



DISCURSO DE ODIO XENÓFOBO EN REDES SOCIALES A TRAVÉS DE LOS MENSAJES DIFUNDIDOS EN LA PLATAFORMA X

SANTA PALELLA STRACUZZI ¹, MARTA SÁNCHEZ ESPARZA ²

¹ EAE Business School, España

² Universidad Rey Juan Carlos (URCJ), España

PALABRAS CLAVE

X
Discurso de odio
Xenofobia
Medios digitales
Redes sociales
Representación social
Inmigrantes

RESUMEN

La red social X facilita la proliferación del discurso de odio. El proyecto Hatemedia (2021-2024) investiga la expansión de expresiones de odio en medios digitales en España, utilizando un Monitor de Odio basado en minería de datos. Se ha desarrollado una librería con cinco tipos de odio: general, político, sexual, misógino y xenóforo. Este estudio analiza mensajes xenófobos en X entre junio y noviembre de 2024, estudiando la representación del colectivo. Se concluye que estos mensajes construyen una narrativa estigmatizadora y que los medios juegan un rol significativo en su difusión, planteando cuestiones sobre responsabilidad editorial y políticas de moderación.

Recibido: 15/ 11 / 2024

Aceptado: 20/ 02 / 2025

1. Introducción

El auge de las redes sociales ha alterado significativamente la comunicación global, lo que permite un rápido intercambio de información y, al mismo tiempo, facilita la propagación del discurso de odio, incluida la retórica xenófoba. Este tipo de discurso tiene un importante impacto en la población, alimentando prejuicios y provocando un aumento de los delitos motivados por el odio. Además, supone la victimización de los individuos que forman parte de un colectivo, convertido en diana de dicho discurso, de lo que se derivan además daños psicológicos para estas personas (Aldamen, 2023).

La exposición al discurso de odio en línea puede reforzar estereotipos negativos y actitudes discriminatorias hacia determinados grupos sociales (Schäfer et al., 2024). Esto se agudiza en entornos con sentimientos nacionalistas preexistentes, donde las redes sociales actúan además como catalizadores para la coordinación y la amplificación de acciones xenófobas. En Rusia, por ejemplo, el aumento de la penetración de las redes sociales se correlacionó con el aumento de los delitos de odio étnico en ciudades con fuertes sentimientos nacionalistas (Bursztyn et al., 2019; Bursztyn et al., 2020). En otro extremo del planeta, en Sudáfrica, algunas investigaciones han verificado cómo las redes sociales reflejan y pueden predecir los incidentes xenófobos, lo que indica su papel a la hora de moldear el sentimiento público y, posiblemente, de incitar a la violencia (Raborife et al., 2024).

La incitación al odio xenófobo en redes sociales no encuentra una respuesta coherente en las legislaciones de los diferentes países, ni tampoco es objeto claro de la autorregulación de las propias plataformas, lo que deja a muchas personas en una situación de extrema vulnerabilidad (Gelashvili, 2018). La falta de una colaboración internacional eficaz y las complejidades técnicas de las plataformas de redes sociales dificultan el desarrollo de estrategias integrales para combatir el discurso de odio xenófobo.

No han faltado los intentos por atajar este problema creciente. En el año 2019, la Organización de las Naciones Unidas publicó un marco de estrategia y plan de acción sobre el discurso de odio. En España, el Ministerio de Inclusión, Seguridad Social y Migraciones emitió en 2017 un Protocolo y sistema de indicadores para la detección del discurso de odio en redes sociales. Observatorios pertenecientes al Gobierno de España, como el Observatorio Español del Racismo y la Xenofobia (OBERAXE, 2017a, 2017b) elaboraron un dossier que reflejaba la evolución del racismo y otras formas de intolerancia en España con las actitudes de la población española hacia la población migrante en las redes sociales. Este observatorio lleva a cabo el Informe Anual de Monitorización del Discurso de Odio en Redes Sociales, en cuya última edición se destaca un aumento preocupante de mensajes de odio dirigidos a grupos vulnerables, principalmente inmigrantes, refugiados y comunidades étnicas específicas (OBERAXE, 2023). El racismo y la xenofobia representaron un porcentaje significativo del total de mensajes monitorizados, siendo las redes sociales una plataforma clave para su propagación.

Entre los hallazgos recogidos por dicho informe, se resalta que el 43,5% de los delitos de odio en España en 2022 tuvieron motivaciones racistas o xenófobas, con un impacto especialmente alto en las comunidades inmigrantes. Las plataformas más involucradas en la difusión de mensajes ligados al discurso de odio incluyen Facebook, Instagram, TikTok, YouTube y X (anteriormente Twitter), con variaciones en la prevalencia según la red social (OBERAXE, 2024).

El informe subraya la importancia de herramientas legislativas, como la Ley de Servicios Digitales de la UE, para garantizar la moderación eficaz del contenido de odio. Además, la difusión de noticias falsas y desinformación fue identificada como un mecanismo común para fomentar el odio, lo que demanda contramedidas tecnológicas y educativas.

El presente trabajo pretende contribuir a la visibilización de este problema, analizando los mensajes xenófobos publicados en la red social X entre junio y noviembre de 2024 extraídos en el marco del proceso de minería de datos del proyecto Hatemedia, con el objetivo de identificar las características de la representación negativa hacia los colectivos inmigrantes y extranjeros en el discurso de odio vertido en la citada red social.

2. Revisión de la literatura

2.1. El discurso de odio

El discurso de odio (*hate speech*) engloba aquellas declaraciones o comunicaciones que contienen un mensaje denigrante, hostil o difamatorio contra ciertos colectivos (Fuentes Osorio, 2024). Naciones Unidas lo define como:

cualquier forma de comunicación de palabra, por escrito o a través del comportamiento, que sea un ataque o utilice lenguaje peyorativo o discriminatorio en relación con una persona o un grupo sobre la base de quiénes son o, en otras palabras, debido a su religión, origen étnico, nacionalidad, raza, color, ascendencia, género u otro factor de identidad. (ONU, 2019)

Según el Consejo de Europa (1997), para que un discurso sea considerado discurso de odio, debe difundir, incitar, promover o justificar actitudes relacionadas con el racismo, la xenofobia, el antisemitismo y otras formas de intolerancia. Asimismo, la Comisión Europea contra el Racismo y la Intolerancia, en su Recomendación General N.º 15, señala que el odio puede estar motivado por factores como raza, color, ascendencia, origen nacional o étnico, entre otras características personales (2016). Las definiciones oficiales de discurso de odio destacan especialmente la discriminación racista y xenófoba como principales causas de rechazo y hostilidad. Por otro lado, el Ministerio del Interior de España llegó a diferenciar hasta 11 categorías de discriminación que afectan a colectivos vulnerables, siendo el racismo y la xenofobia las principales.

Algunos autores, como Valle de Frutos (2024), distinguen entre el discurso de odio y el discurso ofensivo, una subcategoría más leve identificada como toda comunicación en la que se lance un mensaje despectivo hacia una persona o grupo debido a factores culturales como la religión, etnia, nacionalidad, color, género u otro factor identitario. La principal diferencia entre el discurso de odio y el ofensivo radicaría, según este autor, en que mientras que el discurso de odio puede incitar a la violencia física, que lleva a una discriminación de forma directa, el discurso ofensivo realizaría una verbalización y un rechazo sutil, haciendo referencia explícita hacia elementos de la cultura que se rechaza. Se trata de una forma de desacreditar al colectivo, sin insultar ni incitar a la violencia o a otros actos discriminatorios, empleando para ello parámetros de moralidad.

En cualquier caso, este discurso supone la definición y categorización de estos colectivos en función de características como la raza, la religión, el género o la orientación sexual, a menudo mediante el uso de noticias falsas o engañosas (Hyewook, 2023). Este discurso puede manifestarse de forma oral, escrita o visual y se produce en diferentes contextos, incluidos Internet y los medios de comunicación, y, por supuesto, las redes sociales, donde en la actualidad se dirige la conversación social. Investigaciones recientes, como las de Arcila-Calderón et al. (2022b) y Gautam et al. (2024) exponen cómo las tecnologías de inteligencia artificial se están empleando ya para detectar y enfrentar este discurso discriminatorio en Internet, mediante el entrenamiento de modelos de aprendizaje automático.

