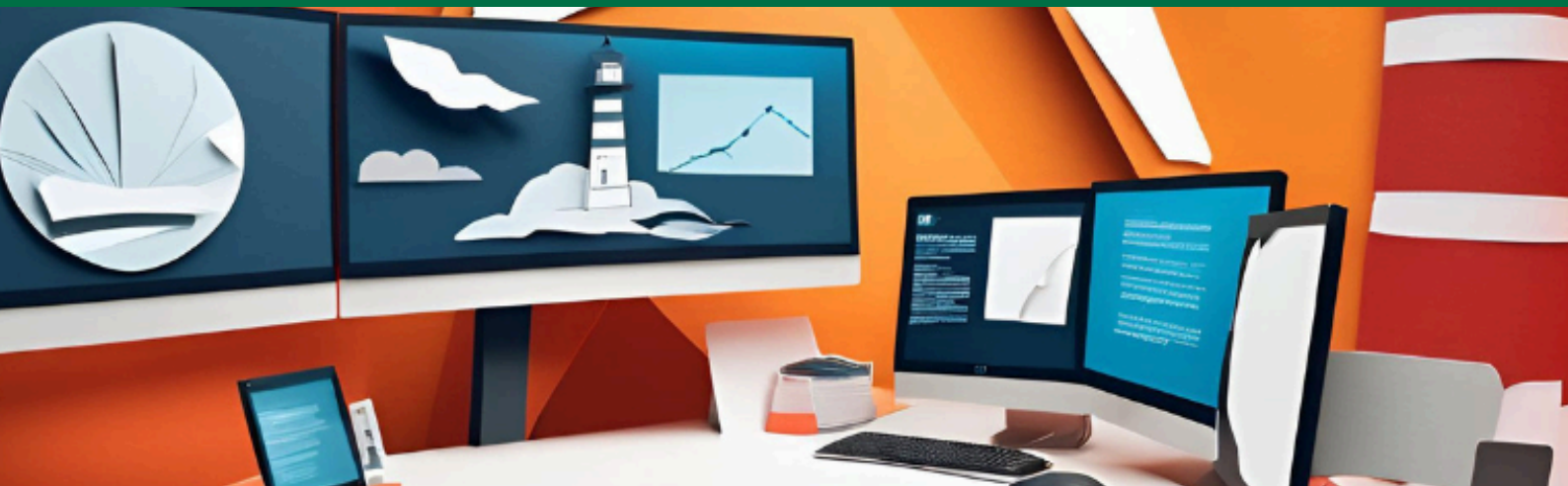


1 ABRIL - 30 JUNIO 2025

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: 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.

Contenidos monitorizados

Durante el periodo entre el 1 de abril y el 30 de junio de 2025, el monitor FARO ha detectado 184.096 contenidos de discurso de odio reportables, y las plataformas han retirado el 34% de los contenidos reportados. El discurso de odio se ha dirigido principalmente a las personas del norte de África y a las musulmanas. Durante el segundo trimestre de 2025, eventos como Eurovisión, la final de la Champions League y otros sucesos mediáticos generaron picos puntuales en los discursos de odio, centrados en narrativas sobre inseguridad, integración cultural y la percepción de ayudas sociales a personas migrantes.

