1 JULIO - 30 SEPTIEMBRE 2025

Boletín de monitorización del discurso de odio en redes sociales

La monitorización del discurso de odio realizada por el Observatorio Español del Racismo y la Xenofobia (OBERAXE) desde el año 2020 consiste en la identificación, análisis y notificación a las plataformas de contenidos de discurso de odio con motivación racista, xenófoba, islamófoba, antisemita y antigitana, publicados en cinco plataformas de redes sociales (Facebook, Instagram, TikTok, YouTube y X); y que puedan ser constitutivos de delito, de infracción administrativa, o que infrinjan las normas de uso de las propias plataformas de prestación de servicios digitales.

La base de la metodología de monitorización parte del modelo establecido en los ejercicios de evaluación del cumplimiento del *Código de Conducta para la lucha contra la incitación ilegal al odio en Internet*, firmado en 2016 por la Comisión Europea junto con las plataformas de prestación de servicios digitales; y renovado en 2025 a través del *Código de Conducta +.*

El convenio de colaboración firmado entre el Ministerio de Inclusión, Seguridad Social y Migraciones y LALIGA, ha permitido al OBERAXE profundizar y multiplicar el alcance del trabajo realizado gracias al Sistema FARO (Filtrado y Análisis de Odio en las Redes Sociales). Un sistema que aplica la inteligencia artificial, entrenada en el Monitor para la Observación del Odio en el Deporte (MOOD) de LALIGA, a la metodología, especialización y experiencia acumulada por el OBERAXE. El Sistema FARO permite identificar y analizar en tiempo real los discursos de odio racistas y xenófobos en redes sociales, facilitando así la detección de los acontecimientos sociopolíticos que suscitan y amplifican estos discursos.

Nota 1: Todos los gráficos y análisis presentados en este boletín fueron elaborados con datos del Sistema FARO (elaboración propia).
Nota 2: Los datos presentados en este boletín deben ser interpretados con cautela, dado que el Sistema Faro se lanzó en marzo de 2025 y todavía está en fase de optimización de la herramienta de inteligencia artificial.

GOBIERNO DE ESPAÑA · MINISTERIO DE INCLUSIÓN, SEGURIDAD SOCIAL Y MIGRACIONES · SECRETARÍA DE ESTADO DE MIGRACIONES · oberaxe Observatorio Español del Racismo y la Xenofobia · LALIGA · FUNDACIÓN LALIGA · Cofinanciado por la Unión Europea

Contenidos monitorizados

Durante el periodo comprendido entre el 1 de julio y el 30 de septiembre de 2025, el monitor FARO detectó 331.817 contenidos de odio reportable, y las plataformas retiraron el 45% de los contenidos reportados. El discurso de odio se dirigió principalmente hacia personas del norte de África y personas musulmanas. A lo largo del trimestre, diversos acontecimientos suscitaron un aumento de estos discursos, lo que sugiere cómo la percepción de la inseguridad y la desinformación pueden amplificarse a través de determinados discursos públicos y mediáticos, favoreciendo un clima de hostilidad que impacta directamente en la convivencia.

