1 OCTUBRE - 31 DICIEMBRE 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

En el periodo comprendido entre el 1 de octubre y el 31 de diciembre de 2025, el monitor FARO detectó 120.990 mensajes de odio reportable, y las plataformas retiraron el 51% de los contenidos reportados. La mayoría de los contenidos analizados se dirigieron hacia las personas del norte de África y las personas musulmanas. Durante este trimestre se ha observado una disminución en el número de contenidos de discurso de odio detectados. Esta tendencia podría estar vinculada al fortalecimiento y la optimización de los mecanismos de moderación y de detección temprana. No obstante, el volumen de mensajes sigue siendo significativo, con un impacto persistente en la polarización y en la erosión de la cohesión y la convivencia social.

120.990
Mensajes detectados

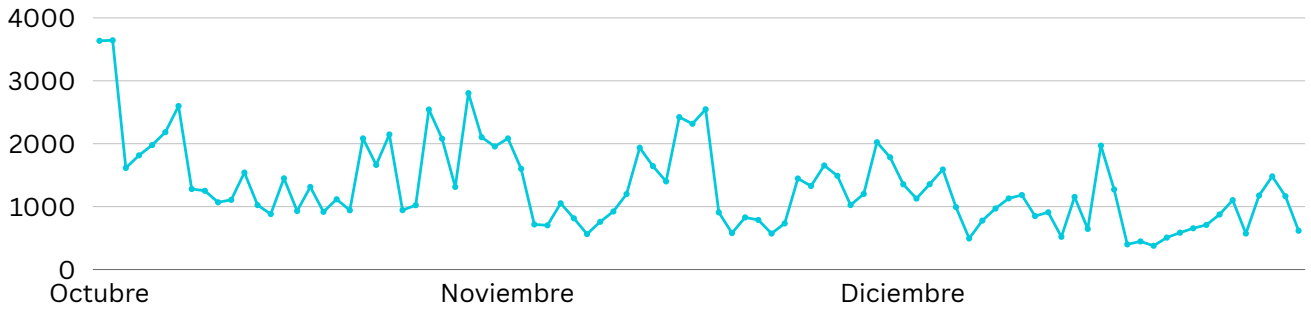
51%
% Mensajes retirados

Evolución de los contenidos detectados

Durante el cuarto trimestre de 2025, se registraron aproximadamente 1.300 mensajes de odio al día.

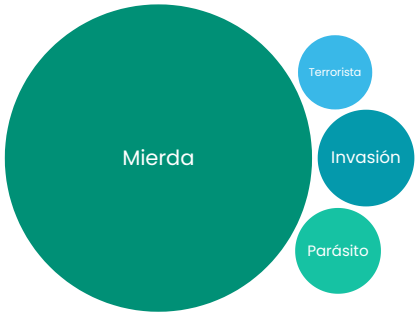
En octubre se observan picos destacados al inicio del mes, con valores superiores a los 3.600 mensajes diarios, seguidos de una tendencia general descendente, mientras que noviembre presenta una evolución más irregular, con un repunte a mediados de mes que supera los 2.000 mensajes diarios. Por último, diciembre presenta niveles inferiores de detección, sin superar en ningún caso los 2.000 mensajes diarios, consolidando la tendencia descendente observada.

Este comportamiento heterogéneo sugiere una posible relación con diversos eventos sociopolíticos, que serán detallados en mayor profundidad en el apartado relativo a los episodios prototípico.



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:

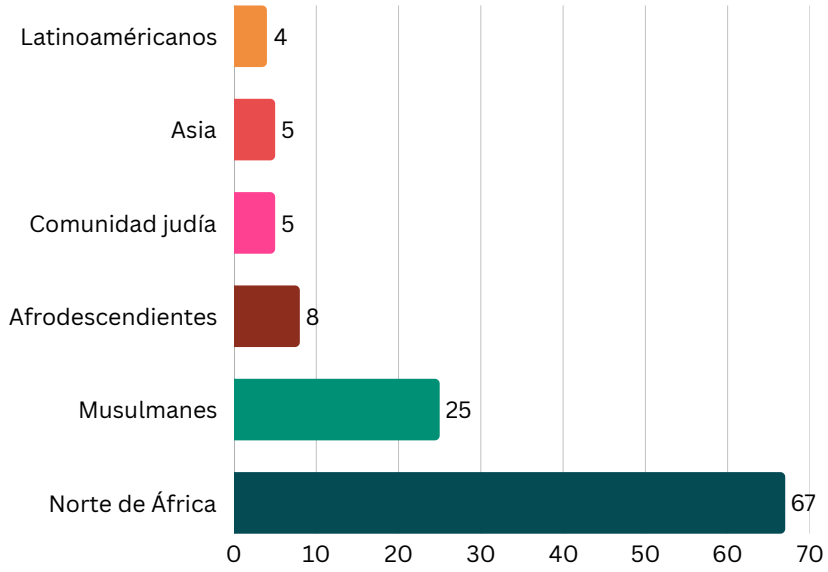


Consulta aquí los datos de monitorización en tiempo real

Características del discurso de odio

Grupo diana

En el cuarto trimestre de 2025, los contenidos de odio se dirigieron principalmente a personas del norte de África, cuya proporción descendió del 79% al 67%, doce puntos porcentuales menos que en el trimestre anterior. A continuación, se sitúan las personas musulmanas, que concentran el 25% de los casos, seguidas por las personas africanas y afrodescendientes, que representan el 8%.



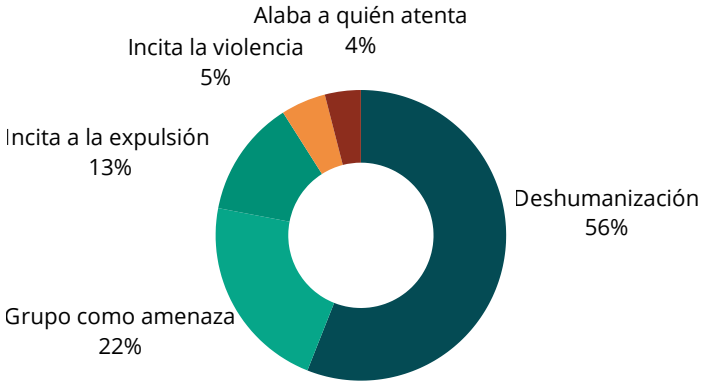
5 %

Los contenidos de odio dirigidos hacia la comunidad judía alcanzaron el 5%, un incremento significativo, posiblemente vinculado tanto con el atentado de Bondi Beach (Sídney) como al conflicto armado entre Israel y Palestina.

Tipo de contenido

Los contenidos de odio adoptan múltiples expresiones de estigmatización que intensifican las actitudes hostiles hacia las personas de origen extranjero, predominando los mensajes que deshumanizan o degradan (56%), que han experimentado un notable incremento de 20 puntos porcentuales con respecto al trimestre anterior.

Por su parte, una proporción significativa de los contenidos de odio incita a la expulsión de los colectivos diana (13%). Asimismo, un 22% de los contenidos presenta a estas personas como una amenaza para la seguridad y la convivencia o incita a la violencia (5%), facilitando la normalización de actitudes de rechazo social. Además, un 4% de los contenidos alaba a quien atenta contra los grupos diana. Este tipo de publicaciones evidencia que la estigmatización no solo se expresa mediante descalificaciones, sino también a través de la celebración explícita de la violencia.



Consulta aquí los datos de monitorización en tiempo real

Expresión del lenguaje

El lenguaje agresivo explícito está presente en el 93% de los contenidos de discurso de odio analizados, mediante insultos, amenazas y descalificaciones que evidencian la normalización de la hostilidad verbal en las redes sociales. Por su parte, el 7% de los mensajes recurre a la ironía o sarcasmo para transmitir contenido discriminatorio en clave humorística, lo que reduce la detectabilidad automática de este tipo de expresiones y dificulta la activación de los mecanismos de control de las plataformas.

Cabe destacar que aproximadamente el 16% de los contenidos reportados utiliza un lenguaje codificado, en el que se combinan letras, números, símbolos o emojis para evitar los mecanismos de censura de las plataformas, con mensajes como: o *"k se lleve a su casa al mierda 🥦"* o *"Islam es♋ancer"*.

Se identifica un empleo reiterado de emojis con carga simbólica que actúan como refuerzo visual de mensajes que promueven la deshumanización, rechazo y violencia contra las personas migrantes. Por ejemplo, los emojis de animales se utilizan para deshumanizar *"los 🐽🐽 chinos me espían el teléfono"*, mientras que las armas y herramientas expresan agresión y violencia contra los grupos diana, como en el caso de "🔫🪓🥓"

El empleo de estos recursos evidencia cómo el discurso de odio se articula mediante códigos visuales y formas simbólicas que facilitan la circulación de mensajes agresivos y contribuyen a normalizar actitudes de rechazo y hostilidad.

