1 ENERO - 31 MARZO 2026

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 enero y el 31 de marzo de 2026, el monitor FARO detectó un total de 105.911 mensajes de odio reportables, y las plataformas retiraron el 55% de los contenidos reportados. La mayoría de los contenidos analizados se dirigieron contra personas del norte de África y personas musulmanas. A lo largo de este trimestre se ha observado una reducción en el volumen de contenidos de discurso de odio detectados, una evolución que podría estar asociada a la mejora progresiva de los sistemas de moderación y detección temprana. No obstante, el número de mensajes continúa siendo relevante, manteniendo un impacto persistente en la polarización social y en el deterioro de la convivencia.

105.911
Mensajes detectados

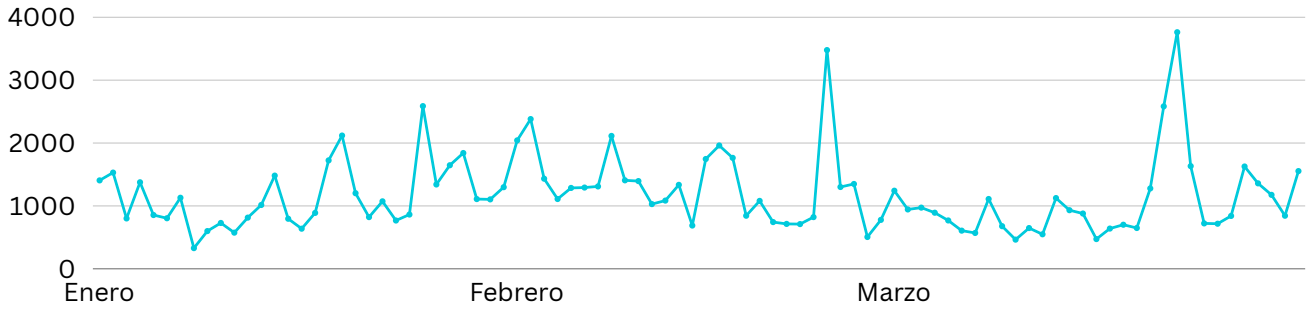
55%
% Mensajes retirados

Evolución de los contenidos detectados

Durante el primer trimestre de 2026, se registró un promedio diario de aproximadamente 1.170 mensajes de odio.

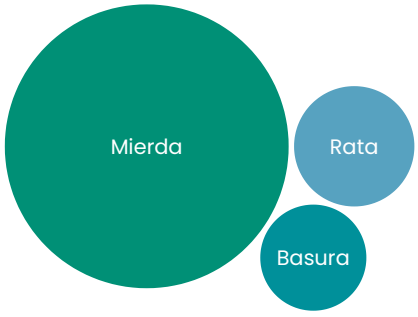
En enero, los niveles de detección se mantuvieron moderados, aunque se registraron algunos picos que superaron los 2.000 mensajes. En febrero se observa un aumento del volumen de mensajes, alcanzando un máximo de más de 3.470 mensajes. Finalmente, marzo presenta un descenso del nivel medio de mensajes, si bien se identifica un pico relevante que supera los 3.700 mensajes.

Este comportamiento variable sugiere una posible relación con eventos sociopolíticos, que se analizarán en el apartado relativo 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:

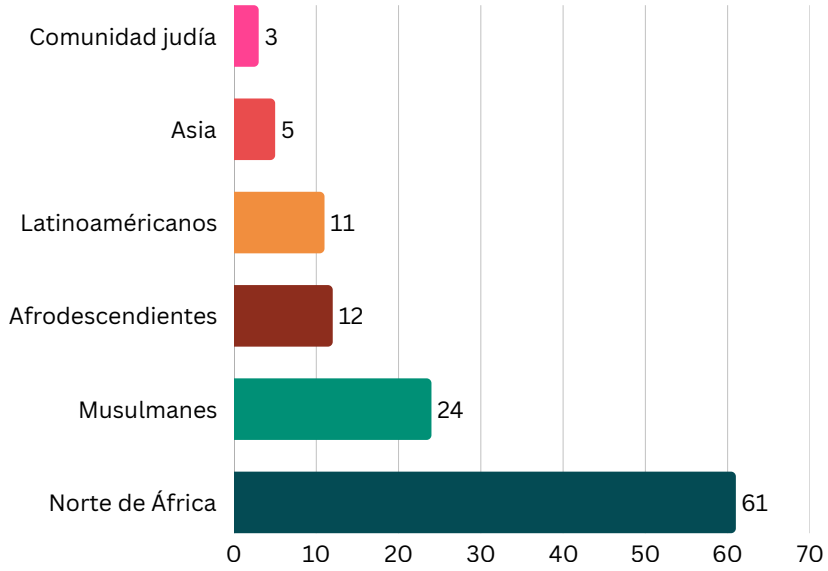


Consulta aquí los datos de monitorización en tiempo real

Características del discurso de odio

Grupo diana

En el primer trimestre de 2026, los contenidos de odio se dirigieron principalmente a personas del norte de África, que concentran el 61% del total, 6 puntos porcentuales menos que el trimestre anterior. A continuación, se sitúan las personas musulmanas (24%) y las personas africanas y afrodescendientes (12%).



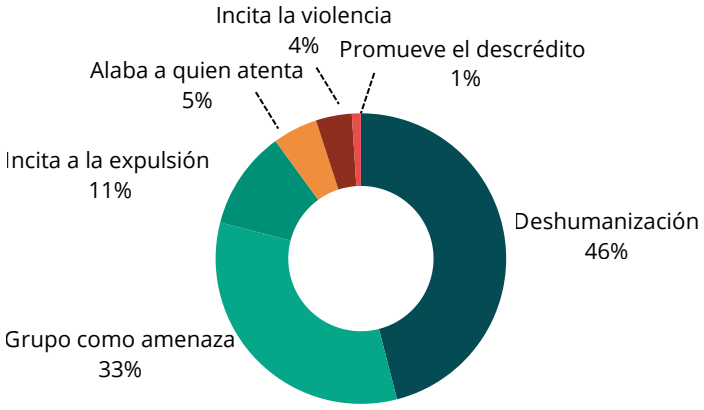
11 %

Los contenidos de odio dirigidos hacia las personas latinoamericanas se han situado este trimestre en un 11%, un incremento de 7 puntos respecto al trimestre anterior, lo que sugiere un aumento de la hostilidad dirigida a este grupo diana.

Tipo de contenido

Los contenidos de odio adoptan diversas formas de estigmatización hacia las personas de origen extranjero, con un predominio de los mensajes que deshumanizan o degradan al grupo diana, que representan el 46% del total y registran un descenso de 10 puntos respecto al trimestre anterior (56%).

Por su parte, los contenidos de odio que presentan a los colectivos diana como una amenaza para la seguridad y la convivencia alcanzaron un 33%. Asimismo, un 11% de los contenidos incita a la expulsión de estas personas, contribuyendo a la reproducción de narrativas de exclusión social. Por otro lado, un 5% de los contenidos alaba a quienes atentan contra los grupos diana, mientras que un 4% incita a la violencia contra estas personas, legitimando la agresión contra ellas. Por último, un 1% de los contenidos promueve el descrédito de los grupos diana.



Expresión del lenguaje

El lenguaje agresivo explícito está presente en el 94% de los contenidos de discurso de odio detectados, fundamentalmente a través del uso de insultos, amenazas y descalificaciones, lo que evidencia una creciente normalización de la violencia verbal en las redes sociales. Asimismo, el 6% de los mansajes recurre a la ironía o sarcasmo, enmascarando el discurso de odio bajo una apariencia humorística, lo que dificulta su detección automática y permite eludir en mayor medida los mecanismos de moderación de las plataformas.

Cabe destacar que aproximadamente el 30% 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: "*Napalm* 🔥" o "*inmediata* ✈️".

Se constata un uso recurrente de emojis con carga simbólica que funcionan como elementos de refuerzo visual en mensajes orientados a promover la deshumanización, el rechazo y la incitación a la violencia. Por ejemplo, los emojis de animales se utilizan para deshumanizar a los grupos diana: "*los* 🐒 *adiestrados bien";* mientras que las armas expresan agresión y violencia: "*invasión* 👴👴👴🏿🔫".

El uso de estos recursos pone de manifiesto la articulación del discurso de odio a través de códigos visuales y contribuyen a la normalización de actitudes de rechazo y hostilidad.

