REVISTA DE CIENCIAS SOCIALES Y HUMANAS XXI
NIVERSITAS

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

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

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

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

Received on: 11/06/2025 **Revised on:** 21/07/2025 **Approved on:** 22/08/2025 **Published on:** 01/09/2025

Suggested citation: Rossel Castagneto, M. L. (2025). Regulating digital hate: between freedom of expression and the protection of vulnerable groups in Chile. *Universitas XXI, 43*, pp. 93-119. https://doi.org/10.17163/uni.n43.2025.04

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 relate 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.

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.

Introduction

The use of artificial intelligence (AI) in political communication has reshaped the ways in which political actors interact with citizens, enabling the automated generation of personalized messages, the circulation of hyper-segmented content, and the amplification of polarizing narratives. Far from being neutral, this technological transformation has also facilitated the proliferation of hate speech in digital environments, especially against historically vulnerable groups such as migrants, women, and sexual dissidents.

The rise of AI-mediated discriminatory expressions raises fundamental legal tensions between freedom of expression and the state's duty to protect equality, human dignity, and democratic deliberation. In this context, the

question that guides this work is: how should the law respond to hate speech amplified by artificial intelligence in political contexts, without compromising freedom of expression?

The main hypothesis argues that the absence of adequate regulation creates a regulatory gap that allows messages harmful to democratic coexistence to circulate with impunity. It is therefore important to rethink the regulatory margins of state intervention, overcoming both the risk of inaction and the punitive overreaction.

This article addresses the problem from a legal-dogmatic and comparative perspective, examining the legal treatment of hate speech in Chile, Germany, and Spain, as well as the standards of the inter-American human rights system. Based on this, a set of normative criteria is proposed to address this phenomenon in a legitimate and proportionate way, in accordance with the principles of democratic constitutionalism.

Materials and method

This study uses a legal-dogmatic methodology, rooted in the tradition of constitutional and criminal law, aimed at identifying, interpreting, and systematizing norms, principles, and categories relevant to the legal treatment of hate speech in digital environments. As Pereznieto Castro (2020) says, this dogmatic approach allows evaluating the internal coherence of the legal system through the structuring principles of criminal law, such as legality, harmfulness, and culpability, which is especially useful for analyzing the legitimate limits of *ius puniendi* in the face of harmful communications. The approach is based on the recognition of hate speech as a controversial legal category, whose definition, limits, and control mechanisms require rigorous normative foundations based on positive law and fundamental rights.

However, as Castillo Morales (2019) warns, the object of criminal law cannot be analyzed in isolation or in a purely formal way. As it is an instrument of social control, its study requires the integration of social, cultural, and political dimensions that go beyond classical dogma. Hence, the author proposes a comparative and interdisciplinary methodology that strengthens the scientific nature of criminal dogma and avoids its technical-formal gap, allowing for a critical evaluation of the legitimacy of punitive power and its suitability for democratic ends. For this reason, a comparative approach is

used, focusing on the legal systems of Germany and Spain. The choice of these systems is justified both by their belonging to the continental legal tradition (based on codification and guarantees) and by their regulatory and jurisprudential experience in the area of freedom of expression and the punishment of hate speech. Germany offers a robust model with criteria of proportionality and exceptional application of criminal law; Spain, on the other hand, has a broader and more controversial criminal offense, subject to doctrinal debate about its limits. This comparison allows us to identify alternative regulatory models that illustrate the risks and virtues of different legal strategies.

The regulatory comparison allows us to respond directly to the second and third specific objectives of the study, related to the evaluation of international standards and the formulation of legal criteria for legitimate and proportionate state intervention.

The comparison criteria focus on three analytical dimensions:

- The criminal classification of hate speech and its substantive limits (typical elements, protected legal rights, degrees of harm).
- The proportionality and effectiveness of state control mechanisms (criminal, administrative, and educational).
- The potential impact on the exercise of fundamental rights, particularly freedom of expression and the principle of equality.

Similarly, the study covers both current legislation and relevant legislative proposals between 2010 and 2025, as well as recent national and international case law. In the case of Chile, Article 31 of Law No. 19,733, legislative bulletins on hate speech and denialism, and their articulation (or lack thereof) with the principle of minimal intervention in criminal law are analyzed.

The work integrates an interdisciplinary operational framework that articulates legal tools with contributions from political theory (deliberative democracy), digital communication (algorithmic virilization, post-truth), and legal sociology (structures of discrimination). These perspectives allow us to problematize the contemporary digital context and its effects on the structuring function of legal language in a democracy.

The primary sources analyzed include constitutional norms, criminal laws, bills, and national, inter-American, and European jurisprudence. A doctrinal and jurisprudential content analysis was used, employing coding matrices that allowed categorizing the levels of regulatory intervention, intensity of

discourse, condition of the affected group, and nature of the state response. The doctrinal and jurisprudential selection is based on criteria of thematic relevance, topicality, and reference in academic and legislative debates. Secondary sources, such as institutional reports, academic articles, and recent empirical studies are used for illustrative and contextual purposes, without aspiring to statistical representativeness, but with analytical value to enrich the legal analysis.