2.2. Plataformas sociales y su rol en la proliferación de discurso xenófobo

Las plataformas sociales son redes y entornos digitales que facilitan la interacción social, la comunicación y el intercambio de información entre los usuarios. Pese a que algunas de estas plataformas como Facebook ya articulaban la conversación pública desde antes de 2010, alcanzando en ese año los 500 millones de usuarios (Bruneel et al., 2013), estas redes ganaron protagonismo tras la pandemia de la COVID-19, ante la radical demanda de contacto personal, trabajo en remoto y digitalización de la enseñanza (Kaur, 2023). Su relevancia en la interacción social ha hecho que las empresas las utilicen como una herramienta fundamental para el marketing, la segmentación de sus clientes y la comercialización de sus productos.

El número de usuarios de plataformas sociales se ha multiplicado en las últimas décadas, y con ello su poder e influencia. En 2024, se superaron por primera vez los cinco billones de usuarios de redes sociales en todo el planeta (We are Social, 2024). Este colosal volumen las ha convertido en un vehículo de articulación de los movimientos sociales y del discurso público, como se ha visto en eventos como la Primavera Árabe y Black Lives Matter (Forrest & Wexler, 2023). Estos y otros acontecimientos han

demostrado su poder a la hora de moldear las percepciones y los comportamientos sociales. Su influencia se extiende incluso a la forma en la que los jóvenes interactúan unos con otros, así como en sus círculos familiares y sociales (Baha, 2022).

Estas herramientas pueden contribuir significativamente al bien social, al facilitar la construcción de comunidades, empoderar a las personas, promover mensajes de interés público, impulsar el crecimiento económico y promover la justicia y el consumo responsable, entre otras muchas cuestiones (Chang & Zhang, 2024). Sin embargo, el acceso masivo a las mismas también ha servido como un vehículo para la proliferación de desinformación y discursos de odio, incluidos los mensajes xenófobos (Forrest & Wexler, 2023). En este contexto, la red social X (anteriormente conocida como Twitter) y otras plataformas similares han desempeñado un papel central, debido a su diseño enfocado en la viralización y al alto volumen de publicaciones diarias que albergan.

La estructura y funcionamiento de la red social X propician un entorno donde los mensajes xenófobos pueden amplificarse rápidamente. La brevedad de las publicaciones, combinada con algoritmos que priorizan la interacción y el contenido polémico, favorecen la circulación de mensajes que apelan a emociones intensas, como el miedo o la ira, características inherentes al discurso de odio. Además, el anonimato parcial que ofrecen estas plataformas reduce las barreras sociales para emitir comentarios discriminatorios, fomentando la proliferación de contenido hostil hacia grupos minoritarios, incluidos los inmigrantes y refugiados.

Algunos estudios recientes sobre la red social X han destacado cómo los picos de actividad xenófoba en la plataforma están correlacionados con eventos sociales o políticos relevantes en el mundo real, ligados de alguna forma al fenómeno de la inmigración o a problemas generados por extranjeros (Santana dos Santos et al., 2022). En este marco, las redes sociales sirven tanto de barómetro como de catalizador de los sentimientos xenófobos, especialmente durante eventos políticos o mundiales importantes. Así, algunas investigaciones han detectado que aproximadamente el 50% de los tuits analizados durante la pandemia de la COVID-19 representaban acoso xenófobo, y una parte importante estaba dirigida a personas percibidas como asiáticas (Dhungana Sainju et al., 2022). En este sentido, trabajos recientes como el de Umarova et al. (2024) establecen las mismas conexiones, tras analizar más de 7.000 tuits relacionados con las políticas de inmigración de EEUU, y ligados a cuentas de personajes influyentes en la esfera pública.

La xenofobia encuentra en X un ecosistema donde expandirse con facilidad. Nasuto y Rowe (2024) llegaron a analizar 220.870 mensajes xenófobos en X (antes Twitter), en torno a la conversación sobre inmigración en el Reino Unido, revelando un alto grado de polarización con respecto a la inmigración en esta red social, donde el sentimiento antiinmigración es más activo y se difunde 1,66 veces más rápido que los mensajes a favor de la inmigración. Esto apunta hacia una presencia significativa de xenofobia en las discusiones relacionadas con los eventos de inmigración.

Durante los períodos en los que se producen eventos relativos a la inmigración, el volumen de publicaciones que demonizan a colectivos extranjeros aumenta significativamente, a menudo acompañado de desinformación y teorías conspirativas. Esto subraya el papel crucial de estas plataformas como espacios donde las tensiones sociales se reflejan y amplifican.

Diversas investigaciones han analizado la relación entre el diseño de plataformas digitales y la amplificación de discursos de odio. El diseño de estas plataformas permite la rápida difusión de contenido que incita al odio, creando cámaras de eco que refuerzan los puntos de vista extremistas (Weber et al., 2023). Esto sucede porque los algoritmos de recomendación, diseñados para maximizar el tiempo de uso y la interacción del usuario, tienden a favorecer contenido polarizador, ya que éste genera mayor *engagement*. Esta lógica algorítmica prioriza la exposición a mensajes controvertidos, incluyendo aquellos cargados de xenofobia. Así sucede con plataformas digitales como X (Twitter), Instagram y YouTube, cuyos algoritmos dan prioridad a la interacción por encima de la moderación del contenido (Dutta, 2024).

Además, las herramientas como los retuits, compartidos o respuestas rápidas permiten la diseminación instantánea de mensajes ofensivos, lo que refuerza la percepción de que las opiniones discriminatorias son más comunes o aceptadas de lo que realmente son.

La virulencia o polarización que se percibe en los mensajes de las redes sociales, sin embargo, se atenúa cuando la población tiene la posibilidad de convivir en la vida real con colectivos de inmigrantes. Al menos así lo apuntan algunos trabajos como los de Arcila et al. (2022a), que han cotejado el nivel de discurso de odio en redes sociales con la mayor o menor presencia de colectivos afectados en los lugares

de los que emanan los mensajes, mediante herramientas de geolocalización. De este modo, ha podido constatar que, en países con mayor presencia de inmigrantes, el volumen de mensajes xenófobos en redes sociales como X disminuye, lo que apunta a una mayor importancia de las mediaciones físicas en el propio entorno a la hora de la construcción de las percepciones sociales sobre un colectivo (Arcila et al., 2022a).

3. Objetivos

Esta investigación tiene como objetivo principal analizar los mensajes xenófobos difundidos en la red social de la plataforma X, abarcando el período de junio a noviembre de 2024.

Para profundizar en esta temática, se establecieron los siguientes objetivos específicos:

- Determinar el nivel de odio de los mensajes xenófobos difundidos en la red social de la plataforma X entre junio y noviembre de 2024.
- Caracterizar los textos de los mensajes de representación negativa que se hace del colectivo extranjero o inmigrante difundidos en la red social de la plataforma X entre junio y noviembre de 2024.
- Conocer el impacto y repercusión que han tenido los mensajes de odio difundidos en la red social de la plataforma X entre junio y noviembre de 2024 en términos de número de *likes*, *retuits*.

4. Metodología

La metodología empleada se enmarca en un diseño mixto o cuali-cuantitativo, que combina técnicas cuantitativas y cualitativas para analizar y clasificar el discurso de odio en redes sociales, específicamente dentro de la plataforma X en el contexto español. Esta investigación es descriptiva y exploratoria, dado que busca describir las características de un fenómeno en particular, en este caso, los mensajes xenófobos y de odio, y explorar las tendencias y patrones en los datos recopilados en 2024. Explora patrones y describe las características del discurso de odio a través de análisis estadísticos y de contenido.

Se utilizó para analizar los datos la estadística descriptiva (frecuencias y porcentajes). Se analizaron los discursos de 5 medios de comunicación en la Plataforma X: ABC, Mundo, El País, La Vanguardia, 20Minutos, desde el mes de junio a noviembre de 2024.

