Universitas-XXI, Revista de Ciencias Sociales y Humanas
ISSN: 1390-3837
ISSN: 1390-8634
revistauniversitas@ups.edu.ec
Universidad Politécnica Salesiana
Ecuador

Rossel-Castagneto, María Lorena

Regular el odio digital: entre la libertad de expresión y la protección de colectivos vulnerables en Chile

Universitas-XXI, Revista de Ciencias Sociales y Humanas,
núm. 43, 2025, Septiembre-Febrero 2026, pp. 95-122
Universidad Politécnica Salesiana
Cuenca, Ecuador

DOI: https://doi.org/10.17163/uni.n43.2025.04

Disponible en: https://www.redalyc.org/articulo.oa?id=476182348004

REVISTA DE CIENCIAS SOCIALES Y HUMANAS XXI
NIVERSITAS

https://doi.org/10.17163/uni.n43.2025.04

Regular el odio digital: entre la libertad de expresión y la protección de colectivos vulnerables en Chile

Regulating of digital hate: between freedom of expression and the protection of vulnerable groups in Chile

María Lorena Rossel-Castagneto
mrossel@udla.cl
https://orcid.org/0000-0003-4085-3000
Universidad de Las Américas, Chile
https://ror.org/0166e9x11

Recibido: 11/06/2025 **Revisado:** 21/07/2025 **Aprobado:** 22/08/2025 **Publicado:** 01/09/2025

Cómo citar: Rossel Castagneto, M. L. (2025). Regular el odio digital: entre la libertad de expresión y la protección de colectivos vulnerables en Chile. *Universitas XXI, 43*, pp. 95-122. https://doi.org/10.17163/uni.n43.2025.04

Resumen
La expansión de la inteligencia artificial (IA) en la comunicación política ha transformado las formas en que los actores políticos interactúan con la ciudadanía, facilitando la generación automatizada de mensajes personalizados y la difusión masiva de información en redes digitales. Esta dinámica ha contribuido a la propagación de noticias falsas y contenidos polarizantes, debilitando el debate público y favoreciendo la reproducción de discursos de odio, especialmente contra colectivos históricamente vulnerables como personas migrantes, mujeres y disidencias sexuales. Casos como *Cambridge Analytica* evidencian cómo el uso intensivo de datos personales puede manipular preferencias políticas, mientras que la jurisprudencia interamericana —como en Azul Rojas Marín vs. Perú— ha comenzado a vincular los discursos de odio con la violencia estructural por prejuicio. En este escenario, se hace cada vez más urgente repensar el rol del derecho frente a expresiones que, bajo la apariencia de libertad de expresión, reproducen exclusiones y discriminación. El artículo analiza esta problemática desde una perspectiva jurídico-comparada, examinando experiencias normativas como las de Alemania y España, y evaluando los vacíos regulatorios del sistema chileno. Se concluye que una respuesta eficaz requiere una regulación clara y proporcionada, orientada a proteger los derechos fundamentales sin recurrir a mecanismos de censura. La investigación propone criterios normativos para una intervención estatal legítima que combine sanciones razonables con medidas preventivas y educativas, fortaleciendo así la deliberación democrática.

Palabras clave
Inteligencia artificial, discursos de odio, libertad de expresión, comunicación política, derechos fundamentales, democracia deliberativa, derecho penal, autoritarismo digital.

Abstract
The expansion of artificial intelligence (AI) in political communication has transformed the ways in which political actors interact with citizens, enabling the automated generation of personalized messages and the massive dissemination of information through digital networks. This dynamic has contributed to the spread of fake news and polarizing content, weakening public debate and fostering the reproduction of hate speech, particularly against historically vulnerable groups such as migrants, women, and sexual minorities. Cases such as Cambridge Analytica reveal how the intensive use of personal data can manipulate political preferences, while inter-American jurisprudence —as in Azul Rojas Marín v. Peru— has begun to link hate speech with structural prejudice-based violence. In this context, it is increasingly urgent to reconsider the role of law in addressing expressions that, under the guise of free speech, perpetuate exclusion and discrimination. This article analyzes the issue from a comparative legal perspective, examining regulatory frameworks such as those of Germany and Spain, and assessing the regulatory gaps within the Chilean legal system. It concludes that an effective response requires clear and proportionate regulation aimed at protecting fundamental rights without resorting to censorship. The study proposes normative criteria for legitimate state intervention that combines reasonable sanctions with preventive and educational measures, thereby strengthening democratic deliberation.

Keywords
Artificial intelligence, hate speech, freedom of expression, political communication, fundamental rights, deliberative democracy, criminal law, digital authoritarianism.

Introducción

La incorporación masiva de inteligencia artificial (IA) en la comunicación política ha reconfigurado las formas en que los actores políticos interactúan con la ciudadanía, posibilitando la generación automatizada de mensajes personalizados, la circulación de contenidos hipersegmentados y la amplificación de narrativas polarizantes. Esta transformación tecnológica, lejos de ser neutra, ha facilitado también la proliferación de discursos de odio en entornos digitales, especialmente contra colectivos históricamente vulnerables como personas migrantes, mujeres y disidencias sexuales.

El auge de expresiones discriminatorias mediadas por IA plantea tensiones jurídicas fundamentales entre la libertad de expresión y el deber estatal de proteger la igualdad, la dignidad humana y la deliberación democrática.

En este contexto, la pregunta que guía este trabajo es: ¿cómo debe responder el derecho a los discursos de odio amplificados por inteligencia artificial en contextos políticos, sin comprometer la libertad de expresión?

La hipótesis central sostiene que la ausencia de una regulación adecuada configura un vacío normativo que permite la circulación impune de mensajes lesivos para la convivencia democrática. Por ello, resulta imperioso repensar los márgenes normativos de intervención estatal, superando tanto el riesgo de inacción como el de sobrerreacción punitiva.

Este artículo aborda el problema desde una perspectiva jurídico-dogmática y comparada, examinando el tratamiento legal de los discursos de odio en Chile, Alemania y España, así como los estándares del sistema interamericano de derechos humanos. A partir de ello, se propone un conjunto de criterios normativos que permitan enfrentar este fenómeno de manera legítima, proporcional y conforme a los principios del constitucionalismo democrático.

Materiales y método

Este estudio adopta una metodología jurídico-dogmática, enraizada en la tradición del derecho constitucional y penal, orientada a identificar, interpretar y sistematizar normas, principios y categorías relevantes para el tratamiento jurídico del discurso de odio en entornos digitales. Tal como sostiene Perez-nieto Castro (2020), esta aproximación dogmática permite evaluar la coherencia interna del ordenamiento jurídico mediante los principios estructurantes del derecho penal, como legalidad, lesividad y culpabilidad, lo que resulta especialmente útil para analizar los límites legítimos del ius puniendi frente a expresiones comunicacionales lesivas. El enfoque parte del reconocimiento de los discursos de odio como una categoría jurídica controvertida, cuya definición, límites y mecanismos de control requieren una fundamentación normativa rigurosa desde el derecho positivo y los derechos fundamentales.

No obstante, como advierte Castillo Morales (2019), el objeto del derecho penal no puede analizarse de forma aislada ni meramente formal. Al tratarse de un instrumento de control social, su estudio exige integrar dimensiones sociales, culturales y políticas que exceden la dogmática clásica. De ahí que el autor proponga una metodología comparada e interdisciplinaria que fortalezca la cientificidad de la dogmática penal y evite su vaciamiento técnico-formal, permitiendo evaluar críticamente la legitimidad del poder punitivo

y su adecuación a fines democráticos. Por ello, se recurre a un enfoque comparado, centrado en los ordenamientos jurídicos de Alemania y España. La elección de estos sistemas se justifica tanto por su pertenencia a la tradición jurídica continental (de base codificada y garantista), como por su experiencia normativa y jurisprudencial en materia de libertad de expresión y sanción de discursos de odio. Alemania ofrece un modelo robusto con criterios de proporcionalidad y aplicación excepcional del derecho penal; España, en cambio, presenta un tipo penal más amplio y controvertido, sujeto a debate doctrinal sobre sus límites. Esta comparación permite identificar modelos regulatorios alternativos que ilustran riesgos y virtudes de distintas estrategias jurídicas.