184.096

Mensajes detectados

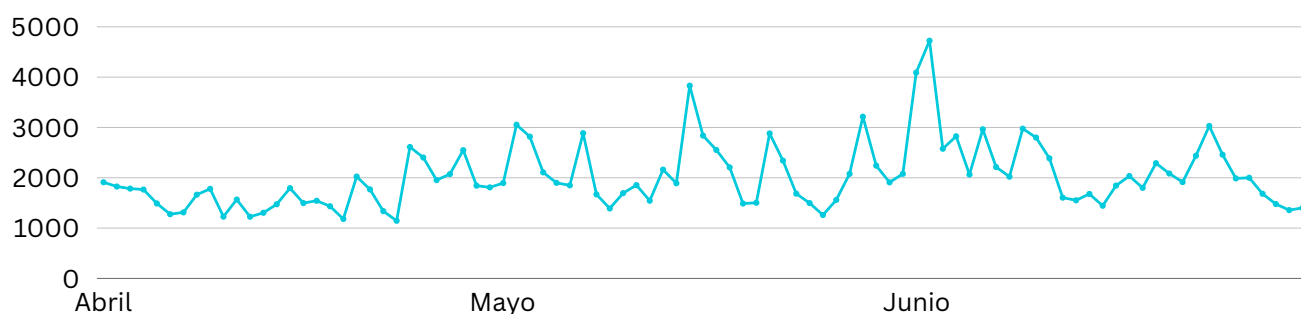
34%

% Mensajes retirados

Evolución de los contenidos detectados

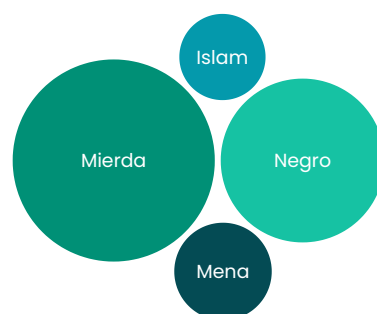
Durante el análisis del segundo trimestre de 2025, se observa una tendencia en la detección de contenidos con una media de 2.500 mensajes de odio.

En abril, los niveles de mensajes detectados se mantuvieron relativamente moderados y estables por debajo de los 2.000 mensajes en la mayoría de los días. En cambio, mayo presenta un incremento de los contenidos detectados con múltiples picos que superan los 3.000 mensajes y alcanza su máximo el 15 de mayo con 3.881 mensajes detectados. El mes de junio, continúa con esta tendencia al alza, con picos que superan los 4000 mensajes, alcanzando un pico máximo de 4.724 el 2 de junio. Sin embargo, a final de mes se observa una clara disminución en el volumen de mensajes. Estas variaciones sugieren la existencia de una relación causal con eventos sociopolíticos que serán detallados en el apartado referente a los episodios prototípicos.



Palabras clave

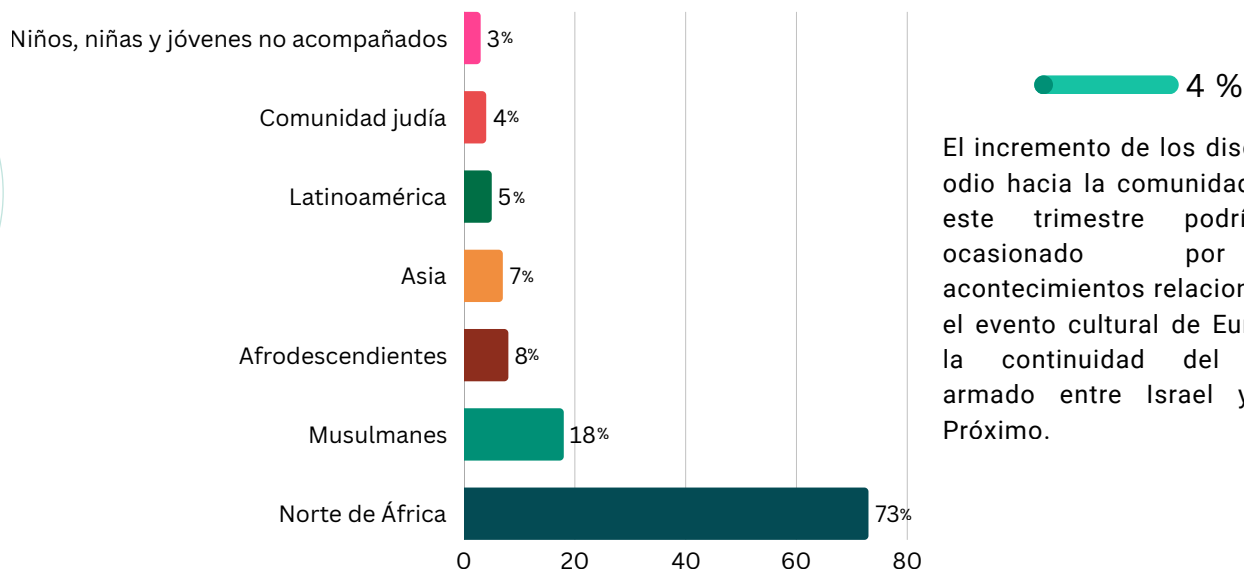
El monitor FARO ha detectado que las principales palabras clave en los contenidos que contienen discurso de odio durante este trimestre son las siguientes:



