Nota de prensa

Observatorio Español del Racismo y la Xenofobia y redes sociales

El OBERAXE publica por primera vez datos de retirada cedidos por las plataformas digitales

- **Las plataformas han ofrecido información sobre sus medidas proactivas de moderación, aquellas que permiten detectar y retirar contenidos de odio antes de recibir denuncias de usuarios**
- **El avance ha sido posible gracias al trabajo coordinado entre el OBERAXE y las plataformas en el Grupo de Trabajo sobre Discurso de Odio en Redes Sociales**
- **"Se trata de un punto de partida y se seguirá avanzando para unificar los datos en próximos informes", explica Tomás Fernández Villazala, director del Observatorio**
- **El OBERAXE detectó casi 121.000 mensajes de odio entre octubre y diciembre de 2025**

Madrid, 5 de marzo de 2026.- El Observatorio Español del Racismo y la Xenofobia (OBERAXE) publica su **Boletín Trimestral** de Monitorización del Discurso de Odio en Redes Sociales caracterizado por un avance sin precedentes en la transparencia y la colaboración con las plataformas digitales.

Por primera vez, Facebook, Instagram, TikTok y YouTube han facilitado al Observatorio datos detallados sobre sus **medidas proactivas de moderación.** Estas medidas incluyen las actuaciones iniciadas de oficio por sus sistemas automatizados o por sus equipos internos, sin necesidad de recibir denuncias de los usuarios.

Aunque los datos aportados presentan variaciones en fechas y niveles de desagregación, lo que no permite ofrecer datos de retirada unificados, abren la puerta hacia una nueva línea de trabajo. "Es un gran avance en trasparencia,

Nota de prensa

gracias al trabajo realizado. Se trata de un **punto de partida** y se seguirá avanzando para unificar los datos en próximos informes", explica Tomás Fernández Villazala, director de OBERAXE.

Durante el último trimestre de 2025, el Observatorio detectó **120.990 mensajes de odio** y se constató una mejora en la tasa de retirada de los mensajes que OBERAXE notifica a las plataformas. Este trimestre se ha logrado un porcentaje del 51% de contenidos eliminados.

Los datos ofrecidos por Facebook, YouTube, Instagram y TikTok corresponden al tercer trimestre de 2025 (del 1 de julio al 30 de septiembre) y son datos de porcentaje de retirada de todo tipo de discurso de odio en todo el mundo. En el caso de Facebook fueron el 66% de las retiradas por discurso de odio. En el caso de Instagram, según los datos de la propia plataforma, fueron un 87% de las intervenciones proactivas por contenido de odio general.

La plataforma TikTok además notificó que un 54% de sus retiradas por discurso de odio se produjeron antes de que el contenido llegara a ver la luz. En el caso de esta red social se facilitaron datos del cuarto trimestre (del 1 de octubre al 31 de diciembre de 2025) y en concreto referidos solo a España. En cuanto a YouTube, el dato facilitado por la plataforma remarca que eliminó el 98% de los contenidos mediante detección automática.

Este avance en transparencia y cooperación es resultado del **Grupo de Trabajo sobre Discurso de Odio en Redes Sociales**, en el que están representadas todas las plataformas, la Secretaría de Estado de Migraciones, LALIGA y el Departamento de Seguridad Nacional (DSN), impulsado por la ministra de Inclusión, Seguridad Social y Migraciones en julio de 2025 y consolidado como un espacio de interlocución estable con las plataformas.

Descenso progresivo del número de mensajes de odio

La detección de **120.990 mensajes de odio** en el trimestre octubre-diciembre 2025, con una media de **1.300 mensajes diarios** y picos asociados a episodios de inseguridad ciudadana, conflicto armado y tensiones deportivas, suponen un descenso muy significativo respecto al trimestre anterior (331.817). Los principales picos de odio en este trimestre sucedieron en octubre, con valores superiores a

Nota de prensa

3.600 mensajes diarios, coincidiendo en el tiempo con la interceptación de la flotilla solidaria a Gaza y con el aniversario de los atentados de Hamás el 7 de octubre.

Las plataformas retiraron el **51% de los contenidos notificados por OBERAXE**, seis puntos más que en el trimestre anterior, consolidando una mejora sostenida en los mecanismos de moderación. La media acumulada de retirada a lo largo del año es de un 41%.

El discurso de odio en el fútbol

Durante el cuarto trimestre de 2025 el discurso de odio en el ámbito deportivo, especialmente en el fútbol, supuso **un 7%** de los mensajes de odio monitorizado. El colectivo diana principal fueron las personas originarias del norte de África.
En diciembre el principal evento generador de odio en redes fue la **concesión del balón de oro a Lamine Yamal.**

Boletín trimestral: https://run.gob.es/otf818fa