159. New York Times, “A third of TikTok’s U. S. users may be 14 or under, raising safety questions”, 14 de agosto de 2020, <https://www.nytimes.com/2020/08/14/technology/tiktok-underage-users-ftc.html>; BBC, “TikTok fined £12.7m for misusing children’s data”, 4 de abril de 2023, <https://bbc.co.uk/news/uk-65175902>

160. Guardian, “TikTok being used by 16% of British toddlers, Ofcom finds”, 29 de marzo de 2022, <https://theguardian.com/technology/2022/mar/29/tiktok-being-used-by-16-of-british-toddlers-ofcom-finds>

161. Autoridad de Protección de Datos de Reino Unido, “ICO fines TikTok £12.7 million for misusing children’s data”, 4 de abril de 2023, <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/04/ico-fines-tiktok-127-million-for-misusing-children-s-data/>

162. Debate con grupos-muestra, Mombasa (Kenia), 17 de marzo de 2023.

de acción de niños, niñas y jóvenes, y se requieren medidas de concienciación y educación para fundamentar sus decisiones y equiparlos mejor para que puedan moverse por los entornos de Internet e identificar riesgos y daños. Sin embargo, es muy preocupante que se haya normalizado la falta de un entorno virtual seguro, lo que puede repercutir en la observación de su derecho a la salud. No debe esperarse que niños, niñas y jóvenes eviten los daños y mitiguen los riesgos de una plataforma que está diseñada para hacer que interactúen con ella a toda costa. La responsabilidad de crear un entorno seguro y libre de daños recae en las empresas de las redes sociales que han diseñado estas plataformas, las administran y obtienen beneficios de ellas.

En el capítulo 7 se analizan con más detalle estas responsabilidades empresariales y su incumplimiento por parte de TikTok.

6. ESPIRALES DESCENDENTES

“Cuando me sentía decaído, creo que el 80% [del contenido] tenía que ver con la salud mental. Es una madriguera en la que te adentras a partir de un solo vídeo. Si un vídeo logra captar tu atención, aunque no te guste, te lo presentan cuando vuelves a abrir TikTok y, como te resulta familiar, lo vuelves a ver. Al verlo de nuevo, su frecuencia en tu feed aumenta exponencialmente”.

Luis, estudiante universitario de 21 años de Manila¹⁶³

6.1 EL FEED ‘PARA TI’ DE TIKTOK LLEVA AL EXTREMO UN SISTEMA INHERENTEMENTE PELIGROSO

El elemento central de la estrategia de TikTok para la interacción es su página “Para ti”, el feed por defecto a través del cual muestra a las personas usuarias un flujo personalizado de vídeos cortos a ritmo rápido. Más allá de su naturaleza adictiva, la función se sitúa también en el centro de los motivos de preocupación acerca del papel que desempeña la plataforma en la exposición de niños, niñas y jóvenes a contenidos potencialmente nocivos.

Los primeros debates sobre la regulación de las plataformas se centraron en cómo exigir a las empresas que eliminen rápidamente las publicaciones ilegales y abusivas, en medio de cifras ingentes de contenidos generados por las personas usuarias. Con las avalanchas de desinformación y de publicaciones incitadoras que recorren las plataformas en contextos políticos polarizados y en situaciones de conflicto, el escrutinio público se ha desplazado hacia las formas en que las plataformas de redes sociales *propagan* activamente los contenidos.¹⁶⁴ Como explicó un empleado anónimo de Facebook:

163. Luis (seudónimo), estudiante universitario de 21 años de Manila, fue entrevistado en persona el 6 de mayo de 2023.

164. Amnistía Internacional, *Myanmar: La atrocidad social* (antes citado); Relatora Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, *La desinformación y la libertad de opinión y de expresión*, 13 de abril de 2021, doc. ONU A/HRC/47/25; *La desinformación y la libertad de opinión y de expresión durante los conflictos armados*, 12 de agosto de 2022, doc. ONU A/77/288; Election Integrity Partnership, *The Long Fuse: Misinformation and the 2020 Election*, 2 de marzo de 2021, <https://www.atlanticcouncil.org/in-depth-research-reports/the-long-fuse-eip-report-read>

“También tenemos datos fehacientes de que las mecánicas de nuestro producto central, como la viralidad, las recomendaciones y la optimización para la interacción, son una parte significativa de la razón por la que estos tipos de lenguaje florecen en la plataforma... La mecánica de nuestra plataforma no es neutra”.¹⁶⁵

Expertos de la ONU, Amnistía Internacional y muchas otras organizaciones de la sociedad civil han documentado el modo en que los principales sistemas de recomendación de contenidos en las redes sociales han contribuido a propagar y amplificar “propaganda, contenidos extremistas y desinformación” en recientes conflictos armados y situaciones de violencia política.¹⁶⁶

Aunque este informe se centra en un tipo diferente de contenido amplificado (contenido depresivo y relativo a autolesiones) y en un conjunto diferente de riesgos (posibles daños a niños, niñas y jóvenes con problemas de salud mental preexistentes que se exponen a dichos contenidos), el origen de estos riesgos es básicamente el mismo: el sistema de recomendación algorítmico de la plataforma de una red social, que está orientado a crear un feed personalizado de contenido viral que lleve a la persona usuaria a interactuar con la aplicación durante el mayor tiempo posible.

A pesar de las pruebas de que la mecánica central de las recomendaciones personalizadas, la viralidad y la “optimización” para maximizar la interacción de las personas usuarias no crean sólo riesgos de abusos y violaciones de los derechos humanos, sino que han contribuido a daños concretos, varias plataformas de redes sociales como TikTok, Instagram y Facebook siguen empleándola para alcanzar su meta. De hecho, el éxito de TikTok al respecto ha llevado a su competidor Meta (empresa matriz de Instagram y Facebook) a emular la confianza de TikTok en las recomendaciones personalizadas dentro de su estrategia sobre interacción y a invertir aún más en perfeccionar sus sistemas algorítmicos.¹⁶⁷

Sin embargo, hasta la fecha los gobiernos no han aprobado legislación para detener la continua “optimización” de las plataformas hacia una personalización basada en inferencias, viralidad y potenciación máxima de la interacción de las personas usuarias. TikTok comercializa abiertamente su capacidad para configurar un feed cada vez más personalizado mediante el análisis de la interacción de las personas usuarias con la plataforma, perfilando a estas personas con eficacia desde el primer momento. En sus propias palabras:

“Desde el momento en que inicias tu recorrido por TikTok, tu primer conjunto de ‘me gusta’, comentarios y reproducciones repetidas dará lugar a recomendaciones, dado que el sistema comenzará a aprender sobre tus preferencias de vídeos... Con el tiempo, estas recomendaciones deberían ser cada vez más personalizadas. Seguir cuentas nuevas, explorar *hashtags*, sonidos, efectos y temas de actualidad son todos ellos actos que contribuyen a refinar tu feed”.¹⁶⁸

165. The Facebook Papers, “What is Collateral Damage?”, 12 de agosto de 2019, citado en Amnistía Internacional, *Myanmar: La atrocidad social* (antes citado).

166. Relatora Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, *La desinformación y la libertad de opinión y de expresión durante los conflictos armados*, 12 de agosto de 2022, doc. ONU A/77/288; Amnistía Internacional, *Myanmar: La atrocidad social* (antes citado); The Atlantic, “Who’s behind #IStandWithPutin?”, 5 de abril de 2022, <https://www.theatlantic.com/ideas/archive/2022/04/russian-propaganda-zelensky-information-war/629475/>; Hannah Porter, “A conversation on fighting disinformation in Yemen”, Yemen Policy Center, marzo de 2022, <https://www.yemenpolicy.org/a-conversation-on-fighting-disinformation-in-yemen/>; Paul Barrett y Justin Hendrix, *A Platform ‘Weaponized’: How YouTube Spreads Harmful Content – And What Can Be Done About It*, NYU Stern Center for Business and Human Rights, junio de 2022, <https://www.stern.nyu.edu/experience-stern/faculty-research/platform-weaponized-how-youtube-spreads-harmful-content-and-what-can-be-done-about-it>; Sara Brown, “In Russia-Ukraine War, Social Media Stokes Ingenuity, Disinformation”, MIT Sloan School of Management, 6 de abril de 2022, <https://mitsloan.mit.edu/ideas-made-to-matter/russia-ukraine-war-social-media-stokes-ingenuity-disinformation>

167. TechCrunch, “Meta expects recommendation models ‘orders of magnitude’ bigger than GPT-4. Why?”, 29 de junio de 2023, <https://techcrunch.com/2023/06/29/metabehavior-analysis-model-is-orders-of-magnitude-bigger-than-gpt-4-why/?guccounter=1>; Platformer, “Meta’s Nick Clegg on how AI is reshaping the feed”, 30 de junio de 2023, <https://www.platformer.news/p/metabehavior-analysis-model-is-orders-of-magnitude-bigger-than-gpt-4-why/?guccounter=1>

168. TikTok, “¿Qué es la página ‘Para ti?’”, <https://www.tiktok.com/creators/creator-portal/es-latam/tiktok-creation-essentials-es-latam/que-es-la-pagina-para-ti/> (consultado en junio de 2023).

En informes anteriores de Amnistía Internacional se ha detallado por qué la recopilación masiva de datos privados que permite este tipo de “optimización” basada en perfilaciones íntimas —en lugar de en señales manifestadas deliberadamente— constituye una forma corporativa de vigilancia masiva que es incompatible con el derecho a la privacidad y se arriesga a causar o contribuir a causar daños en cadena que impliquen aún más abusos contra los derechos humanos.¹⁶⁹ Un documento filtrado de Facebook (ahora Meta) reveló en 2017 que la empresa afirmaba ser capaz de identificar “momentos en los que las personas jóvenes necesitan un refuerzo de confianza” y, al parecer, estaba de hecho dispuesta a vender esos momentos de tristeza o ansiedad como oportunidades de venta a sus anunciantes.¹⁷⁰ Más recientemente, TikTok ha publicitado su capacidad para identificar cuándo las personas usuarias tienen un “vínculo emocional” con el contenido publicitario.¹⁷¹ Los objetivos declarados de Meta y TikTok representan un ataque a los derechos de niños, niñas y jóvenes a la privacidad y a la libertad de pensamiento. Son la prueba de la intención de ambas empresas de analizar e inferir los pensamientos y emociones privados de las personas usuarias y explotarlos con fines comerciales.

Las personas adolescentes no forman simplemente parte de las estrategias para aumentar las cifras de personas usuarias, sino que son especialmente vulnerables a la manipulación. Los investigadores del campo de la neurociencia siguen tratando de comprender mejor cómo afectan el entorno social de la adolescencia y la tecnología moderna al cerebro en esta etapa especialmente maleable del desarrollo. La “mayor sensibilidad a las recompensas” de las personas adolescentes parece tanto promover comportamientos positivos como hacerlas más propensas a participar en conductas de riesgo.¹⁷² El pensamiento racional, en contraste con el procesamiento emocional, no se desarrolla por completo hasta bien entrada la veintena.¹⁷³ Por lo tanto, las personas adolescentes son mucho más propensas que las adultas a actuar impulsivamente y a experimentar reacciones emocionales intensas ante estímulos externos.¹⁷⁴ Eso significa que es necesario actuar con cautela, no solo con respecto a los efectos negativos de la vigilancia de las personas usuarias y su uso para segmentar la publicidad, sino también en la amplificación selectiva de los contenidos por parte de las plataformas de forma más generalizada.

Las plataformas de redes sociales desempeñan un papel enorme en esta etapa clave del desarrollo de las personas jóvenes y en su lucha por comprenderse a sí mismas. Documentos internos de la empresa—entre los que destacan los “Papeles de Facebook”¹⁷⁵— e investigaciones independientes muestran que las principales plataformas ejercen una enorme influencia sobre el estado de ánimo, las percepciones y la autoimagen de niños, niñas y jóvenes a través de los contenidos que recomiendan y los diseños de plataforma que emplean.¹⁷⁶

169. Amnistía Internacional, *Gigantes de la vigilancia* (antes citado). Amnistía Internacional, *“Siento que no se me protege”: Emboscada en la red de vigilancia de TikTok* (antes citado).

170. MIT Technology Review, “Is Facebook targeting ads at sad teens?”, mayo de 2017, technologyreview.com/2017/05/01/105987/is-facebook-targeting-ads-at-sad-teens/.

171. Access Now, “Open letter: TikTok’s ‘Focused View’ feature must respect human rights law”, 7 de febrero de 2023, <https://www.accessnow.org/open-letter-tiktoks-focused-view/>.

172. Zara Abrams, “What neuroscience tells us about the teenage brain”, 25 de agosto de 2022, *Monitor on Psychology*, volumen 53, número 5, <https://www.apa.org/monitor/2022/07/feature-neuroscience-teen-brain>.

173. University of Rochester Medical Centre, Health Encyclopaedia, “Understanding the teen brain”, 2023, <https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=1&contentid=3051> (consultado el 14 de septiembre de 2023).

174. BBC, “The biggest myths of the teenage brain”, 7 de septiembre de 2022, <https://www.bbc.com/future/article/20220823-what-really-goes-on-in-teens-brains>.

175. Wall Street Journal, “Facebook knows Instagram is toxic for teen girls company documents show” (antes citado).

176. Reset Australia, *Surveilling Young People Online: An Investigation Into TikTok’s Data Processing Practices*, julio de 2021, https://au.reset.tech/uploads/resettechaustralia_policymemo_tiktok_final_online.pdf; Jacopo Pruccoli et al., “The use of TikTok among children and adolescents with eating disorders: Experience in a third-level public Italian center during the SARS-CoV-2 pandemic”, 30 de julio de 2022, *Italian Journal of Pediatrics*, volumen 48, <https://doi.org/10.1186/s13052-022-01308-4>; Reset Australia, *Designing for Disorder: Instagram’s Pro-Eating Disorder Bubble in Australia*, abril de 2022, <https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf>; Tech Transparency Project, “‘Thinstagram’: Instagram’s algorithm fuels eating disorder epidemic”, 8 de diciembre de 2021, <https://www.techtransparencyproject.org/articles/thinstagram-instagram-algorithm-fuels-eating-disorder-epidemic>.

177. Comité de los Derechos del Niño, Observación general núm. 25 (antes citado), párr. 62.

178. Reset Australia, *Surveilling young people online* (antes citado).

Como ha argumentado anteriormente el Comité de los Derechos del Niño, el uso que hacen las plataformas de redes sociales de las inferencias basadas en su recopilación de datos masiva para elaborar perfiles de las personas usuarias y “personalizar” las recomendaciones plantea un riesgo muy real de interferir en los pensamientos y emociones de un niño o una niña de manera tal que se menoscabe su libertad de pensamiento.¹⁷⁷

Hay muchos estudios que documentan estos riesgos sistémicos, incluidos informes de investigación recientes que destacan específicamente el papel que desempeña TikTok en la amplificación de contenidos problemáticos y nocivos. En 2021, Reset Australia reveló que el feed “Para ti” de TikTok promovía contenidos que refleja y promueve estereotipos étnicos y de género negativos.¹⁷⁸ Ese mismo año, Wall Street Journal documentó que TikTok dirigía a personas usuarias adolescentes hacia espirales de vídeos sobre trastornos alimentarios¹⁷⁹ y contenidos depresivos.¹⁸⁰ En 2022, un estudio de Center for Countering Digital Hate sobre la amplificación en TikTok de contenidos relativos a trastornos alimentarios y autolesiones descubrió que las nuevas cuentas que incluían el término “loseweight” (perder peso) en su nombre tenían más probabilidades de recibir recomendaciones de ambos tipos de contenidos nocivos.¹⁸¹ En marzo de 2023, otro estudio a pequeña escala descubrió que el feed “Para ti” de TikTok sólo requería una “interacción de bajo nivel” por parte de las personas usuarias jóvenes antes de comenzar rápidamente a recomendar publicaciones extremas relacionadas con el suicidio y con la ideología misógina y violenta denominada “incel”.¹⁸²

Estos hallazgos apuntan a algunos de los graves daños potenciales asociados al sistema de recomendación de TikTok, que añaden riesgos para la salud de los niños, niñas y jóvenes a los ya comentados relativos a los derechos a la privacidad y a la libertad de pensamiento.

Varios progenitores de luto afirman que la selección de contenidos de TikTok ha tenido ya consecuencias mortales. Una demanda presentada en Estados Unidos en junio de 2022 sostiene que TikTok tuvo que ver con la muerte de al menos siete menores que participaron en el llamado “blackout challenge”, reto en el que quienes participan se autoasfixian y comparten vídeos del acto online, vídeos que, según una de las muchas demandas, no sólo se podían buscar, sino que se amplificaban a través del feed “Para ti”.¹⁸³ TikTok se opone a la demanda de los progenitores, manifestando que el reto “es anterior a la plataforma y no ha sido nunca tendencia en TikTok”.¹⁸⁴

Muchos de niños, niñas y personas jóvenes que participaron en la investigación para este informe eran sumamente conscientes de los riesgos asociados a las prácticas de TikTok para fomentar la interacción de las personas usuarias, en relación con los de otras plataformas destacadas.¹⁸⁵ “En todas las plataformas que utilizo hay un cierto grado de inseguridad, pero creo que, entre todas ellas, donde realmente me siento insegura es en TikTok”, explicó “Nikki”, de 24 años, residente en Manila.¹⁸⁶

En respuesta a las preguntas sobre los tipos de contenidos que habían encontrado en sus feeds y que no querían ver, las personas jóvenes participantes mencionaron con frecuencia contenidos sexuales y aquéllos que contenían violencia extrema. Por ejemplo, “María”, alumna de 17 años de Manila, explicó que

179. Wall Street Journal, “‘The corpse bride diet’: How TikTok inundates teens with eating-disorder videos”, 17 de diciembre de 2021 (antes citado).

