The Easta Journal Law and Human Rights
Vol. 3, No. 03, June 2025, pp. 176 - 183
ISSN: 2985-7112, DOI: 10.58812/eslhr.v3i03

The Impact of Content Moderation Policy on the Spread of Fake News on Social Media in Indonesia

Evy Febryani
Universitas Muhammadiyah Palembang

Article Info

Article history:

Received Jun, 2025
Revised Jun, 2025
Accepted Jun, 2025

Keywords:

Content Moderation
Hoax News
Indonesia
Social Media

ABSTRACT

This research investigates the impact of content moderation policies on the spread of hoax news on social media platforms in Indonesia, employing a normative legal analysis approach. The study examines the effectiveness of existing legal frameworks, such as the Electronic Information and Transactions Law (ITE Law), and evaluates content moderation mechanisms employed by both the government and social media platforms. Findings indicate that while the ITE Law provides a legal basis for addressing misinformation, its vague provisions and potential for misuse have raised concerns regarding freedom of speech. Additionally, content moderation efforts by the government and social media platforms face challenges due to technological limitations, inconsistent enforcement, and a lack of public awareness. The study also highlights the need for a more integrated, multi-stakeholder approach to combat hoax news effectively. The research concludes that while current content moderation policies have had some success, their full impact is constrained by legal, technological, and social challenges. To improve the effectiveness of these policies, the study recommends clearer legal frameworks, enhanced technological capabilities, increased media literacy, and stronger coordination among stakeholders.

Corresponding Author:

Name: Evy Febryani
Institution: Universitas Muhammadiyah Palembang
Email: efebryani_1202@yahoo.com

1. INTRODUCTION

The rapid growth of social media in Indonesia has significantly reshaped the way information is shared and consumed, with platforms like Facebook, Twitter, Instagram, and WhatsApp becoming essential tools for communication, news dissemination, and social interaction. However, this increasing reliance on social media has also facilitated the proliferation of hoax news—deliberately fabricated or misleading information intended to deceive the public—which can incite panic, spread hatred, and influence public opinion, particularly during sensitive periods such as political elections, health crises, or social unrest [1]–[3]. Identifying and controlling hoaxes poses serious challenges due to the vast volume of information circulating online, the ease of redistribution through encrypted platforms like WhatsApp, and the involvement of foreign IPs and domains that obscure the origin of the content [1]. The societal impact of hoaxes is profound, as they exacerbate social tensions, foster prejudice, and strain relationships within families and communities [4]. In response, several strategies have been implemented, including legal frameworks such as the Electronic Information and Transaction Law [1], and educational initiatives aimed at improving digital literacy, especially among students. These training programs have demonstrated substantial effectiveness in enhancing individuals' abilities to identify misinformation and manage social media responsibly [5].

In response to the growing concern over misinformation, both social media platforms and the Indonesian government have implemented various content moderation policies aimed at curbing the spread of hoaxes, particularly in sensitive contexts such as the 2024 Indonesian election. These measures typically include the removal of false information, suspension of user accounts, and the integration of fact-checking mechanisms. While these policies are intended to safeguard public trust and electoral integrity, their effectiveness remains contested, especially given the need to balance public protection with concerns over freedom of expression, privacy, and potential overreach by the state. Fact-checking institutions in Indonesia play a vital role in identifying disinformation, yet they often struggle to keep up with the rapid volume and spread of false content, highlighting the need for a more resource-intensive and coordinated response [6]. Government regulations, such as the Ministry of Communications and Informatics (MOCI) Regulation No. 5/2020, empower authorities to request content removal but have drawn criticism for the risk of excessive censorship [7]. Moreover, social media platforms face challenges due to inconsistent guidelines and enforcement practices, revealing gaps in moderation that necessitate tailored approaches for both mainstream and fringe platforms [8]. The persistence of hoaxes and hate speech is often driven by political and economic agendas, posing serious risks to national security, thereby requiring collaborative efforts among the government, private sector, and civil society to ensure a safer digital environment [9].