Contenido viral


moros tasa de delincuencia x32 respecto a los españoles. Que nadie olvide los datos empiecen a difuminarse por los nacidos en ESP.

inviable no ser racista. cuestión de supervivencia.

· **99,5 mil** Visualizaciones

57 832 3 mil 308

Esta publicación de la red social X ha obtenido 100 mil visualizaciones. Se trata de un tweet que surge a raíz de la divulgación de datos sobre el origen de las personas detenidas en el País Vasco.

Este contenido viral instrumentaliza esta información para realizar generalizaciones que criminalizan a las personas migrantes, atribuyendo conductas delictivas exclusivamente a su pertenencia grupal.

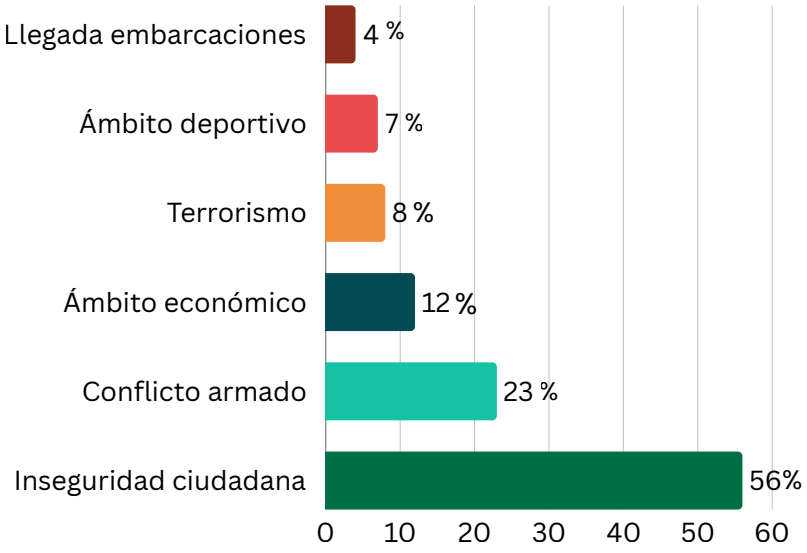
Su amplia circulación contribuye a reforzar estereotipos negativos, alimenta la desconfianza hacia colectivos concretos y favorece un clima social hostil que dificulta su integración y participación plena en la comunidad.

Episodios que suscitan discurso de odio

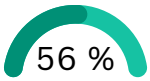
Durante el cuarto trimestre de 2025, el discurso de odio en las principales plataformas ha estado relacionado con los episodios prototípicos de inseguridad ciudadana, conflicto armado y ámbito económico.

La inseguridad ciudadana representó el 56% de los contenidos de odio, manteniéndose como uno de los principales detonantes de discurso de odio, aunque presenta una ligera disminución con respecto al trimestre anterior (61%). Uno de los principales sucesos ocurrido en este trimestre, que ha contribuido al aumento de discurso de odio, fue la noticia de que la policía del País Vasco comenzaría a hacer públicos los datos relativos a la procedencia de las personas detenidas a partir del 13 de noviembre. Este hecho fue aprovechado por numerosos usuarios para volcar su hostilidad contra determinados grupos diana, principalmente, las personas del norte de África, promoviendo la generalización y estigmatización de los mismos y perpetuando estereotipos que vinculan la inmigración con la delincuencia.

Asimismo, a lo largo de este último trimestre de 2025, también se ha observado la proliferación de publicaciones vinculadas a altercados en el espacio público, que tienden a concentrar un volumen significativo de mensajes hostiles y se convierten en vehículos para reforzar narrativas discriminatorias y xenófobas. Su difusión contribuye a legitimar discursos que presentan a los grupos diana como una amenaza y que promueven su exclusión social o expulsión del territorio, como en “Siempre los pelobrocoli, deportación inmediata”.



Los discursos de odio relacionados con terrorismo (8%) aumentan 3 puntos porcentuales respecto al trimestre anterior



La inseguridad ciudadana es el principal detonante de discurso de odio.