El estudio se basa en un conjunto de datos obtenido del proyecto Hatemedia (2021-2024), que utiliza un Monitor de Odio para clasificar mensajes según cinco niveles de discurso de odio (de 0 a 5). Este monitor emplea un enfoque de minería de datos avanzado, analizando grandes volúmenes de publicaciones en redes sociales mediante algoritmos de aprendizaje automático. La metodología del proyecto se estructura en varias fases:

- **Recolección de datos:** Se recopilan mensajes publicados en redes sociales asociadas a medios informativos en España. Esto incluye menciones y comentarios relacionados con noticias.
- **Clasificación automática:** Los mensajes se procesan utilizando un algoritmo que asigna una etiqueta de nivel de odio basada en el lenguaje y el contexto del mensaje.
- **Validación humana:** Un equipo de expertos revisa y valida una muestra de los datos clasificados, asegurando la precisión del algoritmo.
- **Análisis exploratorio:** Se examinan los patrones y tendencias en los datos, como la distribución mensual de mensajes o la relación con eventos mediáticos.

El monitor también utiliza una librería de expresiones asociadas al odio, que incluye hasta cinco categorías: general, político, sexual, misógino y xenófobo. Esto permite identificar y categorizar diferentes formas de discurso de odio con un alto grado de precisión. El trabajo que lleva a cabo el algoritmo del monitor de odio incluye:

- Frecuencia mensual de mensajes de odio, distribuidos por categoría.
- Clasificación de insultos y términos despectivos.

- Análisis por medios de comunicación referenciados.
- Distribución y retuits según niveles de odio.

Un aspecto de extremo interés es la determinación de los niveles de odio por parte del algoritmo entrenado en el Proyecto Hatemedia. En este aspecto se elaboró la publicación Manual para el etiquetado de mensajes de odio (De Lucas et al., 2022), que clasificaba estos discursos en niveles de intensidad del 0 al 5, con el objetivo de analizar y categorizar su impacto y gravedad. Esta metodología asegura una evaluación objetiva, basada en criterios lingüísticos y contextuales, y permite identificar las consecuencias potenciales del discurso analizado.

4.1. Niveles de odio

A continuación, se expone el contenido de cada uno de los 5 niveles de odio utilizados para el etiquetado de los mensajes analizados por el Monitor de Odio:

Nivel 0: Sin odio

El nivel 0 se refiere a aquellos mensajes que no contienen expresiones de odio propiamente dichas, aunque podrían contribuir indirectamente a la generación de discursos de odio. Estos mensajes suelen emplear categorías objetivas o neutras para describir a un grupo social. Por ejemplo:

“Un grupo de marroquíes ha llegado a la frontera”. En este caso, la mención de la nacionalidad no tiene intención despectiva, pero su uso como característica principal podría fomentar prejuicios si se enmarca en un contexto de información sesgada o sensacionalista.

Nivel 1: Estigmatización indirecta

En este nivel, los mensajes no contienen insultos ni violencia verbal evidente, pero presentan hechos o informaciones que pueden reforzar estigmas hacia un grupo social. Estos discursos suelen estar relacionados con la generalización de conductas negativas. Por ejemplo:

“Los inmigrantes se aprovechan de las ayudas sociales”. Aunque no hay un ataque directo, la afirmación generaliza de manera negativa la intención de un colectivo, promoviendo una percepción desfavorable.

Nivel 2: Desprecio manifiesto

Los mensajes clasificados en el nivel 2 atribuyen de manera más directa intenciones negativas a un grupo, a menudo utilizando expresiones abusivas o prejuicios. En este nivel no hay insultos explícitos, pero se construye un discurso de rechazo hacia el colectivo. Por ejemplo:

“Los musulmanes solo vienen a imponer su religión”. Este tipo de mensajes transmite desprecio hacia los integrantes de un grupo, utilizando afirmaciones que buscan desacreditarlos o minimizar su legitimidad.

Nivel 3: Violencia verbal

El nivel 3 abarca expresiones en las que la violencia verbal es evidente, con insultos, humillaciones o desprecios dirigidos a una o varias personas por su pertenencia a un grupo social. Este nivel incluye también intensificadores lingüísticos, como “de mierda” o “pedazo de”, que aumentan la carga negativa del mensaje. Ejemplo:

“Marruecos solo exporta marroquíes de mierda que no quieren trabajar”. Aquí, el uso de calificativos peyorativos y generalizaciones despectivas ejemplifica claramente el odio hacia un grupo.

Nivel 4: Amenazas veladas e incitación implícita

En el nivel 4, los mensajes contienen amenazas indirectas, implícitas o veladas hacia un grupo social o sus miembros. Estas expresiones pueden incluir intimidaciones, alegrías explícitas ante un daño ya consumado o incitaciones no directas a la violencia. Aunque no siempre se da una incitación clara y explícita, el mensaje busca generar miedo, humillación o un ambiente hostil hacia el colectivo objetivo.

Además, pueden aparecer expresiones que manifiesten un deseo de daño hacia el grupo en cuestión, sin llegar a implicar al emisor directamente en las acciones deseadas. Por ejemplo:

"Ten cuidado cuando vayas por la calle; no todos los fachas llegan a casa enteros". Este mensaje utiliza una amenaza implícita que sugiere un riesgo físico para el receptor, sin una incitación directa.

Otro ejemplo sería:

"Me alegra que se haya muerto ese torero; la tauromaquia debería desaparecer junto con ellos". Aquí, aunque no hay una amenaza ni incitación directa, la celebración de la muerte refuerza el ambiente de odio hacia un grupo social vinculado a una actividad cultural.

Nivel 5: Incitación directa a la violencia

El nivel 5 representa el punto más extremo del discurso de odio y está caracterizado por llamamientos explícitos a la violencia física, amenazas claras y directas, o deseos explícitos de que un individuo o grupo social sufra daños graves. Estas expresiones pueden incluir referencias a acciones como matar, golpear, violar o agredir físicamente a los miembros de un colectivo, o atacar instituciones, símbolos o propiedades relacionadas con ellos. Por ejemplo:

"Deberíamos matar a todos esos musulmanes; están destruyendo nuestro país".

Otro ejemplo sería:

"Hay que quemar todas las mezquitas; no son más que nidos de terroristas".

Estos mensajes no solo incitan abiertamente a la violencia, sino que también legitiman acciones agresivas contra los grupos señalados, constituyendo un peligro real de derivar en hechos violentos en la vida real.

Ejemplos adicionales

La metodología también contempla situaciones complejas, como el uso de humor o sarcasmo, que podrían modificar la percepción del odio. Por ejemplo:

"Esos panchitos parecen Gremlins; si se mojan, se reproducen". En este caso, aunque el humor es evidente, el mensaje sigue reforzando estereotipos xenófobos.

La clasificación en niveles expuesta permite diferenciar entre mensajes que fomentan odio de forma directa, aquellos que lo insinúan o contribuyen indirectamente a su proliferación. Este enfoque proporciona un marco claro y sistemático para el análisis del discurso de odio en plataformas digitales y asegura que el etiquetado sea coherente y reproducible. En el caso del presente trabajo, los mensajes extraídos de la red social X entre junio y septiembre de 2024 oscilan entre los niveles 0 y 3, sin encontrarse ninguno de las categorías 4 y 5, que supongan invitaciones abiertas a la violencia.

