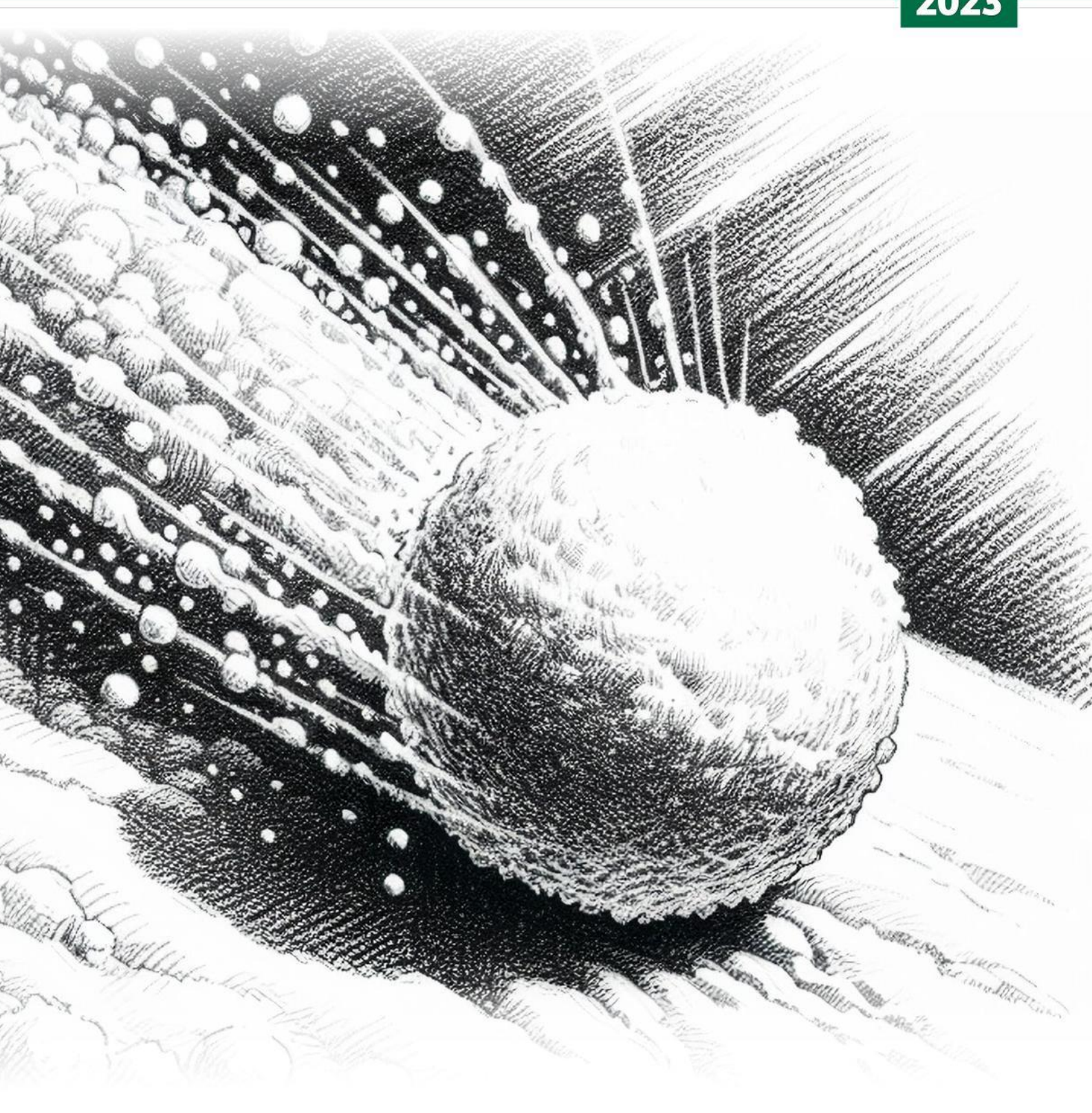


INFORME ANUAL

# MONITORIZACIÓN DEL DISCURSO DE ODIO EN REDES SOCIALES

2023





Catálogo de publicaciones de la Administración General del Estado

<https://cpage.mpr.gob.es>

© Ministerio de Inclusión, Seguridad Social y Migraciones.

Madrid, 2024

Autor/es: Observatorio Español del Racismo y la Xenofobia (OBERAXE)

Edita: Observatorio Español del Racismo y la Xenofobia

Distribuye: Observatorio Español del Racismo y la Xenofobia

Calle María de Guzmán, 52, tercera planta. 28003 Madrid

Correo electrónico: [oberaxe@missm.inclusion.gob.es](mailto:oberaxe@missm.inclusion.gob.es)

Web: <https://www.inclusion.gob.es/oberaxe/es/index.htm>

NIPO: 121-24-009-4

Diseño y maquetación: DISEÑO GRÁFICO GALLEGO Y ASOCIADOS, S. L.

## ÍNDICE

|                                                                                                        |           |
|--------------------------------------------------------------------------------------------------------|-----------|
| <b>ÍNDICE DE TABLAS .....</b>                                                                          | <b>4</b>  |
| <b>ÍNDICE DE GRÁFICOS.....</b>                                                                         | <b>4</b>  |
| <b>PRESENTACIÓN INSTITUCIONAL .....</b>                                                                | <b>5</b>  |
| <b>1. INTRODUCCIÓN.....</b>                                                                            | <b>6</b>  |
| <b>2. OBJETO DE LA MONITORIZACIÓN DEL DISCURSO DE ODIO.....</b>                                        | <b>10</b> |
| 2.1. Origen, contexto y escenario actual de la monitorización.....                                     | 10        |
| 2.2. Objetivos de la monitorización .....                                                              | 11        |
| 2.3. ¿Qué es el discurso de odio?: definición y características principales .....                      | 12        |
| 2.4. Dimensiones del discurso de odio.....                                                             | 13        |
| 2.4.1. El discurso de odio ilegal y el discurso odioso amparado en la libertad de expresión .....      | 13        |
| 2.4.2. El discurso de odio ilegal: los delitos de odio y el discurso de odio ilegal no delictivo ..... | 14        |
| 2.5. El discurso de odio en el régimen jurídico español.....                                           | 15        |
| 2.5.1. Ámbito penal: los delitos de odio .....                                                         | 15        |
| 2.5.2. Ámbito administrativo .....                                                                     | 16        |
| 2.6. ¿Qué monitoriza el OBERAXE? .....                                                                 | 17        |
| <b>3. METODOLOGÍA .....</b>                                                                            | <b>20</b> |
| 3.1. Metodología de recogida de datos para la monitorización .....                                     | 20        |
| 3.1.1. Metodología de búsqueda de contenidos .....                                                     | 20        |
| 3.1.2. Metodología de reporte y denuncia de contenidos.....                                            | 21        |
| 3.1.3. Metodología de anotación de los contenidos.....                                                 | 22        |
| 3.1.4. Equipo de monitorización del discurso de odio en las redes sociales .....                       | 23        |
| 3.2. Metodología de análisis de los datos .....                                                        | 24        |
| 3.2.1. Variables utilizadas en el análisis .....                                                       | 24        |
| <b>4. RESULTADOS.....</b>                                                                              | <b>26</b> |
| 4.1. Contenidos monitorizados y notificados y reacción de las plataformas de internet.....             | 26        |
| 4.1.1. Características del contenido retirado a las 24 horas, a las 48 horas y a la semana             | 27        |
| 4.1.2. Contenido retirado como Trusted Flagger .....                                                   | 28        |
| 4.2. Características del discurso de odio en las redes sociales .....                                  | 29        |
| <b>CONCLUSIONES .....</b>                                                                              | <b>41</b> |
| <b>5. Anexo I: Ejemplos de discurso de odio .....</b>                                                  | <b>44</b> |



## ÍNDICE DE TABLAS

|                                                                                                                              |    |
|------------------------------------------------------------------------------------------------------------------------------|----|
| Tabla 1. Porcentaje de contenidos retirados según el tiempo transcurrido desde la notificación y según plataforma, 2023..... | 27 |
| Tabla 2. Distribución de la tipología del discurso notificado, 2023 .....                                                    | 31 |
| Tabla 3. Distribución de la tipología del discurso de odio (%) en cada grupo diana.....                                      | 32 |
| Tabla 4. Distribución de episodios prototípicos relacionados con el discurso de odio según grupo diana. ....                 | 38 |

## ÍNDICE DE GRÁFICOS

|                                                                                                                                                           |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Gráfico 1. Porcentaje de comunicaciones a cada plataforma .....                                                                                           | 26 |
| Gráfico 2. Porcentaje de contenidos retirados según el tiempo transcurrido desde la notificación al conjunto de las plataformas monitorizadas, 2023 ..... | 27 |
| Gráfico 3. Porcentaje de discurso de odio dirigido a cada grupo diana .....                                                                               | 29 |
| Gráfico 4. Prevalencia de los grupos diana de discurso de odio en cada red social .....                                                                   | 30 |
| Gráfico 5. Distribución de la tipología del discurso de odio (%) en cada grupo diana .....                                                                | 32 |
| Gráfico 6. Distribución de la tipología del discurso de odio según red social .....                                                                       | 33 |
| Gráfico 7. Tipología del discurso de odio según reacción de las plataformas respecto a la retirada .....                                                  | 34 |
| Gráfico 8. Frecuencia de la expresión del discurso de odio .....                                                                                          | 34 |
| Gráfico 9. Distribución de la expresión del discurso de odio según grupo diana al que se dirige ...                                                       | 35 |
| Gráfico 10. Distribución del tipo de expresión del discurso de odio según la plataforma de internet .....                                                 | 36 |
| Gráfico 11. Distribución del discurso de odio notificado a las plataformas según su vinculación con un episodio prototípico .....                         | 36 |
| Gráfico 12. Distribución de episodios prototípicos relacionados con el discurso de odio según grupo diana .....                                           | 39 |
| Gráfico 13. Distribución de la expresión del discurso de odio según el episodio prototípico .....                                                         | 40 |

## PRESENTACIÓN INSTITUCIONAL

Me complace presentar este primer **“Informe Anual de Monitorización del Discurso de Odio en Redes Sociales”** que corresponde a los resultados del año 2023, de la monitorización diaria del discurso de odio racista, xenófobo, islamófobo, antisemita y antigitano que se lleva a cabo desde el Observatorio Español del Racismo y la Xenofobia (OBERAXE), dependiente de la Secretaría de Estado de Migraciones, del Ministerio de Inclusión, Seguridad Social y Migraciones.

Este informe constituye el compendio de la información revisada y consolidada, que cada dos meses publica el OBERAXE en los boletines de monitorización del discurso de odio.

La monitorización, que se realiza desde el año 2020, se desarrolla en cinco plataformas de internet: Facebook, Instagram, TikTok, YouTube y X (antes Twitter), y tiene dos objetivos principales. El primero es hacer un seguimiento del comportamiento de las plataformas de internet en cuanto a la moderación de contenidos: si retiran o no los contenidos de odio que se les notifican, cuánto tiempo tardan en hacerlo, qué información ofrecen a los usuarios respecto a sus notificaciones o si mejoran las estrategias y herramientas de moderación. Todo ello siguiendo la estela de la colaboración iniciada en el *Código de Conducta para la lucha contra la incitación ilegal al odio en Internet*, firmado desde el año 2016 por la Comisión Europea con las principales plataformas de internet.

El segundo objetivo es ofrecer una descripción de las principales características del discurso de odio que detectamos: los grupos de la población más afectados, el tipo de discurso y la gravedad de este, los episodios que desencadenan o a los que se vincula el discurso de odio, entre otras. Hay que recordar que, en la valoración del posible discurso de odio, hay que tener en cuenta dos derechos fundamentales como son la libertad de expresión y la dignidad humana.

Con este primer informe anual pretendemos poner a disposición de la Administración Pública, la sociedad civil, las personas expertas interesadas y la ciudadanía una información que permita acercarnos al conocimiento del estado de la situación, poner de manifiesto el peligro que supone el discurso de odio para la sociedad, y servir como material de formación y sensibilización para aquellos con competencias o interés en esta cuestión.

La Secretaría de Estado de Migraciones y el Ministerio de Inclusión, Seguridad Social y Migraciones tienen la misión de desarrollar e implementar las políticas de extranjería, inmigración y emigración. Un aspecto fundamental de estas políticas es la integración social de las personas inmigrantes que se ve crecientemente amenazada por el racismo y la xenofobia manifestados a través de los discursos y los delitos de odio.

Este informe es una herramienta más, que nos ha brindado una vía de colaboración con las plataformas de internet para mejorar la moderación de contenidos, y que nos ayuda en la senda de cooperación con todos los agentes implicados en la lucha contra los discursos de odio para avanzar en el diseño de políticas que aseguren la protección de la dignidad y los derechos de todas las personas que constituyen la ciudadanía diversa que habita en nuestro país.

Las personas, ante todo. Los derechos humanos por encima de todo.

Pilar Cancela Rodríguez

Secretaria de Estado de Migraciones

## 1. INTRODUCCIÓN

El incremento, en los últimos años, del uso de plataformas digitales ha facilitado tanto la difusión de conocimientos e ideas como la propagación de mensajes de odio racistas, xenófobos y de otras formas de intolerancia. El discurso de odio tiene repercusiones profundas y multifacéticas que abarcan desde el daño sobre la salud mental y física de las víctimas hasta la normalización de la intolerancia y, en consecuencia, la incitación a la discriminación sistemática y a la violencia. El discurso de odio contribuye a la polarización y fragmentación de la sociedad al reforzar prejuicios y estereotipos negativos, fomentando así una sensación de vulnerabilidad en las víctimas, y un entorno social de desconfianza y hostilidad.

Cuando las personas usuarias de las redes sociales ven que los mensajes de odio no son retirados o moderados adecuadamente, o incluso reciben apoyo de la comunidad virtual, pueden percibir estas actitudes como aceptables o incluso como la norma. Esta normalización del discurso de odio hace que las personas que son objeto del mismo se retraigan socialmente, produciendo su aislamiento y reduciendo su participación en la vida comunitaria y pública.

El discurso de odio no se queda en las redes sociales, sino que representa una amenaza en la vida real para los colectivos a los que se dirige, y atenta contra la cohesión social y la seguridad de la ciudadanía. En España, se han incrementado en 2022 los delitos de odio en un 3,72% con respecto al año anterior, según el último [Informe sobre la evolución de los delitos de odio en España](#), publicado por el Ministerio del Interior. Los delitos de odio por racismo y xenofobia, con 755 hechos, son los más numerosos y representan el 43,5 % de total de las denuncias. Le siguen los delitos de odio cometidos en los ámbitos de la orientación sexual e identidad de género (459 hechos).

Del mismo modo, los datos del [informe sobre delitos de odio de 2022 de la Organización para la Seguridad y la Cooperación en Europa \(OSCE\)](#) reflejan que el mayor número de incidentes de odio estuvieron motivados por el antisemitismo, con 3.575 casos, seguidos por el racismo y la xenofobia (2.510), el odio contra personas LGTBI (1.054) y el odio contra personas cristianas (852). Según el tipo de delito o incidente, los ataques son dirigidos con mayor frecuencia hacia las propiedades de los grupos diana, incluyendo daños, actos de vandalismo e incendios; o se trata de amenazas y ataques violentos contra las personas.

Teniendo en cuenta que, en España en 2023, el número de personas extranjeras ascendía a 6.812.698<sup>1</sup>, es imprescindible mantener en la agenda política y social la prevención y la lucha contra el racismo, la xenofobia y la intolerancia asociada; poner de manifiesto el valor de la diversidad y la contribución de las personas migrantes a la sociedad española; y asegurar el respeto a su dignidad y derechos fundamentales.

---

<sup>1</sup> Fuente: Extranjeros con certificado de registro o tarjeta de residencia en vigor SGAM - OPI  
[https://www.inclusion.gob.es/web/opi/estadisticas/catalogo/extranjeros\\_con\\_certificado](https://www.inclusion.gob.es/web/opi/estadisticas/catalogo/extranjeros_con_certificado)

La prevención y la lucha contra los discursos de odio son esenciales para contribuir a mitigar la discriminación que sufren las personas migrantes y para favorecer su integración social. Dada la naturaleza rápida y expansiva de las plataformas digitales, para combatir el discurso de odio se requieren estrategias continuas y multidimensionales que incluyan entre otras: la sensibilización y formación de los operadores jurídicos, otros profesionales, de escolares y jóvenes, y de la ciudadanía en general; herramientas legislativas; e instrumentos tecnológicos que faciliten tanto la detección, retirada y bloqueo de contenidos de odio, como el desarrollo de contranarrativas y discurso alternativo. Además, la cooperación entre las administraciones públicas, las organizaciones de la sociedad civil y las plataformas de internet es crucial para crear un entorno digital más seguro y libre de discursos de odio.