This approach allows us to respond to the objectives of the study: to identify regulatory gaps, evaluate relevant regulatory standards, and propose applicable legal criteria for legitimate state intervention against hate speech in digital environments, particularly in the Chilean context.

Results

This study has identified relevant findings that are organized into three levels: (1) conceptual-doctrinal results; (2) results derived from comparative legal analysis; and (3) propositional results oriented toward normative criteria for legitimate state intervention.

Doctrinal findings: normative challenges of digital hate

- Regulatory gaps in the face of the emergence of artificial intelligence (AI): AI acts as a catalyst for hate speech in digital environments, generating a form of symbolic violence that violates fundamental rights such as informational self-determination, equality, political participation, and human dignity. The absence of specific regulation regarding content generated or amplified by AI leaves a critical legal gap in current legal systems.
- Structural impact of digital hate: it has been observed that hate speech is no longer limited to direct incitement to violence, but operates as a normalized mechanism of exclusion and stigmatization, especially against migrants, women, and sexual dissidents. This discourse affects the effective exercise of democratic judgment, favoring forms of "digital authoritarianism by design".

Comparative findings: contrast between Germany, Spain, and Chile

The main elements are systematized using a comparative matrix focused on three variables:

Dimension analyzed	Germany	Spain	Chile
Criminal classification	§130 StGB (incitement to hatred), defined with restrictive criteria linked to human dignity	Art. 510 CP, broad and imprecise wording, includes speech against "ideology"	Art. 31 Law 19.733: weak administrative offense; no specific criminal offense
State control and role of platforms	NetzDG: requires platforms to remove content within short timeframes; combination of criminal and administrative penalties	Passive role of platforms; no specific law on algorithmic moderation	No regulation of platforms or rules on AI or the viralization of digital hate
Impact on fundamental rights	Guarantee-based model; proportionality and protection of dignity over hate are prioritized	Doctrinal criticism for ambiguity and legal uncertainty; risk of over-criminalization	High risk of impunity; low effectiveness of the legal framework to protect vulnerable groups

This comparative analysis shows that the German model offers a reasonable balance between punishment and guarantees, while the Spanish model presents risks of punitive ambiguity. Chile, on the other hand, exhibits a worrying regulatory gap.

Propositional results: regulatory criteria for state intervention

Based on the analysis carried out, the following regulatory criteria are proposed, aimed at legitimate and proportional regulation of hate speech in digital environments:

1. Differentiation by severity of speech (proposed typology):
    - High severity: direct incitement to violence, justification of crimes against humanity: criminal response.
    - Intermediate severity: misinformation with stigmatizing intention, indirect incitement: administrative response.

- Low severity: non-systematic offensive expressions: educational/preventive response.

2. Limitation of the use of criminal law to the principle of *ultima ratio*:
   - Only applicable when there is objective evidence of real and imminent risk to the fundamental rights of vulnerable groups.
3. Need for reforms to the Chilean framework:
   - Reformulation of Article 31 of Law No. 19,733.
   - Incorporation of an autonomous and limited criminal offense (based on *odium dictum*).
   - Regulation of the role of digital platforms, with obligations for transparent algorithmic moderation.
4. Inclusion of the principle of informational self-determination as a structural guarantee.
   - Especially in relation to the use of personal data for automated political segmentation without consent.
5. Comprehensive state intervention:
   - Combination of proportional sanctions, educational measures, and public policies for digital literacy, without prior censorship or democratic regression.

Conclusions and discussion

Political campaigns and deliberative deterioration in digital contexts

Contemporary political communication is characterized by a radical transformation of the traditional dynamics of interaction between political actors, the media, and citizens, based on the incorporation of digital technologies, artificial intelligence, and big data analysis (Mota, 2023). Indeed, artificial intelligence (AI) has profoundly transformed political communication in election campaigns, especially by intervening in the production, personalization, and dissemination of political discourse. This transformation, as Amaya and Cueva (2025) assert, responds to the use of algorithms that allow for audience segmentation and the generation of highly personalized messages, which redefines the interaction between political actors and citizens (pp. 3-4).

AI enables massive data processing and predictive capabilities that directly influence strategic campaign decision-making (Torres *et al.,* 2024, p. 84). Among other things, this translates into the traceability of campaigns, the detection of fake news, and the emotional manipulation of the electorate (Abrego and Flores, 2021, p. 6).

Ábrego and Flores cite what happened in Mexico as a paradigmatic case of the use of technologies—especially artificial intelligence (AI) and *big* data—for political purposes, highlighting their potential to manipulate information and control voters through what they call *algorithmic governmentality*. This form of power is based on the massive collection of personal data, often without consent, to profile, segment, and condition social and political behaviors for the benefit of governments and corporations (Ábrego and Flores, 2021, pp. 213-216).

Valdez *et al*. (2023) point out that the most relevant mechanisms for manipulating political campaigns and conditioning the electoral behavior of citizens are the use of "bots" and fake profiles, which can replicate messages, position trends on social networks, and attack opposing candidates, creating the illusion of social consensus or rejection (Valdez *et al.*, 2023, p. 33). Likewise, AI allows for the mass analysis of data from social media and other platforms to develop psychographic profiles of voters, as evidenced in the *Cambridge Analytica* case*,* where electoral preferences were manipulated based on the analysis of data obtained without consent (p. 31).