180. Wall Street Journal, “Investigation: How TikTok’s algorithm figures out your deepest desires”, 21 de julio de 2021, <https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477>.

181. Center for Countering Digital Hate, *Deadly by Design*, 15 de diciembre de 2022, <https://counterhate.com/research/deadly-by-design/>.

182. Ekō, *Suicide, Incels, and Drugs: How TikTok’s Deadly Algorithm Harms Kids*, marzo de 2023, https://s3.amazonaws.com/s3.sumofus.org/images/eko_TikTok-Report_FINAL.pdf.

183. The Verge, “The TikTok ‘blackout challenge’ has now allegedly killed seven kids”, 8 de julio de 2022, <https://www.theverge.com/2022/7/7/23199058/tiktok-lawsuits-blackout-challenge-children-death>.

184. The Verge, “The TikTok ‘blackout challenge’ has now allegedly killed seven kids”, 8 de julio de 2022 (antes citado).

185. Sólo entre los grupos-muestra de Kenia, al menos 40 participantes dieron ejemplos de sentir inseguridad en TikTok, por ejemplo en el contexto de lo que consideraban vídeos violentos (incluidos algunos que parecían mostraban una violación y un asesinato) o contenido inapropiado para su edad (por ejemplo, pornográfico), presente en sus feeds.

186. Entrevista online, Manila (Filipinas), 6 de mayo de 2023.

un vídeo incluido en su feed mostraba el asesinato de una mujer, en lo que ella creía que era un crimen real.¹⁸⁷ En otro debate con grupos-muestra de jóvenes en Kisumu (Kenia), la mayoría de las personas participantes comunicaron haberse encontrado con el mismo vídeo viral, en el cual indicaron que aparecía una joven desnudada a la fuerza a manos de un grupo de hombres y a continuación sometida a un acto de mutilación genital femenina.¹⁸⁸

6.2 MAYORES RIESGOS PARA LAS PERSONAS JÓVENES CON PROBLEMAS DE SALUD MENTAL

Múltiples personas jóvenes con problemas de salud mental autodeclarados expusieron al personal investigador de Amnistía Internacional su experiencia de haberse visto arrastradas hacia espirales descendentes de TikTok con contenidos provocadores y relacionados con la salud mental. Consideraban que la amplificación por parte de TikTok de contenidos problemáticos relativos a la salud mental contribuía a exacerbar sus síntomas. Por ejemplo, “Luis”, estudiante universitario de 21 años de Manila diagnosticado con trastorno bipolar, describió su experiencia con el feed “Para ti” de TikTok de este modo:

“Cuando me sentía decaído, creo que el 80% [del contenido] tenía que ver con la salud mental. Es como una espiral. Es una madriguera en la que te adentras a partir de un solo vídeo. Si un vídeo logra captar tu atención, aunque no te guste, te lo presentan cuando vuelves a abrir TikTok y, como te resulta familiar, lo vuelves a ver. Al verlo de nuevo, su frecuencia en tu feed aumenta exponencialmente”.¹⁸⁹

Haciéndose eco de la experiencia de otras personas, “Luis” describió el modo en que los periodos de interacción deliberada con vídeos de TikTok que reflejaban sus propios pensamientos depresivos o ansiosos lo mantenían atrapado en bucles de contenido negativo incluso después de que remitiera la fase depresiva:

“TikTok asume que quieres ver ese contenido concreto durante un tiempo concreto, pero la cuestión es que, para mí, el contenido preferido cambia mucho. Por ejemplo, un día me siento ansioso y quiero ver este contenido. Pero otro estoy muy contento y radiante por algún motivo, y resulta que sigo viendo contenidos muy tristes y deprimentes, y eso me trastoca”.

“Francis”, estudiante de la provincia filipina de Batangas (Filipinas), observó algo parecido:

“Al marcar con un corazón un vídeo triste con el que te identifiques, de pronto toda la página ‘Para ti’ está triste, entras en ‘tristok’. Afecta a mi estado de ánimo”.¹⁹⁰

Otra persona joven participante en un grupo-muestra explicó:

“Como persona que piensa en exceso, el contenido que veo —como vídeos en los que alguien está enfermo o se autodiagnostica— me lleva a pensar en exceso {aún} más. Afecta a mi mentalidad, me hace sentir que tengo los mismos síntomas y agrava mi ansiedad. Ni siquiera busco estos vídeos, sencillamente aparecen en mi feed”.¹⁹¹

Ocho personas jóvenes adultas de Filipinas afirmaron que se les habían mostrado grupos de contenidos provocadores que afectaron a su salud y bienestar. Entre estas ocho personas estaban

187. Entrevista online, Manila (Filipinas), 9 de mayo de 2023.

188. Debate con grupos-muestra, Kisumu (Kenia), 13 de marzo de 2023.

189. Luis (seudónimo), estudiante universitario en Manila de veintipocos años, fue entrevistado en persona el 6 de mayo de 2023.

190. Debate con grupos-muestra en Tanauan, efectuado el 13 de mayo de 2023.

191. Debate con grupos-muestra en Tanauan, efectuado el 13 de mayo de 2023.

“Cris”, un joven profesor que trata de mantener a raya el trastorno por déficit de atención con hiperactividad (TDAH),¹⁹² “Daisy”, de 22 años, que sufre problemas de imagen corporal,¹⁹³ y “Lance”, de 23 años, joven transgénero que lidia con una disforia corporal autodeclarada.¹⁹⁴ Otras personas jóvenes, en particular en Filipinas, comunicaron también los efectos negativos para la salud mental de amplificar propaganda política y lo que consideraban que era desinformación, en el contexto del entorno político ya muy polarizado en el que viven. Si bien es difícil atribuir el estado de salud mental de una persona a un único factor, estos testimonios ponen de relieve que existe el riesgo de que TikTok pueda intensificar los problemas de salud mental preexistentes en niños, niñas y jóvenes.

6.3 EXPLORAR EL SISTEMA DE RECOMENDACIÓN DE TIKTOK SISTEMÁTICAMENTE

La investigación de Amnistía Internacional sobre el sistema de recomendación de TikTok se llevó a cabo para ahondar en las pruebas disponibles de dos modos. En primer lugar, se incorporaron dos países (Kenia y Filipinas) que hasta la fecha están infrarrepresentados en los estudios sobre este tema. En segundo lugar, la auditoría se llevó a cabo con una muestra de mayor tamaño de la utilizada habitualmente en anteriores investigaciones de organizaciones de la sociedad civil sobre este campo.¹⁹⁵

6.3.1 DISEÑO DE INVESTIGACIÓN AUTOMATIZADO (SOCK PUPPETS)

Amnistía Internacional, junto con sus socios técnicos —Instituto de Transparencia Algorítmica e Investigación Forense de la Inteligencia Artificial—, llevó a cabo una auditoría en dos partes del sistema de recomendación de feeds “Para ti” de TikTok, en lo que respecta a contenidos relacionados con la salud mental. La primera consistió en una auditoría automatizada —basada en los llamados “sock puppets” (bots)— de las recomendaciones efectuadas a cuentas de Kenia y Estados Unidos.¹⁹⁶ La segunda fue un experimento de gestión manual para el que se utilizaron una cuenta de Kenia, otra de Filipinas y otra de Estados Unidos.

6.3.2 DISEÑO DE INVESTIGACIÓN AUTOMATIZADO (SOCK PUPPETS)

Para la auditoría automatizada, en la investigación se crearon 40 cuentas automatizadas con 4 personalidades distintas predefinidas, a fin de imitar la conducta de diferentes jóvenes en TikTok. Cada cuenta se configuró para estar activa durante algo menos de 60 minutos en una única sesión diaria a lo largo de 10 días.¹⁹⁷ Las cuentas se dividieron en subgrupos que mostraban cuatro formas diferentes de desplazar la pantalla; de ellas, 20 se configuraron para simular a jóvenes de 13 años en Estados Unidos y otras 20, a jóvenes de 13 años en Kenia (de las cuales 11 se incluyeron en el análisis).¹⁹⁸ La edad indicada en las cuentas (13 años) se eligió con vistas a examinar las recomendaciones

192. Entrevista en Manila, efectuada el 6 de mayo de 2023.

193. Entrevista en Manila, efectuada el 6 de mayo de 2023.

194. Entrevista en Manila, efectuada el 6 de mayo de 2023.

195. Véase por ejemplo Center for Countering Digital Hate, *Deadly by Design*, 15 de diciembre de 2022 (antes citado); Reset Australia, *Surveilling Young People Online: An Investigation Into TikTok's Data Processing Practices*, julio de 2021 (antes citado).

196. En la investigación automatizada no se incluyeron cuentas de Filipinas debido a las limitaciones del proyecto. En su lugar, la investigación se centró en un análisis comparativo de un país infrarrepresentado en anteriores investigaciones en este campo (Kenia) y un mercado y país clave en el que está en curso un debate público sobre los riesgos de las redes sociales (Estados Unidos).

197. Si por problemas técnicos se interrumpía una sesión, se llevaba a cabo una sesión adicional para que la cuenta hubiera estado activa unas 10 horas en total. Dado que TikTok establece el límite voluntario de tiempo en 60 minutos al día para personas menores de 18 años, la experiencia de usuario se probó dentro de ese mismo límite, aunque es inferior al tiempo medio que, según un estudio de mercado estadounidense, pasa una persona adolescente en TikTok.

198. El experimento comenzó con 20 cuentas creadas para representar a las personas usuarias adolescentes de cada país, pero sólo 11 de las cuentas de Kenia funcionaron hasta la conclusión de la investigación. Sólo 11 de las 20 cuentas de Kenia funcionaron durante las 10 sesiones; las demás fueron desactivadas por TikTok antes de la última sesión y, por tanto, se excluyeron del análisis.

ofrecidas al grupo de edad más joven permitido entre quienes usan TikTok, cuentas a las que TikTok aplica medidas de seguridad para adolescentes.¹⁹⁹ Con el fin de limitar el grado en que el estudio promovía contenidos que pudieran ser nocivos para otras personas usuarias, las cuentas no marcaban “me gusta”, no comentaban, no enviaban mensajes, no buscaban nada ni seguían a nadie; solamente veían dos veces las publicaciones recomendadas en el feed “Para ti” si la descripción de la publicación incluía alguno de los términos incluidos en las listas predefinidas:

Subgrupo 1 (Comportamiento depresivo (canalización)): Las cuentas de este subgrupo simulaban a niños y niñas con interés en la salud mental que interactúan con contenidos cada vez más extremos una vez que se les presentan. Las cuentas de esta categoría en un principio volvían a ver vídeos asociados a un conjunto más amplio de 150 términos relativos a la salud mental. A mitad del experimento, se las programó para que volvieran a ver sólo contenidos etiquetados con una lista más corta, de 64 términos, que identificamos antes del experimento como asociados con frecuencia con contenidos claramente depresivos o relativos a las autolesiones.

Subgrupo 2 (Interés amplio en la salud mental): Las cuentas volvieron a ver vídeos etiquetados con una lista de 150 términos relativos a la salud mental.

Subgrupo 3 (Interés general —referido como términos “inofensivos” en los gráficos— e interés en la salud mental): Las cuentas de esta categoría volvieron a ver vídeos que contenían una selección de 480 términos de interés general, así como la lista amplia de 150 términos relativos a la salud mental.

Subgrupo 4 (Interés general —referido como términos “inofensivos” en los gráficos—, grupo de control): Estas cuentas sólo volvieron a ver vídeos asociados a 480 términos de interés general, como *animalsoftiktok*, *comedy* o *traveltiktok*.²⁰⁰

La lista amplia de términos relacionados con la salud mental incluyó términos como: *mentalhealth*, *mentalhealthmatters*, *sadtiktok*, *schooldraining*, *struggling* o *ventingaccount*, así como abreviaturas habituales, errores ortográficos y argot referente a la depresión y las autolesiones.²⁰¹

La lista de términos restringidos relacionados con la salud mental comprende un subconjunto de términos de la lista amplia hallados en los experimentos de prueba, que demostraron estar asociados sobre todo con contenidos problemáticos y nocivos.²⁰²

La pregunta principal de la investigación era hasta qué punto (en su caso) se mostraban a niños, niñas y jóvenes vídeos potencialmente nocivos que representaran, idealizaran o alentaran la depresión, las autolesiones y el suicidio a través de recomendaciones algorítmicas. La hipótesis de trabajo era que, si el efecto de amplificación era potente, en la investigación se observaría un aumento constante del volumen de contenidos potencialmente nocivos relativos a la salud mental que se mostraban a las cuentas de los subgrupos 1 y 2 (conducta depresiva e interés amplio en la salud mental) en relación con los mostrados a las cuentas de intereses más generales.

Aunque el estudio incluyó una muestra mayor en comparación con investigaciones previas de la sociedad civil en este campo, el tamaño muestral de 40 cuentas seguía sin ser lo suficientemente extenso como para efectuar pruebas estadísticas con hipótesis. Por lo tanto, el análisis principal comparó descriptivamente el volumen y la frecuencia de los contenidos relativos a la salud mental que visualizó cada subgrupo. La investigación aspiraba a utilizar como base estudios anteriores, mediante la inclusión de dos grupos de cuasicontrol en forma de subgrupos que únicamente volvían a ver contenidos de interés general (denominados “términos inofensivos” en los gráficos), o de interés general, junto con contenidos sobre salud mental. La inclusión de estos dos grupos

199. Véase el Anexo 3: Respuesta escrita a las preguntas de investigación de Amnistía Internacional.

200. Para consultar las listas completas de términos, véase el Anexo 2.

201. Para consultar las listas completas de términos, véase el Anexo 2.

202. Para consultar las listas completas de términos, véase el Anexo 2.

permitió al personal investigador observar la magnitud de la amplificación por parte del sistema de recomendación, ya que proporcionó una medida de referencia en cuanto al volumen y la frecuencia de los contenidos relativos a la salud mental que muestra TikTok a una persona usuaria “promedio”.

Es muy probable que factores tales como la fecha, la hora del día y la ubicación se utilicen en el algoritmo de recomendación y, por tanto, influyan a la hora de determinar el contenido que TikTok muestra a las personas usuarias. Para garantizar que las comparaciones entre subgrupos fueran lo más sólidas posible, el personal investigador trató de controlar estos factores mediante la creación de cuentas de bots que accedieran a la plataforma en fechas y horas concretas desde una misma ubicación en Kenia y Estados Unidos. Cuestiones técnicas como la desactivación de los bots o fallos en su funcionamiento causaron complicaciones, pero el personal investigador se aseguró, en la medida de lo posible, de que se recopilaran todos los datos en paralelo en cada subgrupo.

6.3.3 DISEÑO DE LA INVESTIGACIÓN DEL EXPERIMENTO MANUAL

Para contrarrestar algunas limitaciones del enfoque automatizado, como la necesidad de predefinir los términos con que la cuenta identificaría el contenido relevante y la incapacidad de las cuentas automatizadas para detectar publicaciones relevantes en función del estado de ánimo, la música o los mensajes de los vídeos en lugar de los hashtags, Amnistía Internacional e Investigación Forense de la Inteligencia Artificial efectuaron también un experimento manual a menor escala, que consistió en grabar la pantalla durante una hora de interacción del personal investigador con el feed “Para ti” en cuentas recién creadas, de nuevo en representación de jóvenes de 13 años. Se configuró una cuenta para cada miembro a través de redes privadas virtuales en Kenia, Filipinas y Estados Unidos. Una vez más, el personal investigador simplemente se desplazó por el feed “Para ti” y vio las publicaciones dos o tres veces si estaban relacionadas con la tristeza o con problemas de salud mental.

6.4 MEDICIÓN Y CATEGORIZACIÓN DE LOS CONTENIDOS NOCIVOS A EFECTOS DE LA INVESTIGACIÓN

6.4.1 CONSTRUCCIÓN DE UNA MEDICIÓN DE AMPLIFICACIÓN

En la bibliografía disponible sobre los riesgos y daños asociados a las redes sociales, rara vez se define el significado preciso de “contenido nocivo”.²⁰³

Esto supone un reto importante, dado que el contenido de TikTok se compone de una mezcla de datos de vídeo, texto y audio y, en este contexto, se recopila a tan gran escala que hace inviables los métodos de categorización manuales. A pesar de estos retos, era vital e integral para la investigación clasificar el contenido para proporcionar una medida de amplificación. Por lo tanto, el personal investigador hizo uso del siguiente enfoque:

1. La medida de la amplificación principal era el porcentaje de vídeos recomendados que contenían un término de la lista amplia de términos de salud mental (ya fuera en un hashtag o un texto asociado). En la investigación se analizaron y representaron gráficamente los términos que aparecían conjuntamente para comprender cómo se relacionan los temas de los vídeos.
2. A partir de la investigación cualitativa con niños, niñas y jóvenes y de los comentarios de los especialistas en salud mental entrevistados, el personal investigador construyó un marco de categorización de contenidos. Se convocó a un grupo de revisión, que clasificó manualmente

203. Por ejemplo, la recomendación del Cirujano general de Estados Unidos para 2023 describe los contenidos nocivos como “por ejemplo, contenidos que fomenten los trastornos alimentarios, la violencia, el abuso de sustancias, la explotación sexual y el suicidio o que hablen de formas de suicidarse”. Véase: Cirujano general de Estados Unidos, *Social Media and Youth Mental Health* (antes citado).

una muestra de 540 vídeos de TikTok con este marco; cada contenido fue revisado independientemente por tres personas revisoras.

