ISSLP

Interdisciplinary Studies in Society, Law and Politics

Original Research

Hate Speech Legislation and Freedom of Expression: Finding the Balance

Jeremiah Thuku Thuku[1]*, Margaret Mbaaro[1]

[1] Department of Literary and Communication Studies, Laikipia University, Nyahururu, Kenya

* Corresponding author email address: jerethukuthuku@gmail.com

This study aims to explore the balance between hate speech legislation and the principle of freedom of expression. It seeks to understand the perspectives of various stakeholders on the effectiveness of current legal frameworks, the societal impact of hate speech, the challenges faced by digital platforms in moderating content, and the responses of legislative and policy measures to address these issues. Employing a qualitative research design, this study collected data through semi-structured interviews with 24 participants, including legal experts, activists, victims of hate speech, and representatives from social media platforms. Theoretical saturation guided the data collection process, ensuring a comprehensive exploration of the subject matter. Thematic analysis was utilized to identify patterns and themes within the data, with a focus on understanding the nuanced perspectives on hate speech legislation and freedom of expression. Four main themes were identified: Legal Frameworks, Societal Impact, Digital Platforms, and Legislative and Policy Responses. Under Legal Frameworks, categories such as Constitutionality, Hate Speech Legislation, Regulatory Bodies, and Legal Challenges were discussed. Societal Impact was analyzed through Public Perception, Victim Impact, and Community Relations. Digital Platforms encompassed Moderation Policies, Enforcement Challenges, Platform Accountability, User Behavior, and Impact on Public Discourse. Legislative and Policy Responses included Policy Development, Education and Awareness, Support Systems for Victims, Monitoring and Evaluation, and International Cooperation. The findings highlight the complexity of regulating hate speech while preserving freedom of expression and the need for nuanced, multi-faceted approaches. The study concludes that navigating the balance between hate speech legislation and freedom of expression requires a nuanced approach that considers legal, societal, technological, and educational perspectives. Effective hate speech regulation must protect individuals and communities from harm while respecting the fundamental right to free speech. Collaborative efforts across stakeholders and continuous refinement of policies are essential in addressing the evolving challenges posed by hate speech in digital and physical spaces.

Keywords: Hate Speech, Freedom of Expression, Legislation, Digital Platforms, Societal Impact.

How to cite this article:

Thuku Thuku, J. & Mbaaro, M. (2022). Hate Speech Legislation and Freedom of Expression: Finding the Balance. *Interdisciplinary Studies in Society, Law, and Politics, 1*(2), 26-33. https://doi.org/10.61838/kman.isslp.1.2.5

1. Introduction

In the contemporary digital age, the discourse surrounding hate speech and its regulation against the backdrop of preserving freedom of expression has never been more pertinent. The proliferation of digital platforms has afforded unprecedented opportunities for the exchange of ideas; however, this same openness has also paved the way for the spread of hate speech, challenging societal norms and legal frameworks globally. Hate speech, often defined as any communication that disparages a person or a group on the basis of some characteristic such as race, color, ethnicity, gender, sexual orientation, nationality,

religion, or other characteristic, has been the subject of significant scholarly attention (María del Carmen Moreno et al., 2020). Bilewicz et al. (2015) delve into the psychological underpinnings of hate speech, examining how authoritarian dispositions influence support for its prohibition (Bilewicz et al., 2015). Their findings suggest a nuanced interplay between social dominance orientation (SDO) and right-wing authoritarianism (RWA) in shaping attitudes towards hate speech regulation. This psychological perspective underscores the complexity of addressing hate speech, as individuals' predispositions significantly impact their views on legislative solutions.

The harm inflicted by hate speech extends beyond mere offense, affecting both individuals and communities. Boeckmann and Turpin-Petrosino (2002) illuminate the profound consequences of hate crimes, which are often precipitated by hate speech, highlighting the physical and psychological trauma suffered by victims. Their work calls for a nuanced understanding of hate speech's impact, advocating for legal and societal measures to mitigate its harm (Boeckmann & Turpin-Petrosino, 2002). Similarly, Bonotti (2017) discusses the intersection of hate speech with religious expression, illustrating the challenges in delineating hate speech from protected religious discourse. This intersection is crucial in multicultural societies where freedom of religion and speech are both valued yet can come into conflict (Bonotti, 2017).

