



BOLETÍN DE MONITORIZACIÓN DEL DISCURSO DE ODIO EN REDES SOCIALES

2023
1 MAYO
30 JUNIO



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 racista y/o xenófobo, antisemita, antigitano e islamófobo que pueden ser constitutivos de delito, infracción administrativa o que violan las normas de conducta de las plataformas.

En el periodo entre el 1 de mayo y el 30 de junio de 2023, las plataformas han retirado el 58% de los contenidos notificados. La xenofobia hacia las personas originarias del norte de África y afrodescendientes predomina en los contenidos analizados. En este bimestre, el 15% de los contenidos está relacionado con acontecimientos en el ámbito del deporte.



408 contenidos de discurso de odio notificados

El número total de contenidos de discurso de odio notificados este bimestre es de 408. En la Figura 1 se observa cierto predominio en el número de contenidos notificados a Twitter, con 119 casos comunicados, seguido de Instagram (90), TikTok (77), Facebook (74), y YouTube (48). El desigual volumen de contenidos notificados obedece al distinto grado de dificultad para su identificación entre redes sociales.

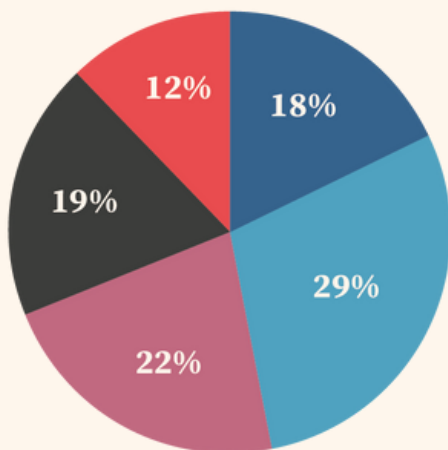
Las plataformas han retirado un 58% de los contenidos notificados

Las plataformas en su conjunto han retirado o bloqueado en España 235 contenidos de los 408 comunicados por el OBERAXE, es decir, el 58%. En este bimestre, la respuesta de las plataformas continúa siendo positiva, ya que el número de casos retirados ha aumentado en tres puntos porcentuales respecto al bimestre anterior (55%).

El 33% de las comunicaciones han sido retiradas a través de la vía de *trusted flagger*

Del total de los 408 contenidos comunicados, se han retirado 102 notificados como usuario normal, lo que supone un 25% del total.

FIGURA 1.
CONTENIDOS COMUNICADOS



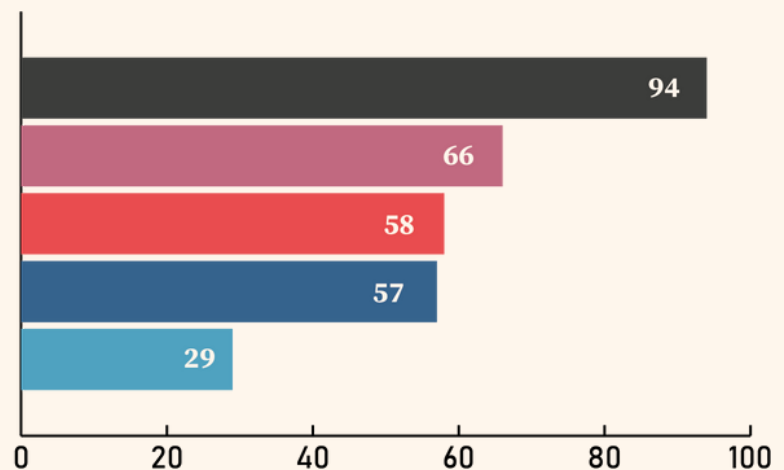
A su vez, mediante la vía *trusted flagger*, empleada este bimestre en 306 ocasiones, las redes han retirado 133 publicaciones más, lo que constituye un 33% adicional.

YouTube aumenta en 24 puntos su tasa de retirada de contenidos

Al analizar los contenidos retirados por cada plataforma en relación al total de casos notificados a cada una de ellas (Figura 2), TikTok es la red social con mayor tasa de retirada (94%). Le sigue Instagram, que ha retirado el 66% de los casos comunicados. YouTube ha retirado más de la mitad de los contenidos notificados (58%), aumentando en 24 puntos porcentuales su tasa de retirada frente al bimestre anterior (34%). Facebook, por el contrario, en este periodo ha reducido su tasa de retirada en 20 puntos (57%). Por último, Twitter se mantiene como la plataforma que menos contenido elimina (29%).

*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 2.
% CONTENIDOS RETIRADOS



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*.

Las plataformas son más eficaces en la retirada si se comunica como *trusted flagger*

Pese a las modificaciones y mejoras introducidas en los procedimientos de notificación y moderación por parte de las plataformas, la notificación como usuario normal sigue sin ser la vía más efectiva para conseguir la retirada de contenido de discurso de odio. En este bimestre, el 43% de los contenidos eliminados se han producido tras la comunicación como usuario normal, frente al 57% de los contenidos retirados a través del mecanismo de *trusted flagger*.

