

BOLETÍN DE MONITORIZACIÓN DEL DISCURSO DE ODIOS EN REDES SOCIALES

1 DE SEPTIEMBRE - 31 DE OCTUBRE (2021)



El **Observatorio Español del Racismo y la Xenofobia (OBERAXE)** realiza desde 2017 ejercicios de monitorización periódicos en cumplimiento del **Código de Conducta para la lucha contra la incitación ilegal al odio en Internet** que firmaron la Comisión Europea y algunas de las empresas prestadoras de servicios de alojamiento de datos con mayor presencia en todo el mundo, como **Twitter, Facebook, Youtube** o, más recientemente, **Instagram y TikTok**.

Además, desde **mayo de 2020** OBERAXE monitoriza diariamente el discurso de odio en España en estas redes sociales, comunicándoles aquellos contenidos considerados "de odio" o que pueden ser constitutivos de delito.

A continuación, se muestran los resultados obtenidos en la monitorización para el periodo comprendido entre el 1 de septiembre y el 31 de octubre de 2021.

NÚMERO DE CONTENIDOS ANALIZADOS, COMUNICADOS Y RETIRADOS EN LAS REDES SOCIALES

El número de **contenidos** de discurso de odio identificados y comunicados a las plataformas Facebook, Twitter, Youtube, Instagram y TikTok durante este periodo ha sido de **704**, lo que supone un **incremento del 16%** con respecto al boletín bimensual anterior (1 julio - 31 de agosto 2021). De estos, **251 contenidos** fueron eliminados al ser notificados como usuario normal, por lo que la tasa de eliminación se ha situado en un **35,7%**, lo que supone un **descenso de 0,7 puntos porcentuales** respecto al periodo anterior.

En todas las redes sociales se ha utilizado, además, la vía de comunicación como **trusted flagger***. Este mecanismo se utiliza cuando ha transcurrido una semana desde las comunicaciones realizadas como **usuario normal**, y los contenidos no han sido eliminados. Tras las notificaciones como **trusted flagger**, las plataformas retiraron **188 contenidos adicionales**, un **26,7%** del total, por lo que la tasa de eliminación total asciende al **62,4%**, **cantidad superior en 6,2 puntos** respecto del periodo anterior.



**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.*

Por plataformas, de las **704 comunicaciones** realizadas, **204** corresponden a **TikTok** (29%), que es la red social a la que más comunicaciones se han realizado, seguida de **Youtube** con **157 comunicaciones** (22,3%) e **Instagram**, con **131** (18,6%). Asimismo, a **Twitter** se han remitido otras **129 comunicaciones** (18,3%) y **83 comunicaciones** (11,8%) a **Facebook**. La **Figura 1** refleja estos datos.