Características del discurso de odio

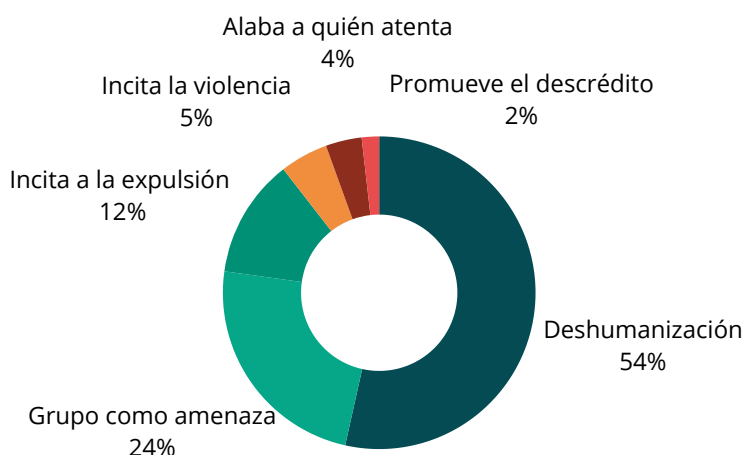
Grupo diana

En el segundo trimestre, las personas del norte de África continúan siendo el grupo diana que recibe la mayor cantidad de contenidos de hostilidad, representando el 73% de los casos. A continuación, se encuentran las personas musulmanas, que concentran el 18%, seguidas por las personas africanas y afrodescendientes, con un 8%.



Tipo de contenido

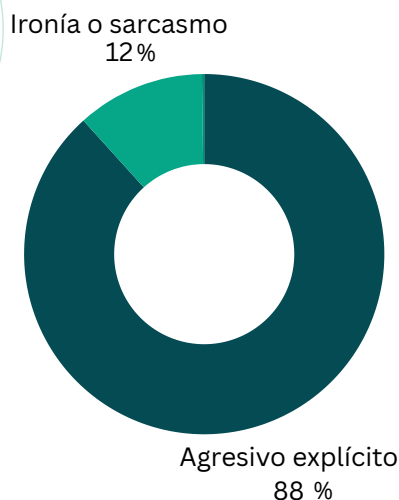
La deshumanización de los diferentes grupos diana vuelve a ser el tipo de contenido predominante (54%) durante el segundo trimestre. El 24% de los contenidos (4 puntos porcentuales más que el trimestre anterior) presenta a estos grupos como una amenaza para la sociedad y se incita a la expulsión de las personas de origen extranjero en un 12%. En un 5% se incita a la violencia contra estas personas y en un 4% se elogia a quien promueve mensajes y acciones violentas. Por otra parte, un 2% de los contenidos promueve el descrédito hacia los grupos diana a los que se dirige.



Expresión del lenguaje

Durante el segundo trimestre de 2025, el lenguaje agresivo y explícito ha estado presente en el 88% de los mensajes detectados. Estos contenidos suelen incluir insultos directos y expresiones de desprecio que llaman a la exclusión de ciertos grupos. Por otro lado, un 12% de los contenidos emplea ironía y sarcasmo para manifestar hostilidad, haciendo alusiones despectivas a hechos o personajes históricos.

Asimismo, es importante reflejar que aproximadamente un 20% de los mensajes reportados utiliza un lenguaje codificado, que incluye combinaciones de números y letras alteradas "#dèpòrt", así como símbolos y emojis específicos, como el emoticono de rata o expresiones que hacen referencia a términos médicos negativos, junto con iconografía religiosa y otras señales visuales "☠️*ANCER 🤢 🙈 🚫 🐭 🔥". Este tipo de codificación busca evadir la detección automática, manteniendo el mensaje de odio de manera cifrada.



Cabe destacar que, durante abril y mayo, hubo una tendencia detectada en plataformas como TikTok en relación con la proliferación del fenómeno conocido como "Brainrot" en la cual, mediante herramientas de inteligencia artificial, se generan videos que reproducen representaciones estereotipadas, caricaturizadas y ridiculizantes, sobre todo de personas procedentes del norte de África, perpetuando discursos xenófobos en formatos audiovisuales de amplia difusión y alto impacto. No obstante, esta tendencia disminuyó notablemente durante el mes de junio.

