# **International Journal of Social Science and Human Research**

ISSN (print): 2644-0679, ISSN (online): 2644-0695

Volume 07 Issue 03 March 2024

DOI: 10.47191/ijsshr/v7-i03-29, Impact factor- 7.876

Page No: 1724-1728

# Empathy and Hate Speech in Social Media: The Case of Indonesia

# Dr. Kus Hanna Rahmi

Psychologist, Faculty of Psychology, Universitas Bhayangkara Jakarta Raya, Indonesia



**ABSTRACT:** Hate speech on social media in Indonesia 10 years ago has become a very interesting topic of discussion. Many of these incidents occur and are perpetrated by teenagers, with implications for violence and their psychological well-being. Hate speech on social media, particularly in the Indonesian context, presents a multifaceted challenge that requires a combination of technological solutions, legal frameworks, and community interventions to effectively combat its spread and impact. By utilizing insights from research on hate speech detection, empathy promotion, and regulatory frameworks, stakeholders can work towards creating a safer and more inclusive online environment in Indonesia. Strengthening digital literacy, as suggested in the references, is essential to empower individuals to identify and respond effectively to hate speech. Legal measures, as emphasized in the references, play an important role in combating hate speech, with the establishment of specialized units such as the Indonesian Virtual Police contributing to monitoring and handling hate speech incidents. Community engagement and educational programs, as discussed in the references, are crucial in promoting dialogue, tolerance, and understanding to foster a culture of mutual respect and empathy.

KEYWORDS: hate speech, social media, mutual respect, community engagement

#### INTRODUCTION

Hate speech on social media is a significant concern in Indonesia, with various studies focusing on understanding, detecting, and addressing this issue. The prevalence of hate speech related to religion and the state, as seen in popular terms like Kadrun and Cebong, has been mapped through qualitative analysis (Sazali et al., 2022). The impact of social changes on the legal framework in Indonesia, particularly in criminalizing hate speech through electronic media, has been highlighted (Santosa, 2021). Machine learning methods, such as the hybrid classifier, have been employed to classify hate speech hashtags on Twitter, reflecting the increasing cybercrimes related to hate speech in Indonesia (Syaifullah & Sibaroni, 2022). Efforts to detect hate speech using advanced technologies like Bidirectional Long Short-Term Memory and Word2vec extraction methods have been discussed, emphasizing the heated discussions on hate speech in Indonesia, especially on social media platforms (Isnain et al., 2020). The role of the Indonesia Virtual Police in countering hate speech underscores the seriousness of this issue in the country (Prabandari et al., 2021). Hate speech is a common occurrence on various online platforms, including social media, with implications for youth violence and psychosocial well-being (Patton et al., 2014). Studies have delved into specific instances of hate speech, such as impoliteness strategies directed at President Joko Widodo, indicating the need for regulatory measures to address hate speech in online media (Widiantho, 2020). Gender and discourse perspectives have been considered in reviewing hate speech content on Indonesian social media, revealing disturbing trends in hate discourse and speech (Wahyuningsih, 2021). The use of dysphemism lexical items in hate speeches has been analyzed, emphasizing the importance of educating individuals, particularly students, on political correctness (Darmawan & Muhaimi, 2020). Various computational methods, including convolutional neural networks and LSTM models, have been proposed for hate speech classification in Indonesian language tweets, reflecting the technological advancements in addressing this issue (Taradhita & Putra, 2021; Arbaatun et al., 2022). The metaphorical use of language on social media, such as Instagram, has been studied to understand the potential and threats of hate speech in public spaces (Mursyldin et al., 2022). Digital literacy has been identified as a crucial factor in combating computational propaganda and hate speech in Indonesia (Salma, 2019). Empathy-based counterspeech is effective in reducing racist hate speech on social media platforms, highlighting the role of empathy in countering online hate (Hangartner et al., 2021). Challenges in hate speech detection on social media have been acknowledged, emphasizing the complexity of identifying and addressing hate speech online (Kovács et al., 2021). The offensive and insulting nature of hate speech has been linked to personality traits, suggesting a hybrid approach for hate speech detection in Arabic social media (Elzayady et al., 2023). In summary, hate speech on social media, particularly in the context of Indonesia,

presents a multifaceted challenge that requires a combination of technological solutions, legal frameworks, and societal interventions to effectively combat its spread and impact.

