

Antisemitismo, (anti)sionismo e islamofobia: precisiones conceptuales, evidencias empíricas y marcos operativos en X, Facebook, Instagram y TikTok (2024)¹

Antisemitism, (Anti-)Zionism and Islamophobia: Conceptual Clarifications, Empirical Evidence and Operational Frameworks on X, Facebook, Instagram and TikTok (2024)

MARTÍN OLLER ALONSO²

CARLOS ARCILA CALDERÓN³

MAXIMILIANO FRÍAS VÁZQUEZ⁴

WILLIAM GONZÁLEZ BAQUERO⁵

Resumen

En este artículo desarrollamos un marco teórico y conceptual que distingue la judeofobia religiosa del antisemitismo moderno y permite separar el odio antijudío de la crítica política legítima a Israel. Para ello, articulamos tres definiciones operativas a partir de la Holocaust Remembrance Alliance (IHRA), la Nexus Task Force (Nexus) y la Jerusalem Declaration on Antisemitism (JDA), con las que operacionalizamos el antisemitismo (conspiraciones sobre “poder judío”, culpabilización colectiva, doble lealtad, negación o banalización de la Shoá, negación del derecho a existir y dobles raseros extremos) y delimitamos aquello que no lo es per se (crítica basada en hechos a las políticas de

¹ Este artículo forma parte del proyecto COIN - Countering Online INTolerance against especially vulnerable groups. Large-scale monitoring and narrative combat of online hatred in Spain. Este Proyecto está financiado por la *European Union's Horizon 2020 research and innovation programme* bajo el programa *Marie Skłodowska-Curie grant agreement No 101034371*.

² Universidad de Salamanca, molleralonso@usal.es <https://orcid.org/0000-0002-2256-5681>

³ Universidad de Salamanca, carcila@usal.es <https://orcid.org/0000-0002-2636-2849>

⁴ Universidad de Salamanca, maxfrias@usal.es <https://orcid.org/0000-0001-9750-6136>

⁵ Universidad de Salamanca, willgb@usal.es <https://orcid.org/0000-0001-5760-9771>

Israel, apoyo a derechos palestinos o BDS no violento sin estereotipos). Sobre esta base analizamos el antisemitismo digital y su entrelazamiento con el (anti)sionismo y la islamofobia en X, Facebook, Instagram y TikTok. Partimos de un corpus de 798.475 mensajes en español geolocalizados en España (abril-agosto de 2024). En nuestro método, combinamos la detección automatizada de toxicidad (API Perspective) con modelos de categorización temática de odio dirigido a cinco grupos vulnerables: antisemitismo, antigitanismo, [islamofobia/racismo antiárabe](#), xenofobia y LGBTIQfobia. Los resultados muestran (i) picos de “neoantisemitismo” ligados a coyunturas relacionadas con la escalada bélica en Israel-Gaza, donde el antisionismo funciona como vehículo de tropos clásicos; (ii) transferencia cruzada de hostilidad entre comunidades judías y musulmanas y solapamientos con otras formas de racismo; (iii) así como diferencias significativas entre plataformas en conspiracionismo y negacionismo, siendo las analogías nazismo-Israel y los dobles raseros extremos marcadores de alto riesgo. Finalmente, mostramos cómo las lógicas algorítmicas de “enragement” y la difusión de contenidos tóxicos convierten estas redes (en especial X) en entornos proclives a la intensificación del (discurso de) odio, más que a su mitigación. Como posibles contraposiciones a esta violencia, proponemos trabajar en la alfabetización conceptual, los protocolos de moderación basados en IHRA-Nexus-JDA, la monitorización de crisis y las medidas de protección para comunidades judías y musulmanas, alineadas con los marcos de la Organización de Naciones Unidas (ONU), la Organización para la Seguridad y la Cooperación en Europa (OSCE) y la Unión Europea (UE), a fin de mitigar el odio (online) sin restringir el disenso y la crítica legítima.

Palabras clave: hate speech, antisemitismo, (anti)sionismo, (anti)islamismo, redes sociales.

Abstract

In this article, we develop a theoretical and conceptual framework that differentiates religious Judeophobia from modern antisemitism and enables the separation of anti-Jewish hatred from legitimate political critique of Israel. To achieve this, we articulate three operational definitions based on the International Holocaust Remembrance Alliance (IHRA), the Nexus Task Force (Nexus), and the Jerusalem Declaration on Antisemitism (JDA). These definitions allow us to operationalise antisemitism- such as conspiracies about “Jewish power,” collective blame, dual loyalty, denial or trivialisation of the Shoah, denial of the right to exist, and extreme double standards- and to clarify what does not qualify as antisemitism per se (fact- based criticism of Israeli policies, support for Palestinian rights, or non- violent BDS, provided it does not rely on stereotypes). Building on this foundation, we analyse digital antisemitism and its entanglement with (anti-)Zionism and Islamophobia on X, Facebook, Instagram, and TikTok. We draw on a dataset of 798,475 Spanish-language messages geolocated in Spain (April–August 2024). Methodologically, we combine automated toxicity detection (Perspective API) with thematic classification models targeting five vulnerable groups: antisemitism, antigypsyism, Islamophobia/anti- Arab racism, xenophobia, and LGBTIQ-phobia. The results reveal (i) spikes of “neo- antisemitism” linked to episodes of military escalation in Israel-Gaza, where anti-Zionism acts as a vehicle for classic tropes; (ii) cross- transfer of hostility between Jewish and Muslim

communities and overlaps with other forms of racism; and (iii) notable cross- platform differences in conspiracy rhetoric and denialism, with Nazi–Israel analogies and extreme double standards emerging as high- risk markers. Finally, we demonstrate how algorithmic “enragement” logics and the circulation of toxic content transform these platforms- especially X- into environments inclined to intensify hate speech rather than mitigate it. As potential countermeasures to this violence, we suggest working on conceptual literacy, moderation protocols grounded in the IHRA–Nexus–JDA, crisis monitoring, and protective measures for Jewish and Muslim communities, aligned with the frameworks of the United Nations (UN), the Organisation for Security and Co- operation in Europe (OSCE), and the European Union (EU), to address (online) hatred without restricting dissent and legitimate criticism.

Keywords: hate speech, antisemitism, (anti-)Zionism, (anti-)Islam/Islamophobia, social media.

1. Introducción

El término antisemitismo es, a la vez, histórico y polémico: designa una ideología moderna —configurada en la segunda mitad del siglo XIX— que trasladó la antigua judeofobia religiosa a un lenguaje pseudocientífico, racializado e irreversible (frente al *Judenhass* premoderno) (Engel 2008; Dahlström 2022; Consonni 2020). La distinción que proponemos en este artículo entre término y concepto evita retroproyecciones anacrónicas y previene equívocos asociados a “semita” como categoría racial inexistente. Los estudios recientes subrayan, además, que el antisemitismo ha mutado históricamente —del religioso al racial y, tras 1945, hacia formas político-ideológicas y negacionistas — sin perder continuidad ni relevancia, al referirse a “el odio más largo” (*the longest hatred*) como una hostilidad extraordinariamente persistente que perdura a través de dichas formas cambiantes (Wistrich 1991; Dahlström 2022).

Hoy en día, el debate se centra en el vínculo entre el antisemitismo y el Estado de Israel y el sionismo (Jaffee 2024). La(s) definición(es) de la International Holocaust Remembrance Alliance (IHRA) (2016) se han consolidado como referencia institucional por su utilidad operativa, a la par que han recibido críticas por el riesgo de politización y por confundir las críticas legítimas dirigidas a las políticas de Israel con el odio antijudío (Stern 2020). Como respuestas correctivas, la Nexus Task Force (Nexus) (Nexus Document 2021) y la Jerusalem Declaration on Antisemitism (JDA 2021) proponen marcos más explícitos para separar (a) las conspiraciones, la culpabilización colectiva, la “doble lealtad”, la negación/banalización de la Shoá y la negación del derecho de los judíos a existir —que sí constituyen antisemitismo— de (b) la crítica legítima basada en hechos a [políticas israelíes](#), el apoyo a derechos palestinos o las tácticas como el Boicot, Desinversión y Sanciones (BDS)—que no lo son per se si no recurren a estereotipos o deslegitimación del pueblo judío—. En paralelo, organismos internacionales como la OSCE/ODIHR (2004–2019), el Consejo de Europa/ECRI (1997, 2020), la UE (2013–2022) y la Comisión Europea (2021) han articulado estrategias, guías y obligaciones penales frente a los delitos de odio, con especial atención al entorno digital.

Para deshacer este intríngulis, reservamos el término antisemitismo para designar ese régimen ideológico moderno —con sus doctrinas, aparatos institucionales y proyectos políticos, incluidos los desarrollos pos-1945 en clave político-ideológica, conspirativa y negacionista— y nos referimos a antijudío de forma más amplia cuando el contenido del odio va dirigido contra personas judías, el judaísmo o lo “judío” como construcción simbólica. Ello abarca desde las formas premodernas de judeofobia hasta las configuraciones nacional-raciales de los siglos XIX y XX. Y, en la posguerra, incluye los discursos antagónicos hacia el antisionismo o la israelfobia cuando lo que se cuestiona no son políticas concretas, sino la legitimidad misma de la existencia judía (colectiva). Por analogía y a partir de la literatura sobre islamofobia, distinguimos entre el régimen ideológico y las formas discursivas del odio hacia lo musulmán. Entendemos la islamofobia como el conjunto de actitudes de temor, hostilidad, odio o discriminación hacia el islam y a quienes se identifican como musulmanes, en la intersección entre la aversión religiosa y el racismo antimusulmán (Bleich 2011; Selod & Embrick 2013). En cambio, empleamos el adjetivo antiislamista para describir los discursos, marcos interpretativos y prácticas que vehiculan este rechazo hacia el islam y las personas musulmanas, sin confundirlo con la discusión específica sobre el “islamismo” como corriente político-religiosa. Así, en el análisis empírico hablamos de discursos antijudíos y antiislamistas para referirnos al repertorio de enunciados e imágenes que encarnan el odio hacia lo judío y lo musulmán, mientras que antisemitismo e islamofobia designan los regímenes ideológicos que estructuran y legitiman históricamente esos discursos en contextos políticos, jurídicos y mediáticos concretos.

Este artículo se inscribe en ese cruce conceptual, normativo, histórico y empírico en el que pocas investigaciones se han adentrado. De ahí que presentemos, además del análisis explicitado en los párrafos anteriores, un estudio computacional de gran escala sobre discurso antisemita, a partir de la (re)significación de lo que es antijudío, antisionista, antiislamista y antimusulmán en redes sociales (X, Facebook, Instagram, TikTok) en España (abril-agosto de 2024), basado en aproximadamente 800.000 mensajes en español. En dicho análisis, combinamos la detección de toxicidad (API Perspective) y de odio con los modelos de categorización (generados en el proyecto COIN) que operacionalizan conjuntamente las definiciones y conceptualizaciones propuestas por la IHRA (2016), Nexus (2021) y JDA (2021), a fin de distinguir odio de crítica política e identificar vectores de neoantisemitismo (Marcus 2015) (por ejemplo, antisionismo con tropos clásicos) e islamofobia (hostilidad hacia musulmanes por serlo, estereotipos civilizacionales, culpabilización colectiva). Este enfoque dialoga con las tipologías psicosociales (Kovács & Fischer 1994)—primario/afectivo, conativo/activo, secundario/latente— para graduar intensidades y modalidades. Además de aportar evidencias empíricas en español —escasas en comparación con las encontradas en inglés—, proponemos un marco reproducible para plataformas y decisores públicos alineado con la OSCE/UE (OSCE 2004; [OSCE/ODIHR 2014](#); Consejo de la UE 2022; Comisión Europea 2021).

Con base en esta perspectiva (i) teórica, conceptual, histórica y legislativa y (ii) empírica, proponemos las siguientes preguntas de investigación.

