

From Hate Speech to Social Exclusion: Rethinking Legal Protection for Religious Minorities in Indonesia's Digital Public Sphere

Indah Sri Utari* 

Universitas Negeri Semarang, Semarang, Indonesia
indahsuji@mail.unnes.ac.id

Ali Masyhar 

Universitas Negeri Semarang, Semarang, Indonesia
alimasyhar@mail.unnes.ac.id

Ridwan Arifin 

Universitas Negeri Semarang, Semarang, Indonesia
Universitat de Barcelona, Barcelona, Spain
rarifiar119@alumnes.ub.edu

DOI: <https://doi.org/10.65815/pacv9b83>

Submitted: 21-09-2025

Revised: 27-12-2025, 27-02-2026

Accepted: 28-03-2026

*Corresponding Author

Abstract

The expansion of digital communication has transformed the production and dissemination of hostility toward religious minorities in Indonesia. Online hate speech can extend beyond verbal harm by reinforcing social stigma, legitimizing discrimination, and contributing to offline intimidation and exclusion. This article examines the relationship between digital hate speech and the broader social marginalization of religious minorities. Using a normative and socio-legal approach, the study analyzes Indonesian legislation governing electronic communication, freedom of expression, religious freedom, and minority protection, together with selected cases and patterns of online discourse. The research applies a rights-based framework to assess the tension between freedom of expression and the state's responsibility to protect minorities from harmful forms of digital hostility. The article argues that existing legal responses frequently concentrate on individual statements and criminal liability while overlooking the structural and cumulative effects of digitally amplified hate. Algorithmic visibility, networked harassment, and repeated stereotyping may create environments in which minority groups are effectively excluded from equal participation in public discourse. The study proposes a broader regulatory framework combining legal accountability, platform responsibility, digital literacy, and effective remedies for targeted communities. It concludes that protecting religious minorities in the digital sphere requires more than criminalizing isolated expressions; it requires addressing the mechanisms through which digital hostility becomes social exclusion and undermines substantive equality.

[Perluasan komunikasi digital telah mengubah produksi dan penyebaran permusuhan terhadap minoritas agama di Indonesia. Ujaran kebencian daring dapat meluas melampaui bahaya verbal dengan memperkuat stigma sosial, melegitimasi diskriminasi, dan berkontribusi pada intimidasi dan pengucilan di dunia nyata. Artikel ini mengkaji hubungan antara ujaran kebencian digital dan marginalisasi sosial yang lebih luas terhadap minoritas agama. Dengan menggunakan pendekatan normatif dan sosio-hukum, penelitian ini menganalisis undang-undang Indonesia yang mengatur komunikasi elektronik, kebebasan berekspresi, kebebasan beragama, dan perlindungan minoritas, bersama dengan kasus-kasus terpilih dan pola wacana daring. Penelitian ini menerapkan kerangka kerja berbasis hak untuk menilai ketegangan antara kebebasan berekspresi dan tanggung jawab negara untuk melindungi minoritas dari bentuk-bentuk permusuhan digital yang berbahaya. Artikel ini berpendapat bahwa respons hukum yang ada seringkali berfokus pada pernyataan individu dan tanggung jawab pidana sambil mengabaikan efek struktural dan kumulatif dari kebencian yang diperkuat secara digital. Visibilitas algoritmik, pelecehan melalui jaringan, dan stereotip berulang dapat menciptakan lingkungan di mana kelompok minoritas secara efektif dikecualikan dari partisipasi yang setara dalam wacana publik. Studi ini mengusulkan kerangka peraturan yang lebih luas yang menggabungkan akuntabilitas hukum, tanggung jawab platform, literasi digital, dan solusi efektif bagi komunitas sasaran. Kesimpulannya, melindungi minoritas agama di ranah digital membutuhkan lebih dari sekadar mengkriminalisasi ekspresi yang terisolasi; hal itu membutuhkan penanganan terhadap mekanisme yang menyebabkan permusuhan digital menjadi pengucilan sosial dan merusak kesetaraan substantif.]