331.817
Mensajes detectados

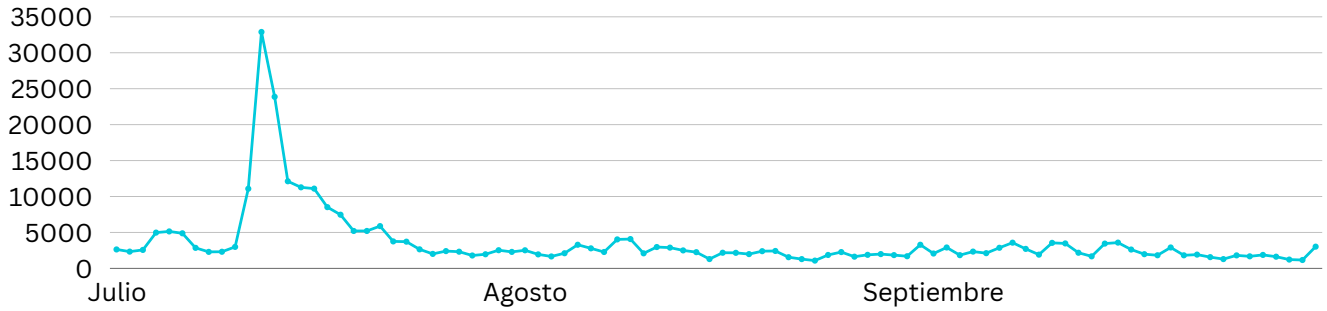
45%
% Mensajes retirados

Evolución de los contenidos detectados

Durante el transcurso del tercer trimestre de 2025, se ha observado una variación significativa en los contenidos de odio detectados, con un incremento especialmente notable en el mes de julio. En ese periodo, se registró un aumento de los contenidos de discurso de odio, alcanzando un máximo de 32.892 contenidos el 12 de julio, coincidiendo con la cobertura mediática del suceso de Torre Pacheco (Murcia), lo que podría estar vinculado a este pico puntual.

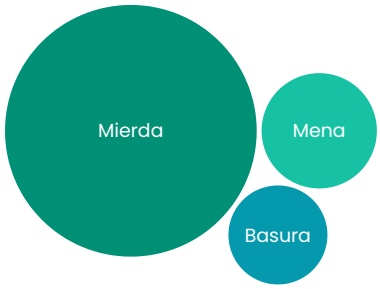
Tras este episodio, el volumen de contenidos descendió de manera progresiva y se mantuvo por debajo de los 5.000 contenidos durante los meses de agosto y septiembre.

En comparación con el trimestre anterior (1 de abril a 30 de junio), durante el cual se registraron 184.096 contenidos, el volumen del tercer trimestre casi se duplicó, alcanzando así el nivel más alto de actividad desde el inicio del año.



Palabras clave

El monitor FARO ha detectado que las principales palabras clave en los contenidos que contienen discurso de odio durante este trimestre son las siguientes:

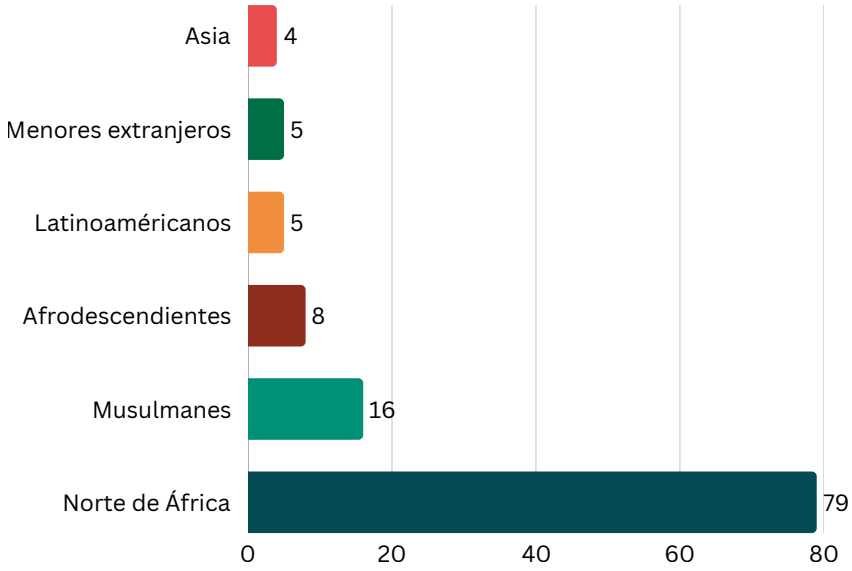


Consulta aquí los datos de monitorización en tiempo real

Características del discurso de odio

Grupo diana

En el tercer trimestre de 2025, los contenidos de odio se dirigieron principalmente a personas del norte de África, cuya proporción pasó del 73% al 79%, seis puntos porcentuales más que en el trimestre anterior. Le siguen las personas musulmanas (16%) y las personas africanas y afrodescendientes (8%)