En esta línea, la Secretaría de Estado de Migraciones promovió la elaboración del [Marco Estratégico de Ciudadanía e Inclusión contra el Racismo y la Xenofobia \(2023 -2027\)](#) aprobado por el Consejo de Ministros el 4 de julio de 2023. Uno de los objetivos tácticos que se contemplan en este Marco Estratégico es impulsar estrategias y planes regionales y locales para prevenir el racismo y la xenofobia, y combatir los delitos de odio y el creciente discurso de odio contribuyendo a proteger los derechos de las personas, la seguridad y la democracia; así como impulsar mecanismos de vigilancia, prevención, detección y eliminación del racismo, la xenofobia y la intolerancia asociada en los medios de comunicación, internet y redes sociales.

Además, en el año 2015 se firmó el [Acuerdo para cooperar institucionalmente en la lucha contra el racismo, la xenofobia, la LGBTIfobia, y otras formas de intolerancia](#), con la participación de: la Fiscalía General del Estado, el Consejo General del Poder Judicial, seis Departamentos ministeriales<sup>2</sup> y el Centro de Estudios Jurídicos. El Acuerdo cuenta además con hasta siete observadores entre plataformas de organizaciones de la sociedad civil (OSC) y órganos consultivos<sup>3</sup>. A través del Acuerdo se lanzó en marzo de 2021 el [“Protocolo para combatir el discurso de odio ilegal en línea”](#) como una herramienta de colaboración entre las instituciones, las OSC y las plataformas de internet<sup>4</sup> para combatir el discurso de odio en línea. El Acuerdo incluye, además, dos grupos de trabajo específicos sobre “sobre discurso de odio en línea” y “discurso de odio en el deporte”.

Asimismo, es necesario destacar los avances en la lucha contra los delitos de odio y los discursos de odio producidos como resultado del lanzamiento del [primer](#) y [segundo](#) Plan de Acción de lucha contra los Delitos de Odio del Ministerio de Interior. También el trabajo continuado de la Fiscalía General del Estado, y particularmente de la Fiscalía de Sala contra los delitos de odio y la discriminación con su red de fiscales provinciales especializados. La red de oficinas de atención a víctimas del delito establecida por el Ministerio de Justicia. La aprobación de la [Ley de Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación](#) promovida por el Ministerio de Igualdad. La atención especial del Ministerio de Educación, Cultura y Deporte a la formación en valores democráticos, convivencia en la red, o la prevención del acoso y el

<sup>2</sup> Ministerios de Presidencia, Relaciones con las Cortes y Justicia; Interior; Educación Cultura y Deporte; Igualdad; Inclusión Seguridad Social y Migraciones; Derechos Sociales, Consumo y Agenda 2030.

<sup>3</sup> el Foro para la Integración Social de los Inmigrantes (FISI); el Consejo para la Eliminación de la Discriminación Racial o Étnica; el Consejo de Víctimas de Delitos de Odio y Discriminación; la Plataforma del Tercer Sector; la FELGTB; el Consejo Estatal del Pueblo Gitano; y la Plataforma de la Infancia.

<sup>4</sup> A lo largo del texto se habla plataformas de servicios de alojamiento de datos, plataformas de internet, y plataformas de redes sociales de manera indistinta.

ciberacoso entre otros. Y la extensión de programas de formación sobre delitos de odio y discurso de odio dirigidos a la carrera judicial organizados por el Centro de Estudios Jurídicos y el Consejo General del Poder Judicial.

En la Unión Europea también se presta una atención especial a los delitos de odio y al discurso de odio, lo que se ha ido plasmando en la normativa comunitaria, entre la que hay que destacar: la [\*Decisión marco de 2008\*](#) *relativa a la lucha contra determinadas formas y manifestaciones del racismo y la xenofobia mediante el Derecho penal*; la [\*Directiva 2012/29/UE\*](#) *del Parlamento Europeo y del Consejo, de 25 de octubre de 2012, por la que se establecen normas mínimas sobre los derechos, el apoyo y la protección de las víctimas de delitos*; [\*la Directiva 2010/13/UE\*](#) *del Parlamento Europeo y del Consejo, de 10 de marzo de 2010, relativa a la coordinación de determinadas disposiciones legales, reglamentarias y administrativas de los Estados miembros relativas a la prestación de servicios de comunicación audiovisual*, que prohíbe la incitación al odio en los servicios de comunicación audiovisual y la promoción de la discriminación en las comunicaciones comerciales audiovisuales; o la [\*la Ley de Servicios Digitales \(DSA\)\*](#) que entró completamente en vigor en febrero de 2024 con el objeto de crear un espacio digital más seguro en el que se protejan los derechos fundamentales de los usuarios además de establecer unas condiciones de competencia equitativas para las empresas.

Además, la Comisión Europea coordina el [\*Grupo de Alto Nivel para Combatir el Discurso de odio y los Delitos de odio\*](#) compuesto por representantes de los Estados miembros, del Parlamento Europeo, organizaciones de la sociedad civil, la Agencia Europea de Derechos Fundamentales, el Consejo de Europa, la OSCE y del Alto Comisionado de Naciones Unidas para los Refugiados. El Grupo tiene como objetivo el intercambio y la diseminación de buenas prácticas entre autoridades nacionales y el debate para identificar el mejor modo de prevenir y combatir el discurso de odio y los delitos de odio.

Por otro lado, desde la Comisión Europea Contra el Racismo (ECRI) del Consejo de Europa, se emiten recomendaciones que se configuran como herramientas esenciales para la lucha contra la discriminación, el antisemitismo, la islamofobia o el antigitanismo. La ECRI emitió su [\*recomendación de política general n.º 15 relativa a la lucha contra el discurso de odio\*](#) en la que resalta que, la libertad de expresión y de opinión no constituyen derechos ilimitados y deben ejercerse de forma que no atenten contra otros derechos de las personas. En mayo de 2022 el Comité de Ministros del Consejo de Europa aprobó la [\*Recomendación CM/REC \(2022\) 16 del Comité de Ministros a los Estados miembros sobre la lucha contra el discurso de odio\*](#).

El artículo 71 de la [\*Ley Orgánica 4/2000\*](#), de 11 de enero, de derechos y libertades de los extranjeros en España y su integración social establece la constitución del Observatorio Español del Racismo y la Xenofobia (OBERAXE) dependiente de la Secretaría de Estado de Migraciones del Ministerio de Inclusión, Seguridad Social y Migraciones. El OBERAXE tiene “funciones de estudio y análisis, y con capacidad para elevar propuestas de actuación, en materia de lucha contra el racismo y la xenofobia”.

El OBERAXE impulsa, participa y colabora en actividades y proyectos para prevenir y combatir el racismo y la xenofobia, y para fomentar la integración de las personas de origen extranjero en



la sociedad española. Una de sus líneas de actuación es la monitorización del discurso de odio en las redes sociales que se lleva a cabo de forma sistemática desde 2020, y que se describe en los apartados siguientes de este informe.

En la monitorización del discurso de odio, el OBERAXE colabora con la Fiscalía General del Estado a través de la denuncia de contenidos que potencialmente podrían ser calificados como delito de odio. Además, colabora con las Fuerzas y Cuerpos de Seguridad del Estado en el marco de sus labores de detección e investigación de incidentes y posibles delitos de odio de mayor complejidad, y se canaliza a través de los Equipos de Respuesta a los Delitos de la Guardia Civil (REDO) y los equipos de investigación de Extremismo Violento y Odio (EVO) de la Policía Nacional.

Otra parte importante de la labor realizada por el OBERAXE para combatir el discurso de odio se lleva a cabo a través de proyectos cofinanciados por la Comisión Europea, en colaboración con la Oficina Nacional de Lucha Contra los Delitos de Odio (ONDOD) del Ministerio del Interior. Se trata de proyectos para mejorar las capacidades de las fuerzas policiales a nivel nacional y local para combatir los delitos de odio y los incidentes xenófobos y racistas en colaboración con la sociedad civil<sup>5</sup>; o proyectos para mejorar las capacidades de las autoridades estatales para identificar, analizar, supervisar y evaluar el discurso de odio en línea, desarrollar estrategias de contranarrativa contra el discurso de odio y generar un mapa de buenas prácticas<sup>6</sup>.

---

<sup>5</sup> Proyecto CISDO: Cooperación Interpolicial y Social contra los Delitos de Odio

<sup>6</sup> Proyecto REAL UP: Discurso de odio, racismo y xenofobia: mecanismos de alerta y respuesta, análisis del discurso *Upstander*

## 2. OBJETO DE LA MONITORIZACIÓN DEL DISCURSO DE ODIO

### 2.1. Origen, contexto y escenario actual de la monitorización

La monitorización del discurso de odio en las redes sociales realizada por el Observatorio Español del Racismo y la Xenofobia (OBERAXE) se inspira 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 YouTube, Twitter, Facebook y Microsoft con la Comisión Europea, y al que se han ido adhiriendo otras plataformas de prestación de servicios digitales, como Instagram y TikTok. Mediante este Código de Conducta, las plataformas de redes sociales adquirieron una serie de compromisos con la Comisión Europea para luchar contra la incitación al odio, como el establecimiento de procedimientos claros y eficaces para evaluar las notificaciones de discurso de odio de los usuarios; y la revisión de la mayoría de las notificaciones en menos de 24 horas, retirando o deshabilitando el contenido de sus servidores si así se considera.

Con el objetivo de examinar la respuesta de las plataformas ante los compromisos adquiridos con la Comisión, anualmente se realizan ejercicios de evaluación del cumplimiento del Código de Conducta, que tienen seis semanas de duración y en los que participa el OBERAXE, entre otras instituciones y organizaciones nacionales y europeas.

Partiendo de este precedente, en el año 2020 el OBERAXE llevó a cabo una monitorización específica impulsada por la necesidad de analizar el discurso de odio contra las personas asiáticas que se estaba produciendo en el contexto de la irrupción de la pandemia de COVID-19. Tras esta experiencia, se decidió ampliar el análisis al discurso de odio con motivación racista, xenófoba, islamófoba, antisemita y antigitana; y continuar con una monitorización diaria de las redes sociales, efectuada durante todo el año y no solo circunscrita al periodo del ejercicio de evaluación del Código de Conducta.

Antes de abordar con detalle el objeto de la monitorización del OBERAXE, es necesario mencionar algunos mecanismos y recursos esenciales para combatir el discurso de odio en internet a nivel europeo y nacional desarrollados en los últimos años, en un escenario en el que nace y se consolida la monitorización del OBERAXE, marcada por la rápida y constante evolución de las herramientas digitales, sus usos y su alcance en la sociedad.

A este respecto, destaca el *Protocolo para Combatir el Discurso de Odio Ilegal en Línea* impulsado por la Secretaría de Estado de Migraciones del Ministerio de Inclusión, Seguridad Social y Migraciones en 2021 con la vocación de ser una guía para la cooperación y la colaboración entre los actores institucionales de la sociedad civil y las empresas de servicios de alojamiento de datos para prevenir, combatir y eliminar el discurso de odio ilegal en línea.

El Protocolo se inspira en el Código de Conducta de la Comisión Europea y en *la Recomendación (UE) 2018/334 de la Comisión de 1 de marzo de 2018 sobre medidas para combatir eficazmente los contenidos ilícitos en línea*; y se basa en la legislación española y en la europea vigente en la materia.

Para el desarrollo de mecanismos y de una respuesta común a nivel europeo frente al aumento de los discursos de odio es fundamental reconocer la labor del Grupo de Alto Nivel (GAN) de la Unión Europea para combatir el discurso de odio y los delitos de odio, coordinado por la Dirección General de Justicia de la Comisión Europea. La representación española en el GAN la ostenta el Ministerio de Inclusión, Seguridad Social y Migraciones (a través del OBERAXE) y el Ministerio de Justicia. Dentro del GAN se han constituido grupos de trabajo para ahondar en ámbitos específicos, como es el caso del grupo de trabajo para combatir el discurso de odio ilegal en línea, que, en 2023, pone el foco en la revisión del Código de Conducta y en el nuevo marco regulatorio de la UE, de forma que se posibilite una respuesta europea coordinada ante el avance de los discursos de odio y la intolerancia.

Por último, el escenario actual para combatir el discurso de odio en internet queda notablemente marcado por el *Reglamento (UE) 2022/2065, del Parlamento y el Consejo Europeo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales*, más conocido como Ley de Servicios Digitales (DSA), que es de aplicación completa desde el 17 de febrero de 2024.

La DSA constituye, por primera vez, un conjunto de normas y obligaciones aplicables a los proveedores de servicios digitales con el objetivo de establecer un marco sólido de transparencia, proteger los derechos fundamentales de los usuarios e impulsar la competitividad en el mercado único europeo a través de sistemas de reporte de contenidos ilegales y de la supervisión de las plataformas por parte de las autoridades.

En este nuevo panorama, la moderación y retirada por parte de las plataformas de redes sociales de los contenidos de discurso de odio ilegal publicados en sus servidores no respondería únicamente a un compromiso, como sucede en el marco del Código de Conducta, sino a unas normas de obligado cumplimiento, acompañadas de un régimen de sanciones aplicable en el caso de haberse infringido las normas establecidas.

Debido a la naturaleza transfronteriza de los servicios digitales, la DSA contempla la creación del coordinador de servicios digitales, una figura que en cada Estado miembro debe actuar como único punto de contacto y contar con funciones específicas para supervisar la aplicación y ejecutar el Reglamento, siendo también el encargado de otorgar la condición de *trusted flagger* o alertador fiable (ver definición en el apartado 3.1.2) a las entidades que lo deseen y cumplan con los requisitos pertinentes. En España, el organismo que desempeña las funciones de coordinador nacional es la Comisión Nacional de los Mercados y la Competencia (CNMC).

## 2.2. Objetivos de la monitorización

Uno de los dos objetivos principales de la monitorización es evaluar la respuesta de las plataformas de prestación de servicios de alojamiento de datos respecto a la retirada de contenidos de discurso de odio ilegal que se les notifican, en virtud de los compromisos adquiridos en el marco del Código de Conducta de la UE y de las obligaciones establecidas en la Ley de Servicios Digitales (DSA).

Con la entrada en vigor de la DSA, la moderación de los contenidos de discurso de odio por parte de las plataformas se articula en torno a dos ejes: la retirada de los contenidos ilegales en base

a la legislación nacional de los Estados miembros, tal y como se establece en el conjunto de obligaciones del Reglamento; y la retirada de los contenidos que infrinjan las normas de uso de la plataforma correspondiente. En este último caso, la moderación de los contenidos no respondería a una obligación legal, sino a la voluntad de la propia plataforma y a los compromisos derivados del Código de Conducta de la UE.

Desde el OBERAXE colaboramos con las plataformas de redes sociales en la identificación y moderación de los contenidos de discurso de odio, a través de diferentes canales de comunicación que incluyen reuniones periódicas en las que se genera un intercambio de información y de impresiones alrededor de la identificación y análisis de los contenidos de discurso de odio.

Además de la notificación de los contenidos y la evaluación de la respuesta de las plataformas, el otro objetivo principal de la monitorización es analizar el discurso de odio racista, xenófobo, islamófobo, antisemita y antigitano y detectar sus tendencias en España, contribuyendo al análisis de la situación y a orientar el desarrollo de políticas públicas.

### 2.3. ¿Qué es el discurso de odio?: definición y características principales

Tomando como referencia la *Recomendación nº15 de la Comisión Europea contra el racismo y la Intolerancia* del Consejo de Europa (ECRI), el discurso de odio se define como “el uso de formas de expresión específicas, como, por ejemplo, la defensa, promoción o instigación del odio, la humillación o el menosprecio de una persona o grupo de personas, así como el acoso, descrédito, difusión de estereotipos negativos o estigmatización o amenaza con respecto a dicha persona o grupo, y la justificación de estas manifestaciones, basadas en una lista no exhaustiva de características personales o estados que incluyen raza<sup>7</sup>, color de piel, idioma, religión o creencias, nacionalidad u origen nacional o étnico, edad, discapacidad, sexo, género, identidad de género y orientación sexual”.

En otras palabras, el discurso de odio es cualquier forma de expresión que denigre, humille o atente contra la dignidad de las personas por razón de alguna de sus características personales o estados, como puede ser su origen étnico, su nacionalidad o su orientación sexual.

A la hora de considerar una expresión como discurso de odio hay que tener en cuenta diferentes elementos. El primero de ellos es el tipo de expresión empleada para difundir el discurso. Por expresión se entienden tanto los discursos orales como escritos en cualquiera de sus formas, incluyendo el uso de medios digitales, imágenes, vídeos, símbolos e iconos, así como el lenguaje no verbal, por ejemplo, a través de gestos.

Las expresiones que se contemplan como discurso de odio han de suponer una clara incitación, promoción o fomento del odio, la violencia, la humillación, el acoso, el menosprecio, el descrédito, la estigmatización o la difusión de estereotipos negativos. Se incluye también la negación, trivialización o justificación de delitos de genocidio, de lesa humanidad o de delitos

---

<sup>7</sup> Todos los seres humanos pertenecen a la misma especie, la ECRI rechaza las teorías que sostienen la existencia de distintas razas. Sin embargo, en la Recomendación, la ECRI emplea el término “raza” a fin de garantizar que las personas que suelen percibirse de forma general y errónea como pertenecientes a otra raza quedan sujetas a la protección que confiere dicho texto.



en caso de conflicto armado cuya comisión haya sido comprobada tras recaer sentencia los tribunales, así como el enaltecimiento de las personas condenadas por haberlos cometido.