(P1) Operacionalización. Basándonos en los criterios diferenciales entre antisemitismo y crítica legítima propuestos por la IHRA (2016), la Declaración Nexus (2021) y la JDA (2021), nos preguntamos: ¿Hasta qué punto un clasificador basado en la IHRA–Nexus–JDA discrimina con precisión entre el antisemitismo y la crítica legítima a Israel y al sionismo?

(P2) Prevalencia y dinámica. Siguiendo los indicadores de registro y monitorización del odio de la [OSCE/ODIHR \(2014\)](#), planteamos: ¿Cuál es la incidencia comparada del antisemitismo y de la islamofobia en España (2024) y cómo varía según la plataforma?

(P3) Escalada coyuntural. Partiendo de las definiciones formuladas en la JDA (2021) y en la Declaración Nexus (2021) sobre el papel del antisionismo en la actualización de tropos antijudíos, proponemos: ¿Qué relación establecen los picos discursivos con los episodios Israel–Gaza y cómo opera el antisionismo como vector de tropos antijudíos clásicos?

(P4) Tropología. Tomando los tropos de antisemitismo político-instrumental recogidos en la IHRA (2016) y aplicando la tipología de intensidades actitudinales de Kovács & Fischer (1994), ¿qué estereotipos (pre)dominan (poder oculto, doble lealtad, negacionismo, comparaciones con el nazismo) y con qué intensidades (primario/afectivo, conativo/activo, secundario/latente) en los discursos de odio online?

(P5) Transferencias de odio. La ODIHR y la Comisión Europea identifican tendencias paralelas de odio hacia judíos y musulmanes y señalan riesgos de polarización, por lo que preguntamos: ¿Se observan patrones de culpabilización colectiva cruzados entre las comunidades judías y musulmanas?

(P6) Implicaciones normativas. A la luz de nuestros resultados y de los marcos regulatorios recientes de la UE (Consejo de la UE 2022; Comisión Europea 2021), ¿qué protocolos de moderación y de alfabetización teórico-conceptual se derivan o se establecen para mitigar el odio sin restringir el disenso político protegido y legítimo?

En el contexto español (2024), analizar la presencia y configuración de los discursos antijudíos y antiislamistas en X, Facebook, Instagram y TikTok, así como sus implicaciones para la delimitación entre discurso de odio y disenso político legítimo en los marcos de gobernanza de contenidos, nos lleva a plantear tres objetivos específicos, mediante los cuales aspiramos a: **(OE1)** refinar la medición computacional del antisemitismo contemporáneo; **(OE2)** aportar un diagnóstico comparado por plataformas y coyunturas; y **(OE3)** traducir nuestros hallazgos teórico/empíricos en pautas aplicables por reguladores (públicos y privados), medios y plataformas en el ecosistema hispanohablante.

2. Redes sociales digitales como enclave sociotécnico, económico y geoestratégico

Las redes sociales digitales constituyen el espacio comunicativo desde el cual se obtienen los datos de este estudio. No son solo espacios de interacción interpersonal, sino también infraestructuras

sociotécnicas orientadas a captar la atención y al rendimiento económico, donde algoritmos de recomendación, sistemas de segmentación publicitaria y métricas de engagement jerarquizan la visibilidad de los mensajes y de los actores. Esta curaduría automatizada y personalizada de contenidos genera entornos informativos parciales —como las conocidas *filter bubbles* y *echo chambers*— que refuerzan creencias previas y minimizan la exposición a perspectivas disonantes (Pariser 2011; Sunstein 2017; Kitchens et al. 2020; Cinelli et al. 2021).

Esta arquitectura online tiene consecuencias directas para la discusión democrática. La combinación de personalización algorítmica y lógica de mercado premia los contenidos más polarizantes, emocionales y conflictivos, que maximizan los clics, los comentarios y el tiempo de visualización, lo que contribuye a la fragmentación del espacio público y a la circulación y difusión del discurso de odio. A ello se suma el carácter geoestratégico de estas plataformas digitales. La literatura sobre [computational propaganda \(propaganda computacional\)](#) define este fenómeno como el ensamblaje de redes sociales, agentes automatizados (bots), macrodatos y curaduría humana, orientado a manipular la opinión pública mediante la difusión coordinada de información engañosa, sesgada (Woolley & Howard 2016) o dañina. En este contexto emergen las llamadas cibertropas: unidades de gobiernos, partidos políticos y otros actores organizados que emplean bots, cuentas títere, troll farms y campañas segmentadas para influir en los debates públicos, enturbiar la deliberación y atacar a colectivos específicos (Bradshaw & Howard 2021), como los judíos y los musulmanes. La “industrialización de la desinformación” transforma las redes sociales en herramientas de guerra informacional y de operaciones psicológicas, a veces en coordinación con aparatos militares y de seguridad.

Este marco es especialmente relevante en coyunturas bélicas y de violencia masiva, como la actual, donde la disputa por el relato —sobre todo en torno a términos como “terrorismo”, “genocidio” o “seguridad”— también se despliega en tiempo real en las plataformas digitales. En estos escenarios, el odio hacia los colectivos judíos o musulmanes se convierte en un fenómeno “orgánico” (expresión de prejuicios preexistentes) y en un producto artificialmente amplificado por la arquitectura algorítmica de las plataformas y por campañas coordinadas de cibertropas que buscan dañar, desinformar, polarizar y deslegitimar las voces críticas. De este modo, las redes sociales digitales funcionan como cámaras de resonancia en las que se entrecruzan dinámicas estructurales de antisemitismo e islamofobia con tácticas estratégicas de desinformación y manipulación política.

Desde esta comprensión de las redes sociales como enclaves sociotécnicos, económicos y geopolíticos, abordamos en las siguientes secciones la genealogía y las tipologías del antisemitismo, del antijudaísmo, del sionismo y de la crítica política, así como la conceptualización de la islamofobia y del prejuicio antimusulmán. En este sustrato se inscriben los discursos antijudíos y antiislamistas que analizamos empíricamente en nuestro estudio.

3. Genealogía y tipologías de antisemitismo, sionismo y crítica política: definiciones operativas y controversias

El antisemitismo es, ante todo, un concepto moderno que cristaliza en la segunda mitad del siglo XIX para resemantizar una hostilidad previa — la judeofobia/Judenhass — en clave secular, pseudocientífica y racializada. Para nosotros, la distinción entre término y concepto en el análisis del odio online es crucial: “antisemitismo” no remite a una supuesta “raza semita” (noción hoy descartada), sino a una ideología politizada que convierte la pertenencia judía en una marca indeleble e inasimilable, desbordando la matriz teológica previa en la que la conversión operaba como una eventual salida (Engel 2008; Klug 2003; Consonni 2022). En ese tránsito y evolución, Wilhelm Marr popularizó este léxico, y la Liga de los Antisemitas contribuyó a su institucionalización en el espacio público alemán. Finalmente, el giro racial-biologicista halló su radicalización en el ideario y las prácticas del nacionalsocialismo (Pulzer 1988; Zimmermann 1986; Wistrich 1991), por todos conocidas.

Desde el punto de vista historiográfico, múltiples propuestas de periodización nos permiten captar ciertas continuidades y rupturas: (i) un antisemitismo antiguo (matrices paganas y étnicas); (ii) un antisemitismo medieval de raíz cristiana (acumulación de imputaciones teológico-jurídicas); (iii) un antisemitismo moderno/racial (siglos XIX–XX), asociado a nacionalismos, biologización de la diferencia y programas de exclusión; y (iv) un antisemitismo contemporáneo/neoantisemitismo, donde el eje Israel/Sionismo y el negacionismo articulan nuevas manifestaciones de rechazo y odio (Pulzer 1988; Wistrich 1991; ECRI 2021; Jaffee 2024). Esta diacronía, sin embargo, no debe entenderse como un continuum uniforme, ya que, de acuerdo con Consonni (2022), el pasaje a lo “racial” constituye una reconfiguración cualitativa, más que una simple intensificación del viejo concepto de Judenhass.

En paralelo, dichas tipologías sincrónicas enriquecen el análisis de las formas y los umbrales del fenómeno del odio online dirigido a “lo judío”. La literatura psicosocial ha distinguido dimensiones y niveles que ayudan a graduar desde el prejuicio hasta la acción: componentes cognitivos/afectivos (estereotipos, hostilidad emocional), conativos (voluntad de restringir derechos/participación) y conductuales (acciones hostiles), junto con expresiones secundarias (p. ej., negación o banalización de la Shoá, traslación de culpas) y un sustrato latente de predisposición social (Kovács & Fischer 1994). Esta gradación es operativa para nuestra investigación y para la generación de políticas públicas, ya que permite diferenciar el umbral lingüístico (discurso), el umbral actitudinal (intencionalidad discriminatoria) y el umbral de daño (conducta y victimización/víctima/victimario).

El debate definicional desarrollado en los últimos años señala la necesidad de mejorar la precisión conceptual para evitar la infla(ma)ción del término (Engel 2008). En este marco, la *Working Definition of Antisemitism* de la IHRA (2016) se ha consolidado como un estándar operativo en educación y política, definiendo el antisemitismo como una “cierta percepción de los judíos que puede expresarse como odio hacia los judíos”⁶, acompañándose de once ejemplos que recogen tanto tropos clásicos

⁶ La IHRA está publicada como documento en línea.

(conspiracionismo, “doble lealtad”, negación/banalización de la Shoá) como manifestaciones contemporáneas relacionadas con Israel (p. ej., negar el derecho del pueblo judío a la autodeterminación, dobles raseros no exigidos a otros Estados, equiparaciones con el nazismo sin base). Aunque no es jurídicamente vinculante, ha sido citada por organismos europeos y multilaterales, como la OSCE/ODIHR, el Consejo de Europa y la Unión Europea, para la formación, el monitoreo y la orientación de políticas públicas (IHRA 2016; OSCE/ODIHR 2017; ECRI 2021; European Parliament 2017; Council of the EU 2018, 2020; European Commission 2021).

La expansión y la aceptación de las definiciones de la IHRA no están exentas de críticas. Stern (2020) alertó sobre los riesgos de politización y de posibles prácticas de disuasión del disenso respecto a las políticas propuestas por el Estado de Israel. En respuesta a dichas críticas, dos marcos complementarios han ganado tracción académica y política en los últimos años: el Nexus Document (Nexus 2021) y la Jerusalem Declaration on Antisemitism (JDA 2021). Nexus precisa cuándo el discurso sobre Israel sí constituye antisemitismo —p. ej., culpabilización colectiva de judíos por acciones del Estado, conspiraciones, negación del derecho del pueblo judío a existir— y cuando no —p. ej., críticas basadas en hechos a políticas e instituciones israelíes, apoyo a derechos palestinos o BDS como táctica no violenta—, siempre que no medien estereotipos, deshumanización o negación de derechos colectivos (Nexus Task Force 2021). La JDA propone una definición concisa (“discriminación, prejuicio, hostilidad o violencia contra judíos como judíos”) y 15 directrices: cinco generales; cinco que identifican el antisemitismo en relación con Israel/Palestina (p. ej., exigir a los judíos que condenen a Israel por ser judíos; atribuir tropos antisemitas al Estado de Israel; presumir lealtad dual y responsabilización colectiva); y cinco que aclaran lo que no constituye antisemitismo (p. ej., rechazo ideológico del sionismo, BDS y calificaciones jurídicas/históricas —incluida la “apartheid”— cuando se formulan con argumentos empíricos y sin estereotipos ni deslegitimación del pueblo judío) (JDA 2021).

En el ámbito europeo, estas definiciones se inscriben en un doble carril normativo. Por una parte, como parte del derecho duro/vinculante: la Decisión Marco 2008/913/JAI obliga a tipificar penalmente formas graves de racismo y xenofobia, y la Directiva 2012/29/UE establece estándares mínimos de derechos y de protección de las víctimas de delitos de odio (Council of the EU 2008; European Parliament & Council 2012). Por otra parte, en lo concerniente al *soft law*/estrategias, el Parlamento Europeo (2017) instó a los Estados a utilizar la IHRA; el Consejo de la UE (2018, 2020) pidió coordinar y transversalizar la lucha contra el antisemitismo y la Comisión Europea adoptó la Estrategia 2021–2030, articulando la prevención, la memoria de la Shoá, la seguridad comunitaria y el fomento de la vida judía (European Parliament 2017; Council of the EU 2018, 2020; European Commission 2021). A ello se suman guías prácticas publicadas por la OSCE/ODIHR y la European Commission against Racism and Intolerance (ECRI) para la prevención, la recogida de datos y la formación especializada (OSCE/ODIHR 2017; ECRI 2021).