5 %

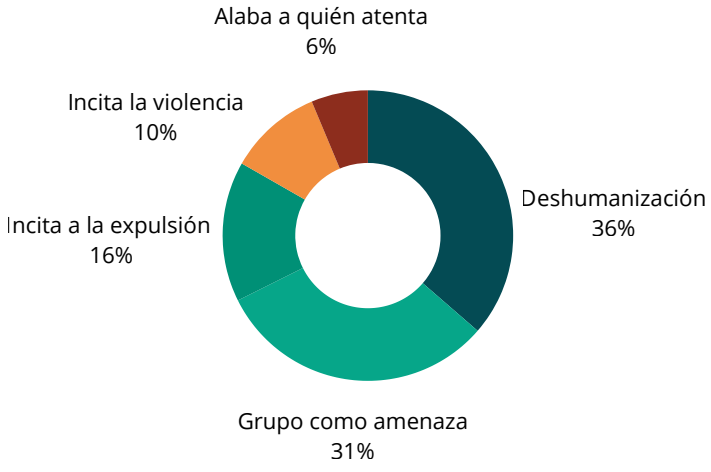
Los contenidos hacia niños, niñas y adolescentes no acompañados también aumentaron, alcanzado el el 5%, posiblemente vinculados al suceso del centro de menores de Hortaleza (Madrid).

Tipo de contenido

Los contenidos de odio muestran diversas formas de estigmatización que refuerzan la hostilidad hacia las personas de origen extranjero, predominando los mensajes que deshumanizan o degradan (36%), aunque han disminuido 18 puntos porcentuales respecto al trimestre anterior.

Una proporción significativa presenta a estas personas como una amenaza para la seguridad y la convivencia (31%) o promueve su expulsión (1%), contribuyendo a normalizar actitudes de rechazo social.

Además, un 10% incita directamente a la violencia, y un 6% elogia a quienes difunden mensajes o acciones violentas, evidenciando cómo la estigmatización y la incitación a la agresión se entrelazan, reforzando patrones de hostilidad que afectan la percepción de convivencia.



Consulta aquí los datos de monitorización en tiempo real

Expresión del lenguaje

El lenguaje agresivo explícito se ha utilizado en el 91% de los contenidos de discurso de odio detectados, recurriendo a insultos, amenazas y descalificaciones que reflejan una normalización de la agresividad verbal en las redes sociales. Por otro lado, el 9% de los mensajes emplea ironía o sarcasmo, ocultando el odio tras el humor, lo que complica su detección automática y evade los mecanismos de control de las plataformas.

Es importante señalar que aproximadamente el 19% de los contenidos reportados utiliza un lenguaje codificado, en el que se combinan letras, números, símbolos o emojis para eludir la censura, con mensajes como: "*el negr☻ suci☻*" o "*bombard3£#ar ese nido de moruegos*".

Se observa un uso recurrente de emojis con carga simbólica para reforzar la deshumanización o la violencia hacia las personas migrantes. Por ejemplo, los emojis de animales se utilizan para deshumanizar ("*Que pinta de 🐷 tiene*"), mientras que los de fuego expresan agresión o violencia contra personas de origen extranjero, como en: "🔥 *vivo y deportado después*" o "🔥 *a todos los moros*". Este tipo de recursos muestra cómo el discurso de odio adopta formas codificadas y visuales, facilitando la difusión de mensajes agresivos y normalizando actitudes hostiles .

Contenido viral

Esta publicación de la red social X ha obtenido 221 mil visualizaciones. Se trata de un video viral en el que se muestran jóvenes portando armas blancas frente a una cámara. El video incluye un texto que criminaliza a niños, niñas y adolescentes no acompañados, utilizando de forma despectiva la expresión "mena" y atribuyéndoles hurtos u otros actos violentos.

