Agosto 2026

Boletín mensual de monitorización del discurso de odio en redes sociales

Este boletín tiene como objetivo analizar los contenidos de discurso de odio que circulan por las redes sociales, monitorizados y notificados a las plataformas de internet (Facebook, Instagram, X, TikTok y YouTube) en el mes de agosto de 2026. Así, los datos permiten explorar tanto las características y patrones del discurso como los grupos de población a los que está dirigido. En particular, se analiza la reacción de las plataformas ante la retirada de contenido, los grupos diana objeto del discurso de odio, los episodios prototípicos que lo suscitan y el tipo de contenido y lenguaje utilizado en los mensajes de discurso de odio.

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 encuentra en fase de optimización de su herramienta de inteligencia artificial.

Contenidos monitorizados

En el mes de agosto de 2026, el monitor FARO detectó un total de 101.025 contenidos de discurso de odio y narrativas discriminatorias en redes sociales, una cifra notablemente superior a la registrada en los meses anteriores.

Sobre la muestra de contenidos de discurso de odio seleccionados y reportados a las plataformas, la tasa de retirada se situó en 73%, lo que supone 3 puntos porcentuales más que el mes anterior (70%).

101.025
Mensajes detectados

73%
% Mensajes retirados

Contenidos retirados

Las plataformas eliminaron el 10% de los contenidos notificados como usuario normal, un descenso de 2 puntos porcentuales respecto al mes de julio (12%). En cuanto a los tiempos de respuesta, el 7% de los contenidos se retiró durante las primeras 24 horas, un 2% adicional en las siguientes 48 horas y el 1% restante a lo largo de la primera semana desde su notificación.

En relación con la vía *trusted flagger*, las plataformas retiraron el 63% de los contenidos reportados, lo que representa un incremento de 5 puntos porcentuales respecto al mes anterior (58%).

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

Los resultados reflejan una mayor eficacia de la vía *trusted flagger* frente a las notificaciones de usuario normal. Para seguir avanzando en la mejora de estos procesos, se mantiene un trabajo conjunto con las plataformas a través del grupo de trabajo sobre discurso de odio en redes sociales, reuniones bilaterales y canales de comunicación directa.

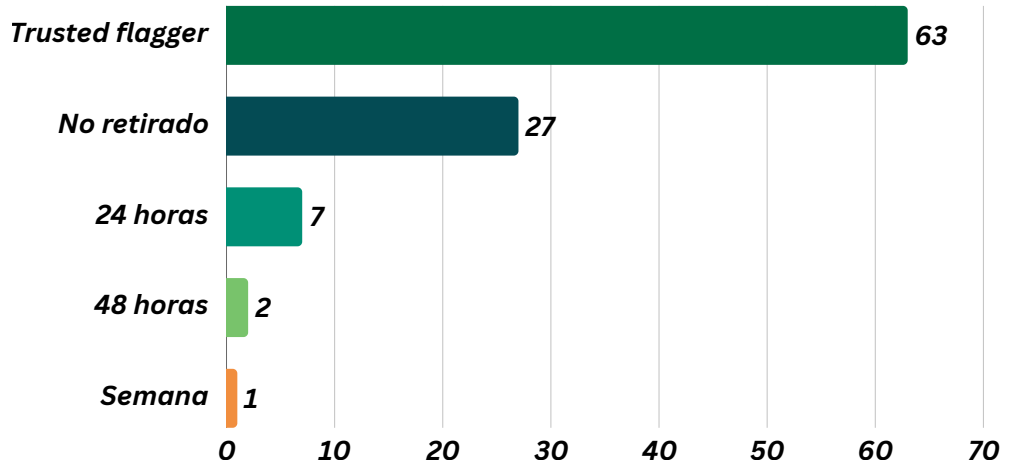


Gráfico reacción de las plataformas ante la retirada de contenido (%)

100 %

TikTok retiró el 100% de los contenidos notificados, de los cuales el 32% se eliminó en las primeras 24 horas tras su notificación como usuario normal.

68 %

Facebook eliminó el 68% de los contenidos notificados, un 94% de ellos mediante *trusted flagger.*

72 %

Instagram alcanzó una tasa de retirada del 72%, el 100% de ellos a través de la vía *trusted flagger*.