Contenido viral



Esta publicación de la plataforma X/Twitter ha obtenido 1.2 millones de visualizaciones. El contenido viral que acompaña la imagen describe una situación violenta, mencionando a "dos individuos del norte de África" como responsables de un acto violento. Asimismo, se incluyen expresiones que exaltan a los "jóvenes talentos fuera de España" y califican despectivamente a otras personas como "basura de otros países", reforzando así un discurso xenófobo y cargado de prejuicios.

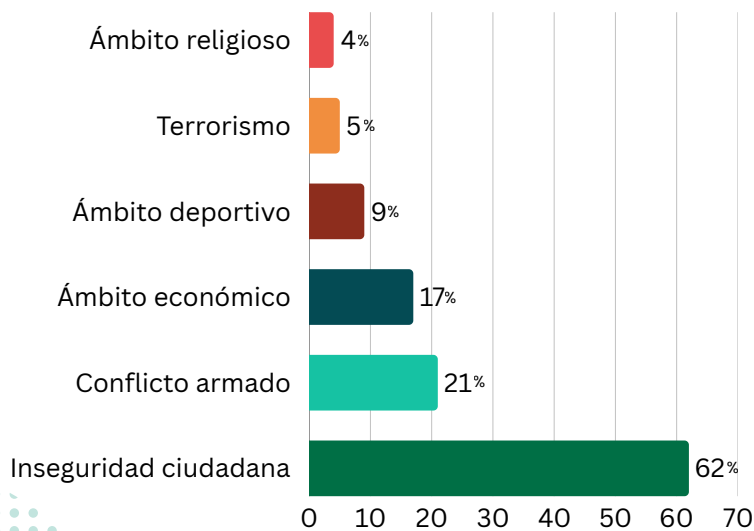
Episodios que suscitan discurso de odio

En el segundo trimestre de 2025, el discurso de odio en redes sociales está relacionado principalmente con los episodios prototípicos de: inseguridad ciudadana, conflicto armado y ámbito económico.

Los mensajes de odio monitorizados durante el segundo trimestre revelan que la inseguridad ciudadana (62%) sigue siendo el episodio prototípico predominante y con una tendencia al alza (aumentando en 5 puntos porcentuales respecto al trimestre anterior). Asimismo, los mensajes siguen siendo dirigidos hacia las personas del norte de África y musulmanas. Algunos casos de inseguridad ciudadana han estado relacionados con los hechos ocurridos el 26 de abril, donde un turista de origen asiático se defendió de un intento de robo presuntamente cometido por una persona del norte de África. Este suceso generó comentarios que vinculaban la inmigración con la delincuencia y que demandaban la expulsión de dicho grupo diana.

Se observa un aumento significativo en los contenidos que hacen referencia al conflicto armado (21%) de 6 puntos porcentuales más respecto al trimestre anterior. Este aumento en los mensajes reportados está relacionado con el evento cultural de Eurovisión durante los días 17 y 18 de mayo, la continuidad de los conflictos internacionales (Israel – Palestina; EE.UU. - Irán) y los disturbios de París tras la victoria del PSG en la final de la Champions League.

En cuanto al ámbito económico, se monitorizaron un 17% de los contenidos, que en su mayoría se relacionan con las subvenciones, ayudas sociales y políticas de inmigración por parte del Gobierno.