Dichas expresiones, para ser consideradas como discurso de odio, han de dirigirse, de manera inequívoca, a una persona o grupo de personas, quedando, por tanto, excluidos del concepto de discurso de odio, los mensajes que fomenten el odio o la hostilidad contra Estados, instituciones, partidos políticos, organizaciones, o los referidos a ideologías, conceptos o posturas políticas. Asimismo, el mensaje o expresión que constituya potencialmente un discurso de odio ha de proferirse en base a las características personales de una persona o grupo, como las ya citadas con anterioridad, y no en función de alguna otra condición o circunstancia.

Es decir, el hecho de proferir, por ejemplo, un insulto o una amenaza podría incurrir en algún tipo de delito o infracción administrativa; sin embargo, no se consideraría discurso de odio si dicha expresión que atenta contra la dignidad de la persona no se realiza por razón de alguna de las características protegidas mencionadas.

## 2.4. Dimensiones del discurso de odio

El discurso de odio es un fenómeno heterogéneo que se reproduce en múltiples manifestaciones que varían en intensidad y forma. Además, a la hora de abordar el discurso de odio desde el punto de vista legal es fundamental prestar atención a las particularidades de cada situación y al propio contexto en el que se difunde, que puede ser muy diferente en función de cada caso. A este escenario se suman las singularidades propias del ámbito digital, como la transformación continua de las redes sociales y las herramientas digitales, y el alcance global de la red, que precisan, en muchos aspectos, de una regularización específica que proteja a las personas usuarias de internet de los discursos discriminatorios e intolerantes a los que se encuentran potencialmente expuestas.

En consecuencia, para indagar con más detalle en el concepto de discurso de odio es necesario abordar sus diferentes dimensiones, que a su vez determinan la elección de las vías y estrategias para combatirlo. En este sentido, conviene puntualizar que no toda expresión discriminatoria entra dentro del concepto de discurso de odio, al igual que no todo discurso de odio es constitutivo de delito.

### 2.4.1. El discurso de odio ilegal y el discurso odioso amparado en la libertad de expresión

En primer lugar, es necesario establecer la diferenciación entre el discurso de odio ilegal y el discurso odioso no ilegal que queda amparado bajo el paraguas de la libertad de expresión. El discurso odioso, que no es considerado ilegal, representa el primer peldaño en la escalera del discurso de odio, si clasificamos este según su nivel de intensidad. Aunque no son ilegales, este tipo de discursos pueden ser igualmente dañinos y peligrosos; ya que de una manera velada contribuyen a generar un clima de odio en la sociedad construido en base a la difusión de estereotipos negativos y de la representación de determinados colectivos como una amenaza.

Es preciso mencionar que en el discurso de odio se produce una colisión de dos derechos fundamentales, como son la libertad de expresión y la dignidad humana. A la hora de considerar un mensaje como discurso de odio se realiza una ponderación entre ambos derechos, que no son absolutos, y se debe valorar, entre otros aspectos, el grado de intensidad de la incitación al odio y a la violencia. En este sentido, la libertad de expresión suele prevalecer cuando el discurso esté basado en juicios y opiniones que no constituyan un ataque directo y grave contra las personas.

También suele prevalecer cuando se interpreta que el mensaje persigue una finalidad aparentemente artística, humorística o satírica, así como cuando el discurso sea parte de un contenido informativo y esté basado en datos o en evidencias. En cualquier caso, para la ponderación de los derechos no existen unos criterios objetivos absolutos, sino que, en gran medida, depende de la interpretación derivada del estudio particular de cada caso.

#### *2.4.2. El discurso de odio ilegal: los delitos de odio y el discurso de odio ilegal no delictivo*

Dentro del discurso de odio ilegal distinguimos entre dos categorías: el delito de odio y el discurso de odio ilegal no delictivo.

En la categoría de discurso de odio ilegal no delictivo se incluye el discurso de odio que vulnera lo establecido en las leyes, pero que no llega a ser considerado delito, por lo que puede ser perseguido por la vía administrativa, al poder constituir una infracción administrativa, y no desde la vía penal.

El delito de odio representa el nivel más elevado de la escalera de intensidad del discurso de odio. Dentro de esta categoría solo se incluyen las expresiones de que son perseguidas desde el ámbito penal. Según lo recogido en documentos como *la Circular 7/2019, de 14 de mayo, de la Fiscalía General del Estado, sobre pautas para interpretar los delitos de odio tipificados en el artículo 510 del Código Penal*; podemos extraer una serie de pautas de interpretación general para considerar una expresión o contenido como delito de discurso de odio.

Uno de los principales criterios es la motivación discriminatoria; es decir, el perpetrador debe seleccionar a sus víctimas por motivos de intolerancia, y no por otros motivos o circunstancias. Por otra parte, en la interpretación de un hecho como posible delito, se tiene en cuenta la relevancia de la conducta. No se persiguen las meras ideas u opiniones, sino sólo aquellas conductas que infrinjan el bien jurídico protegido, la dignidad de la/s persona/s, o que sean susceptibles de generar un riesgo o peligro para el mismo. En este sentido, se toma en consideración que la conducta no solo atemorice a la persona destinataria del mensaje, sino a todo el colectivo al cual pertenece, creando sentimientos de lesión de la dignidad, inseguridad y amenaza.

Al mismo tiempo, se excluyen las manifestaciones “pretendidamente hilarantes” y las que se efectúan desde la “venganza puntual”; el ánimo ha de ser el de agredir. Asimismo, se considera que las expresiones realizadas deben violar las normas básicas de convivencia basadas en el

respeto y la tolerancia, quedando excluidas las opiniones sobre personas de notoriedad pública por su actuación y sometidas a cuestionamiento ciudadano.

A nivel general, fuera del régimen jurídico español, los criterios incluidos en la prueba de umbral del [Plan de Acción de Rabat](#) de las Naciones Unidas sobre la prohibición de la apología de odio nacional, racial o religioso que constituye incitación a la discriminación, la hostilidad o la violencia, son considerados un elemento de referencia en la interpretación de los posibles delitos de odio.

Estos criterios establecen que, para que una expresión o manifestación discriminatoria sea considerada delito, se ha de prestar atención al contexto particular de cada caso; a la posición o estatus social del autor; a la intención del autor; al contenido y forma del discurso, analizando los argumentos utilizados y el grado en el cual el discurso es directo y provocador; a la extensión del discurso, valorando el alcance, la naturaleza pública, así como la magnitud y el tamaño de la audiencia; y a la probabilidad de que el discurso logre incitar una acción real que pueda causar daños contra el colectivo objetivo.

## 2.5. El discurso de odio en el régimen jurídico español

### 2.5.1. *Ámbito penal: los delitos de odio*

En el ámbito penal, el discurso de odio está regulado principalmente a través del artículo 510 del Código Penal español, que contempla penas de prisión de uno a cuatro años y multa de seis a doce meses para quienes “públicamente fomenten, promuevan o inciten directa o indirectamente al odio, hostilidad, discriminación o violencia contra un grupo, una parte del mismo o contra una persona determinada por razón de su pertenencia a aquel, por motivos racistas, antisemitas, antigitanos u otros referentes a la ideología, religión o creencias, situación familiar, la pertenencia de sus miembros a una etnia, raza o nación, su origen nacional, su sexo, orientación o identidad sexual, por razones de género, aporofobia, enfermedad o discapacidad”. (art. 510.1.a).

Las sanciones también se aplican a la elaboración, tenencia y/o difusión de soportes aptos para incitar al odio, hostilidad, discriminación o violencia, comprendiendo cualquier escrito o soporte que engloba a los de carácter audiovisual o electrónico, según se recoge en el artículo 510.1.b. Por otra parte, el artículo 510.1.c. incluye la negación, trivialización grave o enaltecimiento de crímenes contra la humanidad.

A su vez, el artículo 510.2. establece penas de prisión de seis meses a dos años y multa de seis a doce meses para aquellas personas que lesionen la dignidad de las personas mediante acciones que entrañen humillación, menosprecio o descrédito de alguno de los grupos referidos en el apartado anterior; e incluye también la elaboración, tenencia y/o difusión de soportes entre las conductas tipificadas. A su vez, el art. 510.2.b recoge el enaltecimiento o justificación, por cualquier medio de expresión pública o de difusión, de los delitos de odio.

El artículo 510 incluye, asimismo, el tipo agravado de difusión mediática del discurso de odio, a través del cual las penas previstas se impondrán en su mitad superior cuando los hechos se

hubieran llevado a cabo a través de un medio de comunicación social, por medio de internet o mediante el uso de tecnologías de la información, de modo que, aquel se hiciera accesible a un elevado número de personas (art. 510.3).

Además del artículo 510, el Código Penal incluye en su articulado otras dos modalidades típicas de discurso de odio: el delito de enaltecimiento y de menosprecio a las víctimas del terrorismo (art. 578) y la incitación a la comisión de delitos de terrorismo (art.579).

Al margen del art.510, que regula de manera específica los llamados delitos de discurso de odio, en el Código Penal no existe un delito de odio tipificado como tal, sino que se incluyen una serie de tipos que podrían calificarse como delitos de odio. Ejemplo de ello es el agravante genérico de discriminación del artículo 22.4, que consiste en la comisión de un delito “por motivos racistas, antisemitas u otra clase de discriminación referente a la ideología, religión o creencias de la víctima, la etnia, raza o nación a la que pertenezca, su sexo, orientación o identidad sexual, razones de género, la enfermedad que padezca o su discapacidad”.

Otro ejemplo es la modalidad agravada del delito de amenazas (artículo 170 CP) en casos en los que estén “dirigidas a atemorizar a los habitantes de una población, grupo étnico, cultural o religioso, o colectivo social o profesional, o a cualquier otro grupo de personas, y tuvieran la gravedad necesaria para conseguirlo”.

### *2.5.2. Ámbito administrativo*

Fuera de la regulación penal, en el ordenamiento jurídico español existen varias leyes que regulan los incidentes de discurso de odio que no llega a ser tipificados como delito.

En primer lugar, destaca la *Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación*. En su artículo 47, la ley establece infracciones, que acarrearán sanciones administrativas con las cuantías establecidas en el artículo 48, para “los actos u omisiones que constituyan una discriminación, directa o indirecta, por asociación, por error, así como los que constituyan inducción, orden o instrucción de discriminar a una persona por razón de las causas previstas en el apartado 1 del artículo 2 de esta ley”. A su vez, se sanciona toda conducta de represalia por razón de discriminación; así como las conductas de acoso realizadas por razón de alguna de las causas de discriminación previstas en la ley.

La ley entró en vigor en el mes de julio de 2022, aunque la aplicación efectiva de la norma, y, por ende, de las sanciones previstas para las infracciones contempladas, están a la espera del nombramiento de la Autoridad Independiente para la Igualdad de Trato y la No Discriminación.

En el entorno específico de internet es de relevancia lo dispuesto en la [\*Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico\*](#). En el artículo 8, se condiciona la prestación de servicios en línea a que sus proveedores no infrinjan una serie de principios, entre los que se incluye la salvaguarda del orden público, la protección de la juventud y de la infancia, y el respeto a la dignidad de la persona y al principio de no discriminación.

Por último, en el ámbito del deporte, es aplicable lo dispuesto en la [\*Ley 19/2007, de 11 de julio, contra la violencia, el racismo, la xenofobia y la intolerancia en el deporte\*](#).

Por otra parte, a nivel autonómico cabe mencionar la [Ley 19/2020, de 30 de diciembre, de igualdad de trato y no discriminación](#), que regula las conductas de discurso de odio y prevé un régimen de sanciones administrativas cuyo ámbito de aplicación se circunscribe a la Comunidad Autónoma de Cataluña. Asimismo, en otras comunidades autónomas, como Madrid, Canarias o Andalucía, actualmente solo están reguladas algunas causas de discriminación.

## 2.6. ¿Qué monitoriza el OBERAXE?

La monitorización del discurso de odio en las redes sociales que realiza el OBERAXE consiste en la búsqueda, recopilación, análisis y notificación a las plataformas de contenidos de discurso de odio con motivación racista, xenófoba, islamófoba, antisemita y antigitana, 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.

El ámbito que comprende la monitorización son los contenidos de discurso de odio dirigidos a personas o grupos de personas por razón de su origen étnico, procedencia nacional o religión. Por sus propias particularidades, se presta especial atención a determinados grupos vulnerables, como es el caso de las personas inmigrantes; los niños, niñas y juventud no acompañada; y las personas refugiadas.

Si bien ocasionalmente en la labor de monitorización se identifican contenidos de discurso de odio que atacan a las personas por razón de otras características protegidas, como la orientación sexual o la identidad de género, estos no se recopilan ya que quedan fuera de nuestro objeto de análisis como Observatorio Español del Racismo y la Xenofobia. Aun así, estos contenidos de discurso de odio sí son monitorizados cuando la discriminación se presenta de forma múltiple e interseccional, respondiendo, por ejemplo, a una motivación racista y *LGTBIfóbica* al mismo tiempo. Por otra parte, en el análisis de los contenidos de discurso de odio también se presta atención a las publicaciones en las que el autor del mensaje menciona a otros colectivos vulnerables o grupos poblacionales, como el colectivo LGTBI o las mujeres; ya que, en ocasiones, se utiliza la vulnerabilidad de dichos grupos de manera tendenciosa como pretexto para canalizar el discurso de odio con motivación racista, xenófoba, islamófoba, antisemita o antigitana.

Conforme a las características inherentes que ha de presentar una expresión para ser calificada como discurso de odio, citadas con anterioridad, en la monitorización del discurso de odio en redes sociales del OBERAXE se excluyen del objeto de análisis los contenidos referidos a Estados, gobiernos, partidos políticos, instituciones religiosas, ejércitos, entidades y organizaciones.

En este sentido, recordamos que para que una expresión sea considerada discurso de odio, esta ha de estar dirigida a una persona o grupo de personas por razón de sus características personales, por lo que no se monitorizan contenidos que únicamente se refieran a ideologías, corrientes políticas, u otras ideas y conceptos.

Una dimensión particular adquieren los discursos realizados por representantes y figuras políticas, ya sean expresados en el desempeño de sus funciones representativas, por ejemplo, en un parlamento, o en el marco del debate político, especialmente en época de elecciones. La



libertad de expresión alcanza jurídicamente un mayor grado de expansión en el intercambio de ideas políticas en beneficio del sistema democrático. En este escenario, determinados discursos que podrían considerarse intolerantes o de incitación al odio en otros contextos, quedan amparados por la libertad expresión si son emitidos en el citado terreno político. Conforme a este hecho, en la monitorización del discurso de odio tenemos presente este condicionante, ciñéndonos a las expresiones que claramente son discurso de odio y no considerando discursos políticos que, aunque puedan resultar ciertamente polémicos, intolerantes y polarizados, quedarían amparados en el ejercicio de la libertad de expresión.

Asimismo, quedan fuera del objeto de la monitorización los contenidos basados en expresiones discriminatorias o de incitación a la violencia, por muy explícitas o graves que sean, cuando estas no se hayan proferido en función de alguna de las características protegidas o no se identifique a una persona o colectivo vulnerable como destinatario de dicha expresión. Este tipo de mensajes, basados, por ejemplo, en el uso de insultos y amenazas, pueden incurrir en algún tipo de delito o infracción, y no ser permitidas según las normas de uso de las redes sociales, pero no son abordadas desde el ámbito del discurso de odio.

Como ya se expuso con anterioridad, la motivación, en nuestro caso racista, xenófoba, islamófoba, antisemita o antigitana, es un factor determinante a la hora de considerar una expresión como discurso de odio. A este respecto, y en línea con la interpretación aportada por las plataformas de redes sociales respecto a algunos contenidos concretos, no recogemos en nuestra monitorización las publicaciones basadas en expresiones de incitación al odio cuando estas se dirijan a una persona como reacción a su comportamiento delictivo o criminal. Por tanto, no se recogen los contenidos que, según podamos interpretar, no respondan a una motivación de odio sino a una reacción directa motivada por un hecho delictivo.

Al margen del análisis de los contenidos de discurso de odio como tal, como añadido, prestamos atención de manera particular a la difusión de contenidos que hacen uso de expresiones, términos y referencias de carácter supremacista o asociadas al nazismo. Este análisis específico responde al hecho de que estas expresiones no siempre van referidas directamente a una persona o colectivo, sino que, a veces, aparecen de forma aislada y sin estar asociadas a ningún episodio prototípico (*ver definición en apartado 3.2.1*), constituyendo por sí mismas un tipo específico de discurso que promueve el supremacismo y la exaltación del nazismo a través del uso de elementos distintivos como símbolos, imágenes, siglas o números; o *dog whistles*, que son códigos lingüísticos que, mediante términos y palabras clave que funcionan en forma de eufemismos y dobles sentidos, promueven mensajes de odio en determinados círculos con el propósito de no ser detectados por el público general.