Consulta aquí los datos de monitorización en tiempo real

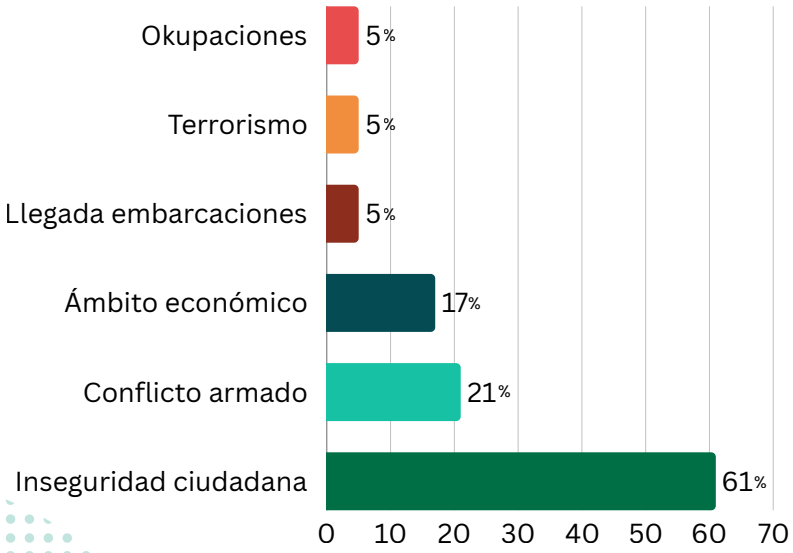
Episodios que suscitan discurso de odio

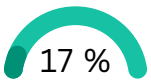
Durante el tercer trimestre de 2025, el discurso de odio en las principales plataformas ha estado relacionado con los episodios prototípicos de inseguridad ciudadana, conflicto armado y ámbito económico.

La inseguridad ciudadana representó el 61% de los contenidos de odio, manteniéndose como uno de los principales detonantes pese a un leve descenso respecto al trimestre anterior. Los mensajes se dirigieron principalmente hacia personas del norte de África, musulmanas, y africanas o afrodescendientes, perpetuando estereotipos que asocian migración con violencia y delincuencia.

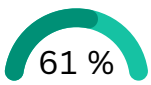
En julio, los sucesos de Torre Pacheco (Murcia), donde un vecino fue agredido presuntamente por un grupo de jóvenes del norte de África, generaron una intensa oleada de contenidos hostiles que promovían la violencia y la expulsión de personas migrantes. Entre los mensajes se incluían llamadas a la creación de patrullas ciudadanas y a ataques directos contra las personas de origen extranjero, alcanzando un pico extraordinario el 12 de julio, con contenidos como: *"localizarlos y acabar con ellos"; "fuera infraseres"; "solo el pueblo puede parar a estas ratas"* o *"pistolas y a la calle para acabar con los moros"*.

Pocos días después, el 16 de julio, la difusión de información falsa sobre la supuesta detención de un joven migrante en La Isleta (Gran Canaria) por la presunta agresión a su pareja menor intensificó aún más la percepción de inseguridad y la hostilidad hacia los grupos diana. La noticia generó mensajes que promovían la violencia y la expulsión, aunque posteriormente se confirmó que el joven había intentado auxiliar a la menor.





Los discursos de odio relacionados con conflicto armado (21%) y ámbito económico (17%) se mantienen sin cambios respecto al trimestre anterior.



La inseguridad ciudadana es el principal detonante de discurso de odio alcanzando su máximo el 12 de julio.