This paper explores the impact of content moderation policies on the spread of hoax news on social media in Indonesia. By employing a normative legal analysis, this study aims to evaluate the legal frameworks and policies that govern content moderation on Indonesian social media platforms, assessing their effectiveness in preventing the spread of false information. It will examine existing laws and regulations such as the Electronic Information and Transactions Law (ITE Law), as well as the role of social media companies in enforcing these policies. Additionally, the paper will investigate the challenges and limitations faced by both regulators and platforms in effectively managing online content without infringing on fundamental rights such as freedom of speech and privacy.

2. LITERATURE REVIEW

2.1 *Hoax News and Its Impact on Society*

Hoax news, or fake news, is a major challenge in the digital era, marked by the intentional manipulation of information to mislead readers, with serious consequences for Indonesian society—including false reports about natural disasters, political events, and public health issues like COVID-19. Its rapid spread is driven by the widespread use of social media and mobile technology, especially among youth, which can influence elections, disrupt public health efforts, and spark social unrest. Key causes include technological advancements that accelerate information sharing [10], political and economic motives aiming for clicks and profit [10], [11], and low media literacy among the public [10], [12]. The impacts range from eroding trust in media and democracy [10], [12], to fostering social and political instability [10], and increasing public health risks [13]. Proposed solutions include improving media literacy [10], [12], promoting fact-checking and ethical journalism [10], and fostering collaboration between media, digital platforms, and the government to effectively regulate and counter hoax news [10].

2.2 *Content Moderation Policies on Social Media*

Content moderation on social media is a complex task that requires balancing the suppression of misinformation with the protection of free speech, a challenge that is especially pronounced in Indonesia, where regulations like the ITE Law aim to curb hoaxes but are often criticized for inconsistent enforcement. Social media platforms have adopted various strategies—such as community guidelines and differentiated approaches for mainstream and fringe platforms—to address misinformation and hate speech, yet these efforts frequently lack transparency and consistency [8], [14]. Facebook, for example, has been criticized for ambiguity in applying its standards, particularly in politically sensitive contexts [14]. Research on moral dilemmas in moderation shows that the public generally favors removing harmful misinformation over protecting free speech when the stakes are high, though opinions are divided along political lines, with Republicans showing less support for content removal than Democrats or Independents [15]. Calls for greater transparency and standardization in moderation practices stress the need for clearer disclosure, distinctions between types of misinformation, and more accountable decision-making processes [14].

2.3 *Legal Frameworks for Content Moderation in Indonesia*

The Indonesian government's strategy to combat hoax news combines legal frameworks, regulatory oversight, and emerging technological innovations. Central to this approach is the ITE Law, particularly Article 28, which criminalizes the dissemination of false information; however, its vague wording has drawn criticism for enabling misuse against dissenting opinions [16], [17]. Court rulings often reference outdated laws, revealing gaps in the legal infrastructure for effectively addressing misinformation, including cases involving fake social media accounts where enforcement is hindered by imprecise indictments [18]. Regulatory agencies such as the National Cyber and Encryption Agency (BSSN), in coordination with Kominfo, are responsible for monitoring online content, yet face difficulties managing the sheer volume and complexity of digital information [16]. Social media platforms are required to comply with Indonesian regulations, but enforcement inconsistencies and the lack of transparency in moderation remain pressing issues. To enhance these efforts, technological solutions like deep learning models—specifically CNN-LSTM hybrids—have shown high accuracy in detecting political hoaxes and offer promising tools for identifying misinformation, particularly during sensitive periods like elections [19].

2.4 *The Role of Fact-Checking and Media Literacy*

Efforts to combat hoax news in Indonesia have increasingly focused on promoting media literacy and developing fact-checking networks, with organizations like MAFINDO playing a pivotal role in verifying information—especially on platforms like WhatsApp, where misinformation on political and trivial topics is widespread [20]. Through initiatives like "turnbackhoax.id," MAFINDO and similar groups aim to educate the public on critically evaluating information and identifying trustworthy sources. Media literacy programs, such as those utilizing short educational videos, have proven effective in reducing the intention to share misinformation by as much as 64% among Indonesian social media users [21], while localized efforts like MAFINDO Solo Raya focus on socialization, education, and training during politically sensitive periods [22]. Despite these advances, challenges remain, including the limited effectiveness of fact-checking in politically polarized contexts, varying levels of trust in fact-checking entities [23], disparities in media literacy due to the digital divide, and the sheer volume of content on social media that complicates comprehensive verification [24].