En lo que respecta al balance conceptual-operativo de nuestro análisis, la IHRA ofrece una herramienta de identificación ampliamente utilizada y el Nexus y la JDA refuerzan la seguridad semántica al trazar

límites positivos y negativos (qué es y qué no es antisemitismo) en el terreno de Israel/Palestina, reduciendo el riesgo de confundir la crítica política legítima con el odio antijudío. Combinadas con el andamiaje jurídico/legislativo europeo (penal, de víctimas, educativo y estratégico), proporcionan criterios operables para la anotación, la investigación y la moderación en entornos digitales polarizados como el actual (González-Esteban et al. 2024), a la vez que preservan el disenso protegido y garantizan una protección efectiva frente al antisemitismo.

De la síntesis propuesta en esta sección, derivamos varias consecuencias: (a) una operación conceptual restrictiva —reservar “antisemitismo” para su configuración decimonónica y sus proyecciones contemporáneas, distinguiéndolo de judeofobia y evitando ambigüedades ligadas a “semita”— y (b) una operacionalización analítica que combina la genealogía histórica, las tipologías psico-sociales y los marcos multifuente (IHRA/JDA/Nexus) para medir intensidades, identificar vectores (p. ej., antisionismo con tropos clásicos) y traducir esos criterios a contextos digitales y *offline*/físicos (IHRA 2016; JDA 2021; Nexus Task Force 2021; Kovács & Fischer 1994; Engel 2008; Klug 2003; Consonni 2022; Pulzer 1988; Zimmermann 1986; Wistrich 1991; ECRI 2021; OSCE/ODIHR 2017; European Parliament 2017; Council of the EU 2008, 2018, 2020; European Parliament & Council 2012; European Commission 2021; Stern 2020).

4. Islamofobia y prejuicio antimusulmán

El término islamofobia alude, en sentido amplio, a las actitudes de temor, odio, hostilidad o discriminación hacia el islam y las personas musulmanas. Suele usarse como equivalente de “prejuicio antimusulmán”, aunque algunos analistas proponen matizar su definición. Una parte del debate académico distingue entre la aversión a la religión islámica en sí (odio antiislámico o musulmanofobia teológica) y el racismo dirigido contra personas o comunidades identificadas como musulmanas (racismo antimusulmán) (Bleich 2011; Selod & Embrick 2013). En la práctica, ambos aspectos suelen entrelazarse, por lo que la *islamofobia* se emplea hoy en día para abarcar un amplio espectro de odio hacia lo musulmán, ya sea presentado como animadversión cultural o religiosa o como racialización de los musulmanes como “el otro” peligroso (Garner & Selod 2015; Meer 2009).

Históricamente, la hostilidad hacia el islam en las sociedades occidentales tiene raíces profundas, que se remontan a las interacciones conflictivas entre Europa cristiana y el mundo musulmán. Desde las Cruzadas medievales hasta la Reconquista ibérica y las guerras contra el Imperio otomano, se forjó en la cristiandad europea una imagen del musulmán como enemigo externo, infiel y amenazante (Spektorowski 2024). Esta herencia de percepciones conflictivas —parte de lo que Edward Said denominó orientalismo, es decir, la construcción estereotipada del “Oriente musulmán” por oposición a Occidente— sentó las bases de prejuicios que perduran (Said 1978). No obstante, el concepto moderno de “islamofobia” es relativamente reciente; si bien su etimología remonta a inicios del siglo XX, adquirió difusión global sobre todo desde la década de 1990 y se consolidó en la literatura comparada en los años posteriores (Bleich 2011).

En la actualidad, los académicos entienden la islamofobia fundamentalmente como una forma de racismo dirigida contra personas percibidas como musulmanas, independientemente de su origen étnico real (Garner & Selod 2015; Meer 2009; Selod & Embrick 2013). Aunque el islam es una religión y no una raza, el fenómeno islamofóbico opera mediante la racialización de ciertos rasgos (nombre, apariencia, origen nacional) asociados estereotipadamente al islam, para luego atribuir características negativas a todo un grupo; por ello, numerosos estudios hablan de “racismo antimusulmán” como sinónimo de islamofobia (Meer 2009; Selod & Embrick 2013). Es importante señalar que la islamofobia contemporánea se ha globalizado y adaptado a distintos contextos locales. Tras los ataques del 11 de septiembre de 2001 y el inicio de la llamada “Guerra contra el Terror”, se inauguró una nueva era de sesgo y hostilidad hacia los musulmanes a escala mundial (Bakali 2024; Kundnani 2012). En países occidentales, esto se tradujo en políticas de seguridad y en narrativas mediáticas que representaban al musulmán (o incluso al árabe y al surasiático) como un potencial enemigo interno, lo que favoreció un clima de sospecha generalizada. En palabras de Naved Bakali, el 11-S homogeneizó la figura del “otro musulmán” como enemigo perpetuo, alimentando un relato islamófobo transnacional que trasciende la división Global Norte/Sur (Bakali 2024).

Los efectos de este fenómeno son tanto interpersonales como institucionales. A nivel interpersonal, la islamofobia se manifiesta en agresiones, discursos de odio, vandalismo en mezquitas y discriminación cotidiana contra individuos musulmanes o percibidos como tales. A nivel institucional, se refleja en leyes y prácticas estatales: perfiles raciales en aeropuertos, vigilancia desproporcionada de comunidades musulmanas, prohibiciones de símbolos religiosos (como el velo islámico) y otras políticas que tratan a los musulmanes como colectivos sospechosos. De hecho, la islamofobia se ha vuelto endémica en la retórica política, la legislación y los aparatos de seguridad estatales de varios países (Kundnani 2012; Selod & Embrick 2013). Esto evidencia su carácter de racismo estructural: no es solo un prejuicio individual, sino que está incrustado en las estructuras de poder. Por ejemplo, en Europa, el auge de partidos de extrema derecha con plataformas antiinmigración e islamófobas ha normalizado un discurso en el que “defender la identidad nacional” implica excluir a la población musulmana. En Asia, naciones como Myanmar o India también han visto narrativas islamofóbicas violentas dirigidas contra minorías musulmanas (Bakali 2024; Garner & Selod 2015). En suma, la islamofobia contemporánea tiene dimensiones globales, variando en sus manifestaciones específicas, pero está unificada por la lógica de estigmatizar al musulmán como “otro” peligroso o inferior.

Cabe aclarar que “prejuicio antimusulmán” es una expresión equivalente a islamofobia, frecuentemente usada en la literatura para enfatizar que el núcleo del problema son el prejuicio y la discriminación (más que un miedo irracional, como sugiere el sufijo -fobia). Algunos autores emplean también términos como “anti-Muslim racism” (racismo antimusulmán) o “anti-Islamic bias”, pero todos apuntan a la misma realidad: la estigmatización sistemática de personas por su adscripción (real o supuesta) al islam (Meer 2009; Selod & Embrick 2013). Por tanto, en este informe utilizaremos islamofobia y prejuicio antimusulmán indistintamente, entendiendo ambos como modalidades específicas de racismo cultural y religioso.

5. Confluencias y confusiones en discursos contemporáneos

A pesar de ser conceptos distintos, el antisemitismo, el antisionismo y la islamofobia (junto con sus afines, el antijudaísmo y el prejuicio antimusulmán) a menudo se entrelazan en los discursos públicos (*online* y *offline*) de manera compleja. En la arena política y mediática actual, es común observar tanto alianzas inesperadas como confusiones intencionales entre estos marcos conceptuales, lo que dificulta el debate franco sobre el racismo y los derechos humanos.

Una intersección histórica significativa es la raíz común que comparten el antisemitismo y la islamofobia en la construcción de la identidad europea. Diversos historiadores y filósofos han señalado que, aunque el judaísmo y el islam se percibieron tradicionalmente de forma distinta —los judíos como un enemigo interno teológico y los musulmanes como un enemigo externo militar—, ambos grupos fueron definidos en Europa como el “otro” frente a una cristiandad homogénea (Spektorowski 2024). Es decir, la idea de una “Europa cristiana exclusiva” generó, a lo largo de los siglos, la exclusión tanto de judíos como de musulmanes, considerándolos ajenos a la identidad nacional o civilizatoria europea. Esta raíz compartida explica ciertas analogías: por ejemplo, se ha comparado la figura del “judío” en el imaginario medieval (acusado de deicidio y asociado a estereotipos de avaricia y conspiración) con la del “musulmán” en el imaginario contemporáneo (acusado de terrorismo y asociado al fanatismo y la barbarie). Ambos prejuicios sirvieron para definir una otredad contra la cual Europa afirmaba sus valores (ya fuesen cristianos en la Edad Media o secular-nacionales en la era moderna) (Spektorowski 2024).

No obstante, las dinámicas recientes también evidencian divergencias. Después del Holocausto, el antisemitismo explícito perdió legitimidad pública en Occidente, mientras que la islamofobia ha crecido sin un tabú social equivalente. En la Europa del siglo XXI, algunos grupos de extrema derecha se declaran defensores de la herencia “judeocristiana” de Occidente, mostrando incluso simpatía hacia Israel, pero esto suele ser estratégico: aprovechan la aceptación de la memoria del Holocausto para presentarse como filosemitas, al mismo tiempo que promueven el odio contra los inmigrantes musulmanes (Spektorowski 2024). Según este mismo autor, dicho fenómeno —a veces asociado a marcos de “choque de civilizaciones”— legitima nuevas exclusiones en nombre de proteger a Europa de un supuesto antisemitismo “importado”, a la vez que refuerza discursos antiislámicos.

Otra convergencia problemática ocurre cuando cualquier crítica a Israel es tildada de antisemitismo. De ahí que “rechazar la confusión entre antisionismo y antisemitismo sea fundamental para promover una educación antiopresiva” (Jaffee 2024: 30). La definición de antisemitismo promovida por la IHRA, adoptada en varios países, incluye ejemplos relacionados con Israel que han sido criticados por su ambigüedad; su uso expansivo puede deslegitimar críticas a los derechos humanos y silenciar el activismo, especialmente cuando lo lideran comunidades musulmanas (Deckers & Coulter 2022; Klug 2013). Esto produce un efecto de enfriamiento (*chilling effect*) en universidades, parlamentos y medios, donde activistas propalestinos son acusados de incitar al odio antijudío simplemente por emplear lemas antisionistas o por denunciar prácticas gubernamentales (Deckers & Coulter 2022; Klug 2013).

En paralelo, cabe notar que el prejuicio antimusulmán a menudo no recibe la misma atención mediática ni política que el antisemitismo, a pesar de estar igualmente extendido. En algunos países occidentales existe una gran sensibilidad (correcta) ante cualquier indicio de antisemitismo, pero una tolerancia relativa hacia comentarios islamófobos en la prensa o en políticas estatales que afectan desproporcionadamente a los musulmanes; esta jerarquización de la discriminación agrava las tensiones intercomunitarias (Garner & Selod 2015; Meer 2009).

6. Islamofobia como racismo estructural e instrumentalización política

La islamofobia, entendida como racismo antimusulmán, constituye hoy una forma de racismo estructural comparable en alcance y gravedad a otras formas históricas de discriminación. Esto significa que no se trata solo de prejuicios individuales, sino de un entramado de prácticas institucionales, narrativas estatales y actitudes socialmente normalizadas que sitúan sistemáticamente a las personas musulmanas (o percibidas como tales) en una posición de desventaja, de sospecha o de marginación (Garner & Selod 2015; Selod & Embrick 2013).