93 %

X registró una tasa de retirada del 93%, concentrándose el 98% de las retiradas en la vía *trusted flagger*.

8 %

YouTube alcanzó una tasa de retirada del 8% sobre los comentarios reportados, correspondiendo el 100% de las retiradas a la vía *trusted flagger*.

Consulta aquí los datos de monitorización en tiempo real

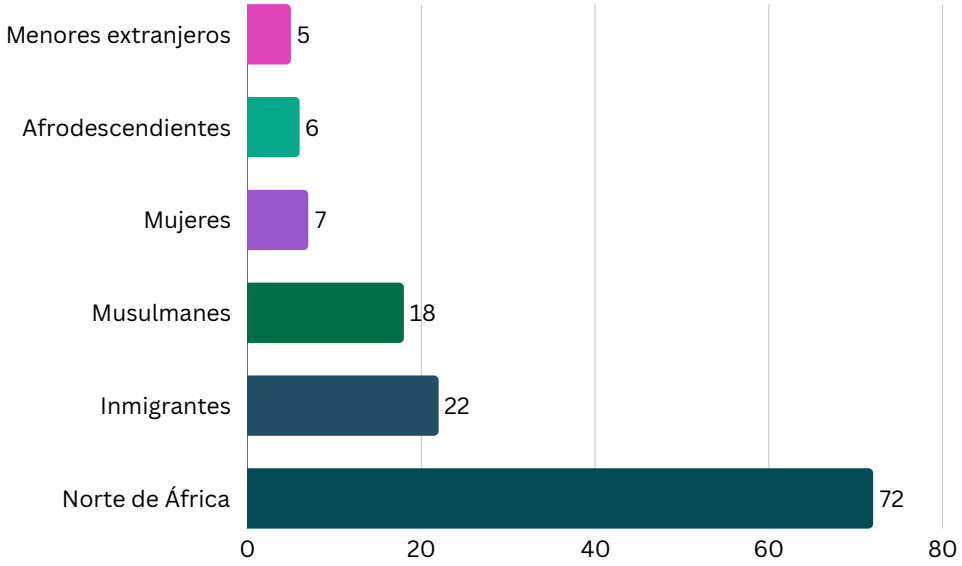
Características del discurso de odio

Grupo diana

Durante el mes de agosto, la hostilidad en redes sociales continuó dirigiéndose principalmente hacia las personas del norte de África, que concentraron el 72% de los contenidos de discurso de odio analizados, 4 puntos porcentuales más que en julio (68%).

Por otra parte, los mensajes dirigidos contra las personas inmigrantes en general aumentaron hasta alcanzar el 22% del total analizado, mientras que las personas musulmanas concentraron el 18% de los contenidos detectados, 2 puntos porcentuales más que el mes anterior.

Asimismo, se detecta una mayor presencia de contenidos de discurso de odio dirigidos contra mujeres y contra niños, niñas y adolescentes no acompañados. Esta variación podría estar relacionada, entre otros factores, con los acontecimientos ocurridos en Ceuta, tal y como se analiza en el apartado de episodios prototípicos.