La comparación normativa permite responder de manera directa al segundo y tercer objetivo específico del estudio, relativos a la evaluación de estándares internacionales y a la formulación de criterios jurídicos para una intervención estatal legítima y proporcional.

El criterio de comparación se centra en tres dimensiones analíticas:

1. La tipificación penal del discurso de odio y sus límites sustantivos (elementos típicos, bienes jurídicos protegidos, grados de lesividad);
2. La proporcionalidad y eficacia de los mecanismos de control estatal (penales, administrativos y educativos);
3. El impacto potencial sobre el ejercicio de derechos fundamentales, particularmente la libertad de expresión y el principio de igualdad.

Del mismo modo, el estudio abarca tanto normas vigentes como proyectos legislativos relevantes, entre los años 2010 al 2025, así como jurisprudencia nacional e internacional reciente. En el caso chileno, se analizan el artículo 31 de la Ley N.º 19.733, los boletines legislativos sobre incitación al odio y negacionismo, y su articulación (o falta de ella) con el principio de intervención mínima del derecho penal.

El trabajo integra un marco interdisciplinario operativo, que articula herramientas del derecho con aportes de la teoría política (democracia deliberativa), la comunicación digital (virilización algorítmica, posverdad) y la sociología jurídica (estructuras de discriminación). Estas perspectivas permiten problematizar el contexto digital contemporáneo y sus efectos sobre la función estructurante del lenguaje jurídico en una democracia.

Las fuentes primarias analizadas incluyen normas constitucionales, leyes penales, proyectos de ley y jurisprudencia nacional, interamericana y euro-

pea. Se utilizó un análisis de contenido doctrinal y jurisprudencial, mediante matrices de codificación que permitieron categorizar los niveles de intervención normativa, la intensidad del discurso, la condición del grupo afectado, y la naturaleza de la respuesta estatal. La selección doctrinal y jurisprudencial se basa en criterios de pertinencia temática, actualidad, y referencia en debates académicos y legislativos. Las fuentes secundarias, como informes institucionales, artículos académicos y estudios empíricos recientes, se utilizan con carácter ilustrativo y contextual, sin aspiración a representatividad estadística, pero con valor analítico para enriquecer el análisis jurídico.

Este enfoque permite responder a los objetivos del estudio: identificar vacíos regulatorios, evaluar estándares normativos relevantes y proponer criterios jurídicos aplicables para una intervención estatal legítima frente al discurso de odio en entornos digitales, particularmente en el contexto chileno.

Resultados

Este estudio ha permitido identificar una serie de hallazgos relevantes que se organizan en tres niveles: (1) resultados conceptuales-doctrinales; (2) resultados derivados del análisis jurídico-comparado; y (3) resultados propositivos orientados a criterios normativos para una intervención estatal legítima.

Hallazgos doctrinales: desafíos normativos del odio digital

- Vacíos normativos ante la irrupción de la inteligencia artificial (IA): la IA actúa como catalizador de discursos de odio en entornos digitales, generando una forma de violencia simbólica que vulnera derechos fundamentales como la autodeterminación informativa, la igualdad, la participación política y la dignidad humana. La ausencia de regulación específica respecto de contenidos generados o amplificados por IA deja un vacío jurídico crítico en los ordenamientos actuales.
- Impacto estructural del odio digital: se constata que los discursos de odio ya no se limitan a la incitación directa a la violencia, sino que operan como mecanismos normalizados de exclusión y estigmatización, especialmente contra personas migrantes, mujeres y disidencias sexuales. Estos discursos afectan el ejercicio efectivo del juicio democrático, favoreciendo formas de "autoritarismo digital por diseño".

Hallazgos comparativos: contraste entre Alemania, España y Chile

Se sistematizan los principales elementos mediante una matriz comparativa centrada en tres variables:

Dimensión analizada	Alemania	España	Chile
Tipificación penal	§130 StGB (incitación al odio), definido con criterios restrictivos y vinculados a la dignidad humana	Art. 510 CP, redacción amplia e imprecisa, incluye discurso contra "ideología"	Art. 31 Ley 19.733: tipo administrativo débil; sin tipo penal específico
Control estatal y rol de plataformas	NetzDG: obliga a plataformas a remover contenido en plazos breves; combinación penal-administrativa	Rol pasivo de plataformas; sin ley específica de moderación algorítmica	Sin regulación de plataformas ni normas sobre IA o viralización de odio digital
Impacto sobre derechos fundamentales	Modelo garantista; se privilegia la proporcionalidad y la protección de la dignidad frente al odio	Críticas doctrinales por ambigüedad e inseguridad jurídica; riesgo de sobre criminalización	Alto riesgo de impunidad; baja efectividad del marco legal para proteger a grupos vulnerables

Este análisis comparativo permite constatar que el modelo alemán ofrece un equilibrio razonable entre sanción y garantías, mientras que el modelo español presenta riesgos de ambigüedad punitiva. Chile, en cambio, exhibe un vacío normativo preocupante.

Resultados propositivos: criterios normativos para la intervención estatal

A partir del análisis realizado, se proponen los siguientes criterios normativos, orientados a una regulación legítima y proporcional del discurso de odio en entornos digitales:

1. Diferenciación por gravedad del discurso (tipología propuesta):
    - Gravedad alta: incitación directa a la violencia, apología de crímenes de lesa humanidad: respuesta penal.

- Gravedad intermedia: desinformación con intención estigmatizante, incitación indirecta: respuesta administrativa.
- Gravedad leve: expresiones ofensivas no sistemáticas: respuesta educativa/preventiva.

2. Limitación del uso del derecho penal al principio de ultima ratio:
   - Solo aplicable cuando existan indicios objetivos de riesgo real e inminente para los derechos fundamentales de colectivos vulnerables.
3. Necesidad de reformas al marco chileno:
   - Reformulación del artículo 31 de la Ley N.º 19.733.
   - Incorporación de un tipo penal autónomo y acotado (con base en *odium dictum*).
   - Regulación del rol de las plataformas digitales, con obligaciones de moderación algorítmica transparente.
4. Inclusión del principio de autodeterminación informativa como garantía estructural.
   - Especialmente frente al uso de datos personales para segmentación política automatizada sin consentimiento.
5. Intervención estatal integral:
   - Combinación de sanciones proporcionales, medidas educativas, y políticas públicas de alfabetización digital, sin censura previa ni regresividad democrática.

Conclusiones y discusión

Campañas políticas y deterioro deliberativo en contextos digitales

La comunicación política contemporánea se caracteriza por la transformación radical de las dinámicas tradicionales de interacción entre actores políticos, medios y ciudadanía, a partir de la incorporación de tecnologías digitales, inteligencia artificial y análisis masivo de datos (Mota, 2023). En efecto, la inteligencia artificial (IA) ha transformado profundamente la comunicación política en las campañas electorales, especialmente al intervenir en la producción, personalización y difusión del discurso político. Esta transformación, como afirman Amaya y Cueva (2025), responde al uso de algoritmos que per-

miten segmentar audiencias y generar mensajes altamente personalizados, lo que redefine la interacción entre actores políticos y ciudadanía (pp. 3-4).

La IA posibilita un tratamiento masivo de datos y una capacidad predictiva que influye directamente en la toma de decisiones estratégicas de campaña (Torres *et al.,* 2024, p. 84). Esto se traduce, entre otros aspectos, en la trazabilidad de las campañas, la detección de noticias falsas y la manipulación emocional del electorado (Abrego y Flores, 2021, p. 6).

Ábrego y Flores, ejemplifican lo ocurrido en México como un caso paradigmático del uso de tecnologías —especialmente la inteligencia artificial (IA) y el *big data*— con fines políticos, destacando su potencial para manipular la información y controlar a los votantes mediante lo que denominan *gubernamentalidad algorítmica*. Esta forma de poder se basa en la recopilación masiva de datos personales, muchas veces sin consentimiento, para perfilar, segmentar y condicionar comportamientos sociales y políticos en beneficio de gobiernos y corporaciones (Ábrego y Flores, 2021, pp. 213-216).

