ISSN 2706-6827 Vol. 5, Issue 8: 62-65, DOI: 10.25236/IJFS.2023.050811

Social Media Companies' Moderation of UGC and Journalistic Content Published on Their Platforms

Jianing Cong

Department of Journalism Studies, The University of Sheffield, Sheffield, S10 2TN, UK

Abstract: *Social media plays a vital role in people's lives today, and its use is widespread. According to a data survey from Statista Research, Facebook had approximately 2.93 billion monthly active users as of the first quarter of 2022. As of January 2022, there were 76.9 million Twitter users in the US. With the widespread use of social media, news content and UGC on social media are used by different people to spread information. Many negative comments or news, such as hate speech, defamation, and fake news. The article analyses the problems faced by the spread of hate speech through social media from three aspects. The article begins by analysing the impact of hate speech and defamation on people. This is followed by a discussion of the impact of fake news from social media on people. Finally, discussion of social media companies vetting the UGC and news content posted on that platform.*

Keywords: *Social media, Hate speech, UGC, Fake news*

1. Introduction

Social media is becoming an important part of people's daily lives. People spend a lot of time on social media, relying on it for entertainment, information search, reading news or simply to spend time [1]. Social media has exploded in the last decade, providing a way for the public to receive information differently from traditional media, and it allows people to participate in the process of news communication [2]. The cheap, accessible and fast dissemination of social media makes people more likely to search and consume news in social media [3]. Social media has made readers also publishers, but the authenticity of information on social media is increasingly being questioned. During COVID-19, about half of the news readers said that the news they read was not related to the facts, and only 28% said that the news content was probably close to the facts [4]. The negative impact of social media features is gradually appearing and the regulation of news and information by social media companies is gaining attention. This article focuses on the negative effects of hate speech, defamation and fake news on people and society, and analyses the current regulatory difficulties encountered by social media companies and their future development.

2. Problems caused by harmful speech spread by social media

Social media has become widely used around the world and has made it easy for people worldwide to express their views. However, the thoughts and opinions of users are not always optimistic and can be undesirable, harmful and may even constitute bullying, offensive content and hate speech. Many governments are increasingly recognising that hate speech is a serious problem and that it is particularly difficult to stop the spread of hatred between countries and minorities on the internet [5]. Not only hate speech, but also defamation can have a significant impact on people. In comparison to traditional media, the anonymous nature of social media has gradually increased the risk of individuals being defamed. Defamation cases in traditional media are relatively rare because traditional media organizations censor their content, usually exercising editorial control and identifying the source and publisher of the information. The impact of misinformation on the public caused by social media when spreading fake news is irreversible in a short period of time. The problems caused by harmful statements spread by social media will affect the public's judgement and the policing of society.

3. Strategies to address the problems

3.1. Strengthening censorship UGC for hate speech

The anonymity, rapidity, and circulating nature of the Internet place people in an environment beyond the reach of traditional law enforcement, making it easy to harass and express hate and becoming a tool for extremists and hate mongers to promote hatred [6]. Although hate speech did not emerge with the rise of social media or the Internet, hate speech existed long before the development of the Internet and social networks. However, the advent of the internet and the subsequent rise of social networks facilitated the spread of hate speech, providing space for this already complex discourse to spread [7]. However, many people use social media to get their news. According to the Pew Research Center, more than half of adults in the United States get their news through social media, even though the content of "news" on social media often contains false and misleading stories [8]. Hate speech is defined as incitement to encourage hatred against a group of people who wish to destroy the target group by harming them. Target groups are differentiated by race, ethnicity, gender, religion, and sexual orientation [9]. Many social media companies have built their business models on attracting attention, and offensive language and behavior often attract attention, with hate speech being more visible to users on social media than on traditional mass media. Regulating harmful speech in social media requires a clear distinction between legitimate freedom of expression and illegal hate speech, which cannot be protected by freedom of expression.

Hate speech stereotypes or prejudices against minorities, and the negative sentiment and negative impact of hate speech have increased over time, e.g., In 2015 messages emerged on Facebook of Palestinian extremists openly recruiting and training terrorists, calling for the murder of Israeli Jews. The negative impact has been extremely bad, causing mental and physical harm to Israeli Jews there [10]. Not only that, but there is a lot of hate speech on Twitter, but at the same time, people who post hate speech suffer specific penalties. Racist comments have been allowed on Twitter, with Liam Stacey tweeting racist comments in 2012 after a Premier League player suffered a cardiac arrest, and Declan McCuish making racist comments about two Rangers players on Twitter in 2014 [11]. While both were punished by the law in the end Liam Stacey was sentenced to 56 days in prison and Declan McCuish was jailed for a year. Such legal punishments will always lag, with the law only coming to restrain users after hate speech has already been spread and impacted. Legal constraints are part of the equation to reduce the impact of hate speech; social media companies should also be vetting the content posted by users. With the widespread use of well-known social media such as Facebook and Twitter, there has been a gradual increase in people using social media to make statements, especially those with influence and visibility. The words they make are more likely to be disseminated and discussed. Social media companies are responsible for publishing user-generated content and should review the user-generated content posted on their platforms.