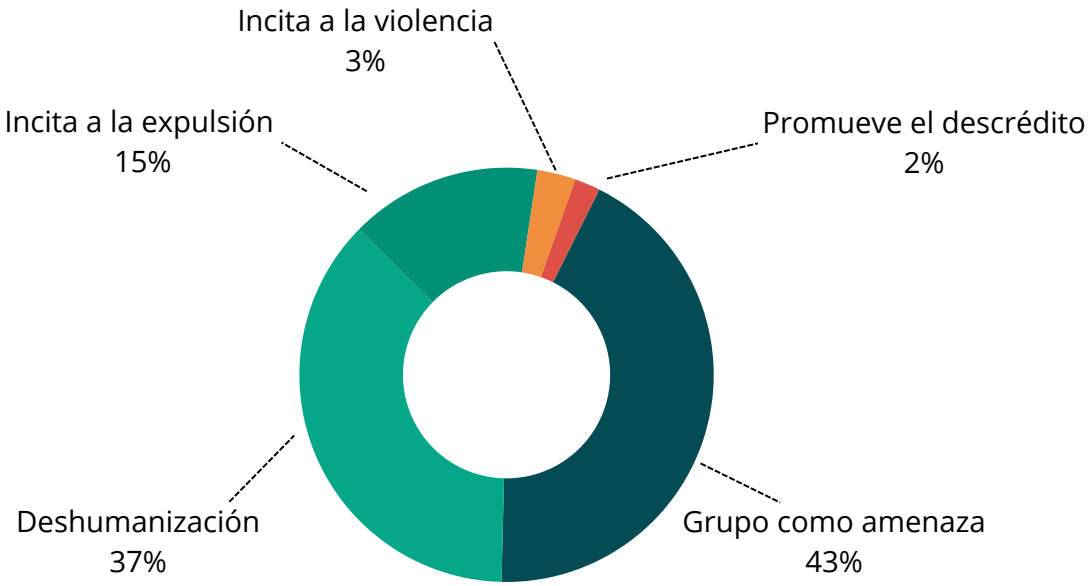
Tipo de contenido

En el mes de agosto, los contenidos que presentan a los grupos diana como una amenaza alcanzan un 43% del total, lo que supone un aumento de 5 puntos porcentuales con respecto al mes de julio (38%).

Asimismo, los mensajes que deshumanizan o degradan gravemente suponen el 37% del total de contenidos, experimentando un descenso de tres puntos porcentuales con respecto al mes de anterior (40%).

Por otro lado, las narrativas que incitan a la expulsión de los grupos diana representan un 15% del contenido, manteniéndose estable con respecto al mes anterior.

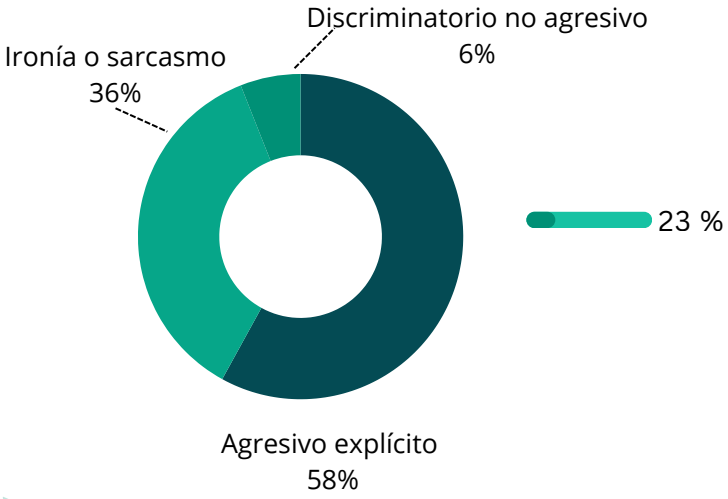
Además, el contenido que incita a la violencia supone un 3% del total, y, por último, las narrativas que promueven el descrédito contra los colectivos diana alcanzan el 2% del total de contenido.



Expresión del lenguaje

El 58% del contenido examinado recurre a un lenguaje agresivo explícito, que se manifiesta a través de insultos, amenazas y descalificaciones dirigidas contra los grupos diana, reflejando un elevado nivel de hostilidad y rechazo. Algunos ejemplos de este tipo de expresiones son: *"tiro al moro en Ceuta", "HP MOROS DE MIERDA", "musulmán muerto, el único bueno", "organicemos la purga"*.

Por otro lado, el 36% de los contenidos analizados emplea la ironía o el sarcasmo como recurso para difundir narrativas discriminatorias bajo una apariencia humorística. Este tipo de mensajes contribuye a la normalización de prejuicios y a la generación de un clima de hostilidad hacia los grupos diana, al tiempo que dificulta su identificación y moderación por parte de los sistemas automáticos de detección de las plataformas, con expresiones como: *"al fondo del mar hay asilo para todos", "venid que tenemos pisos y paguitas para todos", "os aplaudiremos por violar a españolas"*.