Keywords: hate speech, religious minorities, digital rights, social exclusion, freedom of expression

Introduction

The digital transformation of public communication has altered the conditions under which freedom of expression and religious freedom are exercised in Indonesia. Social-media platforms have become important arenas for political debate, religious discussion, identity formation, and civic participation, but they have also created new opportunities for hostility toward religious minorities to become highly visible and rapidly disseminated. Earlier research on Indonesian social media identified significant relationships between religious discourse, political polarization, and hate speech, particularly when religious identity becomes intertwined with political contestation (Sazali et al. 2022). More recent research has similarly shown that online religious discourse may reproduce intolerance through repeated linguistic and ideological patterns on platforms such

as TikTok and Instagram (Hanipah, Djou, and Muslimin 2025). The central legal question is therefore no longer simply whether a particular statement is offensive. It is whether the cumulative digital environment created by hostile expression can undermine the ability of minority communities to participate equally in society.

The Indonesian Constitution provides a strong normative foundation for addressing this problem while simultaneously protecting freedom of expression. Article 28E recognizes freedom of religion, belief, association, assembly, and expression, whereas Article 28F protects the right to communicate and obtain information through available channels. At the same time, Article 28J recognizes that the exercise of rights may be restricted by law to respect the rights of others, morality, religious values, security, and public order in a democratic society. These provisions create a constitutional balancing problem rather than an automatic hierarchy between expression and minority protection. The state cannot treat every offensive statement as a legitimate target of criminalization, but it also cannot remain institutionally passive when expression becomes a vehicle for discrimination, hostility, or violence. The constitutional task is therefore to distinguish protected expression from harmful incitement while ensuring that restrictions satisfy legality, necessity, and proportionality.

Indonesia's statutory framework has attempted to respond to online hostility principally through the Electronic Information and Transactions Law (ITE Law). Article 28(2), together with its criminal sanction, has been used to address electronic information containing expressions intended to incite hatred or hostility based on ethnicity, religion, race, and intergroup relations. The provision has nevertheless attracted sustained criticism because of its broad interpretation and the absence of a consistently applied objective threshold. Analysis of twenty-seven Indonesian judicial decisions demonstrates substantial inconsistency in how courts have assessed hate speech and its social harm (Sahuri 2023). Earlier scholarship likewise identified Article 28(2) as vulnerable to overbreadth and potential interference with freedom of expression (Saleh 2021; Rahmazani 2022). The subsequent amendment of the ITE Law through Law No. 1 of 2024 provides an important opportunity to reconsider the legal framework, but legislative refinement alone cannot solve the broader structural problem of digitally mediated exclusion.

Previous studies can broadly be divided into three groups. The first examines the doctrinal and constitutional problems of Indonesia's digital speech regulations, emphasizing vagueness, overcriminalization, and the danger of suppressing legitimate criticism (Saleh 2021; Rahmazani 2022; Daimah and Krisch 2025). The second examines the production and detection of hate speech through computational or communication approaches. Indonesian-language research demonstrates the scale and complexity of online hate speech while identifying persistent challenges in determining targets, categories, severity, and dissemination patterns (Izzanardi Wijanarko et al. 2024; Kuncoro et al. 2023). The third group examines religion and social media, demonstrating that digital platforms may amplify intolerant narratives but can also support religious tolerance, counter-

narratives, and interfaith solidarity (Rohmawati and Hamidah 2024; Ridho, Damairi, and Matswah 2025). These studies provide valuable foundations, but they rarely integrate legal analysis with the question of how repeated digital hostility can transform into substantive exclusion.

This article addresses that gap by conceptualizing hate speech as a potential process of exclusion rather than merely an unlawful linguistic event. The central argument is that digital hostility can operate through accumulation. A single statement may cause limited harm, but repeated stereotyping, coordinated harassment, exclusionary labeling, threats, and algorithmically amplified narratives may progressively alter the social conditions under which minority groups speak, organize, worship, and participate in public debate. This argument draws upon Lim's concept of algorithmic enclaves, which explains how interactions between users, political actors, and platform architectures can intensify polarized identity formation (Lim 2017). It also engages with social exclusion scholarship, particularly the idea that exclusion concerns restricted participation in social, political, and institutional life rather than merely direct interpersonal discrimination (Silver 1994; Sen 2000).