3. RESEARCH METHODS

This research adopts a normative legal analysis approach to examine Indonesian laws and regulations governing social media content, particularly those aimed at combating hoax news. Normative legal analysis interprets and evaluates legal frameworks based on established norms and principles, with the objective of assessing their alignment with international best practices and their effectiveness in addressing misinformation. The study is further enriched by qualitative analysis through case studies, legal texts, and policy documents to identify patterns, inconsistencies, and implementation challenges within Indonesia's legal infrastructure. Data is drawn from secondary sources, including legal documents such as the ITE Law and related regulations, official government reports from institutions like Kominfo and BSSN, case studies of legal actions and notable hoax incidents, and the content moderation policies of social media platforms like Facebook, Twitter, Instagram, and WhatsApp. The research framework focuses on five key areas: evaluating the clarity and effectiveness of legal frameworks, assessing government and platform-based content moderation mechanisms, identifying challenges such as technological limitations and freedom of speech issues, measuring the impact on hoax news dissemination, and understanding stakeholder perspectives through published literature and reports.

Data analysis is conducted through doctrinal legal research methods for interpreting relevant legal provisions and a critical evaluation of policy effectiveness in real-world implementation. The study investigates the scope and clarity of laws like the ITE Law, and assesses how well content moderation policies—such as content removal, account suspension, and fact-checking—are enforced by both government agencies and digital platforms. Special attention is given to the obstacles faced in moderating content within Indonesia’s diverse, multilingual society, and to the technological constraints that hinder efficient policy enforcement. Additionally, the research adopts a comparative approach by analyzing Indonesia’s strategies in relation to those of countries facing similar misinformation challenges, such as Malaysia and the Philippines. This comparative lens provides insights into Indonesia’s global standing and identifies areas for potential policy improvement and international cooperation.

4. RESULTS AND DISCUSSION

4.1 Evaluation of Legal Frameworks for Content Moderation in Indonesia

The first key finding of this study centers on the limitations of Indonesia’s existing legal frameworks, particularly the Electronic Information and Transactions Law (ITE Law) and Law No. 11 of 2008 (UU ITE), in effectively addressing hoax news. While these laws criminalize the dissemination of false information that could harm the public interest—with penalties including fines and imprisonment—their broad and vague definitions of "false information" have led to arbitrary enforcement and potential misuse. Legal scholars and critics argue that this ambiguity blurs the line between combating misinformation and suppressing dissent, as seen in several cases where activists and journalists have been prosecuted under these laws [16], [25]. Moreover, the legal framework has struggled to keep pace with technological developments, particularly with the rise of encrypted messaging platforms like WhatsApp, where hoaxes commonly spread beyond the reach of conventional monitoring systems [17].

Enforcement of these laws faces further complications due to the lack of clear operational guidelines, especially on decentralized platforms, despite the regulatory powers granted to Kominfo [16]. The absence of specific mechanisms for tracking and moderating content in private digital spaces hinders the government's ability to manage misinformation effectively. This gap necessitates legal reform and clearer, more adaptive policies to better address the nuances of digital communication [17]. Additionally, the criminalization of hoax news, although intended to maintain public order, carries broader implications, including the risk of political abuse and infringement on freedom of expression [26]. In contrast, Islamic Criminal Law offers a normative perspective that emphasizes truthfulness and discourages the spread of falsehoods, presenting an alternative ethical framework in the fight against misinformation [27].

4.2 Effectiveness of Content Moderation Mechanisms

The second major finding of the study highlights the limited effectiveness of content moderation mechanisms in Indonesia, which involve both government regulation and the self-regulation of social media platforms. The Ministry of Communication and Information Technology (Kominfo) is responsible for removing or blocking hoax-related content under frameworks like the ITE Law and the Presidential Regulation No. 22 of 2023, especially during sensitive periods such as elections [28]. However, enforcement is hampered by technological limitations—such as underdeveloped AI systems—and delays that allow hoaxes to circulate widely before any intervention [29]. Despite legal mechanisms being in place, the government's capacity to monitor and act swiftly remains insufficient, particularly on platforms like WhatsApp, where encrypted content in private groups is harder to track and moderate.