3. A continuación, el personal investigador analizó la correspondencia de las categorizaciones manuales con los hashtags y las descripciones de texto de los contenidos. Este enfoque permitió al personal investigador validar nuestra medición de amplificación principal, al evaluar cuánto contenido que incluía un término de la lista amplia se consideraba nocivo en el ejercicio de clasificación manual.

Una limitación clave de la medida de amplificación principal es que los hashtags de vídeo asignados se utilizan tanto para entrenar a los robots e imitar la conducta de las personas usuarias como para medir el volumen de contenido sobre salud mental que se les mostró. Con capacidad adicional, una mayor categorización manual podría haber proporcionado una imagen más profunda y, posiblemente, haber ayudado a perfeccionar los métodos de análisis cuantitativo. Los esfuerzos de investigación de la sociedad civil en este campo siguen viéndose limitados por restricciones presupuestarias y temporales, además de los obstáculos que crea la inexistencia de marcos de acceso a la investigación apropiados y examinados de forma independiente en el momento de redactar este informe (el marco de acceso a la investigación de la LSD aún no se ha aplicado). Lo anterior subraya aún más la necesidad de que se apruebe reglamentación fuera de la UE para imponer obligaciones similares en relación con el acceso a los datos, además de crear obligaciones exigibles a las principales plataformas de redes sociales con el fin de que comuniquen públicamente sus evaluaciones de riesgos y acciones de mitigación.

6.4.2 CLASIFICAR CONTENIDOS

Aunque la investigación de Amnistía Internacional se centró en el efecto acumulativo de la amplificación por parte de TikTok de tipos específicos de contenidos dirigidos a una persona menor a lo largo del tiempo, más que en el efecto de fallos individuales en la moderación de contenidos, la codificación de los datos requirió así y todo que el personal investigador categorizara contenidos individuales. Las categorías utilizadas en este informe reflejan los comentarios de varios médicos y especialistas, pero de todos modos sólo pueden ofrecer una aproximación del contenido observado y de su potencial nocivo para la salud mental de una persona.²⁰⁴ El esquema de categorización que se utilizó pretendía ofrecer una descripción transparente de las dimensiones del contenido potencialmente nocivo para la salud mental que se recomendó a las cuentas de la investigación, en lugar de efectuar una distinción binaria entre contenido “nocivo” y “no nocivo”.

En realidad, más allá de los ejemplos flagrantes de contenidos traumatizantes, las percepciones y las reacciones emocionales ante los contenidos son altamente individuales. Las entrevistas de Amnistía Internacional a niños, niñas y jóvenes pusieron de manifiesto que lo que una persona percibe como provocador o nocivo para su salud mental a menudo depende de sus experiencias y circunstancias personales. También puede variar con el tiempo según los cambios en el estado emocional y mental.

Además, en relación con los contenidos generados por las personas usuarias, los psicólogos y psicólogas tampoco se ponen de acuerdo sobre los pros y los contras de que personas no expertas compartan información sobre salud mental en las redes sociales. Como se ha señalado anteriormente, plataformas como TikTok sí ofrecen oportunidades sin precedentes para la autoexpresión y un sentido de comunidad para muchas personas, incluidas aquellas que comparten los mismos motivos de preocupación sobre la salud mental.²⁰⁵ También pueden conectar con personas que, de otro modo, no tendrían acceso a recursos de salud mental con información diseñada para ser fácil y ampliamente accesible. Sin embargo,

204. La Dra. Bryn Austin (STRIPED/Facultad de Medicina de Harvard), la Dra. Amanda Raffoul (Facultad de Medicina de Harvard), Zeyna Awan, psicóloga residente en Nairobi, Kevin Gachee, psicólogo residente en Nairobi, y Aliya Shah, psicóloga residente en Nairobi, aportaron sus comentarios.

205. John A. Naslund, Ameya Bondre y otros, “Social media and mental health: Benefits, risks, and opportunities for research and practice” (antes citado).

los expertos han expresado serios motivos de preocupación acerca de los fragmentos breves de vídeo que transmiten información y posible desinformación, así como sobre los contenidos que incluyen (supuestos) consejos médicos sin ningún conocimiento de quien los visualiza y de sus necesidades personales, o que animan al autodiagnóstico sobre la base de información incompleta o falsa.²⁰⁶

Un aspecto crucial es que estos contenidos no se consumen de forma aislada, sino en el contexto de un feed de ritmo rápido que, como demuestra la investigación de Amnistía Internacional, puede pronto convertirse en una avalancha de contenidos problemáticos relativos a la salud mental.

Además de ejemplos de contenidos que encajaban claramente en las categorías definidas para el experimento, el personal investigador también identificó publicaciones recomendadas difíciles de categorizar con certeza, por ejemplo, por ser demasiado breves y, por tanto, tener un significado y una intención poco claros. Para mitigar posibles sesgos individuales y validar nuestros hallazgos, tres especialistas de la plataforma Junkipedia²⁰⁷ asignaron una categoría a cada publicación; en los casos en que en esta revisión se asignaron categorías diferentes, la categorización final se basó en una decisión por mayoría.

Tipos de contenidos relacionados con la salud mental con posibles efectos nocivos al amplificarse a gran escala

 <p>Publicaciones de experiencias vividas que hablan de ansiedad, depresión y autolesiones (sin idealizarlas)</p>	 <p>Publicaciones que muestran a personas en situación de angustia emocional</p>	 <p>Publicaciones que presentan los sentimientos de soledad o inadecuación como inevitables o merecidos</p>
 <p>Publicaciones (inclusive citas, texto, voz en off) cuya intención es mostrar las autolesiones, la depresión y los pensamientos suicidas como ineludibles</p>	 <p>Descripciones dramáticas de traumas, sufrimiento y suicidio</p>	 <p>Imágenes traumatizantes</p>
 <p>Publicaciones que glorifican, idealizan o trivializan la depresión, la ansiedad, las autolesiones y el suicidio (incluso a través de dibujos, cómics, música y mensajes)</p>	 <p>Publicaciones que transmiten información errónea relativa a la salud mental, por ejemplo, publicaciones que disuaden a las personas usuarias de solicitar ayuda profesional o de tomar la medicación prescrita, o que fomentan el autodiagnóstico basado en información engañosa</p>	 <p>Publicaciones en las que se mencionan planes de autolesionarse o suicidarse o se explica cómo autolesionarse o suicidarse, o que fomentan de otro modo las autolesiones o el suicidio</p>

206. Joe Martin, "The big issue: Mental health and the TikTok effect", abril de 2023, <https://www.bacp.co.uk/bacp-journals/therapy-today/2023/april-2023/the-big-issue/>

207. Más detalles técnicos en el Anexo 1.

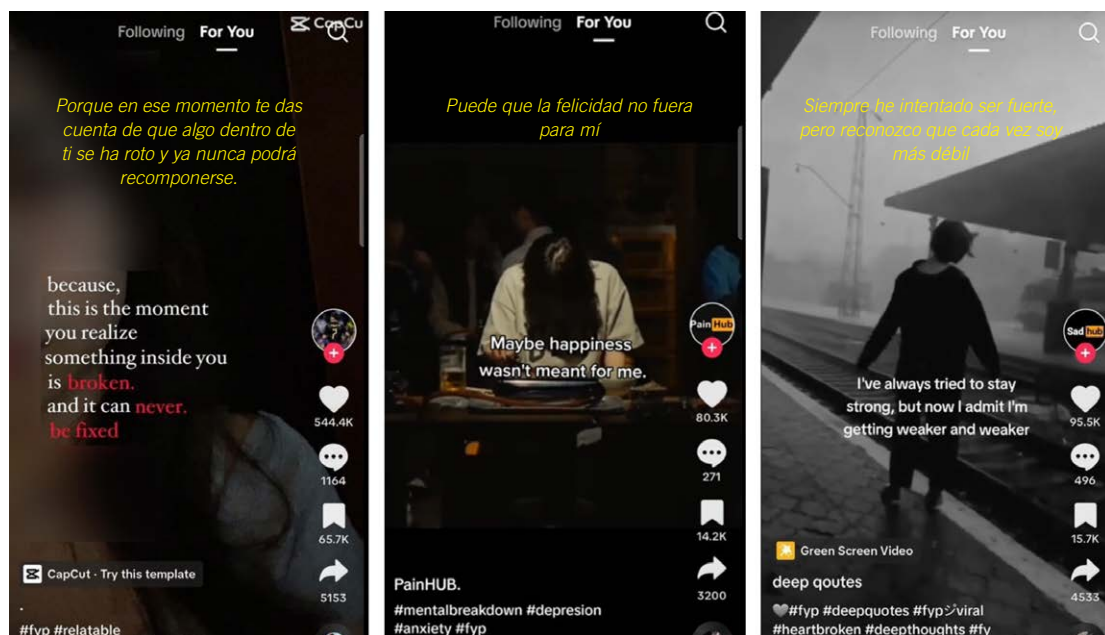
6.5 CONCLUSIONES DESTACADAS

Los resultados de la investigación de Amnistía Internacional sobre TikTok muestran que los niños, niñas y jóvenes que ven contenidos relacionados con la salud mental en la página “Para ti” de TikTok pueden verse arrastrados fácilmente hacia espirales de contenidos potencialmente nocivos, incluidos vídeos que idealizan y fomentan el pensamiento depresivo, las autolesiones y el suicidio.

Así sucedió con cuentas automatizadas configuradas para representar a jóvenes de 13 años en Kenia y Estados Unidos; algunos contenidos se recomendaron tanto a una cuenta como a la otra. Al cabo de 5 o 6 horas en la plataforma, prácticamente la mitad de los vídeos mostrados a las cuentas que habían manifestado interés por la salud mental estaban relacionados con esta y eran potencialmente nocivos, cifra que multiplicaba aproximadamente por 10 el volumen mostrado a las cuentas sin interés por la salud mental.

Estos resultados coincidían con los de la revisión manual de 540 vídeos recomendados a las cuentas de Estados Unidos y Kenia, que mostraron una progresión constante desde el 17% de los vídeos mostrados el primer día categorizados como potencialmente nocivos hasta el 44% del contenido el décimo día.

El experimento manual efectuado con cuentas únicas que representaban a jóvenes de 13 años en Kenia, Estados Unidos y Filipinas mostró un proceso de personalización aún más rápido en todas las ubicaciones, que redujo el feed de la persona usuaria a una espiral aún más pronunciada de contenido sobre salud mental potencialmente nocivo en menos de 21 minutos (en comparación con las 4 a 5 horas requeridas para las cuentas del estudio automatizado).



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

6.6 EL INTERÉS POR LA SALUD MENTAL DESENCADENA EL EFECTO ESPIRAL

Cuando el sistema de recomendación de la página “Para ti” de TikTok detecta que la persona usuaria está interesada en vídeos relativos a la salud mental, es muy probable que convierta este indicio en una pronunciada espiral de contenidos que tratan y fomentan el pensamiento depresivo y las autolesiones.

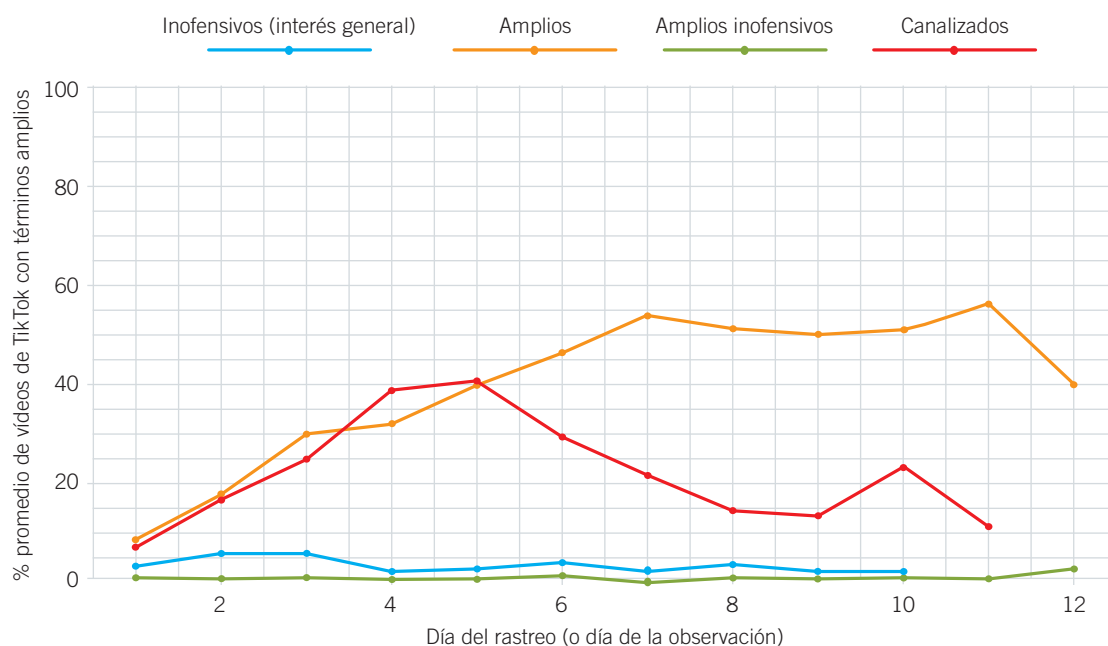
Las limitaciones de capacidad sólo permitieron revisar manualmente una muestra de 540 entre los miles de vídeos recomendados (detalles en el apartado 6.6.4), lo que significa que la validez de los resultados muestrales para el experimento global es limitada. No obstante, la muestra sugiere que los siguientes porcentajes, basados en un análisis cuantitativo de todas las recomendaciones de vídeos asociadas a la lista amplia de términos relativos a la salud mental en todo el experimento, probablemente subestimen el número real de recomendaciones de vídeos con contenido potencialmente nocivo. El porcentaje de vídeos asociados a un término de la lista amplia de términos de salud mental en la muestra fue un 6% inferior al porcentaje real de vídeos clasificados de forma manual como potencialmente nocivos (categorías 1-9) (el análisis de la muestra se tratará con más detalle en el apartado 6.6.4).

Nuestro análisis cuantitativo de los experimentos automatizados demostró que:

- El efecto de amplificación parece ser fuerte. El personal investigador observó un aumento constante durante los primeros cinco a ocho días en el volumen de contenidos sobre salud mental mostrado a las cuentas que manifestaban interés por contenidos sobre salud mental (subgrupos 1 (depresivo) y 2 (interés amplio por la salud mental)) en comparación con las que tenían intereses más extensos (gráficos 1 y 2).
- El volumen de contenido relativo a la salud mental alcanzó su punto máximo entre el quinto y el octavo día, cuando representó entre el 40 y el 55% de todo el contenido mostrado a nuestros subgrupos de conducta depresiva y de interés amplio en la salud mental en Estados Unidos durante este periodo (gráfico 1). Las pruebas de nuestro ejercicio de categorización manual sugieren que 3 de cada 5 vídeos que se identifican como relativos a la salud mental según los términos y hashtags de las descripciones son potencialmente nocivos.
- El volumen de contenidos relativos a la salud mental se mantuvo relativamente bajo (por debajo del 10% de todos los contenidos servidos) y constante a lo largo del tiempo para nuestros subgrupos con intereses más extensos durante todo el periodo del estudio, independientemente de la ubicación (gráficos 1 y 2).
- En las cuentas de canalización de Estados Unidos se observa un descenso de los contenidos relativos a la salud mental después del quinto día, momento en el que las cuentas se programaron para volver a visualizar sólo contenidos que incluyeran un término de la lista restringida (gráfico 1). El análisis sugiere que probablemente se deba a que la frecuencia global de contenidos que incluyen un término de la lista restringida entre el total de contenidos recomendados era muy baja (<1% tanto antes como después de la canalización) para cuatro de las cinco cuentas. Por tanto, las cuentas buscaban contenidos específicos de los que al parecer no había muchos disponibles entre las recomendaciones en el momento del cambio. En ausencia de una señal real, es probable que el sistema de recomendación utilizara otros factores para conformar los feeds “Para ti” de las cuentas (por ejemplo, temas de actualidad, eventos).

- Por el contrario, las cuentas de canalización en Kenia mantuvieron un volumen más constante de contenidos relativos a la salud mental antes y después del quinto día (gráfico 2). El análisis muestra que se debió al bajo número de cuentas de canalización en Kenia que llegaron al final del estudio (sólo dos cuentas no fueron suspendidas), de las cuales una cuenta experimentó un efecto espiral sustancial que probablemente llevara a TikTok a seguir recomendando contenidos relativos a la salud mental tras el cambio a la lista más corta de términos identificados antes del experimento como asociados con frecuencia con contenidos abiertamente depresivos o relacionados con autolesiones.

Gráfico 1: Porcentajes (medias por subgrupo) de vídeos recomendados mostrados a cuentas de Estados Unidos (cuyas descripciones incluyen términos de la lista amplia de términos relativos a la salud mental por día²⁰⁸)



- El efecto de amplificación fue mucho menos pronunciado para las cuentas de Kenia. El subgrupo depresivo y el subgrupo con interés en la salud mental alcanzaron un pico del 25% (gráfico 2).
- Hubo diferencias considerables en la amplificación entre cuentas. Dentro del subgrupo con interés amplio por la salud mental en Estados Unidos, la mayoría de las cuentas alcanzaron picos de más del 60% de vídeos recomendados relativos a temas de salud mental (al día), mientras que una de las cuentas permaneció por debajo del 45% (gráfico 3).
- Observar las cuentas a lo largo de varios días proporcionó información valiosa: En algunas cuentas, hicieron falta entre 500 y 600 vídeos para alcanzar los primeros picos del efecto espiral (durante la segunda o tercera hora/día de visualizar contenidos recomendados). En otras, el proceso de personalización fue más lento y el primer pico se produjo entre los 1.500 y los 2.000 vídeos (véase el gráfico 3, la cuenta estadounidense “Términos amplios EE. UU. 05” basada en una lista amplia de términos de salud mental alcanzó su primer pico el octavo día).