The novelty of this article lies in linking the legal regulation of hate speech with the substantive-equality consequences of digital exclusion. Rather than asking exclusively whether criminal sanctions are sufficiently severe, the article asks whether the legal system recognizes the cumulative harm experienced by targeted communities and whether existing remedies are capable of restoring their equal participation. This perspective also challenges the assumption that hate speech is primarily a problem of individual perpetrators. Digital hostility is frequently produced through interactions among ordinary users, influential accounts, coordinated networks, recommendation systems, moderation policies, and political incentives. Indonesian computational research has already demonstrated that hate-speech detection must account for targets, categories, levels of hostility, and dissemination actors rather than simply identifying offensive words (Kuncoro et al. 2023). Legal analysis should similarly move beyond isolated textual content toward the social architecture through which hostility becomes consequential.

The research employs a normative and socio-legal methodology. The normative component analyzes the 1945 Constitution, Law No. 39 of 1999 on Human Rights, Law No. 12 of 2005 ratifying the International Covenant on Civil and Political Rights, Law No. 1 of 2024 amending the ITE Law, relevant provisions of the Criminal Code framework, and government regulations concerning electronic-system governance. The socio-legal component examines judicial decisions, scholarly studies, documented patterns of religious hostility, and research on digital communication and minority experiences. The analysis applies four principal frameworks: freedom of expression, equality and non-discrimination, proportionality, and substantive participation. Internationally, the Rabat Plan of Action provides a particularly important threshold for distinguishing legitimate

expression from prohibited incitement by considering context, speaker, intent, content, extent, and likelihood of harm (United Nations 2012).

Results and Discussion

1. From Offensive Speech to Cumulative Social Harm

The first analytical finding is that the legal concept of hate speech is often too narrow when it is confined to the linguistic content of a single communication. Social harm in digital environments may emerge through repetition rather than through one exceptionally violent statement. A minority group repeatedly described as dangerous, illegitimate, immoral, or incompatible with national identity may gradually become associated with negative social meanings. This process is particularly significant when hostile narratives circulate across multiple platforms and are repeatedly encountered by large audiences. Research on Indonesian religious discourse indicates that social media has become an important source through which young people encounter religious information, while political and religious content may simultaneously expose them to intolerant narratives (Setianto, Nurjuman, and Handaningtias 2023). The legal significance therefore lies not merely in individual offensiveness but in the cumulative production of an environment in which minority identities are systematically stigmatized.

The distinction between direct harm and structural harm is important. Direct harm may involve a threat, insult, or explicit call for violence against an identifiable person or community. Structural harm occurs when repeated narratives change the social environment in which the targeted group participates. If members of a minority anticipate harassment whenever they disclose their identity, they may withdraw from public discussion, avoid online religious expression, or refrain from defending their community. Such withdrawal represents a form of chilling effect that may be difficult to observe statistically but is highly relevant to substantive equality. The result is paradoxical: a platform may formally permit everyone to speak while practically enabling some groups to speak more safely than others. Equality of formal access therefore does not necessarily produce equality of participation.

This dynamic can be understood through the theory of social exclusion. Silver (1994) conceptualizes exclusion as a multidimensional process involving social isolation and unequal participation, while Sen (2000) emphasizes deprivation of capabilities and opportunities to participate in social life. Applied to digital communication, social exclusion may occur when members of religious minorities face persistent barriers to expressing identity, forming networks, accessing information, or engaging in public debate. Digital hostility becomes legally significant when it contributes to such deprivation. This does not mean every offensive statement creates a legally actionable exclusion. Rather, the law should recognize that repeated and coordinated hostility may generate consequences that cannot be understood through the isolated-message model.

The Indonesian context illustrates why this broader approach is necessary. Studies mapping hate speech on Indonesian social media have identified religious and political identities as strongly interconnected, particularly during periods of political contestation (Sazali et al. 2022). Lim's research on Indonesia similarly demonstrates that social-media architectures can contribute to the formation of algorithmic enclaves in which identity-based polarization is reinforced (Lim 2017). The implication is that digital hate speech may be embedded within wider political and social structures. A legal framework focused exclusively on the individual speaker risks overlooking the networked conditions that amplify hostility. Effective protection therefore requires attention to both expression and the systems through which expression acquires social power.