5. Discusión de resultados

A los fines de dar respuesta y conocer la determinación del nivel de odio de los mensajes xenófobos, caracterizar los textos de representación negativa hacia el colectivo extranjero o inmigrante, y evaluar el impacto de estos mensajes en términos de su recepción y propagación a través de *likes* y *retuits* se presentan los siguientes resultados:

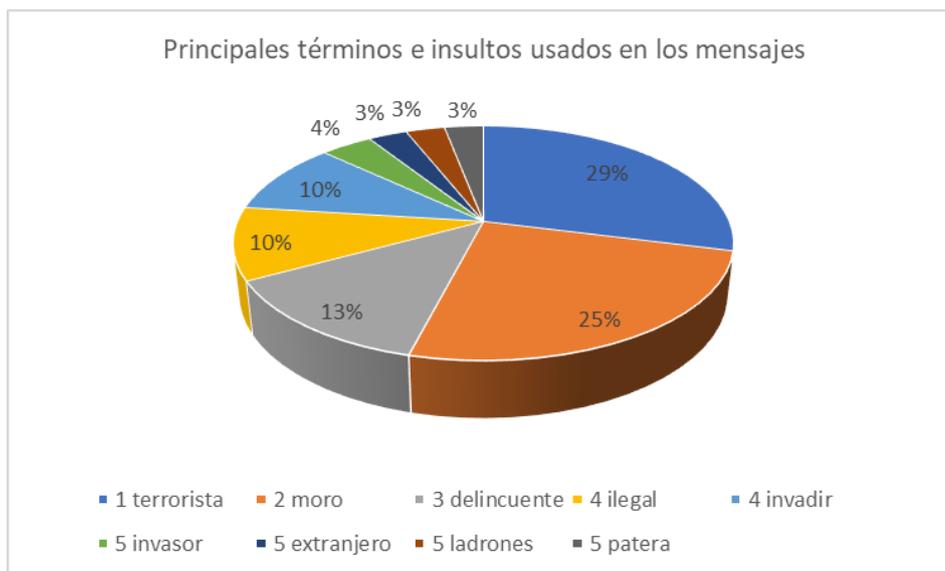
La siguiente tabla 1 y gráfico 1 presentan los términos más frecuentes identificados en los mensajes xenófobos, junto con el número de menciones e interacciones (*retuits* y respuestas).

Tabla 1: Principales términos e insultos usados en los mensajes

Clasificación	Término	Frecuencia	Porcentaje
1	Terrorista	9	29%
2	Moro	8	25%
3	Delincuente	4	13%
4	Ilegal	3	10%
4	Invadir	3	10%
5	Invasor	1	4%
5	Extranjero	1	3%
5	Ladrones	1	3%
5	Patera	1	3%

Fuente: Elaboración propia, 2024.

Gráfico 1. Principales términos e insultos usados en los mensajes



Fuente: Elaboración propia, 2024.

Con respecto a los términos más frecuentes identificados en los mensajes xenófobos, junto con el número de menciones e interacciones y porcentajes de las frecuencias (retuits y respuestas), el gráfico 1 y la tabla muestran una predominancia de términos como "terrorista" y "moro", que constituyen el 29% y el 25% del total respectivamente, lo cual indica una tendencia notable hacia la estigmatización y la categorización negativa en el discurso examinado.

Por otro lado, términos como "delincuente" y "ilegal/invadir", con 13% y 10% respectivamente, aunque menos predominantes, siguen siendo significativos y contribuyen a una narrativa de criminalización y deshumanización. Estos hallazgos enfatizan cómo el lenguaje utilizado en los medios de comunicación puede influir y reforzar las políticas y percepciones públicas hacia los grupos minoritarios, promoviendo una visión sesgada y a menudo negativa.

El análisis realizado sobre los mensajes xenófobos publicados en la red social X entre abril y noviembre de 2024 muestra que ciertos términos despectivos, como 'terrorista', 'moro' y 'delincuente',

son recurrentes y forman parte de una narrativa xenófoba dirigida hacia colectivos inmigrantes y extranjeros. Además, el impacto de estos mensajes, medido a través de interacciones como retuits y menciones, es bajo, pero el volumen de publicaciones refuerza su visibilidad.

Lo anterior resulta coincidente con el resultado de estudios previos en torno a los mensajes desinformativos y estigmatizantes sobre colectivo inmigrante en internet (Sánchez Esparza et al., 2023), donde se constataba la repetición de diferentes tópicos. Los principales inciden en la asociación de la inmigración con la carga económica causada al país anfitrión -donde los inmigrantes estarían quitando beneficios sociales y empleo a los nacionales-, en la amenaza que suponen a la seguridad -culpando al colectivo migrante de la inseguridad en todas las áreas, incluido el terrorismo y la delincuencia-, y en el supuesto peligro para la pervivencia de la identidad de la propia cultura.

En cuanto a los medios de comunicación, se identificó que 'El Mundo' y 'El País' son los medios más referenciados en los mensajes que contienen insultos. Es importante continuar monitoreando estos mensajes para comprender mejor las dinámicas del discurso de odio en redes sociales y su relación con la agenda mediática y los eventos sociopolíticos.

El gráfico 2 muestra el número total de insultos identificados en los mensajes analizados, distribuidos por medio de comunicación. Los medios con mayor cantidad de insultos son 'El Mundo' y 'El País', seguidos por 'ABC'.

Gráfico 2. Insultos distribuidos por medio de comunicación



Fuente: Elaboración propia, 2024.

Conforme a los objetivos del estudio analizar los mensajes xenófobos y determinar el nivel de odio en dichos mensajes, se observa que los medios "Mundo" y "País" presentan las cifras más elevadas, con 11 y 10 insultos respectivamente. Esto podría indicar una mayor frecuencia o intensidad de discursos xenófobos transmitidos a través de estos medios en comparación con otros como "ABC" (5 insultos), "20Minutos" (3 insultos) y "Vanguardia" (1 insulto).

Tabla 2. Texto de los mensajes con mayor nivel de odio (Nivel 3)

Texto del mensaje	Nivel de odio	Medio relacionado
@abc_es @mikelayestaran En morolándia nunca habrá paz. Es lo que hay.	3	ABC - Nov 24
@elmundoes Espero que saquen de España a todos los ucranianos cobardes que nos comimos	3	El Mundo - Nov 24
@elmundoes Más llorones. Si no les gusta donde viven, lo tienen fácil. Vender y a otra parte	3	El Mundo - Nov 24
@elmundoes a mi las narcolanchas me la pela bastante, dan más problemas los saltarines de vallas	3	El Mundo - Nov 24

@elmundoes Por eso traen tanta “mano de obra”, para hacer qué?	3	El Mundo - Nov 24
@elmundoes Entre ellos que se maten.	3	El Mundo - Nov 24
@elmundoes Nos abandonan, roban y también hay que aguantarlos en X 🙄	3	El Mundo - Nov 24
@elmundoes Desagradecidos. Merecen arder en nombre de nuestros valores europeos.	3	El Mundo - Nov 24
@el_pais Cosas de libios...	3	El País - Nov 24
@20m Que le den al cerdo antiespañol que quiere que haya una invasión y se extinga nuestro pueblo	3	20Minutos - Nov 24
@LaVanguardia Desalojen todo el raval de todos lo anti occidente	3	La Vanguardia - Oct 24
@elmundoes La cara que tiene de traidor repugnante. Ojalá lo metan en el talego, pero en el módulo donde están los moros	3	El Mundo - Oct 24
@elmundoes Hay que deportar masivamente, retirar nacionalidades regaladas a delincuentes reincidentes, cerrar a cal y canto las fronteras	3	El Mundo - Oct 24
@elmundoes Judios sois vosotros, usureros de mierda.	3	El Mundo - Oct 24
@elmundoes Judío tenía que ser el cabrón	3	El Mundo - Oct 24
@elmundoes Pues que no hubiesen venido, que no huían de ninguna guerra.A demás menos paguitas.	3	El Mundo - Oct 24
@20m A su puta casa todos ellos.	3	El Mundo - Sep 24
@abc_es Vaya... y yo q creía q iban a ser noruegos y resulta q son marroquíes. Estoy q no quepo en mi asombro.	3	ABC - Sep 24
@abc_es Nos traen los bichos aposta.	3	ABC - Agost 24
@20m Inmigrantes o narcos?	3	20Minutos - Agost 24
@el_pais Que los devuelvan a su país y se dejen de mierdas.	3	ElPaís - Agost 24
@el_pais Que los devuelvan a Marruecos, allí no hay guerras y tienen a su familia	3	ElPaís - Agost 24
@el_pais INVASORES.Se dice INVASORES.INVASORES	3	ElPaís - Agost 24
@elmundoes Nacionalidades de los delincuentes? No hace falta, ya lo sabemos.	3	ElMundo - Agost 24
@elmundoes Es otras palabras: hay 550 delitos al día que quedan impunes en Madrid. Solución del establishment: importar delincuentes de las recientemente vaciadas cárceles nor africanas...	3	ElMundo - Agost 24
@elmundoes Ya somos Africa https://t.co/4uSs48Rxp9	3	ElMundo - Agost 24
@20m Que aprovechen si sirven para CONTROLAR la basura que viene ilegalmente a España	3	20Minutos - Jul 24
@20m Caramba, cómo han cambiado los camiones de la basura.	3	20Minutos - Jul 24
@elmundoes La imagen de España en el mundo moros y más moros jajaja	3	ElMundo - Jul 24
@elmundoes Eso no quita para que se les escapen balas y maten a gente que está cenando tranquilamente. ESTO NO DEBE PASAR.DEPORTACIÓN YA.	3	ElMundo - Jul 24
@elmundoes Pues como se le haya robado a un subsahariano supongo que ya estará en el fondo del mar	3	ElMundo - Jul 24
@elmundoes Habéis convertido Francia en un vertedero, por mucho dinero que tengáis, de África no llega más que mierda	3	ElMundo - Jun 14