208. Cuando, por problemas técnicos, una cuenta funcionó durante un tiempo insuficiente en una sesión prevista para una hora, la sesión incompleta se fusionó con la siguiente, registrada el mismo día. Para la Tabla 2: En un caso concreto, una cuenta de Kenia no completó una sesión y no se registró ninguna otra sesión ese día, por lo que la sesión incompleta se excluyó del análisis.

Gráfico 2: Porcentaje de vídeos recomendados mostrados a cuentas de Kenia (cuyas descripciones incluyen términos de la lista amplia de términos relativos a la salud mental por día)

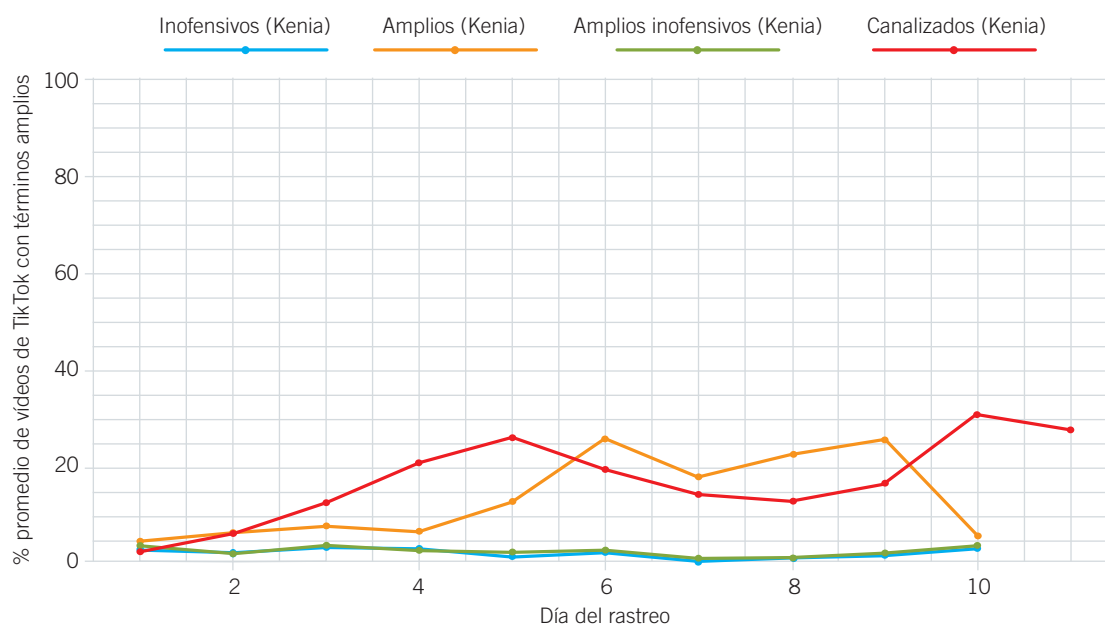
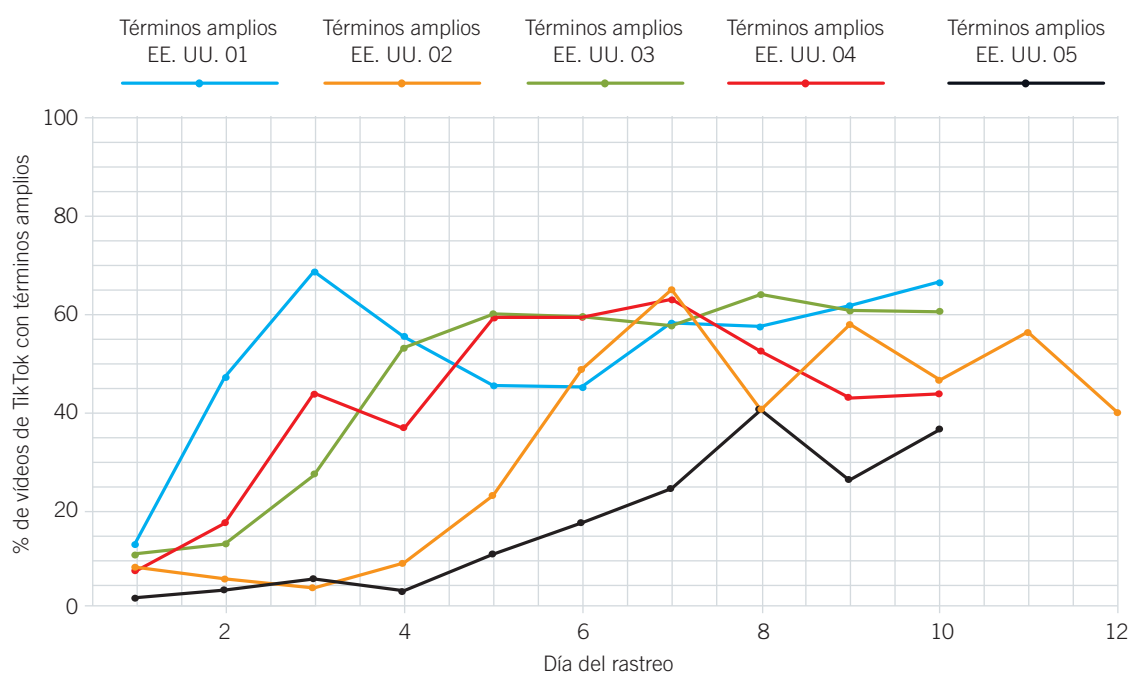


Gráfico 3: Porcentaje de vídeos recomendados mostrados a cuentas de Estados Unidos (grupo 2: interés amplio en la salud mental) cuyas descripciones incluyen términos de la lista amplia de términos relativos a la salud mental por día



Cuatro de cada cinco cuentas registraron picos de más del 60% de contenido relacionado. En todo su feed, al menos el 20% de todos los vídeos recomendados a las cuentas estadounidenses de esta categoría estuvieron relacionados con la salud mental (asociados a términos de la lista amplia).

Para contextualizar estos resultados, “aproximadamente el 13%” de los vídeos del feed de Instagram de la adolescente británica Molly Russell en los seis meses anteriores a su muerte estuvieron —según los informes— “relacionados con el suicidio, la depresión o las autolesiones”.²⁰⁹ El forense del caso halló que este nivel de amplificación personalizada de contenido sobre salud mental había sido suficiente para contribuir significativamente a su muerte.

El porcentaje significativamente mayor de contenidos relativos a la salud mental, la depresión y las autolesiones hallado en los experimentos de Amnistía Internacional²¹⁰ en TikTok como respuesta a tan solo uno de los indicios que suelen dar las personas usuarias (como visualizaciones repetidas, “me gusta” y seguimientos), si se diera en la cuenta de un niño, niña o joven real expondría a esta persona a visión de túnel aún más pronunciada. Especialmente en el caso de los niños, niñas y jóvenes vulnerables, conllevaría el peligro de exacerbar problemas de salud mental preexistentes, con el consiguiente riesgo de daños graves para su salud física. Mientras que el personal investigador que trabajó en este proyecto no marcó con “me gusta”, no buscó contenidos relevantes activamente y no siguió a ningún usuario para evitar contribuir a la amplificación de contenidos potencialmente nocivos, estas señales que envían habitualmente la mayoría de las personas usuarias posiblemente habrían dado lugar a un porcentaje de contenidos relacionados aún mayor.

En el apartado 6.6.2 a continuación se aborda el análisis manual de los feeds de las cuentas de TikTok y se examinan con más detalle los tipos de contenidos que se recomendaron en función del interés de las cuentas en vídeos asociados a términos y hashtags sobre salud mental. Además, los métodos cuantitativos ofrecen información sobre el tema y el tono de las recomendaciones. El siguiente mapa de red muestra los términos amplificados con mayor frecuencia asociados a la salud mental entre todos los vídeos recomendados, así como sus conexiones con los hashtags simultáneos. Los términos más frecuentes asociados a la lista amplia de términos de salud mental son, en primer lugar, expresiones genéricas como “mentalhealth”, “mentalhealthmatters” y “mentalhealthawareness” (a su vez, el 15-20% de esos vídeos incluye también “vent” entre sus otros hashtags y términos), seguidos de “vent”, “anxiety” y “sad”.

209. Politico, “Digital bridge: Platforms on the hook — Transatlantic AI rulebook — Let’s talk data transfers”, 6 de octubre de 2022 (antes citado)

210. Cabe hacer una salvedad: el personal investigador de Amnistía Internacional no pudo revisar la categorización aplicada al feed de Molly Russell, por lo que tal vez no sea directamente comparable.

A diferencia de los resultados observados en relación con las cuentas que sólo manifestaban interés en la salud mental, el personal investigador no observó un efecto espiral para las cuentas que combinaban la lista amplia de términos relativos a la salud mental y la lista de términos inofensivos: del 0,4 al 1,4% de los feeds de Kenia incluyeron videos vinculados a la lista amplia de términos de salud mental y menos del 1% de todos los feeds estadounidenses de ese grupo. Este resultado posiblemente se explique por el peso dado a intereses no relacionados con la salud mental (con tres veces más términos de interés general que relativos a la salud mental) y la probable disponibilidad mucho más alta de contenidos relativos a intereses inofensivos entre todos los videos que podían recomendarse a una persona usuaria.

De forma un tanto contradictoria con lo anterior, Amnistía Internacional descubrió que, a pesar de no mostrar interés activo en los videos asociados a términos relativos con la salud mental, tres de cada cinco cuentas en Estados Unidos programadas para volver a ver sólo contenidos relativos a mascotas, viajes y videos similares de interés general (términos “inofensivos”) observaron que entre el 2% y el 4% de su feed estaba relacionado con términos y hashtags vinculados a la salud mental (lista amplia de términos relativos a la salud mental), con picos de hasta el 12% (a lo largo de 50 videos consecutivos). Las cuentas de ese mismo subgrupo en Kenia vieron entre un 0,6% y un 0,7% de publicaciones relativas a la salud mental. Sería necesario efectuar un análisis más profundo para contextualizar este resultado, pero podría deberse, por ejemplo, a que uno de los términos de interés general apareciera en la descripción de algunos de estos videos, o a que las altas tasas de interacción con estos videos provocaran su amplificación en los feeds de las cuentas.

6.6.1 RESULTADOS DE LOS EXPERIMENTOS MANUALES: EL EFECTO ESPIRAL PUEDE DESENCADENARSE EN 20 MINUTOS

En comparación con las cuentas automatizadas que señalaban su interés por la salud mental, Amnistía Internacional observó un efecto espiral aún más rápido en los experimentos manuales y una tasa aún mayor de contenidos potencialmente nocivos. En la observación manual de las recomendaciones mostradas a una cuenta registrada a nombre de una persona supuestamente de 13 años en Filipinas, el primer video etiquetado con #depressionanxiety [sic] que mostraba a un niño angustiado fue sugerido en los primeros 67 segundos de exploración del contenido recomendado en la página “Para ti”. A partir del minuto 12, el 58% de las publicaciones recomendadas estaban relacionadas con la ansiedad, la depresión, las autolesiones o el suicidio y encajaban en alguna de las categorías mencionadas para contenidos relativos a la salud mental, con efectos potencialmente nocivos para niños, niñas y jóvenes con problemas de salud mental previos.

En el experimento manual llevado a cabo en Estados Unidos, el cuarto video mostrado llevaba la etiqueta #paintok y se centraba en un texto con el mensaje siguiente: “Cuando te das cuenta de que nunca en tu vida has ocupado el primer lugar, sino que eres sólo esa persona que llena un vacío en la vida de otras hasta que dejas de hacerles falta”. A partir del video 20 (en menos de 3 minutos), el 57% de los videos estaban relacionados con problemas de salud mental y al menos 9 de las publicaciones idealizaban, normalizaban o animaban al suicidio en una sola hora.

La cuenta de Kenia en los experimentos manuales fue la que más lentamente avanzó hacia un feed de contenidos relativos a la salud mental. Sin embargo, una vez alcanzado ese punto (a los 20 minutos de comenzar el experimento), el 72% de los videos recomendados en los 40 minutos siguientes estaban relacionados con problemas de salud mental, incluidas al menos 5 referencias a pensamientos suicidas o al deseo de morir de la persona autora del contenido. El primer video que hacía referencia a los problemas de salud mental de la persona creadora del contenido se recomendó a los 3 minutos de comenzar el experimento (25.º video recomendado). Ni un solo video relativo a la salud mental había sido publicado por especialistas en salud mental u organizaciones reconocidas en la materia.

6.6.2 RESULTADOS DE NUESTRA CATEGORIZACIÓN MANUAL: UN ANÁLISIS MÁS DETALLADO DE LOS TIPOS DE CONTENIDOS RELATIVOS A SALUD MENTAL RECOMENDADOS EN EL FEED “PARA TI” DE TIKTOK

Para comprender mejor los tipos de vídeos que TikTok recomendaba a las cuentas, el personal investigador analizó manualmente muestras de 20 minutos vistas por una cuenta en Estados Unidos y una en Kenia, pertenecientes a los dos subgrupos de cuentas que observaron un efecto espiral significativo (canalización e interés amplio por la salud mental) el primer, sexto y décimo día. El personal investigador codificó manualmente 540 contenidos, cada uno de los cuales fue revisado por 3 especialistas por separado y categorizado según una decisión por mayoría.

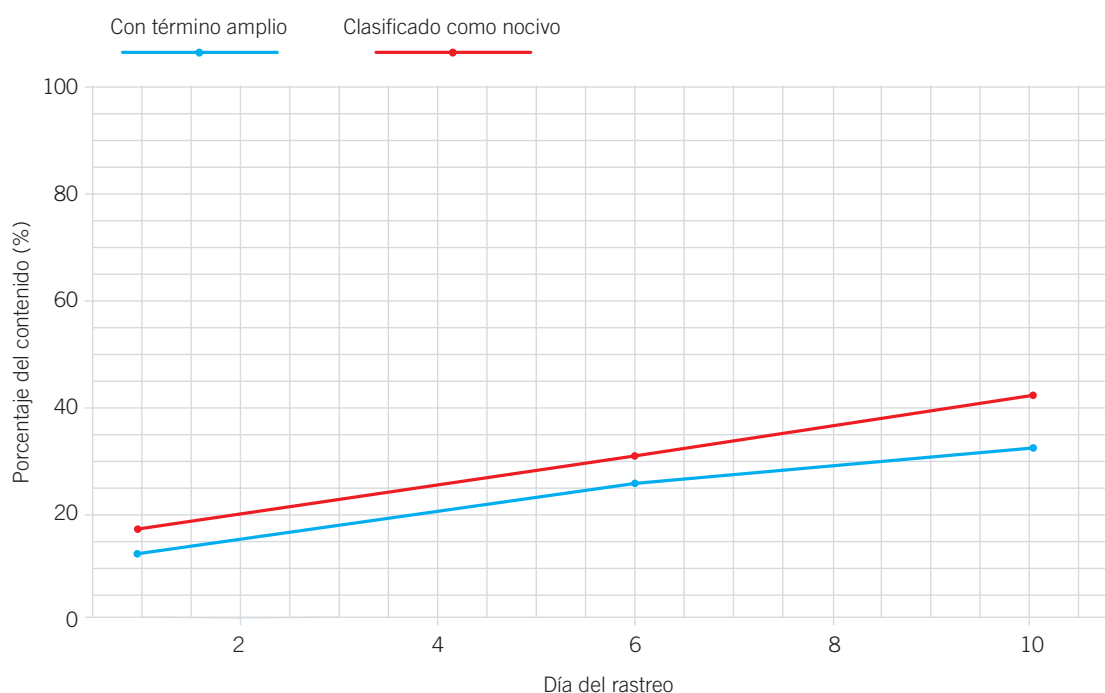
Tabla 1: Categorización manual de muestras del experimento automatizado

Código	Categoría	Núm. contenidos categorizados manualmente	% de contenidos que incluyen un término de la lista amplia
0	No pertinente	393 (73%)	11,2
1	Publicación de experiencias vividas que hablan de ansiedad, depresión y autolesiones (sin idealizarlas)	68 (13%)	44,1
2	Publicación que muestran a personas en situación de angustia emocional	23 (4,3%)	39,1
3	Publicación que presenta los sentimientos de soledad o inadecuación como inevitables o merecidos	8 (1,5%)	37,5
4	Publicación cuya intención es mostrar las autolesiones, la depresión y los pensamientos suicidas como ineludibles	4 (0,7%)	75
5	Descripción dramática de traumas, sufrimiento y suicidio	22 (4%)	59,1
6	Imágenes traumatizantes	0	-
7	Publicación que glorifica, idealiza o trivializa la depresión, la ansiedad, las autolesiones y el suicidio (incluso a través de dibujos, cómics, música y mensajes).	15 (3%)	60
8	Publicación que transmite información errónea relativa a la salud mental , por ejemplo, publicaciones que disuaden a las personas usuarias de solicitar ayuda profesional o de tomar la medicación prescrita, o que fomentan el autodiagnóstico basado en información engañosa	2 (0,4%)	50
9	Publicación en la que se mencionan planes de autolesionarse o suicidarse o se explica cómo autolesionarse o suicidarse, o que fomenta de otro modo las autolesiones o el suicidio	5 (1%)	60

De ellos, el 21% contenía un término de la lista amplia y el 27% se clasificó como potencialmente nocivo (categorías 1-9, tabla 1), lo que concuerda con las conclusiones anteriores del análisis basado en términos, puesto que el personal investigador sólo categorizó contenido de las cuentas de los subgrupos de términos amplios y de canalización. Las publicaciones de experiencias vividas sobre ansiedad, depresión y autolesiones son la categoría más frecuente de los vídeos (el 13% de todo el contenido), que —al amplificarse a esta escala y al presentar un enfoque muy centrado en este tipo particular de experiencia vivida— podrían intensificar problemas de salud mental existentes en niños, niñas y jóvenes.²¹¹ Le siguen las publicaciones que muestran a personas llorando o con angustia emocional (el 4,3% de todos los vídeos) y las descripciones dramáticas de traumas, sufrimiento y suicidio (4%). Otro 4% de los vídeos glorifican o trivializan la depresión, la ansiedad, las autolesiones y el suicidio o hablan de planes para suicidarse, explican cómo autolesionarse o animan a autolesionarse o suicidarse.

