Reporte de Aplicación de Lineamientos de la Comunidad de TikTok, primer trimestre 2021

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

Personas de todo el mundo vienen a TikTok para expresarse creativamente y entretenerse. Para mantener esa confianza, apuntamos a ser transparentes sobre las acciones que tomamos en torno al contenido. Con eso en mente, hoy publicamos nuestro Reporte de Aplicación de Lineamientos de la Comunidad del primer trimestre, mientras que continuamos brindando transparencia al importante trabajo de moderar el contenido para que TikTok sea un lugar seguro y acogedor para nuestra comunidad.

Desde 2019 hemos liberado Informes de Transparencia, incrementando la información que compartimos en cada informe. Esto incluye agregar hallazgos relacionados a las acciones que tomamos para proteger la seguridad e integridad de nuestra plataforma, como reportar la naturaleza de los contenidos eliminados por violar nuestras políticas y el número de anuncios ofensivos rechazados. Hemos escuchado por parte de nuestra comunidad, organizaciones de la sociedad civil y legisladores que esta información es útil para entender cómo TikTok funciona y modera el contenido. Para compartir esta información más frecuentemente, ahora publicaremos trimestralmente los Reportes de Aplicación de Lineamientos de la Comunidad como parte de nuestros esfuerzos más amplios para reportar transparencia. Debio a la naturaleza del procesamiento y respuesta de solicitudes legales, continuaremos publicando esa información bianualmente.

Aquí algunos de los hallazgos clave de nuestro reporte del primer trimestre de 2021:

- Se eliminaron 61,951,327 videos por incumplir nuestras Políticas de Comunidad o Términos de servicio, que son menos del 1% de todos los videos publicados en TikTok durante ese periodo.

- El 82% de estos videos se eliminaron antes de recibir cualquier tipo de visualización, el 91% antes de algún reporte por parte de los usuarios y el 93% dentro de las 24 horas posteriores a su publicación.
- Se rechazaron 1,921,900 anuncios por incumplir las políticas y pautas publicitarias.
- Se eliminaron 11,149,514 cuentas por incumplir con nuestras Políticas de Comunidad o Términos de servicio, de las cuales 7,263,952 se eliminaron por pertenecer potencialmente a una persona menor de 13 años. Esto es menos del 1% del total de cuentas en TikTok.
- Se impidió la creación de 71,470,161 cuentas a través de medios automatizados.

Al evolucionar nuestro reporte, hemos apuntado a ayudar a la industria a avanzar cuando se trata de transparencia y responsabilidad sobre la seguridad del usuario. Para dar más visibilidad a las acciones que tomamos para proteger a los menores, en este informe agregamos la cantidad de cuentas eliminadas por pertenecer potencialmente a personas menores de edad. Todo esto forma parte de nuestros esfuerzos por fortalecer la configuración de privacidad predeterminada para adolescentes, ofrecer herramientas para empoderar a los padres y las familias, y limitar funciones como la mensajería directa y transmisión en vivo sólo para los mayores de 16 años. Para continuar fortaleciendo nuestro acercamiento de mantener TikTok como un lugar para personas mayores de 13 años, apuntamos a explorar nuevas tecnologías que ayuden con el gran desafío de la industria de verificar la edad.

Este informe también incluye nuestro primer resumen de seguridad sobre nuestro programa global de recompensas por fallos, que nos ayuda a identificar y resolver proactivamente las vulnerabilidades de seguridad. Este programa refuerza la madurez en nuestra seguridad general al alentar a los investigadores de seguridad global a identificar y revelar responsablemente los errores a nuestros equipos para que podamos resolverlos antes de que los atacantes los exploten. Como informamos, durante el primer trimestre del 2021, TikTok recibió 33 denuncias válidas y resolvió 29 de ellas. También recibimos y publicamos 8 solicitudes de divulgación pública. En cuanto a la eficiencia de respuesta, TikTok tiene un tiempo promedio de primera respuesta de 8 horas, un tiempo de resolución de 30 días, y en promedio nos toma 3 días para pagar una recompensa.

En el futuro, publicaremos estos datos en nuestro centro de transparencia en línea, mismo que estamos trabajando para convertirlo en un lugar de consulta de nuestros distintos Informes de transparencia y otra información relevante sobre nuestros esfuerzos para proteger la seguridad e integridad de nuestra plataforma y su comunidad.