Meanwhile, social media platforms like Facebook, Instagram, and Twitter implement their own moderation policies using AI and human moderators and often collaborate with local fact-checkers like MAFINDO ("Content moderation issues on social media are evolving", 2023). Yet, enforcement is inconsistent across platforms, with encrypted services such as WhatsApp posing significant challenges due to limited access and jurisdictional issues. Platforms headquartered outside Indonesia may prioritize their internal policies over compliance with local laws, resulting in fragmented enforcement efforts [7]. These challenges highlight the urgent need for a more integrated and localized strategy, combining government oversight with coordinated platform regulation. This should be complemented by enhanced digital literacy campaigns and improved cooperation with law enforcement to address misinformation more effectively [28], [30].

4.3 Challenges in Content Moderation Policies

The study identifies several key challenges that hinder the effectiveness of content moderation policies in Indonesia. One major issue is the ongoing tension between safeguarding freedom of speech and regulating harmful content. The ITE Law, while aimed at curbing misinformation, has been criticized for its vague provisions that enable its misuse to silence political dissent, restrict press freedom, and prosecute journalists and activists who criticize the government [31]. This has led to broader concerns over civil liberties and democratic backsliding. Additionally, excessive moderation practices under MOCI Regulation No. 5/2020 risk further infringing on freedom of expression, underscoring the need for balanced and rights-respecting policies [7].

Technological limitations further undermine enforcement efforts, as the current infrastructure relies heavily on human moderators and rudimentary algorithms that are inadequate for managing the massive and diverse content circulating on digital platforms

[7]. The linguistic and cultural diversity across Indonesia introduces added complexity, requiring more context-aware and sophisticated solutions. Low levels of media literacy among the population also exacerbate the spread of hoaxes, despite efforts by organizations like MAFINDO to raise awareness [32]. Public education on digital literacy and rights remains limited, especially in rural and underserved areas. Moreover, the lack of coordination among the government, social media platforms, and civil society has led to fragmented and ineffective moderation efforts. To address these challenges, a more integrated and collaborative framework is essential, involving all relevant stakeholders in developing consistent and transparent content regulation strategies [32].

4.4 The Impact of Content Moderation on Hoax News

Despite these challenges, some positive outcomes of content moderation policies have been observed. The removal of hoax content from major social media platforms, when it is identified and flagged, has helped reduce the immediate spread of false information. Additionally, the suspension of accounts responsible for disseminating hoaxes has acted as a deterrent for those attempting to spread misinformation.

However, the overall impact of content moderation policies on the widespread dissemination of hoax news remains mixed. While there are short-term gains in limiting the spread of specific hoaxes, the persistent nature of misinformation, driven by both human behavior and algorithmic amplification, means that hoax news continues to circulate, often finding new channels for distribution.

4.5 Recommendations for Improving Content Moderation Policies

Based on the findings of this study, several recommendations can be made to improve the effectiveness of content moderation in combating hoax news:

1. The ITE Law should be revised to provide clearer definitions of hoax news and to better balance freedom of expression with the need to protect public interest. This would ensure that the law is applied consistently and fairly.
2. The Indonesian government and social media platforms should invest in advanced AI tools and machine learning algorithms to detect hoax news more accurately and in real time. Additionally, the development of language-specific and context-sensitive detection systems can help address the multilingual and culturally diverse nature of Indonesia.
3. To reduce the vulnerability of the public to hoax news, widespread media literacy campaigns should be launched, focusing on teaching critical thinking and digital literacy skills, especially in rural areas.
4. A coordinated approach involving the government, social media platforms, civil society organizations, and the public is essential to create a more effective content moderation ecosystem. This could include the establishment of a national task force for combating misinformation that works alongside platforms to monitor, flag, and remove harmful content.

5. CONCLUSION

This research provides a critical evaluation of the impact of content moderation policies on the spread of hoax news in Indonesia. The findings indicate that while the existing legal framework—particularly the ITE Law—serves as a foundation for combating misinformation, its vague provisions risk overreach and may infringe on freedom of expression. Government and platform-based moderation efforts are further hampered by technological limitations, inconsistent enforcement, and low levels of media literacy. These factors collectively weaken the ability to respond promptly and effectively to the rapid spread of hoaxes across digital platforms.