El gráfico 5 muestra el aumento constante del porcentaje de contenidos potencialmente nocivos relativos a la salud mental a lo largo del tiempo. La comparación entre el porcentaje de contenidos categorizados manualmente y el porcentaje de vídeos que incluían un término de la lista amplia de términos relativos a la salud mental muestra que el análisis cuantitativo (basado únicamente en los términos) subestima el volumen real de contenidos potencialmente nocivos, los cuales podrían suponer un riesgo para niños, niñas y jóvenes con problemas de salud mental existentes.

Gráfico 5: Progresión de los contenidos categorizados manualmente en los días 1, 6 y 10 de nuestros experimentos. Mapeamos tanto la progresión de los contenidos clasificados como nocivos (según nuestro marco) como la progresión de los contenidos cuyas descripciones incluían términos de la lista amplia de términos relativos a la salud mental



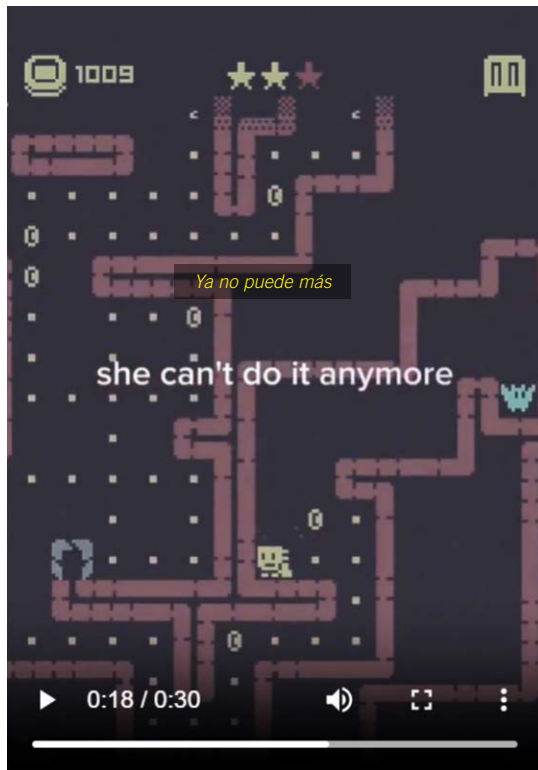
211. Katarzyna Kostyrka-Allchorne, Mariya Stoilova et al., “Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and nonsuicidal self-injury conditions – a systematic review”, febrero de 2023, *Child and Adolescent Mental Health*, volumen 28, número 1 (antes citado); North London Coroner’s Service, “Regulation 28 report to prevent future deaths”, 13 de octubre de 2022 (antes citado).

CONTEXTUALIZACIÓN DEL ANÁLISIS BASADO EN TÉRMINOS MEDIANTE EL EJERCICIO DE CLASIFICACIÓN MANUAL

Para ayudar a contextualizar los resultados del análisis principal basado en términos y hashtags, el personal investigador se propuso comprender en qué parte del marco de clasificación manual se situaba el contenido que incluía términos de la lista amplia de salud mental. De este modo se obtiene simultáneamente una imagen acerca de cuánto contenido amplificado por el sistema de recomendación es potencialmente nocivo y de cuánto contenido potencialmente nocivo omiten las cuentas de bots y no se contabiliza en el análisis basado en términos, debido a que las descripciones no incluyen hashtags ni términos de la lista amplia.

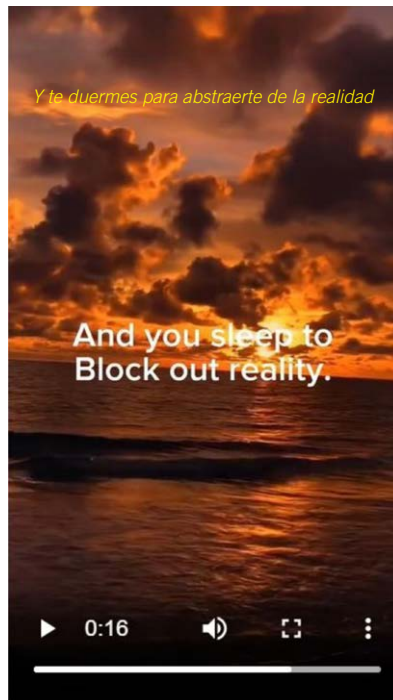
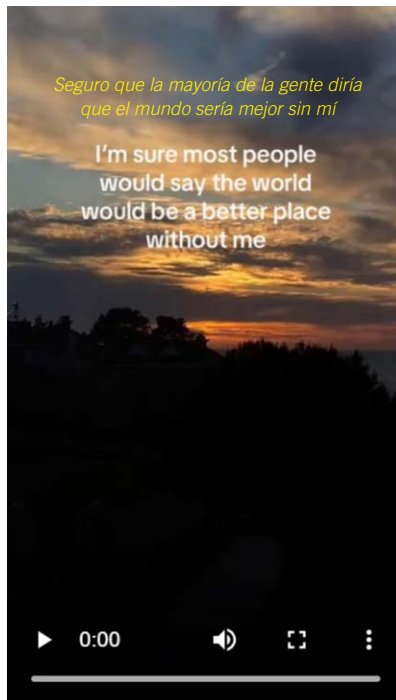
Del contenido que incluía términos de la lista amplia, el 62% se clasificó como potencialmente nocivo. El 18% de los vídeos que no contenían ninguno de los términos se clasificó como potencialmente nocivo. En conjunto, estos datos sugieren que la lista amplia de términos de salud mental es un predictor relativamente bueno de qué contenidos potencialmente nocivos y cuáles no, aunque sigue habiendo: (1) un volumen sustancial de contenido (38%) que incluye un término de la lista amplia y no se clasifica como potencialmente nocivo y (2) un volumen sustancial de contenido potencialmente nocivo que no contiene términos de la lista amplia.

6.7 EJEMPLOS DE PUBLICACIONES RECOMENDADAS



Esta publicación que describe el sufrimiento de una niña y está etiquetada con #broken #vent #ventacc #overthinking #emotionallydrain #ventaccount #ventmentalhealth #yoursafeplace #sadtiktok y otros hashtags se recomendó a tres cuentas de Estados Unidos y a una de Kenia.

Una voz en off generada por inteligencia artificial afirma: ***“No entiendes lo agotada que estás, pero aun así va a la escuela cada día. Sí, siempre sonríe, ¿pero cómo sabes que no le cuesta? Está muy cansada y todas las cosas que los chicos le decís empeoran la situación. Se va a casa llorando en silencio para que nadie la oiga y no le pregunten qué le pasa. Ya no puede más. No veis el dolor que soporta y lo mucho que le cuesta seguir viva. Lo agraváis con toda la basura que le decís, no lo soportará mucho más. Lo intenta con todas sus fuerzas”.***

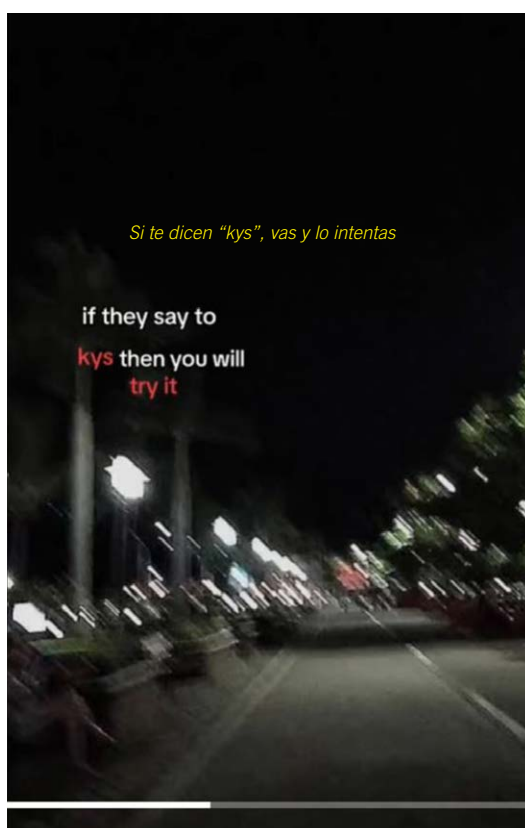
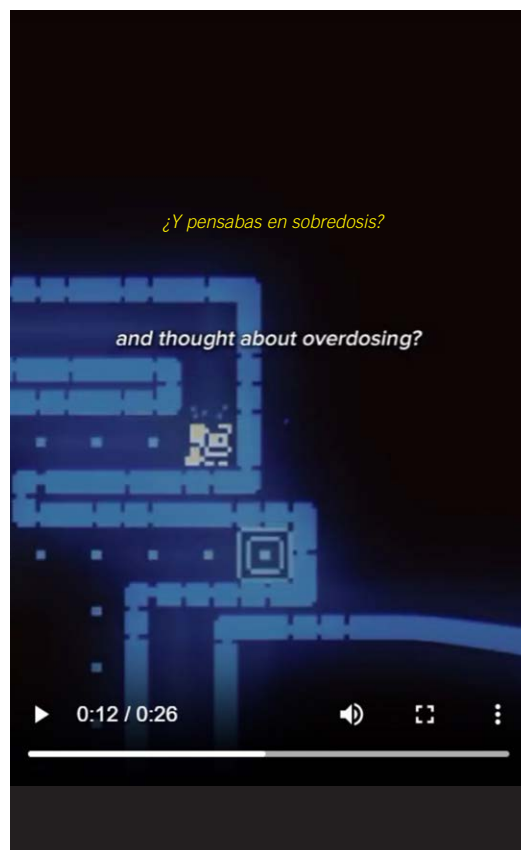
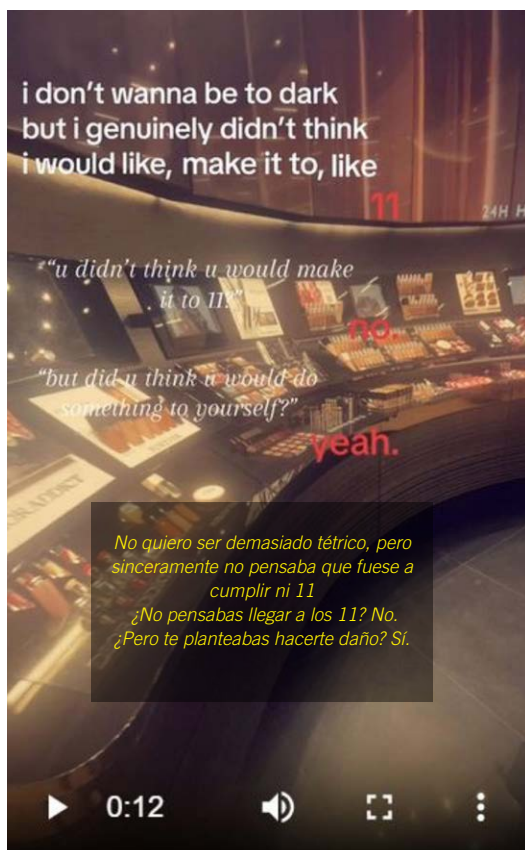


Estas publicaciones suicidas y depresivas de trasfondo romántico se recomendaron a una cuenta de Estados Unidos en el subgrupo de conducta depresiva/canalización.

EMPUJADOS A LA OSCURIDAD

EL FEED “PARA TI” DE TIKTOK FOMENTA LA AUTOLESIÓN Y LA IDEACIÓN SUICIDA

Amnistía Internacional



Ejemplos de publicaciones que tratan el suicidio, incluso con códigos y abreviaturas cuya probable intención es evitar que los detecte la moderación automática de contenidos (“kys” = suicidarse). Las carreteras oscuras y los juegos arcade son fondos habituales en muchas de las publicaciones que rastreamos relacionadas con la depresión y el suicidio.

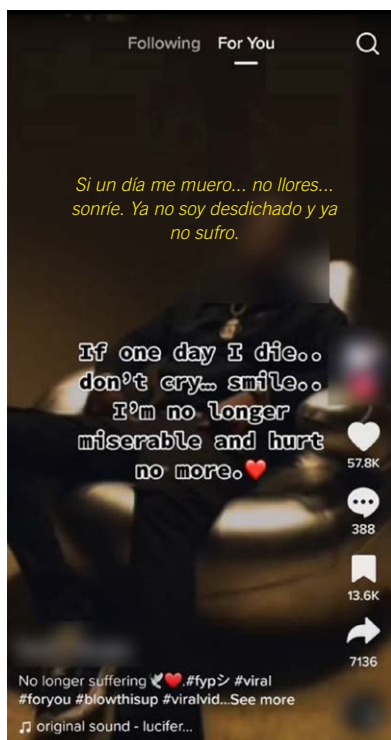
Cita del vídeo de la parte inferior izquierda (etiquetado con #MentalHealthMatters):

“¿Alguna vez has mirado un bote de pastillas y pensado en sufrir una sobredosis? ¿Alguna vez ha deseado que todo llegara a su fin? ¿Alguna vez has querido sencillamente morir? ¿Alguna vez has pensado en la liberación que sería, en todo el dolor que desaparecería? Pero como decía, es sólo un pensamiento”

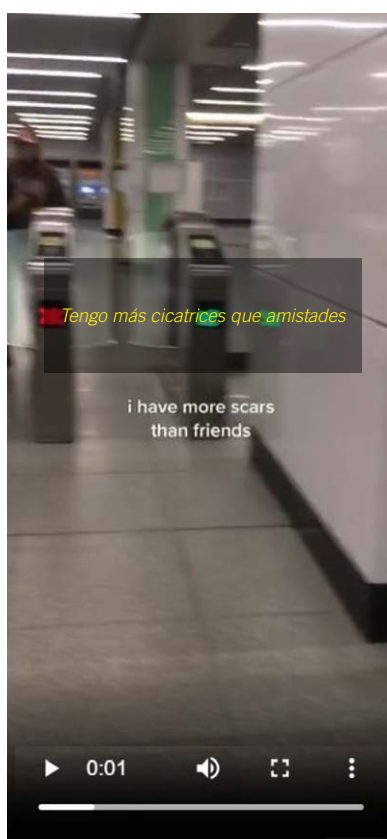
EMPUJADOS A LA OSCURIDAD

EL FEED “PARA TI” DE TIKTOK FOMENTA LA AUTOLESIÓN Y LA IDEACIÓN SUICIDA

Amnistía Internacional



Las publicaciones anteriores recomendadas a las cuentas de adolescentes expresan pensamientos suicidas de modos que idealizan o alientan el suicidio. El vídeo de la derecha es una escena de una película con el comentario superpuesto de la persona autora del vídeo. Muestra a una chica angustiada, que camina hacia los vehículos que se aproximan en una autopista, con gritos de “¡Vamos!”. Se corta justo antes del supuesto choque. La publicación se recomendó a la cuenta de Estados Unidos administrada manualmente.

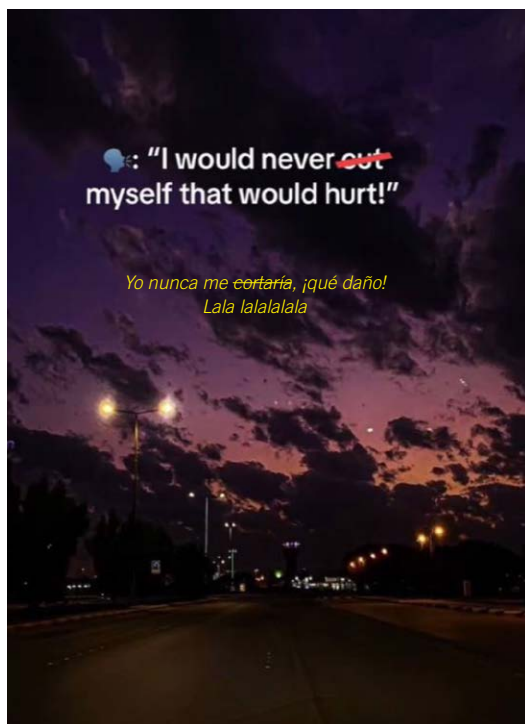


Publicaciones en las que se habla de autolesiones en respuesta a la ira o la decepción por la exclusión social. El vídeo de la derecha parece haberse grabado en Londres. Se recomendó a una cuenta de Kenia entre vídeos de una persona profesional de Reino Unido dedicada a la creación de contenidos sobre salud mental. A las cuentas de Kenia se les mostraban a menudo publicaciones creadas en Reino Unido y Estados Unidos.