Figura 1. Porcentaje de comunicaciones realizadas a cada plataforma

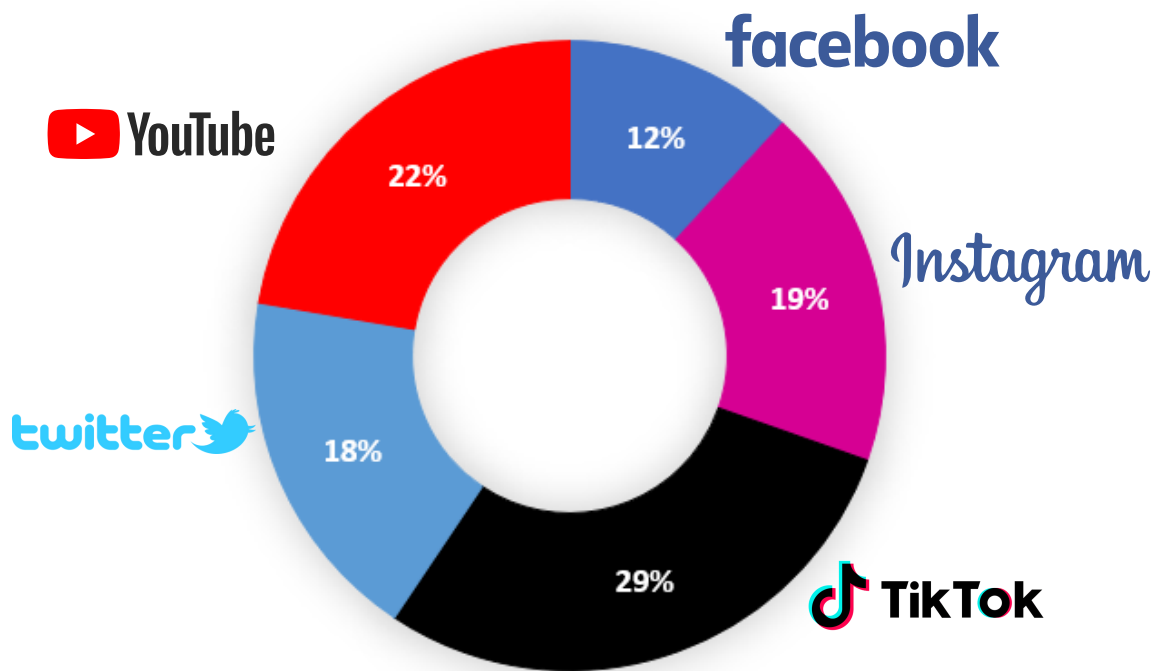
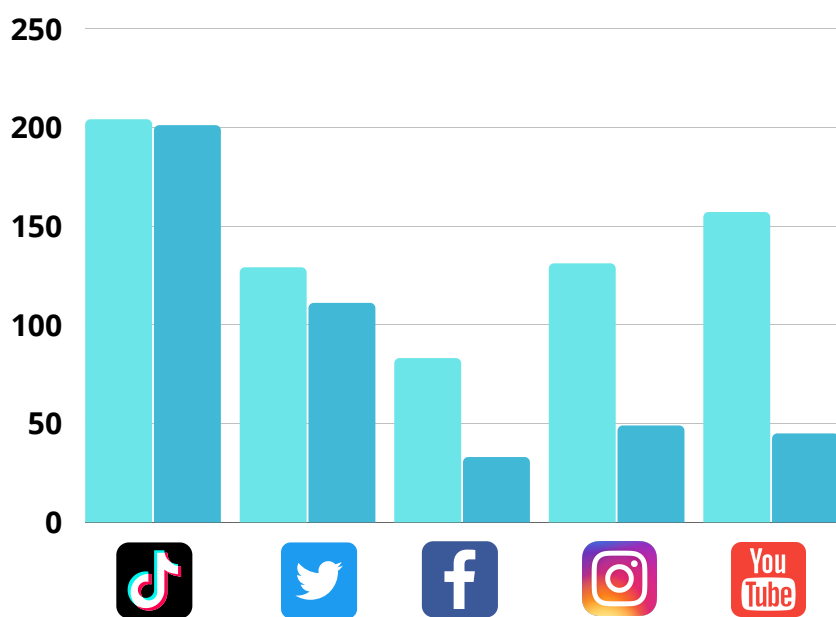


Figura 2. Número de contenidos retirados sobre el total de contenidos comunicados, según plataforma



En la Figura 2 se muestra el total de contenidos comunicados (en celeste) junto al número de contenidos retirados (en azul oscuro)



201 contenidos retirados de 204 comunicados (98,5%)



111 contenidos retirados de 129 comunicados (86%)



33 contenidos retirados de 83 comunicados (39,8%)



49 contenidos retirados de 131 comunicados (37,4%)



45 contenidos retirados de 157 comunicados (28,7%)

