

A Call For Responsibility: Social Media And The Need To Monitor Hate Speech And Fake News For Global Peace

Karim Darban¹, Smail Kabbaj²

^{1,2} ENCG CASABLANCA, University Hassan II, Casablanca, Morocco

ARTICLE INFO

Article history:

Received 1 June
Accepted 29 September
Available online

Keywords:

Social Media,
Online content monitoring,
Hate Speech,
Fake News,
Peace

ABSTRACT

This paper explores the pivotal role of social media in fostering peace and examines responsible monitoring practices to counteract hate speech and fake news propagation. Leveraging an interdisciplinary approach, our research method entailed case studies and a systematic review of existing literature encompassing studies on social media's impact on public opinion, the spread of misinformation, and successful instances of peace promotion through online platforms. Our findings indicate that while social media has the potential to facilitate peacebuilding, the challenges posed by hate speech and misinformation persist. We elucidate the crucial role of artificial intelligence (AI) and human moderation in identifying and mitigating harmful content, underscoring the need for transparent algorithms, human oversight, and continual evaluation. This paper also highlights successful case studies where social media has been instrumental in bridging divides and promoting dialogue. We contend that by addressing these challenges and implementing responsible monitoring strategies, social media should contribute to a more peaceful and inclusive online environment.

© 2023 The Author(s). Published by AIRA.
This is an open access article under the CC BY-SA license
(<http://creativecommons.org/licenses/by-sa/4.0/>).



Corresponding Author:

Karim Darban,
ENCG CASABLANCA, University Hassan II, Casablanca, Morocco.
Email: dr.karim.darban@gmail.com

1. INTRODUCTION

1.1 Explanation of the growing concern regarding hate speech and fake news on social media.

The advent of social media has revolutionized the way people communicate, share information, and consume news. Social media platforms have created a space for people to express their views and opinions, connect with others, and stay informed. However, as social media has grown in popularity, it has also become a breeding ground for hate speech and fake news [1]. Hate speech is a form of expression that promotes or incites hatred, discrimination, or violence against individuals or groups based on their identity, such as race, religion, ethnicity, gender, or sexual orientation [2]. It has been on the rise on social media platforms in recent years, with users often using anonymity to post hateful comments and messages. The impact of hate speech on social media has been significant, with studies showing that it can lead to social exclusion, psychological harm, and even violence [3].

Fake news, on the other hand, is false or misleading information presented as real news. It often spreads rapidly on social media, with users sharing it without fact-checking or verifying its accuracy. The spread of fake news has become a significant concern in recent years, with it often being used to influence public opinion, elections, and even incite violence. The consequences of hate speech and fake news on social media are far-reaching. They can undermine social cohesion, lead to mistrust and division, and even cause physical harm to individuals and communities. The rise of hate speech and fake news on social media has led to growing concerns about the role of social media platforms in promoting peace and ensuring responsible online behavior [3].

1.2 Importance of social media in shaping public opinion and discourse.

Social media has become an integral part of modern society, and its impact on public opinion and discourse is significant. Social media platforms allow users to share their views, connect with others, and engage in discussions about a wide range of topics. Social media has the power to shape public opinion and discourse, with users often relying on social media as a source of information and a platform for expression [4].

The widespread use of social media has led to the democratization of information and the rise of citizen journalism. Social media platforms have given a voice to marginalized communities, allowing them to share their stories and experiences with a wider audience. Social media has also played a crucial role in organizing and mobilizing social movements, such as the Arab Spring, Black Lives Matter, and #MeToo.

The impact of social media on public opinion and discourse is not without its challenges. Social media has been criticized for its role in the spread of hate speech, fake news, and disinformation. Social media algorithms are often designed to prioritize content that is likely to generate engagement, which can lead to the amplification of polarizing and divisive content [5]. Social media has also been accused of promoting filter bubbles and echo chambers, where users are only exposed to information and opinions that confirm their existing beliefs and biases. This can lead to a lack of diversity in perspectives and limit the potential for constructive dialogue and meaningful exchange of ideas[6]. Despite these challenges, social media remains a powerful tool for shaping public opinion and discourse. It has the potential to bring people together, promote understanding, and foster empathy. Social media can also be used to promote peace and social harmony, by creating spaces for dialogue, promoting inclusive discourse, and combatting hate speech and fake news [4].