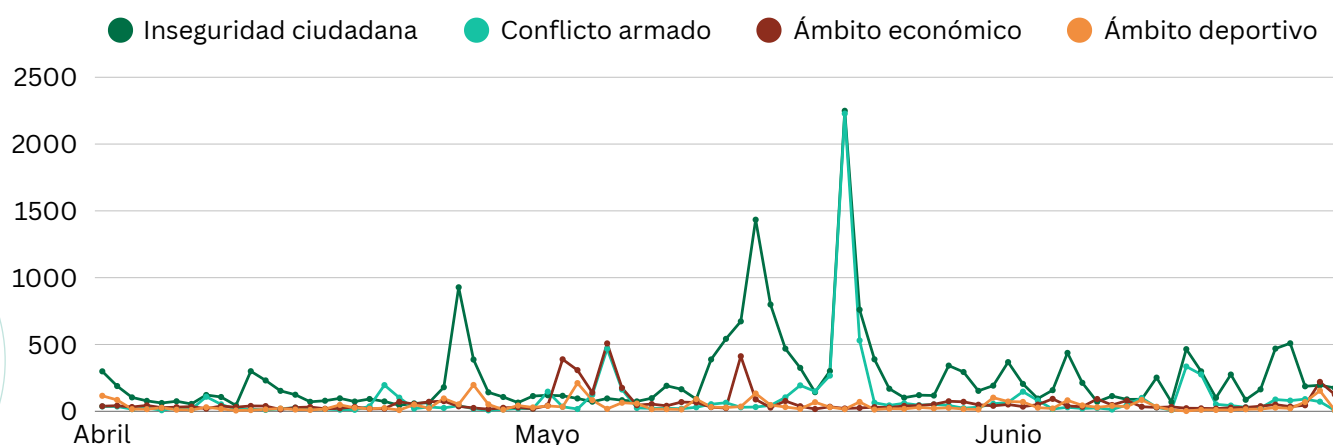
21 %

El aumento del discurso de odio relacionado con el conflicto armado sigue vinculado a los conflictos internacionales

43 %

Contenidos de inseguridad ciudadana que promueven la percepción, pero no están basados en hechos verídicos, actuales y ocurridos en España

Evolución de los episodios prototípicos



A lo largo del segundo trimestre de 2025, el análisis de los mensajes de odio detectados en redes sociales revela una evolución marcada por picos significativos vinculados a episodios concretos. Estos momentos de mayor actividad no son aleatorios, sino que responden a sucesos sociopolíticos o mediáticos que actúan como desencadenantes del discurso de odio. El gráfico refleja estos mensajes en correspondencia con los principales episodios prototípicos, evidenciando cómo ciertos eventos son empleados para fomentar la hostilidad hacia los diferentes grupos diana, los cuales se describen a continuación.

Al igual que el trimestre anterior, la inseguridad ciudadana sigue siendo la narrativa más recurrente que se utiliza como justificación para generar o difundir el discurso de odio. El primer pico relevante ocurrió el 25 de abril, tras la denuncia de una presunta agresión homófoba en el Bando de la Huerta (Murcia), atribuida a una persona del norte de África. La reacción en redes sociales no se limitó a condenar el hecho, sino que se desvió hacia una criminalización generalizada de las personas migrantes con mensajes que deshumanizan y aludiendo a la percepción de amenaza cultural y falta de integración con expresiones como: *“son animales sin domesticar”*.

Por otro lado, el partido de LALIGA entre el RCD Espanyol y el FC Barcelona celebrado el 15 de mayo generó numerosos comentarios de odio dirigidos hacia el jugador Lamine Yamal. Dichos mensajes emplearon un lenguaje explícitamente agresivo y utilizaron el término *“moro”* como un recurso simbólico para cuestionar su legitimidad como jugador del equipo nacional. Asimismo, se observaron expresiones deshumanizadoras que comparaban al jugador con una *“rata de cloaca”*.

También, se ha identificado un aumento significativo en las conversaciones de las diferentes plataformas durante los días 17 y 18 de mayo, vinculado al evento cultural de Eurovisión, donde el debate se centró en la participación de Israel.

Esta controversia dio lugar a la proliferación de mensajes antisemitas e islamófobos relacionados con el conflicto entre Israel y Palestina (Oriente Próximo), incluyendo expresiones como: *"moros y judíos son enemigos de Occidente"*; *"Israel es un estado sionista terrorista, son los mismos que nos traen inmigrantes a España"*; *"Islam fuera de España"*; *"El mundo sería mejor si el bigotón austriaco los hubiese exterminado (refiriéndose a Hitler)"*.

Por otro lado, el 21 de mayo se registra el pico de contenidos más importante del trimestre. Los mensajes están vinculados a una noticia sobre la solicitud de un grupo de personas musulmanas para utilizar la plaza de toros de Algeciras en un evento religioso que incluye el sacrificio ritual de un cordero. Esta noticia generó mensajes con un tono agresivo, en los que se cuestiona la integración de las personas musulmanas y se plantean propuestas de expulsión mediante expresiones despectivas y generalizaciones negativas: *"son salvajes"*; *"primero los violan y después los sacrifican"*; *"a su país, fuera de España"*.