Added to this is the use of generative AI (GenAI) to create falsified visual content—such as war images or deepfakes—that distort the public's perception of reality (Rubio *et al.,* 2024, p. 7; De Rê, 2024, p. 12; Valdez *et al.,* p. 34; Alkiviadou, 2023).

This phenomenon raises serious ethical and legal implications. First, it affects the principle of informational self-determination, recognized by various international human rights standards. Second, it undermines the principle of electoral transparency and the right to receive accurate information, which are essential elements in a deliberative democracy. Finally, the lack of control over these technologies creates a regulatory gap that can enable forms of covert digital authoritarianism, as will be discussed below.

The principle of informational self-determination

The principle of informational self-determination guarantees individuals control over their personal data, allowing them to decide freely and in an

informed way what information can be processed, by whom, and for what purposes. It arises as a legal response to the risks of massive and automated data processing in digital environments mediated by artificial intelligence (AI), where individuals are reduced to objects of algorithmic analysis (Bonilla, 2022, pp. 288-289).

This principle, recognized by the German Federal Constitutional Court in its 1983 census ruling (Volkszählungsurteil), has been incorporated into international instruments such as the European Union's General Data Protection Regulation (GDPR) and the Council of Europe's Convention 108+, which impose obligations on states to ensure transparency, informed consent, and control over the use of personal data (Álvarez Buján, 2023; Bonilla, 2022).

In the context of political campaigns, the use of AI and micro-segmentation systems allows large volumes of data to be collected and processed without informed consent, seriously affecting this right. As Amaya and Cueva (2025) warn, citizens become targets of persuasive strategies designed on the basis of their profiles, which undermines privacy, freedom of information, and voting autonomy (pp. 5-6).

Furthermore, the same mechanisms that enable political manipulation facilitate the spread of hate speech directed against historically vulnerable groups. This reality requires understanding informational self-determination as part of a network of guarantees designed to protect not only individual privacy, but also dignity, equality, and democratic deliberation. In the face of these threats, it is necessary to move toward regulatory frameworks that recognize the structural value of this principle and allow for the sanctioning of expressions that, due to their intensity and effects, are incompatible with a democracy that respects human rights.

Electoral transparency and the right to accurate information

Electoral transparency is a structural condition of democratic constitutionalism, ensuring that the electoral process is conducted under clear rules, with equal access to information and effective oversight of the actors involved. In digital environments, this requirement takes on a new dimension in the face of the opacity of algorithms, the massive personalization of content, and the circulation of false information. As Serra (2023) points out, transparency in contexts of algorithmic governance must be understood as an active princi-

ple, which imposes on public and private actors the obligation to disclose the criteria, sources, and effects of their automated decisions.

This principle is linked to the right to accurate information, understood as access to relevant, pluralistic, and unmanipulated information, a condition for exercising informed and deliberative citizenship. The proliferation of political disinformation—enhanced by the use of AI—violates this right, distorts the electoral debate, and erodes public trust. In this context, transparency must be extended to digital platforms, whose algorithms act as invisible filters that define what users see, amplifying polarizing or discriminatory content.

Both the UN Human Rights Committee and the IACHR Special Rapporteur for Freedom of Expression have emphasized that states must adopt measures to ensure the transparency of digital processes that affect freedom of expression, without resorting to forms of censorship. This includes the duty to regulate digital political advertising, algorithmic moderation mechanisms, and the use of personal data in electoral campaigns.

Thus, electoral transparency in the digital age is not limited to traditional accountability, but it requires regulatory mechanisms that make visible what currently operates in an opaque and automated way, affecting democratic quality and the equal exercise of political rights.

Reflections on covert digital authoritarianism and post-truth

Digital authoritarianism refers to the systematic use of technologies to monitor, manipulate, or repress the population, not only by authoritarian regimes, but also in democracies and technology companies when they operate systems without accountability, creating "authoritarianism by design" (G'sell, 2025, pp. 3-5).

In this environment, disinformation becomes a tool of symbolic control. Lafont (2025) warns that the proliferation of manipulative content weakens the minimum conditions for democratic deliberation and installs echo chambers and algorithmic segmentation that undermine discursive pluralism. When citizens cannot distinguish whether the messages they receive come from fellow citizens or malicious actors, public trust and the legitimacy of the democratic system are eroded (Lafont, 2025, pp. 20-22). From the perspective of constitutional state theory, this prevents the rational recognition of norms, a requirement for the democratic validity of collective decisions. Therefore, strengthening reliable deliberate spaces—such as mini-publics or citizen as-

semblies—is not only an ethical need but also a condition for the survival of democracy (Lafont, 2025, pp. 26-28).

Added to this structural threat is the phenomenon of post-truth, which, according to Astudillo (2023), seeks to install distorted convictions as if they were true, affecting not only freedom of expression but also the possibility of exercising it in a critical and informed manner (pp. 403, 415). This instrumentalization of informational pluralism weakens democratic debate and, in contexts of stigmatization, legitimizes intolerant discourse and discriminatory practices, especially against migrants and sexual dissidents (pp. 417-418). Even rights such as public health can be compromised, as demonstrated by the spread of anti-vaccine discourse without scientific basis (p. 418).