1.3 Exploring the need for social media to take responsibility for promoting peace by monitoring hate content and fake news.

Given the impact of hate speech and fake news, there is a growing need for social media companies to take responsibility for monitoring their platforms and ensuring they do not become a vehicle for promoting hate speech and fake news. Social media platforms have a responsibility to create safe and inclusive online environments and take steps to prevent the spread of harmful content.

This paper will explore the need for social media to take responsibility for promoting peace by monitoring hate content and fake news. It will delve into the challenges that social media companies face in monitoring their platforms, the ethical implications of their actions, and the best practices that can be adopted to promote responsible monitoring.

The paper will also examine the potential of social media in promoting peace and social harmony, by creating spaces for dialogue, promoting inclusive discourse, and combatting hate speech and fake news. It will explore the ways in which social media platforms can be used as a tool for promoting peace and social harmony, and the role of responsible monitoring in ensuring that social media remains a platform for constructive dialogue and meaningful exchange of ideas.

Ultimately, the goal of this paper is to highlight the importance of responsible social media use and the potential of social media platforms in promoting peace and social harmony. By examining the challenges and opportunities of social media in promoting peace, this paper aims to contribute to the ongoing debate about the role of social media in shaping public opinion and discourse, and the responsibility of social media companies in creating a safer and more inclusive online environment.

2. RESEARCH METHOD

In this section, we elaborate on the research methods employed in our study titled "A Call for Responsibility: Social Media and the Need to Monitor Hate Speech and Fake News for Global Peace." Our research methodology was carefully designed to provide a comprehensive and nuanced understanding of the role of social media in fostering peace and the strategies for responsible monitoring of hate speech and fake news.

Interdisciplinary Approach: Our research approach was inherently interdisciplinary, drawing upon insights from various fields including social sciences, communication studies, computer science, and conflict resolution. This interdisciplinary perspective allowed us to analyze the complex interplay between technology, human behavior, and societal impact.

Systematic Literature Review: We conducted a systematic review of existing literature on the topics of social media's impact on public opinion, the dissemination of misinformation, and successful instances of peace promotion through online platforms. Our literature review encompassed peer-reviewed academic articles, books, reports, and relevant grey literature. This extensive review of existing research allowed us to synthesize existing knowledge and identify gaps in the literature, which we addressed in our study.

Case Studies: We conducted a series of case studies to examine real-world instances where social media played a pivotal role in either exacerbating conflict or promoting peace. These case studies provided data that allowed us to explore the nuances and complexities of the relationship between social media and peacebuilding. The selected case studies included contexts such as the Israeli-Palestinian conflict, Myanmar's Rohingya crisis, political polarization in the United States, communal violence in India, and anti-immigrant sentiment in Europe. By analyzing these diverse cases, we were able to identify patterns and trends across different contexts.

3. Results and discussion

3.1 Understanding Hate Content and Fake News

1. Defining hate speech and its impact on social media:

Hate speech refers to any form of speech or expression that promotes discrimination, hostility, or violence against an individual or a group based on their race, ethnicity, religion, gender, sexual orientation, or other characteristics [2]. Hate speech can take various forms, including verbal or written language, gestures, images, and symbols. The impact of hate speech can be devastating, leading to social exclusion, psychological harm, and even physical violence [7]. Social media platforms have become a breeding ground for hate speech, providing a space for individuals and groups to express their prejudices and biases. Social media companies have been criticized for their role in allowing hate speech to flourish on their platforms, with some arguing that they have failed to take sufficient action to prevent the spread of harmful content [8].

The impact of hate speech on social media is significant, with users often experiencing psychological distress, harassment, and even physical harm as a result of exposure to hate speech. The spread of hate speech on social media can also contribute to the normalization of discriminatory attitudes and beliefs, leading to increased social polarization and divisiveness [9]. In response to the growing concern about hate speech on social media, many social media companies have introduced policies and measures aimed at preventing the spread of harmful content. These policies typically involve the removal of content that violates their community standards and the suspension or termination of user accounts that engage in hate speech [10].