# FACTORS CAUSED HATE SPEECH IN SOCIAL MEDIA IN INDONESIA

Hate speech in social media is a significant concern in Indonesia, with various studies shedding light on different aspects of this issue. The prevalence of hate speech related to religion-state relations, such as the use of words like Kadrun and Cebong, has been mapped out (Sazali et al., 2022). This mapping provides insights into the specific language used in hate speech on social media platforms in Indonesia. Additionally, the effectiveness of criminalizing hate speech through electronic media in response to social changes in cyberspace has been discussed, highlighting the impact of information sharing via electronic media on Indonesian law (Santosa, 2021). Furthermore, the classification of hate speech hashtags on Twitter using hybrid classifier methods has been explored, emphasizing the presence of cybercrimes related to hate speech in Indonesia (Syaifullah & Sibaroni, 2022). The role of the Indonesia Virtual Police in countering hate speech on social media has been addressed, indicating the seriousness of hate speech as a problem among Indonesian social media users (Prabandari et al., 2021). Various methods, such as bidirectional long-short-term memory and word-vec extraction, have been proposed for hate speech detection, reflecting the ongoing discussions and efforts to combat hate speech in Indonesia (Isnain et al., 2020). Moreover, impoliteness strategies in hate speech directed at President Joko Widodo have been analyzed, showcasing the need for regulations and interventions to address the spread of hate speech on online platforms (Widiantho, 2020). Hate speech content on Indonesian social media, particularly concerning gender and discourse perspectives, has been reviewed, highlighting the presence of disturbing hate discourse on platforms like Facebook and Instagram (Wahyuningsih, 2021). Dysphemism lexical items of hate speeches have been studied to educate students on political correctness, emphasizing the importance of promoting respectful communication on social media (Darmawan & Muhaimi, 2020). In the context of empathy and prosocial behavior, research has shown a link between prosocial media use, empathy, and helping behavior (Prot et al., 2013). Empathy-based counter-speech has been identified as a strategy to reduce racist hate speech on social media platforms, indicating the potential effectiveness of empathy in mitigating hate speech (Hangartner et al., 2021). Studies have also explored the impact of social media use on empathy, with findings suggesting a complex relationship between social media engagement and empathy levels (James et al., 2017). Various approaches, such as deep learning models and natural language processing techniques, have been proposed to address hate speech detection, underscoring the importance of technological interventions in combating hate speech online (Cao et al., 2020; Arbaatun et al., 2022). Multiple text representations and synonym-based feature expansion have been used to create hate speech detection methods. This shows the variety of approaches used to find and stop hate speech online (Ghozali et al., 202?). In summary, hate speech on Indonesian social media is a multifaceted issue that requires a comprehensive approach involving technological, legal, and social interventions to promote respectful and empathetic online interactions. By leveraging insights from research on hate speech detection, empathy promotion, and regulatory frameworks, stakeholders can work towards creating a safer and more inclusive online environment in Indonesia. Impact of Hate Speech in Indonesia Contemporary hate speech in Indonesia has significant impacts on individuals, communities, and society at large, with potentially far-reaching consequences if not addressed. Exposure to hate speech can lead to harmful implications, such as replacing empathy with intergroup contempt as a dominant response to others (Bilewicz & Soral, 2020). The dynamic effects of derogatory language on intergroup relations and political radicalization highlight how hate speech can both motivate and result from derogatory language, exacerbating tensions within society (Bilewicz & Soral, 2020). Hate speech based on religion in Indonesia has a particularly influential impact on the younger generation, especially school-age individuals, potentially shaping their attitudes and behaviors (Hamdi & Musthofa, 2020). This form of hate speech can contribute to social inequality, non-respectful attitudes, and national decline, further perpetuating divisions within society (Irawan, 2018). The widespread and fast-growing nature of hate speech on social media platforms can escalate into open conflicts and national disintegration if not promptly addressed (Prabandari et al., 2021). Additionally, there may be a direct link between hate speech and actual hate crimes, increasing the risk of being the target of hate speech and abusive language (Alshalan et al., 2020). The impact of online hate speech is not limited to virtual spaces; it can spill over into real-world actions, leading to conflicts, riots, and other harmful consequences (Sidiq et al., 2019). Additionally, hate speech can elicit strong emotional reactions and have a depressing impact on social self-esteem, which has an impact on people's psychological health, especially those who are the target of it (Boeckmann & Liew, 2002). The rise of hate speech in Indonesia, facilitated by the development of internet-based social media, has negative implications for online interactions and societal harmony (Ghozali et al., 2023). The experience of hostility, hate speech, and adversarial behavior in online spaces has increased substantially, with documented health and mental health outcomes for users (Cover, 2022). Hate speech not only offends and assaults the dignity of individuals but also undermines cultural diversity and democratic processes, posing a threat to social cohesion (Tefăniță & Buf, 2021). In summary, hate speech in Indonesia has multifaceted impacts on individuals, communities, and society, ranging from psychological distress and intergroup tensions to societal divisions and conflicts. Addressing hate speech requires a comprehensive approach that considers this issue's psychological, social, and legal dimensions to foster a more inclusive and respectful online and offline environment in Indonesia.

