



BOLETÍN DE MONITORIZACIÓN DEL DISCURSO DE ODIO EN REDES SOCIALES

2024
1 JULIO
31 AGOSTO

El Observatorio Español del Racismo y la Xenofobia (OBERAXE) colabora desde 2017 con la Comisión Europea en los ejercicios de monitorización que se realizan en cumplimiento del Código de Conducta para la lucha contra la incitación ilegal al odio en Internet, que fue firmado con las empresas prestadoras de servicios de alojamiento de datos con mayor presencia en la Unión Europea.



Desde mayo de 2020, el OBERAXE monitoriza a diario el discurso de odio en España en cinco redes sociales (Twitter, Facebook, YouTube, Instagram y TikTok) y notifica aquellos contenidos considerados de odio o que pueden ser constitutivos de delito.

En el periodo entre el 1 de julio y el 31 agosto de 2024 se han realizado 506 notificaciones a las plataformas, las cuales han retirado el 37% de los contenidos. Una parte importante del discurso de odio identificado en este periodo se vincula a la desinformación y a la difusión de bulos que criminalizan a las personas inmigrantes. Asimismo, repuntan los contenidos de discurso de odio en el ámbito del deporte.



CONTENIDOS ANALIZADOS

506 contenidos de discurso de odio notificados

El número total de contenidos de discurso de odio notificados es de 506. En la Figura 1 se observa cierto predominio en el número de contenidos notificados a TikTok con 133 casos comunicados, seguido de X (122), Facebook (112), Instagram (103), y YouTube (36). El desigual volumen de contenidos notificados obedece al distinto grado de dificultad para su identificación entre redes sociales.

El conjunto de las plataformas incrementa la tasa de retirada de contenido en 13 puntos porcentuales

Las plataformas en su conjunto han retirado 187 contenidos del total de las 506 notificaciones realizadas por el OBERAXE, es decir, el 37%. La tasa de retirada ha aumentado en trece puntos porcentuales respecto al bimestre anterior (24%).

No obstante, esta tasa sigue siendo insuficiente ya que una gran parte del discurso de odio identificado permanece en la red. Asimismo, el número de contenidos retirados tras su notificación como usuario normal sigue siendo bajo. A través de esta vía, las plataformas han eliminado 67 contenidos, lo cual supone solo el 13% respecto al total de las notificaciones.

A su vez, mediante la vía *trusted flagger*, empleada este bimestre en 439 ocasiones, las plataformas han retirado 120 publicaciones más, lo que supone un 24% adicional.

TikTok e Instagram retiran el 65% de los contenidos de discurso de odio notificados

Al analizar los contenidos retirados por cada plataforma en relación con el total de casos notificados a cada una de ellas (Figura 2), TikTok e Instagram han sido, en este periodo, las redes con mayor tasa de retirada, con un 65% cada una de ellas. Muy por debajo se sitúan YouTube (22%), X (12%) y Facebook (9%).

*Conforme a la RECOMENDACIÓN (UE) 2018/334 DE LA COMISIÓN de 1 de marzo de 2018 sobre medidas para combatir eficazmente los contenidos ilícitos en línea se entiende como "comunicante fiable" (*trusted flagger*): aquella persona física o jurídica que un prestador de servicios de alojamiento de datos considere que posee competencias y responsabilidades particulares a efectos de la lucha contra los contenidos ilícitos en línea.

FIGURA 1.
CONTENIDOS COMUNICADOS

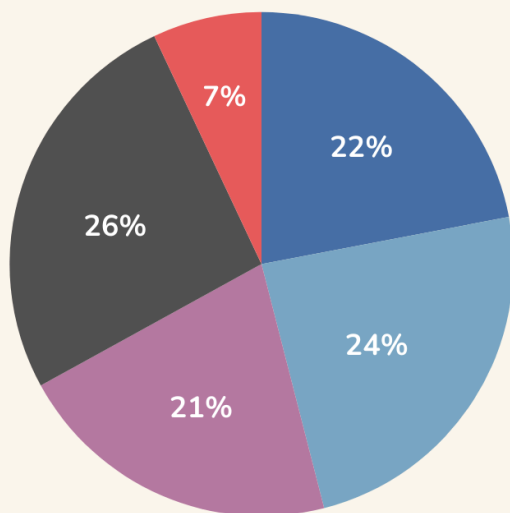
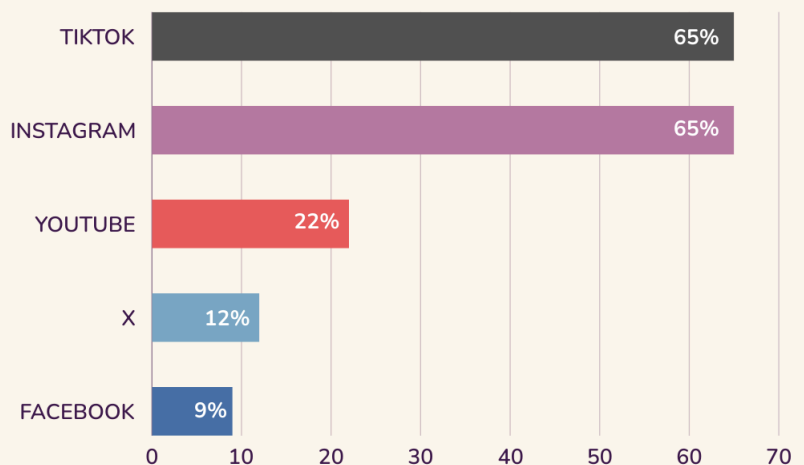