2. Religious Minorities and the Unequal Digital Public Sphere

The digital public sphere is often described as an open space in which citizens can communicate without the gatekeeping associated with traditional media. Yet openness does not necessarily produce equality. The capacity to participate meaningfully depends upon visibility, safety, network access, credibility, and the ability to withstand harassment. Religious minorities may possess the same formal right to create accounts and publish opinions as majority users, but their participation can become more costly when digital hostility repeatedly targets their identity. The concept of the public sphere must therefore be connected to substantive equality. A public space is not genuinely inclusive merely because everyone can enter it; it must also allow participants to remain present without facing systematically unequal burdens.

Recent Indonesian research supports this distinction. A nationally representative survey of 3,820 Indonesian Muslims found no statistically significant relationship between social-media exposure for religious and political information and higher levels of religious intolerance, suggesting that social media may amplify the visibility of intolerant discourse without necessarily changing underlying attitudes (Studia Islamika 2025). This finding is important because it cautions against technologically deterministic explanations. Social media should not automatically be treated as the cause of intolerance. Instead, digital platforms may function as amplification infrastructures through which pre-existing political, religious, and social divisions become more visible and consequential. Legal regulation should consequently focus on concrete harms and mechanisms rather than assuming that all digital religious disagreement is inherently dangerous.

The experience of minority communities also demonstrates that digital spaces can produce both exclusion and resistance. Research on religious minority narratives in Indonesia finds that social media can simultaneously amplify discriminatory discourse and facilitate minority advocacy (Alimuddin and Lubis 2025). Similarly, research on grassroots feminist and interfaith networks demonstrates how digital activism can construct counter-narratives against religious intolerance (Ridho, Damairi, and Matswah 2025). These findings complicate a

purely restrictive regulatory model. If state intervention becomes overly broad, it may suppress not only hate speech but also the counter-speech through which minority communities challenge discrimination. The objective of law should therefore be to preserve the communicative capacity of vulnerable communities while restricting expression that crosses the threshold into incitement or serious rights violations.

The concept of digital citizenship is particularly relevant here. Equal citizenship requires the capacity to participate in public communication, contest dominant narratives, and communicate one's identity without unreasonable exclusion. When religious minorities withdraw from digital spaces because of repeated harassment, citizenship becomes formally equal but substantively differentiated. This distinction mirrors broader constitutional debates about minority protection in Indonesia, where legal recognition does not always translate into equal institutional treatment. The digital environment adds a new dimension because exclusion can occur without a formal prohibition. A community may legally possess every right to speak but be practically discouraged from speaking because of threats, mass reporting, doxing, ridicule, or coordinated harassment. The legal response should therefore include protection against these indirect mechanisms.

3. The Limits of Criminalization Under the ITE Law

Indonesia's reliance on criminal law presents a significant tension. Criminal sanctions can be justified where expression constitutes genuine incitement to discrimination, hostility, or violence. However, criminalization is a coercive intervention that must satisfy legality, necessity, and proportionality. Article 20(2) of the ICCPR requires prohibition of advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence, while Article 19 protects freedom of expression. The Rabat Plan of Action provides a structured approach for reconciling these provisions through a contextual threshold. Courts should consider the social and political context, speaker's position, intent, content and form, extent of dissemination, and likelihood of resulting harm (United Nations 2012). This framework provides a more precise alternative to content-based criminalization.

Indonesian judicial practice has demonstrated the difficulty of applying such a threshold consistently. Analysis of twenty-seven decisions involving Article 28(2) of the ITE Law found that courts have interpreted hate speech broadly and inconsistently, particularly when assessing the nature and degree of harm (Sahuri 2023). The problem is not merely doctrinal uncertainty. Broad interpretation can create a chilling effect because citizens cannot predict which controversial statements may result in criminal liability. This uncertainty is especially problematic for religious minorities and activists because robust criticism of dominant institutions may be mistaken for hostility toward a protected religious group. A rights-compatible hate-speech framework must therefore distinguish between

criticism, satire, religious disagreement, offensive expression, discriminatory advocacy, and incitement.

The problem is reinforced by the broader history of the ITE Law. Scholarship has identified Articles 27(3) and 28(2), as well as government content-control powers, as potential sources of excessive restriction on expression (Setyawan et al. 2025; Rahmazani 2022). The amendment of the ITE Law through Law No. 1 of 2024 is therefore significant because it represents an attempt to clarify and reform the digital legal framework. Nevertheless, statutory amendment cannot by itself guarantee rights-compatible enforcement. Administrative and prosecutorial practices, police interpretation, judicial reasoning, and platform cooperation all influence the practical meaning of digital rights. The central reform challenge is consequently institutional rather than merely textual.