Valdez *et al*. (2023) señalan que los mecanismos más relevantes para manipular campañas políticas y condicionar el comportamiento electoral de la ciudadanía es el uso de "*bots*" y perfiles falsos, que pueden replicar mensajes, posicionar tendencias en redes sociales y atacar a candidatos adversarios, creando la ilusión de consenso o rechazo social (Valdez *et al*., 2023, p. 33). Asimismo, la IA permite analizar masivamente datos provenientes de redes sociales y otras plataformas para elaborar perfiles psicográficos de los votantes, como lo evidenció el caso de *Cambridge Analytica*, donde se manipularon preferencias electorales a partir del análisis de datos obtenidos sin consentimiento (p. 31).

A ello se suma el uso de IA generativa (GenAI) para crear contenidos visuales falsificados —como imágenes bélicas o deepfakes— que distorsionan la percepción pública de la realidad (Rubio *et al.,* 2024, p. 7; De Rê, 2024, p. 12; Valdez *et al.*, p. 34; Alkiviadou, 2023).

Este fenómeno plantea serias implicancias éticas y jurídicas. En primer lugar, afecta el principio de autodeterminación informativa, reconocido por diversos estándares internacionales en materia de derechos humanos. En segundo lugar, debilita el principio de transparencia electoral y el derecho a recibir información veraz, elementos esenciales en una democracia deliberativa. Finalmente, la ausencia de control sobre estas tecnologías genera un vacío normativo que puede habilitar formas de autoritarismo digital encubierto, como se desarrollará a continuación.

El principio de autodeterminación informativa

El principio de autodeterminación informativa garantiza a las personas el control sobre sus datos personales, permitiéndoles decidir de forma libre e informada qué información puede ser tratada, por quién y con qué fines. Surge como una respuesta jurídica frente a los riesgos del tratamiento masivo y automatizado de datos en entornos digitales mediados por inteligencia artificial (IA), donde los individuos son reducidos a objetos de análisis algorítmico (Bonilla, 2022, pp. 288-289).

Este principio, reconocido por el Tribunal Constitucional Federal Alemán en la sentencia del censo poblacional de 1983 (Volkszählungsurteil), ha sido incorporado en instrumentos internacionales como el Reglamento General de Protección de Datos (RGPD) de la Unión Europea y el Convenio 108+ del Consejo de Europa, que imponen a los Estados obligaciones para garantizar la transparencia, el consentimiento informado y el control sobre el uso de datos personales (Álvarez Buján, 2023; Bonilla, 2022).

En contextos de campañas políticas, el uso de IA y sistemas de microsegmentación permite recopilar y procesar grandes volúmenes de datos sin consentimiento consciente, afectando gravemente este derecho. Como advierten Amaya y Cueva (2025), los ciudadanos se convierten en blancos de estrategias persuasivas diseñadas a partir de sus perfiles, lo que socava la privacidad, la libertad informativa y la autonomía del voto (pp. 5-6).

Además, los mismos mecanismos que permiten manipulación política facilitan la difusión de discursos de odio dirigidos contra grupos históricamente vulnerables. Esta realidad exige comprender la autodeterminación informativa como parte de una red de garantías destinadas a proteger no solo la privacidad individual, sino también la dignidad, la igualdad y la deliberación democrática. Frente a estas amenazas, se requiere avanzar hacia marcos normativos que reconozcan el valor estructurante de este principio y permitan sancionar aquellas expresiones que, por su intensidad y efectos, resulten incompatibles con una democracia respetuosa de los derechos humanos.

Transparencia electoral y derecho a la información veraz

La transparencia electoral constituye una condición estructural del constitucionalismo democrático, al garantizar que el proceso electoral se desarrolle bajo reglas claras, acceso igualitario a la información y fiscalización efectiva

de los actores involucrados. En los entornos digitales, esta exigencia adquiere una nueva dimensión frente a la opacidad de los algoritmos, la personalización masiva de contenidos y la circulación de información falsa. Como señala Serra (2023), la transparencia en contextos de gobernanza algorítmica debe entenderse como un principio activo, que impone a los actores públicos y privados la obligación de revelar los criterios, fuentes y efectos de sus decisiones automatizadas.

Este principio se articula con el derecho a la información veraz, entendido como el acceso a información relevante, plural y no manipulada, condición para ejercer una ciudadanía informada y deliberante. La proliferación de desinformación política —potenciada por el uso de IA—, vulnera este derecho, distorsiona el debate electoral y erosiona la confianza pública. En este contexto, la transparencia debe extenderse a las plataformas digitales, cuyos algoritmos actúan como filtros invisibles que definen lo que los usuarios ven, amplificando contenidos polarizantes o discriminatorios.

Tanto el Comité de Derechos Humanos de la ONU como la Relatoría para la Libertad de Expresión de la CIDH han subrayado que los Estados deben adoptar medidas para asegurar la transparencia de los procesos digitales que afectan la libertad de expresión, sin caer en formas de censura. Esto incluye el deber de regular la publicidad política digital, los mecanismos de moderación algorítmica y el uso de datos personales en campañas electorales.

Así, la transparencia electoral en la era digital no se limita a la rendición de cuentas tradicional, sino que exige mecanismos normativos que hagan visible lo que hoy opera de forma opaca y automatizada, afectando la calidad democrática y el ejercicio igualitario de los derechos políticos.

Reflexiones en torno al autoritarismo digital encubierto y la posverdad

El autoritarismo digital designa el uso sistemático de tecnologías para vigilar, manipular o reprimir a la población, no solo por regímenes autoritarios, sino también en democracias y empresas tecnológicas cuando operan sistemas sin rendición de cuentas, configurando un "autoritarismo por diseño" (G'sell, 2025, pp. 3-5).

En este entorno, la desinformación se vuelve una herramienta de control simbólico. Lafont (2025) advierte que la proliferación de contenidos manipuladores debilita las condiciones mínimas de deliberación democrática e ins-

tala cámaras de eco y segmentación algorítmica que socavan el pluralismo discursivo. Cuando los ciudadanos no pueden distinguir si los mensajes que reciben provienen de conciudadanos o de actores maliciosos, se erosiona la confianza pública y la legitimidad del sistema democrático (Lafont, 2025, pp. 20-22). Desde la teoría del Estado constitucional, ello impide el reconocimiento racional de las normas, requisito para la validez democrática de las decisiones colectivas. Por tanto, el fortalecimiento de espacios deliberativos fiables —como los mini públicos o asambleas ciudadanas— no es solo un imperativo ético, sino una condición de posibilidad para la subsistencia del orden democrático (Lafont, 2025, pp. 26-28).

A esta amenaza estructural se suma el fenómeno de la posverdad, que según Astudillo (2023), busca instalar convicciones distorsionadas como si fueran verdad, afectando no solo la libertad de expresión, sino la posibilidad de ejercerla de manera crítica e informada (pp. 403, 415). Esta instrumentalización del pluralismo informativo debilita el debate democrático y, en contextos de estigmatización, legitima discursos intolerantes y prácticas discriminatorias, especialmente contra migrantes y disidencias sexuales (pp. 417-418). Incluso derechos como la salud pública pueden verse comprometidos, como lo demuestra la difusión de discursos antivacunas sin base científica (p. 418).

En este escenario de manipulación informativa y deterioro deliberativo, los discursos de odio adquieren una peligrosidad particular. Las plataformas digitales han facilitado su circulación y legitimación bajo la retórica de la libertad de expresión, pese a que refuerzan estructuras de exclusión y violencia simbólica. Como muestran los vínculos entre posverdad y autoritarismo digital, el problema ya no es solo la falsedad del contenido o la opacidad de su origen, sino su capacidad para reproducir dinámicas de discriminación estructural. Por ello, urge abordar su regulación no como censura, sino como una respuesta proporcional y necesaria para resguardar la dignidad y los derechos fundamentales de los colectivos vulnerables.

Libertad de expresión y discursos de odio en clave comparada: aproximación al sistema interamericano y sistema europeo de derechos humanos

En los entornos digitales contemporáneos, el discurso de odio se ha convertido en una amenaza estructural para la dignidad humana, la igualdad y

el pluralismo democrático, afectando de forma especialmente intensa a colectivos históricamente vulnerables como mujeres, personas migrantes y disidencias sexuales.