Por último, conviene citar que una parte importante del discurso de odio que identificamos en las redes sociales se relaciona directamente con la cuestión de la desinformación. En ocasiones, el discurso de odio se canaliza a través de publicaciones que difunden noticias, vídeos y datos falsos, manipulados o difícilmente verificables. La difusión de bulos y *fake news* contribuye a promover la hostilidad contra determinados grupos, como, por ejemplo, las personas inmigrantes y; además, es un fenómeno difícilmente abordable desde el prisma del discurso de odio ya que dichos contenidos constituyen en sí mismos un mensaje informativo y describen un hecho, sea cierto o no, y no una expresión que pueda exceder el límite de la libertad de expresión.

Desde el OBERAXE, a la hora de identificar las tendencias del discurso de odio, ponemos el foco en la utilización de los contenidos informativos como vía para promover los discursos de odio, con especial atención a la difusión de bulos y noticias falsas. En el caso de detectar contenidos potencialmente falsos o manipulados, las plataformas de redes sociales nos permiten, como usuarios, enviar un reporte específico sobre información falsa, a través del cual moderan el contenido tras realizar las correspondientes acciones de verificación.

## 3. METODOLOGÍA

### 3.1. Metodología de recogida de datos para la monitorización

La monitorización del discurso de odio que realiza el OBERAXE consiste en la búsqueda diaria, recopilación, análisis y notificación 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/Twitter).

#### 3.1.1. Metodología de búsqueda de contenidos

El procedimiento de búsqueda de los contenidos se realiza de forma manual a través de diferentes vías, que pueden variar en función de los usos y características de cada red social. En la monitorización se contemplan tanto publicaciones principales como comentarios y respuestas a estas, ya sean en formato de texto, imagen o vídeo, incluyendo los mensajes que hacen uso de emoticonos.

La búsqueda se limita a la observación de perfiles abiertos en las cinco redes sociales mencionadas, quedando fuera de nuestro ámbito los contenidos publicados en grupos privados, como es el caso de Telegram. La participación en este tipo de grupos, que precisan de una solicitud previa a los administradores, requiere en cierta medida adoptar un rol activo y de intervención, con los riesgos que esto supone, por lo que estas acciones no se ajustan al papel de observación y están más relacionadas con las labores de investigación llevadas a cabo por las Fuerzas y Cuerpos de Seguridad del Estado (FCSE) en el área de la criminalidad informática.

Dentro del ámbito de las FCSE, cabe mencionar la labor ejercida por los Equipos de Respuesta a los Delitos de la Guardia Civil (REDO), que investigan las conductas de odio y desarrollan actuaciones preventivas y operativas, realizando también acciones sensibilización y formación dentro del cuerpo. Por otra parte, los equipos de investigación de Extremismo Violento y Odio (EVO) de la Policía Nacional realizan tareas de investigación específicas para la persecución y esclarecimiento de los delitos de odio. La principal guía de actuación de estos equipos se encuentra recogida en el [Protocolo de Actuación de las Fuerzas y Cuerpos de Seguridad para los Delitos de Odio y conductas que vulneran las normas legales sobre la discriminación](#).

Retomando la metodología de la monitorización del OBERAXE, una de las principales vías para la búsqueda de contenidos son los perfiles y redes de usuarios que difunden de manera recurrente contenidos de discurso de odio; que promueven el clima de odio en sus publicaciones o que en ellas tratan temas relacionados con los episodios prototípicos (definidos en el apartado 3.2. Metodología de análisis de los datos) que desencadenan el discurso de odio. Otra vía para la identificación de los contenidos es la búsqueda en noticias, temas de actualidad y tendencias de las redes sociales vinculadas a los episodios prototípicos.

Ambas vías se complementan con la búsqueda de palabras clave seleccionadas de un glosario de términos, de elaboración propia que se actualiza semanalmente e incluye conceptos generales, términos discriminatorios por sí mismos, lenguaje en código y *dog whistles*. Si bien la búsqueda por palabras da sus frutos en redes como X/Twitter, no es tan efectiva en otras plataformas como YouTube o TikTok.

La metodología basada en la búsqueda manual, aunque no permite trabajar con un volumen elevado de datos, posibilita la identificación de determinados contenidos de discurso de odio cuya detección a través de herramientas informáticas plantea aún importantes retos, debido a que requieren de un análisis “humano” del contexto de cada caso y de la relación entre los diferentes elementos que intervienen en el discurso.

El hecho de prestar atención diariamente a las tendencias generales del discurso de odio en las redes y de analizar cada caso de manera pormenorizada, posibilita trabajar con una muestra pequeña, pero de “calidad”, en la que se otorga especial importancia a la categorización de cada contenido en función de las características que ha de presentar el discurso de odio.

Cabe mencionar que la labor del OBERAXE en la monitorización se circunscribe a la observación, identificación de contenidos de discurso de odio y reporte a las plataformas en el marco de colaboración con las propias plataformas y la Comisión Europea. El OBERAXE se reconoce como un participante activo en el trabajo de eliminación del discurso de odio en internet, aunque no adoptamos el papel de moderador, ya que el volumen de discurso de odio en las redes es, por desgracia, inabarcable con los mecanismos existentes. Los resultados de la monitorización no representan, por tanto, el total del discurso de odio en línea, sino que responden a la perspectiva de análisis del OBERAXE, y están basados de manera específica en la muestra recogida.

### *3.1.2. Metodología de reporte y denuncia de contenidos*

Además de la recopilación y el análisis de los contenidos de discurso de odio, el otro área que vertebra la monitorización, y de la que nace el proyecto, es la evaluación de la respuesta de las plataformas ante el discurso de odio difundido en sus servidores y el reporte de los contenidos ilegales o que infrinjan las normas de las plataformas.

Como ya hemos mencionado, la monitorización del OBERAXE se inspira en la metodología aplicada por la Comisión Europea en los ejercicios anuales de evaluación del Código de Conducta. A través de estos ejercicios, con la colaboración de instituciones y organizaciones de la sociedad civil de los diferentes Estados miembros, se examina si las plataformas cumplen con el compromiso adquirido de retirar el contenido de discurso de odio ilegal en línea y de atender a los reportes de los usuarios a través de procedimientos claros y eficaces.

Los ejercicios de evaluación se realizan desde el año 2016 con una periodicidad anual, aunque llevan paralizados desde 2022 ante la aprobación de la DSA y la adaptación al nuevo marco regulatorio europeo. Durante el periodo que dura el ejercicio, las instituciones y organizaciones notifican a la Comisión Europea los contenidos de discurso de odio reportados a las plataformas, y analizan la respuesta de estas ante los reportes, informando de si han retirado el contenido reportado o no, con atención al tiempo que ha transcurrido desde el reporte hasta la retirada.

El OBERAXE, en su monitorización diaria, aplica una metodología basada en el procedimiento contemplado en el Código de Conducta, adaptándolo al contexto nacional y al objetivo de disponer de un mecanismo de reportes y de seguimiento estable y sostenible en el tiempo.

Cuando identificamos en una red social un contenido de discurso de odio ilegal, ya sea potencialmente constitutivo de delito, que pueda constituir una infracción administrativa, o que infrinja las normas de uso de la propia plataforma en la que se haya publicado, lo reportamos a la plataforma siguiendo el procedimiento que establecen las condiciones generales de comunicación y denuncia de cada red social.

En línea con la metodología propuesta por la Comisión Europea, el reporte de los contenidos se realiza, en primer lugar, como usuarios normales de la red social. Una vez reportado, analizamos si el contenido ha sido retirado a las 24 horas, a las 48 horas, o a la semana desde que se envió la notificación.

Si habiendo transcurrido una semana, el contenido reportado permanece público en la red social, empleamos la vía cualificada de reporte conocida como *trusted flagger* (comunicador o alertador fiable). Los *trusted flagger* son individuos o entidades acreditadas especialmente por el prestador de servicios de alojamiento de datos por entender que cuenta con la cualificación o conocimiento necesario para notificar contenido de discurso de odio, al dedicar su actividad a cuestiones relacionadas con la lucha contra la discriminación y atesorar la experiencia necesaria en este ámbito.

Los reportes enviados a través de los formularios habilitados para los *trusted flagger* son revisados por las plataformas de manera preferente respecto a los reportes realizados por los “usuarios normales”. A diferencia de estos últimos, el perfil de los *trusted flagger* no es de carácter anónimo.

Ante las notificaciones remitidas, ya sea a través del procedimiento establecido para los usuarios como para los *trusted flagger*, las plataformas prestadoras de servicios de alojamiento de datos valoran el contenido reportado procediendo, si así lo consideran, a adoptar diferentes medidas como la retirada del contenido, la limitación de visibilidad de la publicación, el bloqueo en una región geográfica determinada, o la eliminación del perfil del usuario infractor.

Si identificamos contenidos de mayor intensidad de odio, que puedan ser constitutivos de delito de odio, estos se denuncian a la Fiscalía General del Estado a través de la Fiscalía de Sala de Delitos de Odio y Discriminación, que ejerce como Punto de Contacto atendiendo a los principios de unidad de actuación y dependencia jerárquica que rigen la actuación del Ministerio Fiscal, y a la normativa y protocolos de actuación existentes.

### 3.1.3. Metodología de anotación de los contenidos

Para la recogida de la información relativa a los contenidos de discurso de odio identificados en las redes, el OBERAXE hace uso, desde el mes de junio de 2022, de ALERTODIO, una aplicación desarrollada en colaboración con la Cátedra de Transparencia y Gestión de Datos de la Universidad Politécnica de Valencia, que tiene el fin de facilitar el registro sistemático de información en la actividad de monitorización. En la actualidad, el soporte y desarrollo técnico



de ALERTODIO lo realiza la Gerencia de Informática de la Seguridad Social (GISS), del Ministerio de Inclusión, Seguridad Social y Migraciones.

La aplicación cuenta con un apartado para la anotación de los contenidos de discurso de odio identificados por parte de cada persona monitorizadora. También incluye otro apartado para registrar las palabras clave identificadas (glosario) y otro para la generación de gráficos de resultados con los datos recogidos.

Las variables que se recogen en cada anotación de los contenidos son las siguientes:

P01. Identificación del monitorizador; P02. Identificador del caso; P02a.URL; P02b. Texto; P02c. Captura de pantalla; P03. Fecha y hora de publicación; P04. Fecha y hora de alta de la anotación; P05. Red social; P06. Id del usuario; P07. ¿El mensaje va dirigido a una persona o a un colectivo?; P08. ¿Se utiliza una supra categoría para referirse al grupo víctima?; P09. Grupo atacado; P10. Marca de sexo/género de la víctima; P11. Otros grupos victimizados; P12. Otros grupos o personas no victimizadas; P13. Episodio prototípico principal; P13b. Episodio prototípico secundario; P14. Contenido del mensaje de odio; P15. Lenguaje utilizado; P15b. ¿Incluye imágenes, videos o memes?; P16. Normas que podría infringir el contenido; P17. Reacción de la plataforma; P17b. Fecha de reporte (*trusted flagger*); P17c. Código TF; P17d. Respuesta de la plataforma (TF); P17e. Fecha de retirada (TF); P18. Comentarios.

#### *3.1.4. Equipo de monitorización del discurso de odio en las redes sociales*

El equipo del OBERAXE encargado de la monitorización del discurso de odio en línea cuenta, habitualmente, con ocho personas. Seis de ellas se dedican a monitorizar, asignándose una red social para cada una, con la posibilidad de reforzar alguna de las redes, debido a su complejidad o al volumen de contenidos, con otra persona monitorizadora. Dos personas cuentan con funciones de coordinación y análisis de los datos.

Con el objetivo de dar homogeneidad al trabajo de monitorización desempeñado en cada una de las plataformas, las personas encargadas de una red social dedican a su labor una hora y media diaria. En cuanto a la distribución del tiempo, se dedica una hora a la búsqueda y anotación de los contenidos, y media hora al seguimiento de los casos anotados y su actualización en ALERTODIO.

Una cuestión fundamental en la recogida de los datos es contar con cierta uniformidad de criterios por parte de las personas que integran el equipo a la hora de identificar, interpretar y analizar los contenidos de discurso de odio. En este sentido, semanalmente se celebra una reunión en la que participan todas las personas del equipo con el objetivo principal de revisar el trabajo semanal de monitorización, poner en común los temas y tendencias del discurso observadas, consolidar conocimientos y resolver dudas y otras cuestiones.

Las reuniones semanales también se plantean como un espacio seguro en el que poder compartir con el resto del equipo impresiones, sensaciones y preocupaciones derivadas de la tarea de monitorización, puesto que la exposición diaria a mensajes de incitación al odio, y en ocasiones a contenidos sensibles, explícitos y violentos, pueden generar efectos negativos a

nivel mental y emocional. Ante esta situación, el apoyo en otras personas y la comunicación con el resto del equipo y las personas responsables es esencial para poder buscar soluciones.

### 3.2. Metodología de análisis de los datos

Los contenidos de monitorización del discurso de odio se han analizado mediante una metodología cuantitativa utilizando los datos disponibles en la aplicación ALERTODIO correspondientes al año 2023. Los resultados que se muestran en los apartados siguientes responden a los dos objetivos de la monitorización: la descripción de la respuesta de las cinco plataformas de redes sociales (Facebook, Instagram, TikTok, YouTube y X/Twitter) al reporte de contenidos que realiza el OBERAXE; y la descripción de las características del discurso de odio según las categorías de interés del OBERAXE y los grupos de población diana que sufren el discurso de odio, y que se han considerado en la monitorización.

Para el análisis se ha utilizado el paquete estadístico SPSS. Con el objetivo de medir las variables cuantitativas y su distribución se han generado estadísticas descriptivas: frecuencias y medias. También se ha calculado la desviación típica para evaluar la dispersión de los datos y todas las variables se han medido con un nivel de confianza del 95% para determinar la precisión de las estimaciones.

También, se han creado tablas cruzadas para explorar la relación entre variables como la frecuencia de ocurrencia de los episodios prototípicos o la reacción de las plataformas de redes sociales ante contenidos que presentan a los grupos diana como una amenaza para la sociedad.

Por último, se ha realizado un análisis para identificar patrones recurrentes en el discurso de odio. Se han clasificado los datos en categorías relevantes (tipología del contenido discriminatorio y ofensivo, episodios prototípicos, expresión del lenguaje) y se han extraído resultados cualitativos.

El análisis, que se ha realizado sobre el discurso de odio notificado, ha proporcionado una comprensión integral del discurso de odio analizado y ha permitido obtener información significativa a partir de conjuntos de datos complejos y heterogéneos. Se ha tratado de identificar patrones y relaciones que pueden informar sobre estrategias de prevención y mitigación del discurso de odio en línea.

Hay que tener en cuenta que los datos globales presentados en este informe para el año 2023 pueden variar respecto a los boletines que se publican cada dos meses con resultados parciales de la monitorización ya que, tras el reporte de contenidos como *trusted flagger*, se hace un seguimiento del caso durante un mes adicional, lo que, en algún caso, puede modificar los resultados con posterioridad a la publicación del boletín bimestral.

#### 3.2.1. Variables utilizadas en el análisis

Las *categorías de discurso de odio* que se han analizado son: racismo, xenofobia, islamofobia, antisemitismo y antigitanismo.

Los *grupos de población diana* objeto de estas categorías de discurso que se han considerado son: personas inmigrantes en general; personas europeas; personas procedentes del norte de África; personas latinoamericanas; personas afrodescendientes; personas de origen asiático; personas refugiadas; niñas y niños y juventud no acompañada; comunidad gitana; comunidad judía. La islamofobia se ha considerado cuando se mencionaba el islam o las personas musulmanas. Se ha considerado una categoría “otros” cuando se explicitan otras personas o grupos de población; y una categoría “ninguno” cuando no se explicitan personas o grupos de población a los que se dirige el discurso de odio.

*Tipología del discurso de odio.* El discurso de odio se puede presentar como un acto claramente voluntario (un argumento) o como una expresión que podría parecer involuntaria (un insulto fruto de una discusión). En su contenido, el discurso de odio supone una promoción o instigación al odio, la humillación o el menosprecio. Puede incluir una variedad de conductas en cuanto a la tipología del contenido y los efectos que puede tener sobre los grupos diana se pueden categorizar en: los que suponen una amenaza para la integridad física o psíquica de una persona o grupo: *incitación a la violencia*, en este grupo se han incluido las categorías “amenazas directas e indirectas”, “alaba a quien atenta contra el grupo diana”, “incita a la expulsión del grupo” y “deshumaniza o degrada gravemente”; o los que suponen una discriminación incompatible con el principio de igualdad: *incitación a la discriminación*, donde se han incluido las categorías “presenta al grupo como una amenaza por sus acciones” y “promueve el descrédito en base a atributos personales del grupo”.