Criminal law should also be considered a remedy of last resort rather than the universal response to digital hostility. The United Nations Strategy and Plan of Action on Hate Speech emphasizes that addressing hate speech does not simply mean reducing speech; it requires tackling root causes, supporting counter-narratives, strengthening education, and involving governments, civil society, and private-sector actors (United Nations 2019). This approach is particularly relevant to Indonesia because social exclusion cannot be criminalized away. A prosecution may punish an offender without repairing the social relationships, information environment, and institutional vulnerabilities that allowed exclusion to develop. Legal policy should therefore combine targeted criminal liability with civil remedies, platform accountability, community protection, digital literacy, and accessible complaint mechanisms.

4. Algorithmic Amplification and Platform Responsibility

A central feature distinguishing digital hate speech from conventional speech is the role of platform architecture. Online content does not circulate solely according to individual choices. Recommendation systems, engagement-based ranking, sharing mechanisms, hashtags, and network structures can determine which messages become visible and which remain peripheral. Lim's analysis of Indonesian political communication introduced the concept of algorithmic enclaves to explain how users and algorithms mutually reinforce polarized identities (Lim 2017). More recent research on Indonesia's digital ecosystem has identified automated content production, personalization, echo chambers, and multimodal manipulation as mechanisms that can amplify disinformation and hate speech (Sonni 2025). These findings suggest that platform governance is legally relevant even when platforms are not the original authors of hostile content.

The legal significance of amplification lies in the distinction between hosting and actively increasing reach. A platform that merely provides neutral infrastructure raises different regulatory questions from a system that recommends content because it generates high engagement. When harmful narratives become highly visible because they stimulate outrage or identity conflict, the social harm may be

partly produced by platform design. This does not justify general censorship or state control over algorithms. It does, however, support a regulatory expectation that platforms conduct reasonable risk assessment, provide transparent moderation procedures, and establish effective mechanisms for reporting serious identity-based harassment. Platform responsibility should therefore be proportionate to the degree of control that platforms exercise over content visibility.

Indonesian research on hate-speech detection also reveals the technological complexity of this problem. Kuncoro et al. (2023) demonstrate that detection systems need to identify not only whether a message is hateful but also its target, category, severity, and the actors involved in dissemination. Similarly, Wijanarko et al. (2024) emphasize that hate-speech propagation is context-dependent and that robust classification requires attention to the cultural and social characteristics of targeted groups. These findings have legal implications because automated moderation systems cannot be treated as neutral substitutes for legal judgment. Algorithms may reproduce linguistic biases, misunderstand religious context, or incorrectly classify minority advocacy as hateful. Platform governance must therefore combine automated detection with human review and meaningful appeal mechanisms.

Platform responsibility should also include transparency toward targeted communities. Victims of coordinated harassment often do not know why their content is removed, why abusive accounts remain active, or how to appeal decisions. An effective rights-based model would require platforms to provide accessible complaint channels, explanations for significant moderation decisions, emergency pathways for credible threats, and mechanisms for challenging systematic harassment. Such procedures are especially important for minority communities because unequal moderation can itself reproduce exclusion. If platforms rapidly remove minority counter-speech while allowing hostile stereotypes to remain visible, formal neutrality becomes substantively discriminatory.

5. Proportionality, Remedies, and the Protection of Minority Speech

A rights-based framework must distinguish between protecting minorities and silencing controversial expression. The danger of overregulation is particularly acute in religious contexts because disagreement about theology, doctrine, proselytization, political ideology, and religious institutions is often intense. A democratic legal order cannot criminalize speech merely because it is blasphemous, offensive, unpopular, or critical of a religious majority or minority. Restrictions become more legitimate when expression crosses into advocacy of discrimination, hostility, or violence and when the relevant harm is sufficiently concrete. The Rabat threshold is valuable precisely because it requires context and proportionality rather than relying solely on offensive content (United Nations 2012).