Consulta aquí los datos de monitorización en tiempo real

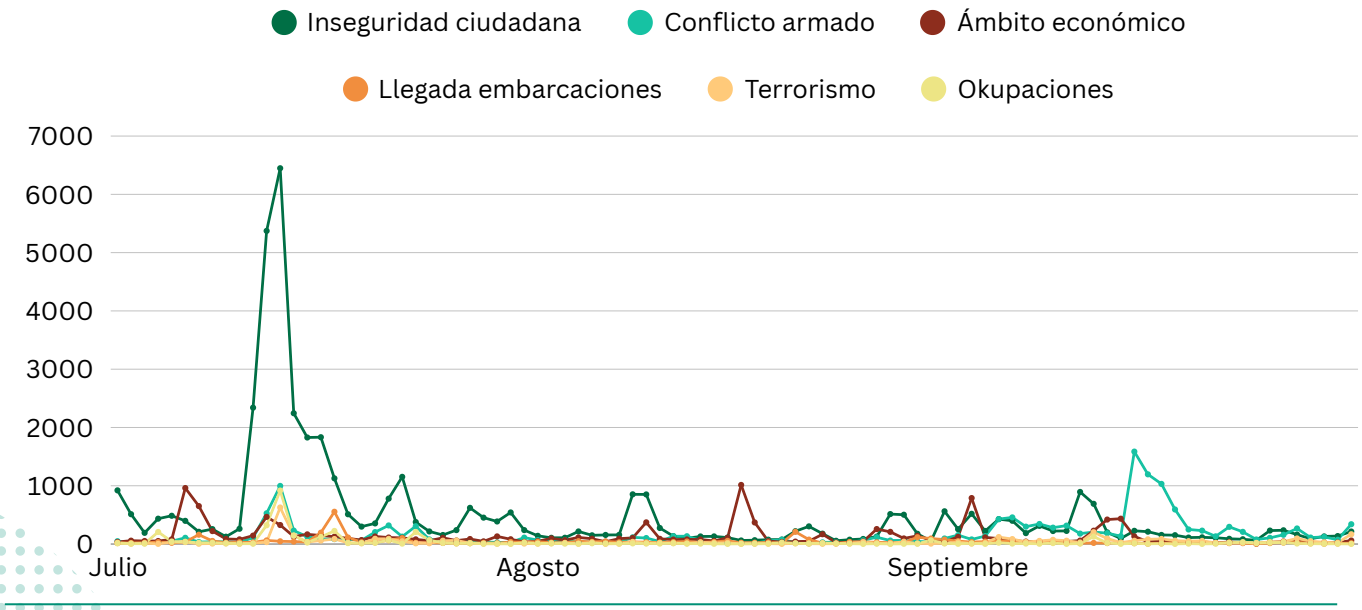
En agosto, otro suceso como la detención de un menor residente en el centro de acogida de Hortaleza (Madrid) por su presunta implicación en una agresión sexual, provocó una gran concentración de comentarios que relacionaban a la inmigración con la delincuencia y la inseguridad. La noticia continuó generando mensajes violentos durante septiembre, incitando a la expulsión de personas migrantes y reforzando estereotipos negativos con comentarios como: *"fuera menas" o "fuera parásitos invasores", "den caza al moro" o "fuera todos los moros"*.

El conflicto armado entre Israel y Palestina, junto con la cobertura de manifestaciones pro-palestinas, la Vuelta Ciclista España y la "flotilla" humanitaria a Gaza, generó un 21% de los contenidos de odio. Los mensajes adoptaron un tono islamófobo y antisemita, criminalizando a la población y justificando la violencia. Muchos de ellos expresaban hostilidad hacia los grupos diana y vinculaban su identidad religiosa o nacional con actos terroristas *"los terrorista son los palestinos"* o *"judíos genocidas"*, evidenciando cómo los conflictos internacionales pueden reforzar prejuicios locales y polarizar debates en redes sociales.

En el ámbito económico (17%), se observaron picos de discurso de odio relacionados con la percepción de que los migrantes reciben recursos públicos, intensificados durante la gestión de incendios en distintas comunidades y tras el anuncio de ayudas a la población palestina. Comentarios como *"Invasores que viven de la paguita y que traen delincuencia"* o *"Son parásitos que quieren dinero gratis"* reflejan cómo las narrativas económicas pueden instrumentalizarse para generar rechazo social hacia grupos vulnerables.

Finalmente, la llegada de personas migrantes en embarcaciones representó un 5% de los contenidos de odio, con un pico el 17 de julio. Los mensajes criminalizaban e incitaban a la expulsión y violencia, incluyendo expresiones extremadamente deshumanizadoras: *"moro ilegal asesino"*; *"despellejarlos vivos y bañarlos en lejía"* o "d*eportación masiva en patera*".