# PREVENTION AND TREATMENT OF HATE SPEECH IN INDONESIA

A multifaceted approach involving technological, legal, educational, and social interventions is essential to prevent and address hate speech in Indonesia. Leveraging research and strategies from various disciplines can contribute to mitigating the spread of hate speech and fostering a more inclusive online environment. Here are some key strategies based on the potential references provided: Utilizing Advanced Technology: Research by Waseem & Hovy (2016), Davidson et al. (2017), Luu et al. (2022), Arbaatun et al. (2022), and Paul et al. (2022) highlight the importance of leveraging advanced technologies such as natural language processing, deep learning models, and language detection algorithms to detect and combat hate speech on social media platforms effectively. Implementing these technologies can enhance the efficiency of hate speech detection and enable prompt responses to problematic content. Enhancing Digital Literacy: Research by Salma (2019) emphasizes the significance of improving digital literacy among the population to combat hate speech in Indonesia. Educating individuals on identifying and responding to hate speech can empower them to navigate online spaces responsibly and contribute to creating a more respectful online discourse. Legal Framework and Enforcement: Research by Royani (2021), Prabandari et al. (2021), and Silambi et al. (2018) underscores the importance of legal measures in addressing hate speech. Strengthening regulations, such as criminal and social law approaches, and establishing specialized units like the Indonesia Virtual Police can aid in monitoring, reporting, and taking action against hate speech perpetrators. Community Engagement and Education: Research by Anis et al. (2018), Jääskeläinen (2019), and Sya'bani et al. (2021) highlights the role of community engagement, arts education, and philosophical discussions in countering hate speech. Promoting dialogue, tolerance, and understanding through community initiatives and educational programs can foster a culture of respect and empathy, mitigating the prevalence of hate speech. Monitoring and Response Mechanisms: Research by Brown (2017) and Yuliyanti et al. (2020) suggests the importance of implementing monitoring systems to track hate speech content and responses. Developing efficient reporting mechanisms and swift responses to hate speech incidents can help prevent the escalation of harmful content and protect individuals from its negative impacts. Empowering Youth and Media Literacy: Research by Putri et al. (2022; Febryanti et al. (2022); and Sya'bani et al. (2021) emphasize the role of youth empowerment and media literacy in addressing hate speech. Educating young individuals on responsible online behavior, critical thinking, and media analysis can equip them with the skills to discern and counter hate speech effectively. By integrating these strategies and drawing insights from research across various domains, Indonesia can work towards creating a safer, more inclusive online environment and combating the detrimental effects of hate speech on individuals and society.