FIGURA 2.
CONTENIDOS RETIRADOS



● INSTAGRAM ● FACEBOOK ● X ● TIKTOK ● YOUTUBE

EVOLUCIÓN EN LA RETIRADA DE CONTENIDOS

La Figura 3 presenta los plazos en los que las plataformas retiran el contenido que se les ha notificado*: 24 horas, 48 horas o una semana. Asimismo, incorpora el número de contenidos retirados tras la comunicación como *trusted flagger*.

El 19% de las notificaciones se ha retirado a las 24 horas

Del total de contenidos comunicados al conjunto de las plataformas, estas han retirado el 19% de las notificaciones a las 24 horas; el 6%, a las 48 horas; y el 11%, a la semana de haber sido notificados.

Al analizar los plazos en los que las plataformas eliminan los contenidos reportados, observamos que X retira más de la mitad de sus casos (53%) en un plazo inferior a las 48 horas. TikTok, por su parte, ha eliminado el 28% de los contenidos en 24 horas.

Por el contrario, en Facebook predomina la evaluación y retirada de los contenidos en el plazo de una semana, eliminando el 60% de sus casos en este periodo de tiempo; mientras que YouTube en este bimestre ha comenzado a actuar sobre los casos a la semana de haber sido notificados.

La implementación del comunicante fiable es crucial para la eliminación efectiva del contenido de odio

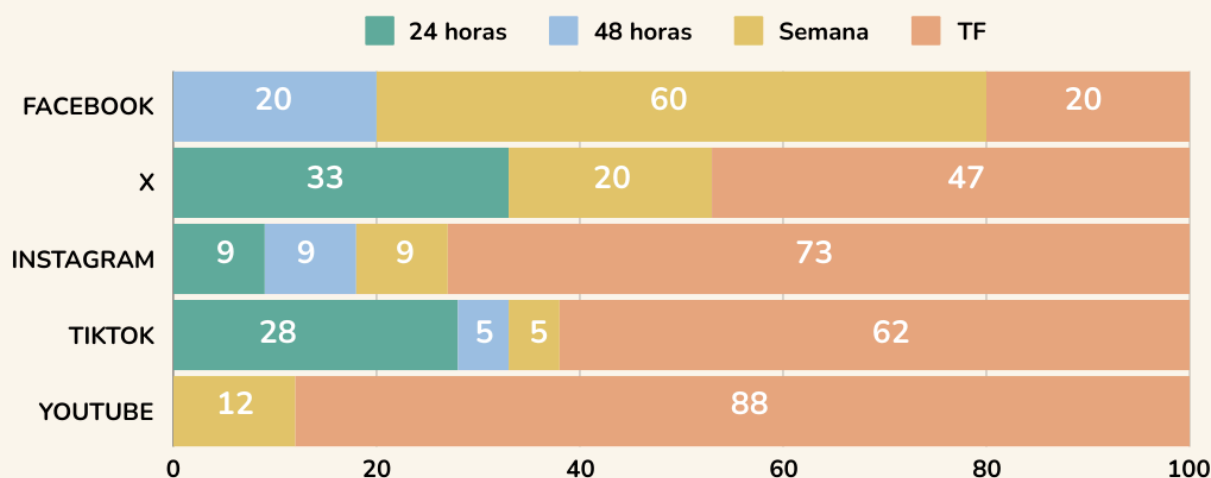
En el conjunto de las plataformas, el 64% del total de casos retirados (187) ha sido a través de la vía de *trusted flagger*.

La proporción de casos retirados tras su reporte como *trusted flagger*, en comparación con los retirados a través de la notificación como usuario normal, muestra la importancia de la figura de los comunicantes fiables en la identificación y acción sobre los contenidos de discurso de odio que pueden infringir las normas de las plataformas o ser constitutivos de delito o de una infracción administrativa.