Finalmente, el 23% del contenido analizado incluye imágenes, vídeos, memes o el uso de lenguaje codificado, adoptando un formato de carácter visual. Este tipo de contenidos posee una mayor capacidad de difusión y de impacto y, simultáneamente, dificulta la identificación automática y precoz del discurso de odio por parte de las plataformas. Algunos ejemplos de este tipo de contenido son: *"cuantos 🐖 🧕", "hay que 🪨 🪨", "🪖⚔️tic tac", "moros 🤮🤮🤮"*

Consulta aquí los datos de monitorización en tiempo real

Episodios que suscitan discurso de odio

Durante el mes de agosto, el 39% de los contenidos analizados estuvo asociado al episodio de cruce fronterizo en Ceuta. En este sentido, la llegada de miles de personas procedentes de Marruecos a la Ciudad Autónoma de Ceuta a finales de julio, continuó generando un elevado volumen de mensajes en agosto, dirigidos especialmente contra las personas migrantes y las personas del norte de África, así como contra niños, niñas y adolescentes no acompañados, un grupo especialmente vulnerable. En este contexto, se detectaron numerosos contenidos deshumanizantes y degradantes, con mensajes como *"animales que no controlan instintos", "ratas islámicas"* o *"Marruecos descarga la basura en España"*. También se identificaron mensajes que promovían la expulsión de las personas migrantes, con referencias explícitas a la violencia, con expresiones como "*expulsión por las buenas o por las malas"* o *"si no los aceptan de vuelta, hay que devolverlos al mar*". Del mismo modo, se detectaron contenidos que incitaban a la violencia o la justificaban con mensajes como *"momento de salir a cazar", "balas contra la invasión" o "DESDE LA FRONTERA RATATATA"*. A finales de agosto, la quema de asentamientos en Ceuta generó además mensajes que celebraban estos ataques y promovían nuevas acciones hostiles, con mensajes como *"poco está ardiendo", "meterle fuego a todos", "fuera gentuza aunque sea quemándalos" o "cuando duermen, gasolina*". Las narrativas identificadas en torno a este episodio, incluidas aquellas que justifican o celebran la violencia contra las personas migrantes, refuerzan percepciones de amenaza e invasión que contribuyen a la estigmatización de los grupos diana y legitiman discursos de exclusión y violencia.