The principle of proportionality should therefore structure the entire legal response. First, restrictions must have a clear legal basis. Second, the state must demonstrate a legitimate objective, such as protecting individuals from discrimination or preventing violence. Third, the restriction must be necessary to achieve that objective. Fourth, the measure must be proportionate to the severity and likelihood of harm. Criminal prosecution should be reserved for the most serious cases, particularly those involving intentional incitement and a realistic possibility of discriminatory or violent consequences. Less serious forms of hostility may be addressed through platform moderation, civil remedies, education, mediation, or counter-speech. This graduated approach reduces the risk that hate-speech regulation becomes a general mechanism for controlling political or religious disagreement.

Effective remedies must also be victim-centered. Current legal discussions often focus on identifying and prosecuting perpetrators, while the needs of targeted communities receive less attention. Victims may require rapid removal of threats, protection against coordinated harassment, correction of defamatory information, restoration of unlawfully removed content, psychological support, legal assistance, and opportunities to publish counter-narratives. The remedy should correspond to the harm. If the harm is networked and reputational, a purely individual criminal sentence may not restore equality. The law should therefore recognize remedies that address the informational environment as well as the offender.

Counter-speech represents an important complementary strategy. Indonesian civil-society organizations have demonstrated that digital platforms can be used to promote tolerance and challenge exclusionary narratives (Rohmawati and Hamidah 2024). Research on grassroots feminist and interfaith networks similarly shows that marginalized actors can use digital communication to contest intolerant discourse (Ridho, Damairi, and Matswah 2025). The state should support these efforts through digital-literacy programs, public-interest information, and educational initiatives rather than relying exclusively on removal and prosecution. The objective is not to replace harmful speech with state-approved speech but to expand the communicative resources available to citizens so that minority voices can remain visible.

6. Toward a New Legal Framework for Digital Minority Protection

The analysis suggests that Indonesia requires a shift from a statement-centered model toward a harm-and-participation model of digital minority protection. The statement-centered model asks whether particular words satisfy statutory elements. The harm-and-participation model asks whether expression, combined with its context and mode of dissemination, creates a serious risk of discrimination, hostility, violence, or systematic exclusion. This does not eliminate the importance of statutory elements. Instead, it requires courts and enforcement institutions to interpret them through constitutional equality and international human-rights

standards. Such an approach would make the legal system more capable of addressing cumulative harms while reducing arbitrary criminalization.

The proposed framework should contain four layers of state responsibility. The first is prevention, including digital literacy, public education, research, and support for counter-narratives. The second is platform governance, requiring proportionate moderation, transparency, complaint mechanisms, and safeguards against algorithmic amplification of severe identity-based hostility. The third is legal accountability, using criminal law for serious incitement and other clearly defined offenses while preserving legitimate expression. The fourth is remedy, ensuring that targeted minorities can obtain rapid protection, legal assistance, correction, restoration, and other appropriate forms of redress. These layers correspond with the UN Strategy on Hate Speech, which emphasizes coordinated action involving governments, civil society, private actors, and citizens rather than relying on criminal law alone (United Nations 2019).

The framework should also recognize the possibility of discriminatory enforcement. A hate-speech regime can itself become a source of minority exclusion if legal institutions selectively prosecute minority speakers while ignoring hostile speech by majority actors. This is why neutrality cannot be assessed solely through statutory wording. Enforcement patterns must also be examined. If the same expression receives different legal treatment depending on the identity of the speaker or target, formal equality becomes illusory. Independent oversight, transparent prosecutorial standards, judicial reasoning, and publicly available enforcement data are therefore essential. Digital-rights regulation should be evaluated not only by how much harmful content it removes but also by whether it distributes expressive burdens fairly.

Finally, legal protection must extend beyond the moment when hate speech becomes a criminal offense. Religious minorities need institutional mechanisms capable of responding to early warning signs of exclusion. Coordinated harassment, repeated stereotyping, threats against religious institutions, and systematic disinformation may indicate escalating hostility even when individual messages do not satisfy the criminal threshold. Civil-society monitoring, platform reporting, human-rights institutions, and community organizations can provide early intervention. The objective is preventative rather than punitive: to stop digital hostility from developing into physical intimidation, institutional discrimination, or withdrawal from public participation. Such an approach is more consistent with substantive equality because it treats minority protection as an ongoing institutional obligation.