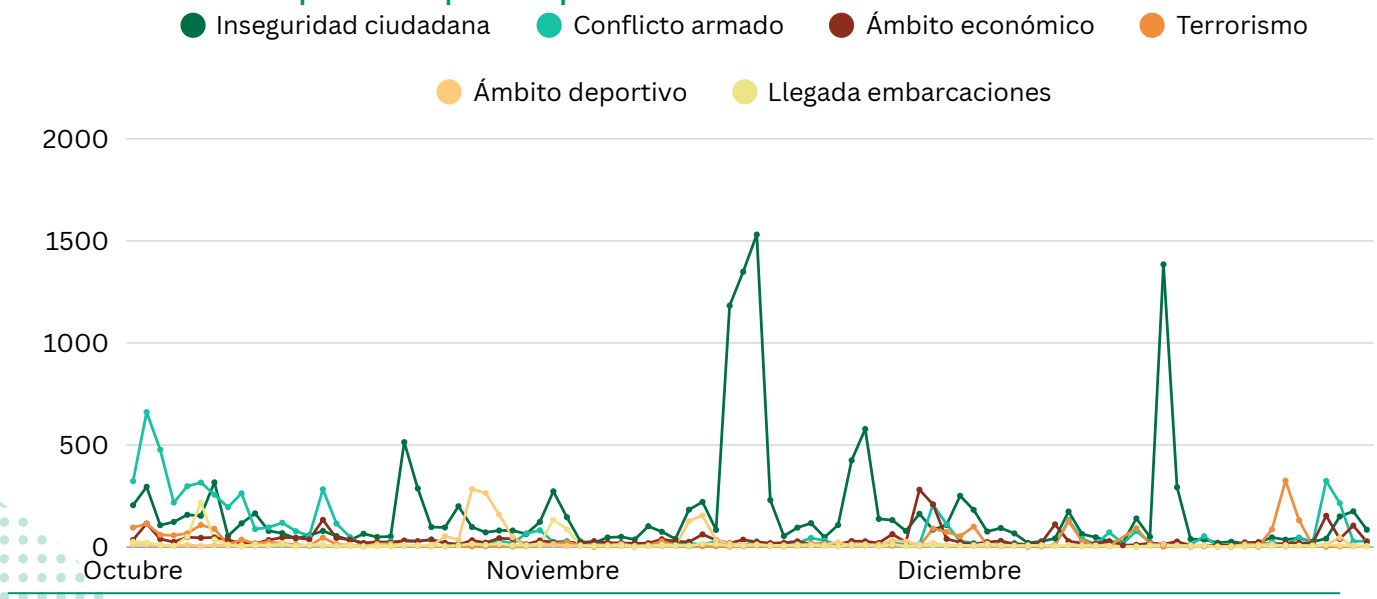
Los contenidos con discurso de odio relacionados con el conflicto armado representaron el 23% del total analizado durante el último trimestre de 2025. Además de la cobertura mediática sobre la guerra entre Israel y Palestina, las noticias relativas a la interceptación de la "flotilla" humanitaria con destino a Gaza generaron un notable aumento de mensajes con contenido islamófobo y antisemita. A ello se sumaron diferentes actos de protesta y la intensificación del debate público en los días previos y posteriores al 7 de octubre, fecha en la que se cumplió dos años de los ataques de Hamás, lo que reactivó el volumen de comentarios hostiles. Entre los contenidos detectados predominan mensajes que criminalizan a la población y justifican la violencia, con expresiones como "*todos fuera de este mundo*" o "*aniquilarlos por completo*", así como comentarios que presentan la identidad religiosa como una amenaza: "*los moros y los judíos, amenaza para el mundo*".

A partir de mediados del mes de octubre, el número de mensajes se estabilizó y, durante prácticamente todo noviembre, se mantuvo en niveles más bajos aunque constantes, con una media aproximada de 27 contenidos diarios. No obstante, a finales de noviembre se registraron nuevos picos, asociados principalmente a la difusión de noticias sobre protestas en apoyo a Palestina que tuvieron lugar en mercados navideños de distintas ciudades europeas.

En cuanto al ámbito económico, que supone el 12% del total de contenidos de discurso de odio, uno de los episodios que concentró un volumen más elevado de mensajes fue el desalojo de 400 personas inmigrantes que residían de forma irregular en el edificio B9 en Badalona, ocurrido el 17 de diciembre. Este hecho provocó la difusión de narrativas que presentan al grupo diana como una amenaza económica y social, contribuyendo a un clima de hostilidad a nivel local, en detrimento de la integración e incrementando la vulnerabilidad de las personas atacadas, con mensajes como "*los moros okupas al fresco como toca*".