While these measures have been effective in some cases, they have also been criticized for being too subjective and unfair. The challenge for social media companies is to strike a balance between preventing the spread of harmful content and promoting free speech [11]. To address the issue of hate speech on social media effectively, it is essential to understand the underlying factors that contribute to its spread. This includes examining the role of social and political factors, such as the rise of nationalism and populism, in fueling the spread of hate speech [12] [13]. It is also important to recognize the impact of hate speech on different communities, including marginalized groups who are most vulnerable to its effects. This includes exploring the intersectionality of hate speech and the ways in which it can reinforce existing inequalities and power imbalances.

3.2 Understanding fake news and its spread on social media:

The spread of fake news on social media is facilitated by the fact that social media platforms are designed to prioritize content that generates engagement and shares, rather than the accuracy or reliability of the information. This means that false or misleading information can often spread rapidly on social media, reaching a wide audience before it can be corrected or debunked [14].

The impact of fake news on social media can be significant, with users often forming opinions and making decisions based on false information. Fake news can also contribute to the spread of conspiracy theories and undermine trust in institutions and the media. The spread of fake news on social media is often driven by individuals and organizations who seek to manipulate public opinion or advance their interests. This includes foreign actors who use fake news to interfere in elections and sow discord in other countries [14].

To address the issue of fake news on social media effectively, it is essential to understand the underlying factors that contribute to its spread. This includes examining the role of social and political factors, such as the erosion of trust in traditional media, in fueling the spread of fake news.

It is also important to recognize the impact of fake news on different communities, including those who are most vulnerable to its effects. This includes exploring the ways in which fake news can reinforce existing biases and stereotypes and contribute to the marginalization of certain groups [15].

In response to the growing concern about fake news on social media, many social media companies have introduced measures aimed at preventing the spread of false or misleading information. These measures typically involve the use of algorithms and fact-checking tools to identify and flag potentially false information [15]. While these measures have been effective in some cases, they have also been criticized for not sufficiently proactive in preventing the spread of fake news [16]. The challenge for social media companies is to develop more effective strategies for identifying and preventing the spread of false or misleading information on their platforms.

3.3 Examples of hate speech and fake news leading to conflict and violence :

Hate speech and fake news have been implicated in numerous instances of conflict and violence around the world. In this chapter, we will examine several examples of how hate speech and fake news have contributed to conflict and violence in different contexts.

Palestine & Israel: Hate speech and fake news have been used by social media users in the Israeli-Palestinian conflict to demonize and dehumanize the other, contributing to a climate of hostility and violence. False claims about Palestinian terrorism and Israeli aggression have been circulated on social media, contributing to a lack of trust and a failure to reach a negotiated settlement [17] [18] [19] [20].

Myanmar: In Myanmar, hate speech and fake news on social media played a significant role in the persecution of the Rohingya Muslim minority. False rumors and misinformation circulated on social media about the Rohingya, portraying them as a threat to national security and as terrorists. This led to widespread violence and displacement, with hundreds of thousands of Rohingya forced to flee their homes [21]

United States: Hate speech and fake news have contributed to a climate of political polarization and violence. Claims about election fraud and conspiracy theories about the COVID-19 pandemic have been circulated on social media, leading to mass polarizations [22]

India: The spread of fake news on social media has contributed to a rise in communal violence, particularly targeting religious minorities. False information about alleged attacks on Hindu temples and the slaughter of cows, which are considered sacred in Hinduism, has fueled mob violence and lynchings of Muslims and other minorities [23]

Europe: In Europe, hate speech and fake news have been used to target refugees and migrants, fueling xenophobic and anti-immigrant sentiment. False claims about the alleged criminality and threats posed by refugees have been circulated on social media, contributing to a rise in hate crimes and political movements that seek to restrict migration [24].

These examples demonstrate the significant impact that hate speech and fake news can have on social and political dynamics, contributing to conflict and violence in different contexts. They also underscore the need for social media companies to take responsibility for monitoring and regulating hate speech and fake news on their platforms.