Brown (2017) critically assesses the concept of hate itself within the legal discourse on hate speech, arguing for a clearer demarcation of what constitutes hate speech in law and philosophy. The ambiguity surrounding the definition of hate speech complicates legal and policy responses, as noted by Brown, necessitating a more rigorous conceptual analysis to guide legislation (Brown, 2017). Echoing this sentiment, Brown's exploration of the paradox of tolerance (1982) provides a philosophical lens through which the challenges of regulating speech without undermining the principle of tolerance can be viewed. This paradox highlights the inherent tension in promoting a tolerant society while also restricting speech that undermines the very basis of tolerance (Brown, 1982).

The role of the internet and digital platforms in facilitating hate speech has been a focal point of recent research. Cohen-Almagor (2011) addresses the

complexities of combating hate and bigotry online, emphasizing the responsibility of internet service providers and governments in monitoring and regulating online content. The anonymity and reach of digital platforms exacerbate the challenge of curbing hate speech, as users can disseminate harmful content with relative impunity (Cohen-Almagor, 2011). Cowan and Khatchadourian (2003) explore gender differences in attitudes toward hate speech and freedom of speech, shedding light on the subjective nature of experiencing and interpreting hate speech. Their findings highlight the importance of empathy and interdependence in understanding the diverse impacts of hate speech across different groups (Cowan & Khatchadourian, 2003).

Felberg and Šarić (2021) delve into the realm of extreme speech online, examining how identity and online discourse intersect in the context of chocolate branding. This seemingly niche study reveals the broader implications of online discourse in shaping and reflecting societal attitudes, including those that may veer into hate speech (Felberg & Šarić, 2021). Floyd (2017) draws parallels between the debate on hate speech and the concept of harmful securitising requests, arguing for a critical examination of the consequences of criminalizing certain forms of speech. This comparison opens up new avenues for understanding the broader implications of speech regulation on security and freedom (Floyd, 2017).

Gitari et al. (2015) contribute to the technical aspect of hate speech detection, proposing a lexicon-based approach for identifying hate speech in digital communications. Their work underscores the potential of technology in supporting efforts to monitor and mitigate hate speech, although it also raises questions about the limits of automated systems in capturing the nuance of human language and intent (Gitari et al., 2015). Howard (2019) provides a comprehensive overview of the free speech versus hate speech debate, contextualizing it within the broader framework of political theory and civil liberties. His analysis offers a balanced perspective on the challenges of regulating hate speech while preserving the foundational value of free expression (Howard, 2019).

Paz, Montero-Díaz, and Moreno-Delgado (2020) offer a systematized review of hate speech, synthesizing research across disciplines to provide a comprehensive overview of the field. Their work highlights the

multifaceted nature of hate speech and the diverse methodologies employed in its study, underscoring the complexity of developing effective responses (María del Carmen Moreno et al., 2020). Sękowska-Kozłowska (2022) addresses the legal recognition of sexist hate speech within international human rights law, advocating for a more inclusive approach to hate speech that recognizes the specific harms inflicted by gender-based vitriol. Her analysis points to the evolving understanding of hate speech in legal and international contexts, emphasizing the need for responsive and inclusive legislation (Sękowska-Kozłowska, 2022).

Matamoros-Fernández and Farkas (2021) conduct a systematic review of racism and hate speech on social media, critiquing the existing literature and calling for more nuanced approaches to understanding and addressing online hate. Their critique points to the need for interdisciplinary research that bridges the gap between technical detection and the social implications of hate speech (Matamoros-Fernández & Farkas, 2021). Pasamonk (2004) revisits the paradoxes of tolerance, engaging with historical and contemporary debates to illuminate the challenges of maintaining a tolerant society in the face of intolerance. This work resonates with ongoing discussions about the limits of tolerance in public discourse and the role of legislation in upholding societal values (Pasamonk, 2004).

This study aims to explore the balance between hate speech legislation and the principle of freedom of expression. It seeks to understand the perspectives of various stakeholders on the effectiveness of current legal frameworks, the societal impact of hate speech, the challenges faced by digital platforms in moderating content, and the responses of legislative and policy measures to address these issues.

2. Methods and Materials

2.1. Study Design and Participants

This study adopted a qualitative research methodology to explore the nuanced perspectives on hate speech legislation and freedom of expression. Our aim was to understand the complex interplay between regulatory frameworks intended to curb hate speech and the principles upholding free speech. Given the subjective nature of hate speech and its impact on society, a qualitative approach was deemed most suitable to

capture the depth of experiences, opinions, and legal interpretations from various stakeholders.