Aunque carece de una definición jurídica única, organismos internacionales y estudios doctrinales lo describen como toda expresión que incita o justifica la violencia, discriminación o estigmatización en razón de características identitarias. En redes sociales, estos discursos adoptan formas explícitas y sutiles, se amplifican mediante algoritmos y distorsionan la opinión pública, deteriorando la deliberación democrática y reforzando prejuicios estructurales. Su proliferación ha generado efectos diferenciados: desde la salida de víctimas de plataformas digitales hasta la normalización de narrativas patriarcales en contextos escolares. Todo ello evidencia la urgencia de respuestas normativas y sociales que combinen regulación proporcional, moderación responsable y formación ciudadana en entornos digitales (Alkiviadou, 2023; Prabhu y Seethalakshmi, 2025; Sibrián *et al.,* 2024; Marolla *et al.,* 2024; Sánchez *et al.,* 2023).

La regulación de los discursos de odio presenta un dilema fundamental: por un lado, la responsabilidad estatal de proteger la dignidad humana, la convivencia pacífica y la democracia, basándose en principios constitucionales de igualdad y no discriminación; por otro, la necesidad de salvaguardar la libertad de expresión, que impone la prohibición de censura previa y limita la regulación del contenido de la expresión (Abramovich, 2022, pp. 88, 89)

El Sistema Interamericano de Derechos Humanos (SIDH) ha desarrollado una doctrina sobre la libertad de expresión que se caracteriza por su amplio alcance y un enfoque considerado "hiperprotector" y comprendería la libertad de "buscar, recibir y difundir informaciones e ideas de toda índole" (Jacoby, 2020, p. 149).

En la Opinión consultiva 5/85, la Corte Interamericana de Derechos Humanos (en adelante Corte IDH), al interpretar el artículo 13 de la Convención Americana de Derecho Humanos (en adelante CADH) ha reconocido que la libertad de expresión posee una doble dimensión. La dimensión individual, que se refiere al derecho de cada persona a expresar sus opiniones y difundir su pensamiento por cualquier medio y la dimensión social, que se vincula con su función instrumental para el intercambio de ideas e informaciones, y para la comunicación masiva entre las personas.

Es crucial que ambas dimensiones sean garantizadas de manera simultánea. La Relatoría Especial para la Libertad de Expresión (RELE-OEA, 2004)

ha enfatizado que no se puede "menoscabar una de ellas invocando como justificación la preservación de la otra", lo que diferencia este modelo del principio de ponderación predominante en el sistema europeo.

Agrega Jacoby (2020) que un pilar fundamental de la libertad de expresión en el SIDH es la prohibición de la censura previa, contenido en el artículo 13.2 de la CADH. El ejercicio de este derecho no puede estar sujeto a censura previa, salvo en situaciones específicamente delimitadas y bajo condiciones muy estrictas. En su lugar, el sistema privilegia un régimen de responsabilidades ulteriores. Estas responsabilidades deben estar expresamente fijadas por ley, ser necesarias para asegurar el respeto a los derechos o la reputación de terceros, o para proteger la seguridad nacional, el orden público o la salud y moral públicas (Jacoby, 2020, pp. 150-154).

Adicionalmente, el artículo 13.5 de la CADH hace referencia explícita al discurso de odio, prohibiendo "toda propaganda en favor de la guerra y toda apología del odio nacional, racial o religioso que constituyan incitaciones a la violencia o cualquier otra acción ilegal similar contra cualquier persona o grupo de personas, por ningún motivo, inclusive los de raza, color, religión, idioma u origen nacional".

Destaca Abramovich (2022) que, de la interpretación de estas normas, la Relatoría y la Corte IDH, se desprende una clasificación de los discursos de odio, para determinar los niveles de injerencia estatal, distinguiendo entre discurso no protegido, discurso protegido y discurso especialmente protegido.

El modelo europeo, y específicamente la jurisprudencia del Tribunal Europeo de Derechos Humanos (TEDH), en cambio, parte del supuesto de que existe un *continuum* entre la emisión del discurso y los daños a los derechos fundamentales, lo que justifica la limitación del discurso como una forma directa de protección de esos derechos. Este modelo busca un equilibrio, ponderando la libertad de expresión y otros derechos fundamentales, así como valores públicos como la paz y la convivencia social (Rodríguez Zepeda, 2018, p. 47).

Abramovich sostiene que el TEDH reconoce la necesidad de sancionar y prevenir todas las formas de expresión que propaguen, inciten, promuevan o justifiquen el odio basado en la intolerancia. Para determinar los niveles de injerencia estatal y el tipo de responsabilidad, el TEDH y la Recomendación CM/Rec(2022)16 del Comité de Ministros a los Estados miembros sobre la lucha contra el discurso de odio, establecen una serie de factores para evaluar la gravedad del discurso de odio, tales como el contenido del discurso, el con-

texto político y espacial, la intención del orador, el papel o estatus del orador en la sociedad, forma en que se difunde o amplifica la expresión, la capacidad de expresión para provocar consecuencias perjudiciales, y la naturaleza y el tamaño de la audiencia (Recomendación CM/Rec(2022)16, párr. 32).

A diferencia del modelo europeo, el Sistema Interamericano de Derechos Humanos (SIDH) no ha desarrollado una doctrina tan explícita y profunda sobre el discurso de odio, sino que ha privilegiado los mecanismos de responsabilidad ulteriores (Jacoby, 2020, p. 150). No obstante, un pronunciamiento clave donde la Corte IDH se refirió específicamente a la naturaleza de los discursos de odio en relación con la violencia en el Caso Azul Rojas Marín y otra vs. Perú. En esta sentencia, la Corte IDH profundizó en la comprensión de la violencia por prejuicio, especialmente contra las personas LGBTI, vinculándola directamente con los discursos de odio, ya que a juicio de la Corte "esta violencia, alimentada por discursos de odio, puede dar lugar a crímenes de odio" (CIDH, 2020, párr. 93).

En este caso, la Corte IDH determinó que en la sociedad peruana existían y persisten fuertes prejuicios contra la población LGBTI que a menudo conducen a la violencia, incluso por parte de agentes estatales y que la violencia contra las personas LGBTI posee un fin simbólico: comunicar un mensaje de exclusión o subordinación a un grupo específico (CIDH, 2020, párr. 50-93). Concluye la Corte, que respecto de estas situaciones el Estado tiene la obligación positiva de adoptar medidas para revertir o cambiar situaciones discriminatorias existentes en sus sociedades y ejercer una protección especial frente a las actuaciones de terceros que, bajo su tolerancia o aquiescencia, vulneren los derechos de grupos en situación de vulnerabilidad (CIDH, 2020, párr. 89).

Por su parte, la RELE promueve activamente mecanismos no sancionatorios y políticas públicas integrales, enfocándose más bien en medidas alternativas al derecho penal, como la generación de políticas públicas y la creación de comisiones o comités gubernamentales para su seguimiento, fomentar la sensibilización y capacitación y asegurándose de que cualquier restricción legal sea clara, precisa y proporcional, evitando la vaguedad que podría llevar a abusos (Jacoby, 2020, p. 161).

A pesar de la disparidad de criterios entre el Sistema Interamericano de Derechos Humanos, que privilegia un enfoque de protección ulterior y rechaza de manera enfática la censura previa, y el modelo europeo, que admite la limitación anticipada del discurso como mecanismo de protección directa frente a la intolerancia, ambos sistemas reconocen el potencial lesivo de los

discursos de odio cuando estos refuerzan estructuras de discriminación y exclusión. Como lo ha expresado la Corte IDH en el caso Azul Rojas Marín vs. Perú, este tipo de expresiones no solo reflejan prejuicios sociales persistentes, sino que pueden incidir directamente en la comisión de crímenes de odio, especialmente contra colectivos históricamente vulnerados (CIDH, 2020, párr. 93). De ahí que, más allá de las diferencias doctrinales en torno a la ponderación o la responsabilidad ulterior, resulta indispensable avanzar hacia una regulación sustantiva de los discursos de odio, en particular aquellos que afectan a grupos en situación de vulnerabilidad. Esta regulación debe equilibrar la protección de la libertad de expresión con la necesidad de garantizar condiciones de igualdad y dignidad, evitando que las narrativas discriminatorias continúen operando como mecanismos normalizados de violencia simbólica, exclusión social y legitimación de la intolerancia. En este contexto, tanto la Corte IDH como el TEDH han coincidido en que los Estados tienen una obligación positiva de prevenir y sancionar estas expresiones cuando su contenido, contexto o impacto puede traducirse en agresiones reales o simbólicas que amenacen la convivencia democrática y los derechos humanos fundamentales.