Del mismo modo, los días 1 y 2 de junio se registra un pico significativo en el volumen de comunicaciones, coincidiendo con los disturbios ocurridos en París tras la victoria del Paris Saint-Germain (PSG) en la final de la UEFA Champions League. En este contexto, se identificaron numerosos mensajes que asociaban a personas originarias del norte de África, así como a población africana o afrodescendiente, con comportamientos violentos y situaciones de inseguridad ciudadana. En muchos casos, los contenidos compartidos presentaban un marcado tono xenófobo, incitando incluso a la expulsión de estas personas con expresiones como: *"deberían deportarlos a todos"* o *"putos moros"*.

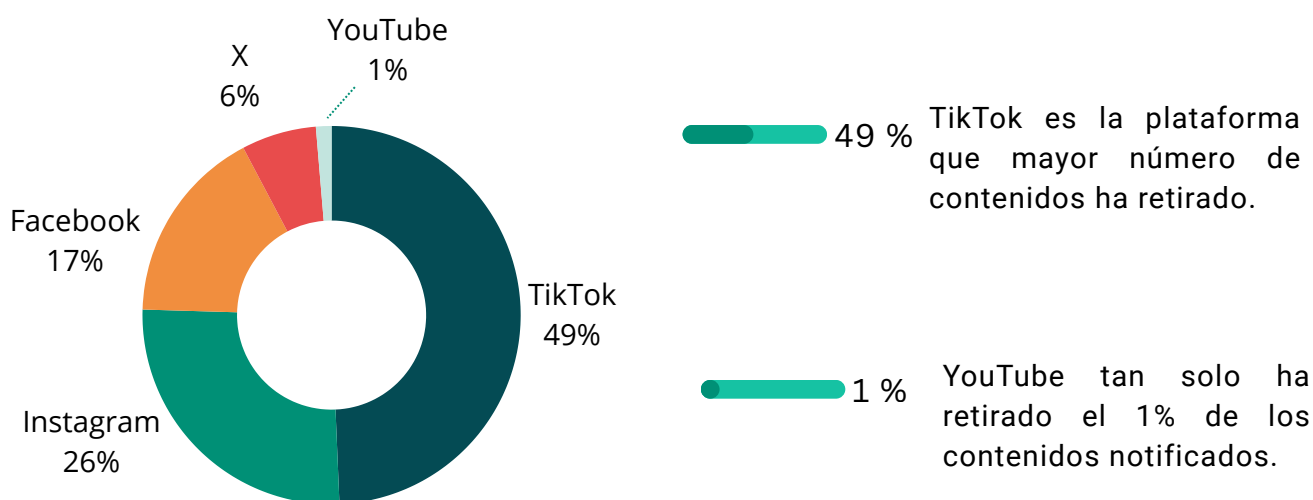
Asimismo, los días 19 y 20 de junio se observa un nuevo repunte, en este caso vinculado a la detención de un agente de policía en Torrejón de Ardoz (Madrid) tras el fallecimiento de un ciudadano del norte de África. Los mensajes analizados en redes sociales durante este episodio incluyen celebraciones del hecho y manifestaciones que justifican la violencia, recurriendo a estereotipos y prejuicios sobre la población migrante. En varios casos, se exalta la actuación policial mientras se deshumaniza a la víctima mediante expresiones ofensivas como *"un moro menos"*. Este tipo de discursos refuerza narrativas de odio y contribuye a la estigmatización de los diferentes grupos diana con contenidos como *"vienen a delinquir y a no adaptarse"*.

Por último, aunque los mensajes relacionados con el ámbito económico se mantienen estables y sin grandes fluctuaciones a lo largo del trimestre, cabe señalar que una de las principales narrativas generadoras de discursos de odio es la que alude a que las personas inmigrantes alojadas en dispositivos de acogida perciben ayudas sociales o prestaciones económicas.

Reacciones de las plataformas

En el segundo trimestre, las plataformas en su conjunto han retirado el 34% de las notificaciones realizadas por el OBERAXE. Se han retirado como usuario normal el 11% de los mensajes reportados y un 23% adicional a través de la vía de *trusted flagger*.