Un rasgo definitorio de la islamofobia contemporánea es su institucionalización. Tras el 11-S, muchos estados occidentales implementaron leyes antiterroristas y medidas de seguridad que, bajo la apariencia de neutralidad, se dirigieron casi en exclusiva a vigilar, controlar o reprimir a poblaciones musulmanas, tanto ciudadanas como migrantes. Este sesgo se refleja en “listas de vigilancia”, perfiles raciales en las fronteras, espionaje policial de mezquitas y comunidades, e incluso en programas de “prevención de la radicalización” (p. ej., *Prevent* en el Reino Unido) aplicados de forma desproporcionada (Kundnani 2012; Selod & Embrick 2013). Tales políticas envían el mensaje de que ser musulmán constituye, en sí mismo, un factor de riesgo para la seguridad nacional. En el ámbito político, líderes populistas y partidos nacionalistas han explotado abiertamente sentimientos islamófobos para ganar apoyo, presentándose como defensores de la nación frente a una amenaza musulmana, interna o externa; este clima ha impulsado restricciones específicas a símbolos, prácticas y a la presencia pública musulmana (Bakali 2024; Garner & Selod 2015).

Además de ser estructural, la islamofobia ha sido instrumentalizada políticamente en múltiples contextos para apuntalar conflictos militares, reprimir a minorías o desacreditar adversarios ideológicos, operando como una narrativa que cohesiona a los públicos bajo el miedo e inhibe las demandas de derechos (Bakali 2024; Deckers & Coulter 2022). Reconocer la islamofobia como racismo estructural implica también entender que combatirla requiere más que la tolerancia interpersonal; exige cambios en las políticas estatales, los marcos legales y las narrativas públicas para dismantelar la arquitectura de exclusión que la sustenta (Garner & Selod 2015; Meer 2009).

7. El caso de Gaza: genocidio, islamofobia y silenciamiento del activismo propalestino

La reciente guerra en Gaza y sus repercusiones internacionales ilustran cómo interactúan estos conceptos y cómo se emplean para moldear la opinión pública y acallar voces disidentes. Un conjunto creciente de trabajos sobre genocidio ha argumentado que las políticas y actos en Gaza encajan en la definición legal de genocidio, atendiendo tanto a patrones de conducta (matanza masiva, destrucción de infraestructura vital e imposición de condiciones de vida incompatibles con la supervivencia) como a indicios de intención (Hurwitz & White 2024; El-Affendi 2024; Verdeja 2025). En esta línea, análisis recientes en ciencia política y relaciones internacionales discuten las implicaciones jurídicas y políticas de dichos hallazgos, así como su impacto en la justicia penal internacional (Brooks & Griffiths 2025; Jacob 2024; Nijim 2024).

En este contexto, las redes sociales digitales constituyen un frente adicional del conflicto. Diversos estudios sobre cibertropas y propaganda computacional han mostrado cómo gobiernos, actores militares y organizaciones políticas utilizan campañas coordinadas, bots y cuentas falsas para moldear la percepción internacional de los conflictos armados, desacreditar fuentes de información alternativas y amplificar narrativas que deshumanizan al enemigo (Bradshaw & Howard 2021; Woolley & Howard 2016). En el caso de Gaza, ello implica que los picos de discurso de odio antimusulmán y antijudío que circulan en redes sociales no solo reflejan reacciones espontáneas de los usuarios, sino también la edición algorítmica de los contenidos y la difusión planificada de mensajes por parte de actores organizados. Esta dimensión comunicacional planificada refuerza la necesidad de leer los datos de odio en clave estructural y estratégica, más allá de la mera suma de expresiones individuales.

Frente a acusaciones tan graves, se ha observado una estrategia de confluencia: desplazar la atención hacia el antisemitismo y etiquetar como extremistas a quienes denuncian (de forma vehemente) las violaciones, lo que produce un efecto de enfriamiento/institucionalización (“*wokismo*”) sobre el activismo propalestino, especialmente cuando lo lideran comunidades musulmanas (Deckers & Coulter 2022; Klug 2013). En campus universitarios y espacios públicos, este marco ha favorecido suspensiones, censuras y campañas de difamación, en ocasiones justificadas mediante definiciones expansivas de antisemitismo y tropos islamófobos (Deckers & Coulter 2022; Klug 2013).

Por otro lado, la islamofobia, como combustible del conflicto, opera a través de la deshumanización de los habitantes de Gaza: representaciones que los presentan como intrínsecamente fanáticos o “terroristas” (o, al menos, corresponsables por apoyar y/o votar a Hamás - حماس) reducen la empatía pública, rebajan el umbral de tolerancia frente a la devastación y normalizan políticas de castigo colectivo (Garner & Selod 2015; Meer 2009). En contraste, la solidaridad internacional liderada por comunidades musulmanas ha sido clave para visibilizar la tragedia, aunque a menudo sea el principal blanco de las estrategias de silenciamiento (Bakali 2024).

8. Metodología

8.1. Enfoque y diseño

Para este artículo adoptamos un diseño mixto que integra (i) una fase histórico-conceptual-político-legislativa, en la que delimitamos el antisemitismo y el resto de las acepciones y conceptualizaciones en su genealogía moderna y en las definiciones operativas debatidas en la literatura especializada, y (ii) una fase empírica computacional a gran escala.

La primera fija los criterios normativos para distinguir la incivildad o toxicidad del odio antijudío y del odio antimusulmán (islamofobia) y separa la crítica política legítima de la hostilidad ilegítima (según el marco teórico desarrollado en el artículo). La segunda operacionaliza esas distinciones mediante modelos lingüísticos y métricas automáticas y computacionales comparables entre algunas de las principales redes sociales y los colectivos vulnerables y estigmatizados estudiados.

8.2. Fase histórico-teórico-conceptual

Esta fase histórico-conceptual se diseñó con un propósito estrictamente metodológico: destilar el acervo teórico-normativo sobre antisemitismo y antiislamismo en criterios observables para la anotación humana y la posterior modelización automática (segunda fase). Para ello, partimos de la distinción analítica entre término y concepto de antisemitismo —su configuración moderna y su mutabilidad histórica— para fijar el alcance del fenómeno que medimos (Engel 2008; Dahlström 2022; [Consonni 2022](#); Wistrich 1991), traducéndolo a un codebook operativo alineado con la Working Definition de la IHRA (2016), con las precisiones de la JDA (2021) y el Nexus Document (2021), incorporando también las advertencias sobre posibles usos politizados del marco de la IHRA ([Stern 2020](#)). En paralelo, garantizamos la coherencia normativa con los estándares europeos y de soft law: las guías de [OSCE/ODIHR \(2014\)](#), la Recomendación General n.º 9 (revisada) de ECRI (2021), la Decisión Marco 2008/913/JAI del Consejo de la UE (Council of the EU 2008), la Resolución del Parlamento Europeo sobre antisemitismo (European Parliament 2017), la Estrategia de la Comisión Europea 2021–2030 (European Commission 2021) y las Conclusiones del Consejo de la UE (Council of the EU 2022).

El resultado de esta primera fase es un *codebook* histórico-conceptual reproducible que: (i) formaliza los criterios positivos de antisemitismo (culpabilización colectiva; tropos de poder oculto/control de medios o finanzas; “doble lealtad”; negación/banalización de la Shoá; negación del derecho del pueblo judío a existir; analogías nazis deshumanizadoras) y los criterios negativos (crítica factual sin estereotipos ni deslegitimación); (ii) define el anti-islamismo conforme a OSCE/ODIHR/ECRI; y (iii) mapea tipologías y niveles psico-sociales (primario/afectivo, conativo/activo, secundario/latente; cf. Kovács & Fischer 1994) a indicadores computacionales. Este mapa conceptual enriquece el entrenamiento de los clasificadores, guía la interpretación de métricas automáticas (p. ej., “ataque a la identidad”, “insulto”) y sostiene, a lo largo de toda la cadena metodológica, la separación analítica entre la incivildad tóxica, la crítica política y el odio dirigido contra las comunidades judías y musulmanas.

8.3. Corpus y recolección de datos: fase empírica computacional

El punto de partida de nuestro análisis empírico es el corpus de contenidos recopilado en el proyecto COIN: 798.475 mensajes públicos en español, publicados entre el 1 de abril y el 13 de agosto de 2024, en X, Facebook, Instagram y TikTok. Este censo se generó mediante consultas léxicas por colectivo vulnerable (migrantes, personas LGBTIQ, musulmanes, judíos, gitanos) y operadores booleanos, restringiendo la búsqueda a contenidos geolocalizados en España o asociados a perfiles con localización declarada en España y a la detección automática del idioma “es”. A partir de este universo común se han derivado distintos casos de estudio; en este artículo trabajamos únicamente con las puntuaciones de antisemitismo y antiislamismo asignadas por los clasificadores a dichos mensajes. Los corpus anotados de 5.044 tuits (antisemitismo) y 14.367 tuits (islamofobia) constituyen los conjuntos de entrenamiento y evaluación de los modelos, no el universo analizado. La mayor cantidad de ejemplos anotados de islamofobia se debe a decisiones prácticas del proceso de anotación en el proyecto COIN (distintas fases y prioridades de trabajo) y no debe interpretarse como un indicador directo de prevalencia comparada.

La recolección se efectuó mediante las APIs oficiales de X Developer y TikTok Research, y mediante CrowdTangle para Facebook e Instagram. En X se aplicó un muestreo estratificado en franjas de dos horas (limitación del plan básico de la API), lo que implica que el corpus de X es amplio, pero no exhaustivo; algunos días presentan lagunas puntuales debido a incidencias técnicas. En TikTok se recuperó la descripción textual completa del vídeo (incluidos los hashtags en ese campo), pero no se incluyeron los comentarios de los usuarios. En Facebook/Instagram, CrowdTangle cubre páginas, grupos y perfiles públicos monitorizados por la herramienta, por lo que quedan fuera los perfiles estrictamente personales; esta restricción de cobertura se considera una limitación externa del estudio. En todas las plataformas se eliminaron los duplicados exactos y las republicaciones idénticas para evitar la sobrerepresentación de cadenas virales.

La selección temática se basó en diccionarios léxicos por colectivo, elaborados colaborativamente por el equipo COIN a partir de la literatura previa sobre discurso de odio, de glosarios de organizaciones especializadas y de exploraciones inductivas del propio corpus. Se trata de diccionarios abiertos que incluyen tanto denominaciones neutras (“judío”, “musulmana”, “palestinos”, “sin papeles”) como etnónimos peyorativos y eufemismos habituales; un ejemplo ilustrativo del caso judío sería la combinación de “judío/judía/judíos”, “sionista/sionistas” y variantes de expresiones históricamente antisemitas. La versión completa de estos listados se documenta en el material suplementario del proyecto para garantizar la trazabilidad y la replicabilidad (<https://coinusal4excellence.com/coin-tecnicas-de-seguimiento-e-investigacion/>). La interseccionalidad se modeló identificando mensajes que contienen términos presentes en ≥ 2 diccionarios del mismo texto; es decir, las coocurrencias entre colectivos se determinan por el léxico y no por la salida de los clasificadores.

En cuanto a la validez interna y externa de los datos, el criterio de muestreo introduce diversas limitaciones. Al tratarse de mensajes filtrados por diccionarios, no captamos todo el universo de odio

ni de incivilidad, sino solo el segmento que coocurre con los términos seleccionados. Además, el período de observación coincide con una fase de extrema violencia en Gaza y de alta conflictividad geopolítica, en la que es razonable suponer tanto un aumento “orgánico” del volumen de mensajes como la posible amplificación artificial por parte de campañas coordinadas (cibertropas, bots, cuentas institucionales altamente activas). No contamos con una detección sistemática de bots ni excluimos cuentas institucionales: analizamos todas las cuentas presentes en el corpus como parte del ecosistema comunicativo real. Esto refuerza el carácter contextual y acotado de los resultados: describen el comportamiento de estas plataformas en ese intervalo específico y bajo esas condiciones, más que una “medida verdadera” del antisemitismo o de la islamofobia en la sociedad española.