El desafío de sancionar el odio: eficacia y límites del derecho penal en Alemania y España

Valdés (2021, pp. 35-47) señala que, aunque la utilización del derecho penal para limitar la libertad de expresión es generalmente vista con sospecha, existe un sentimiento generalizado de rechazo hacia los discursos de odio y una aceptación de su sanción penal en muchos ordenamientos contemporáneos. En sistemas jurídicos como el alemán, francés, italiano o austriaco, la dignidad humana desempeña un papel protagónico, permitiendo sanciones penales contra expresiones que constituyen un ataque directo a esta, o al orden constitucional. Estos sistemas son descritos como "militantes", pues excluyen determinadas expresiones por su contenido y por el peligro potencial que significan, más allá de un peligro real. Comparten la sensibilidad de proteger bienes jurídicos fundamentales, como el orden público (entendido como paz pública) y la dignidad de la persona, incluso con el derecho penal como *ultima ratio* (Valdés, 2021, p. 47).

Sin embargo, Valdés enfatiza que cualquier restricción a la libertad de expresión implica un riesgo para el orden democrático y para el derecho in-

dividual. Para que una conducta sea prohibida y castigada penalmente, debe implicar afectaciones graves a bienes jurídicos, lo que convierte las expresiones o manifestaciones castigadas penalmente en excepciones muy puntuales; la regla general es que la libertad de expresión no puede ser restringida penalmente.

Alemania tiene una de las legislaciones más robustas en Europa en esta materia. En Alemania, los discursos de odio se encuentran tipificados en el Código Penal (Strafgesetzbuch, StGB), particularmente en los §§ 130 y 185-187.

El §130 StGB, titulado Volksverhetzung (incitación al odio popular), sanciona penalmente a quien incite al odio contra partes de la población o llame a la violencia o a medidas arbitrarias contra ellas, así como a quien atente contra la dignidad humana mediante la injuria, el desprecio o la calumnia de grupos por motivos de origen étnico, religión, nacionalidad o características similares. Este precepto también contempla la negación del Holocausto como una forma específica de incitación al odio (González Ruiz, 2023, pp. 87-88).

Adicionalmente, la Network Enforcement Act (NetzDG), obliga a las plataformas digitales a eliminar discursos de odio de manera expedita, bajo sanción administrativa, lo que muestra un enfoque integral que abarca tanto el ámbito penal como el administrativo y digital (González Ruiz, 2023, pp. 89-90).

Por su parte, el Código penal español tipifica los discursos de odio en el artículo 510, que sanciona tanto a quienes fomenten o inciten directa o indirectamente al odio o violencia en contra de un grupo o personas por motivos como la raza, la religión, el género, la discapacidad, etc., como a quienes produzcan o difundan material que fomente el odio violencia en contra de estos grupos o personas, o a quienes nieguen o hagan exaltación de genocidios y crímenes contra la humanidad.

Si bien esta regulación tiende a proteger a grupos históricamente discriminados ha sido objeto de críticas por los términos ambiguos y amplios que utiliza el legislador para sancionar los discursos de odio o el negacionismo. En efecto, Tapia (2021) sostiene que la nueva redacción del artículo 510.1.a) es excesivamente amplia e imprecisa, lo que genera inseguridad jurídica y permite una expansión desmedida del derecho penal. Esto se evidencia, por ejemplo, en la equiparación de discursos radicales o controvertidos —como los expresados por artistas, activistas o usuarios de redes sociales— con incitaciones reales a la violencia o la discriminación (Tapia, 2021, pp. 284-290).

Adicionalmente, según la autora, el término "ideología" como circunstancia sospechosa de discriminación ha servido en su opinión para criminalizar

expresiones disidentes o políticamente incorrectas. Ello ha llevado a considerar como delitos de odio discursos dirigidos contra toreros, policías, miembros de la Corona o incluso personas con ideología neonazi, lo que desvirtúa el sentido original de los delitos antidiscriminatorios y trivializa su función protectora frente a colectivos históricamente marginados (Tapia, 2021, pp. 297-304). Por lo anterior, Tapia comparte a propuesta del Grupo de Estudios de Política Criminal de reformar el tipo penal, limitando la punibilidad a los casos en que exista una incitación directa y pública a la comisión de delitos concretos (contra la vida, la integridad, la libertad, etc.) por motivos discriminatorios, y solo cuando exista un riesgo inminente de su comisión. Esta delimitación permitiría, a su juicio, recuperar la legitimidad del tipo penal y respetar los principios de intervención mínima y proporcionalidad (Tapia, 2021, p. 314).

En este mismo sentido, Alastuey cuestiona la orientación actual del artículo 510. En primer lugar, sostiene que la amplitud del tipo penal, con expresiones como "fomentar un clima de odio", afecta el principio de legalidad penal al introducir cláusulas valorativas poco precisas (Alastuey, 2024, p. 497). La consecuencia es la judicialización del discurso histórico o ideológico, incluso cuando no existe una incitación directa, concreta ni un riesgo claro para la paz pública.

En segundo lugar, advierte sobre el riesgo de instrumentalización política del derecho penal, en tanto, el uso de tipos como el de negacionismo puede responder más a fines simbólicos y reactivos frente a la presión social o mediática, que a una necesidad real de tutela penal (Alastuey, 2024, pp. 507-508).

En consecuencia, Alastuey no niega la gravedad del discurso negacionista, pero aboga por una interpretación del tipo penal restrictiva y conforme al principio de intervención mínima, compatible con los estándares del Tribunal Europeo de Derechos Humanos.

En el mismo sentido, Fuentes Osorio (2022) sostiene que la redacción excesivamente amplia, del artículo 510 genera ambigüedad jurídica y permite una aplicación desigual y, en ocasiones, arbitraria. Esta amplitud dificulta distinguir con claridad entre expresiones que incitan efectivamente al odio o a la violencia y aquellas que, si bien pueden resultar ofensivas, se encuentran amparadas por la libertad de expresión. Esta ambigüedad, lejos de fortalecer la protección de los colectivos vulnerables, puede debilitarla, ya que permite que los tribunales resuelvan los casos sin incorporar adecuadamente el contexto estructural de discriminación en que dichas expresiones ocurren (Fuentes Osorio, 2022, pp. 2-3).

Asimismo, el autor critica la judicialización del discurso de odio por su enfoque excesivamente individualista y descontextualizado, que no reconoce las condiciones estructurales de vulnerabilidad de los grupos afectados. En muchos casos, los jueces privilegian una visión formalista de la libertad de expresión, restando importancia al daño social que produce la reproducción de discursos estigmatizantes. A esto se suma una práctica judicial selectiva, que atiende con mayor frecuencia los casos más mediáticos, dejando sin respuesta institucional muchas expresiones de odio cotidiano y sistemático. En consecuencia, la estrategia penal, por sí sola, no garantiza una protección efectiva de los derechos fundamentales, y debe ir acompañada de políticas públicas orientadas a combatir la discriminación estructural y promover una cultura de derechos humanos (Fuentes Osorio, 2022, pp. 6-9).

A la luz del análisis comparado, resulta evidente que la sanción penal de los discursos de odio representa una herramienta legítima y, en muchos casos, necesaria para la protección de los derechos fundamentales de colectivos históricamente discriminados. No obstante, esta respuesta punitiva debe articularse de manera proporcional y excepcional, respetando el principio de mínima intervención penal y sin sacrificar garantías básicas como la libertad de expresión. La experiencia alemana, con su enfoque sistemático y acotado, contrasta con el caso español, donde la amplitud e imprecisión del tipo penal ha suscitado fundadas críticas en torno a la inseguridad jurídica y la potencial instrumentalización del derecho penal. En este contexto, cualquier intento de reprimir el discurso de odio debe equilibrar cuidadosamente los fines de protección frente a la discriminación con los riesgos de restricción indebida del debate público, asegurando que las respuestas penales no se conviertan en mecanismos de censura o represión ideológica. Solo así es posible sostener una política penal que sea eficaz, legítima y acorde con los estándares democráticos y de derechos humanos.

En este sentido resulta interesante la propuesta de Kaufman (2015), quien plantea un modelo gradual y proporcional que busca compatibilizar la protección de la libertad de expresión con la necesidad de sancionar aquellas manifestaciones que vulneran gravemente los derechos fundamentales y el orden democrático.