@elmundoes Argelinos, cameruneses, marroquíes, senegalés, malienses... los invasores de Francia quieren que la africanización de Europa siga adelante	3	ElMundo - Jun 14
@elmundoes Europa es un estercolero, primero habría que limpiarlo y luego pues ya se vería.	3	ElMundo - Jun 14
@elmundoes Hay mucha chusma en España	3	ElMundo - Jun 14

Fuente: Elaboración propia, 2024.

Como se observa en la tabla 2, el análisis de los mensajes publicados en la plataforma X clasificados como nivel 3 de odio por el Proyecto Hatemedia, muestra un patrón de hostilidad explícita en la comunicación online, especialmente concentrado en ciertos medios. En particular, "El Mundo" destaca por la repetida aparición de mensajes agresivamente xenófobos y discriminatorios, lo cual podría apuntar hacia una audiencia que responde activamente a este tipo de retórica.

Por ejemplo, varios mensajes de "El Mundo" en noviembre de 2024, como "En morolándia nunca habrá paz. Es lo que hay." y "Espero que saquen de España a todos los ucranianos cobardes que nos comimos", no solo deshumanizan a los grupos mencionados, sino que también promueven una narrativa de exclusión y hostilidad. Estos mensajes comparten un desprecio manifiesto hacia los inmigrantes y minorías, utilizando términos peyorativos y un lenguaje que incita al rechazo y la violencia verbal.

Comparativamente, otros medios como "El País" y "La Vanguardia" también presentan mensajes de odio, pero en una cantidad significativamente menor. Un ejemplo de "El País" sería "Cosas de libios...", que, aunque menos explícito que los mensajes de "El Mundo", igualmente contribuye a estereotipos y percepciones negativas. Esto sugiere que, aunque el discurso de odio está presente en varios medios, la intensidad y la frecuencia varían considerablemente.

Estas diferencias en la cantidad y naturaleza de los mensajes podrían reflejar variaciones en las políticas editoriales o en las estrategias de moderación de los comentarios de cada medio. Este patrón subraya la importancia de que los medios de comunicación implementen prácticas de moderación más efectivas y desarrollen políticas claras contra la diseminación de discursos de odio, para mitigar el impacto negativo de estos mensajes en la cohesión social y el respeto hacia la diversidad.

La tabla 3 muestra el número total de mensajes y retuits clasificados según los niveles de odio (0 a 3), donde el nivel 0 representa discurso apenas perceptible y el nivel 3 incluye expresiones explícitas de odio.

Tabla 3. Mensajes y retuits por nivel de odio

Nivel de odio	Número de mensajes	Número de retuits
0	2906	158
1	395	6
2	251	13
3	955	38

Fuente: Elaboración propia, 2024.

En la tabla 3 vemos la distribución de mensajes y sus retuits según el nivel de odio, desde junio a noviembre de 2024, en la red social X. En el siguiente desglose por niveles se evidencia una visión clara de la dinámica de la interacción del usuario con el discurso de odio en la plataforma.

Con respecto al nivel 0 (sin odio) que incluye mensajes neutros o no ofensivos, muestra la cantidad más alta de mensajes, con 2906, y un número relativamente alto de retuits, 158. Esto indica que, aunque estos mensajes son prevalentes, su capacidad de generar una reacción o viralización es moderada, sugiriendo que los usuarios están más dispuestos a compartir contenidos que no contienen odio explícito.

En el nivel 1 (estigmatización indirecta) son menos frecuentes, con solo 395, el número de retuits es excepcionalmente bajo (6). Esto podría interpretarse como una menor tendencia de los usuarios a interactuar con contenidos que, aunque no son abiertamente agresivos, pueden contener insinuaciones o generalizaciones negativas.

Con relación al nivel 2 (desprecio manifiesto) con 251 mensajes, también recibe una respuesta baja en términos de retuits (13), lo que puede reflejar una resistencia de la audiencia a difundir mensajes

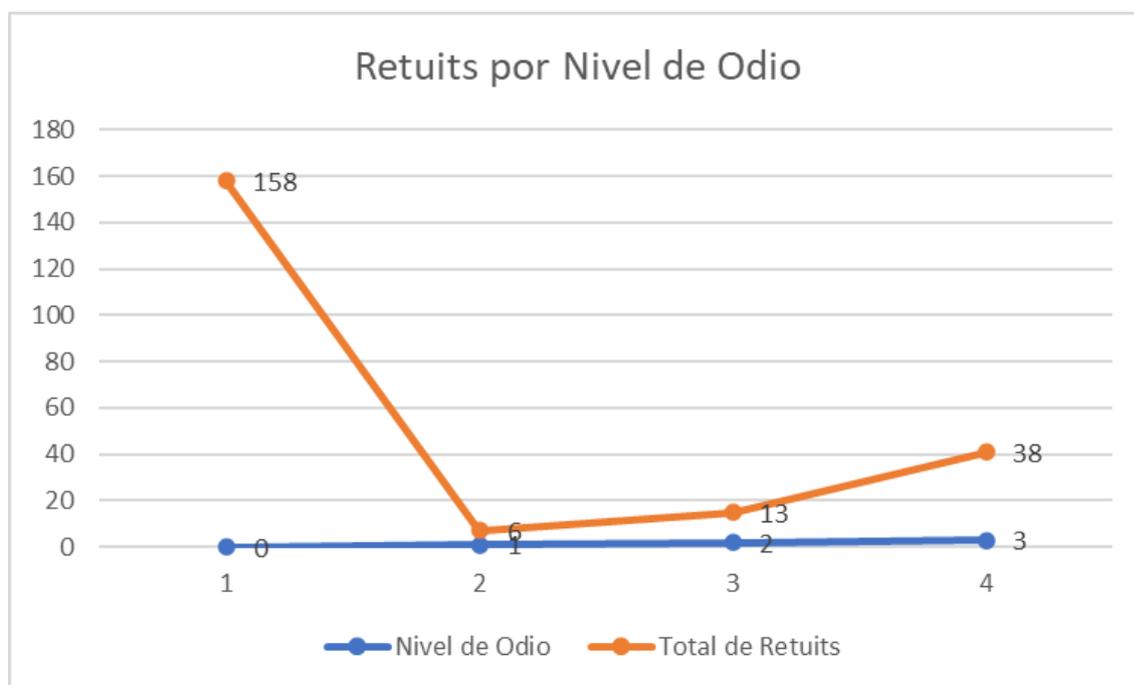
que expresan desprecio o rechazo hacia grupos o individuos, a pesar de la presencia de un discurso negativo más claro.

El nivel 3 (violencia verbal) es el segundo nivel más alto en cuanto a la cantidad de mensajes (955), pero con solo 38 retuits, sugiere que, aunque estos mensajes de odio explícito son numerosos, la comunidad no los amplifica con frecuencia. Esta es una señal alentadora de que el contenido más extremadamente negativo no es ampliamente aceptado o promovido por los usuarios.