3.4 The role of social media in promoting peace

Exploring the potential of social media in promoting peace :

Social media has emerged as a powerful tool for shaping public opinion and discourse, and it has the potential to play a significant role in promoting peace. Despite its negative impact on the spread of hate speech and fake news, social media can also be used to counter these issues and promote positive messages of peace and understanding.

Monitoring and deleting hate speech and fake news:

Social media companies have a responsibility to monitor and delete hate speech and fake news on their platforms. By removing harmful content, social media companies can help to prevent the spread of disinformation and inflammatory messages that can exacerbate tensions and fuel conflict. Facebook, for example, has implemented a system of content moderation that relies

on both human moderators and artificial intelligence to detect and remove hate speech and fake news [10]. This approach, while imperfect, has helped to limit the spread of harmful content on the platform.

Building bridges between communities:

Social media can be used to build bridges between communities that are divided by conflict or mistrust. By providing a platform for dialogue and communication, social media can help to break down barriers and foster understanding between different groups. This can be particularly effective in conflict zones, where traditional means of communication may be limited or unsafe [25].

Promoting positive narratives:

Social media can also be used to promote positive narratives about peace and conflict resolution. By highlighting stories of individuals and groups who are working towards peace and understanding, social media can inspire others to follow their lead. This can help to shift the focus away from divisive and negative messages, and towards positive and constructive ones [26].

Providing a space for activism:

Social media can provide a space for activism and advocacy, allowing individuals and groups to mobilize around issues related to peace and conflict. By providing a platform for individuals to share their stories and experiences, social media can help to amplify voices that are often marginalized or ignored [27].

Encouraging empathy and understanding:

Social media can also encourage empathy and understanding between different groups, by allowing individuals to share their perspectives and experiences. By exposing individuals to different viewpoints and experiences, social media can help to break down stereotypes and promote understanding between different groups [28].

Despite these potential benefits, social media is not a panacea for promoting peace. It can also be used to spread hate speech and fake news, and can exacerbate tensions and conflict if not used responsibly. As such, it is important for social media companies and users to take responsibility for promoting positive messages of peace and understanding, and to take steps to address the spread of hate speech and fake news on their platforms.

Identifying the challenges in using social media for peacebuilding :

While social media has the potential to promote peace, it also presents a number of challenges that must be addressed in order to effectively leverage its power. The following are some of the challenges in using social media for peacebuilding:

Filter bubbles and echo chambers:

Social media algorithms often prioritize content that aligns with a user's existing beliefs and interests, creating filter bubbles and echo chambers that can reinforce existing biases and prevent exposure to diverse perspectives [6]. This can limit the potential for social media to foster understanding and promote peace between different groups.

Lack of accountability:

Social media companies have faced criticism for their lack of accountability in addressing harmful content on their platforms. Without effective regulation and enforcement, social media may continue to be used to spread messages of hate and disinformation, undermining efforts to promote peace [29].

Online harassment:

Social media can be a breeding ground for online harassment and trolling, particularly against individuals or groups with opposing views or those advocating for peace. This harassment can create a hostile environment that discourages constructive dialogue and undermines efforts to promote peace [30].

Polarization and division:

Social media can amplify and exacerbate existing divisions between different groups, particularly in contexts of conflict. This can lead to increased polarization and hostility, further hindering efforts to promote peace [22].

Manipulation and misinformation:

Social media can be used to manipulate public opinion and disseminate misinformation, particularly in the context of political or social unrest. This can undermine efforts to promote peace by perpetuating falsehoods and sowing distrust between different groups (Wu, L., et al., 2019).

Limitations of online communication:

While social media can facilitate communication and dialogue, it also has limitations. Online communication lacks many of the nuances and nonverbal cues of face-to-face interaction, making it easier for misunderstandings to occur. Additionally, social media may not be able to fully address the complex issues underlying conflicts, such as historical grievances or power imbalances [31]

Limited access to technology:

Not all communities have equal access to technology, including social media. This can create a digital divide that further marginalizes some groups, particularly those in conflict zones or those without the means to access social media. This lack of access can limit the effectiveness of social media as a tool for peacebuilding.