In this scenario of information manipulation and deliberative deterioration, hate speech becomes particularly dangerous. Digital platforms have facilitated its circulation and legitimization under the rhetoric of freedom of expression, even though they reinforce structures of exclusion and symbolic violence. As the links between post-truth and digital authoritarianism show, the problem is no longer just the falsity of the content or the opacity of its origin, but its ability to reproduce dynamics of structural discrimination. It is therefore urgent to address its regulation not as censorship, but as a proportionate and necessary response to safeguard the dignity and fundamental rights of vulnerable groups.

Freedom of expression and hate speech in comparative terms: an approach to the inter-American and European human rights systems

In contemporary digital environments, hate speech has become a structural threat to human dignity, equality, and democratic pluralism, affecting historically vulnerable groups such as women, migrants, and sexual dissidents particularly intensely.

Although there is no single legal definition, international organizations and doctrinal studies describe it as any expression that incites or justifies violence, discrimination, or stigmatization on the basis of identity characteristics. On social media, this discourse takes explicit and subtle forms, it is amplified by algorithms, and distorts public opinion, undermining democratic deliberation and reinforcing structural prejudices. Its proliferation has had

varying effects, from victims leaving digital platforms to the normalization of patriarchal narratives in school settings. All of this highlights the urgent need for regulatory and social responses that combine proportional regulation, responsible moderation, and citizen education in digital environments (Alkiviadou, 2023; Prabhu and Seethalakshmi, 2025; Sibrián *et al.,* 2024; Marolla *et al.,* 2024; Sánchez *et al.,* 2023).

The regulation of hate speech presents a fundamental dilemma: on the one hand, the state's responsibility to protect human dignity, peaceful coexistence, and democracy, based on constitutional principles of equality and non-discrimination; on the other, the need to safeguard freedom of expression, which prohibits prior censorship and limits the regulation of the content of expression (Abramovich, 2022, pp. 88, 89).

The Inter-American Human Rights System (IAHRS) has developed a doctrine on freedom of expression that is characterized by its broad scope and an approach considered "hyperprotective" and would include the freedom to "seek, receive, and communicate ideas of all kinds" (Jacoby, 2020, p. 149).

In Advisory Opinion 5/85, the Inter-American Court of Human Rights (hereinafter IACHR), in interpreting Article 13 of the American Convention on Human Rights (hereinafter ACHR), recognized that freedom of expression has a dual dimension. The individual dimension refers to the right of each person to express their opinions and disseminate their thoughts by any means, and the social dimension is linked to its instrumental function for the exchange of ideas and information, and for mass communication between people.

It is crucial that both dimensions be guaranteed simultaneously. The Special Rapporteur for Freedom of Expression (RELE-OEA, 2004) has emphasized that "one cannot undermine one of them by invoking the preservation of the other as justification," which differentiates this model from the principle of balancing that predominates in the European system.

Jacoby (2020) adds that a fundamental pillar of freedom of expression in the inter-American human rights system is the prohibition of prior censorship, contained in Article 13.2 of the American Convention on Human Rights. The exercise of this right cannot be subject to prior censorship, except in specifically defined situations and under very strict conditions. Instead, the system favors a regime of subsequent responsibilities. These responsibilities must be expressly established by law, be necessary to ensure respect for the rights or reputation of third parties, or to protect national security, public order, or public health and morals (Jacoby, 2020, pp. 150-154).

Additionally, Article 13.5 of the ACHR explicitly refers to hate speech, prohibiting "any propaganda in favor of war and any advocacy of national, racial, or religious hatred that constitutes incitement to violence or any other similar illegal action against any person or group of people, for any reason, including those of race, color, religion, language, or national origin".

Abramovich (2022) points out that from the interpretation of these norms by the Rapporteur and the Inter-American Court, a classification of hate speech emerges to determine the levels of state interference, distinguishing between unprotected speech, protected speech, and specially protected speech.

The European model, and specifically the jurisprudence of the European Court of Human Rights (ECHR), on the other hand, assumes that there is a *continuum* between the expression of speech and the infringement of fundamental rights, which justifies the limitation of speech as a direct form of protection of those rights. This model seeks a balance, weighing freedom of expression and other fundamental rights, as well as public values such as peace and social coexistence (Rodríguez Zepeda, 2018, p. 47).

Abramovich argues that the ECHR recognizes the need to punish and prevent all forms of expression that spread, incite, promote, or justify hatred based on intolerance. To determine the levels of state interference and the type of responsibility, the ECHR and Recommendation CM/Rec(2022)16 of the Committee of Ministers to member states on combating hate speech establish a series of factors for assessing the seriousness of hate speech, such as the content of the speech, the political and spatial context, the speaker's intention, the speaker's role or status in society, the way in which the expression is disseminated or amplified, the capacity of the expression to cause harmful consequences, and the nature and size of the audience (Recommendation CM/Rec(2022)16, para. 32).