Para la operacionalización del odio, entrenamos cinco clasificadores/modelos independientes (xenofobia, LGBTIQfobia, islamofobia, antisemitismo, antigitanismo), todos ellos obtenidos mediante fine-tuning de BERT-base-multilingual-uncased sobre corpus anotados en español y particiones 60/20/20 (entrenamiento/validación/prueba). En este artículo, utilizamos exclusivamente los modelos de antisemitismo (5.044 tuits anotados) e islamofobia (14.367 tuits anotados). En ambos casos, la clase positiva (mensajes de odio) es minoritaria respecto a la negativa, en línea con la baja prevalencia de odio explícito; la distribución exacta por clase y las métricas detalladas (precisión, exhaustividad y F1 por clase, así como micro y macro-F1) se presentan en una tabla específica de rendimiento del modelo (WP4 de COIN). Los datos se anotaron siguiendo el codebook histórico-conceptual descrito en la sección 8.2, que instruye a los codificadores a identificar tanto formas explícitas (insultos, deshumanización directa, negación de derechos) como implícitas (eufemismos, alusiones veladas, culpabilización colectiva sin mención directa del grupo), con ejemplos contextualizados de cada categoría. Los modelos devuelven una probabilidad continua entre 0 y 1 por mensaje, que reportamos como porcentaje para facilitar la interpretación.

Respecto a la dimensión lingüística, se procesaron mensajes detectados como españoles; al restringir la recolección a contenidos emitidos desde España (cuando fue posible), cabe esperar una predominancia de variedades peninsulares, aunque no puede descartarse la presencia de formas dialectales latinoamericanas, especialmente en cuentas institucionales o mediáticas. No se tradujeron los textos antes de su análisis. Para la medición complementaria de “toxicidad”, empleamos la Perspective API en su versión en español, subdividida en seis atributos (TOXICITY, SEVERE_TOXICITY, IDENTITY_ATTACK, INSULT, PROFANITY, THREAT) y evaluados en una escala de 0 a 1. Esta API es un modelo entrenado principalmente con datos de medios digitales, y su documentación reconoce posibles sesgos por idioma y contexto (no realizamos una validación cruzada independiente, por lo que sus puntuaciones se utilizan aquí como indicadores comparativos y exploratorios, no como criterio normativo definitivo). El umbral de 0,5 se adopta para binarizar variables en algunos análisis (p. ej., “toxicidad alta” $\geq 0,5$), ya que representa el punto medio de la escala de probabilidad; los resultados principales se reportan siempre como valores continuos para evitar artefactos de corte.

Para finalizar, las variables de difusión se armonizaron para permitir comparaciones razonables entre plataformas. Se trabajó con ‘likes’, comentarios y una suma estandarizada de ambas (engagement_

sum), transformados en puntuaciones z por plataforma. En el caso de TikTok, las “vistas” presentan órdenes de magnitud mucho mayores que las interacciones observadas en otras redes; por ello, se analizaron solo de manera descriptiva en TikTok, sin compararlas directamente con las métricas de engagement de X, Facebook o Instagram. De este modo, las comparaciones entre plataformas se basan en medidas relativas y no en cifras absolutas difícilmente homologables.

8.4. Alcance concreto de la investigación

Este artículo presenta una investigación de carácter observacional, descriptivo-analítico y no experimental, centrada en el caso de España y en un período temporal acotado. El universo empírico lo constituyen mensajes públicos en español publicados entre el 01/04 y el 13/08/2024 en cuatro redes sociales digitales (X, Facebook, Instagram y TikTok); la unidad de análisis es el mensaje individual, no las personas usuarias ni las comunidades. A partir de este universo, trabajamos con un corpus filtrado mediante diccionarios léxicos, colectivos y operadores booleanos, de modo que los resultados se refieren exclusivamente al discurso producido en dichas plataformas, en ese intervalo temporal y bajo esos criterios de selección.

El alcance sustantivo del artículo se circunscribe a (i) la operacionalización del antisemitismo y el antiislamismo en un *codebook* aplicable al odio online; (ii) la medición comparada de estos fenómenos en las cuatro plataformas analizadas; y (iii) la exploración de su relación con indicadores de toxicidad y difusión. No pretendemos estimar la prevalencia del antisemitismo o la islamofobia en la sociedad española en su conjunto, ni en todas las esferas digitales (p. ej., mensajería privada), sino ofrecer una radiografía acotada de cómo se configuran estos discursos en redes sociales abiertas bajo las condiciones indicadas y analizadas. En consecuencia, cualquier generalización debe entenderse limitada al contexto, al período y a las plataformas estudiados, así como a los márgenes de error asociados a los modelos de clasificación empleados.

9. Resultados: del andamiaje histórico-conceptual a la medición computacional del antisemitismo (y del antiislamismo) en redes

El andamiaje histórico-conceptual desarrollado en las secciones previas se traduce en variables observables en esta sección de resultados. A partir del *codebook* normativo (IHRA, JDA, Nexus y marcos europeos), se entrenaron clasificadores específicos de antisemitismo y antiislamismo, y se establecieron criterios para distinguir entre crítica política legítima y discurso de odio dirigido a judíos o musulmanes. En los apartados siguientes, presentamos los resultados de aplicar dichos modelos a los 798.475 mensajes del corpus, combinando (i) las probabilidades específicas de odio por colectivo generadas por los clasificadores BERT y (ii) los seis atributos de incivildad lingüística (toxicidad, toxicidad severa, insultos, blasfemia, ataque a la identidad, amenazas) proporcionados por la API de Perspective.

Los modelos de antisemitismo e islamofobia se entrenaron sobre un corpus anotado siguiendo el protocolo descrito en el apartado 8.2 (doble codificación, resolución de desacuerdos y cálculo del α de Krippendorff). El modelo de antisemitismo alcanza $\alpha=0,79$ y un F1 global de 0,89; el de islamofobia obtiene un F1 de 0,86. Para no sobrecargar el texto, las métricas desagregadas (micro-/macro-F1 y F1 por clase) se reportan en una tabla técnica específica (en el WP4 del proyecto COIN) y aquí se emplea “F1” como abreviatura del rendimiento medio del modelo. Todas las probabilidades que se presentan a continuación deben entenderse como salidas de los modelos, no como medidas “censales” del antisemitismo o de la islamofobia en la sociedad en su conjunto.

9.1. Intensidad de odio y perfiles de toxicidad por plataforma

Las probabilidades medias de antisemitismo e islamofobia muestran diferencias marcadas entre las plataformas (Tabla 1). Para el antisemitismo, la probabilidad media estimada por el modelo es del 33,4% en X, del 68,4% en Facebook, del 31,3% en Instagram y del 90,9% en TikTok. Para islamofobia, los valores correspondientes son 31,7%, 24,3%, 7,4% y 27,2%. En términos de toxicidad lingüística, los atributos de Perspective con valores más altos son “insulto” y “ataque a la identidad”, mientras que las amenazas presentan valores medios claramente inferiores. Con respecto a los mensajes clasificados como tóxicos sobre judíos, los promedios de toxicidad, insultos y ataques a la identidad por plataforma son: X (41 %, 38 %, 36 %), Facebook (16 %, 10 %, 13 %), Instagram (11 %, 7 %, 10 %) y TikTok (17 %, 12 %, 16 %). Mientras que para los mensajes tóxicos sobre musulmanes quedan así: X (34 %, 31 %, 28 %), Facebook (5 %, 4 %, 3 %), Instagram (4 %, 3 %, 2 %) y TikTok (7 %, 4 %, 5 %).

Estas diferencias no deben leerse como una “clasificación” de plataformas en términos absolutos de odio, sino como indicadores relativos, condicionados por el diseño muestral y por el propio comportamiento de los modelos. En Facebook e Instagram, trabajamos únicamente con páginas, grupos y perfiles públicos monitorizados por CrowdTangle, lo que tiende a sobrerrepresentar cuentas institucionales o verificadas, previsiblemente menos propensas a insultos directos. En X, por el contrario, la muestra incluye usuarios anónimos, cuentas políticas y perfiles muy activos en controversias, lo que se asocia con un discurso más crudo y polarizado. En TikTok, solo se analizó la descripción textual de los videos, generalmente muy breve y con alta densidad de hashtags; es posible que el modelo BERT otorgue un peso especial a determinados tokens repetitivos o etiquetas polémicas, lo que infla la probabilidad media en comparación con textos más largos. Por tanto, los valores numéricos son informativos en cada plataforma, pero su comparación directa entre redes debe hacerse con cautela.

En todas las redes, las seis dimensiones de Perspective están fuertemente correlacionadas entre sí (por ejemplo, toxicidad e insultos presentan $r \approx .93-.98$, $p < .01$, según la plataforma), lo que indica que los diferentes indicadores captan un “núcleo” común de incivilidad y toxicidad léxica: cuando aparecen insultos, suele aumentar también la valoración de la toxicidad global y del ataque a la identidad.

9.2. Desacople parcial “toxicidad–odio” y especificidad de plataforma

Al analizar la relación entre la toxicidad lingüística y el odio específico, observamos un “acoplamiento” parcial. Aquí entendemos por acoplamiento la correlación de Pearson entre la toxicidad global (media de los seis atributos de Perspective) y la probabilidad de odio asignada por el clasificador. En el caso del antisemitismo, las correlaciones son $r=.32$ (X), $-.02$ (Facebook), $.20$ (Instagram) y $.04$ (TikTok); para la islamofobia, $r=.28$, $.14$, $.11$ y $.04$, respectivamente ($p<.01$, salvo los coeficientes próximos a cero).

En X, la asociación moderada indica que buena parte de los mensajes con alta probabilidad de odio también presenta un lenguaje explícitamente agresivo. En Facebook e Instagram, en cambio, encontramos un número apreciable de mensajes que el modelo clasifica como antisemitas o islamófobos con niveles bajos de toxicidad textual. Se trata de mensajes que recurren a eufemismos, marcos conspirativos o imputaciones colectivas, sin insultos directos.

También se observan casos inversos [alta toxicidad con baja probabilidad de odio] — en los que se producen insultos genéricos sin referencia a colectivos protegidos (por ejemplo, descalificaciones entre usuarios en discusiones políticas). Estos ejemplos ilustran por qué la toxicidad es un indicador complementario, pero no suficiente, para captar el antisemitismo o la islamofobia: el odio puede expresarse con un lenguaje relativamente neutro y, a la inversa, no toda incivilidad o toxicidad constituye un discurso de odio. En TikTok, el desacoplamiento se intensifica debido a la brevedad de las descripciones y al peso del componente audiovisual, que no está presente en nuestro análisis textual [Tabla 1].

Tabla 1. Antisemitismo y antiislamismo por plataforma: intensidad de odio, toxicidad y acoplamiento

Plataforma	Antisemitismo Odio (%)	Toxicidad (%)	At. identidad (%)	r(tox, odio)	Islamofobia Odio (%)	Toxicidad (%)	At. identidad (%)	r(tox, odio)
X	33,4	41	36	.32	31,7	34	28	.28
Facebook	68,4	16	13	-.02	24,3	5	3	.14
Instagram	31,3	11	10	.20	7,4	4	2	.11
TikTok	90,9	17	16	.04	27,2	7	5	.04

Notas: Odio = probabilidad media de discurso de odio (0–1) expresada en %; toxicidad y “ataque a la identidad” = promedios de Perspective en %; r = Pearson entre toxicidad y odio ($p<.01$ salvo valores cercanos a 0).

Fuente: elaboración propia.

9.3. Interseccionalidad del odio y difusión

Un 2,5 % de los mensajes del corpus (19.896 posts) menciona al menos dos colectivos vulnerables; entre ellos, la combinación más frecuente es Islam+Judaísmo (6.397 mensajes), seguida de cruces como “musulmanes+migrantes” o “judíos+LGBTIQ”. X concentra la mayor proporción de mensajes interseccionales (4,5 %) y TikTok, la menor (1,1 %).

Los mensajes interseccionales presentan, en promedio, niveles de toxicidad significativamente mayores que los de los mensajes que se refieren a un solo colectivo (pruebas t de Student; tamaños de efecto pequeños a medianos, con un *d* global de aproximadamente 0,40). Esto se aprecia con especial claridad en Facebook e Instagram, mientras que en X y TikTok las diferencias son más moderadas. A modo de ejemplo (paráfrasis), un mensaje que combina referencias a “musulmanes inmigrantes” y “judíos globalistas” tiende a acumular insultos, imputaciones de amenazas y acusaciones de control económico o político, lo que se traduce en puntuaciones elevadas tanto de toxicidad como de probabilidad de odio en ambos clasificadores.