La distribución de los mensajes y los retuits según los niveles de odio refleja una posible conciencia o rechazo por parte de la comunidad de usuarios hacia el discurso de odio más explícito y agresivo. Sin embargo, el número elevado de mensajes de odio, especialmente en los niveles 2 y 3, subraya la necesidad de estrategias efectivas de moderación y educación digital para reducir la presencia y el impacto de este tipo de contenido en las plataformas sociales.

El gráfico 3 compara los retuits generados por mensajes en los niveles 0, 1, 2 y 3 de discurso de odio. Los mensajes de nivel 0, aunque menos explícitos, generan más retuits, mientras que los de nivel 3, más explícitos, generan menos interacción. En tal sentido, observamos que los mensajes neutros (Nivel 0) tienen la mayor cantidad de retuits, lo que podría indicar una preferencia por compartir contenido que no contiene odio explícito. En contraste, los mensajes con un grado de odio más intenso (Nivel 3) también son compartidos, pero en menor medida.

Gráfico 3. Retuits por nivel de odio



Fuente: Elaboración propia, 2024.

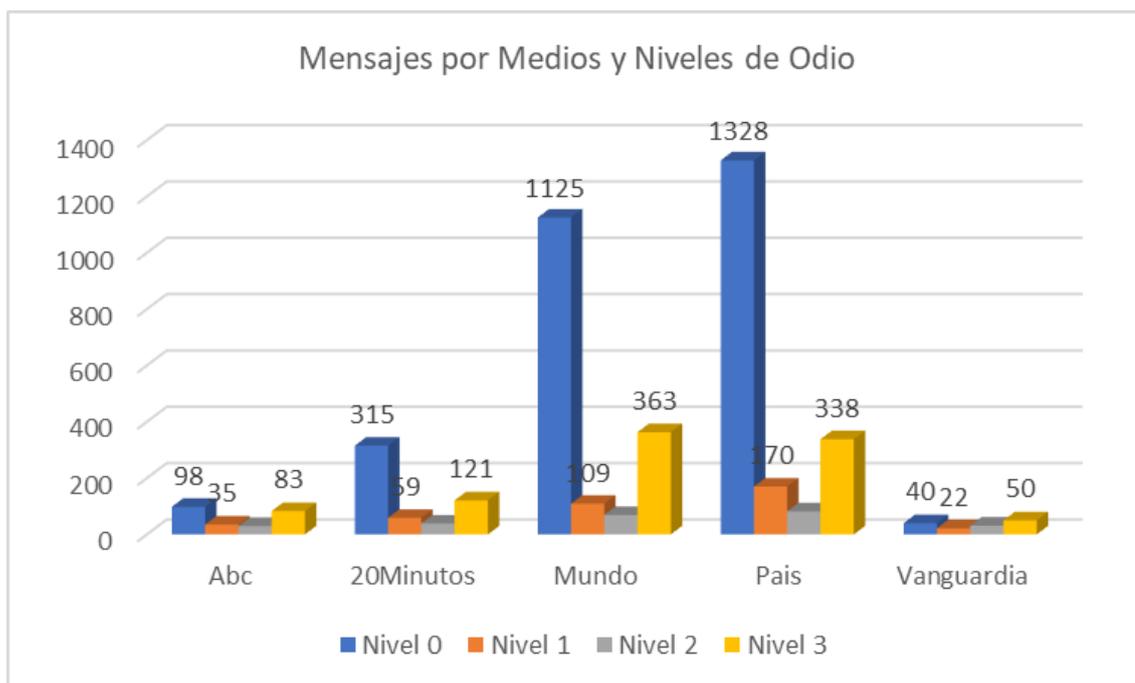
La siguiente tabla 4, y gráfico 4 combina los niveles de odio (0 a 3) y los medios de comunicación, mostrando el número de mensajes clasificados en cada nivel y para cada medio.

Tabla 4. Mensajes por medios y niveles de odio

Medio	Nivel 0	Nivel 1	Nivel 2	Nivel 3
Abc	98	35	30	83
20Minutos	315	59	39	121
Mundo	1125	109	69	363
País	1328	170	82	338
Vanguardia	40	22	31	50

Fuente: Elaboración propia, 2024.

Gráfico 4. Mensajes por medios y niveles de odio



Fuente: Elaboración propia, 2024.

El análisis muestra que el nivel 0, aunque menos explícito, genera la mayor cantidad de mensajes y retuits, lo que sugiere que el discurso de odio sutil tiene mayor probabilidad de ser compartido. Los niveles más altos, como el nivel 3, presentan menos retuits, posiblemente por su carácter explícito que limita la aceptación o visibilidad.

Los medios de comunicación como El País y El Mundo el mayor volumen de mensajes xenófobos. Los discursos se amplifican a través de menciones a noticias publicadas por estos medios y concentran la mayor cantidad de mensajes en todos los niveles, destacando su influencia en la narrativa mediática que se replica en redes sociales

Con relación a la frecuencia mensual de mensajes xenófobos, el gráfico de distribución muestra un aumento significativo en los meses de junio y septiembre de 2024. Estos picos podrían guardar relación con acontecimientos públicos o noticias de interés general relacionadas con temáticas de inmigración. En este sentido se han localizado algunos acontecimientos en los meses de mayor volumen de mensajes xenófobos:

1. Junio 2024:

- Nueva Ley de Asilo: En junio, España aprobó una nueva ley de asilo que facilitó la obtención de permisos de trabajo y vivienda para los solicitantes de asilo. Esta ley generó un amplio debate en los medios de comunicación y en las redes sociales.
- Incremento de llegadas por mar: En junio, se registró un aumento significativo en el número de migrantes que llegaron a las costas españolas por mar, lo que generó preocupación y discusiones sobre la capacidad de acogida del país.

2. Septiembre 2024:

- Crisis migratoria en Canarias: En septiembre, las Islas Canarias experimentaron un aumento en la llegada de migrantes, lo que llevó a una crisis humanitaria y a un debate sobre la gestión de la inmigración en España.
- Inmigración como principal preocupación: Según el barómetro del Centro de Investigaciones Sociológicas (CIS) de septiembre, la inmigración se convirtió en la principal preocupación para los españoles, superando al desempleo. Este cambio en la percepción pública también se reflejó en un aumento de mensajes xenófobos en las redes sociales.

Para finalizar, los resultados obtenidos de la plataforma X entre junio y noviembre de 2024 ilustran claramente la prevalencia de términos cargados de odio y xenofobia, como "terrorista" y "moro", que dominan el discurso con altos porcentajes de menciones e interacciones. Esta predominancia de términos específicos subraya una tendencia hacia la estigmatización de ciertos grupos étnicos y religiosos, una dinámica que Fuentes Osorio (2024) y Naciones Unidas (2019) reconocen como indicativos de una sociedad que facilita la discriminación a través de su lenguaje en los medios y plataformas públicas. La frecuencia de estos términos no solo refleja percepciones arraigadas, sino que también, según el Consejo de Europa (1997), promueve actitudes de racismo y xenofobia, reforzando así barreras invisibles contra la integración y el respeto mutuo.

Además, el análisis de la interacción de estos mensajes, específicamente los retuits y respuestas, indica que, aunque el volumen de publicaciones xenófobas es significativo, la interacción general es relativamente baja, lo que podría sugerir una resistencia o desaprobación tácita por parte de la audiencia mayoritaria. Sin embargo, el hecho de que estos mensajes persistan y sean visibles en plataformas mediáticas como El Mundo y El País refleja la necesidad urgente de políticas de moderación más rigurosas dentro de los medios, tal como sugieren investigaciones sobre el impacto social de los discursos de odio (Valle de Frutos, 2024).

La implicación de los grandes medios en la difusión de mensajes xenófobos y su rol en la formación de la opinión pública no pueden ser subestimados. La relación entre la narrativa mediática y la amplificación del discurso xenófobo, destacada en los mensajes analizados, ilustra cómo los medios pueden actuar como amplificadores de prejuicios existentes, facilitando un entorno en el que el odio no solo se perpetúa, sino que se normaliza (Hyewook, 2023). Este fenómeno es especialmente crítico en el contexto de la influencia de la inteligencia artificial y el aprendizaje automático en la detección y gestión del discurso de odio, como muestran estudios de Arcila-Calderón et al. (2022) y Gautam et al. (2024).