Unlike the European model, the Inter-American Human Rights System (IAHR) has not developed such an explicit and in-depth doctrine on hate speech, but has instead favored mechanisms of subsequent accountability (Jacoby, 2020, p. 150). However, a key ruling in which the Inter-American Court specifically referred to the nature of hate speech in relation to violence was in the case of Azul Rojas Marín *et al*. vs. Peru. In this ruling, the Inter-American Court of Human Rights deepened its understanding of prejudice-based violence, especially against LGBTI people, linking it directly to hate speech, since, in the Court's opinion, "this violence, fueled by hate speech, can lead to hate crimes" (CIDH, 2020, para. 93).

In this case, the Inter-American Court determined that strong prejudices against the LGBTI population existed and persist in Peruvian society, often leading to violence, including by state agents, and that violence against LGBTI people has a symbolic purpose: to communicate a message of exclusion or subordination to a specific group (CIDH, 2020, paras. 50-93). The Court concludes that regarding these situations, the State has a positive obligation to adopt measures to reverse or change existing discriminatory situations in its societies and to exercise special protection against the actions of third parties who, with its tolerance or acquiescence, violate the rights of vulnerable groups (CIDH, 2020, para. 89).

For its part, the RELE actively promotes non-punitive mechanisms and comprehensive public policies, focusing instead on alternatives to criminal law, such as the development of public policies and the creation of government commissions or committees to monitor them, promoting awareness and training, and ensuring that any legal restrictions are clear, precise, and proportionate, avoiding vagueness that could lead to abuse (Jacoby, 2020, p. 161).

Despite the disparity in criteria between the Inter-American Human Rights System, which favors a subsequent protection approach and emphatically rejects prior censorship, and the European model, which allows for the advance limitation of speech as a mechanism for direct protection against intolerance, both systems recognize the harmful potential of hate speech when it reinforces structures of discrimination and exclusion. As the Inter-American Court of Human Rights has stated in the case of Azul Rojas Marín vs. Peru, this type of expression not only reflects persistent social prejudices, but can also directly influence the commission of hate crimes, especially against historically vulnerable groups (CIDH, 2020, para. 93). Therefore, beyond doctrinal differences regarding the weighing of interests or subsequent liability, it is essential to move toward substantive regulation of hate speech, particularly the one that affects vulnerable groups. This regulation must balance the protection of freedom of expression with the need to guarantee conditions of equality and dignity, preventing discriminatory narratives from continuing to operate as normalized mechanisms of symbolic violence, social exclusion, and legitimization of intolerance. In this context, both the Inter-American Court of Human Rights and the European Court of Human Rights have agreed that states have a positive obligation to prevent and punish such expressions when their content, context, or impact can translate into real or symbolic attacks that threaten democratic coexistence and fundamental human rights.

The challenge of punishing hate: effectiveness and limits of criminal law in Germany and Spain

Valdés (2021, pp. 35-47) points out that, although the use of criminal law to limit freedom of expression is generally viewed with suspicion, there is a widespread rejection toward hate speech and an acceptance of its criminal punishment in many contemporary legal systems. In legal systems such as those of Germany, France, Italy, and Austria, human dignity plays a leading role, allowing criminal sanctions against expressions that constitute a direct attack on it or on the constitutional order. These systems are described as "militant" because they exclude certain expressions due to their content and the potential danger they pose, beyond any real danger. They share a sensitivity to protecting fundamental legal rights, such as public order (understood as public peace) and human dignity, even with criminal law as *ultima ratio* (Valdés, 2021, p. 47).

However, Valdés emphasizes that any restriction on freedom of expression poses a risk to democratic order and individual rights. For conduct to be prohibited and criminally punished, it must seriously affect legal rights, making criminally punishable expressions or statements as very specific exceptions; the general rule is that freedom of expression cannot be restricted criminally.

Germany has one of the most robust legislations in Europe in this area. In Germany, hate speech is classified in the Criminal Code (Strafgesetzbuch, StGB), particularly in §§ 130 and 185-187.

§130 StGB, entitled Volksverhetzung (incitement to popular hatred), criminally punishes anyone who incites hatred against parts of the population or calls for violence or arbitrary measures against them, as well as anyone who violates human dignity through insults, contempt, or slander against groups on the basis of ethnic origin, religion, nationality, or similar characteristics. This provision also considers Holocaust denial to be a specific form of incitement to hatred (González Ruiz, 2023, pp. 87-88).

Additionally, the Network Enforcement Act (NetzDG) requires digital platforms to promptly remove hate speech, under administrative penalty, which shows a comprehensive approach that covers both the criminal and administrative and digital spheres (González Ruiz, 2023, pp. 89-90).

For its part, the Spanish Criminal Code defines hate speech in Article 510, which punishes both those who directly or indirectly promote or incite hatred or violence against a group or individuals on grounds such as race, religion,

gender, disability, etc., as well as those who produce or disseminate material that promotes hatred or violence against these groups or individuals, or those who deny or glorify genocide and crimes against humanity.

Although this regulation aims to protect historically discriminated groups, it has been criticized for the ambiguous and broad terms used by the legislator to punish hate speech or denialism. Indeed, Tapia (2021) argues that the new wording of Article 510.1.a) is excessively broad and imprecise, creating legal uncertainty and allowing for an excessive expansion of criminal law. This is evident, for example, in the equating of radical or controversial speech—such as that expressed by artists, activists, or social media users—with actual incitement to violence or discrimination (Tapia, 2021, pp. 284-290).