Contenido viral

Esta publicación de la red social X ha obtenido más de 6.000 visualizaciones. Se trata de un tweet que surge en el contexto de las políticas migratorias y la gestión de los centros de acogida de menores extranjeros no acompañados en España.

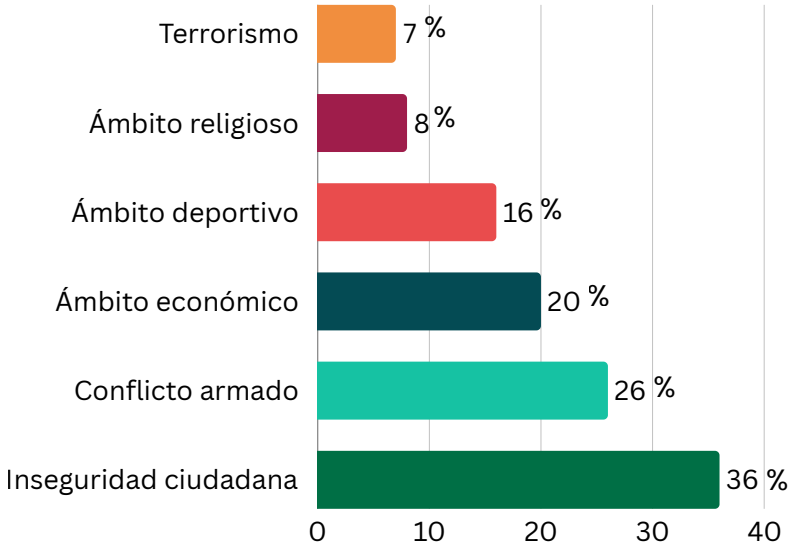
Este contenido viral pone de manifiesto cómo los menores no acompañados, especialmente aquellos en contextos de mayor vulnerabilidad, son presentados como beneficiarios injustos de recursos públicos, así como objeto de estigmatización y discriminación persistentes.

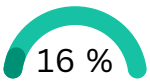
Episodios que suscitan discurso de odio

A lo largo del primer trimestre de 2026, el discurso de odio en las principales plataformas se ha vinculado principalmente a episodios prototípicos de inseguridad ciudadana, conflicto armado, ámbito económico y ámbito deportivo.

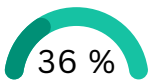
La inseguridad ciudadana representó el 36% de los contenidos de discurso de odio, siendo las personas del norte de África el principal grupo diana afectado, seguidas de las personas musulmanas y de las personas africanas o afrodescendientes. Entre los distintos episodios registrados durante el trimestre, destaca la difusión de un vídeo viral grabado en el aeropuerto de Valencia, en el que se observa a un hombre encaramado al techo de un avión. Este contenido fue instrumentalizado para atribuirle un origen extranjero, pese a no existir constancia de una relación con la inmigración. A partir de este suceso se generaron mensajes que incitaban explícitamente a la violencia con expresiones como "*una cacería ya*" o "*la cabeza como trofeo*".

Asimismo, se observa la circulación de publicaciones y materiales descontextualizados vinculados a altercados en el espacio público, que concentran un volumen notable de mensajes hostiles y contribuyen a reforzar narrativas xenófobas y deshumanizadoras. Este tipo de contenidos legitima discursos que vinculan inmigración e inseguridad ciudadana, promoviendo su exclusión social o expulsión del territorio e, incluso, la violencia contra ellos, con mensajes como *"todos mierdas, deportación"* o *"piedra al cuello y al mar"*.





Los discursos de odio relacionados con el ámbito deportivo (16%) aumentan 9 puntos porcentuales respecto al trimestre anterior