EMPUJADOS A LA OSCURIDAD

EL FEED “PARA TI” DE TIKTOK FOMENTA LA AUTOLESIÓN Y LA IDEACIÓN SUICIDA

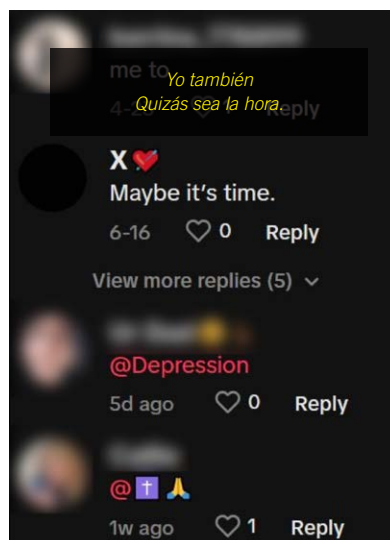
Amnistía Internacional



Una publicación que trivializa las autolesiones. El vídeo reutiliza un sonido común, dos voces que cantan “la la lalalalala”. El sonido se corta en la segunda imagen. El vídeo se recomendó a una cuenta de Estados Unidos (subgrupo de conducta depresiva/canalización) y se etiquetó con #ventaccount #mentalhealthmatters #ventacc #vent #deepvibes #sadtok #sadtiktok #sadvibes #MentalHealth #nightvibes #chilltok #feelinglonely #trustissues #chilltiktok #iwanttobealone #help. Ya no estaba disponible (lo retiró quien lo había creado o TikTok) cuando lo comprobamos unas semanas después de que se recomendara a la cuenta.



“Empecé a pensar que la vida de todo el mundo sería mejor sin mí”



EMPUJADOS A LA OSCURIDAD

EL FEED “PARA TI” DE TIKTOK FOMENTA LA AUTOLESIÓN Y LA IDEACIÓN SUICIDA

Amnistía Internacional

Estos ejemplos ilustran el tipo de publicaciones a que pueden exponerse niños, niñas y jóvenes si muestran interés por contenidos relativos a la salud mental. Las publicaciones que idealizan el pensamiento depresivo y suicida o que trivializan las autolesiones, así como la publicación y los comentarios anteriores, que parecen fomentar las autolesiones y el suicidio, dan una idea de los riesgos sistémicos detectados en este estudio. El sistema de recomendación de TikTok amplificó muchas de estas publicaciones hasta el punto de atraer cientos de miles de visitas.

En su conjunto, estos resultados demuestran que TikTok no se limita a no moderar adecuadamente contenidos individuales, sino que su sistema de recomendación amplifica dichos contenidos activamente a gran escala, con lo que se arriesga a agravar problemas de salud mental en niños, niñas y jóvenes.

6.8 LIMITACIONES DEL PROYECTO

Una limitación clave de la investigación técnica fue el enfoque centrado exclusivamente en contenidos en lengua inglesa. Aunque muchas de las entrevistas con niños, niñas y jóvenes de Kenia y Filipinas señalan el inglés como principal lengua de interacción con la plataforma, muchas personas jóvenes también publican, buscan y consumen contenidos en otras lenguas regionales. El discurso sobre la salud mental, vinculado a un lenguaje más clínico, parece compartirse predominantemente en inglés en ambos países,²¹² pero las publicaciones de las personas usuarias sobre experiencias vividas también podrían crearse y compartirse en otros idiomas. En informes anteriores se ha señalado que TikTok y Meta dedican muchos menos recursos a la moderación de los contenidos que no están en inglés.²¹³ Por lo tanto, es posible que el personal investigador hubiese observado mayores tasas de contenidos problemáticos y nocivos amplificados en idiomas distintos del inglés.

LIMITACIONES DE LAS AUDITORÍAS AUTOMATIZADAS

Como se ha señalado anteriormente en este informe, TikTok y otras plataformas de redes sociales han puesto obstáculos a la investigación independiente, por ejemplo prohibiendo la recopilación automatizada de datos (*scraping*) en sus Términos de Servicio. Durante la investigación técnica de Amnistía Internacional, se suspendieron muchas de las cuentas de Kenia —posiblemente debido al uso de una red privada virtual para simular la ubicación— antes de llegar al final de nuestro periodo de observación y, por tanto, hubo que excluirlas del análisis.

Más allá de estas complicaciones, el problema central de las cuentas automatizadas —para llevar a cabo un experimento como éste a gran escala sin exponer al personal investigador a la tensión de tener que interactuar con contenido angustioso durante largos periodos— radica en las limitaciones de los bots para representar la conducta humana. El experimento manual a pequeña escala puso de manifiesto dichas limitaciones; mientras que un bot sólo puede reaccionar ante rasgos preidentificados de vídeos relevantes (limitados en este caso a términos y hashtags en las descripciones de los vídeos), el personal investigador o las personas usuarias identificarían el contenido relevante basándose en el conjunto formado por las descripciones, el contenido hablado, los elementos visuales, el texto mostrado, la música, los sonidos y el nombre del perfil. Por tanto, es probable que las cuentas

212. Este hecho parece estar relacionado con factores tanto socioeconómicos como culturales. Por ejemplo, el tagalo, una de las lenguas más habladas en Filipinas, no dispone de una traducción exacta para la palabra “depresión”. El término más próximo que se utiliza en tagalo es la palabra referente a “tristeza”. Es más probable que terapeutas y pacientes, a menudo de clase acomodada, utilicen el inglés en el ámbito de las conversaciones sobre salud mental. Las organizaciones locales sin ánimo de lucro dedicadas a la salud mental con las que hablamos en este proyecto y que utilizan TikTok de forma activa también publican en inglés.

213. Mozilla Foundation, *From Dance App to Political Mercenary: How Disinformation on TikTok Gaslights Political Tensions in Kenya*, 8 de junio de 2022, <https://foundation.mozilla.org/en/blog/new-research-disinformation-on-tiktok-gaslights-political-tensions-ahead-of-kenyas-2022-elections>; Amnesty International, *Myanmar: La atrocidad social* (antes citado).

automatizadas no manifestaran interés en muchas publicaciones que sí eran relevantes, entre otras cosas porque el análisis manual y la exploración manual previa mostraron que algunos de los contenidos más extremos relativos a las autolesiones y el suicidio no incluyen descripciones ni hashtags, y en su lugar a menudo parecen asociarse a otros contenidos relevantes a través de sonidos comunes.

FALLOS EMPRESARIALES

7.1 FALTA DE LA DILIGENCIA DEBIDA ADECUADA

El caso de Molly Russell, la niña británica de 14 años que murió por autolesiones tras haber visto contenidos depresivos en Instagram, muestra de la forma más trágica que exponer a una persona joven con síntomas depresivos a un feed de redes sociales compuesto por un alto volumen de publicaciones que normalizan o incluso idealizan el pensamiento depresivo, las autolesiones y el suicidio tiene el potencial de intensificar problemas de salud mental preexistentes en personas usuarias jóvenes y puede contribuir potencialmente a la comisión de actos nocivos e incluso devastadores en el mundo físico.²¹⁴ Como ya se ha comentado, los sistemas de recomendación algorítmicos que emplean TikTok y otras empresas de redes sociales para maximizar la interacción de las personas usuarias tienen el potencial de inferir los pensamientos y emociones de la persona usuaria. A continuación, el sistema de recomendación responde a estos estados mentales inferidos mostrando contenidos más personalizados y atractivos que predice que atraerán a la persona usuaria, con potencial para perpetuar el estado mental o emocional inferido y atrapar a las personas usuarias en un bucle de retroalimentación. De este modo, el sistema de recomendación tiene el potencial de influir, interferir o incluso manipular los pensamientos de un niño, una niña o una persona joven. Este hecho representa una forma de poder sobre niños, niñas y jóvenes mucho más selectiva y personal que la ejercida anteriormente por los medios de comunicación tradicionales, los cuales difundían sus contenidos a un público amplio.

Tal interferencia o manipulación constituye un abuso contra el derecho a la libertad de pensamiento y tiene el potencial de exacerbar los problemas de salud mental preexistentes en personas usuarias jóvenes, lo que supone graves riesgos para la salud física e incluso para la vida.²¹⁵ Dadas las bien documentadas vulnerabilidades emocionales de niños, niñas y jóvenes,²¹⁶ casos como el de Molly Russell, y la amplia evidencia basada en informes previos de la sociedad civil y los medios de comunicación, por no mencionar las revelaciones contenidas en los “Papeles de Facebook”, TikTok debería saber y haber identificado que sus sistemas de recomendación algorítmicos crean el riesgo de exponer a las personas usuarias jóvenes a espirales de publicaciones potencialmente nocivas, que podrían exacerbar problemas de salud mental preexistentes.

Como se ha comentado en el apartado 4.4, para cumplir con sus responsabilidades según lo establecido en los Principios rectores de la ONU sobre las empresas y los derechos humanos

214. BBC News, “Molly Russell: Coroner’s report urges social media changes” (antes citado).

215. Katarzyna Kostyrka-Allchorne, Mariya Stoilova et al., “Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and nonsuicidal self-injury conditions – a systematic review”, febrero de 2023, *Child and Adolescent Mental Health*, volumen 28, número 1 (antes citado); North London Coroner’s Service, “Regulation 28 report to prevent future deaths”, 13 de octubre de 2022 (antes citado).

216. University of Rochester Medical Centre, “Understanding the teen brain”, 2023 (antes citado); Cirujano general de Estados Unidos, *Social Media and Youth Mental Health* (antes citado).

(Principios rectores de la ONU), TikTok debería llevar a cabo la adecuada diligencia debida en materia de derechos humanos para identificar, prevenir, mitigar y dar cuenta del modo en que aborda los daños potenciales y reales.

En su respuesta escrita a las preguntas que envió Amnistía Internacional el 12 de julio de 2023 a TikTok, la empresa no reveló qué riesgos específicos, incluidos los relativos a los derechos de la infancia, había identificado durante la fase de diseño y desarrollo al crear la aplicación, ni el proceso que siguió para identificar los riesgos y las respectivas estrategias de mitigación aplicadas antes del lanzamiento. Una vez ya en activo, TikTok debería haber continuado y mantenido una diligencia debida proactiva.

En su carta, la empresa afirma que su enfoque acerca de la seguridad de las personas jóvenes se sustenta en su compromiso con los derechos humanos, en particular con los Principios rectores de la ONU, “y su llamamiento a llevar a cabo la diligencia debida en materia de derechos humanos”. Este compromiso se incluye en su sitio web, pero TikTok no cuenta con una política de derechos humanos a disposición del público, aparte de este compromiso general, la cual es parte esencial de la diligencia debida en materia de derechos humanos, tal como se indica en los Principios rectores de la ONU y en la Guía de la OCDE de debida diligencia.²¹⁷

Amnistía Internacional dio seguimiento a la comunicación por escrito y preguntó a TikTok si cuenta con una política de ese tipo, que abarque un amplio rango de derechos humanos incluido el derecho a la privacidad, a la libertad de pensamiento y a la salud, y si el acceso a esa política es público. El equipo de investigación preguntó también quién es la persona del equipo directivo de la empresa responsable de su aplicación. TikTok respondió señalando a Amnistía Internacional el compromiso con los derechos humanos que aparece en su sitio web y que, según la empresa, se basa en varios marcos de los derechos humanos entre los que figuran los Principios rectores de la ONU, que ha prometido respetar. Sin embargo, TikTok no facilitó la política de derechos humanos que Amnistía Internacional le había solicitado ni la información sobre quién, a nivel directivo, es responsable de su aplicación y cómo se integra, desde ese nivel, en todas las funciones de la empresa.²¹⁸

TikTok sí declaró en su respuesta escrita de 12 de julio de 2023 que cuenta con un equipo centralizado que dirige su labor en materia de derechos humanos, así como con otro personal en la empresa que, según afirma, está “facultado por la dirección para abordar cualquier riesgo identificado”.²¹⁹ Amnistía Internacional preguntó a TikTok sobre el tamaño de este equipo y la persona del equipo directivo de la cual depende. La organización preguntó también sobre la facultad que estas personas tiene para abordar cualquier riesgo identificado, el proceso que se sigue cuando identifican un riesgo y las medidas que pueden tomar. TikTok no respondió a esta pregunta.²²⁰

Las Normas de la comunidad de TikTok incluyen un subapartado sobre las políticas de TikTok en materia de seguridad y bienestar de los menores, supervisadas por una subsección del equipo de Política del Producto sobre Confianza y Seguridad. Además, según TikTok, su “equipo de Equidad en la plataforma está especializado en temas de derechos humanos, equidad e inclusión y aplica un enfoque multifacético a la revisión y reforma de las políticas, las funciones del producto y los sistemas algorítmicos”.

Sin embargo, TikTok no detalló en sus cartas qué implica este “enfoque multifacético”, ni qué riesgos específicos trata de detectar o ha identificado, en particular en relación con su sistema de

217. OCDE, Guía de la OCDE de debida diligencia para una conducta empresarial responsable, <https://mneguidelines.oecd.org/Guia-de-la-OCDE-de-debida-diligencia-para-una-conducta-empresarial-responsable.pdf> (consultado el 18 de julio de 2023)

218. Véase el Anexo 4: Respuesta escrita de TikTok, 29 de octubre de 2023.

219. Véase el Anexo 3 del presente informe: Respuesta escrita de TikTok, 12 de julio de 2023.

220. Véase el Anexo 4: Respuesta escrita de TikTok, 29 de octubre de 2023 (en inglés).

recomendación. Sí afirmó haber incorporado un enfoque de derechos humanos en sus Normas de la comunidad, contar con consejos asesores globales que incluyen a expertos en derechos de la infancia y ser miembro de WeProtect Global Alliance y Tech Coalition,²²¹ donde trata con sus homólogos temas relativos a los riesgos actuales y futuros. En junio de 2023, TikTok anunció también la creación de un Consejo Juvenil,²²² el cual, según afirma, proporcionará “una oportunidad más estructurada y regular para que los jóvenes brinden sus puntos de vista a la plataforma” sobre la creación de políticas, productos y programas.

Estas iniciativas tienen potencial para lograr un mayor relacionamiento continuado por parte de TikTok con titulares de derechos, especialistas, organizaciones de la sociedad civil y otras partes interesadas. Sin embargo, sigue sin estar claro exactamente cómo se utilizará esta relación para sentar las bases de la diligencia debida y las evaluaciones de riesgos de TikTok y con qué frecuencia se consultará a las partes interesadas.

En su respuesta a Amnistía, TikTok no mencionó el riesgo específico para los derechos a la privacidad y a la libertad de pensamiento en relación con sus prácticas de recogida de datos, perfilación y sistemas de recomendación algorítmicos, ni el riesgo para el derecho a la salud que suponen los feeds de contenidos altamente personalizados. La empresa declaró en su carta de 12 de julio que “en el futuro [...] llevará a cabo evaluaciones de impacto periódicas en colaboración con terceros, que pueden adoptar la forma de evaluaciones de toda la empresa, evaluaciones de equipos y evaluaciones específicas de productos o líneas de negocio”. TikTok no proporcionó detalles sobre la fecha de inicio de las evaluaciones o su regularidad ni indicó si se harán públicos los resultados o no. Lo anterior contrasta con las orientaciones de la ONU, que exigen “evaluaciones periódicas y exhaustivas de los efectos [...] sobre tales derechos” (los derechos humanos) como “elemento esencial de los procesos de diligencia debida en materia de derechos humanos”.²²³

Amnistía Internacional solicitó más detalles a TikTok sobre sus políticas y prácticas de diligencia debida en materia de derechos humanos. TikTok respondió que “realiza consultas con una serie de partes interesadas para orientar el ejercicio de la diligencia debida y está aplicando varias recomendaciones sobre las actividades de la empresa en materia de confianza y seguridad, resultantes del trabajo sobre derechos humanos que lleva a cabo con la empresa consultora Article One. El equipo de Equidad en la plataforma aplica esas recomendaciones en colaboración con un grupo de trabajo sobre derechos humanos formado por miembros de los equipos de toda la empresa. La evaluación recomendó que TikTok debe llevar a cabo una evaluación del impacto en materia de derechos de la infancia, que se pondrá en marcha en colaboración con Article One”.²²⁴

Aunque es positivo que cuente con un equipo de trabajo sobre derechos humanos formado por personal de los equipos de toda la empresa y que realice evaluaciones del impacto en materia de derechos de la infancia, el hecho de que TikTok no haya efectuado hasta el momento esas evaluaciones —teniendo en cuenta la enorme popularidad de su plataforma entre las personas menores de 18 años— es un descuido importante. La carencia de esas evaluaciones del impacto pone en duda que TikTok haya podido evaluar adecuadamente los riesgos que su plataforma supone para los niños y niñas, y refleja que la diligencia debida no se ha llevado a cabo de manera adecuada.

221. WeProtect Global Alliance y Tech Coalition son dos iniciativas en las que participan las principales plataformas tecnológicas, cuyo objetivo es proteger a niños y niñas frente a la explotación y los abusos sexuales en Internet.

222. TikTok Newsroom, “TikTok actualiza su funcionalidad de Sincronización Familiar y establece el Consejo Juvenil de TikTok”, 27 de julio de 2023, <https://newsroom.tiktok.com/es-es/tiktok-actualiza-su-funcionalidad-de-sincronizacion-familiar-y-establece-el-consejo-juvenil-de-tiktok>

223. Alta Comisionada de las Naciones Unidas para los Derechos Humanos, El derecho a la privacidad en la era digital: informe, 15 de septiembre de 2021 (antes citado).