Additionally, according to the author, the term "ideology" as a suspicious circumstance of discrimination has served to criminalize dissident or politically incorrect expressions. This has led to speeches directed against bullfighters, police officers, members of the Crown, or even people with neo-Nazi ideologies being considered hate crimes, which distorts the original meaning of anti-discrimination crimes and trivializes their protective function against historically marginalized groups (Tapia, 2021, pp. 297-304). For this reason, Tapia agrees with the proposal of the Criminal Policy Study Group to reform the criminal classification, limiting punishment to cases where there is direct and public incitement to commit specific crimes (against life, integrity, freedom, etc.) for discriminatory reasons, and only when there is an imminent risk of their commission. According to the author, this delimitation would allow the criminal offense to regain its legitimacy and respect the principles of minimal intervention and proportionality (Tapia, 2021, p. 314).

Likewise, Alastuey questions the current orientation of Article 510. First, the author argues that the breadth of the criminal offense, with expressions such as "fostering a climate of hatred," affects the principle of criminal legality by introducing vague evaluative clauses (Alastuey, 2024, p. 497). The consequence is the judicialization of historical or ideological discourse, even when there is no direct, concrete incitement or clear risk to public peace.

Secondly, the author warns of the risk of political instrumentalization of criminal law, insofar as the use of offenses such as denialism may respond more to symbolic and reactive purposes in the face of social or media pressure than to a real need for criminal protection (Alastuey, 2024, pp. 507-508).

Consequently, Alastuey does not deny the seriousness of denialist discourse, but advocates a restrictive interpretation of the criminal offense in

accordance with the principle of minimum intervention, compatible with the standards of the European Court of Human Rights.

Similarly, Fuentes Osorio (2022) argues that the excessively broad wording of Article 510 creates legal ambiguity and allows for unequal and, at times, arbitrary application. This breadth makes it difficult to clearly distinguish between expressions that effectively incite hatred or violence and those that, while offensive, are protected by freedom of expression. This ambiguity, far from strengthening the protection of vulnerable groups, can weaken it, as it allows courts to resolve cases without adequately incorporating the structural context of discrimination in which such expressions occur (Fuentes Osorio, 2022, pp. 2-3).

The author also criticizes the judicialization of hate speech for its excessively individualistic and decontextualized approach, which fails to recognize the structural conditions of vulnerability of the affected groups. In many cases, judges privilege a formalistic view of freedom of expression, downplaying the social harm caused by the stigmatizing discourse. Added to this is a selective judicial practice, which most often deals with the most high-profile cases, leaving many everyday expressions and systematic hate unaddressed by institutions. Consequently, the criminal strategy does not guarantee effective protection of fundamental rights and must be accompanied by public policies aimed at combating structural discrimination and promoting a culture of human rights (Fuentes Osorio, 2022, pp. 6-9).

In light of the comparative analysis, it is clear that the criminal punishment of hate speech is a legitimate and, in many cases, necessary tool for protecting the fundamental rights of historically discriminated groups. However, this punitive response must be proportionate and exceptional, respecting the principle of minimum criminal intervention and without sacrificing basic guarantees such as freedom of expression. The German experience, with its systematic and limited approach, contrasts with the Spanish case, where the breadth and imprecision of the criminal offense has given rise to well-founded criticism regarding legal uncertainty and the potential instrumentalization of criminal law. In this context, any attempt to suppress hate speech must carefully balance the goals of protection against discrimination with the risks of undue restriction of public debate, ensuring that criminal responses do not become mechanisms of censorship or ideological repression. Only in this way is it possible to maintain a criminal policy that is effective, legitimate, and in line with democratic and human rights standards.

In this regard, Kaufman's (2015) proposal is interesting, as proposing a gradual and proportional model that seeks to reconcile the protection of freedom of expression with the need to punish those expressions that seriously violate fundamental rights and democracy.

The author points out that the notion of "*hate speech*" is reconceptualized as *odium dictum.* This Latin term, and its variants *(odium dicta, odia dictum, odia dicta)*, seeks to denote a dogmatic, unjustified, and destructive opinion directed at historically discriminated groups or individuals because of their membership in these groups (Kaufman, 2015, p. 47). Unlike a mere expression of hatred, an *odium dictum* implies a malicious and premeditated intention to humiliate, denigrate, or incite others to marginalize or exclude the victims (Kaufman, 2015, p. 139). Kaufman emphasizes that this precise conceptualization is crucial to differentiate it from other "expressions of hatred" which, while unpleasant, may be protected by freedom of expression (Kaufman, 2015, p. 43).