El autor señala que la noción de "discurso de odio" (*hate speech*) se reconceptualiza como *odium dictum*. Este término en latín, y sus variantes *(odium dicta, odia dictum, odia dicta*), busca denotar una opinión dogmática, injustificada y destructiva dirigida a grupos históricamente discriminados o a

personas por su pertenencia a estos (Kaufman, 2015, p. 47). A diferencia de una mera expresión de odio, un *odium dictum* implica una intención maligna y premeditada de humillar, denigrar o incitar a otros a la marginalización o exclusión de las víctimas (Kaufman, 2015, p. 139). Kaufman subraya que esta conceptualización precisa es crucial para diferenciarlo de otras "expresiones de odio" que, si bien desagradables, podrían estar protegidas por la libertad de expresión (Kaufman, 2015, p. 43).

Kaufman propone una clasificación tripartita de los *odium dicta* según su gravedad, con el fin de establecer respuestas jurídicas y sociales diferenciadas. Los de mayor gravedad, como la incitación directa al genocidio, al terrorismo o la apología de crímenes de lesa humanidad, justifican sanciones penales estrictamente delimitadas, bajo el principio de *ultima ratio*. Los de entidad intermedia, que si bien no incitan directamente a la violencia pueden persuadir a terceros a discriminar o excluir, deberían ser abordados por vías administrativas, a través de organismos especializados en materia antidiscriminatoria. Finalmente, los de menor gravedad, asociados a expresiones insensibles o socialmente ofensivas, no ameritan sanción jurídica, pero sí estrategias de respuesta a largo plazo, como la educación en derechos humanos y campañas de sensibilización. Esta propuesta busca proteger la dignidad de los grupos históricamente vulnerados, sin comprometer indebidamente la libertad de expresión, y exige, para una sanción penal, la concurrencia de múltiples criterios que acrediten una voluntad de humillar o excluir con impacto real y significativo (Kaufman, 2015, pp.179 ss.).

Discursos de odio en Chile. Límites y vacíos en la protección de los colectivos vulnerables

En Chile, el debate sobre la penalización de los discursos de odio ha adquirido una creciente relevancia en el contexto del fortalecimiento de los mecanismos jurídicos de protección frente a la discriminación y la violencia simbólica contra colectivos históricamente vulnerables. En efecto, si bien el Artículo 31[1] de la Ley sobre Libertades de Opinión e Información y Ejercicio del Periodismo (Ley N° 19.733) sanciona a quienes, por cualquier medio de comunicación social, realice publicaciones o transmisiones destinadas

1 Artículo 31, Ley N° 19.733 "El que, por cualquier medio de comunicación social, realizare publicaciones o transmisiones destinadas a promover odio u hostilidad respecto de personas o colectividades en razón de su raza, sexo, religión o nacionalidad, será penado con multa de veinticinco a cien unidades tributarias mensuales. En caso de reincidencia, se podrá elevar la multa hasta doscientas unidades tributarias mensuales"

a promover odio u hostilidad respecto de personas o colectividades por su raza, sexo, religión o nacionalidad, con multas que pueden aumentar en caso de reincidencia, esta norma ha tenido escasa aplicación ya que las personas afectadas por discursos de odio suelen optar por vías judiciales más eficaces, como el recurso de protección o acciones penales por injurias y calumnias.[2] [3]

Debido a lo inorgánico de las normas chilenas que reprimen estas conductas y a la necesidad de proteger a colectivos vulnerables, se han desarrollado diversas iniciativas legislativas orientadas a modificar dicho artículo, o incorporar normas penales con el propósito de ampliar su cobertura y dotarla de mayor eficacia sancionatoria.

Una de estas iniciativas de reforma ha sido el Proyecto de Ley contenido en el Boletín N.º 7130-07, que buscaba no solo redefinir el artículo 31, sino también introducir en el Código Penal un nuevo tipo penal autónomo de incitación al odio, así como agravantes específicas para delitos cometidos con móviles discriminatorios. Esta iniciativa legislativa dio lugar a un amplio debate jurídico y político, en el cual organismos como el Instituto Nacional de Derechos Humanos (INDH) manifestaron su apoyo a la necesidad de establecer sanciones frente a discursos de odio, pero advirtieron sobre los riesgos de una tipificación excesivamente amplia o ambigua, que pudiera afectar el ejercicio legítimo de la libertad de expresión.

Actualmente, en el Congreso chileno se encuentra en tramitación el Proyecto de Ley contenido en el Boletín N.º 11.949-17, presentado en 2019, cuyo objetivo es tipificar penalmente tanto la incitación a la violencia y a la discriminación, como el negacionismo de violaciones a los derechos humanos ocurridas en Chile entre 1973 y 1990. Esta iniciativa propone modifica-

2 Adicionalmente, diversas normas sancionan conductas discriminatorias, aunque no tipifican expresamente la incitación al odio. Entre ellas cabe destacar: la Ley N.º 20.609, que establece medidas contra la discriminación y agrega una agravante al artículo 12 del Código Penal por móviles discriminatorios; la Ley N.º 19.253, que sanciona la discriminación intencionada contra personas indígenas; la Ley N.º 20.422, que contempla sanciones por actos discriminatorios hacia personas con discapacidad; el Decreto Ley N.º 1.094, que impide el ingreso al país a extranjeros que fomenten doctrinas violentas o discriminatorias; y la Ley N.º 18.603, que sanciona a partidos y organizaciones que promuevan el odio o la discriminación.

3 Una revisión de la base jurisprudencial del Poder Judicial (Pjud) revela que, de un total de 54 sentencias de la Corte Suprema que mencionan la Ley N° 19.733, 39 corresponden a recursos de protección —en su mayoría rechazados— y el resto a causas por injurias y calumnias, que conllevan sanciones penales más severas, incluyendo penas privativas de libertad. Sin embargo, no se registra ninguna causa en que se haya invocado o aplicado el delito previsto en el artículo 31 de dicha ley, relativo a la incitación al odio.

ciones a la Ley N.º 20.609, al Código Penal y a la Ley N.º 20.393, estableciendo sanciones para quienes, mediante medios de difusión pública, inciten a la violencia física, promuevan el menosprecio o deshonra hacia personas por motivos discriminatorios (raza, orientación sexual, religión, entre otros), o nieguen, aprueben o justifiquen crímenes de lesa humanidad. Asimismo, contempla la responsabilidad penal de personas jurídicas cuando se valgan de estas expresiones para su beneficio. A pesar de su contenido innovador y su alineación con estándares internacionales, el proyecto permanece en su primer trámite constitucional desde su ingreso (Boletín N.º 11.949-17, 2019).

Si bien la iniciativa es loable en su objetivo de proteger bienes jurídicos como la dignidad, la honra y la paz social, el proyecto de ley presenta deficiencias conceptuales y problemas de graduación sancionatoria que lo alejan de los estándares doctrinales e internacionales que debieran orientar la regulación de esta materia.

En efecto, el proyecto de ley propone sancionar la incitación a la violencia física (Artículo 161-C) y la promoción de la deshonra o menosprecio (artículo 161-D) contra una vasta lista de categorías, incluyendo "ideología, opinión o filiación política o deportiva, la sindicación o participación en organizaciones gremiales o la falta de ellas, el trabajo que realiza". Kaufman advierte explícitamente contra listas tan amplias para la aplicación de sanciones penales. Sostiene que la protección legal contra los *odium dicta* debe circunscribirse a "grupos históricamente discriminados" que han sufrido "exclusiones estructurales en el largo plazo" (Kaufman, 2015, p. 157). La inclusión de la "opinión o filiación política o deportiva" o la "sindicación" como motivos para una sanción penal por discursos de odio es precisamente lo que Kaufman critica ya que puede ser utilizada para perseguir "ideas políticas alternativas, a los disidentes ideológicos o, peor aún, para amedrentar a quienes denuncian corrupción" (Kaufman, 2015, p. 47). En su visión, los políticos, por ejemplo, deben "mostrar un mayor grado de tolerancia" a la crítica, aunque sea ácida. La propuesta del proyecto, al criminalizar discursos basados en estas categorías, podría exceder el objetivo de proteger a los "débiles y perseguidos, las víctimas históricas de siempre" (Kaufman, 2015, p. 156).