#### CONCLUSION

In conclusion, addressing hate speech in Indonesia requires a comprehensive and multi-faceted approach that integrates technological advancements, legal frameworks, educational initiatives, and community engagement strategies. Leveraging advanced technologies such as natural language processing and deep learning models, as highlighted in references, can enhance the detection and mitigation of hate speech on social media platforms. Strengthening digital literacy, as suggested in reference, is crucial to empowering individuals to identify and respond to hate speech effectively. Legal measures, as emphasized in references, play a significant role in combating hate speech, with the establishment of specialized units like the Indonesia Virtual Police contributing to monitoring and addressing hate speech incidents. Community engagement and educational programs, as discussed in the references, are essential in promoting dialogue, tolerance, and understanding to foster a culture of respect and empathy. Monitoring systems, as highlighted in the references, are vital for tracking hate speech content and implementing timely responses. Empowering these strategies and drawing insights from research across various disciplines, Indonesia can work towards creating a safer and more inclusive online environment, combating the detrimental effects of hate speech on individuals and society.

#### REFERENCES

- Calvin Erico Rudy Salim, Derwin Suhartono.("Long Short-Term Memory for Hate Speech and Abusive Language Detection on Indonesian Youtube Comment Section (2021). Long short-term memory for hate speech and abusive language detection on Indonesian YouTube comment section.. https://doi.org/10.18178/wcse.2021.06.029
- Alshalan, R., Al-Khalifa, H., AlSaeed, D., Al-Baity, H., & Alshalan, S. (2020). Detection of hate speech in COVID–19– related tweets in the Arab region: deep learning and topic modeling approach. Journal of Medical Internet Research, 22(12), e22609. https://doi.org/10.2196/22609
- Anis, M., Sari, L., & Yuliarti, M. (2018). Reception messages of hate speech among teenagers in the global world: an attempt to gain the good quality of labor demands. The Social Sciences, 3(11), 1318 https://doi.org/10.18502/kss.v3i11.2850
- Arbaatun, C., Nurjanah, D., & Nurrahmi, H. (2022). Hate speech detection on Twitter through natural language processing using the LSTM model. Building of Informatics Technology and Science (Bits), 4(3). https://doi.org/10.47065/bits.v4i3.2718