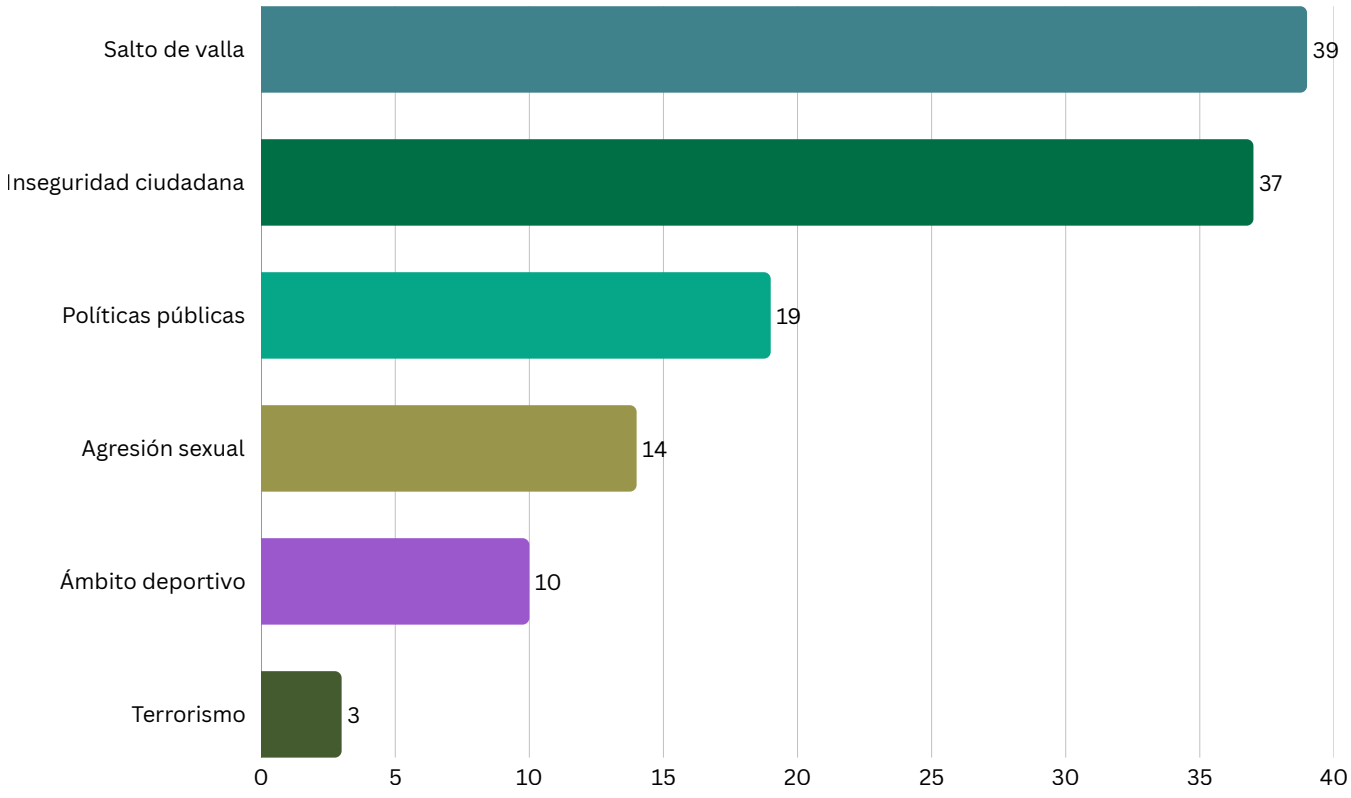
La inseguridad ciudadana representó el segundo episodio prototípico más frecuente, estando presente en el 37% de los contenidos analizados. En este ámbito, los contenidos generados en torno a la llegada de personas migrantes a Ceuta, y su permanencia en la ciudad, fueron utilizados para reforzar narrativas que presentaban a las personas migrantes como una amenaza para la seguridad y la convivencia, favoreciendo su criminalización colectiva mediante mensajes como *"solo han venido a invadir y a hacer daño", "vienen a matar cristianos y blancos" o "terroristas"*. Asimismo, se detectó la difusión de contenidos que asociaban la presencia de personas migrantes con riesgos para la salud pública, contribuyendo a reforzar estereotipos negativos y percepciones de alarma social. Este tipo de narrativas estigmatizantes combinan con frecuencia referencias a la inseguridad ciudadana con atribuciones relativas a enfermedades o conductas incívicas, reforzando la construcción de los grupos diana como una amenaza para la población, con mensajes como *"moros violadores con tuberculosis"*. Además, continuó observándose la circulación de contenidos descontextualizados sobre altercados, conflictos en el espacio público o incidentes ocurridos en otros lugares, que fueron utilizados para reforzar percepciones de inseguridad y alimentar narrativas hostiles hacia las personas migrantes.

En cuanto al discurso de odio asociado al episodio prototípico de las políticas públicas, este representó el 19% del total de los contenidos detectados durante el mes de agosto. Una parte relevante de los contenidos monitorizados surgió en torno al debate sobre los recursos públicos destinados a la atención y acogida de las personas migrantes llegadas a Ceuta. En este contexto, se difundieron narrativas que presentaban a las personas migrantes como beneficiarias indebidas de ayudas y servicios públicos, cuestionando la legitimidad de las medidas de acogida, con comentarios como "*paguita para ellos y a robar*" o "*son parásitos mantenidos*".

Asimismo, se identificaron mensajes que retrataban a las personas migrantes como una carga económica para las administraciones y para la población local, reforzando percepciones de agravio comparativo y contribuyendo a su estigmatización y exclusión social.

El episodio prototípico de agresión sexual concentró el 14% del total de los contenidos monitorizados. Una parte importante de estos mensajes se vinculó a informaciones difundidas en el contexto de la llegada de personas migrantes a Ceuta, así como a distintos sucesos ampliamente compartidos en redes sociales relacionados con agresiones sexuales. Estos contenidos fueron utilizados para asociar de forma generalizada la violencia sexual con las personas de origen extranjero. En particular, persisten las narrativas que presentan a las personas migrantes como una amenaza para la seguridad de las mujeres, mediante expresiones como "*violadores de nacimiento*", "*¿acogerles para delinquir y violar?*" o *"violadores que llegan de África"*. Estas publicaciones contribuyen a la criminalización colectiva de los grupos diana y refuerzan discursos xenófobos y de exclusión.

Nota: Por razones de protección de datos y responsabilidad institucional, los ejemplos de mensajes han sido omitidos o adaptados para evitar la reproducción literal de expresiones de odio.



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