Además, el proyecto parece carecer de una graduación clara y proporcional en las sanciones. Mientras que la incitación directa a la violencia física (Artículo 161-C) se alinea con la categoría de "mayor gravedad" de Kaufman que justifica sanciones penales severas, el Artículo 161-D sanciona con prisión la difusión de ideas para "promover la deshonra o menosprecio".

Kaufman clasifica los discursos "susceptibles de humillar y excluir" o "susceptibles de persuadir a terceros de discriminar, humillar y excluir" como de "mediana entidad", sugiriendo que sean tratados por la administración pública o "instancias administrativas especializadas" para evitar la "saturación de los tribunales" (Kaufman, 2015, pp. 145 y ss.).

Para corregir estas deficiencias, la propuesta debería revisar y acotar las categorías protegidas para la sanción penal, limitándolas estrictamente a grupos históricamente discriminados. Es crucial implementar un modelo de respuesta graduada, donde las penas privativas de libertad sean la *ultima ratio* para los casos de mayor gravedad (como la incitación directa a la violencia o el negacionismo), y los casos de "mediana entidad" —del artículo 161-D— sean gestionados mediante sanciones administrativas. Por su parte, los casos de menor gravedad —*dictum* insensible frente a la vulnerabilidad— deberían ser abordados mediante acciones educativas y campañas de sensibilización (Kaufman, 2015, p. 145). De este modo, el proyecto se beneficiaría de una clarificación en los umbrales de intencionalidad y malignidad para la criminalización. Kaufman exige que, para la aplicación penal, el emisor tenga una "intención deliberada de humillar o excluir" o que la "promoción del odio sea la consecuencia probable de su expresión". Por otro lado, la inclusión del negacionismo de graves violaciones a los derechos humanos (artículos 161-E y 161-F) sí es coherente con el marco de Kaufman, quien lo considera de la "mayor gravedad" y justificativo de "sanciones penales severas", en línea con la jurisprudencia europea que protege la "verdad histórica" y la dignidad de las víctimas. Este aspecto del proyecto refleja la importancia de evitar la impunidad en discursos que atentan contra la dignidad y la integración social (Kaufman, 2015, pp. 99 y ss.).

En resumen, si bien el proyecto chileno avanza en la protección contra discursos dañinos, su articulación, especialmente en las tipificaciones penales, podría beneficiarse de una mayor especificidad y una aplicación más estratificada de las sanciones, reservando el derecho penal para los discursos más graves que atentan directamente contra la dignidad de grupos históricamente vulnerables, y utilizando herramientas administrativas y educativas para el resto de las manifestaciones. Esto no solo sería más coherente con lo expresado por la doctrina, sino también con los principios de mínima intervención y proporcionalidad en derecho penal.

Conclusiones

La urgencia de una regulación clara y proporcional del discurso de odio en Chile no es meramente teórica. En los últimos años, diversas campañas políticas han evidenciado cómo, en ausencia de límites normativos efectivos, proliferan expresiones que estigmatizan a colectivos históricamente vulnerables, trivializan la violencia sexual, relativizan derechos fundamentales y reivindican discursos negacionistas o autoritarios. Estos ejemplos demuestran que, sin un marco jurídico que establezca límites precisos, el espacio político puede transformarse en un escenario de violencia simbólica impune, incompatible con la dignidad, la igualdad y la deliberación democrática. Si bien existen proyectos de ley orientados a sancionar el discurso de odio y el negacionismo, estos adolecen de deficiencias conceptuales y carecen de una adecuada graduación sancionatoria, lo que impone la necesidad de avanzar hacia una regulación más coherente con los estándares internacionales de derechos humanos.

Si bien la libertad de expresión constituye una garantía esencial en las democracias contemporáneas, su protección no puede traducirse en impunidad frente a discursos que refuerzan la discriminación estructural, legitiman la exclusión o incitan a la violencia. El desafío consiste en diseñar mecanismos de intervención estatal que, sin caer en censura previa ni en una expansión irrazonable del derecho penal, permitan sancionar aquellas expresiones que lesionan gravemente la dignidad de los grupos protegidos por el principio de igualdad.

Con base en el análisis realizado, se proponen como criterios normativos orientadores:

- La necesidad de distinguir entre tipos de discursos de odio, de modo que solo aquellos que impliquen un riesgo real e inminente —como la incitación directa a la violencia, la apología del genocidio o el negacionismo— sean objeto de sanción penal, conforme al principio de *ultima ratio*. La incorporación de sanciones administrativas y medidas reparadoras en casos de discurso de odio de mediana o baja gravedad, especialmente cuando afecten derechos de grupos históricamente discriminados.
- El fortalecimiento de políticas públicas en materia de educación en derechos humanos, alfabetización digital y cultura democrática como herramientas preventivas no punitivas.

En síntesis, la regulación penal del discurso de odio es necesaria, pero insuficiente por sí sola. Se requiere una respuesta integral que combine sanciones proporcionales con medidas educativas y preventivas, evitando tanto la inacción estatal como la sobrerreacción punitiva. Toda intervención debe guiarse por los principios de legalidad, necesidad y proporcionalidad, asegurando la protección efectiva de los colectivos vulnerables sin menoscabar las libertades esenciales del debate democrático. Aunque este estudio adopta un enfoque dogmático-normativo, sus hallazgos abren camino a futuras investigaciones empíricas y comparadas sobre el impacto real del discurso de odio y la eficacia de su regulación en entornos digitales.

Referencias bibliográficas

Ábrego Molina, V. H. y Flores Mérida, A. (2021). Datificación crítica: práctica y producción de conocimiento a contracorriente de la gubernamentalidad algorítmica. Dos ejemplos en el caso mexicano. *Revista Administración Pública y Sociedad*, (11), 211-229. https://acortar.link/1ykPo9

Abramovich, V. (2021). El límite democrático de las expresiones de odio. En V. Abramovich, M. J. Guembe y M. Capurro Robles (coords.), *El límite democrático de las expresiones de odio* (pp. 17-57). Tesseo y Universidad Nacional de Lanús.

Alkiviadou, N. (2023). Artificial intelligence and online hate speech moderation. *Sur International Journal on Human Rights, 32,* 105-117. https://acortar.link/1ykPo9

Almonacid-Díaz, C. (2022). Ética y actividad política. Discursos de odio, negacionismo y desinformación en la Convención Constitucional de Chile. *Revista Palabra y Razón,* (22), 60-62. https://doi.org/10.29035/pyr.22.56

Álvarez Bujan, M. V. A. (2023). Inteligencia artificial y medidas cautelares en el proceso penal: tutela judicial efectiva y autodeterminación informativa en potencial riesgo. *Revista Española de Derecho Constitucional, 127,* 177-207. https://doi.org/10.18042/cepc/redc.127.06

Alastuey Dobón, C. (2024). El Derecho penal ante el negacionismo. Comentario a la SAP de Barcelona de 9 de septiembre de 2024 (Caso de la Librería Europa III). *Revista de Derecho Penal y Criminología,* (32), 483-519. https://acortar.link/1ykPo9

Amaya López, C. y Cueva Gaibor, D. (2025). Inteligencia artificial y comunicación política en campañas electorales: mirada crítica, implicaciones y desafíos. *Revista Social Fronteriza, 5*(2), e687. https://doi.org/10.59814/resofro.2025.5(2)687

Astudillo Muñoz, J. L. (2023). Notas sobre la posverdad, los discursos de odio y la regresión de lo democrático. *Revista de Filosofía Jurídica, Social y Política,* (19), 401-427. https://acortar.link/1ykPo9

Boletín N.º 11.949-17. (2019). Modifica el Código Penal, y las leyes N.º 20.393 y 20.609, para sancionar el negacionismo respecto de las violaciones a los derechos humanos cometidas en Chile, y la incitación a la violencia y a la discriminación contra personas o grupos de personas. Congreso Nacional de Chile.