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

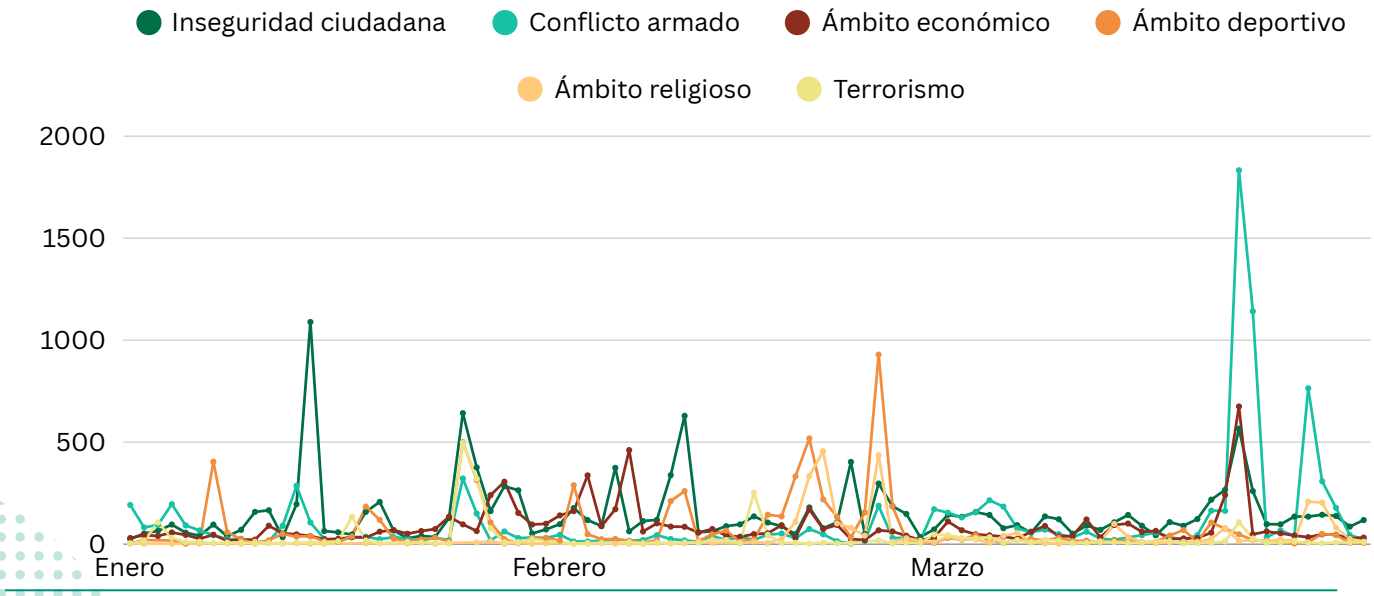
Consulta aquí los datos de monitorización en tiempo real

Los contenidos de discurso de odio asociados al conflicto armado representaron el 26% del total analizado durante el primer trimestre de 2026. La operación militar llevada a cabo el 3 de enero por Estados Unidos en Venezuela generó un volumen relevante de comentarios dirigidos contra la población venezolana, incluyendo mensajes que promovían su expulsión del territorio, con expresiones como "*que alegría por los venezolanos invasores, que vuelvan a su país* 🛫 ". Posteriormente, en el mes de marzo, la intensificación de los conflictos armados en Gaza e Irán contribuyó a un aumento de la polarización en redes sociales, con la aparición de contenidos antisemitas que legitiman la violencia contra las personas judías, con comentarios como "*más misiles y que se jodan los judíos*". Asimismo, se identificaron narrativas islamófobas que criminalizan y deshumanizan a la población musulmana, con comentarios como "*los seguidores del islam son todos unos asesinos*".

En cuanto al ámbito económico, que representó el 20% del total de contenidos de discurso de odio, uno de los episodios que concentró un volumen más elevado de mensajes fue la tramitación del real decreto para la regularización extraordinaria de personas extranjeras que ya residen España, con mensajes que presentan a las personas inmigrantes como una amenaza económica, social y de seguridad, con comentarios como "*van a venir a violar a nuestras hijas*" o "*ratas parasitarias*". Por su parte, la propuesta de prohibición del burka y el niqab en el espacio público intensificó la difusión de contenidos islamófobos, especialmente contra las mujeres, con mensajes como "*deportación para acabar con el burka*" o "*invasión islámica*". Finalmente, la aprobación del real decreto que refuerza la universalidad del acceso a la sanidad pública reactivó discursos que cuestionan el acceso de las personas inmigrantes a los servicios sanitarios, con comentarios como: "*y mientras españoles en lista de espera, necesitamos purga*" o "*sabandijas africanas*".