Social media is a platform for disseminating many messages, and controlling the spread of hate speech management and censorship at the source can be very practical and more effective in maintaining the social environment. Many developed countries have laws against hate speech, and those convicted of breaking the law often face significant fines or even imprisonment. These laws have prompted social media and websites to develop regulations against hate speech. However, social media sites face severe difficulties identifying and censoring questionable posts. This is a difficult task due to their size and the inability of social media sites to block or edit all hate speech users [12]. Not only this, but the management system of social media companies is highly confusing and the lack of clarity in the platform's policies, procedures and values lead to significant differences in the interpretation of user experience on the same site. Content control in a web environment consisting of vague rules and opaque procedures is very difficult. However, different people's perceptions of platforms for content management differ. Governments and other actors say that there should be stricter controls on certain speech, while other members of society demand that platforms increase their online freedom of expression [13]. Social media platforms face many difficulties in managing the content posted on their platforms, both those posed by the nature of the platform itself and those posed by human beings.

3.2. Regulating UGC for defamation

Social media companies should review user-generated content posted on their platforms, and the features of social media allow defamatory statements to spread faster than ever before.The anonymity that characterizes social media allows users to speak freely and without restriction on social media. The anonymity of social media is widely recognised as a feature that enhances the freedom of communication, especially online. This feature has led to many cases of defamation on social media. However, the rapid dissemination and anonymity of social media make the legal issues associated with social media more complex than traditional media.. Because of the spread of information on the internet, particularly on social media, it is easy for information to cross borders; however, different legal issues can arise if uninformed users disseminate defamatory statements [14]. There are many cases of defamation through social media today. For example, two days after the wife of John Bercow, the Speaker of the British House of Commons, wrongly linked a "Tory leader" to sexual assault allegations on BBC Newsnight, a tweet by Sally Bercow about Tory MP Lord McAlpine's tweet was defamatory. Ms. Bercow said she had learned "the hard way" that comments can sometimes be "found to be grossly defamatory, even if you don't mean to be defamatory and no clear allegation is made" [15].

Many people in social media defamation cases are unaware that their actions will cause deformation and that the tweets they publish will cause distress to the person concerned. The fact that everyone on social media is free to express themselves as they please contributes to a certain extent to the occurrence of defamation. It is important to note that neither malicious slander nor unintentional defamation has anything to do with freedom of expression. Freedom of expression is defined as the right to freedom of expression, which includes holding opinions and receiving and imparting information and ideas without interference from public authorities and regardless of national boundaries. These freedoms are exercised because they carry with them duties and responsibilities. It cannot be used to protect against defamation and hate speech. Social media companies should review the UGC content posted on their platforms and moderate inaccurate statements and tweets that relate to the privacy of others and those with hate speech.

3.3. Avoiding fake news content

The inherent fast-spreading nature of social media makes the screening of news lax. Most people believe in their first impressions and once influenced by fake news spread on social media, the user's opinion of something is difficult to be changed in a short period. This is why social media companies should regulate the content of news posted on their platforms, not only in terms of content but also in distribution. Fake news, like hate speech, is not a new concept. Fake news emerged before the rise of the internet, and the rise of the internet and social media has accelerated the spread of fake news. This is coupled with publishers using false and misleading information to amplify their interests in order to gain attention [16]. Fake news is defined as news articles that are deliberate, verifiable, and likely to mislead readers [17]. However, the reason for fake news on social media is the problematic vetting of content posted by platforms and the poor regulation of the creation of social media accounts. The low cost of creating social media accounts has also encouraged the emergence of malicious user accounts such as social robots and trolls. A social robot is a social media account controlled by a computer algorithm that automatically generates the content and interacts with humans (or other bot users) [3]. For example, Social bots massively distorted online discussions of the 2016 US presidential election. In the run-up to Election Day, a large number of bot accounts tweeted in support of Trump or Clinton. And there were plenty of trolls to disrupt the internet landscape. Trolls, that is, aim to provide an internet environment for spreading fake news on social media by disrupting the order of the online environment with the aim of generating emotions in real users [3].