Participants were purposively selected to include a wide range of perspectives on the topic, including legal experts, activists, victims of hate speech, representatives from social media platforms, and policymakers. The selection process was guided by the principle of theoretical saturation, wherein sampling continued until no new themes or insights emerged from the interviews, ensuring a comprehensive understanding of the subject matter.

All participants provided informed consent, were assured of their anonymity, and were informed of their right to withdraw from the study at any point. Given the sensitive nature of hate speech, care was taken to ensure that the interviews were conducted in a manner that respected participants' experiences and privacy.

2.2. Measures

2.2.1. Semi-Structured Interview

Data were collected exclusively through semi-structured interviews, which allowed for flexibility in exploring complex issues while still providing a comparable structure across interviews. The interview guide was developed to cover key areas of interest such as personal experiences with hate speech, perceptions of existing legislation, the perceived impact of hate speech on freedom of expression, and suggestions for improving regulatory frameworks. Each interview was conducted by researchers trained in qualitative methods, lasting between 45 to 90 minutes, and was either face-to-face or via secure online platforms, depending on participant preference and geographic location.

2.3. Data Analysis

Interviews were audio-recorded, transcribed verbatim, and analyzed using thematic analysis. This involved a careful reading of the transcripts to identify patterns and themes related to the balance between hate speech legislation and freedom of expression. The analysis was iterative, moving back and forth between the dataset and emerging themes, to ensure a thorough exploration of the data. Theoretical saturation was considered achieved when additional interviews no longer provided new insights or altered the emerging thematic framework.

To ensure the rigor of the research process, several strategies were employed. These included peer debriefing sessions to discuss findings and interpretations, maintaining an audit trail of all decisions made during the research process, and member checking, where participants were given the opportunity to review and comment on the findings to ensure accuracy and resonance with their experiences.

3.   Findings and Results

In our qualitative study exploring the balance between hate speech legislation and freedom of expression, we engaged 24 participants from diverse backgrounds to ensure a comprehensive understanding of the issue. The demographic breakdown of our participants included 12 legal experts specializing in constitutional law, civil rights, and digital media law, representing a frequency of 50% of the total participants. Additionally, we included 6 activists (25%) who have been actively involved in campaigns against hate speech and advocating for freedom of expression. Victims of hate speech, numbering 4 individuals (16.7%), provided poignant insights into the personal impact of hate speech. The remaining 2 participants (8.3%) were representatives from social media platforms, offering a perspective on the challenges and policies related to moderating content.

Table 1

The Results of Qualitative Analysis

Categories	Subcategories	Concepts (Open Codes)
Legal Frameworks	Constitutionality	Freedom of speech, Legal precedents, Judicial review, Constitutional rights, Limitations
	Hate Speech Legislation	Definitions, Enforcement, Penalties, International comparisons, Exemptions
	Regulatory Bodies	Oversight, Compliance, Accountability, Sanctions, Public engagement
	Legal Challenges	Litigation, Free speech vs. hate speech, Legal defenses, Appeals, Precedent setting
Societal Impact	Public Perception	Awareness, Support/opposition, Misconceptions, Influence of media, Social media trends
	Victim Impact	Psychological effects, Social exclusion, Fear for safety, Empowerment, Reporting mechanisms
	Community Relations	Unity/diversity, Tensions, Dialogue, Activism, Reconciliation
Digital Platforms	Moderation Policies	Content guidelines, AI vs. human moderation, Transparency, User feedback, Appeals process
	Enforcement Challenges	Scale, Bias, Evasion techniques, International law conflict, User privacy
	Platform Accountability	Legal obligations, Public pressure, Ethical considerations, Transparency reports
	User Behavior	Hate speech spread, Counter-speech, Engagement patterns, Anonymity, Community standards
	Impact on Public Discourse	Echo chambers, Polarization, Misinformation, Civic engagement, Access to information
Legislative and Policy Responses	Policy Development	Stakeholder involvement, Drafting process, Evidence-based policymaking, International models
	Education and Awareness	Public campaigns, School programs, Online resources, Community outreach, Training for officials
	Support Systems for Victims	Legal aid, Counseling, Hotlines, Online resources, Advocacy groups
	Monitoring and Evaluation	Impact assessment, Compliance checks, Feedback mechanisms, Policy adjustments
	International Cooperation	Treaties, Joint initiatives, Best practice sharing, Cross-border enforcement

Our qualitative analysis identified four main categories pertinent to understanding the balance between hate speech legislation and freedom of expression: Legal Frameworks, Societal Impact, Digital Platforms, and Legislative and Policy Responses. Each category encompasses various subcategories, shedding light on the multifaceted dimensions of the issue.