Por último, el 8% de los mensajes analizados se refirieron a contenidos relacionados con sucesos terroristas. A raíz del atentado perpetrado durante la celebración de la festividad judía de Janucá en Sídney, el 14 de diciembre, se registraron picos de discurso de odio en los días posteriores.

Evolución de los episodios prototípicos



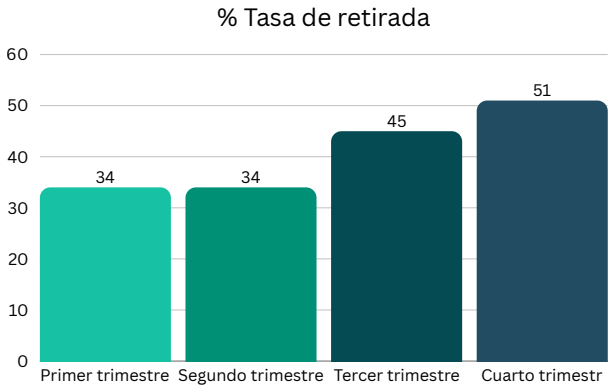
Consulta aquí los datos de monitorización en tiempo real

Reacciones de las plataformas

Durante el cuarto trimestre, las plataformas han retirado el 51% del contenido notificado por OBERAXE, lo que implica un incremento de 6 puntos porcentuales respecto al trimestre pasado (45%).

Este aumento refleja el compromiso sostenido de las plataformas con la eliminación del discurso de odio, consolidado tras la mejora continua de los mecanismos de moderación y el refuerzo de los canales de colaboración institucional. Esta cooperación contribuye a mejorar la coordinación con las plataformas y a agilizar la respuesta frente a la difusión de discurso de odio.

La media acumulada ponderada de la tasa de retirada de contenidos notificados por el OBERAXE hasta el 31 de diciembre alcanza el 41%.



En cuanto al tiempo de respuesta, se mantiene la misma tendencia que en el trimestre anterior, ya que el 8% de los contenidos fueron retirados antes de 24 horas, y un 1% en las siguientes 48 horas. En el plazo de una semana se retiró el 2%, mientras que los contenidos retirados mediante la vía *trusted flagger* alcanzaron el 40%.

En conjunto, la vía de comunicante fiable (*trusted flagger*) resulta decisiva y pone de manifiesto la mayor eficacia de este mecanismo frente a la retirada por usuario normal.

A continuación, se presenta el gráfico que recoge la tasa de retirada de cada plataforma en relación con las comunicaciones de contenidos de discurso de odio que han recibido:

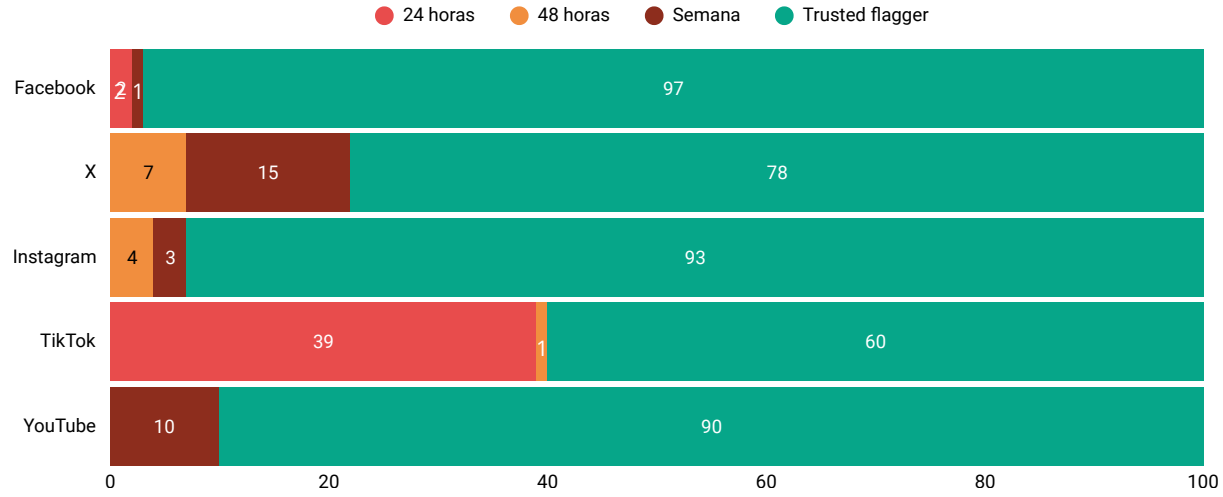
TikTok retira el 90% de los contenidos notificados, situándose como la plataforma más eficaz en este aspecto.