Kaufman proposes a three-part classification of *odium dicta* according to their severity, with the aim of establishing differentiated legal and social responses. The most serious cases, such as direct incitement to genocide, terrorism, or the glorification of crimes against humanity, justify strictly defined criminal penalties, under the principle of *ultima ratio*. Those of intermediate severity, which, although they do not directly incite violence, may persuade third parties to discriminate or exclude, and should be addressed through administrative channels, via agencies specializing in anti-discrimination matters. Finally, those of lesser severity, associated with insensitive or socially offensive expressions, do not warrant legal sanctions, but rather long-term response strategies, such as human rights education and awareness campaigns. This proposal seeks to protect the dignity of historically vulnerable groups without unduly compromising freedom of expression, and requires, for criminal punishment, the concurrence of multiple criteria that demonstrate a willingness to humiliate or exclude with real and significant impact (Kaufman, 2015, pp. 179 ff.).

Hate speech in Chile. Limits and gaps in the protection of vulnerable groups

In Chile, the debate on the criminalization of hate speech has become increasingly relevant in the context of strengthening legal mechanisms to protect

historically vulnerable groups from discrimination and symbolic violence. Indeed, although Article 31[1] of the Law on Freedom of Opinion and Information and the Exercise of Journalism (Law No. 19,733) punishes those who, through any means of social communication, make publications or broadcasts intended to promote hatred or hostility towards individuals or groups on the basis of their race, sex, religion, or nationality, with fines that may increase in the event of recidivism, this rule has been rarely applied since those affected by hate speech often opt for more effective legal remedies, such as appeals for protection or criminal actions for libel and slander. [2][3]

Due to the ineffectiveness of Chilean regulations that punish such conduct and the need to protect vulnerable groups, various legislative initiatives have been developed to amend this article or incorporate criminal regulations with the aim of broadening its coverage and making it more effective in terms of sanctions.

One of these reform initiatives was the bill contained in Bulletin No. 7130-07, which sought not only to redefine Article 31, but also to introduce into the Criminal Code a new autonomous criminal offense of incitement to hatred, as well as specific aggravating circumstances for crimes committed with discriminatory reasons. This legislative initiative gave rise to a broad legal and political debate, in which organizations such as the National Human Rights Institute (INDH) expressed their support for the need to establish penalties for hate speech, but warned of the risks of an excessively broad or

1 Article 31, Law No. 19,733 "Anyone who, through any means of social communication, makes publications or broadcasts intended to promote hatred or hostility towards individuals or groups on the basis of their race, sex, religion, or nationality, shall be punished with a fine of twenty-five to one hundred monthly tax units. In the event of a repeat offense, the fine may be increased to two hundred monthly tax units."

2 In addition, various regulations punish discriminatory conduct, although they do not expressly classify incitement to hatred as a crime. Among these, the following are noteworthy: Law No. 20,609, which establishes measures against discrimination and adds an aggravating factor to Article 12 of the Criminal Code for discriminatory motives; Law No. 19,253, which punishes intentional discrimination against indigenous people; Law No. 20,422, which provides for penalties for discriminatory acts against people with disabilities; Decree Law No. 1,094, which prevents foreigners who promote violent or discriminatory doctrines from entering the country; and Law No. 18,603, which punishes parties and organizations that promote hatred or discrimination.

3 A review of the jurisprudential basis of the Judiciary (Pjud) reveals that of a total of 54 Supreme Court rulings that mention Law No. 19,733, 39 correspond to appeals for protection—most of which were rejected—and the rest to cases of libel and slander, which carry more severe criminal penalties, including imprisonment. However, there is no record of any case in which the offense provided for in Article 31 of that law, relating to incitement to hatred, has been invoked or applied.

ambiguous classification, which could affect the legitimate exercise of freedom of expression.

Currently, the Chilean Congress is considering the bill contained in Bulletin No. 11.949-17, presented in 2019, which aims to criminalize both incitement to violence and discrimination, as well as the denial of human rights violations that occurred in Chile between 1973 and 1990. This initiative proposes amendments to Law No. 20,609, the Criminal Code, and Law No. 20,393, establishing penalties for those who, through public media, incite physical violence, promote contempt or dishonor toward people on discriminatory grounds (race, sexual orientation, religion, among others), or deny, approve, or justify crimes against humanity. It also contemplates the criminal liability of legal entities when they use these expressions for their own benefit. Despite its innovative content and alignment with international standards, the bill remains in its first constitutional stage since its introduction (Bulletin No. 11,949-17, 2019).

While the initiative is commendable in its objective of protecting legal rights such as dignity, honor, and social peace, the bill has conceptual deficiencies and problems with the severity of penalties that distance it from the doctrinal and international standards that should guide the regulation of this matter.

In effect, the bill proposes to punish incitement to physical violence (Article 161-C) and the promotion of dishonor or contempt (Article 161-D) against a vast list of categories, including "ideology, political or sports opinion or affiliation, union membership or participation in trade organizations or lack thereof, and the work performed." Kaufman explicitly warns against such broad lists for the application of criminal sanctions. He argues that legal protection against *odium dicta* should be limited to "historically discriminated groups" that have suffered "long-term structural exclusion" (Kaufman, 2015, p. 157). The inclusion of "political or sporting opinion or affiliation" or "union membership" as grounds for criminal punishment for hate speech is precisely what Kaufman criticizes, as it can be used to persecute "alternative political ideas, ideological dissidents, or, worse still, to intimidate those who denounce corruption" (Kaufman, 2015, p. 47). In his view, politicians, for example, should "show a greater degree of tolerance" to criticism, even if it is harsh. By criminalizing speech based on these categories, the bill's proposal could exceed the objective of protecting "the weak and persecuted, the historical victims of all time" (Kaufman, 2015, p. 156).