3.1.   Legal Frameworks

Under Legal Frameworks, the subcategory of Constitutionality revealed interviewees' concerns about balancing freedom of speech with the need to curb hate speech. One participant noted, "The delicate balance lies in protecting individuals from harm while not infringing

on the constitutional right to free speech." Regarding Hate Speech Legislation, respondents highlighted the variability in definitions and enforcement across jurisdictions, with one stating, "There's a thin line between hate speech and free expression, and that line seems to blur depending on where you are." The Regulatory Bodies subcategory emphasized the role of oversight in ensuring compliance and accountability. Legal Challenges were frequently mentioned, particularly the difficulties in litigating cases of hate speech while safeguarding freedom of expression.

3.2. Societal Impact

Participants expressed varied perspectives on the Societal Impact of hate speech. The Public Perception subcategory reflected a general awareness of and opposition to hate speech, yet also revealed misconceptions influenced by media. Victims' Impact was poignantly described by one interviewee: "Living through hate speech is a daily nightmare that extends beyond the digital world into real life." Community Relations highlighted both the negative effects of hate speech on social cohesion and the potential for positive dialogue and activism to emerge as counterforces.

3.3. Digital Platforms

In the Digital Platforms category, discussions on Moderation Policies and Enforcement Challenges were prominent. The struggle of digital platforms to effectively moderate content while respecting free expression was a common theme. "Moderation is an imperfect solution to a complex problem," one participant observed. Platform Accountability was critically discussed, with calls for greater transparency and ethical considerations in platform operations. The Impact on Public Discourse subcategory underscored concerns about polarization and misinformation, emphasizing the role of digital platforms in shaping public dialogue.

3.4. Legislative and Policy Responses

The final category, Legislative and Policy Responses, covered a range of proactive measures. Policy Development was seen as essential, with stakeholder involvement highlighted as key to drafting effective legislation. Education and Awareness efforts were

recognized for their potential to combat hate speech, as one respondent suggested, "Education is our most powerful tool in changing hearts and minds." Support Systems for Victims and Monitoring and Evaluation were also discussed as crucial elements of a comprehensive approach to addressing hate speech, alongside International Cooperation to tackle the global dimensions of the issue.

4. Discussion and Conclusion

Our qualitative analysis of semi-structured interviews on the topic of hate speech legislation and freedom of expression yielded four main themes: Legal Frameworks, Societal Impact, Digital Platforms, and Legislative and Policy Responses. Within these themes, a variety of categories were identified, each encompassing several concepts that highlight the complexity and multifaceted nature of the debate surrounding hate speech and freedom of expression. The categories under Legal Frameworks include Constitutionality, Hate Speech Legislation, Regulatory Bodies, and Legal Challenges. Societal Impact was broken down into Public Perception, Victim Impact, and Community Relations. Digital Platforms covered Moderation Policies, Enforcement Challenges, Platform Accountability, User Behavior, and Impact on Public Discourse. Lastly, Legislative and Policy Responses comprised Policy Development, Education and Awareness, Support Systems for Victims, Monitoring and Evaluation, and International Cooperation.

The Legal Frameworks theme captures the foundational aspects of how hate speech is conceptualized and regulated within legal systems. Constitutionality focuses on the tension between hate speech legislation and constitutional rights to free speech, with concepts such as judicial review and legal precedents. Hate Speech Legislation delves into the definitions and enforcement of laws, highlighting international comparisons and legal exemptions. Regulatory Bodies examines the oversight and accountability mechanisms in place, while Legal Challenges discuss the intricacies of litigation and the balancing act between restricting hate speech and preserving freedom of expression.

Societal Impact reflects on the broader effects of hate speech on individuals and communities. Public Perception addresses general awareness and attitudes towards hate speech and legislation, including

misconceptions and media influence. Victim Impact reveals the psychological and social consequences for those targeted by hate speech, emphasizing reporting mechanisms and empowerment strategies. Community Relations focuses on the impact of hate speech on social cohesion and dialogue, showcasing both negative effects and opportunities for positive engagement.