Evolución de los episodios prototípicos

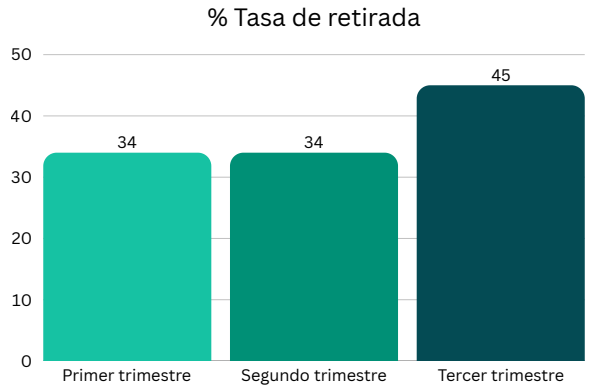


Consulta aquí los datos de monitorización en tiempo real

Reacciones de las plataformas

En el tercer trimestre, las plataformas han retirado el 45% de las notificaciones realizadas por el OBERAXE, lo que supone un incremento de más de 11 puntos porcentuales respecto al trimestre anterior (34%).
Este aumento refleja que, en el último trimestre, las plataformas eliminaron casi la mitad de los contenidos notificados, lo que evidencia un refuerzo en sus mecanismos de moderación, intensificado tras los sucesos ocurridos en Torre Pacheco (Murcia) en julio de 2025. En esa ocasión, el notable incremento de discursos de odio en redes sociales y el conflicto social derivado del suceso llevó a la Ministra de Inclusión, Seguridad Social y Migraciones a convocar con carácter urgente una reunión con las principales plataformas digitales y otras instituciones públicas, con el fin de coordinar medidas para frenar la difusión de contenidos que incitan al odio y la violencia hacia las personas de origen extranjero. En septiembre de 2025, una segunda reunión del grupo de trabajo permitió evaluar los avances y acordar encuentros trimestrales regulares para mantener un seguimiento constante, así como reuniones bilaterales con cada una de las plataformas.

La media acumulada ponderada de la tasa de retirada de contenidos notificados por el OBERAXE hasta el 30 de septiembre alcanza el 38%.

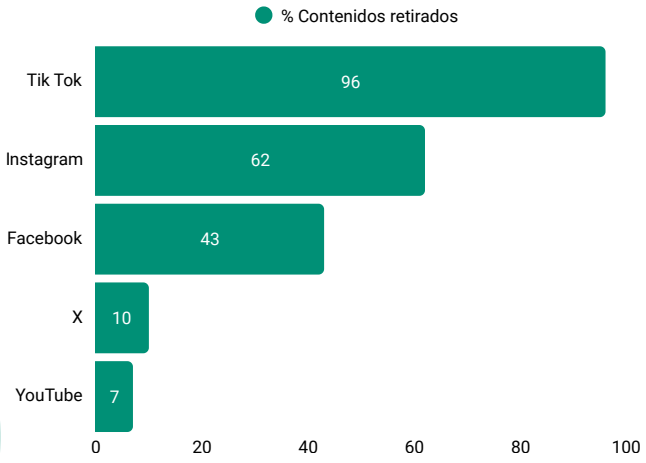


La experiencia derivada de los episodios de Torre Pacheco ha servido como un punto de inflexión para fortalecer los mecanismos de prevención y reacción frente a los discursos de odio en redes sociales, poniendo de manifiesto la importancia de la cooperación entre autoridades y plataformas para preservar la convivencia democrática, proteger a los grupos diana y garantizar que las redes sociales sean un espacio seguro, libre de violencia y discriminación.

En cuanto a el tiempo de respuesta que constituye un indicador relevante de eficacia, solo un 8% de los contenidos fueron retirados en menos de 24 horas, y un 1% adicional en las siguientes 48 horas. Un 3% fue eliminado en el plazo de una semana, mientras que el 33% restante se retiró a través de la vía de *trusted flagger*.