Despite these obstacles, the study notes some progress, such as the removal of harmful content and suspension of accounts spreading misinformation. However, overall effectiveness remains limited, as misinformation continues to proliferate due to both human behavior and algorithm-driven amplification. To improve outcomes, the study recommends refining legal definitions, developing more sophisticated detection technologies, and expanding public digital literacy education. Most importantly, it emphasizes the need for stronger collaboration among the government, social media platforms, and civil society to build an integrated, transparent, and accountable framework for content moderation in Indonesia's digital landscape.

REFERENCES

[1] A. S. Sastrosubroto and S. Pratama, "Fake news and information in Social Media: Problems and challenges, the case in Indonesia.," *J. Sci. Temper*, vol. 7, no. 1&2, pp. 61–68, 2019.

[2] E. Effendi, "Turnitin User Behaviour And Hoax Information On Social Media Case Of Indonesia," *J. Stud. Komun.*, vol. 7, no. 3, pp. 930–943, 2023.

[3] E. Indartuti, I. Murti, and K. Kusnan, "Analisis Penyebaran Hoaks COVID-19 di Indonesia," *Society*, vol. 12, no. 2, pp. 251–278, 2024.

[4] A. M. Arif and A. Miswar, "An Analysis on How Hoax News Spread through Social Media," *Lit. Trends Libr. Dev.*, vol. 1, no. 2, pp. 42–51, 2020, doi: https://doi.org/10.24252/literatify.v1i2.14964.

[5] A. R. Cindoswari *et al.*, "Peningkatan Literasi Digital Dan Pengelolaan Media Sosial Sebagai Pencegahan Penyebaran Hoax Pada Siswa/I Di Sma Immanuel Kota Batam: Digital Literacy Improvement and Social Media Management as Prevention of the Spread of Hoax to Students in Immanuel High," *PUAN Indones.*, vol. 5, no. 1, pp. 207–218, 2023, doi: https://doi.org/10.37296/jpi.v5i1.166.

[6] P. Alamsyah, L. N. Hakim, G. Wijaya, and A. Wicaksono, "Debunking disinformation on YouTube: a fact check on the 2024 Indonesian election," *J. Stud. Komun.*, vol. 8, no. 3, pp. 547–560, 2024, doi: https://doi.org/10.25139/jsk.v8i3.8348.

[7] P. Audrine and I. Setiawan, "Impact of Indonesia's Content Moderation Regulation on Freedom of Expression," *Policy Pap.*, no. 38, 2021.

[8] M. Singhal *et al.*, "SoK: Content moderation in social media, from guidelines to enforcement, and research to practice," in *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*, 2023, pp. 868–895.

[9] B. Gunawan and B. M. Ratmono, "Social Media, Cyberhoaxes and National Security: Threats and Protection in Indonesian Cyberspace," *Int. J. Netw. Secur*, vol. 22, no. 1, pp. 93–101, 2020.

[10] M. Deddy Satria and Hairunnisa, "The Phenomenon of Fake News (Hoax) in Mass Communication: Causes, Impacts, and Solutions," *Open Access Indones. J. Soc. Sci.*, vol. 6, no. 3, pp. 980–988, 2023.

[11] P. Wannamaker, "A Working Definition of Fake News," 2022.

[12] M. de Vilhena, S. Moreira, and I. S. Guedes, "Fake News: Fatores de suscetibilidade e de disseminação," *Sociol. Rev. da Fac. Let. da Univ. do Porto*, vol. 48, 2024.

[13] A. Park, M. Montecchi, K. Plangger, and L. Pitt, "Understanding fake news: A bibliographic perspective," *Def. Strateg. Commun.*, vol. 8, pp. 141–172, 2020, doi: https://doi.org/10.30966/2018.RIGA.8.

[14] A. Zornetta, "Online misinformation: improving transparency in content moderation practices of social media companies," 2022.

[15] A. Kozyreva *et al.*, "Free speech vs. harmful misinformation: Moral dilemmas in online content moderation," in *Proceedings of the National Academy of Sciences of the United States of America*, 2022.

[16] S. N. Sari, "Konstruksi Tindak Pidana Berita Bohong (Hoax) Dalam Undang-Undang ITE," *Badamai Law J.*, vol. 6, no. 2, pp. 371–399, 2021.