*Expresión del discurso de odio.* En la expresión del discurso de odio se pueden utilizar diferentes tipos de lenguaje. En este sentido, se han analizado tres categorías: “discurso agresivo explícito”, “discurso discriminatorio no agresivo” y “tono irónico o sarcástico”. El *discurso agresivo explícito* se caracteriza por su tono hostil, ataques directos y expresiones de violencia verbal o física hacia las personas o grupos diana. El *discurso discriminatorio no agresivo*, no utiliza necesariamente insultos o amenazas directas, pero si perpetúa estereotipos, prejuicios y exclusión hacia ciertos grupos en función de las características como la raza, género, religión, etc. El *tono irónico o sarcástico*. Se emplea para enmascarar contenidos de odio a través del humor o sarcasmo. También tiene un impacto negativo, pero por su carácter poco explícito se hace más difícil la retirada de este tipo de contenido.

*Episodio prototípico.* El discurso de odio es muy sensible a incidentes de diversa índole que ocurren en el corto plazo. Puede ser desencadenado o producirse un repunte de discurso de odio por la ocurrencia de determinados eventos que, a efectos de la monitorización del OBERAXE, hemos denominado episodios prototípicos. La agenda mediática y política del momento, así como la desinformación influyen en los episodios prototípicos. Los episodios prototípicos que hemos definido son los siguientes: salto de valla; llegada de pateras; inseguridad ciudadana (violencia y/o robo); okupaciones; conflicto armado; terrorismo; elecciones; políticas públicas: acción de un gobierno (subvención, apoyo, etc.); algún acontecimiento vinculado a la salud (pandemia, vacunas, etc.); algún acontecimiento vinculado a la economía (paro, etc.); agresión sexual; eventos deportivos; eventos religiosos; entre otros.

## 4. RESULTADOS

### 4.1. Contenidos monitorizados y notificados y reacción de las plataformas de internet

En 2023, se han notificado 2655 contenidos considerados de odio racista, xenófobo, antisemita, antigitano o islamófobo, que podían ser constitutivos de delito, de infracción administrativa o que violan las normas de conducta de las plataformas de internet, a las cinco redes sociales monitorizadas (Facebook, Instagram, TikTok, YouTube y X/Twitter). Las plataformas retiraron un total de 1313 contenidos, un 49,4% de los que les fueron notificados. 499 (18,8%) de los contenidos comunicados se retiraron cuando se notificaron como usuario normal y 814 (30,7%) cuando se notificaron como *trusted flagger*.

En la distribución de comunicaciones realizadas a cada plataforma (gráfico 1) hay un predominio de las realizadas a X/Twitter con 806 casos (30,4% del total), seguido de Instagram con 568 casos (21,4%), Facebook con 545 (20,5%), TikTok con 380 (14,3%) y YouTube con 356 (13,4%). Este desigual volumen de notificación de contenidos obedece, principalmente, al distinto grado de dificultad para su identificación en cada red social.

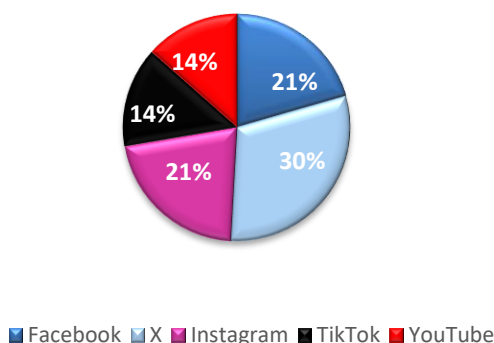


Gráfico 1. Porcentaje de comunicaciones a cada plataforma

La reacción de las plataformas en relación con el tiempo de retirada del contenido notificado se muestra en el gráfico 2 donde se observa que la mayoría de los reportes son retirados cuando se notifican como *trusted flagger* (30,7%) y que hay una proporción que no son retirados (50,0%). La eficiencia y la rapidez en la retirada de contenidos a las 24h, a las 48h, a la semana, o por la vía de *trusted flagger* se muestra en la tabla 1, y los resultados difieren entre las cinco plataformas, aunque, todas ellas son más receptivas a la eliminación de contenido cuando se emplea la vía de *trusted flagger*. La plataforma más eficiente es TikTok que retiró el 77,7% del total del contenido que se le notificó, seguida por Instagram (62,5%), Facebook (47,5%), YouTube (42,2%) y X/Twitter (31,5%).

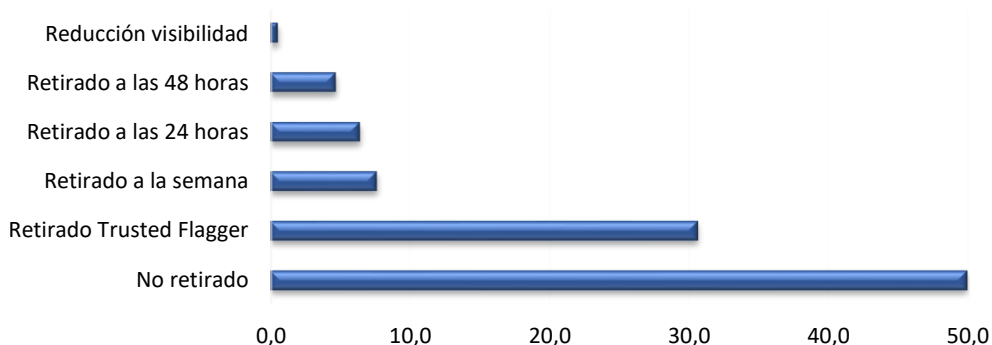


Gráfico 2. Porcentaje de contenidos retirados según el tiempo transcurrido desde la notificación al conjunto de las plataformas monitorizadas, 2023

|              | Total contenido retirado | Retirado a las 24 horas | Retirado a las 48 horas | Retirado a la semana | Retirado Trusted Flagger | Reducción visibilidad |
|--------------|--------------------------|-------------------------|-------------------------|----------------------|--------------------------|-----------------------|
|              | %                        | %                       | %                       | %                    | %                        | %                     |
| Facebook     | 47,5                     | 5,3                     | 0,7                     | 5,9                  | 35,6                     | 0,0                   |
| X/Twitter    | 31,5                     | 2,7                     | 2,1                     | 2,7                  | 23,9                     | 1,9                   |
| Instagram    | 62,5                     | 6,3                     | 5,1                     | 11,1                 | 40,0                     | 0,0                   |
| TikTok       | 77,7                     | 15,3                    | 12,4                    | 18,7                 | 31,3                     | 0,0                   |
| YouTube      | 42,2                     | 7,3                     | 7,9                     | 4,2                  | 22,8                     | 0,0                   |
| <b>Total</b> | <b>49,4</b>              | <b>6,4</b>              | <b>4,7</b>              | <b>7,6</b>           | <b>30,7</b>              | <b>0,6</b>            |

Tabla 1. Porcentaje de contenidos retirados según el tiempo transcurrido desde la notificación y según plataforma, 2023

#### 4.1.1. Características del contenido retirado a las 24 horas, a las 48 horas y a la semana

De las publicaciones que han sido comunicadas como usuario normal, las plataformas en su conjunto han retirado el 6,4% en las primeras 24 horas tras la notificación. TikTok es la red social que más contenido retiró en este periodo temporal (15,3%), seguida de YouTube (7,3%), Instagram (6,3%), Facebook (5,3%), y por último X/Twitter que solo ha retirado un 2,7% de sus contenidos en 24 horas.

Casi el 50% de los contenidos retirados a las 24 horas contienen lenguaje que deshumaniza o degrada o que expresa agresividad (54%). El grupo diana mayoritario de estos contenidos retirados a las 24 horas ha sido el de las personas originarias del norte de África (33,3% casos) y predomina la inseguridad ciudadana como episodio prototípico (38%).

El conjunto de plataformas ha retirado el 4,7% de los contenidos en el plazo de 48 horas desde la notificación. En los contenidos retirados en este plazo predomina el uso del lenguaje agresivo en un 58,4% y en un 36% el lenguaje discriminatorio no agresivo. El episodio prototípico principal

sigue siendo la inseguridad ciudadana en 44,8% en las notificaciones retiradas, y el grupo diana también son las personas originarias del norte de África (49,6%).

La plataforma que más contenido ha retirado a las 48 horas ha sido TikTok, que ha eliminado el 12,4% de los contenidos que se le notificaron, seguida de YouTube (7,9%), Instagram (5,1%), X/Twitter (2,1%) y Facebook (0,7%). Respecto al contenido retirado en el plazo de una semana, el conjunto de las plataformas ha retirado el 7,6% de las notificaciones, siendo TikTok la plataforma que más contenido ha retirado en este plazo temporal (18,7%), seguida de Instagram (11,1%), Facebook (5,9%), YouTube (4,2%) y X/Twitter (2,7%). Cabe resaltar que el 23,6% de los casos eliminados a la semana no están vinculados a ningún episodio prototípico, y el 53,2% de las comunicaciones retiradas en ese plazo contienen un lenguaje agresivo explícito.

#### 4.1.2 Contenido retirado como *Trusted Flagger*

La vía *trusted flagger* es la que más confianza ofrece a las plataformas para retirar contenido. Como muestra, el 30,7% de las 2156 notificaciones realizadas fueran retiradas por esta vía, frente al 18,8% cuando se notificó como usuario normal.

Al analizar la eficacia de cada plataforma ante la vía *trusted flagger*, se observa una variación respecto a la retirada como usuario normal. En este caso ya no es TikTok, sino Instagram la red social que más contenido de discurso de odio retira por la vía de *trusted flagger* (40% de los contenidos comunicados), le sigue Facebook (35,5%), TikTok (31,3%), X/Twitter con un 23,9% y por último YouTube (22,8%).

#### 4.1.3. Características del contenido no retirado

El porcentaje de contenido no retirado ha sido del 50,6% (1342 casos). Este porcentaje incluye un 0,6% de notificaciones para las que se ha reducido la visibilidad por parte de la red social X/Twitter, que ha establecido este mecanismo a finales de 2023 como acción positiva para disminuir el efecto del contenido de odio, que, aunque continúa circulando por la red, es menos visible para las personas usuarias.

Sin embargo, la retirada de contenido de discurso de odio por parte de las redes sociales sigue siendo insuficiente teniendo en cuenta que en un 79% de las comunicaciones se infringen las propias normas establecidas por cada una de las plataformas.

Al analizar el contenido no retirado (1342 notificaciones) se observa que en un 53% de los casos deshumaniza o degrada gravemente al colectivo al que se dirige el discurso; en un 44,2% promueve el descrédito; en el 24,2% incita a la expulsión del colectivo diana; en un 22,8% incita a la violencia; en un 7% alaba a quien atenta contra el grupo diana y un 0,5% presenta al grupo como una amenaza. La mayor parte de contenido no retirado estaba dirigido hacia las personas originarias del norte de África (445 casos) y predomina la inseguridad ciudadana como episodio prototípico.



## 4.2. Características del discurso de odio en las redes sociales

### 4.2.1. Grupos diana

Uno de los objetivos de la monitorización es identificar de qué manera están afectados los grupos de población diana del discurso de odio para prevenir, sensibilizar y reforzar acciones positivas. En 2023, en el 78,6% de las comunicaciones notificadas por el OBERAXE a las plataformas de redes sociales el mensaje iba dirigido a un colectivo, frente al 21,4% en el que se mencionaba a una persona en concreto.

Atendiendo al colectivo al que se dirige el discurso de odio (gráfico 3), predomina la hostilidad hacia las personas originarias del norte de África (33,7%). A continuación, se encuentra la islamofobia con un 26,2%. Cabe resaltar, además, que un 6,8% de los contenidos islamófobos tienen una marca femenina. El tercer colectivo más alcanzado por el discurso de odio es el de las personas afrodescendientes con un 23,7%. La categoría inmigrante se ha utilizado en un 14,4% de los casos, es decir, además de hacer mención a grupos específicos de población inmigrante se alude también a las personas inmigrantes en general.

El discurso de odio hacia la comunidad judía se observa en un 4,1% de los contenidos y se ha producido un aumento a raíz del conflicto en Oriente Próximo (Israel – Palestina). De hecho, el 41,4% de los contenidos dirigidos hacia este colectivo está vinculado a este acontecimiento. Por otro lado, cabe resaltar que el 78,3% de las comunicaciones que no están vinculadas a ningún colectivo específico están enmarcadas en un contexto supremacista, a través de la utilización de imágenes y simbología o expresiones relacionadas con el nazismo.

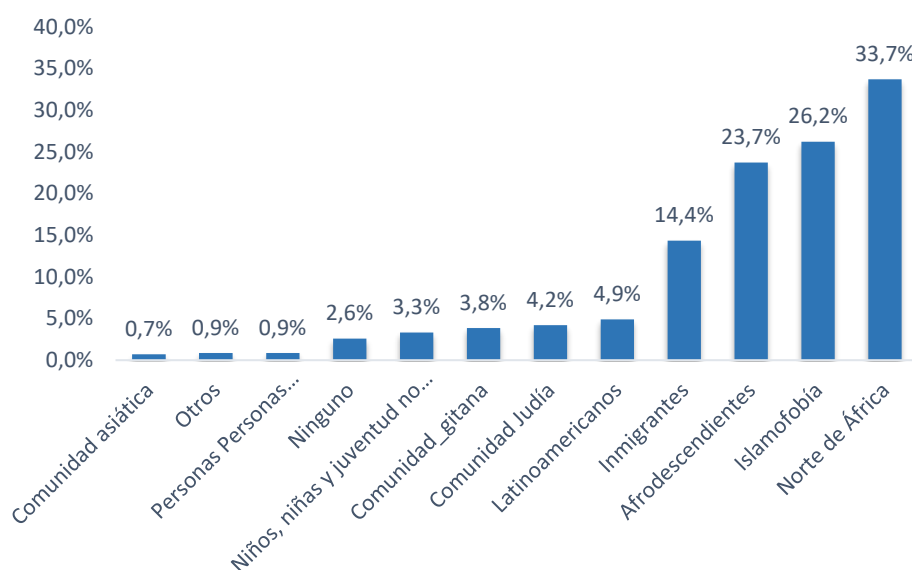


Gráfico 3. Porcentaje de discurso de odio dirigido a cada grupo diana

En referencia al grupo diana al que se dirige el discurso de odio según la plataforma en la que este se difunde de manera predominante, se constata que varía sin que se haya podido identificar ninguna relación causal (gráfico 4). Los discursos de odio dirigidos a los niños y niñas menores no acompañados (46,6%), comunidad judía (44,1%), personas afrodescendientes (33,4%), personas originarias del norte de África (30,5%), latinoamericanas (25,4%), e islamofobia (27,6%) predominan en la plataforma X/Twitter. Los discursos dirigidos a las personas europeas se encuentran más en YouTube (41,2%), y los contenidos dirigidos hacia la comunidad gitana se muestran con mayor frecuencia en TikTok (37,3%). Por último, el discurso dirigido a la categoría inmigrantes en general predomina en Facebook (30,2%) lo que ocurre también con el discurso dirigido a las personas refugiadas (52,2%).

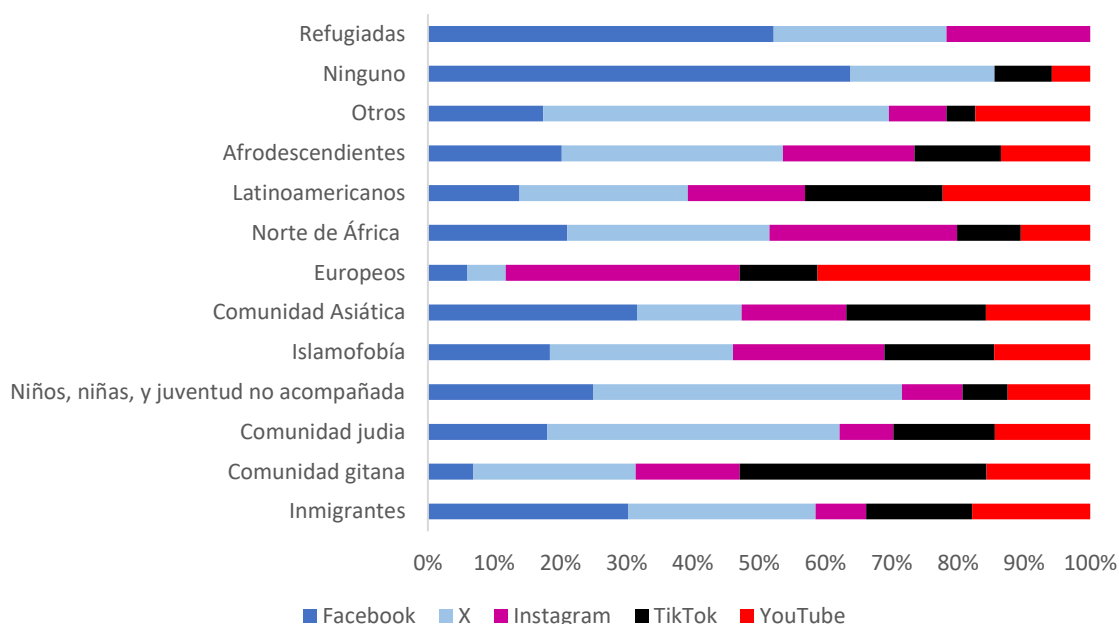


Gráfico 4. Prevalencia de los grupos diana de discurso de odio en cada red social

#### 4.2.2. Tipología del discurso de odio

Entre las tipologías de discurso observadas (tabla 2), la **deshumanización o degradación grave** de personas o colectivos vulnerables se encuentra en el 53,7% de las comunicaciones dirigidas a la población de origen extranjero. Esto fomenta la intolerancia y el prejuicio, así como la vulneración del derecho a la dignidad, igualdad y no discriminación. Algunos ejemplos de casos notificados expresaban lo siguiente: “Con una machine gun con cargadores de 300 balas cada uno se haría una limpieza en menos de media hora”; o “Hasta que no se abra la veda de la caza del moro en patrullas ciudadanas, esto aumentará exponencialmente”.