En términos de difusión, X es la única plataforma donde los mensajes interseccionales obtienen, en promedio, más interacción (*likes*, comentarios y *engagement* combinado) que los mensajes no interseccionales, mientras que, en Facebook y TikTok, la comención de colectivos se asocia a una ligera reducción del *engagement*. En el conjunto del corpus, las diferencias en los *likes* no alcanzan significación estadística, pero sí se observan diferencias significativas en el número de comentarios, lo que sugiere que los mensajes interseccionales generan más controversia discursiva que adhesión explícita.

9.4. Síntesis interpretativa

A la luz de estos resultados pueden extraerse tres conclusiones principales: (i) Toxicidad $\neq$ odio, (ii) Especificidad de plataforma y (iii) Efecto interseccional.

- (i) La toxicidad y el odio no son iguales ni equivalentes: la toxicidad identifica con precisión insultos, blasfemias y ataques directos a la identidad, pero es un indicador parcial del antisemitismo y del antiislamismo, especialmente en plataformas donde el odio se expresa de manera más implícita o argumentativa. Desde el punto de vista metodológico, esto refuerza la necesidad de modelos que integren señales semánticas y pragmáticas (y, en su caso, multimodales) además de la detección de lenguaje malsonante.
- (ii) La plataforma condiciona la forma del odio: las diferencias entre X, Facebook, Instagram y TikTok reflejan tanto prácticas comunicativas diversas como sesgos de cobertura y de modelo. X, con usuarios mayoritariamente anónimos y altamente politizados, presenta un acoplamiento relativamente alto entre la toxicidad y el odio. Facebook e Instagram, muestreados mediante CrowdTangle, incluyen, sobre todo, páginas y perfiles públicos, donde el discurso hostil tiende a adoptar formas más contenidas e implícitas. En TikTok, los textos breves y cargados de *hashtags* hacen que el modelo sea especialmente sensible a ciertas combinaciones léxicas. Por ello, las comparaciones interplataforma deben leerse en clave relativa y estar acompañadas de las limitaciones metodológicas señaladas.
- (iii) El odio interseccional presenta una mayor densidad de toxicidad: las comenciones de colectivos (por ejemplo, “migrante musulmán”, “judío LGBTIQ”) se asocian con una mayor toxicidad y, en el caso de X, con una mayor interacción. Analíticamente, esto sugiere que la hostilidad hacia judíos y musulmanes no opera en compartimentos estancos, sino que se entrelaza con otros ejes de estigmatización (migración, género, sexualidad), lo cual resulta relevante tanto para la detección automática (características cruzadas, criterios combinados) como para el diseño de políticas públicas de prevención.

En conjunto, los hallazgos empíricos son coherentes con la evolución histórica del antisemitismo y del prejuicio antimusulmán, tal como se describe en el marco teórico. En muchos casos, el odio se expresa hoy mediante marcos político-ideológicos, alusiones a la seguridad o narrativas civilizatorias, más que mediante insultos abiertos. De ahí la importancia de utilizar definiciones normativas explícitas (IHRA, JDA, Nexus) como referencia para distinguir la crítica legítima de la hostilidad dirigida a judíos y musulmanes, y de adoptar una medición multidimensional sensible tanto a las especificidades de cada plataforma como a las formas interseccionales del odio.

10. Discusión

Anclar el análisis computacional y digital en la genealogía que distingue la judeofobia (*Judenhass*) del antisemitismo moderno y, dentro de los marcos operativos de las IHRA–Nexus–JDA, ha resultado decisivo para leer y comprender con precisión nuestros hallazgos, sin confundir la incivildad con el odio dirigido. El dato central —el desacople parcial entre toxicidad léxica y probabilidad de odio, especialmente en Facebook ($r \approx -.02$ para el antisemitismo) y TikTok ($r \approx .04$)— muestra que una fracción relevante del neoantisemitismo y del antiislamismo contemporáneos se articula mediante eufemismos, insinuaciones y marcos conspirativos (poder oculto, doble lealtad, negación del derecho colectivo a existir). Es decir, en el mismo terreno semántico que las definiciones operativas buscan distinguir de la crítica política legítima dirigida a Israel o al sionismo. En contrapartida, el núcleo/binomio “insulto–ataque a la identidad” covaría con la alta consistencia interna, lo que confirma que la incivildad es mensurable y útil como señal, pero insuficiente para captar el odio cuando este se enmascara en gramáticas alusivas o implícitas.

Este patrón se entiende plenamente cuando se ubica sobre el telón de fondo que hemos descrito en la sección 2: las redes sociales digitales no son un simple canal neutro, sino un enclave sociotécnico, económico y geoestratégico donde los algoritmos de recomendación, las métricas de engagement y los modelos de negocio basados en la atención privilegian el contenido polarizante. La personalización algorítmica, las cámaras de eco y las *filter bubbles* reducen la exposición a perspectivas disonantes y aumentan la probabilidad de que el antisemitismo y el antiislamismo circulen en entornos ideológicamente afines, donde los tropos hostiles pueden expresarse con escasas restricciones normativas. Además, la presencia de cibertropas, *bots*, cuentas coordinadas y campañas de desinformación implica que una parte del odio que medimos no es solo “orgánico” (expresión espontánea de prejuicios), sino también un producto estratégicamente amplificado para moldear la opinión pública y tensionar la deliberación democrática. De modo que los resultados empíricos obtenidos deben leerse no solo como un reflejo de actitudes sociales, sino también como el efecto combinado de los sesgos estructurales de las plataformas y de los usos políticos y militares de la comunicación digital.

La especificidad de la plataforma tensiona aún más las lecturas de nuestro estudio. X concentra el mayor acoplamiento “léxico–odio” y la mayor intensidad de hostilidad verbal; Facebook atenúa la incivildad, pero exhibe probabilidades elevadas de antisemitismo; Instagram permanece relativamente

bajo en ambos ejes, y TikTok combina máximos de antisemitismo con una baja señal tóxica textual, coherente con su naturaleza multimodal y con la limitación del corpus a descripciones de vídeo. En conjunto, el patrón sugiere que allí donde las infraestructuras sociales y de contenido incentivan la codificación indirecta (humor, insinuación, *dog-whistles*), el antisemitismo y el antiislamismo pueden florecer sin insultos explícitos, lo que refuerza la necesidad de criterios semánticos y contextuales.

El contexto Israel-Gaza actúa como catalizador coyuntural de picos de hostilidad, y el antisionismo opera con frecuencia como vector de tropos clásicos: deslegitimación del derecho colectivo del pueblo judío, doble lealtad, culpabilización colectiva de la diáspora y analogías con el nazismo como recurso deshumanizador. Nuestro marco analítico permite diferenciar este desplazamiento del disenso legítimo — crítica basada en hechos, políticas de Israel, apoyo a los derechos palestinos o estrategias no violentas como BDS— cuando, como hemos mencionado en varias ocasiones en este artículo, no recurre a estereotipos ni a la negación del pueblo judío. En paralelo, el antiislamismo se expresa como “esencialización civilizacional” y la atribución de culpas colectivas a los musulmanes como tales, con patrones retóricos que, de nuevo, no requieren un léxico abiertamente ofensivo para causar daño simbólico.

La interseccionalidad agrava el cuadro/encuadre. Las comenciones del islam y del judaísmo —aunque minoritarias (~2,5%)— presentan una toxicidad media significativamente superior y, en X, presentan una mayor interacción. Esta convergencia de sesgos sugiere transferencias de odio entre colectivos, en los que episodios geopolíticos polarizados reactivan repertorios de larga duración y se retroalimentan mediante lógicas algorítmicas de amplificación. En términos de tipologías psicosociales, observamos una convivencia de componentes afectivos (rechazo emocional explícito) e indirectos (negacionismo, comparaciones históricas instrumentales) que se encajan en configuraciones secundarias y latentes del antisemitismo y del antiislamismo.

Desde el punto de vista metodológico, demostrar que “toxicidad $\neq$ odio” fuera de X tiene consecuencias de primera magnitud: las políticas de moderación basadas solo en palabras malsonantes o insultos omiten un volumen crítico de hostilidad codificada. Nuestros clasificadores por colectivo ($F1 \geq 0,86$; $\alpha = 0,79$ en antisemitismo), guiados por reglas positivas y negativas derivadas de IHRA-Nexus-JDA, reducen falsos positivos frente a la crítica dura y capturan formas implícitas de antisemitismo/antiislamismo, pero también muestran que los costes de error varían por plataforma: riesgo de falsos negativos en Facebook/TikTok y de escalada léxica en X. El *pipeline* —codebook histórico-conceptual → anotación guiada → modelos por grupo → lectura conjunta con métricas de toxicidad → análisis interseccional y de difusión— se confirma como transferible y como base para auditorías externas y longitudinales.

En el plano normativo y operativo, los resultados avalan un enfoque de “medición multidimensional sensible a la plataforma”: se debe (i) combinar atributos de toxicidad con detectores semánticos específicos (antisemitismo, antiislamismo); (ii) incorporar módulos interseccionales que identifiquen combinaciones identitarias; (iii) aplicar umbrales y criterios diferenciados por entorno (X vs. Facebook/

Instagram vs. TikTok); y (iv) sostener la moderación con criterios públicos basados en IHRA/JDA/Nexus, tal como recomiendan OSCE/ODIHR, ECRI y la Estrategia UE 2021–2030, para proteger el disenso y, a la vez, mitigar el odio. Esto implica —más allá del filtrado léxico— la formación de (i) equipos de revisión para la deliberación conceptual, (ii) protocolos de apelación que evalúen la evidencia fáctica de la crítica política, y (iii) la monitorización de las crisis para anticipar escaladas en torno a Israel/Gaza.

La distinción entre antisemitismo, antijudaísmo (religioso-cultural), antiislamismo y antisionismo no es un matiz académico, sino la línea de flotación de cualquier política proporcional. Nuestro estudio muestra, al mismo tiempo, que no toda crítica a Israel o al sionismo es antisemitismo y que una parte del antisemitismo contemporáneo se expresa precisamente a través de vectores políticos que reciclan tropos clásicos en un entorno mediado por algoritmos, cámaras de eco y campañas coordinadas de desinformación. Ignorar esta doble evidencia —histórica y computacional, sociotécnica y semántica— conduce tanto a censuras indebidas como a zonas ciegas donde el odio prospera con baja toxicidad. La contribución de este trabajo consiste, precisamente, en ofrecer un puente operativo entre historia, teorías de las redes sociales, definiciones normativas y detección automática que permite moderaciones más precisas y garantistas, refuerza la protección efectiva de las comunidades judías y musulmanas y preserva el disenso democrático indispensable en el debate sobre Israel, Gaza y, en general, sobre la convivencia en sociedades plurales y mediadas por plataformas digitales [Tabla 2].