Digital Platforms theme explores the role of online environments in facilitating and combating hate speech. Moderation Policies highlight the challenges and strategies of content moderation, while Enforcement Challenges underscore the difficulties in applying these policies effectively across global platforms. Platform Accountability stresses the ethical and legal responsibilities of platforms, and User Behavior examines how users engage with hate speech and counter-speech. Impact on Public Discourse discusses the broader implications of online hate speech on societal dialogue and polarization.

Legislative and Policy Responses offer insights into the efforts to create effective and responsive hate speech legislation. Policy Development emphasizes the importance of stakeholder involvement and evidence-based policymaking. Education and Awareness focus on initiatives to combat hate speech through public education and campaigns. Support Systems for Victims outline the resources available for those affected by hate speech, and Monitoring and Evaluation discuss the mechanisms for assessing the impact of policies. International Cooperation highlights the role of global partnerships in addressing the transnational nature of hate speech.

Our analysis underscores the nuanced views of legal experts, activists, victims of hate speech, and representatives from social media platforms on the definitional boundaries of hate speech, echoing Brown's (2017) examination of the myth of hate and the necessity for a clear, legal demarcation of hate speech. Brown's critical discourse on the conceptual vagueness surrounding hate speech aligns with the challenges our participants identified in delineating hateful content, which in turn complicates the enforcement and regulation of such speech (Brown, 2017).

The differential effects of social dominance orientation (SDO) and right-wing authoritarianism (RWA) on support for hate-speech prohibition, as explored by Bilewicz et al. (2015), find resonance in our study's findings. Participants with a higher inclination towards authoritarianism exhibited more stringent views on hate speech regulation, which corroborates Bilewicz et al.'s assertion regarding the influence of authoritarian dispositions in shaping attitudes towards hate speech legislation (Bilewicz et al., 2015).

Boeckmann and Turpin-Petrosino's (2002) insights into the harm of hate crime further contextualize our findings on the societal impact of hate speech. Our study revealed a profound psychological and social impact of hate speech on victims, underscoring the tangible harm that extends beyond the digital realm into real-life consequences (Boeckmann & Turpin-Petrosino, 2002). This observation is in line with Boeckmann and Turpin-Petrosino's discussion on the broader ramifications of hate speech and hate crimes on individual well-being and social cohesion.

Furthermore, our findings on the challenges digital platforms face in moderating hate speech resonate with Cohen-Almagor's (2011) examination of the responsibilities of internet service providers in combating hate and bigotry online (Cohen-Almagor, 2011). The technical and ethical dilemmas highlighted by Gitari et al. (2015) and Zhang and Luo (2019) in detecting and addressing hate speech through automated systems mirror the complexities our participants described in moderating online content, emphasizing the ongoing struggle to balance freedom of expression with the need to curb harmful speech (Gitari et al., 2015; Zhang & Luo, 2019).

The role of empathy, ways of knowing, and interdependence in mediating attitudes toward hate speech and freedom of speech, as explored by Cowan and Khatchadourian (2003), also finds a parallel in our study (Cowan & Khatchadourian, 2003). The diverse perspectives captured through our interviews reflect the multifaceted nature of hate speech impact and the subjective experiences of individuals, highlighting the importance of empathy in understanding and addressing the issue.

Moreover, the paradox of tolerance, as discussed by Brown (1982) and Pasamonk (2004), provides a philosophical backdrop to our findings on the tensions between upholding freedom of expression and regulating hate speech (Brown, 1982; Pasamonk, 2004). The paradoxical challenge of maintaining a tolerant society while restricting intolerant speech underscores

the delicate balance that laws and policies must achieve, a theme that is central to our study and the broader discourse on hate speech legislation.

This study underscores the delicate equilibrium that must be navigated in crafting and implementing hate speech legislation, highlighting the critical need for a nuanced approach that respects freedom of expression while protecting individuals and groups from the harms of hate speech. The insights garnered from our participants illuminate the complex dynamics at play and underscore the importance of inclusive, comprehensive strategies in addressing the challenges posed by hate speech in diverse societies.

Despite its contributions, this study is not without limitations. The reliance on semi-structured interviews, while valuable for gaining in-depth insights, limits the generalizability of the findings. The sample size, though purposively selected to encompass a range of perspectives, may not fully capture the breadth of experiences and views that exist on this complex issue. Furthermore, the evolving nature of digital communication and legislative frameworks may necessitate ongoing research to stay abreast of new developments and their implications for hate speech regulation.