- 5) Bilewicz, M. and Soral, W. (2020). Hate speech epidemic. the dynamic effects of derogatory language on intergroup relations and political radicalization. Political Psychology, 41(S1), 3-33. https://doi.org/10.1111/pops.12670
- Boeckmann, R. and Liew, J. (2002). Hate speech: Asian American students' justice judgments and psychological responses. Journal of Social Issues, 58(2), 363-381. https://doi.org/10.1111/1540-4560.00265
- 7) Brown, A. (2017). What is so special about online (as compared to offline) hate speech? Ethnicities, 18(3), 297-326. https://doi.org/10.1177/1468796817709846
- 8) Cao, R., Lee, R., & Hoang, T. (2020). Deephate: hate speech detection via multi-faceted text representations. https://doi.org/10.1145/3394231.3397890
- 9) Cover, R. (2022). Digital hostility, subjectivity, and ethics: theorizing the disruption of identity in instances of mass online abuse and hate speech. Convergence the International Journal of Research into New Media Technologies, 29(2), 308-321. https://doi.org/10.1177/13548565221122908
- 10) Darmawan, I. and Muhaimi, L. (2020). Dysphemism lexical items of hate speeches: towards the education of students for political correctness. https://doi.org/10.2991/assehr.k.200827.061
- 11) Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. Proceedings of the International AAAI Conference on Web and Social Media, 11(1), 512-515. https://doi.org/10.1609/icwsm.v11i1.14955
- 12) Elzayady, H., Mohamed, M., Badran, K., & Salama, G. (2023). A hybrid approach based on personality traits for hate speech detection in Arabic social media. International Journal of Electrical and Computer Engineering (IJECE), 13(2), 1979. https://doi.org/10.11591/ijece.v13i2.pp1979-1988
- 13) Febryanti, D., Mahmud, Z., Putri, S., Ovalia, F., & Sekarwangi, Y. (2022). Tipologi hate speech di twitter terkait kebijakan pemerintah selama pandemi covid-19. Jurnal Komunikasi Global, 11(2), 274-299. https://doi.org/10.24815/jkg.v11i2.26733
- 14) Fortuna, P. and Nunes, S. (2018). A survey on automatic detection of hate speech in text. Acm Computing Surveys, 51(4), 1-30. https://doi.org/10.1145/3232676
- 15) Ghozali, I., Sungkono, K., Sarno, R., & Abdullah, R. (2023). Synonym-based feature expansion for Indonesian hate speech detection. International Journal of Electrical and Computer Engineering (IJECE), 13(1), 1105. https://doi.org/10.11591/ijece.v13i1.pp1105-1112
- 16) Hamdi, S. and Musthofa, K. (2020). The concept of lita'arofu as a multicultural education model to respond to religious hate speech. Attarbiyah Journal of Islamic Culture and Education, 5(2), 91-104. https://doi.org/10.18326/attarbiyah.v5i2.91-104
- 17) Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., ... & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. Proceedings of the National Academy of Sciences, 118(50). https://doi.org/10.1073/pnas.2116310118
- Irawan, I. (2018). Hate speech di Indonesia. Mawa Izh Jurnal Dakwah Dan Pengembangan Sosial Kemanusiaan, 9(1), 1-17. https://doi.org/10.32923/maw.v9i1.712
- 19) Isnain, A., Sihabuddin, A., & Suyanto, Y. (2020). Bidirectional long short-term memory method and word2vec extraction approach for hate speech detection. Ijccs (Indonesian Journal of Computing and Cybernetics Systems), 14(2), 169. https://doi.org/10.22146/ijccs.51743
- 20) Jääskeläinen, T. (2019). Countering hate speech through arts and arts education. Policy Futures in Education, 18(3), 344-357. https://doi.org/10.1177/1478210319848953
- 21) James, C., Davis, K., Charmaraman, L., Konrath, S., Slovák, P., Weinstein, E., ... & Yarosh, L. (2017). Digital life and youth well-being, social connectedness, empathy, and narcissism. Pediatrics, 140(Supplement\_2), S71-S75. https://doi.org/10.1542/peds.2016-1758f
- 22) Kovács, G., Alonso, P., & Saini, R. (2021). Challenges of hate speech detection in social media. Sn Computer Science, 2(2). https://doi.org/10.1007/s42979-021-00457-3
- 23) Lim, M. (2017). Freedom to hate: social media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia. Critical Asian Studies, 49(3), 411-427. https://doi.org/10.1080/14672715.2017.1341188
- 24) Luu, S., Nguyen, K., & Nguyen, N. (2022). Impacts of transformer-based language models and imbalanced data for hate speech detection on Vietnamese social media texts.. https://doi.org/10.21203/rs.3.rs-2242843/v1
- 25) Mursyidin, M., Nuthihar, R., Hasan, R., Herman, R., Rizki, A., & Wahdaniah, W. (2022). The metaphor of Aceh language on Instagram social media: potential and threats of hate speech in the public space. Retorika Jurnal Bahasa Sastra Dan Pengajarannya, 15(1). https://doi.org/10.26858/retorika.v15i1.23694
- 26) Patton, D., Hong, J., Ranney, M., Patel, S., Kelley, C., Eschmann, R., ... & Washington, T. (2014). Social media as a vector for youth violence: a review of the literature. Computers in Human Behavior, 35, 548-553. https://doi.org/10.1016/j.chb.2014.02.043