Asimismo, el descrédito atribuido únicamente a **características personales** del grupo o sin dar ningún argumento más que la pertenencia a este está presente en el 41,6% de los contenidos de odio analizados. Este hecho es clave para entender el prejuicio racista o xenófobo que se otorga a una característica biológica (el origen étnico o racial).

Del mismo modo, los resultados reflejan que en un 28,2% de los contenidos monitorizados se presenta a las personas migrantes y/o de origen extranjero como una **amenaza para la sociedad** por sus acciones (*“Urge limpieza étnica o seremos un país tercermundista en un par de décadas”*); y se incita a la **expulsión del colectivo** al que se dirige el discurso en un 21,6% de las comunicaciones (*“fuera los que no comen jamón vamos a rociar España con  ”*). Además, en un 7,6% se alaba al usuario/a que atenta contra el grupo diana (*“El pintor austriaco tenía que a ver terminado su trabajo del exterminio”*), esto puede provocar un aumento de los actos violentos y da legitimación a las personas que realizan el discurso de odio por el apoyo público que reciben.

| Tipología del discurso                                   | Sí % | No % |
|----------------------------------------------------------|------|------|
| Incita a la violencia con amenazas directas o indirectas | 24,4 | 75,6 |
| Deshumaniza o degrada gravemente                         | 53,7 | 46,3 |
| Alaba a quien atenta contra el grupo diana               | 7,6  | 92,4 |
| Incita a la expulsión del colectivo                      | 21,6 | 78,4 |
| Promueve el descrédito en base a atributos personales    | 41,6 | 78,4 |
| Presenta al grupo como una amenaza por sus acciones      | 28,2 | 71,8 |

Tabla 2. Distribución de la tipología del discurso de odio notificado, 2023

En cuanto a la distribución de los grupos diana según la tipología del discurso de odio, el gráfico 5 y la tabla 3 revelan que prevalecen los contenidos de **deshumanización** hacia los grupos de personas originarias del norte de África (70,9%), personas refugiadas (60,9%) y personas latinoamericanas (63,1%). Por otro lado, en el discurso hacia la comunidad gitana prepondera el **descrédito** en el 62,7% de los contenidos, en base a atributos personales o del colectivo, y que responde a la estigmatización social que sufre este grupo.

Cabe resaltar que en un 47,7% de los contenidos dirigidos hacia la comunidad judía se **alaba al usuario/a que ejerce el comentario**, dado que la mayoría de estos hacen alusión al nazismo. En cuanto al colectivo de niños, niñas y juventud no acompañada, en un 45,5% de las notificaciones se presenta al grupo como una **amenaza** para la sociedad.

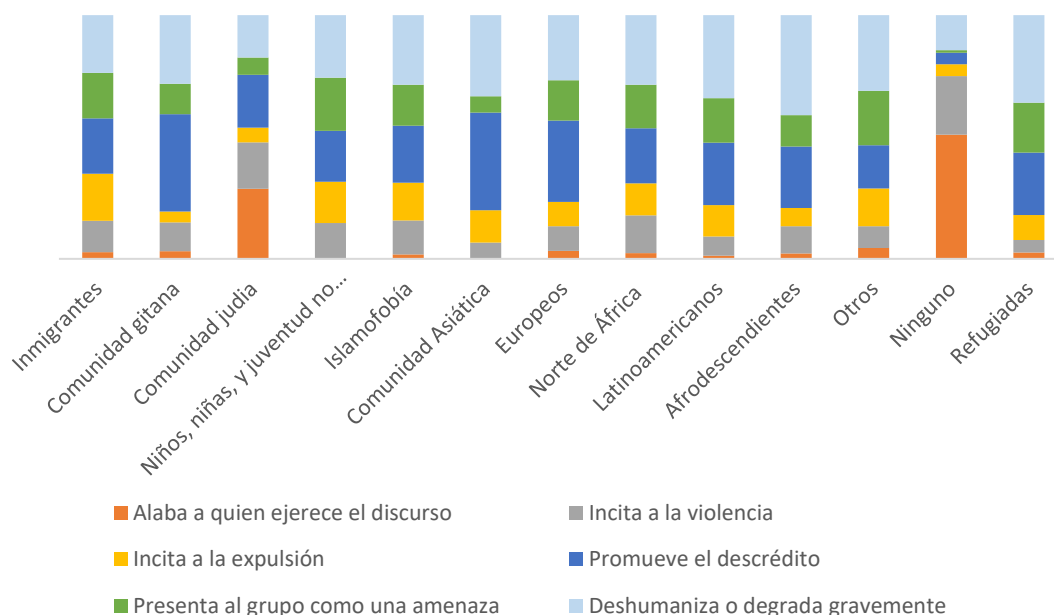


Gráfico 5. Distribución de la tipología del discurso de odio (%) en cada grupo diana

|                                        | Alaba a quien ejerce el discurso % | Incita a la violencia % | Incita a la expulsión % | Promueve el descrédito % | Presenta al grupo como una amenaza % | Deshumaniza o degrada gravemente % |
|----------------------------------------|------------------------------------|-------------------------|-------------------------|--------------------------|--------------------------------------|------------------------------------|
| Inmigrantes                            | 5,5                                | 25,5                    | 38,6                    | 45,1                     | 37,0                                 | 47,2                               |
| Comunidad gitana                       | 4,9                                | 18,6                    | 6,9                     | 62,7                     | 19,6                                 | 44,1                               |
| Comunidad judía                        | 47,7                               | 31,5                    | 9,9                     | 36,0                     | 11,7                                 | 28,8                               |
| Niños, niñas, y juventud no acompañada | 0,0                                | 30,7                    | 35,2                    | 43,2                     | 45,5                                 | 53,4                               |
| Islamofobia                            | 3,2                                | 24,7                    | 27,5                    | 41,4                     | 29,8                                 | 50,6                               |
| Comunidad Asiática                     | 0,0                                | 10,5                    | 21,1                    | 63,2                     | 10,5                                 | 52,6                               |
| Europeos                               | 5,9                                | 17,6                    | 17,6                    | 58,8                     | 29,4                                 | 47,1                               |
| Norte de África                        | 4,6                                | 30,1                    | 25,7                    | 44,0                     | 34,6                                 | 55,7                               |
| Latinoamericanos                       | 2,3                                | 14,6                    | 23,8                    | 47,7                     | 33,8                                 | 63,1                               |
| Afrodescendientes                      | 3,8                                | 19,4                    | 12,9                    | 43,7                     | 22,3                                 | 70,9                               |
| Otros                                  | 8,7                                | 17,4                    | 30,4                    | 34,8                     | 43,5                                 | 60,9                               |
| Ninguno                                | 76,8                               | 36,2                    | 7,2                     | 7,2                      | 1,4                                  | 21,7                               |
| Refugiadas                             | 4,3                                | 8,7                     | 17,4                    | 43,5                     | 34,8                                 | 60,9                               |

Tabla 3. Distribución de la tipología del discurso de odio (%) en cada grupo diana. **Nota:** los porcentajes de las filas pueden sumar más de 100 ya que un mismo contenido puede responder a varias tipologías.

Por otro lado, al analizar los contenidos de discurso de odio se ha revisado si también se mencionaban otros colectivos vulnerables además de los grupos de población diana objeto de la monitorización del OBERAXE (mujeres, personas LGTBIQ+, etc.). En la mayoría de las notificaciones (94,5%) esto no ha sucedido, salvo en un 3,7% donde el discurso va dirigido también contra las mujeres.

En el gráfico 6 se muestra la tipología del discurso de odio más predominante en cada plataforma. En TikTok y Facebook predomina el descrédito en base a atributos personales, con un 45,5% y un 53,6% respectivamente. La deshumanización de las personas está más presente en Instagram (58,2%) y YouTube (61,8%). Por último, el contenido de incitación a la violencia con amenazas directas o indirectas está más presente en X/Twitter (34,3%) y en Instagram (25,5%).

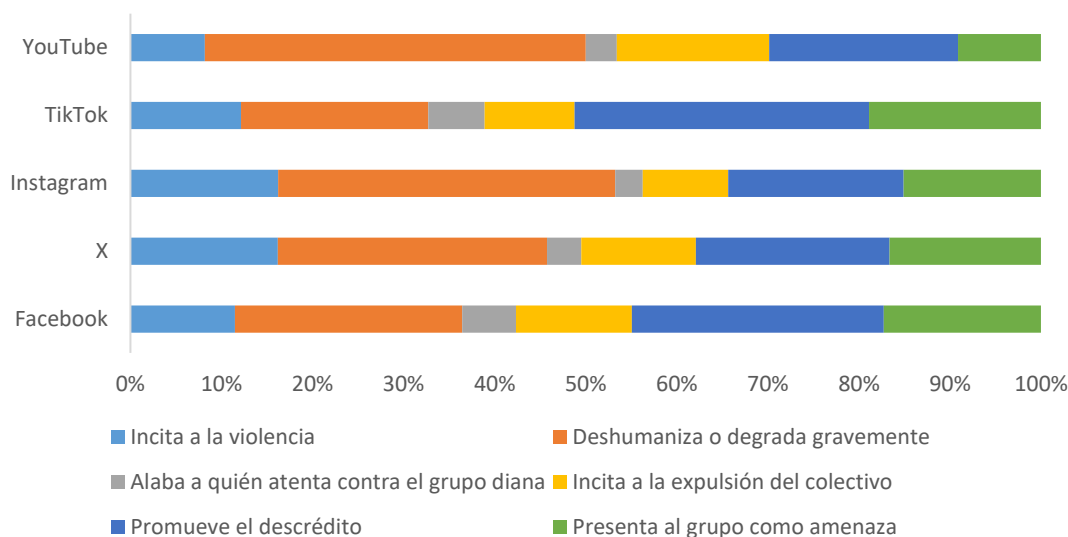


Gráfico 6. Distribución de la tipología del discurso de odio según red social

La tipología del discurso de odio en función de la reacción de las plataformas de retirarlo se muestra en el gráfico 7. Como era de esperar, el contenido que más se retira es el que incita a la violencia (53,0%) seguido del contenido que deshumaniza o degrada gravemente (50,6%). Sin embargo, hay que tener en cuenta que esto deja casi la mitad de este grave contenido sin retirar, y que la mayor parte de las plataformas de redes sociales sólo lo retira cuando se notifica como *trusted flagger* (28,35% del contenido que incita a la violencia y 33,08% del contenido que deshumaniza o degrada gravemente).

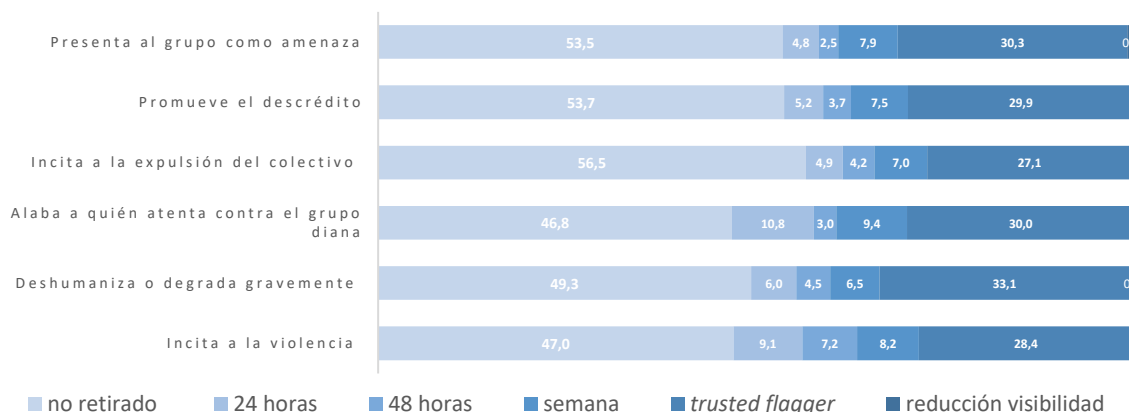


Gráfico 7. Tipología del discurso de odio según reacción de las plataformas respecto a la retirada

#### 4.2.3. Expresión del discurso de odio

De las tres categorías que se han considerado para la expresión del discurso de odio, los resultados muestran que el **discurso agresivo explícito** es el más frecuente ya que se observa en el 54% de los contenidos notificados. Ejemplos de comunicaciones encontradas: “*el mejor gitano...el gitano muerto*”; “*el mejor marroquí es el que más está ahorcado*”; “*ratas inmundas que deben ser expulsadas*”. El **discurso discriminatorio no agresivo** se observa en el 39,5% de los contenidos monitorizados con ejemplos como los siguientes: “*judaísmo=cáncer*”; o “*menuda limpieza hay que hacer en este país entre los de dentro y ahora monos*”. Y el **tono irónico o sarcástico** está presente en el 6,5% de los contenidos (gráfico 8). Algunos de los contenidos monitorizados son: “*mira un negro que habla xd 😊*”; “*yo digo que revivamos al pintor austriaco, para que termine de hacer su trabajo y que los de Israel deje de estar jugando al vivo*”; “*hay costumbres que son de la selva y de ahí no deberían salir*”.

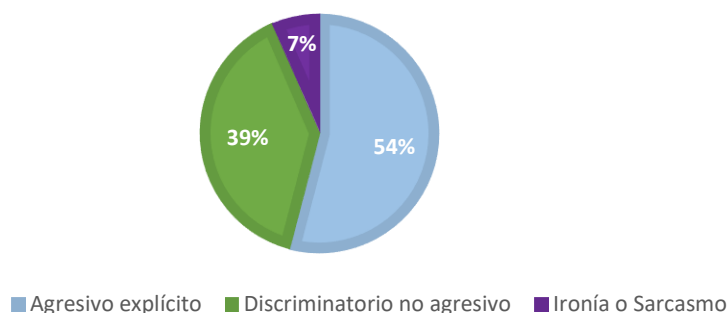


Gráfico 8. Frecuencia de la expresión del discurso de odio



En cuanto al género utilizado en los contenidos del discurso de odio, en un 94,1% de los casos notificados al conjunto de las plataformas se emplea el masculino genérico. Solo se emplea el femenino en un 3,3% de los casos y en un 2,1% el masculino individual.

En referencia a la expresión del discurso de odio empleado según el grupo diana al que va dirigido, se aprecian algunas diferencias, aunque en la gran mayor parte de los casos predomina el uso del discurso agresivo explícito. Siendo este utilizado en un 61,2% de los casos dirigidos hacia la población originaria del norte de África, un 60,9% en las personas refugiadas, 56,9% en personas afrodescendientes y un 55,1% en la islamofobia. Sin embargo, el gráfico 9 muestra como en el discurso de odio dirigido hacia la comunidad asiática (73,7%), la población inmigrante (49,3%) y la comunidad gitana (48%), en general prevalece más el discurso discriminatorio no agresivo.