Future research should aim to broaden the scope of inquiry by incorporating quantitative methods to complement the qualitative insights of this study, potentially enhancing the generalizability of the findings. Longitudinal studies could provide valuable information on the long-term effects of hate speech legislation and its impact on freedom of expression. Additionally, cross-cultural comparisons could offer a more nuanced understanding of how different societal values and legal systems navigate the balance between curbing hate speech and upholding free speech.

The findings of this study have important implications for policymakers, legal practitioners, and digital platform administrators. It is imperative to engage in continuous dialogue with a broad spectrum of stakeholders to refine and adapt hate speech legislation. Education and awareness campaigns, informed by the insights from this and similar studies, can play a crucial role in shaping public understanding and attitudes towards hate speech and freedom of expression. For digital platforms, the development of more sophisticated and contextually aware moderation tools, guided by

ethical considerations and respect for user rights, is crucial. In practice, a collaborative, multi-stakeholder approach that includes legal, educational, and technological strategies offers the most promising path forward in addressing the challenges posed by hate speech while safeguarding the fundamental right to free expression.

Authors' Contributions

Authors contributed equally to this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

Declaration of Interest

The authors report no conflict of interest.

Funding

Ethics Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants.

References

Bilewicz, M., Soral, W., Marchlewska, M., & Winiewski, M. (2015). When Authoritarians Confront Prejudice. Differential Effects of SDO and RWA on Support for Hate-Speech Prohibition. *Political Psychology*. https://doi.org/10.1111/pops.12313

Boeckmann, R. J., & Turpin-Petrosino, C. (2002). Understanding the Harm of Hate Crime. *Journal of Social Issues*. https://doi.org/10.1111/1540-4560.00257

Bonotti, M. (2017). Religion, Hate Speech and Non-Domination. *Ethnicities*. https://doi.org/10.1177/1468796817692626

Brown, A. (2017). What Is Hate Speech? Part 1: The Myth of Hate. *Law and Philosophy*. https://doi.org/10.1007/s10982-017-9297-1

Brown, G. L. (1982). The Paradox of Tolerance. *Higher Education Quarterly*. https://doi.org/10.1111/j.1468-2273.1982.tb01343.x

Cohen-Almagor, R. (2011). Fighting Hate and Bigotry on the Internet. *Policy & Internet*. https://doi.org/10.2202/1944-2866.1059

Cowan, G., & Khatchadourian, D. (2003). Empathy, Ways of Knowing, and Interdependence as Mediators of Gender Differences in Attitudes Toward Hate Speech and Freedom of Speech. *Psychology of Women Quarterly*. https://doi.org/10.1111/1471-6402.00110

Felberg, T. R., & Šarić, L. (2021). Chocolate, Identity, and Extreme Speech Online. *Fleks - Scandinavian Journal of Intercultural Theory and Practice*. https://doi.org/10.7577/fleks.4168

Floyd, R. (2017). Parallels With the Hate Speech Debate: The Pros and Cons of Criminalising Harmful Securitising Requests. *Review of International Studies*. https://doi.org/10.1017/s0260210517000328

Gitari, N. D., Zhang, Z., Hanyurwimfura, D., & Long, J. (2015). A Lexicon-Based Approach for Hate Speech Detection. *International Journal of Multimedia and Ubiquitous Engineering*. https://doi.org/10.14257/ijmue.2015.10.4.21

Howard, J. W. (2019). Free Speech and Hate Speech. *Annual Review of Political Science*. https://doi.org/10.1146/annurev-polisci-051517-012343

María del Carmen Moreno, P., Montero-Díaz, J., & Moreno-Delgado, A. (2020). Hate Speech: A Systematized Review. *Sage Open*. https://doi.org/10.1177/2158244020973022

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, Hate Speech, and Social Media: A Systematic Review and Critique. *Television & New Media*. https://doi.org/10.1177/1527476420982230

Pasamonk, B. (2004). The Paradoxes of Tolerance. *The Social Studies*. https://doi.org/10.3200/tsss.95.5.206-210

Sękowska-Kozłowska, K. (2022). Sexist Hate Speech and the International Human Rights Law: Towards Legal Recognition of the Phenomenon by the United Nations and the Council of Europe. *International Journal for the Semiotics of Law - Revue Internationale De Sémiotique Juridique*. https://doi.org/10.1007/s11196-022-09884-8

Zhang, Z., & Luo, L. (2019). Hate Speech Detection: A solved Problem? The Challenging Case of Long Tail on Twitter. *Semantic Web*. https://doi.org/10.3233/sw-180338