Por último, el 16% de los mensajes analizados se refirieron al ámbito deportivo. Varios encuentros deportivos han servido para catalizar mensajes de discurso de odio contra jugadores de origen extranjero. Asimismo, en torno a publicaciones de Lamine Yamal relacionadas con su práctica del Ramadán se detectaron mensajes que cuestionaban su pertenencia, con comentario como "*a su puto país* 😤", "*y en la selección española, que asco*" o "*hijo puta, no me representa*".

Evolución de los episodios prototípicos



Consulta aquí los datos de monitorización en tiempo real

Reacciones de las plataformas

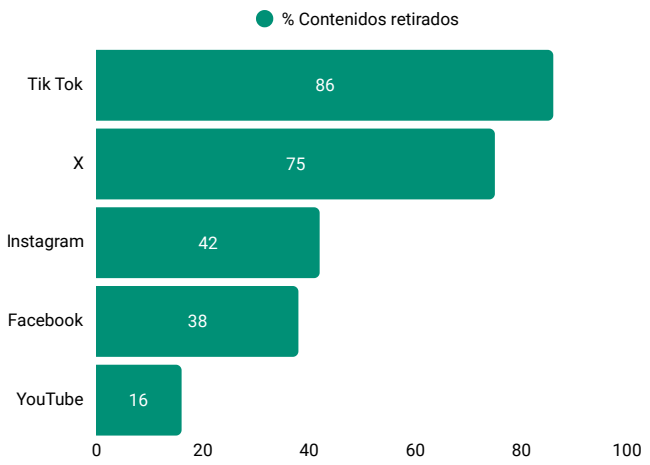
En el primer trimestre, las plataformas retiraron el 55% del contenido notificado por OBERAXE, lo que supone un aumento de 4 puntos porcentuales en comparación con el trimestre anterior (51%).

Este aumento pone de relieve el compromiso sostenido de las plataformas con la eliminación del discurso de odio, resultado del fortalecimiento progresivo de los mecanismos de moderación y de la colaboración institucional. En esta línea, en enero de 2026 se celebró la segunda reunión de seguimiento del grupo de trabajo con las principales plataformas de redes sociales, con el objetivo de seguir afianzando la colaboración existente. Esta colaboración contribuye a mejorar la coordinación con las plataformas y a agilizar la respuesta frente a la difusión de discurso de odio.

En cuanto a los tiempos de respuesta, se mantiene una distribución similar a la del trimestre anterior, ya que el 10% de los contenidos fueron retirados antes de 24 horas, y un 1% en las siguientes 48 horas. En el plazo de una semana se eliminó el 3%, mientras que los contenidos retirados a través de la vía *trusted flagger* representaron el 41% del total.

En conjunto, la vía de comunicante fiable (*trusted flagger*) destaca por su mayor eficacia frente a la retirada de contenidos a través de denuncias de 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:



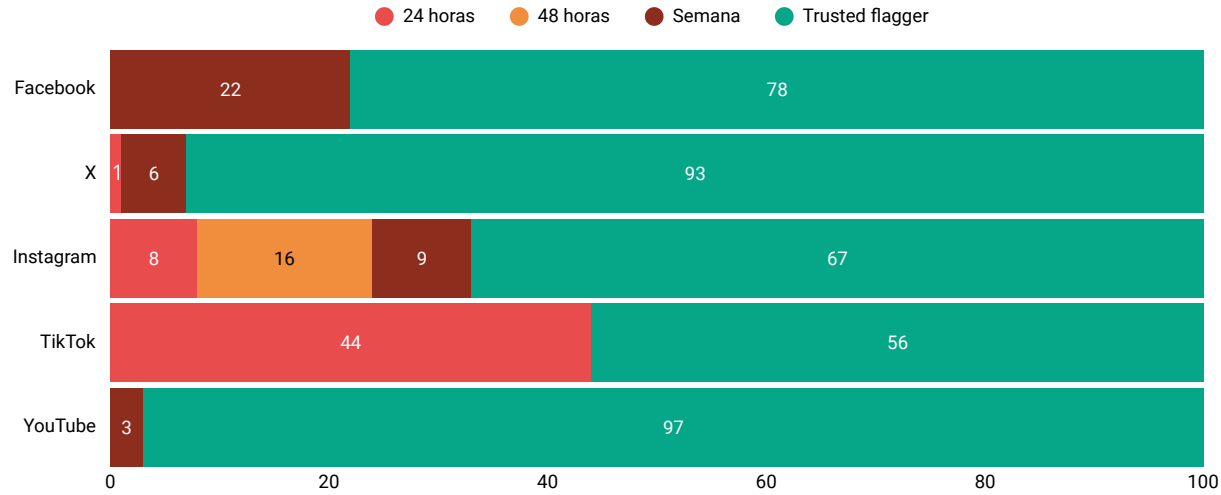
X ha alcanzado una tasa de retirada del 75%, más del doble de la registrada en el trimestre anterior (36%)