Conclusion

This article has argued that online hate speech directed at religious minorities should be understood not only as an issue of offensive communication or individual criminal liability but as a potential mechanism of social exclusion. Digital hostility becomes particularly consequential when it is repeated, networked, amplified, and

connected to existing social divisions. Under these conditions, the harm may extend beyond emotional injury to include stigmatization, reduced participation, self-censorship, networked harassment, and unequal access to the digital public sphere. Indonesian scholarship demonstrates that religion, politics, and online hostility frequently intersect, while computational studies show that hate-speech dissemination is shaped by targets, severity, context, and network structures. The legal system should therefore recognize the cumulative and structural dimensions of digital harm without assuming that every controversial religious expression constitutes unlawful hate speech.

The article further concludes that Indonesia's legal response requires a more carefully calibrated relationship between freedom of expression and minority protection. The ITE Law remains important, particularly for serious incitement to discrimination, hostility, and violence, but criminalization cannot be the sole regulatory strategy. The Rabat Plan of Action provides a useful proportionality framework, while the UN Strategy on Hate Speech demonstrates the importance of prevention, education, counter-speech, and multi-stakeholder responsibility. A rights-based Indonesian framework should therefore combine precise criminal thresholds with platform accountability, transparent moderation, accessible remedies, digital literacy, and protection for minority counter-speech. Future research should empirically examine how online harassment affects the participation of specific religious minorities, how Indonesian platforms implement content-moderation policies, and whether algorithmic recommendation systems disproportionately amplify identity-based hostility. Such research would allow legal reform to move from assumptions about digital hate toward evidence-based protection of equal citizenship in the digital public sphere.

References

- Alimuddin, Alimuddin, and Muhammad Zulfikar Satria Lubis. 2025. "Religious Minority Rights and Social Media Narratives in Indonesia: A Case Study of Online Hate Speech." *Indonesian Minority Justice Review*.
- Daimah, Daimah, and Nico Krisch. 2025. "The Cyber Footprint of Digital Activism: The Legal Boundary between Freedom of Expression and Hate Speech in Indonesia." *Indonesian Cyber Law Review* 2 (1).
- Hanipah, Heni, Dakia N. Djou, and Muslimin. 2025. "When Intolerant Language Crowds the Feed: How Digital Discourses Reconstruct Religious Freedom in Indonesia." *Harmoni* 24 (2).
- Izzanardi Wijanarko, Musa, Lucky Susanto, Prasetya Anugrah Pratama, Ika Karlina Idris, Traci Hong, and Derry Tanti Wijaya. 2024. "Monitoring Hate Speech in Indonesia: An NLP-Based Classification of Social Media Texts." *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*.

- Kastolani. 2020. "Understanding the Delivery of Islamophobic Hate Speech via Social Media in Indonesia." *Indonesian Journal of Islam and Muslim Societies*.
- Kuncoro, A. P., et al. 2023. "Hate Speech and Abusive Language Detection in Indonesian Social Media: Progress and Challenges." *Heliyon* 9 (8): e18647.
- Lim, Merlyna. 2017. "Freedom to Hate: Social Media, Algorithmic Enclaves, and the Rise of Tribal Nationalism in Indonesia." *Critical Asian Studies* 49 (3): 411–427.
- Rahmazani. 2022. "Problematisasi Hukum Penerapan Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) di Indonesia." *Mimbar Hukum* 34 (1).
- Ridho, Akhsin, Moh. Uzaer Damairi, and Akrimi Matswah. 2025. "Grassroots Feminist Networks and Interfaith Solidarity against Religious Intolerance: A Digital Ethnography from Indonesia." *Harmoni* 24 (1).
- Riadi, Imam, Abdul Fadlil, and Murni. 2023. "Identifying Hate Speech in Tweets with Sentiment Analysis on Indonesian Twitter Utilizing Support Vector Machine Algorithm." *Khazanah Informatika* 9 (2).
- Rohmawati, Wiwin Siti Aminah, and Kamilia Hamidah. 2024. "Religious Tolerance Narratives on Social Media: Network Analysis of Indonesian Civil Society Organizations on Instagram (2019–2023)." *Islamic Review: Jurnal Riset dan Kajian Keislaman* 14 (1).
- Sahuri, S. 2023. "Hate Speech and the Harm in Indonesian Judicial Decisions." *Cogent Education* 10.
- Saleh, Gazalba. 2021. "The Constitutionality of the Electronic Information and Transaction Law: Towards Overcoming SARA Conflict on Social Media." *Jurnal Konstitusi*.
- Sazali, Hasan, Umar Abdur Rahim SM, Rustono Farady Marta, and Al Rynne Gabonada Gatcho. 2022. "Mapping Hate Speech Relationships Indonesia's Religion and State in Social Media." *Communicatus: Jurnal Ilmu Komunikasi* 6 (2).
- Sen, Amartya. 2000. *Social Exclusion: Concept, Application, and Scrutiny*. Manila: Asian Development Bank.
- Setianto, Yearry, Husnan Nurjuman, and Uliviana Handaningtias. 2023. "Remaja, Media Sosial dan Ujaran Kebencian: Studi Konsumsi Online Religious Content di Banten." *Interaksi: Jurnal Ilmu Komunikasi* 12 (1): 125–145.
- Setyawan, Adisti Puspa, Antonius Sambodo Adi A., I Dewa BS Wirantaya, Pina Mustika, and William Gomarga. 2025. "Indonesia's Electronic Information and Transactions Law: History, Impact, and Challenges." *Journal Research of Social Science, Economics, and Management* 4 (12).
- Silver, Hilary. 1994. "Social Exclusion and Social Solidarity: Three Paradigms." *International Labour Review* 133 (5–6): 531–578.
- Sonni, AF Alem Febri. 2025. "AI-Based Disinformation and Hate Speech Amplification: Analysis of Indonesia's Digital Media Ecosystem." *Frontiers in Communication*.