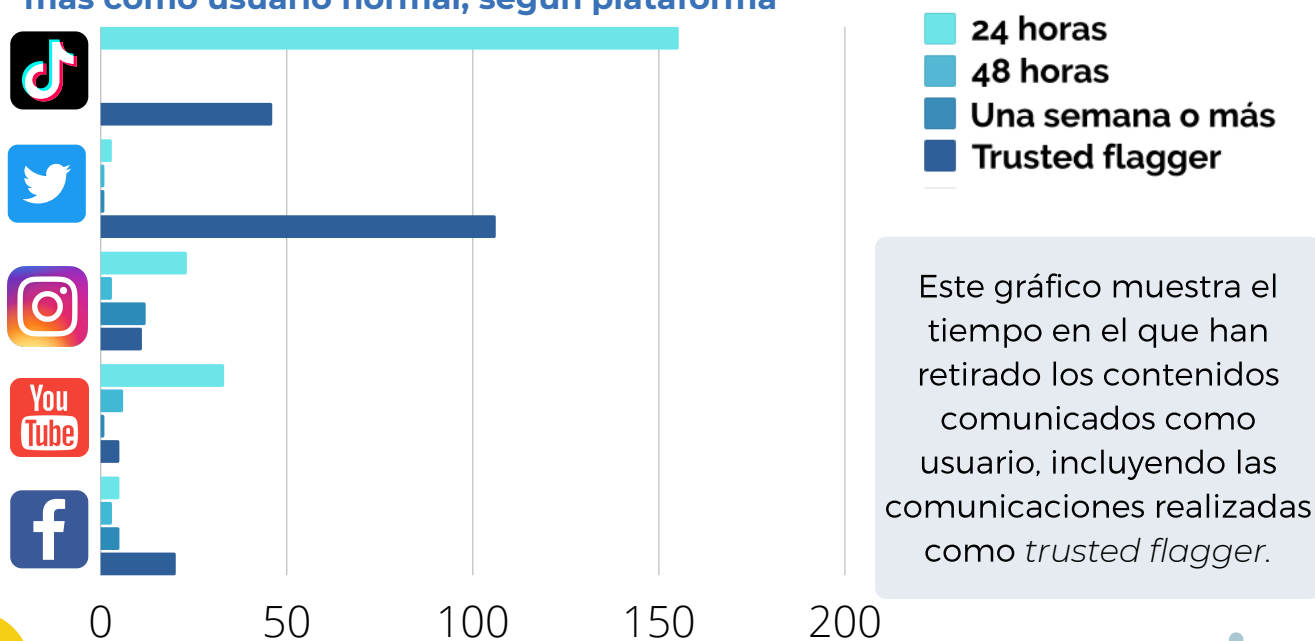
NÚMERO DE CONTENIDOS RETIRADOS A LAS 24H, 48H O UNA SEMANA Y COMO TRUSTED FLAGGER

En la monitorización, se realiza un seguimiento para comprobar cuánto tiempo les lleva a las plataformas tomar una decisión sobre la retirada o no del contenido comunicado. La Figura 3 recoge el **plazo temporal en el que las plataformas eliminan el contenido**: 24 horas, 48 horas, una semana o más. También se incorpora el número de comunicaciones eliminadas como *trusted flagger**.

Durante este ejercicio la eliminación de los contenidos ha sido más dispersa en el tiempo, y hay importantes cambios en algunas plataformas. No obstante, la tendencia a eliminar el grueso de los contenidos a las 24 horas desde su notificación permanece, a excepción de Twitter que retiró la mayoría de contenidos tras comunicarlos como *trusted flagger*. Por plataformas, **TikTok** eliminó el **76%** de los contenidos comunicados en el primer día, seguido de **Youtube (21%)**, **Instagram (17,6%)**, **Facebook (6%)** y **Twitter (2,3%)**.

En el presente ejercicio destaca que plataformas como **Facebook y Twitter retiraron gran parte de los contenidos comunicados como *trusted flagger*** y no como usuario normal. Para la plataforma Twitter esta tendencia ha sido constante a lo largo del año (2021). En cambio para Facebook supone **un cambio en tendencia preocupante** dado que hasta el momento gran parte del contenido se retiraba a las 24 horas de haber sido comunicado en calidad de usuario normal.

Figura 3. Contenidos retirados a las 24 horas, a las 48 horas, a la semana o más como usuario normal, según plataforma



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



CONTENIDOS DE DISCURSO DE ODIOS SEGÚN MOTIVO DE DISCRIMINACIÓN EN CADA RED SOCIAL

Los resultados de la monitorización muestran que el discurso de odio en las plataformas estudiadas sigue estando influido en gran medida por la agenda mediática y política del momento.

Los casos de **anti-inmigración** (Figura 4) han aumentado ligeramente con respecto al boletín anterior (30,8% en julio-agosto). Los contenidos contra los **menores extranjeros no acompañados** no solo han pasado a ser el tercer motivo de discriminación más común, sino que también **descienden en 5,8 puntos** respecto al periodo anterior (30% en julio-agosto).

Destaca el aumento de **contenido xenófobo** dirigido hacia la población **magrebí**, que **se incrementa en 7,6 puntos porcentuales** (16% en el boletín anterior). Los contenidos de xenofobia hacia las personas de origen europeo (0,5% en julio-agosto) y religión musulmana (14,5% en julio-agosto) vuelven a descender, junto con los contenidos anti-semitas (2,6% en julio-agosto).

Figura 4. Contenidos de discurso de odio según motivo de discriminación