These challenges are significant, but they do not negate the potential for social media to promote peace. Rather, they underscore the importance of using social media responsibly and addressing these challenges in order to fully leverage its potential for peacebuilding. By recognizing these challenges and taking steps to overcome them, we can work towards a more peaceful and just world.

Highlighting successful instances of social media promoting peace between Palestine and Israel :

While there are certainly challenges in using social media for peacebuilding, there are also many examples of successful efforts to use social media to promote peace and reconciliation. The following are some examples of successful instances of social media promoting peace:

The Parents Circle - Families Forum:

Is a joint Israeli-Palestinian organization made up of families who have lost loved ones in the conflict. The organization has a strong social media presence, with over 43k followers on Facebook. They use social media to share stories of their personal experiences and to promote their message of reconciliation and peace.

Combatants for Peace:

Is an Israeli-Palestinian organization made up of former Israeli soldiers and Palestinian fighters who have renounced violence and are now working together for peace. The organization has a strong social media presence, with over 26k followers on Facebook. They use social media to share stories of their joint activities and to promote their message of peace and reconciliation.

Repair the World:

Is a Jewish organization that aims to inspire and mobilize the Jewish community to take action and help address the most pressing social issues of our time. They use social media to share stories of individuals and organizations who are working to make a difference, and to encourage others to get involved.

The Hand in Hand Center for Jewish-Arab Education:

Is an Israeli organization that promotes Jewish-Arab coexistence through education. The organization has a strong social media presence, with over 13k followers on Facebook. They use social media to share stories of successful Jewish-Arab partnerships and to promote their programs and events.

The MyJihad Campaign:

Was started by a group of American Muslims in response to the negative stereotypes and misconceptions surrounding the word "jihad" (that literally translates to "struggle" or "effort"). The campaign encourages Muslims to share their own personal struggles and challenges on social media using the hashtag #MyJihad. The campaign has been successful in promoting understanding and breaking down stereotypes.

These examples demonstrate that social media can be a powerful tool for promoting peace and reconciliation. By leveraging the power of social media to share messages of peace, promote understanding, and raise awareness of important issues, we can work towards a more peaceful and just world.

3.5 Best Practices for Responsible Monitoring

The role of artificial intelligence in detecting hate speech and fake news :

With the growing concern over hate speech and fake news on social media, it is becoming increasingly important for social media platforms to take responsibility for monitoring and removing such content. One tool that has been increasingly used for this purpose is artificial intelligence (AI). In this chapter, we will explore the role of AI in detecting hate speech and fake news, and the best practices for responsible monitoring.

AI has the potential to be a powerful tool in detecting hate speech and fake news on social media. By analyzing large amounts of data and patterns of behavior, AI algorithms can quickly identify and flag potentially harmful content. Some of the techniques used by AI for detecting hate speech and fake news include:

Natural Language Processing (NLP):

NLP is a branch of AI that deals with the interaction between computers and human language. By analyzing the words and phrases used in social media posts, NLP algorithms can identify patterns of hate speech and misinformation (Al-Makhadmeh, Z., & Tolba, A., 2020).

Machine Learning:

Machine learning is another AI technique that involves the use of algorithms to analyze data and identify patterns. By training machine learning algorithms on large datasets of social media content, they can learn to recognize patterns of hate speech and fake news [32]

Network Analysis:

Network analysis involves the analysis of social media networks to identify patterns of behavior and influence. By analyzing the networks of individuals and groups who are spreading hate speech and fake news, AI algorithms can identify and flag potentially harmful content

While AI has great potential for detecting hate speech and fake news, there are also several challenges and limitations to its use. Some of these challenges include:

Contextual understanding:

One of the biggest challenges in using AI to detect hate speech and fake news is the need for contextual understanding. Words and phrases that may be harmless in one context may be considered hate speech or fake news in another context (Al-Makhadmeh, Z., & Tolba, A., 2020). AI algorithms need to be able to understand the context of a social media post in order to accurately identify potentially harmful content.

