

# Nuestra continua lucha contra el odio y el acoso

*Por Cormac Keenan, director de Confianza y Seguridad de TikTok*

Hoy publicamos nuestro [Informe de aplicación de las normas de la comunidad del segundo trimestre del 2021](#), el cual detalla el volumen y la naturaleza de los contenidos infractores y las cuentas eliminadas de TikTok para proteger la seguridad de nuestra comunidad y la integridad de nuestra plataforma. También estamos compartiendo actualizaciones en nuestro trabajo para proteger a las personas de los comportamientos abusivos.

## **Un vistazo a nuestro último Informe de cumplimiento de las normas de la comunidad**

Como se detalla en [nuestro informe](#), entre abril y junio se eliminaron 81,518,334 videos en todo el mundo por infringir nuestras Normas de la Comunidad o las Condiciones del Servicio, lo que supone menos del 1% de todos los videos subidos a TikTok. De ellos, identificamos y eliminamos el 93% en las 24 horas siguientes a su publicación y el 94.1% antes de que un usuario los denunciara. El 87.5% de los contenidos eliminados tenían cero visualizaciones, lo que supone una mejora desde nuestro último informe (81.8%).

Seguimos avanzando de forma constante en la detección proactiva de conductas de odio, intimidación y acoso. Por ejemplo, el 73.3% de los videos de acoso e intimidación se eliminaron antes de que se presentaran denuncias, en comparación con el 66.2% del primer trimestre de este año, mientras que el 72.9% de los videos de comportamientos de odio se eliminaron antes de que se presentaran denuncias, en comparación con el 67% del último informe. Este progreso es atribuible a las mejoras continuas de nuestros sistemas que marcan proactivamente los símbolos de odio, las palabras y otras señales de abuso para una revisión posterior por nuestros equipos de seguridad.

El acoso en general, y la incitación al odio en particular, son cuestiones muy matizadas y contextuales que pueden ser difíciles de detectar y moderar correctamente cada vez. Por ejemplo, la reapropiación de un término no es una violación de nuestras políticas, pero usar ese término para atacar o abusar de otra persona violaría nuestra política de comportamiento de odio. El *bullying* puede ser muy personal y requiere un contexto fuera de línea que no

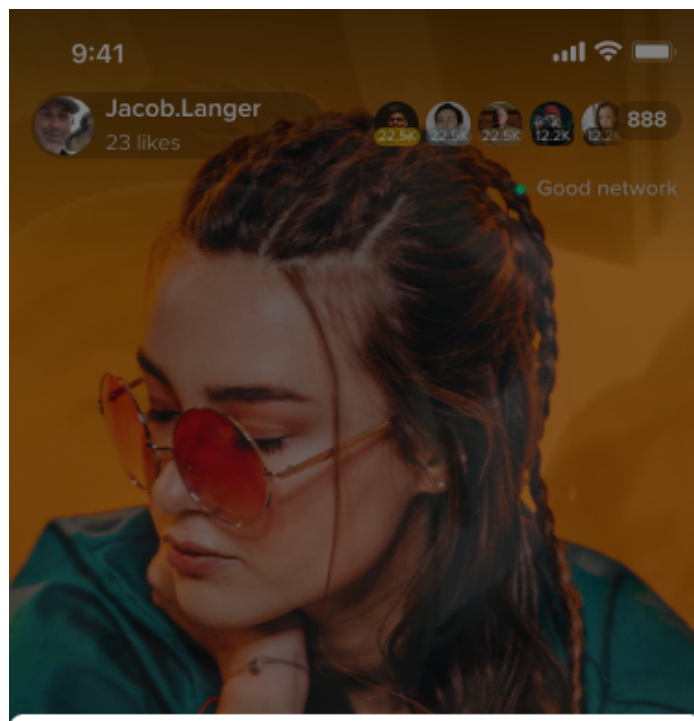
siempre está disponible. Para aplicar mejor nuestras políticas, formamos y guiamos regularmente a nuestro equipo a entender las diferencias entre, por ejemplo, la reapropiación y las calumnias o la sátira y el *bullying*. También hemos contratado a expertos en políticas de derechos civiles, equidad e inclusión. A medida que realizamos continuas mejoras de nuestros mecanismos de detección, nos esforzamos por resolver estos problemas críticos para nuestra comunidad. Animamos a las personas a denunciar las cuentas o los contenidos que puedan infringir nuestras Normas de la Comunidad.

### **Incorporando medidas contra el abuso en nuestro producto**

Además de eliminar contenidos, permitimos a los usuarios personalizar su experiencia con una serie de herramientas y [recursos](#), entre los que se incluyen formas eficaces de [filtrar los comentarios sobre sus contenidos](#), [eliminar o denunciar varios comentarios a la vez](#) y bloquear cuentas en forma masiva.

También hemos añadido [avisos](#) que animan a la gente a considerar el impacto de sus palabras antes de publicar un comentario potencialmente desagradable o violatorio. El efecto de estos avisos ya se ha notado, ya que casi 4 de cada 10 personas han optado por retirar y editar su comentario, y la mayoría de las ediciones no infringen nuestras políticas. Aunque no todo el mundo opta por cambiar sus comentarios, nos alienta el impacto de características como esta y y continuamos desarrollando y probando nuevas acciones para prevenir posibles abusos.

Hoy estamos ampliando estas funciones con la mejora de la configuración de silencio para los comentarios y preguntas durante las transmisiones EN VIVO. Las retransmisiones en directo en TikTok son una forma emocionante de conectar a creadores y espectadores, y estamos mejorando la experiencia de seguridad a través del diseño. Ahora, el anfitrión o su [ayudante de confianza](#) pueden silenciar temporalmente a un espectador poco amable durante unos segundos o minutos, o durante la duración del EN VIVO. Si una cuenta es silenciada por cualquier cantidad de tiempo, también se eliminará todo el historial de comentarios de esa persona. Los anfitriones ya pueden [desactivar los comentarios](#) o limitar los comentarios potencialmente perjudiciales mediante un filtro de palabras clave. Esperamos que estos nuevos controles ayuden a los anfitriones y a la audiencia a realizar transmisiones en directo seguras y entretenidas.



### Seleccionar duración

- 5 segundos
- 30 segundos
- 1 minuto
- 5 minutos
- Video LIVE actual

Silenciar

Para más información sobre las medidas que estamos tomando para proteger la seguridad de nuestra comunidad y la integridad de nuestra plataforma, te animamos a leer [nuestro segundo informe trimestral](#).