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 44% de retirada en las primeras 24 horas. Por su parte, Facebook es la plataforma que más contenido retiró en el transcurso de una semana con un 22%, seguida de Instagram con un 9%.

Por otro lado, X y YouTube tienen mayor efectividad cuando utilizan la vía *trusted flagger,* con tasas de retirada superiores al 90% en ambos casos; mientras que en el caso de Facebook este porcentaje es del 78%, Instagram 67% y TikTok 56%.



Medidas proactivas de las plataformas

A continuación se presentan los datos facilitados por las plataformas en relación con las medidas adoptadas para mejorar la moderación de contenidos durante el cuarto trimestre de 2025 (1 octubre - 31 diciembre), correspondientes a su actividad a escala mundial. La difusión de esta información se enmarca en el trabajo desarrollado por el grupo de trabajo sobre discurso de odio en redes sociales, constituido en julio de 2025 en colaboración con las propias plataformas, con el objetivo de reforzar de manera coordinada las actuaciones destinadas a prevenir y combatir la propagación de contenidos de odio en el entorno digital.

74 % En Facebook, el 74% de las medidas aplicadas por discurso de odio fueron proactivas.

84 % En Instagram, el porcentaje de actuaciones proactivas por contenido de odio general alcanzó el 84%.

98 % En el caso de Tiktok, durante el primer trimestre de 2026 (1 enero - 31 marzo), el porcentaje de retirada proactiva de discurso de odio alcanzó un 98%.

42 % Por su parte, YouTube eliminó el 42% del contenido que infringía sus normas antes de recibir visualizaciones, a través de su sistema de denuncia automática.

Consulta aquí los datos de monitorización en tiempo real

El discurso de odio en el fútbol

Durante el primer trimestre de 2026, el discurso de odio vinculado al ámbito deportivo supuso un 16% del total del contenido analizado. En este contexto, el ámbito futbolístico se mantuvo como un espacio significativo para la expresión de narrativas xenófobas y racistas en redes sociales, siendo las personas del norte de África el grupo diana que concentra la mayor parte de estos contenidos de discurso de odio.

En el periodo analizado, Lamine Yamal y Vinícius Júnior concentran el mayor volumen de mensajes con discurso de odio en redes sociales, convirtiéndose en los futbolistas más atacados. Entre los contenidos detectados se incluyen comentarios como "*puto moro, fuera de la selección*" o "*negro de mierda*".

Los datos del trimestre muestran además cómo determinados episodios deportivos actúan como catalizadores de picos de actividad discriminatoria. En enero, la celebración de la Copa Africana de Naciones y su retransmisión, así como las celebraciones de distintas aficiones en ciudades españolas, generaron un aumento de mensajes hostiles vinculados al origen racial o étnico de jugadores y seguidores. Por otro lado, los enfrentamientos entre aficiones durante el partido de Liga entre el FC Barcelona y el RCD Espanyol, disputado el 3 de enero, derivaron en la estigmatización de aficionados a los que se atribuía un origen extranjero, con comentarios como "*tiro al moro*". Por último, el partido de Champions League entre el Benfica y el Real Madrid, celebrado el 17 de febrero, en el que se activó el protocolo contra el racismo por insultos dirigidos a Vinícius Júnior, actuó como detonante de una intensa circulación de mensajes de odio contra el jugador.

Más allá de los ataques dirigidos a figuras individuales, estos mensajes ponen de relieve tensiones sociales más amplias, en las que el fútbol aparece como un espacio de alta visibilidad para la activación de narrativas racistas y xenófobas. La reiteración de este tipo de expresiones en contextos de especial exposición mediática muestra cómo determinados eventos deportivos amplifican discursos de exclusión y hostilidad hacia grupos concretos.

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