Bias and accuracy:

Another challenge with AI is the potential for bias and accuracy issues. AI algorithms are only as accurate as the data they are trained on. If the training data contains biases, those biases will be reflected in the algorithm's output. Additionally, algorithms may make mistakes in identifying hate speech and fake news, leading to the removal of legitimate content [32]

Adaptability:

Social media is constantly changing, with new platforms and new types of content emerging all the time. AI algorithms need to be adaptable to these changes in order to continue to be effective in detecting hate speech and fake news [32]

Despite the challenges and limitations of AI in detecting hate speech and fake news, there are several best practices that social media platforms can follow to ensure responsible monitoring. These include:

Transparent algorithms:

Social media platforms should be transparent about the algorithms they use for detecting hate speech and fake news. This includes disclosing how the algorithms were trained, what data they use, and what metrics they use to identify potentially harmful content [33]

Human oversight:

While AI can be effective in detecting hate speech and fake news, it should not be relied on exclusively. Human moderators should be used to review content flagged by the AI algorithms to ensure accuracy and fairness [34]

Continual monitoring and evaluation:

Social media platforms should continually monitor and evaluate their AI algorithms to ensure that they are effective and accurate. This includes regularly reviewing the training data and adapting the algorithms to changes in social media behavior [34]

Collaboration with experts:

Collaboration with experts from various fields is essential in developing effective AI tools for monitoring hate speech and fake news on social media. Experts in linguistics, psychology, sociology, and other related fields can help in creating algorithms that accurately detect hate speech and fake news. Furthermore, working with experts in conflict resolution and peacebuilding can provide valuable insights into how to use AI tools to promote peace and prevent conflicts on social media. These experts can help in developing strategies that leverage the power of social media to build bridges between communities and foster constructive dialogue. Collaboration with NGOs and civil society groups is also crucial in developing AI tools that are grounded in ethical principles and respect for human rights. These organizations can provide feedback on the impact of AI tools on marginalized communities and help in identifying potential biases in the algorithms used for hate speech and fake news detection [34]

Human moderation and its impact on monitoring hate speech and fake news :

While AI tools can be effective in detecting hate speech and fake news, they are not infallible. AI algorithms may miss certain nuances or contexts that are essential in determining whether a particular post contains hate speech or fake news. Moreover, some types of hate speech and fake news may be more challenging for AI algorithms to detect, such as hate speech that is disguised through sarcasm or irony.

Therefore, human moderation is still essential in monitoring hate speech and fake news on social media. Human moderators can provide a more nuanced understanding of language and context and can quickly identify false information or hate speech that may slip through AI algorithms.

However, there are also potential drawbacks to human moderation. Human moderators may have biases or preconceived notions that could influence their judgment. Therefore, it is essential to have a diverse team of moderators who are trained in

identifying and avoiding personal biases. Additionally, human moderation can be time-consuming and expensive, especially for social media platforms with a vast user base. This may make it challenging for smaller social media platforms to invest in human moderation.

The importance of transparency and accountability in responsible monitoring :

Transparency and accountability are essential components of responsible monitoring on social media. Social media platforms must be transparent about their content moderation policies and processes to ensure that their users understand how their content is being monitored and what standards they need to follow [33]. Transparency can help to build trust with users and create a sense of accountability for the platform.

Platforms should be transparent about how they determine what constitutes hate speech and fake news, and how they enforce their policies. They should also be clear about the consequences for violating those policies, including the removal of content or suspension of accounts. This can help users understand what is and is not acceptable behavior on the platform, and it can also discourage the spread of hate speech and fake news.

Moreover, social media platforms should be accountable for their content moderation decisions. Users should have a way to appeal moderation decisions and receive an explanation of why their content was flagged or removed. This can help to ensure that content moderation is fair and impartial, and that users' rights to free speech and expression are protected [34]. Finally, platforms should be transparent about how they handle user data, including how they collect, store, and use it. Users should have control over their data, and platforms should be accountable for how they use it [35].