Por último, la continua monitorización y análisis de estas interacciones se hacen esenciales para comprender mejor las dinámicas del discurso de odio en las redes sociales y su relación con eventos sociopolíticos y mediáticos. La educación digital y las estrategias de intervención efectivas deben ser priorizadas para combatir la proliferación del discurso de odio, asegurando así una plataforma más inclusiva y respetuosa. Este enfoque no solo responde a la necesidad inmediata de mitigar el impacto del odio en línea, sino que también prepara el terreno para una reflexión más profunda y soluciones a largo plazo que aborden las raíces culturales y sociales del problema.

6. Conclusiones

La investigación sobre el discurso de odio en la plataforma X ha permitido alcanzar con éxito los objetivos establecidos para el estudio, proporcionando una perspectiva integral y detallada de la naturaleza y el impacto de estos mensajes. Primero, al determinar el nivel de odio de los mensajes xenófobos, se identificó que términos altamente negativos como "terrorista" y "moro" predominan en la comunicación, representando respectivamente el 29% y el 25% del total de las menciones. Estos hallazgos indican un nivel considerable de hostilidad dirigida hacia colectivos específicos, reflejando un ambiente donde el prejuicio y la estigmatización son prominentes.

En relación con la caracterización de los textos, la investigación reveló que el lenguaje utilizado en los mensajes no solo categoriza de manera peyorativa a los inmigrantes y extranjeros, sino que también promueve una narrativa de criminalización y deshumanización. Términos como "delincuente", "ilegal" y "invadir", aunque menos frecuentes que los términos más agresivos, contribuyen significativamente a una percepción distorsionada y negativa de estos colectivos. Esta caracterización de los mensajes portadores de odio destaca la urgencia de abordar el análisis sobre cómo el lenguaje en los medios digitales puede influir en las actitudes sociales y perpetuar la discriminación.

Los hallazgos anteriores muestran cómo se reproducen tópicos en torno al colectivo inmigrante que aparecen de forma recurrente en los bulos y noticias falsas que, de forma intencionada y coordinada, se vierten en las redes sociales de los países de acogida, como sucede en Italia, Grecia o España (Sánchez Esparza et al., 2024). De esta forma la figura del inmigrante es asociada a atributos como los de vándalo, antisocial, usurpador o delincuente, provocando la construcción de sistemas de significado (mitos) e imaginarios sociales que estigmatizan a este colectivo y lo dibujan como un elemento inconveniente en la sociedad de acogida.

Al evaluar el impacto y la repercusión de estos mensajes, se observa que, a pesar del alto volumen de publicaciones xenófobas, la interacción en términos de likes y retuits es relativamente baja. Esto podría interpretarse como una resistencia por parte de la mayoría de los usuarios a incrementar los niveles del discurso de odio, aunque la presencia constante de estos mensajes sigue siendo una preocupación. Es notable que, aunque los mensajes de alto contenido de odio como los clasificados en el nivel 3 son frecuentes, su tasa de retuits no es proporcionalmente alta, lo que sugiere una posible disonancia entre la visibilidad del contenido y la aprobación del mismo por parte de la comunidad en línea.

Los resultados también implican que los principales medios de comunicación, como “El Mundo” y “El País”, juegan un rol significativo en la difusión de estos discursos, lo que plantea cuestiones sobre la responsabilidad editorial y las políticas de moderación de contenido en plataformas de gran alcance. La asociación entre la intensidad del discurso de odio y estos medios indica la necesidad de estrategias para limitar la propagación de mensajes xenófobos, que no solo afectan a los individuos directamente afectados, sino que también envenenan el ambiente social más amplio.

Este estudio subraya la importancia de seguir monitoreando y analizando el discurso de odio en las redes sociales para comprender mejor sus mecanismos y efectos en la sociedad. De cara al futuro, es esencial implementar y reforzar las medidas educativas, de moderación y políticas públicas que puedan mitigar los efectos dañinos del odio online, promoviendo un entorno digital más inclusivo y respetuoso. Así, se podrá no solo reducir la prevalencia de estos discursos perjudiciales sino también fomentar una cultura de respeto y tolerancia en la esfera digital global.

Cabe destacar, que el estudio se centró en mensajes recopilados específicamente de la plataforma X, limitando la generalización de los resultados a otras plataformas de redes sociales donde los patrones de discurso de odio pueden variar debido a diferencias en la demografía de los usuarios, las políticas de moderación y el diseño de la plataforma. Aunque el uso de algoritmos de aprendizaje automático para clasificar el discurso de odio proporciona una herramienta poderosa, estos pueden carecer de la capacidad para entender completamente el contexto y el tono, especialmente en casos de sarcasmo o ironía, lo que puede resultar en clasificaciones erróneas.

Finalmente, este estudio sobre el discurso de odio en la plataforma X muestra la complejidad y la necesidad de abordar este fenómeno en nuestras sociedades digitales. A pesar de las limitaciones metodológicas y de alcance, los resultados obtenidos subrayan la necesidad de intervenciones multidisciplinarias que incluyan mejoras tecnológicas, políticas públicas más rigurosas y estrategias educativas eficaces para combatir el discurso de odio. La persistencia de términos peyorativos y mensajes xenófobos, y su impacto en la cohesión social, desafían a reflexionar sobre nuestra responsabilidad colectiva en la creación de espacios digitales más inclusivos y respetuosos. Este esfuerzo requiere no sólo una vigilancia constante de las plataformas y medios de comunicación, sino también un compromiso activo de cada usuario para fomentar un diálogo constructivo y respetuoso, asegurando que la tecnología sirva como puente y no como barrera para el entendimiento mutuo y la dignidad humana.

7. Agradecimientos

El presente trabajo es resultado parcial del proyecto ‘Hatemia: Taxonomía, presencia e intensidad de las expresiones de odio en entornos digitales vinculados a los medios informativos profesionales españoles’. Hatemia (PID2020-114584GB-I00), financiado por MCIN/AEI/ 10.13039/5011000110