Consulta aquí los datos de monitorización en tiempo real

Evolución en la retirada de contenidos

En cuanto a los tiempos de reacción, TikTok destaca con un 39% de retirada en las primeras 24 horas. Por su parte, X es la red social que más mensajes retiró en el transcurso de una semana, con un 15%, seguido de Youtube con un 10%.

Por otra parte, tanto Facebook, Instagram y Youtube tienen mayor efectividad al utilizar la vía de comunicación fiable (*trusted flagger*), con tasas de retiradas superiores al 90% en los tres casos; mientras que en el caso de X este porcentaje es del 78% y en el caso de TikTok del 60%.



Medidas proactivas de las plataformas

A continuación se presentan los datos facilitados por las plataformas sobre la moderación de contenidos correspondientes al tercer trimestre de 2025 (1 de julio–30 de septiembre), relativos a su actividad a nivel mundial. Esta información se enmarca en el trabajo desarrollado con las plataformas en el grupo de trabajo sobre discurso de odio en redes sociales, constituido en julio de 2025, y que busca reforzar de manera conjunta las medidas de moderación destinadas a combatir la difusión de contenidos de odio. Estos datos reflejan las medidas proactivas de moderación, entendidas como aquellas iniciadas y aplicadas directamente por las propias plataformas mediante sistemas automatizados y revisión interna sin mediar denuncia previa de los usuarios.

66 % En Facebook, el 66% de las medidas aplicadas por discurso de odio fueron proactivas.

87 % En Instagram, el porcentaje de actuaciones proactivas por contenido de odio general alcanzó el 87%.

54 % En el caso de Tiktok, en España, el 54% de las retiradas por discurso de odio se produjo antes de recibir visualizaciones.

98 % En lo que respecta a la moderación de contenidos en general, en Youtube, el 98% se eliminó por detección automática.

Por último, se incluye información adicional del periodo del 1 octubre al 31 diciembre de 2025, durante el cual, según los datos de TikTok se eliminaron, en total, **440.121** comentarios y **34.477** vídeos por discurso de odio.

El discurso de odio en el fútbol

Durante el cuarto trimestre de 2025, el discurso de odio vinculado al ámbito deportivo supuso un 7% con respecto al total del contenido monitorizado. En este sentido, el ámbito futbolístico se mantuvo como un espacio significativo para la expresión de narrativas xenófobas y racistas en redes sociales, siendo el colectivo diana de las personas del norte de África el que concentra la mayoría de contenido de discurso de odio.

Tal y como muestran las últimas tendencias, los futbolistas más atacados y que generan más contenido de discurso de odio en redes sociales son Lamine Yamal y Vinícius Júnior, con comentarios como "*moro desnutrido*" o "*el negro de mierda quiere jugar al fútbol*".

Los datos mensuales del trimestre muestran además cómo determinados episodios deportivos actúan como catalizadores de picos de actividad discriminatoria. En octubre, el clásico entre el Real Madrid y el FC Barcelona se asoció a un notable aumento de mensajes de odio, coincidiendo con el desarrollo del encuentro y su amplia repercusión mediática. En ese contexto, se detectó un incremento de insultos de carácter racista dirigidos contra Lamine Yamal y Vinícius Júnior, con contenidos que reproducían estereotipos y discursos xenófobos. En noviembre, Lamine Yamal volvió a ser objeto de ataques tras su ausencia en la convocatoria de la Selección Española, un episodio que desencadenó comparaciones deshumanizantes y comentarios que instrumentalizan su ascendencia norteafricana para justificar discursos de exclusión. Por último, la concesión del Balón de Oro a Lamine Yamal en diciembre actuó como un detonante que intensificó también la circulación de mensajes de odio contra él, aunque en menor medida que en los anteriores episodios. Estos patrones refuerzan la tendencia sostenida a utilizar el origen del jugador como eje de hostilidad en el debate futbolístico.

Más allá de los ataques directos a futbolistas concretos, este tipo de mensajes refleja tensiones sociales más amplias, en las que el deporte opera como un escenario simbólico para la reproducción de estereotipos y prejuicios. La reiteración de estas expresiones durante eventos de alta visibilidad mediática muestra cómo determinados episodios se convierten en detonantes que amplifican sentimientos de exclusión y desconfianza hacia determinados grupos diana, concentrando estas dinámicas especialmente en los deportistas.

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