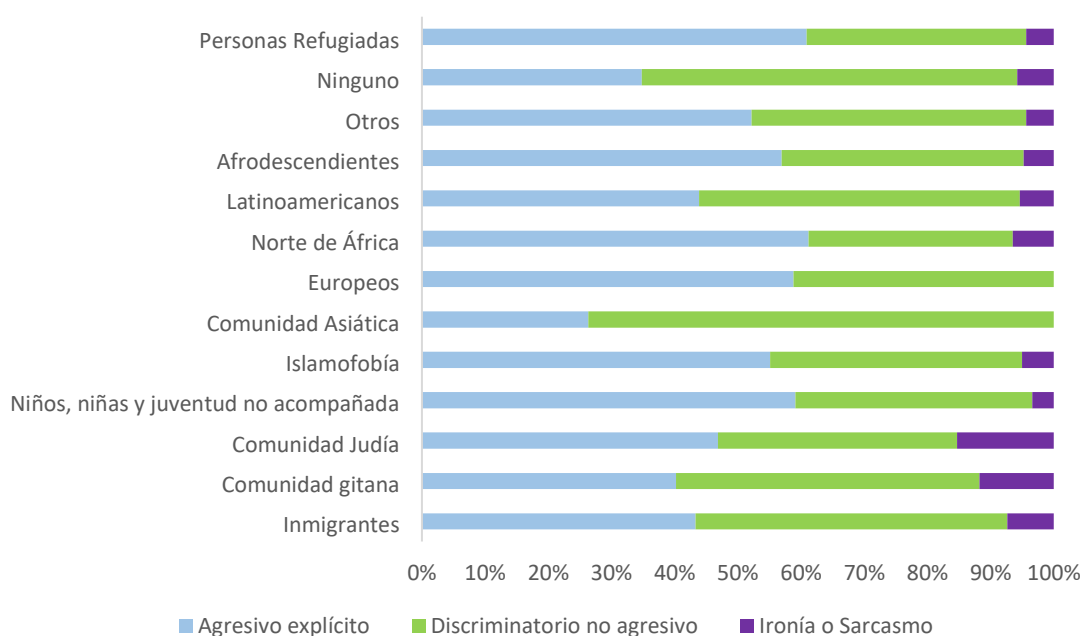


Gráfico 9. Distribución de la expresión del discurso de odio según grupo diana al que se dirige

Por otro lado, en el gráfico 10 se muestra el tipo de expresión de discurso de odio empleado según su presencia en cada una de las plataformas. En X/Twitter e Instagram predomina el discurso agresivo explícito, con un 38,7% y un 29,5% de los contenidos notificados a estas plataformas respectivamente. El discurso discriminatorio no agresivo es más predominante en Facebook (27,7%) y en YouTube (22,8%). Y el tono irónico es más frecuente en TikTok produciéndose en un 34,9% de los casos notificados.

La complejidad que supone detectar del discurso de odio cuando se usa una expresión irónica o sarcástica se refleja en los porcentajes de retirada por parte de las plataformas. Del total de notificaciones realizadas (2655), 172 son contenido irónico y se ha retirado el 41,3% frente al 58,7% no retirado. Sin embargo, los contenidos de discurso agresivo se han retirado en una

proporción del 51,8%, es decir 10,5 puntos porcentuales más que el contenido sarcástico, aunque en esta diferencia puede influir también la gravedad del lenguaje empleado.

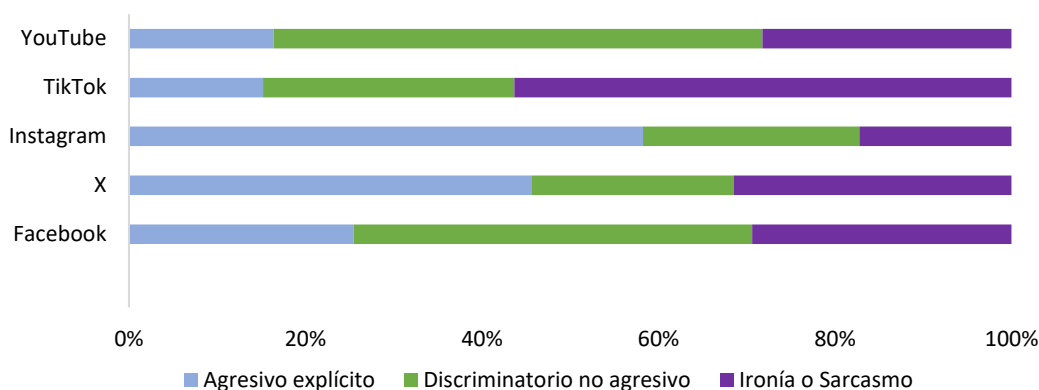


Gráfico 10. Distribución del tipo de expresión del discurso de odio según la plataforma de internet

#### 4.2.4. Episodio prototípico al que se vincula el discurso de odio

En el gráfico 11 se muestra la proporción de contenidos notificados a las plataformas según el episodio prototípico que los ha desencadenado o al que se vinculan. El principal episodio prototípico con el que se relacionan los contenidos notificados es la **inseguridad ciudadana** (real o supuesta) relacionada con la violencia y/o robo, que se observa en el 39,5% de los contenidos, y que se asocia sin necesidad de pruebas con un colectivo determinado.

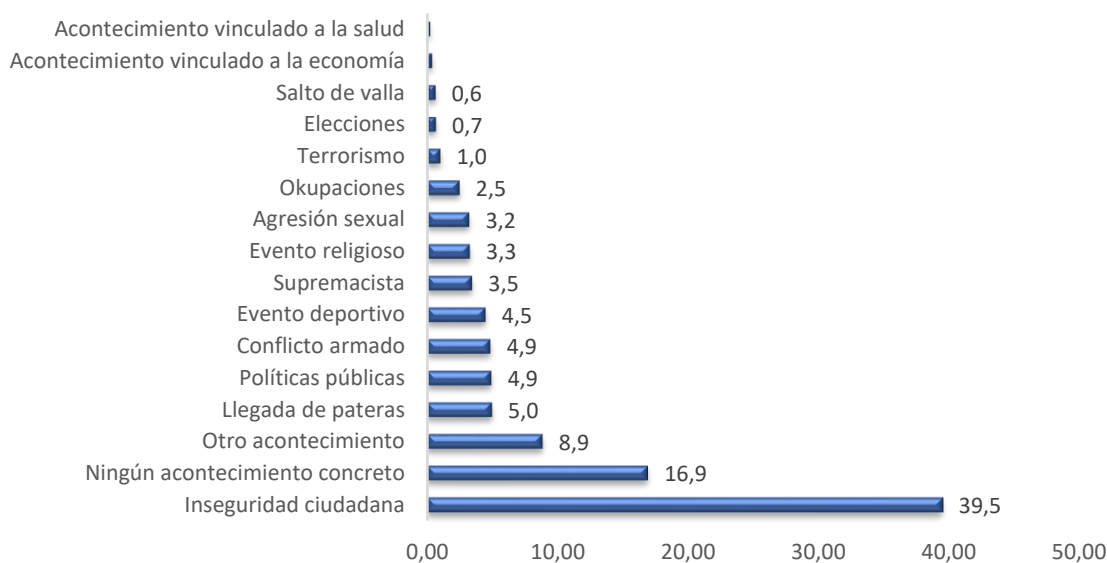


Gráfico 11. Distribución del discurso de odio notificado a las plataformas según su vinculación con un episodio prototípico

Un 4,5% de los contenidos monitorizados en 2023 están relacionados de forma directa e indirecta con el ámbito **deportivo**, en concreto en el fútbol. No obstante, el foco del discurso de odio ligado a este episodio tuvo un pico en el bimestre de mayo a junio donde el porcentaje de incidentes llegó al 15% a raíz de los sucesos relacionados con el caso del futbolista Vinicius. Este caso evidenció el racismo y xenofobia que existe en el ámbito deportivo, y que se difunde o promueve a través de las redes sociales.

Otro de los incidentes que han marcado el año 2023 han sido las **elecciones** generales, las autonómicas y municipales en España. El discurso de odio hacia la población inmigrante puede influir en las elecciones al polarizar el debate político y movilizar a ciertos segmentos de la población. El uso de la retórica antiinmigrante en las campañas electorales puede aumentar el discurso de odio en redes sociales y amplificar actitudes racistas y xenófobas. Por otra parte, un 4,9% de los contenidos están vinculados con las acciones que se desarrollan en materia de **políticas públicas** (subvenciones, apoyo/recursos para la integración, etc.), y que son un foco para promover la polarización política a través del discurso de odio. Además, ha sido en el periodo electoral cuando este suceso prototípico ha tenido mayor alcance llegando al 5%.

La **llegada de embarcaciones (pateras)** a las costas de las Islas Canarias ha suscitado el discurso de odio, con un 5% de los contenidos monitorizados en este año vinculado a este suceso, sobre todo, a partir del mes de octubre. En el 36,3% de los casos relacionados con la llegada de embarcaciones a las costas canarias se presenta al colectivo de inmigrantes como una amenaza. Además, en un 12,1% de los casos vinculados a este episodio prototípico se hace alusión al Gobierno central (5,3%), a las políticas públicas (4,5%), y al gobierno autonómico (2,3%).

En relación con el episodio prototípico **conflicto armado** se han analizado las comunicaciones relacionadas con la guerra entre Ucrania y Rusia y el conflicto en Oriente Próximo (Israel – Palestina). En 2023, el 4,9% de las notificaciones están relacionadas con estos dos acontecimientos. De ese porcentaje, un 29,4% ha estado ligado a la guerra entre Ucrania y Rusia. El resto está vinculado al conflicto en Oriente Próximo, a raíz del ataque terrorista del movimiento Hamás el 7 de octubre de 2023.

Asimismo, el 16,9% de los contenidos de discurso de odio no está relacionado con **ningún episodio prototípico**. Esto sugiere que el discurso anti inmigratorio que se difunde a través de las redes sociales, principalmente a través de *fake news*, también está suscitado, en una proporción de casos, solamente por actitudes individuales negativas hacia las personas inmigrantes y por estereotipos.

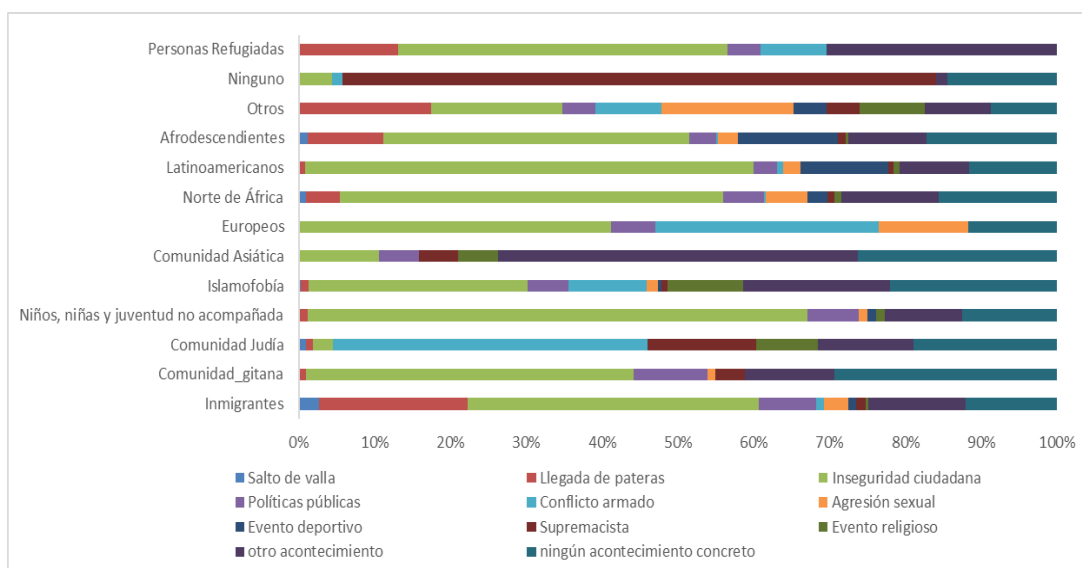
La distribución de los episodios prototípicos según el grupo de población diana al que afectan más, se recoge en la tabla 4 y gráfico 12. La inseguridad ciudadana es el episodio prototípico que predomina en todos los colectivos, aunque es más relevante en los discursos dirigidos hacia los niños, niñas y jóvenes no acompañados en los que el 65,9% de los casos notificados relativos a este grupo están ligados este episodio prototípico. Del mismo modo que ocurre con las personas de Latinoamérica (59,2%) y del norte de África (50,6%).

En cuanto al episodio prototípico de conflicto armado, se observa que un 41,4% de los contenidos dirigidos hacia la comunidad judía están relacionados con el conflicto en Oriente Próximo.

Por otro lado, los contenidos supremacistas predominan en la categoría “ningún colectivo” siendo un 78,3% de los casos. Se observa que la simbología nazi se emplea para atacar en general a la población de origen extranjero sin vincularla a ningún grupo diana específico.

|                            |          | Salto de valla | Llegada de pateras | Inseguridad ciudadana | Políticas públicas | Conflicto armado | Agresión sexual | Evento deportivo | Supremacista | Evento religioso | otro acontecimiento<br>(ocupaciones, economía,<br>salud, salto de valla,<br>elecciones, terrorismo y<br>otros) | ningún<br>acontecimiento<br>concreto | Total  |
|----------------------------|----------|----------------|--------------------|-----------------------|--------------------|------------------|-----------------|------------------|--------------|------------------|----------------------------------------------------------------------------------------------------------------|--------------------------------------|--------|
| Inmigrantes                | Recuento | 10             | 75                 | 146                   | 29                 | 4                | 12              | 4                | 5            | 1                | 49                                                                                                             | 46                                   | 381    |
|                            | %        | 2,6%           | 19,7%              | 38,3%                 | 7,6%               | 1,0%             | 3,1%            | 1,0%             | 1,3%         | 0,3%             | 12,9%                                                                                                          | 12,1%                                | 100,0% |
| Comunidad gitana           | Recuento | 0              | 1                  | 44                    | 10                 | 0                | 1               | 0                | 4            | 0                | 12                                                                                                             | 30                                   | 102    |
|                            | %        | 0,0%           | 1,0%               | 43,1%                 | 9,8%               | 0,0%             | 1,0%            | 0,0%             | 3,9%         | 0,0%             | 11,8%                                                                                                          | 29,4%                                | 100,0% |
| Comunidad Judía            | Recuento | 1              | 1                  | 3                     | 0                  | 46               | 0               | 0                | 16           | 9                | 14                                                                                                             | 21                                   | 111    |
|                            | %        | 0,9%           | 0,9%               | 2,7%                  | 0,0%               | 41,4%            | 0,0%            | 0,0%             | 14,4%        | 8,1%             | 12,61%                                                                                                         | 18,9%                                | 100,0% |
| Niños, niñas y juventud no | Recuento | 0              | 1                  | 58                    | 6                  | 0                | 1               | 1                | 0            | 1                | 9                                                                                                              | 11                                   | 88     |
|                            | %        | 0,0%           | 1,1%               | 65,9%                 | 6,8%               | 0,0%             | 1,1%            | 1,1%             | 0,0%         | 1,1%             | 10,2%                                                                                                          | 12,5%                                | 100,0% |
| Islamofobia                | Recuento | 2              | 7                  | 201                   | 37                 | 72               | 10              | 3                | 6            | 69               | 135                                                                                                            | 153                                  | 695    |
|                            | %        | 0,3%           | 1,0%               | 28,9%                 | 5,3%               | 10,4%            | 1,4%            | 0,4%             | 0,9%         | 9,9%             | 19,4%                                                                                                          | 22,0%                                | 100,0% |
| Comunidad Asiática         | Recuento | 0              | 0                  | 2                     | 1                  | 0                | 0               | 0                | 1            | 1                | 9                                                                                                              | 5                                    | 19     |
|                            | %        | 0,0%           | 0,0%               | 10,5%                 | 5,3%               | 0,0%             | 0,0%            | 0,0%             | 5,3%         | 5,3%             | 47,4%                                                                                                          | 26,3%                                | 100,0% |
| Europeos                   | Recuento | 0              | 0                  | 7                     | 1                  | 5                | 2               | 0                | 0            | 0                | 0                                                                                                              | 2                                    | 17     |
|                            | %        | 0,0%           | 0,0%               | 41,2%                 | 5,9%               | 29,4%            | 11,8%           | 0,0%             | 0,0%         | 0,0%             | 0,0%                                                                                                           | 11,8%                                | 100,0% |
| Norte de África            | Recuento | 8              | 40                 | 452                   | 48                 | 3                | 49              | 23               | 8            | 9                | 114                                                                                                            | 140                                  | 894    |
|                            | %        | 0,9%           | 4,5%               | 50,6%                 | 5,4%               | 0,3%             | 5,5%            | 2,6%             | 0,9%         | 1,0%             | 12,8%                                                                                                          | 15,7%                                | 100,0% |
| Latinoamericanos           | Recuento | 0              | 1                  | 77                    | 4                  | 1                | 3               | 15               | 1            | 1                | 12                                                                                                             | 15                                   | 130    |
|                            | %        | 0,0%           | 0,8%               | 59,2%                 | 3,1%               | 0,8%             | 2,3%            | 11,5%            | 0,8%         | 0,8%             | 9,2%                                                                                                           | 11,5%                                | 100,0% |
| Afrodescendientes          | Recuento | 7              | 63                 | 254                   | 22                 | 2                | 16              | 83               | 7            | 2                | 65                                                                                                             | 108                                  | 629    |
|                            | %        | 1,1%           | 10,0%              | 40,4%                 | 3,5%               | 0,3%             | 2,5%            | 13,2%            | 1,1%         | 0,3%             | 10,3%                                                                                                          | 17,2%                                | 100,0% |
| Otros                      | Recuento | 0              | 4                  | 4                     | 1                  | 2                | 4               | 1                | 1            | 2                | 2                                                                                                              | 2                                    | 23     |
|                            | %        | 0,0%           | 17,4%              | 17,4%                 | 4,3%               | 8,7%             | 17,4%           | 4,3%             | 4,3%         | 8,7%             | 8,7%                                                                                                           | 8,7%                                 | 100,0% |
| Ninguno                    | Recuento | 0              | 0                  | 3                     | 0                  | 1                | 0               | 0                | 54           | 0                | 1                                                                                                              | 10                                   | 69     |
|                            | %        | 0,0%           | 0,0%               | 4,3%                  | 0,0%               | 1,4%             | 0,0%            | 0,0%             | 78,3%        | 0,0%             | 1,5%                                                                                                           | 14,5%                                | 100,0% |
| Personas Refugiadas        | Recuento | 0              | 3                  | 10                    | 1                  | 2                | 0               | 0                | 0            | 0                | 7                                                                                                              | 0                                    | 23     |
|                            | %        | 0,0%           | 13,0%              | 43,5%                 | 4,3%               | 8,7%             | 0,0%            | 0,0%             | 0,0%         | 0,0%             | 30,4%                                                                                                          | 0,0%                                 | 100,0% |
|                            | Recuento | 17             | 132                | 1050                  | 131                | 129              | 86              | 119              | 92           | 87               | 235                                                                                                            | 450                                  | 2655   |

**Tabla 4. Distribución de episodios prototípicos relacionados con el discurso de odio según grupo diana. Nota:** Se han agrupado los episodios prototípicos. En “otro acontecimiento” se incluyen los casos relacionados con ocupaciones, economía, salud, elecciones, terrorismo y otros.