[17] S. T. Santoso, F. Tanuwijaya, and I. Suarda, "Criminal Liability of Spreading Fake News on Social Media: Indonesian Criminal Law Perspective," *Indon. JLS*, vol. 4, no. 126–149, 2023.

[18] M. Rezky and A. L. Ibrahim, "Fake Accounts on Social Media as a Criminal Act of Electronic Information Manipulation in Indonesia," *Yuridika*, vol. 37, no. 3, p. 615, 2022.

[19] Y. Sibaroni, S. Mahadzir, S. S. Prasetiyowati, and A. F. Ihsan, "Combating Misinformation: Leveraging Deep Learning for Hoax Detection in Indonesian Political Social Media," *J. INFOTEL*, vol. 16, no. 2, pp. 413–426, 2024.

[20] D. Rahmawan, I. Garnesia, and R. Hartanto, "Content Analysis of MAFINDO's Verified WhatsApp-Related Misinformation in Indonesia," *J. Kaji. Jurnalisme*, vol. 8, no. 1, pp. 99–114, 2024.

[21] T. W. Ford, M. Yankoski, M. Facciani, and T. Weninger, "Online Media Literacy Intervention in Indonesia Reduces Misinformation Sharing Intention," *J. Media Lit. Educ.*, vol. 15, no. 2, pp. 99–123, 2023.

[22] R. D. Maqruf, "Bahaya Hoaks Dan Urgensi Literasi Media: Studi Pada Mafindo Solo Raya," *Acad. J. Da'wa Commun.*, vol. 2, no. 1, pp. 121–150, 2021.

[23] M. Facciani, I. Adinugroho, D. Apriliawati, and T. Weninger, "Tackling Misinformation in Indonesia: Assessing Fact-

Checking and Media Literacy," 2024.

[24] F. J. Umar, M. Martinihani, and N. R. Marsuki, "Waspada Dunia Maya: Strategi Mengidentifikasi dan Mengatasi Hoaks," *J. Pendidik. DAN ILMU Sos.*, vol. 2, no. 2, pp. 114–122, 2024.

[25] I. Rosyadi, "Criminal Liability Against Perpetrators of HOAX Spread in Indonesia," *Int. J. Law Dyn. Rev.*, vol. 1, no. 1, pp. 41–53, 2023.

[26] C. B. Devina, D. C. Iswari, G. C. B. Goni, and D. K. Lirungan, "Tinjauan Hukum Kriminalisasi Berita Hoax: Menjaga Persatuan vs. Kebebasan Berpendapat," *Kosmik Huk.*, vol. 21, no. 1, pp. 44–58, 2021.

[27] R. Maghfiroh and R. Abbas, "Studi komparasi penyebaran berita bohong (hoax) perspektif uu ite dan hukum pidana Islam," *Rechtenstudent*, vol. 1, no. 2, pp. 154–165, 2020.

[28] R. F. Putra and A. A. Nugroho, "The Role of The Ministry of Communications and Information In Preventing The Spread of Hoaxes During The 2024 Election," *J. Law, Polit. Humanit.*, vol. 4, no. 4, pp. 1018–1028, 2024.

[29] A. Y. Sulistyawan and S. A. G. Pinilih, "The Reality of Spreading Hoaxes on Social Media: A Sociolegal Approach," in *2nd International Conference on Indonesian Legal Studies (ICILS 2019)*, 2019, pp. 113–117.

[30] K. M. Herawati, "Pengaturan Pemblokiran Konten Penyebaran Kampanye Hitam Melalui Media Sosial," *KERTHA WICAKSANA*, vol. 18, no. 2, pp. 62–70, 2024.

[31] A. Fakih, "Media Under the Law: Press Freedom Challenges in Indonesia," *Indones. Media Law Rev.*, vol. 3, no. 1, 2024.

[32] H. Alvina, L. Julianti, A. A. P. W. Sugiantari, and I. W. W. W. Udytama, "The State of Digital Freedom in Indonesia an Assessment of Online Censorship, Privacy, and Free Expression," *J. Digit. Law Policy*, vol. 1, no. 3, pp. 141–152, 2022.