Tabla 2. Hallazgos, evidencias e implicaciones metodológicas y normativas del “antisemitismo”

Eje	Hallazgo clave	Evidencia/Indicadores	Implicación metodológica	Implicación normativa/política
Distinción conceptual	No confundir la incivilidad (toxicidad) con el odio dirigido (antisemitismo/antislamismo).	Desacople parcial tox-odio: p , η , $r \approx -0.02$ (Facebook) y $r \approx 0.04$ (TikTok) en antisemitismo; alta coherencia interna entre “insulto” y “ataque a la identidad”.	Métricas duales: atributos de Perspective + clasificadores por colectivo; lectura conjunta y no sustitutiva.	Criterios de IHRA/JDA/Nexus para distinguir la crítica legítima del odio y evitar la sobremoderación mediante un lenguaje duro, sin tropos.
Plataforma	El patrón varía por red.	X: mayor acoplamiento léxico-odio; Facebook: alta probabilidad de antisemitismo y baja toxicidad; Instagram: bajos en ambos; TikTok: máximos de antisemitismo y baja señal textual.	Umbrales y modelos sensibles a la plataforma; priorizar la multimodalidad en TikTok.	Políticas diferenciadas por entorno; auditorías por red; transparencia en el enforcement.
Contexto Israel-Gaza	Escala los picos de hostilidad y reactiva los tropos.	Antisionismo como vector de doble lealtad, deslegitimación del derecho colectivo y analogías con el nazismo.	Monitorización coyuntural y detección de tropos; listas de eventos para ventanas de riesgo.	Protocolos de crisis y alertas tempranas; protección de las comunidades durante las ventanas de escalada.
Antisemitismo “implícito”	Hostilidad codificada sin insulto frontal.	Eufemismos, insinuaciones y marcos conspirativos de baja toxicidad y alta probabilidad de generar odio (Facebook/TikTok).	Características semánticas y contextuales; reglas positivas y negativas del codebook.	Formación de revisores sobre los criterios de IHRA/JDA/Nexus; vías de apelación ante críticas de hecho.
Antislamismo	Esencialización civilizacional y culpabilización colectiva de los musulmanes.	Patrones retóricos de odio con baja toxicidad explícita; alineación moderada en X.	Clasificador específico de islamofobia + atributos de identidad.	Integrar los estándares de la OSCE/ODIHR-ECRI en las guías y en la recopilación de datos.
Interseccionalidad	La convergencia de sesgos intensifica el daño y, en X, la difusión.	$\approx 2.5\%$ de mensajes: comenciones de Islam+Judaísmo más frecuentes; mayor toxicidad media, más interacción en X.	Detectores interseccionales (features cruzadas, umbrales compuestos) y análisis de engagement.	Priorización de casos interseccionales en moderación y en apoyo a las víctimas.
Rendimiento modelos	Buena discriminación con marcos operativos.	F1_antisemitismo=0.89; F1_islamofobia=0.86; $\alpha=0.79$ (antisemitismo).	Entrenamiento guiado por reglas positivas y negativas (IHRA-JDA-Nexus); evaluación en la plataforma.	Estándares de auditoría externa y de reporting de FP/FN por colectivo y por red.
“Toxicidad ≠ odio”	La toxicidad es una señal útil pero insuficiente fuera de X.	Núcleo de insultos/identidad consistente; acoplamiento bajos con odio en Facebook/TikTok.	Pipeline híbrido: léxico + semántica + (cuando proceda) multimodalidad.	Evitar listas negras puramente léxicas; incorporar el contexto y el propósito en las decisiones.
Recomendaciones a plataformas	Moderación proporcional y garantista.	Hallazgos comparativos por red y comerciones.	Modelos híbridos y multimodales, umbrales por red, módulos interseccionales, rutas de apelación.	Alineación con la Estrategia de la UE 2021-2030, la Decisión Marco 2008/913/JAI y la Directiva 2012/29/UE.
Recomendaciones a reguladores	Medición y educación como palancas.	Necesidad de estadísticas comparables y alfabetización pública.	Observatorios y series temporales; codebooks abiertos y reproducibles.	Implementar las guías de la OSCE/ODIHR y del ECRI (2021); educación sobre la Shoá y los tropos conspirativos.
Limitaciones	Riesgo de subdetección multimodal y de sesgos de muestreo.	TikTok: solo descripciones; API de X con restricciones.	Incluir audio e imagen, ampliar las ventanas temporales y calibrar los costes de FP/FN.	Exigir un acceso seguro a los datos para la investigación y la evaluación independientes.

Fuente: elaboración propia.

11. Conclusiones

Este artículo parte de un objetivo general claro: analizar, en el contexto español de 2024, la presencia y la configuración de los discursos antijudíos y antiislamistas en redes sociales, así como sus implicaciones para la delimitación entre el discurso de odio y el disenso político legítimo. Para ello, formulamos tres objetivos específicos (OE1–OE3) y seis preguntas de investigación (P1–P6). En las páginas anteriores, hemos articulado un recorrido que abarca desde la genealogía histórico-conceptual hasta la modelización computacional. A la luz de los resultados obtenidos, en estas conclusiones podemos precisar cómo se responden esas preguntas y en qué medida se han alcanzado los objetivos propuestos.

En relación con la P1 y el OE1 (refinar la medición del antisemitismo contemporáneo), el trabajo demuestra que es posible construir clasificadores guiados por los criterios diferenciales de IHRA, Nexus y JDA que distingan de forma razonablemente fiable entre el antisemitismo y el antiislamismo, y entre la crítica política legítima y el antisemitismo. El modelo de antisemitismo (α de Krippendorff = 0,79; F1 global = 0,89) y el de islamofobia (F1 = 0,86) identifican con alta precisión los mensajes que contienen tropos antijudíos y antimusulmanes –culpabilización colectiva, poder oculto, doble lealtad, negación o banalización de la Shoá, deslegitimación del derecho colectivo a existir, asociación sistemática entre islam y terrorismo– y, al mismo tiempo, clasifican como “no odio” un volumen significativo de mensajes críticos con las políticas del Estado de Israel o con actores islamistas cuando se formulan sobre base fáctica y sin estereotipos. El desacople parcial entre la toxicidad léxica y la probabilidad de odio, especialmente visible en Facebook, Instagram y TikTok, confirma que este *codebook* normativo aporta información sustantiva más allá de los indicadores de “lenguaje malsonante”: una gran parte del antisemitismo y del prejuicio antimusulmán circula hoy codificada en marcos argumentativos o conspirativos con baja incivilidad/toxicidad textual, y nuestro enfoque permite detectarla sin penalizar el disenso político informado, dando así respuesta a la P1.

Las P2 y P3, vinculadas al OE2 (ofrecer un diagnóstico comparativo por plataformas y coyuntura), se abordan mediante el análisis de las probabilidades medias de odio y de su evolución temporal en los 798.475 mensajes del corpus. Mostramos que la intensidad estimada de antisemitismo e islamofobia no es homogénea entre redes: X y TikTok presentan las mayores probabilidades de antisemitismo, mientras que la islamofobia es relativamente más visible en X y Facebook, y más residual en Instagram. Estos patrones deben leerse como indicadores relativos, condicionados tanto por el diseño muestral (diccionarios léxicos, APIs, cobertura de CrowdTangle) como por la naturaleza de cada plataforma (anonimato y politización en X, institucionalidad en Facebook/Instagram, brevedad y multimodalidad en TikTok). En el periodo de evaluación, se observa que los picos de probabilidad de antisemitismo y de antiislamismo se alinean con episodios clave del ciclo Israel–Gaza; en esos momentos, el modelo detecta un aumento de mensajes en los que el antisionismo funciona efectivamente como vector de tropos antijudíos clásicos –por ejemplo, cuando convierte al “sionismo” en esencia maligna o traslada al conjunto de “los judíos” la responsabilidad por las acciones del gobierno israelí–, sin dejar por ello de identificar como legítimas las críticas basadas en vulneraciones de derechos humanos o

en diagnósticos jurídicos sobre la situación en Palestina. De este modo, la P2 queda respondida al ofrecer un mapa comparativo de la incidencia relativa por plataforma, y la P3 pone de manifiesto la relación entre la escalada coyuntural y la actualización de repertorios discursivos.

Las P4 y P5 se centran en la composición interna del odio detectado. A partir de las anotaciones y de la salida de los clasificadores, describimos qué estereotipos (pre)dominan y cómo se distribuyen sus intensidades. En el antisemitismo, el “poder oculto/contubernio”, la “doble lealtad” y el negacionismo/banalización de la Shoá aparecen como núcleos recurrentes, junto con analogías infundadas con el nazismo. En el antiislamismo, los ejes principales son la asociación del islam con la violencia y el terrorismo, la supuesta incompatibilidad cultural con Europa y la deshumanización de la población de Gaza. La tipología de Kovács y Fischer —dimensiones primarias-afectivas, conativas-activas y secundarias-latentes— se traduce operativamente en un continuo que abarca desde el insulto abierto hasta la normalización de clichés en mensajes aparentemente argumentativos. Esto permite responder a la P4, mostrando qué formas de antisemitismo y de prejuicio antimusulmán predominan en el ecosistema analizado y cómo se relacionan con los niveles de toxicidad textual.

En cuanto a la P5, la identificación de mensajes interseccionales ($\approx 2,5$ % del corpus) y, en particular, de la combinación Islam+Judaísmo como la más frecuente (6.397 mensajes) evidencia la existencia de patrones de culpabilización cruzada entre ambas comunidades. Estos mensajes presentan, en promedio, mayor toxicidad que los de un solo eje y, en el caso de X, también mayor interacción, lo que sugiere que el odio dirigido simultáneamente a judíos y musulmanes constituye un foco específico de polarización discursiva. Aquí se cumplen tanto la P5 como la parte del OE2: el análisis no solo constata la presencia de estas transferencias de odio, sino que también cuantifica su intensidad y su visibilidad relativa en cada plataforma.

Todo ello desemboca en la P6 y en el OE3, ambos orientados a las implicaciones normativas. A partir de la evidencia empírica, proponemos un conjunto de pautas directamente trasladables a la gobernanza de contenidos y a las políticas públicas: (i) la necesidad de mediciones multidimensionales que combinen clasificadores específicos de odio y métricas de toxicidad, calibradas por plataforma; (ii) la conveniencia de desarrollar detectores interseccionales sensibles a combinaciones identitarias; (iii) la pertinencia de monitorizar especialmente las coyunturas de alta conflictividad, en las que el riesgo de escalada y de campañas coordinadas es mayor, y la importancia de formar a los equipos de moderación —tanto humanos como automáticos— en los criterios de IHRA, JDA y Nexus, incluyendo cláusulas explícitas de protección del disenso político legítimo en torno a Israel/Palestina. Estas recomendaciones dialogan con los marcos existentes de OSCE/ODIHR, ECRI y la Estrategia Europea 2021–2030, y constituyen la respuesta concreta a la P6.

A pesar de que creemos firmemente que los hallazgos obtenidos en COIN y en este artículo abren nuevas vías de análisis, somos conscientes de que las limitaciones del estudio matizan el alcance de estas conclusiones. El foco geográfico (España) y lingüístico (español) restringe la extrapolación a otros contextos; el muestreo se basa en diccionarios léxicos y está condicionado por las limitaciones

de las APIs (especialmente la cobertura de CrowdTangle para Facebook/Instagram y el muestreo estratificado de X); en TikTok analizamos solo descripciones textuales, con una probable subdetección de la dimensión audiovisual; y la ironía, el sarcasmo o los *dog-whistles* muy contextuales siguen siendo desafíos importantes para los modelos actuales. Además, las cifras que presentamos son probabilidades medias de odio, no tasas absolutas de incidencia en la población general.

Precisamente por ello, las líneas futuras se presentan nítidas: (i) avanzar hacia modelos multimodales que integren audio, imagen y texto; (ii) ampliar la ventana temporal y aplicar diseños longitudinales o cuasiexperimentales que permitan medir con mayor precisión los efectos de determinadas coyunturas; (iii) evaluar sistemáticamente los costes de error por plataforma y ajustar umbrales adaptativos en diálogo con actores comunitarios; (iv) extender el enfoque a otros idiomas y países mediante auditorías externas; y (v) combinar el desarrollo técnico con intervenciones de alfabetización crítica frente a tropos conspirativos y negacionistas. En suma, hemos mostrado que no toda crítica política es antisemitismo ni islamofobia, pero también que una fracción relevante de ambos odios circula hoy en formas semántico-políticas de baja toxicidad léxica y con patrones de amplificación diferenciados según la plataforma. Vincular la genealogía histórica, los marcos normativos y las herramientas computacionales (nos) permite avanzar hacia formas de evaluación y moderación más precisas y proporcionadas, que refuercen la protección efectiva de las comunidades judías y musulmanas sin erosionar el disenso democrático imprescindible en sociedades abiertas, plurales y democráticas.