YouTube, en este periodo, ha retirado el 88% de las comunicaciones a través de la vía de *trusted flagger*. Les sigue Instagram (73%), TikTok (62%), X (47%), y Facebook (20%).

*Si al cabo de una semana no se ha realizado la retirada por parte de la plataforma del contenido notificado como usuario normal, se procede a utilizar la vía de comunicación cualificada o *trusted flagger*.

FIGURA 3.
% TIEMPO DE RETIRADA DE CADA PLATAFORMA



● 24 HORAS ● 48 HORAS ● SEMANA O MÁS ● TRUSTED FLAGGER

CONTENIDO DE DISCURSO DE ODIO

El 44% de los contenidos notificados deshumaniza o degrada gravemente a las personas de origen extranjero

La deshumanización y la degradación son las formas predominantes de discurso de odio en las redes (44%), seguidas de los mensajes destinados a promover el descrédito de las personas sin dar ningún argumento más que la pertenencia al grupo (35%).

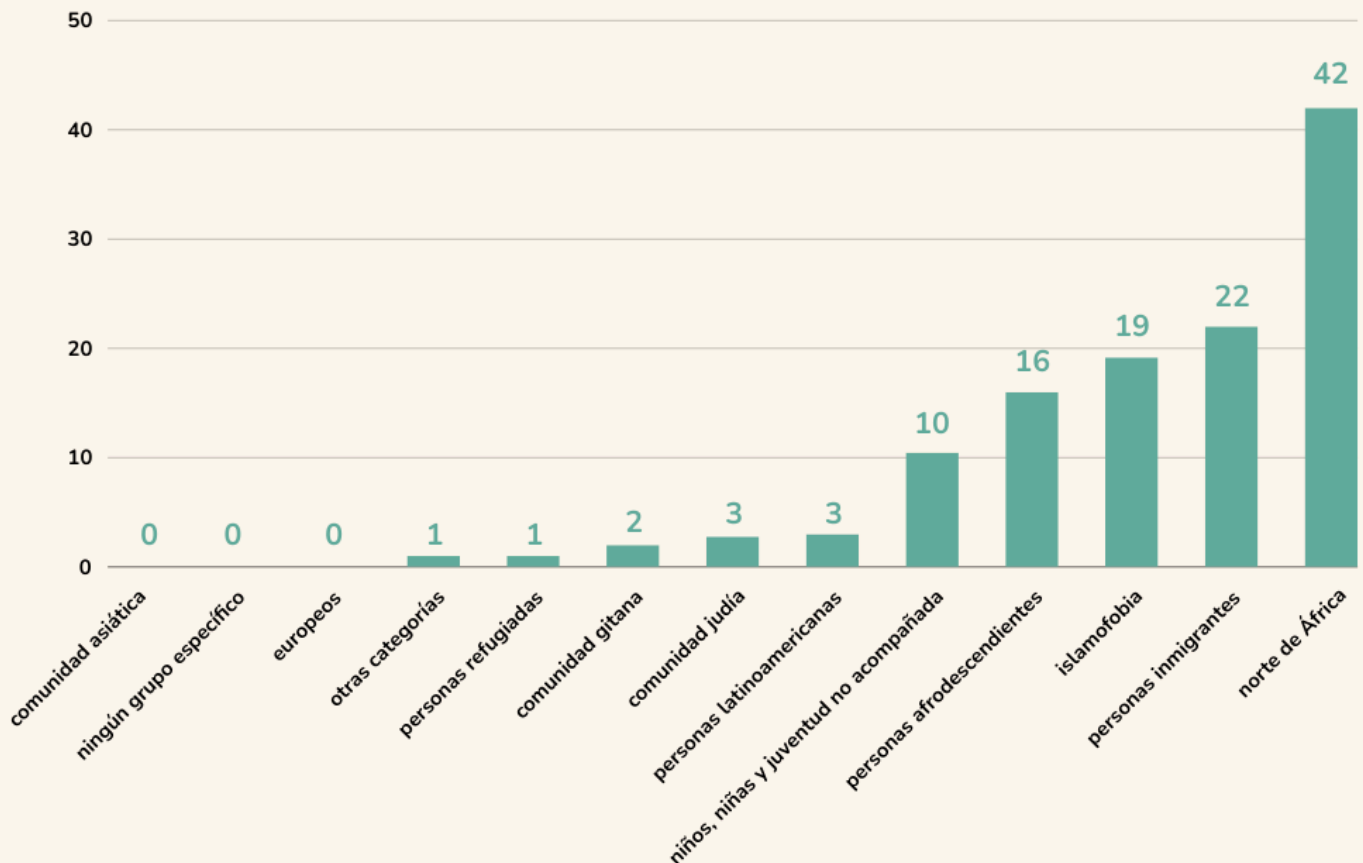
El calado de este tipo de discurso en la sociedad genera un clima de odio y hostilidad que puede desembocar en conductas agresivas y violentas, especialmente en situaciones de tensión social. De hecho, un 29% de los contenidos incita a la violencia con amenazas directas o indirectas.