Bonilla Gutiérrez, J. C. . (2024). IA y Privacidad: Protegiendo la Autodeterminación Informativa en la Era Digital. *Revista de la Facultad de Derecho de México, 74*(290), 125-148. https://doi.org/10.22201/fder.24488933e.2024.290.89719

Castillo Morales, J. P. (2019). Metodología y comparación jurídica en el Derecho penal. La incidencia del Derecho comparado en la estructura de la dogmática jurídico-penal. Revista de Derecho, (246), 13-47. http://dx.doi.org/10.4067/S0718-591X2019000200013

Comité de Ministros del Consejo de Europa. (2022). Recomendación CM/Rec (2022)16 del Comité de Ministros a los Estados miembros sobre la lucha contra el discurso de odio. Consejo de Europa.

Corte Interamericana de Derechos Humanos. (2020). Caso Azul Rojas Marín y otra Vs. Perú. Sentencia de 12 de marzo de 2020.

Echiburú, G. G., Vargas, C. B. y Letelier, G. M. (2019). Análisis Crítico del Discurso: posicionamiento valorativo y discurso de odio en la discusión parlamentaria sobre la Ley de identidad de género. *Revista Latinoamericana de Estudios del Discurso, 19*(2), 71-92. https://acortar.link/1ykPo9

Fuentes Osorio, M. (2022). Hateful speech, vulnerabilidad y judicialización: los límites del derecho penal frente a los discursos de odio. *Quaderns de Dret,* (89), 1-17. https://doi.org/10.34617/ATPS-JJ84

García-García, J. (2025). Respuestas normativas para enfrentar el discurso del odio desde la experiencia europea [Regulatory responses to address hate speech from the European experience]. *European Public & Social Innovation Review,10*, 01-12. https://doi.org/10.31637/epsir-2025-1904

González Ruiz, M. J. (2023). *Evaluación de la regulación de los delitos de odio en el sistema penal chileno* [Tesis de grado, Universidad de Chile]. Repositorio Académico. https://acortar.link/1ykPo9

G'sell, F. Digital Authoritarianism: from state control to algorithmic despotism (January 30, 2025). http://dx.doi.org/10.2139/ssrn.5117399

Jacoby, A. X. (2020). Más que palabras: libertad de expresión y discurso de odio en el Sistema Interamericano de Derechos Humanos. *Eunomía. Revista en Cultura de la Legalidad, 18*, 148-163. https://doi.org/10.20318/eunomia.2020.5268

Instituto Nacional de Derechos Humanos. (2014). Informe sobre el Proyecto de Ley que tipifica el delito de incitación al odio racial y religioso (Boletín N.º 7130-07). Santiago de Chile. pp. 9-11.

Kaufman, G. (2015). *Odium dicta. Libertad de expresión y protección de grupos discriminados en internet.* Consejo nacional para prevenir la discriminación. https://acortar.link/1ykPo9

Lafont, C. (2025). La democracia deliberativa tras la transformación digital de la esfera pública. *Revista Latinoamericana sobre Democracia,* (0), 20-29. https://doi.org/10.22201/iis.rld.2025.00.4

Matamoros-Fernández, A. y Farkas, J. (2021). Racism, hate speech, and social media: A systematic review. *First Monday, 26*(2). https://acortar.link/1ykPo9

Mota, C. (2023). Inteligencia artificial y comunicación política: desafíos democráticos en la era algorítmica. Fundación Friedrich Ebert – Observatorio de Medios y Plataformas. https://acortar.link/1ykPo9

Nazmine, N., Tareen, M. K. y Noreen, S. (2021). Hate speech and social media: A systematic review. *International Journal of Innovation, Creativity and Change, 15*(8), 5285-5294.

Paúl Díaz, Á. (2011). La penalización de la incitación al odio a la luz de la jurisprudencia comparada. *Revista chilena de derecho, 38*(3), 573-609. https://dx.doi.org/10.4067/S0718-34372011000300007

Pereznieto Castro, L. (2020). La dogmática jurídica, con especial referencia al derecho internacional privado. Universidad Nacional Autónoma de México, Instituto de Investigaciones Jurídicas. https://revistas.juridicas.unam.mx

Prabhu, R. y Seethalakshmi, V. (2025). A comprehensive framework for multi-modal hate speech detection in social media using deep learning. *Scientific Reports, 15*, 13020. https://doi.org/10.1038/s41598-025-94069-z

Relatoría Especial para la Libertad de Expresión de la Comisión Interamericana de Derechos Humanos (RELE-OEA). (2009). *Marco jurídico interamericano*

sobre el derecho a la libertad de expresión. Organización de los Estados Americanos.

Rodríguez Zepeda, J. (2018) El peso de las palabras: libre expresión, no discriminación y discursos de odio. *El prejuicio y la palabra, 27*. https://acortar.link/1ykPo9v

Rubio, R., Alvin, F. F. y Andrade, V. (2024). *Inteligencia artificial y campañas electorales. Disfunciones informativas y amenazas sistémicas de la nueva Comunicación política.* Centro de Estudios Políticos y Constitucionales.

Sánchez Sánchez, A. M., Ruiz-Muñoz, D. y Sánchez Sánchez, F. J. (2023). Research trends in the control of hate speech on social media for the 2016-2022 time frame. *Cuadernos.info*, (56), 89-112. http://dx.doi.org/10.7764/cdi.55.60093

Sarasqueta, G., Ferrero, M., Olmedo, S., Rojas Montiel, E. R., Castillo Peñaherrera, C., Martínez Rodríguez, R. Maíz de Sotomayor, N., Salinas Goytia, J. R. y Ames Tineo de Saavedra, A. C. (2025). La construcción de narrativas políticas en campañas electorales sudamericanas: un análisis de las redes sociales [The construction of political narratives in South American electoral campaigns: a social media analysis]. *Revista Latina de Comunicación Social, 83*,1-22. https://www.doi.org/10.4185/RLCS-2025-2442

Serra Cristóbal, R. (2023). *Fake news* y el derecho a recibir información veraz: entre la libertad de expresión y la desinformación en la era digital. *Revista de Derecho Político 116*, enero-abril, 13-46. https://doi.org/10.5944/rdp.116.2023.37147

Sibrian, N., Alfaro, A. y Núñez, J. C. (2024). Validación de instrumento sobre exposición a discursos de odio de comunidades migrantes en el ecosistema mediático chileno: resultados preliminares [Validation of an instrument on exposure of migrant communities to hate speech in the Chilean media ecosystem: preliminary results]. *Revista Latina de Comunicación Social, 82,* 01-23. https://www.doi.org/10.4185/RLCS-2024-2226

Sibrian, N. y Labrador, M. J. (2024). Desinformación y marcos regulatorios en América Latina: desafíos en torno a discursos de odio, eliminación masiva de datos y derecho al olvido. Democracia y desinformación. Nuevas formas de polarización, discursos de odio y campañas en redes. Respuestas regulatorias de Europa y América Latina, 139-162. https://acortar.link/1ykPo9 v

Sosa Huapaya, A. (2024). Conflictos entre la autodeterminación informativa y la segmentación de perfiles a través de la publicidad programática online en el Perú. *Derecho & Sociedad,* (63), 285-298. https://doi.org/10.18800/dys.202402.019

Tapia Ballesteros, P. (2021). El discurso de odio del art. 510.1.a) del Código Penal español: la ideología como un Caballo de Troya entre las circunstancias sospechosas de discriminación. *Política Criminal, 16*(31), 284-320. https://acortar.link/1ykPo9

Valdés-Rivera, J. (2021). Libertad de expresión y derecho penal: el caso de los discursos de odio. In *Estándares para la protección de periodistas y personas defensoras de derechos humanos* (pp. 35-52). Tirant lo blanch.

Valdez Zepeda, A., Aréchiga, D. y Daza Marco, T. (2024). Inteligencia artificial y su uso en las campañas electorales en sistemas democráticos. *Revista Venezolana de Gerencia, 29*(105), 63-76. https://doi.org/10.52080/rvgluz.29.105.5

Declaración de Autoría - Taxonomía CRediT	
Autor	**Contribuciones**
María Lorena Rossel Castagneto	Conceptualización, Metodología, Software, Validación, Análisis formal, Investigación, Recursos, Curaduría de datos, Escritura-borrador original, Escritura-revisión y edición, Visualización, Supervisión, Administración del proyecto, Adquisición de fondos.

Declaración de Uso de Inteligencia Artificial
La autora DECLARA que la elaboración del artículo ***Regular el odio digital: entre la libertad de expresión y la protección de colectivos vulnerables en Chile,*** no contó con el apoyo de Inteligencia Artificial (IA).