Fake news takes advantage of the characteristics of social media and is widely spread on social platforms. The spread of fake news can bring a lot of adverse effects, and social media has a particular responsibility for the spread of fake news. Social media should set up a strict regulatory system and a regulatory organization to monitor fake news on social media to prevent it from being exploited for political and financial gains. The underlying reason for the spread of fake news is not the fake news itself, which, as mentioned earlier in the article, has been around for a long time and has been spread due to the rapid spread and lack of censorship of social media. The greatest impact of fake news is not the false reporting of an event, but the greater harm that can be caused by widespread sharing and distribution through social media or the internet. The use of social bot accounts in social media drives the sharing of fake news and further exacerbates the spread of fake news. Ultimately leading to this content being shared automatically making it difficult for technical staff to detect [18]. Technology companies on social media platforms use many different policies to conduct relevant content reviews. Two of the most prominent aspects are terms of service and community guidelines [19]. This is how social media companies constrain users' content posting. However, content moderation is often described as a form of limiting the user's voice; it uses an algorithm of keywords to remove textual content posted by users and block them from accessing channels in the event of account suspension. However, social media companies face the difficulty of censoring the content of news in the face of fake news/ while also vetting the users who create their accounts.

4. Conclusion

In conclusion, social media's anonymity, immediacy, and global nature make it easy for hate speech, defamation, and fake news to be disseminated. Social media companies should exercise a degree of censorship over the content published on their platforms. However, the difficulties faced in supervising content posted on social media are indeed difficult to resolve in a short time. For example, there is limited way to restrict content such as defamation legally, and the platform faces dilemma to stop other users from re-posting content by blocking accounts that post hate speech. This research proposes that social media platforms use algorithms and other screening techniques to censor the content posted by users containing sensitive words to curb the spread of negative information. In addition, the relevant legal system should be strengthened to protect the victims and punish the purveyors. However, the scale of the law and the limits of speech need to be negotiated diligently by all parties in order to achieve positive results.

References

[1] Giannakos, M. N., Chorianopoulos, K., Giotopoulos, K., & Vlamos, P. (2013). Using facebook out of habit. Behaviour & Information Technology, 32(6), 594-602.

[2] Rutsaert, P., Regan, Á., Pieniak, Z., McConnon, Á., Moss, A., Wall, P., & Verbeke, W. (2013). The use of social media in food risk and benefit communication. Trends in Food Science & Technology, 30(1), 84-91.

[3] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

[4] Nielsen, R. K., Kalogeropoulos, A., & Fletcher, R. (2020). Social media very widely used, but use for news and information about Covid-19 is declining. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=3708405

[5] Mondal, M., Silva, L. A., & Benevenuto, F. (2017, July 4-7). A measurement study of hate speech in social media Proceedings of the 28th ACM Conference on Hypertext and Social Media, Prague, Czech Republic. https://doi.org/10.1145/3078714.3078723

[6] Banks, J. (2010). Regulating hate speech online. International Review of Law, Computers & Technology, 24(3), 233-239.

[7] Alkiviadou, N. (2019). Hate speech on social media networks: Towards a regulatory framework? Information & Communications Technology Law, 28(1), 19-35.

[8] Ott, B. L. (2017). The age of twitter: Donald j. Trump and the politics of debasement. Critical studies in media communication, 34(1), 59-68.

[9] Herz, M., & Molnár, P. (2012). The content and context of hate speech: Rethinking regulation and responses. Cambridge University Press.

[10] Guiora, A., & Park, E. A. (2017). Hate speech on social media. Philosophia, 45, 957-971.

[11] Burnap, P., & Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. Policy & Internet, 7(2), 223-242.

[12] Şahi, H., Kılıç, Y., & Sağlam, R. B. (2018, 20-23 Sept. 2018). Automated detection of hate speech towards woman on twitter. 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, Bosnia and Herzegovina.

[13] Roberts, S. T. (2018). Digital detritus:'Error'and the logic of opacity in social media content moderation. First Monday, 23(3).

[14] Mills, A. (2017). Choice of law in defamation and the regulation of free speech on social media: Nineteenth century law meets twenty-first century problems. Edward Elgar Publishing.

[15] Sherwin, A. (2013). Twitter libel: Sally bercow says she has 'learned the hard way' as she settles with tory peer lord mcalpine over libellous tweet. https://www.independent.co.uk/news/uk/crime/twitter-libel-sally-bercow-says-she-has-learned-the-hard-way-as-she-settles-with-tory-peer-lord-mcalpine-over-libellous-tweet-8630653.html

[16] Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. Procedia Computer Science, 141, 215-222.

[17] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211-236.

[18] Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2021). Trends in combating fake news on social media–a survey. Journal of Information and Telecommunication, 5(2), 247-266.

[19] Myers West, S. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. New Media & Society, 20(11), 4366-4383.