Por otra parte, el 24% de los mensajes presentan al grupo como una amenaza; y el 14% promueve la expulsión de las personas de origen extranjero.

El 48% de los mensajes contienen un lenguaje agresivo explícito

Casi la mitad de las comunicaciones analizadas contienen lenguaje explícitamente agresivo, incluyendo insultos y expresiones hostiles. Por otro lado, en un 40% se emplean expresiones discriminatorias no agresivas; y en un 12% se utiliza la ironía y el sarcasmo como recurso para encubrir los comentarios con motivación racista y/o xenófoba.

FIGURA 4.
% COLECTIVOS VICTIMIZADOS EN EL DISCURSO DE ODIO



Aumenta la hostilidad hacia las personas del norte de África y las personas inmigrantes

En este periodo ha crecido la hostilidad hacia las personas con origen del norte de África (42%) y hacia las personas inmigrantes (22%), aumentando, respectivamente, en nueve y cuatro puntos porcentuales respecto al bimestre anterior.

También se ha incrementado, en dos puntos porcentuales, el discurso de odio hacia los niños, niñas y juventud no acompañada (10%). Por el contrario, la islamofobia y el discurso contra personas afrodescendientes han mostrado una ligera disminución de ocho y siete puntos porcentuales, respectivamente.

La desinformación alimenta un discurso de odio que criminaliza a las personas inmigrantes

El 43% de las publicaciones de discurso de odio racista y/o xenófobo se suscitan a raíz de incidentes de inseguridad ciudadana, como violencia, agresiones y robos, atribuidos, ya sea de manera verídica o falsamente, a personas de los diferentes colectivos analizados.

Cabe señalar que el 58% de los contenidos referentes a incidentes violentos (219) promueven la percepción de inseguridad ciudadana, pero no están basados en hechos verídicos, actuales y ocurridos en España; por lo que se refieren a información falsa, bulos o incidentes descontextualizados.

Un claro ejemplo de los bulos difundidos con el objeto de criminalizar a las personas de origen inmigrante han sido las publicaciones asociadas al asesinato de Mateo en Mocejón (Toledo) producido el pasado 18 de agosto. El 3% de los contenidos de discurso de odio identificados en este bimestre hacían referencia directa a este suceso, promoviendo el alarmismo social, la percepción de las personas inmigrantes como una amenaza y la fragmentación social, en un período marcado también por el debate en torno a las políticas migratorias.

En este bimestre, el principal el colectivo atacado en relación con el episodio de percepción de inseguridad ciudadana es el de las personas con origen del norte de África (28%), seguido de las personas inmigrantes (15%).

En este bimestre, otro acontecimiento específico que ha influido en la difusión de discurso de odio son los incidentes y protestas producidas en Reino Unido en contra de las personas inmigrantes. El 4% del total de los contenidos se vincula a estos hechos.

Aumenta en 12 puntos porcentuales el discurso de odio en el ámbito deportivo

Las competiciones deportivas celebradas este verano han tenido una incidencia destacada en el porcentaje de contenidos de discurso de odio notificados. El 13% del total de los casos se vincula al ámbito del deporte, mientras que el 11% se relaciona directamente con los Juegos Olímpicos, con contenidos que deshumanizan a deportistas en base a su origen étnico.

Otra de las narrativas observadas en este ámbito es la difusión de mensajes de odio contra deportistas que, en sus respectivas disciplinas, representan a España y tienen un origen étnico diverso, algo que ocurre especialmente con atletas españolas y españoles afrodescendientes y procedentes del norte de África.

Acciones del OBERAXE en el ámbito del deporte

Estos datos ponen en relieve cómo los eventos deportivos pueden ser un terreno fértil para la producción de discursos intolerantes que trascienden del plano deportivo y repercuten en toda la sociedad.

Por ello, desde el OBERAXE se trabaja para erradicar el racismo en este ámbito y emplear el deporte como un medio valioso para promover la inclusión, con iniciativas como el proyecto europeo SCORE.

Otra vía es la colaboración entre LALIGA y el Ministerio de Inclusión, Seguridad Social y Migraciones, con la cesión de la herramienta MOOD y el trabajo conjunto para mejorar la monitorización del discurso de odio en redes.