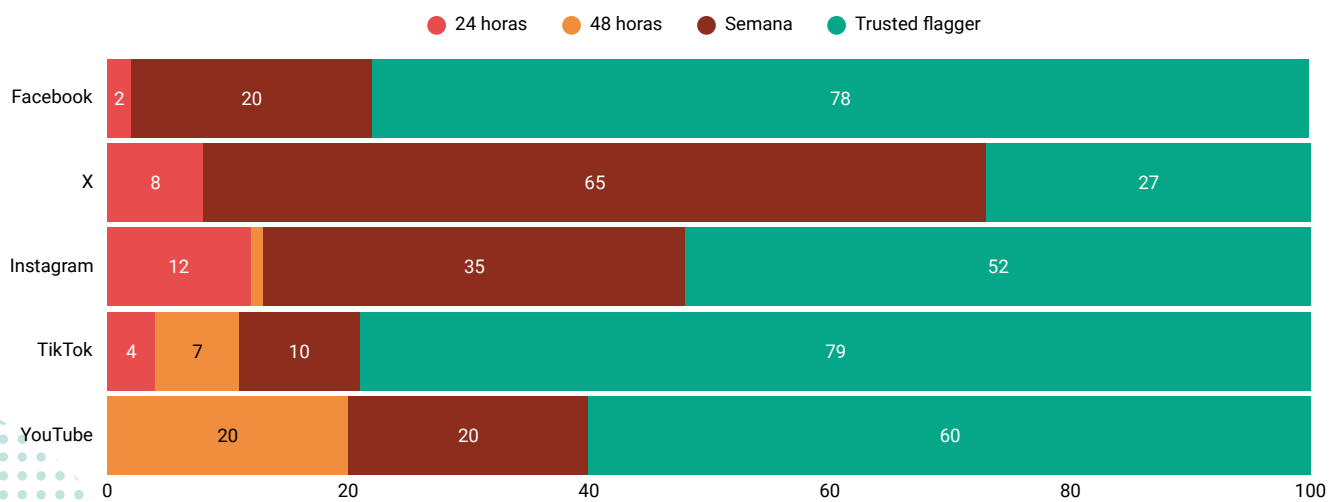
A continuación, se muestra el gráfico de contenidos retirados por cada plataforma en relación con las notificaciones efectuadas a cada una de ellas.



Evolución en la retirada de contenidos

Al analizar los tiempos de retirada de contenido en las distintas plataformas, Instagram destaca con un 12% de contenido retirado en las primeras 24 horas, mientras que YouTube no ha retirado ningún contenido en menos de 24 horas al ser reportado como usuario normal. X elimina la mayor parte de los reportes en el plazo de una semana, con un 65%.

Por otra parte, Facebook y TikTok presentan una mayor efectividad en la retirada de contenido cuando se utiliza la vía *trusted flagger*.



El discurso de odio en el fútbol

Durante el segundo trimestre de 2025, el discurso de odio vinculado al ámbito futbolístico se mantuvo como un espacio significativo para la expresión de narrativas xenófobas y racistas en redes sociales.

Más allá de los insultos directos a futbolistas, estos mensajes reflejan tensiones sociales más amplias, donde el deporte funciona como un escenario simbólico para reproducir estereotipos y prejuicios. La recurrencia de estos discursos durante eventos deportivos evidencia cómo episodios mediáticos se convierten en detonantes que amplifican sentimientos de exclusión y desconfianza hacia determinados grupos diana, enfocando estas expresiones principalmente en los deportistas.

Por otro lado, la iniciativa de LALIGA para medir y combatir los discursos de odio en redes sociales, a través del monitor MOOD que sirve para:

- Monitorizar el discurso de odio en sus diversas tipologías y observar su evolución a lo largo de la temporada de los partidos de LALIGA.
- Impactar generando contenido relevante para las aficiones y los clubes.
- Educar y concienciar, ayudando a las organizaciones deportivas a sensibilizarse y tomar medidas contra la violencia en el deporte.
- Promover valores impulsando la tolerancia, el respeto y la inclusión como pilares dentro de las competiciones.

Durante el segundo trimestre del año, la jornada con peor nota es la jornada 30, situándose en un 7.0. Por el lado positivo, destaca la nota de la jornada 38, con un 4,1. A continuación, se muestran los datos del monitor MOOD.

5,9

Mood de la jornada

4.801K

Menciones

1.961K

Usuarios

Mood de la jornada: es una puntuación del 1 al 10 que mide el grado de violencia discursiva en redes sociales alrededor de la competición bajo el amparo de La Liga Nacional de Fútbol Profesional.