224. Véase el Anexo 4: Respuesta escrita de TikTok, 29 de octubre de 2023 (en inglés).

De hecho, TikTok expone en su carta del 29 octubre de 2023 dirigida a Amnistía Internacional que otra recomendación es “elaborar un proceso de diligencia debida en materia de derechos humanos para toda la empresa que incluirá la realización periódica de evaluaciones de impacto en ese ámbito”, que la empresa está ahora elaborando, y “en el que se plantearán los factores que determinan cuándo debe realizarse una evaluación”. El hecho de que TikTok no cuente actualmente con un proceso de diligencia debida en materia de derechos humanos para toda la empresa es un incumplimiento claro de su responsabilidad de respetar los derechos humanos. Tampoco queda claro a qué se refiere la empresa con los factores que determinarán cuándo debe realizarse una evaluación. Teniendo en cuenta los riesgos graves para los derechos humanos de los niños y niñas que supone la plataforma, el gran número de personas menores de edad que la utilizan en todo el mundo y la importancia que tiene en la vida de muchas de ellas, es indispensable que la empresa realice evaluaciones del impacto sobre los riesgos para los derechos de la infancia de forma periódica en el marco de su ejercicio de la diligencia debida y siempre que implemente nuevas características de diseño.

En su respuesta por escrito, TikTok no incluyó una lista de los riesgos concretos identificados. Teniendo en cuenta que TikTok empieza ahora a elaborar un proceso de diligencia debida en materia de derechos humanos que, según dice, cumplirá las normas internacionales de derechos humanos, incluidos los Principios rectores de la ONU, es lógico que la empresa no haya facilitado una lista de los riesgos concretos que ha identificado ni las medidas adoptadas para prevenirlos y mitigarlos porque, hasta la fecha, viene incumpliendo su responsabilidad de respetar los derechos humanos mediante el ejercicio adecuado de la diligencia debida en materia de derechos humanos.

Esta falta de un proceso de diligencia debida adecuado es incoherente con los Principios rectores de la ONU, que exigen “que las empresas cuenten con políticas y procesos para saber y hacer saber que respetan los derechos humanos en la práctica”.²²⁵ Los Principios rectores de la ONU también aclaran que “hacer saber implica comunicar, ofrecer transparencia y rendir cuentas a las personas o grupos que puedan verse afectados”.²²⁶ Otro elemento esencial de la diligencia debida, como se indica en la Guía de la OCDE de Devida Diligencia, implica comunicar cómo se han abordado los impactos. En sus cartas, TikTok señalaba distintas medidas encaminadas a promover la seguridad, la salud y el bienestar infantiles, aunque algunas de ellas no abordan la cuestiones tratadas en este informe.²²⁷ Además, como se comenta a continuación, muchas de las medidas que la empresa ha aplicado, y que pueden tener por objeto abordar cualquier posible riesgo de enviar a niños, niñas y jóvenes a espirales de contenido nocivo para la salud mental, son inadecuadas.

7.2 RIESGOS GRAVES COMBINADOS CON RESPUESTAS INADECUADAS

Al responder a preguntas concretas sobre las medidas que toma TikTok para prevenir los riesgos para la salud mental de las personas jóvenes, en su respuesta escrita a Amnistía Internacional hizo referencia a sus Normas de la comunidad, que establecen qué tipos de contenido están prohibidos y, por tanto, en caso de que se denuncien o detecten de otro modo, se eliminan de la plataforma.²²⁸ Entre ellas se la prohibición de “Mostrar, promover o proporcionar instrucciones de suicidio y autolesiones, o cualquier desafío, reto, juego o pacto relacionado”, “Mostrar o promover las bromas sobre el suicidio y las autolesiones” y “Compartir planes de suicidio y autolesiones”.²²⁹

225. Principios rectores de la ONU (antes citados), Principio 21 y comentario.

226. Principios rectores de la ONU (antes citados), Principio 21 y comentario.

227. OCDE, Guía de la OCDE de Devida Diligencia para una Conducta Empresarial Responsable (antes citada).

228. Véase el Anexo de este informe: Respuesta escrita de TikTok, 12 de julio de 2023.

229. TikTok, “Normas de la comunidad”, actualizadas en marzo de 2023, tiktok.com/community-guidelines/es/mental-behavioral-health/

Sin embargo, el análisis manual llevado a cabo por Amnistía Internacional sobre muestras de publicaciones incluidas en los feeds “Para ti” de cuentas que denotaban interés por la salud mental descubrió que aproximadamente un 1% de los vídeos recomendados fomentaba las autolesiones o el suicidio y otro 3% glorificaba, idealizaba o trivializaba la depresión, la autolesión o el suicidio. Estos vídeos no sólo estaban disponibles para las personas usuarias que buscaban este tipo de contenidos, sino que —como se muestra en nuestra investigación— el sistema de recomendación de TikTok también los amplificaba activamente.

Aunque TikTok nunca ha reconocido públicamente los riesgos para los derechos humanos asociados a la creación de un feed altamente personalizado, sí ha introducido herramientas que parecen sustentarse en la posibilidad de que existan tales riesgos. En 2021, la empresa anunció que estaba trabajando para diversificar las recomendaciones de contenidos con el fin de evitar la producción de feeds excesivamente temáticos, al afirmar: “En TikTok, reconocemos que demasiado de cualquier tema —ya sean animales, consejos de fitness o proyectos de bienestar personal— no encaja con la experiencia de descubrimiento variada que pretendemos crear”.²³⁰

En julio de 2022, se informó de que TikTok estaba a punto para desplegar una herramienta que facilitara esa diversificación.²³¹ En su carta de julio a Amnistía Internacional, TikTok también afirmó que “interrumpe los patrones repetitivos, de modo que el contenido que puede ser adecuado si se ve de forma ocasional, como el relativo a fitness extremo o dietas, no se visualice en bloques, que podrían ser más problemáticos”.²³²

Sin embargo, la investigación de Amnistía Internacional no halló pruebas de que las espirales descendentes a las que se enviaron las cuentas fueran interrumpidas por contenidos diversificados hasta el punto de poder mitigar los riesgos. Por el contrario, el personal investigador observó picos repetidos en el efecto espiral a lo largo del tiempo (véase, por ejemplo, el gráfico 3 del apartado 6.6.1).

TikTok introdujo otra herramienta nueva en 2022 para que las personas usuarias pudieran especificar palabras o hashtags para excluir de su feed los vídeos etiquetados con esos términos, como complemento del botón “No me interesa” ya existente, a través del cual las personas usuarias pueden expresar el deseo de que no se les recomienden vídeos similares. Las personas usuarias han puesto en duda la eficacia de la herramienta “No me interesa”,²³³ observación compartida por varias de las personas jóvenes participantes en la investigación. Que dijeron que habían utilizado la herramienta con la esperanza de evitar más contenidos provocadores, que percibían como nocivos para su salud mental, pero que descubrieron que su efecto era limitado en recomendaciones posteriores. En palabras de “Nikki”, activista de 24 años de Manila: “Me rendí. Ya no hago clic en ellas [las herramientas de control de la persona usuaria en distintas plataformas], porque por mucho que haga clic, denuncie y oculte, al final vuelven a aparecer en mi feed”.²³⁴

La investigación de Amnistía Internacional no puso a prueba la eficacia de la nueva herramienta de filtrado de palabras clave, pero la gran variedad de hashtags y términos utilizados por la “comunidad creadora” de TikTok para contenidos relacionados con la ansiedad, la depresión, tipos de autolesiones y el suicidio —incluidas faltas de ortografía deliberadas y códigos— probablemente limitarían su eficacia.

La carta de respuesta a nuestros hallazgos que la empresa envió en octubre de 2023 señala también la función de “sincronización familiar”, que permite a los progenitores personalizar el límite de uso

230. TikTok, “An update on our work to safeguard and diversify recommendations”, 16 de diciembre de 2021, <https://newsroom.tiktok.com/en-us/an-update-on-our-work-to-safeguard-and-diversify-recommendations>

231. The Verge, “TikTok is giving you new ways to control your For You page”, 13 de julio de 2022, <https://www.theverge.com/2022/7/13/23205795/tiktok-algorithm-hashtag-filters-safety-content-levels>

232. Véase el Anexo de este informe: Respuesta escrita de TikTok, 12 de julio de 2023.

233. Mashable, “TikTok doesn’t seem to care if I’m ‘not interested’”, 28 de abril de 2022, <https://mashable.com/article/tiktok-not-interested>

234. Nikki (seudónimo), 24 años, Manila, entrevistada el 8 de mayo de 2023.

diario de sus hijos e hijas y utilizar la herramienta de filtrado de palabras clave para que dejen de ver contenidos relacionados con palabras clave en sus feeds “Para ti” y “Siguiendo”.²³⁵ Esos controles parentales pueden agregar una capa de protección, sobre todo si los niños y niñas se consideran incapaces de adoptar medidas protectoras por sí mismos, aunque sea limitada por la misma dificultad descrita anteriormente de excluir todos los términos que guardan relación con un tema o un tipo de contenidos. Sin embargo, los controles parentales también crean nuevas complejidades y pueden entrañar riesgos para el cumplimiento de los derechos de la infancia en situaciones en las que el interés superior del niño o niña —por ejemplo para acceder a información sobre un tema concreto— no se ajuste a las creencias y opiniones de los progenitores.

En su respuesta escrita a Amnistía Internacional,²³⁶ TikTok también destacó su función “Actualizar”, que permite a las personas usuarias restablecer su feed y entrenar de nuevo al sistema de recomendación.²³⁷ Sin embargo, una vez más, los resultados de la investigación de Amnistía Internacional parecen sugerir que esta función tendría una utilidad limitada para un niño, una niña o una persona joven que manifieste interés en la salud mental, dado que probablemente volverían a entrenar al sistema de recomendación para que amplificara contenidos potencialmente nocivos en un breve periodo, a menos que omitieran activamente todos los contenidos relativos a la salud mental. Además, aunque omitieran el contenido sobre salud mental, los resultados de la investigación técnica efectuada como parte de la investigación sugieren que las personas usuarias que no muestran interés activo en dichos contenidos siguen recibiendo recomendaciones de contenidos relacionados, aunque menos que quienes manifiestan interés activamente.

Por último, el personal investigador también puso a prueba la utilidad de la herramienta de transparencia “Por qué estás viendo este vídeo” de la aplicación de TikTok, que se supone, según TikTok, que “ayuda a dar más contexto a los contenidos recomendados en los feeds Para ti”.²³⁸ Sin embargo, los paneles de información que se mostraron al personal investigador como parte de los experimentos manuales sólo incluían explicaciones simplistas e inespecíficas como “este vídeo es popular en tu país” o “este vídeo es más largo, parece que te gustan los vídeos más largos”, poco útiles para explicar por qué se había mostrado un vídeo en concreto a la cuenta o cómo había evolucionado el efecto de personalización de la cuenta en cuestión, que había experimentado un pronunciado efecto espiral.

Es evidente que se necesitan herramientas de control más eficaces para las personas usuarias. Sin embargo, aunque estas herramientas fueran eficaces, son medidas de protección aisladas que trasladan la responsabilidad de crear un entorno seguro a quien las usa, incluidas personas jóvenes y vulnerables que tal vez no sean capaces de distanciarse activamente de contenidos provocadores o nocivos. Por tanto, constituyen una respuesta inadecuada a los riesgos sistémicos asociados al sistema de recomendación de TikTok.

TikTok no aplica las medidas que cabe esperar de una diligencia debida adecuada en materia de derechos humanos y, como demuestra la investigación de Amnistía Internacional, no previene los graves riesgos que supone su plataforma para la salud mental y física de personas usuarias jóvenes con problemas de salud mental ya existentes.

235. Véanse el Anexo 4 del presente informe: Respuesta escrita de TikTok, 29 de octubre de 2023; y TikTok, “Seguridad del usuario: ¿Qué es la sincronización familiar?”, consultado el 30 de octubre de 2023, <https://support.tiktok.com/es/safety-hc/account-and-user-safety/user-safety#4>

236. Véase el Anexo 3 del presente informe: Respuesta escrita de TikTok, 12 de julio de 2023.

237. TikTok, “TikTok presenta una nueva forma de actualizar Feed Para Ti”, 16 de marzo de 2023, <https://newsroom.tiktok.com/es-es/tiktok-presenta-una-nueva-forma-de-actualizar-feed-para-ti>

238. TikTok, “Descubre por qué se te recomienda un vídeo en el Feed Para Ti”, 20 de diciembre de 2022, <https://newsroom.tiktok.com/es-es/por-que-este-video>

8 CONCLUSIÓN Y RECOMENDACIONES

CONCLUSIÓN

La página “Para ti” de TikTok y el sistema algorítmico de recomendación que hay detrás de ella han ayudado a catapultar la plataforma a su ubicuidad actual en la vida de los niños, niñas y jóvenes. Mientras que quienes trabajan en marketing se asombran de su capacidad para mantener a la gente enganchada y las plataformas competidoras de TikTok se afanan por emular las características de diseño que impulsan su éxito, los niños, niñas y jóvenes quedan expuestos a un sistema que convierte sus vulnerabilidades psicológicas en un medio para aumentar al máximo la “interacción de las personas usuarias”.

Mediante la personalización impecable y en tiempo récord del feed de las personas usuarias, TikTok ha creado una plataforma muy adictiva y expone a quienes la utilizan a importantes riesgos para la salud. TikTok resulta particularmente peligrosa para los niños, niñas y jóvenes de todo el mundo que sufren depresión, ansiedad y otros problemas de salud mental porque, como muestra este informe, su feed “Para ti” puede hacer caer rápidamente a quienes manifiestan un interés por la salud mental en espirales de contenidos depresivos, incluidos vídeos que idealizan, trivializan y fomentan, las autolesiones y el suicidio.

Pese al mayor control de los sistemas algorítmicos de las plataformas de redes sociales existente tras la publicación de los “Papeles de Facebook” y la investigación del pesquisidor de Reino Unido sobre la muerte de Molly Russell, TikTok y otras plataformas de redes sociales continúan persiguiendo la interacción de las personas usuarias independientemente del daño que puedan provocarles. Lo que hacen, en cambio, es priorizar la viralidad y el tiempo dedicado a la plataforma sobre la seguridad de sus sistemas. Con ello, TikTok y otros gigantes de las redes sociales han normalizado la vigilancia y manipulación de las personas usuarias de todo el mundo, y menoscaban e incumplen el derecho de éstas a la privacidad y la libertad de pensamiento.

Estos abusos contra los derechos humanos pueden parecer abstractos e intangibles, pero tienen consecuencias en la vida real de los niños, niñas y jóvenes de todo el mundo. Al convertir la vigilancia con fines comerciales en un medio para que la vista se mantenga fija en la pantalla, en algunos casos mediante un flujo de contenidos depresivos y prohibidos de salud mental, TikTok pone a las personas jóvenes que la utilizan en situación de riesgo para la salud física y mental.

Aún no se conoce por completo el alcance total de estos riesgos, sobre todo porque los datos de investigación y el interés de las empresas presentan un injusto sesgo hacia Europa y América del Norte. Sin embargo, TikTok continúa creciendo y, cada día, miles de niños, niñas y jóvenes más se registran en una plataforma que podría hacer su vida menos segura.

TikTok debe revisar de forma urgente sus procesos de recopilación de datos y amplificación, y ejercer de manera integral la diligencia debida en materia de derechos humanos. No obstante, las medidas particulares de tan sólo una empresa no bastan para refrenar un modelo empresarial que es intrínsecamente incompatible con los derechos humanos. Por consiguiente, los Estados deben controlar las “grandes empresas tecnológicas” como TikTok de conformidad con las leyes y normas internacionales de derechos humanos a fin de proteger y hacer efectivos los derechos de los niños, niñas y jóvenes.

Esto requiere un cambio de enfoque por parte de los Estados, que durante demasiado tiempo se han centrado primordialmente en exigir a las plataformas que eliminen el contenido ilegal o —en el caso de los gobiernos más represivos— indeseado de sus plataformas. En lugar de ello, los Estados deben abordar la causa fundamental de la proliferación de contenidos nocivos online, que es el modelo empresarial basado en la maximización de la interacción en detrimento de los derechos humanos de las personas usuarias. Mientras se sigue debatiendo a nivel mundial sobre los riesgos de la inteligencia artificial, es fundamental que todos los gobiernos adopten e implementen normas para prevenir y mitigar los daños que provocan algunos de los sistemas logarítmicos más generalizados que ya están dominando nuestros espacios públicos digitales.

RECOMENDACIONES

RECOMENDACIONES PARA TIKTOK

- TikTok y otras empresas tecnológicas que dependen de la invasiva actividad basada en datos, constitutiva de vigilancia empresarial masiva, deben cambiar progresivamente a un modelo de empresa respetuoso con los derechos humanos. Como primera medida, deben garantizar que sus políticas y procesos de diligencia debida en materia de derechos humanos abordan los efectos generalizados y sistémicos que sus modelos empresariales tienen en los derechos humanos, concretamente el derecho a la privacidad, el derecho a la libertad de opinión y de pensamiento y el derecho a la salud. Y, además, deben ser transparentes en cuanto a los riesgos, incluidos los relativos a los derechos humanos, que han identificado y la manera en que se han abordado.
- TikTok debe dejar de aumentar al máximo la “interacción de las personas usuarias” en detrimento de la salud y otros derechos humanos de quienes las utilizan, a la vista de la evidencia disponible de las consecuencias negativas que el uso compulsivo de las plataformas tiene especialmente para la gente joven. En el marco de su proceso de diligencia debida en materia de derechos humanos, TikTok, en colaboración con las personas usuarias —incluidos niños, niñas y jóvenes— y especialistas independientes, debe identificar los elementos de diseño que estimulan el uso adictivo de las plataformas y la comparación social, y sustituirlos por una utilización que se centre en la “la seguridad desde el diseño” y el interés superior de la infancia.
- TikTok debe ejercer la diligencia debida en materia de derechos humanos de forma continua y proactiva durante todo el ciclo vital de las tecnologías algorítmicas, antes y después de la implantación de sistemas y características de diseño nuevos, para que los riesgos puedan identificarse durante la fase de desarrollo y los abusos contra los derechos humanos y otros efectos nocivos puedan detectarse de forma inmediata en cuanto esas tecnologías se hayan puesto en marcha.
- TikTok debe involucrar a los niños, niñas y jóvenes, a personas expertas del mundo académico y la sociedad civil y a otras partes interesadas pertinentes en sus procesos continuos de diligencia debida en materia de derechos humanos. Los niños, niñas y jóvenes también deberían

desempeñar un papel fundamental en la implementación de “la seguridad desde el diseño” mediante su participación en el proceso de desarrollo de las herramientas y características de las plataformas de redes sociales.