- Studia Islamika. 2025. "Social Media and Religious Intolerance in Indonesia." *Studia Islamika* 33 (1).
- United Nations. 2012. *Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred That Constitutes Incitement to Discrimination, Hostility or Violence*. Geneva: Office of the United Nations High Commissioner for Human Rights.
- United Nations. 2019. *United Nations Strategy and Plan of Action on Hate Speech*. New York: United Nations.
- United Nations Human Rights Committee. 2011. *General Comment No. 34: Article 19, Freedoms of Opinion and Expression*. Geneva: United Nations.
- United Nations. 1966. *International Covenant on Civil and Political Rights*. New York: United Nations.
- Republic of Indonesia. 1945. *The 1945 Constitution of the Republic of Indonesia*.
- Republic of Indonesia. 1999. *Law No. 39 of 1999 on Human Rights*.
- Republic of Indonesia. 2008. *Law No. 11 of 2008 on Electronic Information and Transactions*.
- Republic of Indonesia. 2016. *Law No. 19 of 2016 on Amendment to Law No. 11 of 2008 on Electronic Information and Transactions*.
- Republic of Indonesia. 2023. *Law No. 1 of 2023 on the Criminal Code*.
- Republic of Indonesia. 2024. *Law No. 1 of 2024 on the Second Amendment to Law No. 11 of 2008 on Electronic Information and Transactions*.
- United Nations Educational, Scientific and Cultural Organization. 2021. *The Rabat Plan of Action on the Prohibition of Incitement to Hatred*. Paris: UNESCO.
- Kastolani. 2020. "Understanding the Delivery of Islamophobic Hate Speech via Social Media in Indonesia." *Indonesian Journal of Islam and Muslim Societies*.
- Pamungkas, Martinus Danang, Zulfikri Suleman, and Anang Dwi Santoso. 2024. "On Digital Ethnography: Nonviolent Campaign Movement of the Instagram Community @Toleransi.id in Realizing Religious Moderation." *Al-Qalam* 30 (1).
- Surahmat, Sailal Arimi, Novi Eka Susilowati, and Fardan Mahmudatul Imamah. 2025. "From Invocation to Indictment: Analyzing Hate Speech Elements in a Social Media Prayer." *Rainbow: Journal of Literature, Linguistics and Culture Studies* 14 (1).

“The claim "hate speech is not free speech" implies "free" is a type of speech, as opposed to how speech is treated in a free society.”

Michael Malice
An American Author

Acknowledgment

None

Funding Information

None

Conflicting Interest Statement

The authors state that there is no conflict of interest in the publication of this article.

Publishing Ethical and Originality Statement

All authors declared that this work is original and has never been published in any form and in any media, nor is it under consideration for publication in any journal, and all sources cited in this work refer to the basic standards of scientific citation.

Generative AI Statement

N/A