Referencias

- Aldamen, Y. (2023). Xenophobia and Hate Speech towards Refugees on Social Media: Reinforcing Causes, Negative Effects, Defense and Response Mechanisms against That Speech. *Societies*, 13(4), 83. <https://doi.org/10.3390/soc13040083>
- Arcila-Calderón, C., Sánchez-Holgado, P., Quintana-Moreno, C., Amores, J., & Blanco-Herrero, D. (2022). Hate speech and social acceptance of migrants in Europe: Analysis of tweets with geolocation. [Discurso de odio y aceptación social hacia migrantes en Europa: Análisis de tuits con geolocalización]. *Comunicar*, 71, 21-35. <https://doi.org/10.3916/C71-2022-02>
- Arcila-Calderón, C., Amores, JJ, Sánchez-Holgado, P., Vrysis, L., Vryzas, N., y Oller Alonso, M. (2022). ¿Cómo detectar el odio en línea hacia migrantes y refugiados? Desarrollo y evaluación de un clasificador de discurso de odio racista y xenófobo mediante aprendizaje superficial y profundo. *Sustainability*, 14 (20), 13094. <https://doi.org/10.3390/su142013094>
- Baha, O. A. (2022). Nurturing Communication through Social Online Platforms. *International Journal of Humanities and Education Development (IJHED)*, 4(1), 155-159. <https://doi.org/10.22161/>
- Bruneel, S., De Wit, K., Verhoeven, J. C., & Elen, J. (2013). Facebook: When education meets privacy. *Interdisciplinary Journal of e-Skills and Lifelong Learning*, 9, 125-148. <https://doi.org/10.28945/1868>
- Bursztyn, L., Egorov, G., Enikolopov, R., & Petrova, M. (2019). Social media and xenophobia: Evidence from Russia (Working Paper No. 26567). *National Bureau of Economic Research*. <https://doi.org/10.3386/w26567>
- Bursztyn, L., Egorov, G., Enikolopov, R., & Petrova, M. (2020). Social Media and Xenophobia: Evidence from Russia. *CEPR Discussion Paper No. DP14877*, Available at SSRN: <https://ssrn.com/abstract=3628198>
- Chang, L., & Zhang, X. (2024). Introduction: Platforms for social good. *Global Media and China*, 9(3), 253-259. <https://doi.org/10.1177/20594364241241777>
- Comisión Europea contra el Racismo y la Intolerancia. *Recomendación de política general N.º 15 de la ECRI sobre la lucha contra el discurso de odio*; Consejo de Europa: Estrasburgo, Francia, 2016.
- Consejo de Europa. *Recomendación nº R 20 del Comité de Ministros a los Estados miembros sobre el "discurso de odio"*; Consejo de Europa: Estrasburgo, Francia, 1997.
- De Lucas Vicente, A; Römer Pieretti, M.; Izquierdo, D.; Montero-Díaz, J.; Said-Hung, E. (2022). Manual para el Etiquetado de mensajes de odio. figshare. Presentation. <https://doi.org/10.6084/m9.figshare.18316313.v4>
- Dhungana Sainju, K., Zaidi, H., Mishra, N., & Kuffour, A. (2022). Xenophobic Bullying and COVID-19: An Exploration Using Big Data and Qualitative Analysis. *International Journal of Environmental Research and Public Health*, 19(8), 4824. <https://doi.org/10.3390/ijerph19084824>
- Dutta, M. J. (2024). Digital platforms, Hindutva, and disinformation: Communicative strategies and the Leicester violence. *Communication Monographs*, 1–29. <https://doi.org/10.1080/03637751.2024.2339799>
- Forrest, K. B., & Wexler, J. (2023). Social media platforms. In K. B. Forrest & J. Wexler (Eds.), *Is justice real when reality is not?* (pp. 147-159). Academic Press. <https://doi.org/10.1016/B978-0-323-95620-8.00001-3>
- Fuentes Osorio, J. L. (2024). Hateful speech. La expansión del discurso de odio. *Revista Electrónica de Criminología*. 02-08. 1-30.
- Gautam, A., Singh, A., Verma, A., & Sinha, J. (2024). Hate speech detection using deep learning. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 12(5), 196–202. <https://doi.org/10.22214/ijraset.2024.61475>
- Gelashvili, T. (2018). *Hate speech on social media: Implications of private regulation and governance gaps* (Master's thesis, Lund University). Faculty of Law, Lund University.
- Hyewook J. (2023). Hate Speech, Subject Agency and Performativity of Bodies. *The Journal of Criticism and Theory*, 28(1), 271-313. [10.19116/theory.2023.28.1.271](https://doi.org/10.19116/theory.2023.28.1.271)
- Kaur, T. (2023). Digital platforms: Social media platforms, knowledge platforms, media sharing platforms, service-oriented platforms. *International Journal for Research in Applied Science & Engineering Technology*, 11(VII), 2031. doi: 10.22214/ijraset.2023.54879

- Martínez-Rolán, X., Sierra, J., & Ring Carlson, C. (2024). Discursos de odio, populismo digital y autoritarismos en red. *Revista Latina De Comunicación Social*, (82). Recuperado a partir de <https://nuevaepoca.revistalatinacs.org/index.php/revista/article/view/2298>
- Ministerio del Interior de España (2020). Informe de Evolución de los Delitos de Odio en España. <https://bit.ly/3PGN4ys>
- Ministerio de Inclusión, Seguridad Social y Migraciones. (2017). *Protocolo y sistema de indicadores para la detección del discurso de odio en redes sociales*. <https://bit.ly/3QImKnr>
- Nasuto, A. & Rowe, F. (2024a). Exposing Hate - Understanding Anti-Immigration Sentiment Spreading on Twitter. *arXiv.org, abs/2401.06658* doi: 10.48550/arxiv.2401.06658
- Nasuto A, Rowe F (2024b) Understanding anti-immigration sentiment spreading on Twitter. *PLoS ONE* 19(9): e0307917. <https://doi.org/10.1371/journal.pone.0307917>
- Observatorio Español del Racismo y la Xenofobia (Oberaxe). (2017a). *Evolución del racismo, la xenofobia y otras formas de intolerancia en España*. Ministerio de Trabajo, Migraciones y Seguridad Social.
- Observatorio Español del Racismo y la Xenofobia (Oberaxe), Centro de Investigaciones Sociológicas. (2017b). *Actitudes hacia la inmigración X*. Estudio Nº 3190. <https://bit.ly/3ohwSqD>
- Observatorio Español del Racismo y la Xenofobia [OBERAXE]. (2024). *Informe anual de monitorización del discurso de odio en redes sociales 2023*. Ministerio de Inclusión, Seguridad Social y Migraciones. <https://www.inclusion.gob.es/oberaxe/es/index.htm>
- Organización de las Naciones Unidas (ONU) (2019). *La estrategia y plan de acción de las Naciones Unidas para la lucha contra el discurso de odio*. <https://bit.ly/3ibIWpU>
- Raborife, M., Ogbuokiri, B., Aruleba, K. (2024). The Role of Social Media in Xenophobic Attack in South Africa. (2024). *Journal of the Digital Humanities Association of Southern Africa* , 5(1). <https://doi.org/10.55492/dhasa.v5i1.5026>
- Sánchez Esparza, M., Diéguez, I. V., & Arribas, A. M. (2023). Mapping Stigmatizing Hoaxes Towards Immigrants on Twitter and Digital Media: Case Study in Spain, Greece, and Italy. In E. Said Hung & J. Diaz (Eds.), *News Media and Hate Speech Promotion in Mediterranean Countries* (pp. 136-161). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-6684-8427-2.ch008>
- Sánchez Esparza, M., Vázquez Diéguez, I., & Merino Arribas, D. (2024). La semiótica del odio en los bulos sobre inmigrantes detectados por plataformas de fact checking en España, Grecia e Italia. *Revista ICONO 14. Revista científica De Comunicación Y Tecnologías Emergentes*, 22(1), e2083. <https://doi.org/10.7195/ri14.v22i2.2083>
- Santana dos Santos, A.R.; De Oliveira Rodrigues, C.M. & Summer de Melo, H.B. (2022). Identifying Xenophobia in Twitter Posts Using Support Vector Machine with TF/IDF Strategy. In *Proceedings of the XVIII Brazilian Symposium on Information Systems (SBSI '22)*. Association for Computing Machinery, New York, NY, USA, Article 37, 1–7. <https://doi.org/10.1145/3535511.3535548>
- Schäfer, S., Rebasso, I., Boyer, M. M., & Planitzer, A. M. (2024). Can We Counteract Hate? Effects of Online Hate Speech and Counter Speech on the Perception of Social Groups. *Communication Research*, 51(5), 553-579. <https://doi.org/10.1177/00936502231201091>
- Umarova, K., Okorafor, O., Lu, P., Shan, S., Xu, A., Zhou, R., Otiono, J., Lyon, B., & Leshed, G. (2024). Xenophobia Meter: Defining and Measuring Online Sentiment toward Foreigners on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 18(1), 1517-1530. <https://doi.org/10.1609/icwsm.v18i1.31406>
- Valle De Frutos, S. (2024). Discurso de odio y ofensivo en la red social Twitter hacia el colectivo chino. Análisis de la sinofobia: del rechazo cultural encubierto al explícito. *Anuario Electrónico de Estudios en Comunicación Social "Disertaciones"*, 17(1). <https://doi.org/10.12804/revistas>
- We Are Social. (2024). *Digital 2024*. We Are Social. <https://wearesocial.com/es/blog/2024/01/digital-2024/>
- Weber, I., Vandebosch, H., Poels, K., & Pabian, S. (2023). Features for hate? Using the Delphi method to explore digital determinants for online hate perpetration and possibilities for intervention. *Cyberpsychology, Behavior, and Social Networking*, 26(7). <https://doi.org/10.1089/cyber.2022.0195>