4. CONCLUSION

This paper underscores the substantial influence of social media on shaping public discourse and peacebuilding efforts. While social media holds promise in promoting peace, it remains plagued by the persistent issues of hate speech and fake news. The research has emphasized the essential role of artificial intelligence and human moderation in content monitoring, advocating for transparency, accountability, and continual evaluation as vital components. Furthermore, the paper showcases successful instances where social media has been a catalyst for dialogue and understanding. As we look ahead, the responsibility falls on both social media platforms and policymakers to enhance content monitoring, educate users, and prioritize peace and security in the digital realm. Ultimately, responsible monitoring practices can harness social media's potential as a potent instrument for a more harmonious and tolerant society.

REFERENCES

- [1] G. Di Domenico, J. Sit, and D. Ishizaka Alessio and Nunan, "Fake news, social media and marketing: A systematic review," *J. Bus. Res.*, vol. 124, pp. 329–341, Jan. 2021.
- [2] R. K. Whillock and D. Slayden, *2455 Teller Road*. Thousand Oaks, CA 91320: SAGE Publications, Inc, 1995.
- [3] A. Matamoros-Fernández and J. Farkas, "Racism, hate speech, and social media: A systematic review and critique," *Telev. New Media*, vol. 22, no. 2, pp. 205–224, Feb. 2021.
- [4] N. Anstead and B. O'Loughlin, "Social media analysis and public opinion: The 2010 UK general election," *J. Comput. Mediat. Commun.*, vol. 20, no. 2, pp. 204–220, Mar. 2015.
- [5] M. Petrescu and A. S. Krishen, "The dilemma of social media algorithms and analytics," *J. Mark. Anal.*, vol. 8, no. 4, pp. 187–188, Dec. 2020.
- [6] L. Terren, R. Borge-Bravo, Open University of Catalonia, and Open University of Catalonia, "Echo chambers on social media: A systematic review of the literature," *Rev. Commun. Res.*, vol. 9, pp. 99–118, 2021.
- [7] M. A. Paz, J. Montero-Díaz, and A. Moreno-Delgado, "Hate Speech: A systematized review," *Sage Open*, vol. 10, no. 4, p. 215824402097302, Oct. 2020.
- [8] S. A. Castaño-Pulgarín, N. Suárez-Betancur, L. M. T. Vega, and H. M. H. López, "Internet, social media and online hate speech. Systematic review," *Aggress. Violent Behav.*, vol. 58, no. 101608, p. 101608, May 2021.
- [9] M. Mondal, L. A. Silva, and F. Benevenuto, "A measurement study of hate speech in social media," in *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, New York, NY, USA: ACM, Jul. 2017.
- [10] "Facebook." Accessed: May 05, 2023. [Online]. Available: <https://transparency.fb.com/en-gb/policies/community-standards/hate-speech/>
- [11] S. Tiwari and G. Ghosh, "Social media and freedom of speech and expression: Challenges before the Indian law," *SSRN Electron. J.*, 2014.
- [12] M. Lim, "Freedom to hate: social media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia," *Crit. Asian Stud.*, vol. 49, no. 3, pp. 411–427, Jul. 2017.