**Gráfico 12. Distribución de episodios prototípicos relacionados con el discurso de odio según grupo diana** *Nota: Se han agrupado los episodios prototípicos. En “otro acontecimiento” se incluyen los casos relacionados con ocupaciones, economía, salud, elecciones, terrorismo y otros.*

La relación entre la expresión del discurso de odio y el episodio prototípico que lo ha desencadenado se muestra en la tabla 5. Como ya se ha descrito el episodio prototípico más frecuente es el de “inseguridad ciudadana” y la expresión del discurso de odio que se usa con mayor frecuencia en estos casos es el agresivo explícito, que se observa en un 43,3% de los contenidos. El lenguaje discriminatorio no agresivo, es el más frecuente en el discurso relacionado con “llegada de pateras” aunque se muestra con una frecuencia elevada en discursos de odio que no se han vinculado a ningún episodio prototípico (20,4%). En los discursos vinculados a “evento deportivo”, “conflicto armado” y “políticas públicas” predomina el lenguaje irónico con un 9,3%, 8,1% y 6,4% respectivamente.

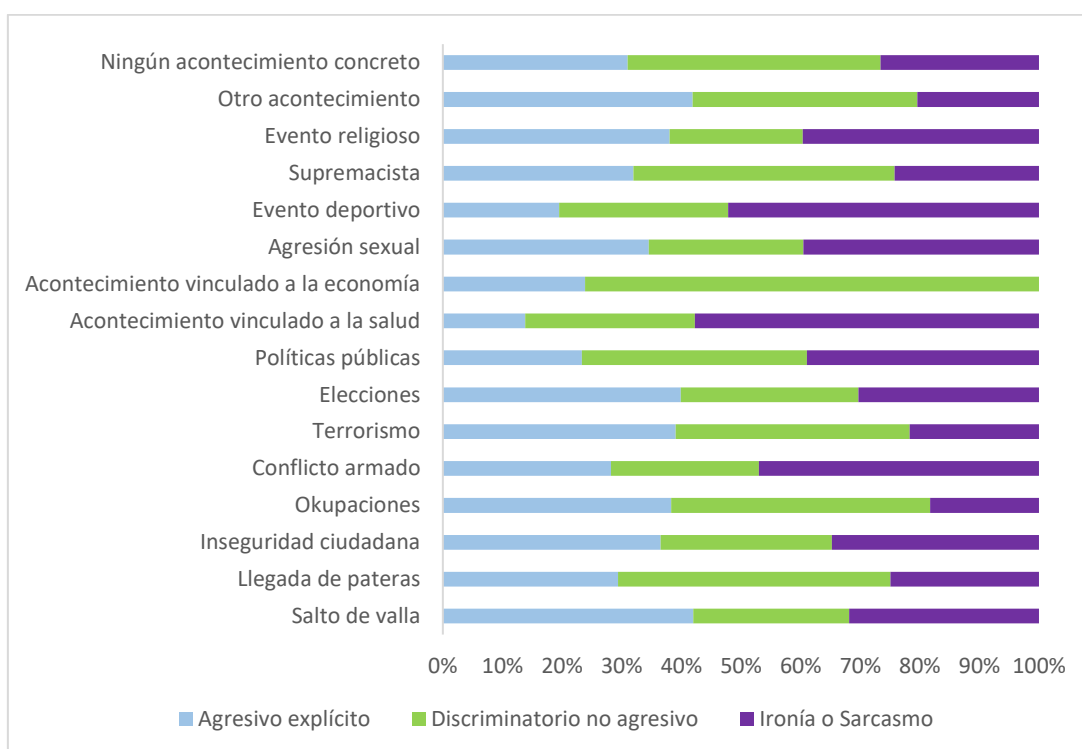


Gráfico 13. Distribución de la expresión del discurso de odio según el episodio prototípico

## CONCLUSIONES

Los resultados que se ofrecen en este informe son el fruto de la monitorización diaria del discurso de odio en línea que se lleva a cabo en el OBERAXE. Como ya se ha descrito, el trabajo se realiza de una forma manual, sistemática y protocolizada, utilizando un banco de palabras que se actualiza semanalmente y con una puesta en común continuada de las personas que monitorizan. Sin embargo, hay que tener en cuenta que los contenidos recogidos solo suponen una pequeña muestra del universo del discurso de odio que circula por las redes sociales y, además, la actuación de monitorización manual puede estar influenciada por algún tipo de sesgo de las personas que la realizan. A pesar de estas limitaciones, consideramos que los resultados mostrados son valiosos y ofrecen una aproximación a las características del discurso racista, xenófobo, islamófobo, antisemita y antigitano, y en consecuencia pueden contribuir al desarrollo de estrategias para combatirlo.

Los datos obtenidos muestran que las cinco plataformas monitorizadas (X/Twitter, Facebook, Instagram, TikTok y YouTube) retiraron menos de la mitad del contenido que se les notificó en 2023 (49,4% de los 2655 contenidos reportados), que podía ser constitutivo de delito, de infracción administrativa o que violaba sus propias normas de conducta internet. La mayor parte de ellos fueron retirados cuando se notificaron como *trusted flagger* (30,7% sobre el total notificado) y mucho menos cuando se notificaron como usuario normal (18,8% sobre el total notificado), lo que significa que los reportes de potencial discurso de odio que haga la ciudadanía como usuario normal tienen pocas probabilidades de ser retirados. No obstante, el OBERAXE mantiene una relación de colaboración con las plataformas de internet para compartir hallazgos y contribuir a mejorar su moderación de contenidos.

Las personas originarias del norte de África, seguido de la islamofobia son las dos categorías diana a las que más contenido de odio se dirige (33,7% y 26,2% respectivamente sobre el total monitorizado), lo que afecta principalmente a la población de origen marroquí que es la segunda nacionalidad extranjera más numerosa en España (787.317 personas<sup>8</sup>). Esto se puede relacionar con un prejuicio secular en nuestro país que es alimentado en las redes sociales aprovechando las circunstancias actuales de inmigración desde el norte de África y los ataques terroristas yihadistas ocurridos en algunos países. Para asegurar la inclusión social, la prosperidad y la convivencia pacífica de todos es necesario combatir de manera efectiva este discurso y proteger a toda la ciudadanía independientemente de su origen o características personales.

De los 2655 casos de contenido de odio monitorizados y reportados, en un 53,7% se deshumaniza o degrada gravemente a las personas a las que va dirigido, y en un 21,6% se incita a la expulsión del colectivo inmigrante. El episodio prototípico de “inseguridad ciudadana” (real o supuesta) constituye el desencadenante más frecuente del discurso de odio observado (39,5% de los casos) y generalmente se vincula a los grupos diana de manera gratuita. Esta

---

<sup>8</sup> Observatorio Permanente de la Inmigración: *Extranjeros con certificado de registro o tarjeta de residencia en vigor 31 diciembre 2023*. [Nota Extranjeros con certificado.pdf \(inclusion.gob.es\)](#)



estigmatización genera un ambiente hostil. Además, puede alimentar el miedo y resentimiento hacia ciertos grupos de la población, arrastrándonos a una mayor fragmentación y conflicto social. Además, la expresión de discurso de odio que más se asocia al episodio prototípico “inseguridad ciudadana” es el discurso agresivo explícito (43,3%).

Llama la atención que los niños/as y jóvenes no acompañados sean presentados como una amenaza para la sociedad (45,5% del discurso del que son objeto), se les vincula con la inseguridad ciudadana (65,09% de los contenidos de discurso que se dirige a este grupo), y se utiliza un lenguaje agresivo explícito contra ellos (52% de los casos). Todo ello obviando que son niños y niñas sin la tutela de sus padres o familiares y en una situación de extrema vulnerabilidad.

Los principales hallazgos de la monitorización del OBERAXE muestran los riesgos del mundo virtual que ha desencadenado una exaltación del racismo, la xenofobia y la intolerancia asociada. Se produce de manera impune un uso del discurso de odio que tiene como objetivo generar temor en la ciudadanía, y normalizar la idea de que hay personas y grupos de población que son peligrosos y sin valor como seres humanos, lo que favorece la incitación a la violencia y trata de llamar a su expulsión. Como afirman Soler y Flecha *las palabras juegan un papel trascendental no solo en la comunicación, sino en la construcción de la realidad social; y el habla, y otros signos pueden tener como objetivo una comunicación que busca acuerdos, o ser un simple instrumento para alcanzar otro objetivo diferente a la comunicación*<sup>9</sup>.

La cultura del discurso del odio que se propaga por las redes sociales tiene un impacto devastador en los colectivos más vulnerables de la sociedad. Los discursos de odio privan a las personas que los sufren de su humanidad y las culpabiliza de su destino. Además, los mensajes negativos tienen un impacto psicológico y físico, produciendo en ocasiones patologías, aumentan la vulnerabilidad y la sensación de inseguridad, y dificultan el desarrollo normal de la vida de las víctimas.

En los últimos años, la superposición de crisis de personas refugiadas, emergencias migratorias, crisis económicas y sanitarias, así como el uso extensivo de internet y las redes sociales, han dejado al descubierto el potencial que tienen los discursos de odio para promover la polarización de la sociedad y para vulnerar a personas de determinados grupos, especialmente las de origen inmigrante. En relación con ello, hay que señalar que la cultura del odio y, en particular, los discursos de odio son un mecanismo de la necropolítica (Butler, 2006; 2010; Mbembe, 2011) que produce activamente hostilidad hacia estas poblaciones, tratando de presentarlas como una amenaza para justificar una determinada construcción de Europa.

Es necesario avanzar en estrategias de contranarrativa que cambien el peso de los contenidos en las redes. Impulsar el rechazo activo de la ciudadanía hacia los discursos de odio y hacia los *haters*, contribuye al descrédito de estos y a la reparación de las víctimas. La responsabilidad de las plataformas de redes sociales, y la mejora de sus estrategias de retirada de contenidos, va

---

<sup>9</sup> Marta Soler, Ramón Flecha. Desde los actos de habla de Austin a los actos comunicativos. Perspectivas desde Searle, Habermas y CREA. [https://www.scielo.cl/scielo.php?pid=S0718-09342010000400007&script=sci\\_arttext&tlng=en](https://www.scielo.cl/scielo.php?pid=S0718-09342010000400007&script=sci_arttext&tlng=en)

más allá de lo que sucede en el mundo *online* puesto que tiene un efecto directo en el tipo de sociedad que estamos construyendo. Desde las instituciones, las organizaciones de la sociedad civil, las propias plataformas de internet y la Academia es necesario avanzar en el análisis, las estrategias de vigilancia y de actuación, pero también en normativas efectivas y recursos para combatir esta lacra que daña a la sociedad en su conjunto y amenaza la supervivencia de los sistemas democráticos.

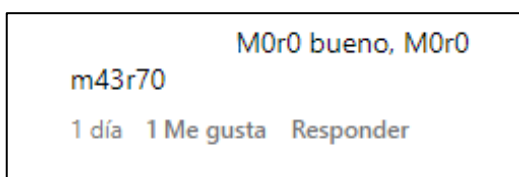
## 5. Anexo I: Ejemplos de discurso de odio

A continuación, se presenta una serie de contenidos de discurso de odio ilegal en línea monitorizados a lo largo del año 2023 que complementan a los recogidos a lo largo del texto.

### 1. Episodio prototípico: inseguridad ciudadana



### 2. Uso de lenguaje en código



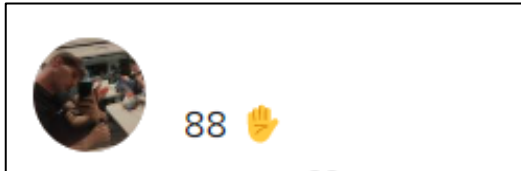
\*Se combinan caracteres numéricos y letras para componer una palabra. Ejemplo: el número 0 equivale a la letra "o" y el 3 equivale a la letra "e".

### 3. Discurso de odio contra personas africanas y afrodescendientes

Negros negros portaos bien o a recoger algodón si queréis vivir aquí.

\*se realiza una referencia a las plantaciones de algodón, elemento asociado a la esclavitud en el contexto histórico de Estados Unidos.

#### 4. Lenguaje en código con simbología nazi



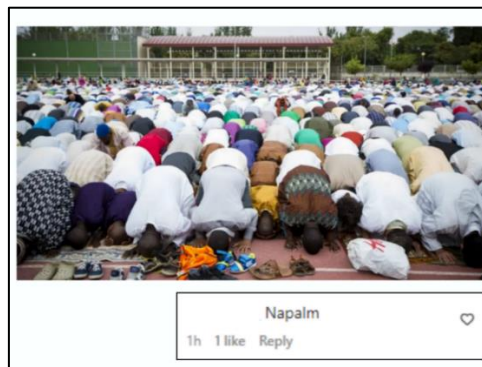
\* 88 significa 'HH' (Heil Hitler). Se acompaña del emoji de una mano saludando, que representa el saludo fascista.

#### 5. Mensaje construido con emojis

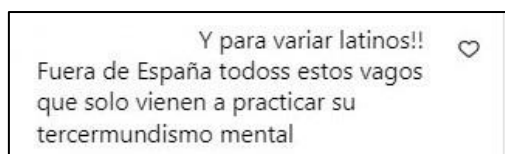


\*el mensaje constituye una incitación a la violencia contra las personas musulmanas. Observamos, entre otros, los elementos del cerdo (referencia al islam) junto con otros emojis, como la sangre, el cuchillo y la calavera.

#### 6. Episodio prototípico: evento religioso



#### 7. Discurso de odio contra personas latinoamericanas



## 8. Episodio prototípico: evento deportivo



\*deshumanización del futbolista Vinicius Jr en el contexto de los incidentes racistas producidos en varios partidos de La Liga Española de Fútbol contra el futbolista brasileño en 2023.

## 9. Discurso de odio contra niños, niñas y juventud no acompañada

Vienen a violar y robar y encima hay que gastar dinero en ellos, fuera menas de mierda de España

\*se presenta a los menores extranjeros no acompañados como una amenaza y se incita a su expulsión.

## 10. Antisemitismo

Judio = a jabón



\*el término “jabón” se emplea como referencia al Holocausto.

## 11. Episodio prototípico: llegada de pateras



Esta es toda la mierda que tenemos en España es bajar de las pateras y a otro día lla le dan la alluda y encima entra a las casas a robar esto es una vergüenza

## 12. Discurso de odio contra personas procedentes del norte de África / personas musulmanas

Yo pondría de moda CAZAR AL MORO en España. Dar de hostias a todos los **moros** que veáis por ahí. Que se vayan a su puto país. Por asesinos, violadores, ladrones, parásitos. Tenéis toda la razón del mundo, por cada hostión bien dado que les déis. Por ser de una SECTA SATÁNICA.

\*se utiliza la palabra “moro” sin distinguir entre religión y procedencia nacional. Se refiere tanto al origen nacional “a su puto país” como a la religión “secta satánica”.

## 13. Discurso antiinmigración



## 14. Episodio prototípico: políticas públicas

PORTADA INTERNACIONAL OPINIÓN ESPAÑA SOCIEDAD EL

**El Gobierno ultima la apertura de casi 11.000 plazas en cuarteles, hoteles y hostales para acoger a los migrantes llegados a Canarias**

MARÍA MARTÍN / MIGUEL GONZÁLEZ | Madrid

Migraciones instalará alojamientos en cuatro terrenos de Defensa en Madrid, Alcalá de Henares, Sevilla y Cartagena para unas 4.000 personas

Un bombazo ya

#### 15. Se promueve el descrédito en base a atributos personales del grupo



\*se hace uso de una narrativa en la que presenta a las mujeres musulmanas y/o marroquíes como unas “parásitas” (que se aprovechan supuestamente de las ayudas y de la Seguridad Social de España) por el único motivo de ser musulmanas y/o marroquíes.