En conjunto, la vía de comunicante fiable ha sido determinante, retirándose casi tres cuartas partes de los contenidos a través de esta vía (74%). En contraste, las notificaciones retiradas por un usuario normal apenas alcanzaron el 12%. Este patrón evidencia como las plataformas son más efectivas cuando se emplea el canal de comunicante fiable, así como la limitada capacidad del usuario común para promover la retirada de contenidos hostiles.

A continuación, se muestra el gráfico de la tasa de retirada de cada plataforma en relación con las comunicaciones de contenidos de discurso de odio efectuadas a cada una de ellas.

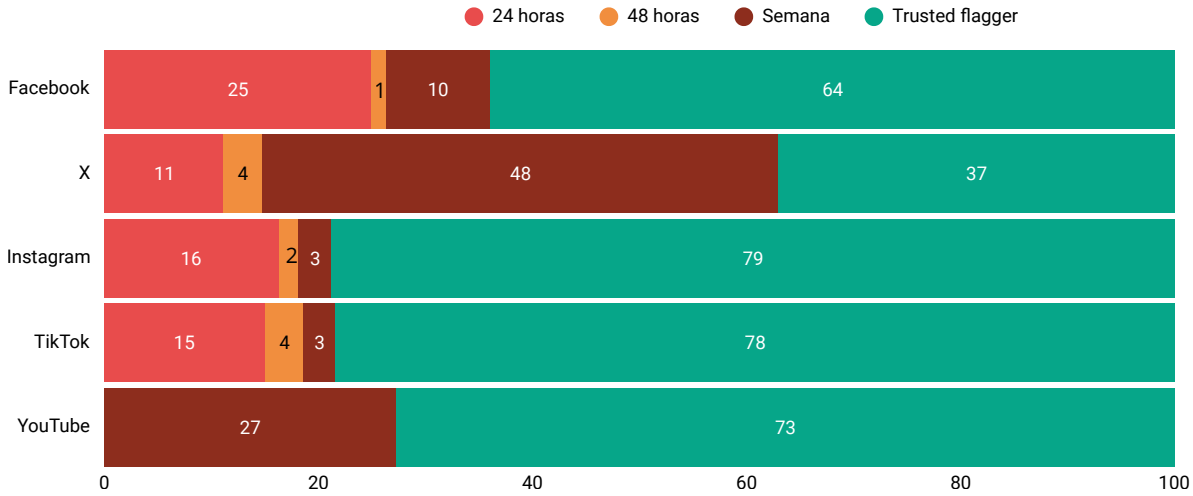


TikTok se sitúa como la plataforma más eficaz, retirando el 96% de los contenidos notificados.

Evolución en la retirada de contenidos

Al realizar un análisis de los tiempos de retirada de contenidos en las distintas plataformas, se observa que Facebook destaca con una tasa de retirada del 25% de los mensajes y/o imágenes en las primeras 24 horas. Por otra parte, X es la red social que más contenidos retira en el transcurso de una semana con un 48%, seguido de YouTube con un 27%.

Por otro lado, Instagram y TikTok presentan más efectividad cuando utilizan la vía de comunicante fiable (*trusted flagger*), con tasas de retirada superiores al 78%. YouTube y Facebook también muestran una elevada proporción de retiradas a través de esta vía, con un 73% y un 64% respectivamente, mientras que en X esta cifra es considerablemente menor (37%).



Edita y distribuye: Observatorio Español del Racismo y la Xenofobia (OBERAXE)

NIPO: 121-23-006-3

GOBIERNO DE ESPAÑA
MINISTERIO DE INCLUSIÓN, SEGURIDAD SOCIAL Y MIGRACIONES
SECRETARÍA DE ESTADO DE MIGRACIONES
oberaxe Observatorio Español del Racismo y la Xenofobia
LALIGA
FUNDACIÓN LALIGA
Cofinanciado por la Unión Europea