Furthermore, the bill seems to lack clear and proportional gradation in its penalties. While direct incitement to physical violence (Article 161-C) aligns with Kaufman's category of "greater severity" that justifies severe criminal penalties, Article 161-D punishes the dissemination of ideas to "promote dishonor or contempt" with imprisonment. Kaufman classifies speech that is "likely to humiliate and exclude" or "likely to persuade third parties to discriminate, humiliate, and exclude" as "medium severity," suggesting that it be handled by the public administration or "specialized administrative bodies" to avoid "saturation of the courts" (Kaufman, 2015, pp. 145 et seq.).

To correct these deficiencies, the proposal should review and narrow the categories protected by criminal sanctions, limiting them strictly to historically discriminated groups. It is crucial to implement a graduated response model, where custodial sentences are the *last resort* for the most serious cases (such as direct incitement to violence or denialism), and cases of "medium severity"—under Article 161-D—are handled through administrative sanctions. For their part, less serious cases—insensitive *dictum* in the face of vulnerability—should be addressed through educational actions and awareness campaigns (Kaufman, 2015, p. 145). In this way, the bill would benefit from clarification of the thresholds of intent and malice for criminalization. Kaufman requires that, for criminal prosecution, the speaker have a "deliberate intention to humiliate or exclude" or that "the promotion of hatred be the likely consequence of their expression." On the other hand, the inclusion of denial of serious human rights violations (Articles 161-E and 161-F) is consistent with Kaufman's framework, which considers it to be of the "utmost gravity" and justifying "severe criminal penalties," in line with European case law that protects "historical truth" and the dignity of victims. This aspect of the bill reflects the importance of preventing impunity in speech that undermines dignity and social integration (Kaufman, 2015, pp. 99 et seq.).

In summary, although the Chilean bill advances in the protection against harmful speech, its articulation, especially in criminal classifications, could benefit from greater specificity and a more stratified application of sanctions, reserving criminal law for the most serious speech that directly attacks the dignity of historically vulnerable groups, and using administrative and educational tools for other forms of expression. This would not only be more consistent with legal doctrine, but also with the principles of minimal intervention and proportionality in criminal law.

Conclusions

The urgency of clear and proportional regulation of hate speech in Chile is not merely theoretical. In recent years, various political campaigns have shown how, in the absence of effective regulatory limits, expressions that stigmatize historically vulnerable groups, trivialize sexual violence, relativize fundamental rights, and vindicate denialist or authoritarian discourse. These examples demonstrate that, without a legal framework that establishes precise limits, the political arena can become a stage for unpunished symbolic violence, incompatible with dignity, equality, and democratic deliberation. Although there are bills aimed at punishing hate speech and denialism, these suffer from conceptual deficiencies and lack adequate sanctions, which makes it necessary to move toward regulation that is more consistent with international human rights standards.

Although freedom of expression is an essential guarantee in contemporary democracies, its protection cannot translate into impunity for speech that reinforces structural discrimination, legitimizes exclusion, or incites violence. The challenge is to design mechanisms for state intervention that, without resorting to prior censorship or an unreasonable expansion of criminal law, make it possible to punish expressions that seriously harm the dignity of groups protected by the principle of equality.

Based on the analysis carried out, the following normative guidelines are proposed:

- The need to distinguish between types of hate speech, so that only those that pose a real and imminent risk—such as direct incitement to violence, advocacy of genocide, or denialism—are subject to criminal sanctions, in accordance with the principle of *ultima ratio*. The incorporation of administrative sanctions and remedial measures in cases of medium or low severity hate speech, especially when it affects the rights of historically discriminated groups.
- Strengthening public policies on human rights education, digital literacy, and democratic culture as preventive, non-punitive tools.

In short, criminal regulation of hate speech is necessary but insufficient on its own. A comprehensive response is required that combines proportionate sanctions with educational and preventive measures, avoiding both state inaction and punitive overreaction. All interventions must be guided by the principles

of legality, necessity, and proportionality, ensuring the effective protection of vulnerable groups without undermining the essential freedoms of democratic debate. Although this study adopts a dogmatic-normative approach, its findings pave the way for future empirical and comparative research on the real impact of hate speech and the effectiveness of its regulation in digital environments.

References

Ábrego Molina, V. H. and 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. and 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. and 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. and 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. and 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. and Seethalakshmi, V. (2025). A comprehensive framework for multimodal 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. and 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. and 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. and 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. and 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. and 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. and 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

Declaration of Authorship - CRediT Taxonomy	
Author	**Contributions**
María Lorena Rossel Castagneto	Conceptualization, Methodology, Software, Validation, Formal Analysis, Research, Resources, Data Curation, Original Draft Writing, Writing-Revision and Editing, Visualization, Supervision, Project Management, Fund Acquisition.

Declaration of Use of Artificial Intelligence
The author DECLARES that the article ***Regulating digital hate: between freedom of expression and the protection of vulnerable groups in Chile*** was not created with the support of Artificial Intelligence (AI).