Bibliografía

- Naved Bakali. 2024. "The typologies of global Islamophobia and its manifestations across the Global North and South". *Journal of Muslim Philanthropy & Civil Society*, 8(2): 124–143.
- Erik Bleich. 2011. "What is Islamophobia and how much is there? Theorizing and measuring an emerging comparative concept". *American Behavioral Scientist*, 55(12): 1581–1600. <https://doi.org/10.1177/0002764211409387>
- Samantha Bradshaw, Hannah Bailey y Philip N. Howard. 2021. *Industrialized disinformation: 2020 global inventory of organized social media manipulation (COMPROP Working Paper Series)*. Programme on Democracy & Technology, Oxford Internet Institute.
- Andrew Brooks y Mark Griffiths. 2025. "Genocidal warfare in Gaza and the issue of balance in political geography". *Political Geography*, 122: 103387. <https://doi.org/10.1016/j.polgeo.2025.103387>
- Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi y Michele Starnini. 2021. "The echo chamber effect on social media". *Proceedings of the National Academy of Sciences*, 118(9): e2023301118. <https://doi.org/10.1073/pnas.2023301118>
- Manuela Consonni. 2022. "Upping the Antis: Addressing the conceptual ambiguities surrounding 'antisemitism'". *Società e Politica*, 13(25): 75–90. <https://doi.org/10.6092/issn.2239-2464/14486>
- Council of Europe, Committee of Ministers. 1997, 30 de octubre. Recommendation No. R (97) 20 on 'Hate Speech'. <https://go.coe.int/8Z6xK>
- Council of Europe, European Commission against Racism and Intolerance (ECRI). 2021. General Policy Recommendation No. 9 (revised) on preventing and combating antisemitism. <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/recommendation-no.9>
- Council of the European Union. 2008, 28 de noviembre. Council Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia by means of criminal law. *Official Journal of the European Union*, L 328: 55–58. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32008F0913>
- Council of the European Union. 2018, 6 de diciembre. Council Declaration on the fight against antisemitism and the development of a common security approach to better protect Jewish communities and institutions in Europe (15213/18). <https://data.consilium.europa.eu/doc/document/ST-15213-2018-INIT/en/pdf>

Council of the European Union. 2020, 2 de diciembre. Council Declaration on mainstreaming the fight against antisemitism across policy areas. <https://data.consilium.europa.eu/doc/document/ST-13637-2020-INIT/en/pdf>

Council of the European Union. 2022, 4 de marzo. Council Conclusions on combating racism and antisemitism (6406/22). <https://data.consilium.europa.eu/doc/document/ST-6406-2022-INIT/x/pdf>

David Dahlström. 2022. Antisemitism as a political weapon: A discourse analysis of claims of antisemitism in relation to Palestine/Israel (Master's thesis, Stockholm University). Stockholm University DiVA portal. <https://su.diva-portal.org/smash/record.jsf?pid=diva2:1755271>

Jan Deckers y Jonathan Coulter. 2022. "What is wrong with the International Holocaust Remembrance Alliance's definition of antisemitism?". *Res Publica*, 28(4): 733–752. <https://doi.org/10.1007/s11158-022-09553-4>

Abdelwahab El-Affendi. 2024. "The futility of genocide studies after Gaza". *Journal of Genocide Research*, 26(5): 683–696. <https://doi.org/10.1080/14623528.2024.2305525>

David Engel. 2008. "Away from a definition of antisemitism: An essay in the semantics of historical description". En Jeremy Cohen y Moshe Rosman (eds.), *Rethinking European Jewish history*, 30–53. Liverpool University Press. <https://doi.org/10.3828/liverpool/9781904113560.003.0003>

European Commission. 2016. Code of Conduct on Countering Illegal Hate Speech Online. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

European Commission. 2021. EU Strategy on combating antisemitism and fostering Jewish life (2021–2030) (COM(2021) 615 final). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2021:615:FIN>

European Parliament and Council of the European Union. 2012, 25 de octubre. Directive 2012/29/EU establishing minimum standards on the rights, support and protection of victims of crime. Official Journal of the European Union, L 315: 57–73. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012L0029>

European Parliament and Council of the European Union. 2018, 14 de noviembre. Directive (EU) 2018/1808 amending Directive 2010/13/EU on audiovisual media services. Official Journal of the European Union, L 303: 69–92. <https://eur-lex.europa.eu/eli/dir/2018/1808/oj>

European Parliament and Council of the European Union. 2021, 29 de abril. Regulation (EU) 2021/784 on addressing the dissemination of terrorist content online. Official Journal of the European Union, L 172: 79–109. <https://eur-lex.europa.eu/eli/reg/2021/784/oj>

European Parliament. 2017, 1 de junio. Resolution on combating antisemitism (2017/2692(RSP)). https://www.europarl.europa.eu/doceo/document/TA-8-2017-0243_EN.html

European Parliament. 2018, 25 de octubre. Resolution on the rise of neo-fascist violence in Europe. https://www.europarl.europa.eu/doceo/document/TA-8-2018-0428_EN.html

European Union Council. 1996, 15 de julio. Joint Action 96/443/JHA concerning action to combat racism and xenophobia. Official Journal of the European Union, L 185: 5–7. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31996F0443>

European Union Council. 1997, 2 de junio. Council Regulation (EC) No 1035/97 establishing a European Monitoring Centre on Racism and Xenophobia (EUMC). Official Journal of the European Union, L 151: 1–7. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31997R1035>

Steve Garner y Saher Selod. 2015. "The racialization of Muslims: Empirical studies of Islamophobia". *Critical Sociology*, 41(1): 9–19. <https://doi.org/10.1177/0896920514531606>

Gobierno de España – Ministerio de Asuntos Exteriores, Unión Europea y Cooperación. 2020, 22 de julio. España refrenda la definición operativa de "antisemitismo" de la IHRA. <https://www.exteriores.gob.es>

Gobierno de España – Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática. 2023. Plan nacional para la implementación de la Estrategia europea de lucha contra el antisemitismo (2023–2030). <https://www.mpr.gob.es>

José-Luis González-Esteban, Carmen María López-Rico, Lorraine Morales-Pino y Federico Sabater-Quinto. 2024. "Intensification of hate speech, based on the conversation generated on TikTok during the escalation of the war in the Middle East in 2023". *Social Sciences*, 13(1): 49. <https://doi.org/10.3390/socsci13010049>

Deena R. Hurwitz y Walter H. White Jr. 2024, November 25. "Recognizing Islamophobia, anti-Palestinian racism, and antisemitism". *Human Rights Magazine*. American Bar Association. <https://www.americanbar.org/groups/crsj/resources/human-rights/2024-november/recognizing-islamophobia-anti-palestinian-racism-and-antisemitism/>

International Holocaust Remembrance Alliance (IHRA). 2016. Working Definition of Antisemitism. <https://holocaustremembrance.com/resources/working-definition-antisemitism>

Cecilia Jacob. 2024. "A new politics of international criminal justice: Accountability in the Israel–Palestine conflict". *International Affairs*, 100(6): 2563–2584.

Laura J. Jaffee. 2024. "Anti-Zionism is not antisemitism: The centrality of Palestinian liberation in the struggle for anti-oppressive education". *Critical Education*, 15(1): 29–51. <https://doi.org/10.14288/ce.v15i1.186771>

Jerusalem Declaration on Antisemitism (JDA). 2021. JDA on Antisemitism. <https://jerusalemdeclaration.org/>

- Brent Kitchens, Steven L. Johnson y Peter Gray. 2020. "Understanding echo chambers and filter bubbles: The impact of social media on diversification and partisan shifts in news consumption". *MIS Quarterly*, 44(4): 1619–1649.
- Brian Klug. 2003. "The collective Jew: Israel and the new antisemitism". *Patterns of Prejudice*, 37(2): 117–138. <https://doi.org/10.1080/0031322032000087973>
- Brian Klug. 2013. "Interrogating new anti-Semitism". *Ethnic and Racial Studies*, 36(3): 468–482. <https://doi.org/10.1080/01419870.2013.734388>
- András Kovács y György Fischer. 1994. *Anti-Semitism Among Hungarian College and University Students*. American Jewish Committee. <https://archive.jpr.org.uk/object-1738>
- Arun Kundnani. 2012. "Radicalisation: The journey of a concept". *Race & Class*, 54(2): 3–25.
- Kenneth L. Marcus. 2015. *The Definition of Anti-Semitism*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199375646.001.0001>
- Nasar Meer. 2009. "Refutations of racism in the 'Muslim question'". *Patterns of Prejudice*, 43(3–4): 335–354.
- Nexus Task Force. 2021. *The Nexus Document: Understanding antisemitism at its nexus with Israel and Zionism*. Nexus Project. <https://nexusproject.us/nexus-resources/the-nexus-document/>
- Mohammed Nijim. 2024. "Gazacide: Palestinians from refugeehood to ontological obliteration". *Critical Sociology*. Advance online publication. <https://doi.org/10.1177/08969205241286530>
- OSCE. 2004. *Berlin Declaration on Antisemitism*. <https://www.osce.org/cio/31432>
- OSCE Office for Democratic Institutions and Human Rights (OSCE/ODIHR). 2014. *Hate Crime Data Collection and Monitoring: A Practical Guide*. <https://odihr.osce.org/odihr/datacollectionguide>
- OSCE Office for Democratic Institutions and Human Rights (OSCE/ODIHR). 2017. *Understanding Anti-Semitic Hate Crimes and Addressing the Security Needs of Jewish Communities: A Practical Guide*. <https://www.osce.org/odihr/317166>
- OSCE Office for Democratic Institutions and Human Rights (OSCE/ODIHR) y UNESCO. 2018. *Addressing Anti-Semitism through Education: Guidelines for Policymakers*. <https://unesdoc.unesco.org/ark:/48223/pf0000263702>
- Eli Pariser. 2011. *The filter bubble: What the internet is hiding from you*. Penguin Press.
- Peter G. J. Pulzer. 1988. *The Rise of Political Anti-Semitism in Germany and Austria 1867–1914* (Rev. ed.). Harvard University Press.
- Edward W. Said. 1978. *Orientalism*. Pantheon.
- Saher Selod y David G. Embrick. 2013. "Racialization and Muslims: Situating the Muslim experience in race scholarship". *Sociology Compass*, 7(8): 644–655. <https://doi.org/10.1111/soc4.12057>
- Alberto Spektorowski. 2024. "Anti-Semitism, Islamophobia and Anti-Zionism: Discrimination and political construction". *Religions*, 15(1): 74.
- Kenneth S. Stern. 2020. *The Conflict over the Conflict: The Israel/Palestine Campus Debate*. New Jewish Press (University of Toronto Press). <https://utppublishing.com/doi/book/10.3138/9781487507367>
- Cass R. Sunstein. 2017. *#Republic: Divided democracy in the age of social media*. Princeton University Press.
- UNESCO. 2012. *Holocaust Education in a Global Context*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000219279>
- UNESCO. 2013. *Why Teach about the Holocaust?* UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000222837>
- UNESCO. 2017. *Education about the Holocaust and Preventing Genocide: A Policy Guide*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000248071>
- United Nations General Assembly. 1948, 9 de diciembre. *Convention on the Prevention and Punishment of the Crime of Genocide*. United Nations Treaty Series, 78: 277. https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-1&chapter=4
- United Nations General Assembly. 2005, 1 de noviembre. *Resolution 60/7: Holocaust remembrance*. <https://undocs.org/en/A/RES/60/7>
- United Nations General Assembly. 2019. *Elimination of all forms of religious intolerance* (Report of the Special Rapporteur A/74/358). <https://undocs.org/en/A/74/358>
- Ernesto Verdeja. 2025. "The Gaza genocide in five crises". *Journal of Genocide Research*. Advance online publication. <https://doi.org/10.1080/14623528.2025.2452707>
- Robert Solomon Wistrich. 1991. *Antisemitism: The Longest Hatred*. Pantheon Books.
- Samuel C. Woolley y Philip N. Howard. 2016. "Political communication, computational propaganda, and autonomous agents: Introduction". *International Journal of Communication*, 10: 4882–4890.
- Moshe Zimmermann. 1986. *Wilhelm Marr: The Patriarch of Anti-Semitism*. Oxford University Press.