- Las evaluaciones del impacto sobre los derechos humanos deberían hacerse públicas con regularidad y deberían incluir información sobre los riesgos y las medidas de mitigación adoptadas con respecto a determinados países (especialmente si los sistemas pueden tener un mayor efecto debido a situaciones de conflicto político o emergencia humanitaria), determinadas categorías de personas usuarias tales como los niños, niñas y jóvenes, y determinados cambios en los productos.
- Para respetar la privacidad y proporcionar a las personas usuarias la posibilidad real de elección y control, la existencia de un ecosistema de redes sociales exento del uso de perfiles no debería simplemente ser una opción sino lo normal. Por consiguiente, los algoritmos determinantes de contenidos que usan TikTok y otras plataformas online no deberían basarse por defecto en la elaboración de perfiles (por ejemplo, a partir del tiempo de visión o la interacción); y deben solicitar la inclusión voluntaria en lugar de la exclusión voluntaria, de manera tal que el consentimiento de inclusión voluntaria se otorgue libremente y sea específico, informado (lo que incluye utilizar un lenguaje adaptado a la infancia) e inequívoco.
- TikTok debería dejar de recopilar datos personales privados y hacer inferencias sobre los intereses, el estado emocional o el bienestar de las personas usuarias a partir del tiempo de visión y su interacción con la finalidad de “personalizar” las recomendaciones de contenido y segmentar la publicidad. En lugar de utilizar la vigilancia generalizada para adaptar los feeds a los intereses de las personas usuarias, TikTok debería permitir a éstas expresar sus intereses mediante avisos explícitos (por ejemplo, se les podría pedir que indicasen intereses concretos si quisieran recibir recomendaciones personalizadas) y únicamente si ello fuera acompañado de su consentimiento informado, específico y libremente otorgado.
- TikTok debe aplicar medidas adicionales para impedir que la personas usuarias en situación de riesgo caigan en constantes de uso compulsivo y espirales de contenidos potencialmente nocivos. Estas medidas podrían consistir en un límite diario obligatorio sobre el número de recomendaciones personalizadas ofrecidas a los niños y niñas y un listado actualizado de términos relacionados con los contenidos dudosos sobre salud mental, que se consideren aptos para realizar búsquedas sobre ellos pero no para amplificarlos en el feed “Para ti”.
- Introducir medidas de “fricción” como, por ejemplo, una estrategia de mitigación. En el marco de sus procesos de diligencia debida en materia de derechos humanos, TikTok debería invertir en investigación para identificar e incorporar medidas orientadas a limitar la rápida y a menudo desproporcionada amplificación algorítmica de los contenidos dudosos.
- Como medida provisional, mejorar con urgencia la eficacia de las actuaciones orientadas a diversificar las recomendaciones de contenido en el feed “Para ti”, lo que incluye poner herramientas de control eficaces a disposición de las personas usuarias. Estas herramientas deberían resultar fáciles de encontrar y comprender, y deberían ofrecer una forma eficaz de suprimir recomendaciones de contenido futuras relativas a determinados temas, etiquetas o personas usuarias.
- Mejorar radicalmente la transparencia en cuanto a la utilización de algoritmos que determinan y moderan los contenidos, garantizando que su mecánica puede consultarse libremente y también se explica a lo largo del proceso de uso de la plataforma y en términos claramente comprensibles y adaptados a la infancia en todos los idiomas pertinentes.

- Garantizar la coherencia en la toma de decisiones relacionada con la moderación de contenidos, y también una supervisión humana adecuada de los contenidos automatizados y la inversión suficiente en la dotación de recursos para la moderación en todos los idiomas.
- TikTok debería generar campañas públicas e informar en su plataforma sobre las diferentes características de seguridad que las personas usuarias pueden habilitar. Estas campañas podrían promocionarse entre las personas usuarias a través de diferentes canales, como publicaciones promovidas en los feeds y notificaciones internas de la aplicación que animen a aprender a usar con confianza distintas herramientas de seguridad.
- Permitir a las personas investigadoras independientes acceder y examinar todos los sistemas algorítmicos pertinentes y, cuando sea necesario, acceder a la información necesaria para realizar investigaciones independientes sobre los riesgos sistémicos y las vulneraciones de los derechos humanos. La Ley de Servicios Digitales de la UE ha creado un marco de acceso para la investigación que debería servir de modelo para los marcos de acceso para la investigación de TikTok y otras grandes plataformas de redes sociales en cualquier otro lugar. Cuando la ausencia de regulación suponga la inexistencia de directrices claras sobre los criterios de acceso y los procesos de verificación de contenidos, TikTok y las demás redes sociales deberían tratar de crear organismos consultivos independientes que puedan garantizar un proceso eficaz de verificación.

RECOMENDACIONES PARA LOS ESTADOS

Recomendaciones para que la protección de datos y la regulación de las plataformas sean efectivas

Los Estados deben:

- Garantizar que el acceso a los servicios e infraestructuras digitales básicos tales como TikTok y otras plataformas de redes sociales, y su uso, no se basen en la vigilancia omnipresente de los niños, niñas, jóvenes y personas adultas que los utilizan. Para ello será preciso promulgar y/o aplicar una legislación integral sobre la protección de datos, de conformidad con las leyes y normas internacionales, para prohibir la publicidad dirigida que se basa en prácticas de seguimiento invasivas. Estas leyes deberían restringir la cantidad y el alcance de los datos personales que pueden recopilarse, limitar de forma estricta la finalidad para la que las empresas tratan esos datos y garantizar que las inferencias sobre las personas obtenidas a parte de la recopilación y tratamiento de los datos personales quedan protegidas. Además, deberían exigir que las empresas ofrezcan información clara desde un principio a las personas usuarias sobre la finalidad de la recolección de sus datos personales y que no traten posteriormente esos datos de forma incompatible con finalidad expresada o la responsabilidad que tienen de respetar los derechos humanos.
- Como primera medida, deben impedir que las empresas condicionen el acceso a sus servicios a que las personas usuarias “den su consentimiento” a la recopilación, tratamiento y difusión de sus datos personales para la segmentación de contenidos o con fines comerciales o publicitarios.
- Regular las empresas de redes sociales para que garanticen que los algoritmos que usan las plataformas online a fin de determinar los contenidos no se basan por defecto en la elaboración de perfiles; y que esos algoritmos solicitan la inclusión voluntaria en lugar de la exclusión voluntaria, de manera tal que el consentimiento de inclusión voluntaria se otorga libremente y es específico, informado e inequívoco. La recopilación y el uso de los datos personales sensibles inferidos (por ejemplo, las recomendaciones basadas en el tiempo de visión y los “me gusta” que permiten inferir información sensible) para personalizar los anuncios y los contenidos deberían estar prohibidos. En cambio, las personas usuarias deberían controlar los intereses declarados

o manifestados que desean que la plataforma tome en cuenta en la configuración de su feed. Quienes prefieran un feed basado en recomendaciones personalizadas deben tener la posibilidad de comunicar sus intereses personales a la plataforma basándose en el consentimiento expreso, libremente otorgado e informado, y en avisos expresados en lenguaje adaptado a la infancia.

- Los procesos normativos deben conllevar consultas significativas con los grupos afectados, incluidos los niños, niñas y jóvenes, así como con especialistas independientes y organizaciones de la sociedad civil.
- Garantizar que se creen organismos nacionales reguladores de la protección de datos, que la independencia de esos organismos se garantiza por ley y que se los dota de suficientes recursos, conocimientos técnicos y facultades para investigar y sancionar de forma significativa —de conformidad con las leyes y normas internacionales— los abusos de los reglamentos de las empresas de redes sociales. Esos organismos deben poder garantizar la supervisión eficaz e independiente del diseño de la plataforma, así como del diseño, el desarrollo y la implementación de los sistemas algorítmicos, a fin de garantizar que las empresas rinden cuentas ante la justicia por la identificación, prevención y mitigación de las vulneraciones de los derechos humanos asociadas con esos sistemas.
- Promulgar o aplicar marcos normativos para garantizar que las personas pueden ejercer en la práctica su derecho a elegir modelos empresariales respetuosos con los derechos humanos alternativos a los basados en la vigilancia. Esto incluye adoptar medidas para garantizar la interoperabilidad (la posibilidad de comunicarse con los contactos existentes utilizando otra plataforma compatible) y no la simple portabilidad de los datos, de modo que las personas puedan cambiar de unos servicios a otros sin sufrir perjuicios sociales, y para reducir los efectos de red.
- Exigir a TikTok y otras empresas de redes sociales que ofrezcan explicaciones adecuadas a cada edad para los niños y niñas, o los progenitores o personas cuidadoras de niños y niñas de muy corta edad, de sus condiciones de servicio. En el caso de los niños y niñas, estas explicaciones deberían expresarse en un lenguaje sencillo y claro, ofrecer información transparente a lo largo del proceso de uso y no sólo al principio, y ofrecer explicaciones claras de las opciones de control por parte de las personas usuarias y la configuración predeterminada. Las empresas de redes sociales deberían también estar obligadas a facilitar soluciones no textuales —como imágenes, vídeos y animaciones— para ayudar a comprender sus condiciones de servicio, y deberían estar fácilmente localizables para aclarar posible dudas.
- Otorgar al interés superior de la infancia una consideración primordial a la hora de regular las acciones de publicidad y marketing dirigidas a personas menores de edad y accesibles para éstas. El patrocinio, la publicidad indirecta y todas las demás formas de contenidos con fines comerciales deben estar claramente diferenciados de todos los demás contenidos y deben evitar perpetuar los estereotipos discriminatorios (como los basados en el género, el origen racial, la edad, la situación de discapacidad, etc.).

Recomendaciones sobre la debida diligencia debida en materia de derechos humanos

Los Estados deben:

- Exigir por ley a las empresas tecnológicas que ejerzan de forma continua y proactiva la diligencia debida en materia de derechos humanos con el objetivo de identificar y abordar los riesgos y los efectos negativos para los derechos humanos de sus actividades a nivel global, incluidos los vinculados a sus sistemas de algoritmos o derivados del conjunto de su modelo empresarial. Si su actividad se dirige a menores de edad o cuenta con menores de edad como personas usuarias

finally, companies should be obliged to incorporate the rights of the child in their due diligence processes, in concrete to carry out and put at public disposal evaluations of the impact on the rights of children and girls, paying special attention to the differentiated effects —and in serious cases— that the digital environment has on this group of population. Additionally, they should adopt the adequate measures to prevent, monitor, investigate and sanction the abuses against the rights of the child committed by companies.

Recomendaciones sobre medidas de reparación efectivas

Los Estados deben:

- Invertir in the implementation of effective digital education programs to guarantee that people know their rights —including the right to seek an effective remedy in the event of a violation of data protection, privacy or any other human right— when they access digital services.
- Guarantee access to an effective remedy in case of abuses against human rights linked to the effects that companies' technological activities, including the harms derived from the activities of their subsidiaries (foreign or national). The mechanisms of reparation should be easily accessible and understandable so that people can file complaints when their rights have been violated.

ANEXO

1. PREPARACIÓN DE LA INVESTIGACIÓN TÉCNICA

Para esta indagación utilizamos una versión relativamente reciente de la aplicación móvil de TikTok que funciona en dispositivos Android. Nuestras cuentas se manejaban de forma automática, y nuestro equipo interceptaba la comunicación de la aplicación con los servidores de TikTok para llevar a cabo la recopilación de datos desde el lado del cliente. Luego enviamos esos datos a los servidores de Junkipedia, el sistema de auditoría creado por nuestro socio técnico Instituto de Transparencia Algorítmica, en los que enriquecimos la recopilación que habíamos realizado con metadatos adicionales, como el texto de las transcripciones de audio de los vídeos de TikTok, los datos sobre la interacción y los datos de los medios de comunicación.

Las cuentas de Estados Unidos las creamos directamente en dispositivos Android ubicados en ese país. En el caso de Kenia, creamos las cuentas en otro país, pero después las manejamos con una red privada virtual (VPN) keniana. Tras la creación de las cuentas y la puesta en marcha de los experimentos, 9 de las 20 cuentas kenianas fueron deshabilitadas por TikTok antes de que terminaran los experimentos correspondientes. Hemos excluido esos casos del análisis.

El proceso de creación de las cuentas requería a veces resolver una prueba de validación (captcha), y ocasionalmente tuvimos que verificar la dirección de correo electrónico que habíamos utilizado para crear la cuenta. En la práctica, verificamos los correos electrónicos de todas las cuentas que creamos. Las fechas de nacimiento de todas las personas usuarias se establecieron en días aleatorios de mayo de 2010, por lo que TikTok configuró todas cuentas como privadas. De vez en cuando, TikTok nos mostraba un mensaje que informaba de que nuestra cuenta era de configuración privada, y en muchos casos aparecía también una ventana emergente titulada “Aspectos destacados sobre la privacidad para adolescentes” con una serie de vídeos relativos a la recopilación y el uso de datos por parte de TikTok.

2. LISTAS DE TÉRMINOS UTILIZADOS EN EL EXPERIMENTO AUTOMATIZADO

Broad Mental Health Interest:

alwayshereforyou❤️❤️
anxiety
bipolar
blade
blowup
bpd
broken
canileave
cantfightanylonger
childhoodtrama
cry

cryforhelp
crying
darkvibes
deppresed
deppresion
depression
depressed
depressedtiktok
depressing
depressingslideshow
depression

depressionanxiety
depressionawareness
depressions
dontknowwhattodo
drained
drainedmentally
emotionallydrained
endpoint
harm
heartbroken
hurting

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

EMPUJADOS A LA OSCURIDAD

EL FEED "PARA TI" DE TIKTOK FOMENTA LA AUTOLESIÓN Y LA IDEACIÓN SUICIDA

Amnistía Internacional

Narrow Mental Health Interest:

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

General Interest:

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

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

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

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

3. RESPUESTA ESCRITA DE TIKTOK, 12 DE JULIO DE 2023

TikTok response page 1



July 12th, 2023

Dear Ms. Abdul Rahim,

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

DUE DILIGENCE

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

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

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

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

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

NOT allowed

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

Age-restricted (18 years and older)

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

For You Feed ineligible

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

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

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

DATA COLLECTION

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

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

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

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

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

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

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

AGE GATE AND MINIMUM AGE APPEALS

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

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

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

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

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

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

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

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

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

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

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

Encouraging Supportive Conversations

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

#MentalHealthAwareness Creator Spotlight

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

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

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

With warm regards,
TikTok Trust & Safety

4. RESPUESTA ESCRITA DE TIKTOK, 29 DE OCTUBRE DE 2023

TikTok response page 1



October 29, 2023

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

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

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

Privacy and Advertising Policies

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

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

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

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

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

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

Human Rights

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

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

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

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

Teen Safety and Mental Health

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

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

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

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

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


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

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

Sincerely,



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



**AMNISTÍA INTERNACIONAL
ES UN MOVIMIENTO GLOBAL
DE DERECHOS HUMANOS.
LAS INJUSTICIAS QUE
AFECTAN A UNA SOLA
PERSONA NOS AFECTAN A
TODAS LAS DEMÁS.**

CONTACTO



info@amnesty.org



+44 (0)20 7413 5500

ÚNETE A LA CONVERSACIÓN



www.facebook.com/AmnestyGlobal



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

EMPUJAR A LA OSCURIDAD

EL FEED “PARA TI” DE TIKTOK FOMENTA LA AUTOLESIÓN Y LA IDEACIÓN SUICIDA

Durante la pandemia de COVID-19, TikTok se convirtió en una plataforma global y atrajo a centenares de millones de niños, niñas y jóvenes en gran medida gracias a su página “Para ti”, un feed de sugerencias personalizadas de vídeos que puede desplazarse indefinidamente, y el sistema algorítmico de recomendación que hay detrás de ella.

Mediante una hiperpersonalización impecable, TikTok ha creado una plataforma adictiva pese a la creciente evidencia científica de los graves riesgos para la salud asociados con el uso compulsivo de las redes sociales por parte de los niños y niñas. Al examinar otros riesgos de la segmentación de contenidos de la plataforma, la investigación de Amnistía Internacional muestra que la página “Para ti” de TikTok puede arrastrar fácilmente a los niños, niñas y jóvenes que manifiestan un interés por la salud mental hacia espirales de contenidos potencialmente nocivos, incluidos vídeos que idealizan y fomentan el pensamiento depresivo, las autolesiones y el suicidio. TikTok podría agravar los problemas que este grupo de población tiene con la depresión, la ansiedad y las autolesiones, y poner en peligro la salud física y mental de las personas jóvenes.

TikTok debe revisar de forma urgente sus procesos de recopilación de datos y amplificación, y ejercer de manera integral la diligencia debida en materia de derechos humanos. No obstante, las medidas particulares de tan sólo una empresa no bastan para refrenar un modelo empresarial que es intrínsecamente incompatible con los derechos humanos. Los Estados deben controlar las “grandes empresas tecnológicas” de conformidad con las leyes y normas internacionales de derechos humanos a fin de proteger y hacer efectivos los derechos de los niños, niñas y jóvenes.