- [13] P. Bhat and O. Klein, "Covert hate speech: White nationalists and dog whistle communication on twitter. Twitter, the public sphere, and the chaos of online deliberation," pp. 151–172, 2020.
- [14] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, May 2017.
- [15] Y. M. Rocha, G. A. de Moura, G. A. Desidério, F. D. de Oliveira Carlos Henrique and Lourenço, and L. D. de Figueiredo Nicolete, "The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review," *Z. Gesundh. Wiss.*, vol. 31, no. 7, pp. 1–10, Oct. 2021.
- [16] E. C. Tandoc Jr, D. Lim, and R. Ling, "Diffusion of disinformation: How social media users respond to fake news and why," *Journalism (Lond.)*, vol. 21, no. 3, pp. 381–398, Mar. 2020.
- [17] S. Masharqa, *Fake News in Palestine: exploratory research into the content, channels, and responses*. 2020.
- [18] M. Zahoor and N. Sadiq, "Digital public sphere and Palestine-Israel conflict: A conceptual analysis of news coverage," *Lib. Arts Soc. Sci. Int. J.*, vol. 5, no. 1, pp. 168–181, May 2021.
- [19] J. Salminen, M. Milenkovic, S.-G. Sengiu Sercan and Jung, and B. J. Jansen, "Weaponizing words: Analyzing fake news accusations against two online news channels," in *2021 8th International Conference on Behavioral and Social Computing (BESC)*, IEEE, Oct. 2021.
- [20] G. Hitman and M. Zwilling, "Normalization with Israel: An analysis of social networks discourse within gulf states," *Ethnopolitics*, vol. 21, no. 4, pp. 423–449, Aug. 2022.
- [21] C. Fink, "Dangerous speech, anti-Muslim violence, and Facebook in Myanmar," *J Int Aff*, vol. 71, no. 1, pp. 43–52, 2018.
- [22] L. Grimminger and R. Klinger, "Hate towards the political opponent: A twitter corpus study of the 2020 US Elections on the basis of offensive speech and stance detection," Mar. 2021.
- [23] A. M. Ishmam, J. Arman, and S. Sharmin, "Towards the development of the Bengali language corpus from public Facebook pages for hate speech research," in *Proceedings of Asian CHI Symposium 2019: Emerging HCI Research Collection*, New York, NY, USA: ACM, May 2019.
- [24] C. Arcila-Calderón, D. Blanco-Herrero, M. Fr\`ias-Vázquez, and F. Seoane-Pérez, "Refugees welcome? Online hate speech and sentiments in Twitter in Spain during the reception of the boat Aquarius," *Sustainability*, vol. 13, no. 5, p. 2728, Mar. 2021.
- [25] H. Baytiyeh, "Social media's role in peacebuilding and post-conflict recovery," *Peace Rev.*, vol. 31, no. 1, pp. 74–82, Jan. 2019.
- [26] P. C. Ragandang III, "Social media and youth peacebuilding agency: A case from Muslim Mindanao," *J. Peacebuilding Dev.*, vol. 15, no. 3, pp. 348–361, Dec. 2020.
- [27] M.-G. Chon and H. Park, "Social media activism in the digital age: Testing an integrative model of activism on contentious issues," *Journal. Mass Commun. Q.*, vol. 97, no. 1, pp. 72–97, Mar. 2020.
- [28] T. Blakemore and K. Agllias, "Social media, empathy and interpersonal skills: social work students' reflections in the digital era," *Soc. Work Educ.*, vol. 39, no. 2, pp. 200–213, Feb. 2020.
- [29] M. K. Feeney and G. Porumbescu, "The limits of social media for public administration research and practice," *Public Adm. Rev.*, vol. 81, no. 4, pp. 787–792, Jul. 2021.
- [30] F. Stevens, J. R. C. Nurse, and B. Arief, "Cyber stalking, cyber harassment, and adult mental health: A systematic review," *Cyberpsychol. Behav. Soc. Netw.*, vol. 24, no. 6, pp. 367–376, Jun. 2021.
- [31] J. M. Basch, K. G. Melchers, A. Kurz, M. Krieger, and L. Miller, "It takes more than a good camera: Which factors contribute to differences between face-to-face interviews and videoconference interviews regarding performance ratings and interviewee perceptions?," *J. Bus. Psychol.*, vol. 36, no. 5, pp. 921–940, 2021.
- [32] O. Oriola and E. Kotze, "Evaluating machine learning techniques for detecting offensive and hate speech in south African tweets," *IEEE Access*, vol. 8, pp. 21496–21509, 2020.
- [33] U. Reisach, "The responsibility of social media in times of societal and political manipulation," *Eur. J. Oper. Res.*, vol. 291, no. 3, pp. 906–917, Jun. 2021.
- [34] B. Ganesh and J. Bright, "Countering extremists on social media: Challenges for strategic communication and content moderation," *Policy Internet*, vol. 12, no. 1, pp. 6–19, Mar. 2020.
- [35] H. Felzmann, E. F. Villarronga, and A. Lutz Christoph and Tamò-Larrioux, "Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns," *Big Data Soc.*, vol. 6, no. 1, p. 205395171986054, Jan. 2019.