Instagram y TikTok, las redes que más contenido retiran tras la notificación de usuario normal

Instagram es la plataforma que más contenidos ha eliminado este bimestre tras ser notificados como usuario normal (59% del total de casos retirados). Le sigue TikTok, con un 51%.

El 8% de los contenidos han sido retirados a las 24 horas tras su notificación

Tan solo el 8% de los casos reportados se elimina a las 24 horas, siendo Instagram la plataforma que más contenido retira en este plazo (19%), seguida de TikTok (18%), YouTube (14%), Twitter (9%) y Facebook (7%). A las 48 horas, las redes sociales retiran el 7% del total de contenidos.

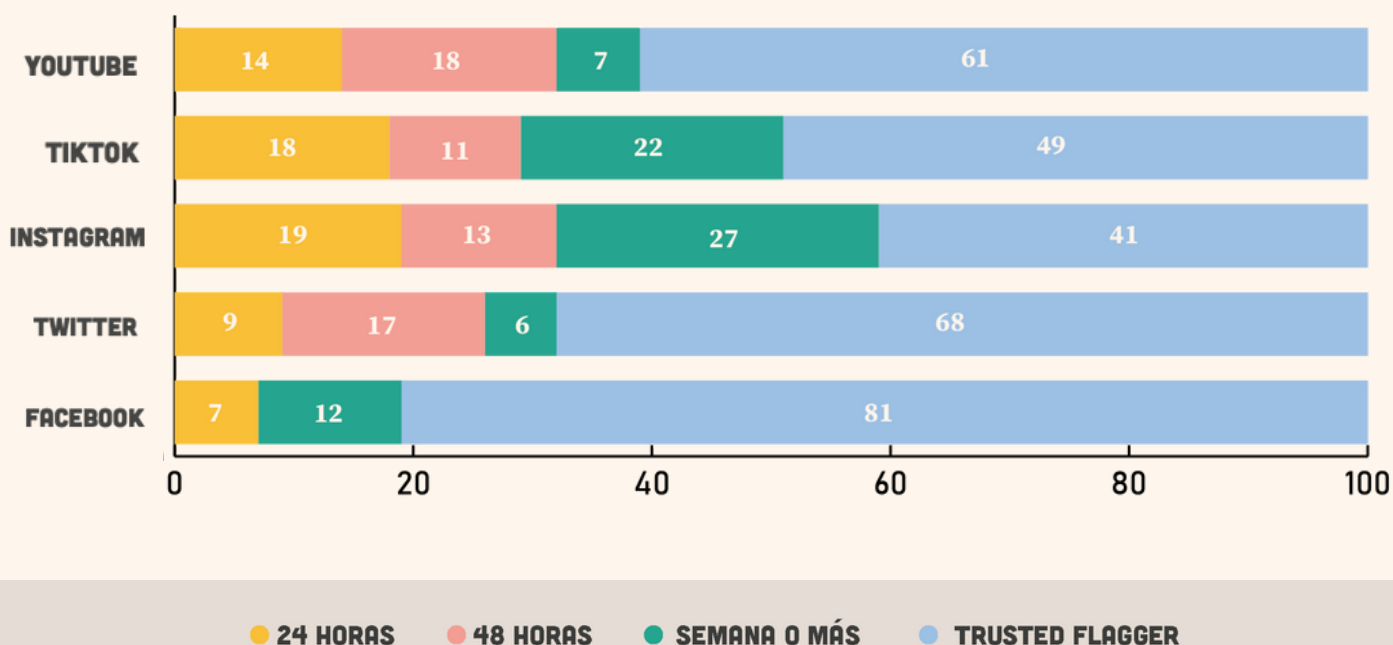
Las plataformas, por lo general, no actúan con tanta celeridad. Un 10% de los contenidos se retiran tras el reporte como usuario normal, pero permanecen visibles en la red alrededor de una semana tras su notificación hasta ser eliminados.

Facebook elimina más contenido a través de la vía de comunicante fiable

El procedimiento de notificación de retirada a través de la vía *trusted flagger* prevalece más en Facebook. El 81% de los casos eliminados por esta plataforma han sido retirados tras la notificación como comunicante fiable.

*Si transcurrida una semana, la plataforma no retira el contenido denunciado como usuario normal, se procede a utilizar el procedimiento de comunicante fiable o *trusted flagger*.