- 27) Paul, R., Seelawi, H., Nozza, D., Zeerak, T., & Vidgen, B. (2022). Multilingual hate check: functional tests for multilingual hate speech detection models.. https://doi.org/10.18653/v1/2022.woah-1.15
- 28) Prabandari, A., Cahyaningtyas, I., & Wibawa, K. (2021). The role of Indonesia virtual police in countering hate speech on social media.. https://doi.org/10.4108/eai.29-6-2021.2312584
- 29) Prot, S., Gentile, D., Anderson, C., Suzuki, K., Swing, E., Lim, K., ... & Lam, B. (2013). Long-term relations among prosocial-media use, empathy, and prosocial behavior. Psychological Science, 25(2), 358-368. https://doi.org/10.1177/0956797613503854
- 30) Putri, L., Annisarizki, A., & Andika, D. (2022). The journalists have built self-meaning: spreading hate speech in the presidential election 2019 in Banten. Aspiration Journal, 1(2), 175-199. https://doi.org/10.56353/aspiration.v1i2.13
- 31) Royani, Y. (2021). Hate speech during the caliphate of 'alī ibn abī tālib and its relevance to regulations in Indonesia: the study of the Islamic law history. Al-Ahkam, 31(2), 223-240. https://doi.org/10.21580/ahkam.2021.31.2.8956
- 32) Salma, A. (2019). Defining digital literacy in the age of computational propaganda and hate spin politics. Kne Social Sciences. https://doi.org/10.18502/kss.v3i20.4945
- 33) Santosa, D. (2021). The effectiveness of criminalizing hate speech through electronic media in dealing with social changes of communicating in cyberspace. Pancasila and Law Review, 2(2), 79-90. https://doi.org/10.25041/plr.v2i2.2354
- 34) Sazali, H., SM, U., Marta, R., & Gatcho, A. (2022). Mapping hate speech on social media: religion-state relations in Indonesia. Communicatus Jurnal Ilmu Komunikasi, 6(2), 189-208. https://doi.org/10.15575/cjik.v6i2.20431
- 35) Sidiq, H., Kusumo, D., & Sardi, I. (2019). Mendeteksi cyberhate pada twitter menggunakan text classification dan crowdsourced labeling. Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (Jnteti), 8(4), 315. https://doi.org/10.22146/jnteti.v8i4.530
- 36) Silambi, E., Azis, Y., Alputila, M., & Septarini, D. (2018). Legal testing on hate speech through social media.. https://doi.org/10.2991/icss-18.2018.297
- 37) Ștefăniță, O. and Buf, D. (2021). Hate speech in social media and its effects on the lgbt community: a review of the current research. Romanian Journal of Communication and Public Relations, 23(1), 47-55. https://doi.org/10.21018/rjcpr.2021.1.322
- 38) Sya'bani, T., Fitriawati, C., Ratnasari, D., & Habibah, S. (2021). Responses to the practice of hate speech at the Pancasila youth organization in Sidoarjo. Jed (Jurnal Etika Demokrasi), 6(2), 292-298. https://doi.org/10.26618/jed.v6i2.4515
- 39) Syaifullah, A. and Sibaroni, Y. (2022). Hate speech hashtag classification on Twitter using the hybrid classifier method. Jurikom (Jurnal Riset Komputer), 9(4), 828. https://doi.org/10.30865/jurikom.v9i4.4548
- 40) Taradhita, D. and Putra, I. (2021). Hate speech classification in Indonesian language tweets by using a convolutional neural network. Journal of ICT Research and Applications, 14(3), 225-239. https://doi.org/10.5614/itbj.ict.res.appl.2021.14.3.2
- 41) Wahyuningsih, S. (2021). Reviewing hate speech in Indonesian social media content: gender and discourse perspectives. Elt-Lectura, 8(1), 48-55. https://doi.org/10.31849/elt-lectura.v8i1.5573
- 42) Waseem, Z. and Hovy, D. (2016). Hateful symbols or hateful people? predictive features for hate speech detection on Twitter.. https://doi.org/10.18653/v1/n16-2013
- 43) Widiantho, Y. (2020). Hate speech addressed to President Joko Widodo in online media: impoliteness strategies analysis. Lingpoet Journal of Linguistics and Literary Research, 1(1), 14-19. https://doi.org/10.32734/lingpoet.v1i1.4692
- 44) Yuliyanti, T., Subyantoro, S., & Pristiwati, R. (2020). Form of hate speech comments on najwa Shihab YouTube channels in the general election campaign of the president and vice president of the Republic of Indonesia 2019. Seloka Jurnal Pendidikan Bahasa Dan Sastra Indonesia, 9(3), 254-263. https://doi.org/10.15294/seloka.v9i3.42513



There is an Open Access article, distributed under the term of the Creative Commons Attribution – Non Commercial 4.0 International (CC BY-NC 4.0)

(https://creativecommons.org/licenses/by-nc/4.0/), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.