FIGURA 3.
% TIEMPO DE RETIRADA DE CADA PLATAFORMA



El 15% de los contenidos se vincula a acontecimientos en el ámbito deportivo

El 15% de los contenidos notificados están relacionados, de forma directa o indirecta, con los incidentes racistas acontecidos en estadios de La Liga durante el periodo analizado. Esto refleja la amplia difusión de este tipo de incidentes en redes, ya sea para propagar mensajes de odio o para promover la contranarrativa. No obstante, el incidente prototípico principal al cual aluden los contenidos es la inseguridad ciudadana (39%), ya sean hechos reales o *fake news*. El 6% se refiere a la ocupación. Asimismo, el 5% se sitúa en el marco de las elecciones, al coincidir el bimestre con las elecciones municipales y autonómicas del 28 de mayo.

El 66% de los contenidos degrada o deshumaniza a las personas

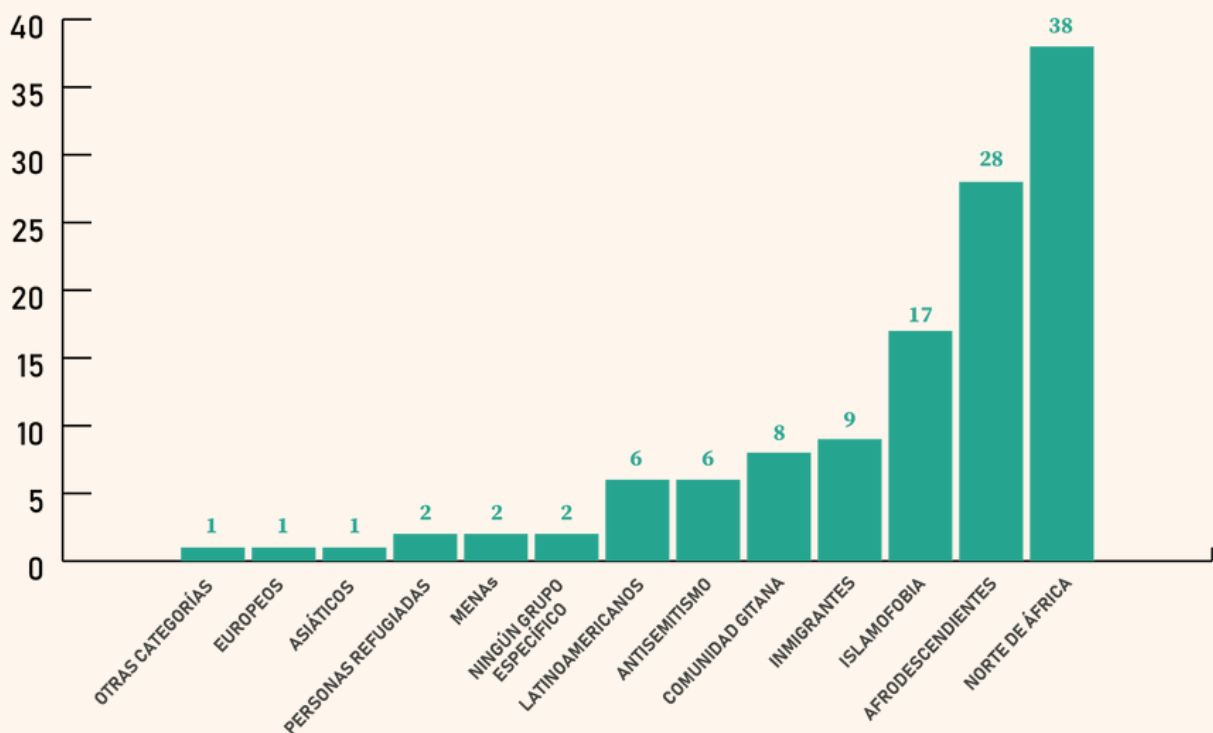
El 61% del discurso de odio notificado emplea insultos o expresiones agresivas.

El 66% de los contenidos emplea mensajes y/o imágenes que deshumanizan o degradan gravemente a las personas, el 50% promueve el descrédito en base a los atributos personales del grupo atacado, el 23% incita a la expulsión, y el 21% presenta al grupo como una amenaza.

Disminuye la islamofobia en nueve puntos porcentuales

Las personas originarias del norte de África (38%) siguen siendo el grupo que más discurso de odio recibe, con un aumento de cuatro puntos porcentuales respecto al bimestre anterior. Del mismo modo, aumenta el discurso de odio contra las personas afrodescendientes (28%) y contra la comunidad gitana (8%). Por otro lado, la islamofobia (17%) disminuye en nueve puntos porcentuales. Asimismo, el 7% de los contenidos se basan en un discurso supremacista y hacen uso de simbología y referencias al nazismo, sin dirigirse a un grupo diana específico.

FIGURA 4.
% COLECTIVOS VICTIMIZADOS EN EL DISCURSO DE ODIO



Edita y distribuye: Observatorio Español del Racismo y la Xenofobia.

NIPO: 121-22-039-0