Original Paper

Ethical Thoughts of the Regulation of False Information in Social Media: Based on Legitimacy and Social Justice

Chien-wei Kung[1*]

[1] School of Communication, Hong Kong Baptist University, Hong Kong SAR, China

[*] Chien-wei Kung, School of Communication, Hong Kong Baptist University, Hong Kong SAR, China

Abstract

It is now common for social media to regulate rumors. They claim that the purpose of this action is to safeguard social interests. However, some cases have proved that the regulation of rumor has exceeded the necessary limit, but also showed the partiality of the regulation object and the irrationality of the "rumor" standard. Although freedom of speech has boundaries, the regulation of social media is much stricter than it, which hurts social media to play its role as a public sphere. We can't ask individuals to take too much responsibility for rumor spreading. At the same time, we can't easily take harsh regulatory measures such as deplatforming and even legal sanctions against individuals, because this will lead to the lack of legitimacy of the regulatory behavior of social media and the aftereffect of injustice.

Keywords

social media, false information, regulation, ethics, justice, legitimacy

1. Introduction

Nowadays, social media and social media regulation coexist in the world. Although the control methods may be different in different places, it seems that control has long been a "matter of course". There are various reasons for regulation, but in most cases, "cracking down on rumors" is a legitimate reason. But is it morally blameless to regulate social media in the name of "Combating rumors"? I'm afraid not.

2. Cases and Problems

Firstly, we use a case to lead to the problem of this paper. In December 2017, People's Daily, China Daily, Global Times, and other media sent a piece of news on weibo and other social media accounts,

entitled "a 6-year-old computer programming student who has been admitted to MIT at the age of 14", "A 14-year-old middle school student has signed an enrollment agreement with MIT. He said he wants to go to the United States to learn a top technology and go home to serve his motherland."

At first glance, it seems that this is just ordinary news about "gifted youth", but facts have proved that this is completely false news, and its content is meaningless from beginning to end. Not only that, but the people around the middle school student also showed some "invitation letters from Tsinghua University, South China University of Technology and Southern University of Science and Technology" to prove that he is a genius. However, the evidence is also full of mistakes: the format is wrong, the content is wrong, Chinese and English are mixed, and even the signer under the email is William Jefferson Clinton—the full name of former US President Bill Clinton.

As we all know, "don't believe rumors and don't spread rumors" is a very popular sentence on China's Internet in recent years. Many Internet opinion leaders and media also call themselves "rumor refuters" and are keen to "rumor refutation" in history, current politics, and many other aspects. However, in this kind of social news that most people are keen on, they collectively act as "rumor mongers", and spread such obvious and low-level rumors. After spreading rumors, the accounts of these official media did not (in fact, it is impossible) accept any punishment, which is very different from the consequences of spreading rumors by ordinary people. Because of this, the question of this paper is ready to come out: is it ethically justified to regulate social media in the name of "cracking down on rumors", and why do "rumor crackers" can take no responsibility for their rumor spreading behavior, and what deep-seated problems does this reflect?

3. Theoretical Framework of Social Media Regulation

3.1 "Gatekeeper" and Self-Censorship

The regulation of social media is a kind of "gateway", that is, a special "door guard" decides to "enter or exit" (Lewin, 1947). When it comes to social media, the gatekeeper should be the platform management of social media. However, because of the substantial non-independence of social media and the influence of political factors, some official or semi-official institutions, such as government agencies or some major media, can usually also play the role of gatekeeper. Even if these institutions do not issue clear instructions to social media platforms, social media managers may also make censorship based on the preferences of these institutions. But unlike external gate-keeping, when regulation exists for a certain period, people tend to act as gatekeepers themselves before they speak, i.e., to self-censor. Whether it is individuals in general, public figures, or the news media, if their statements are blocked or deleted due to censorship, it is difficult for them to achieve the purpose of disseminating information. Therefore, self-censorship to ensure that they pass the "gate keeping" process has become a conscious and unconscious action taken by the majority of social media users. Of course, the purpose of self-censorship is likely not just to pass the social media gatekeeper; on the one hand, in some areas, the consequences of some speeches are not just punished by social media. This

2

potential legal deterrent will induce them to conduct self-examination with higher standards and even make a lot of false statements (Robinson & Tannenberg, 2018; Shen & Truex, 2021); on the other hand, psychological factors such as fear of isolation (i.e., the spiral of silence) (Bar-Tal, 2017), economic factors (Gray & BISHOP KENDZIA, 2009), etc., may also lead to self-censorship, so the fear of being punished by the platform or furthermore by other institutions is not the only factor of self-censorship. However, in the context of this paper, we believe that the contribution of gatekeeping to self-censorship is evident, as evidenced by the relevant studies cited earlier.

3.2 The Limits and Standards of Freedom of Speech

When it comes to regulating the media, it is necessary to mention freedom of expression. Take the United States as an example. The legal root of freedom of speech in the United States is the first amendment to the Constitution: "Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech." The above-mentioned freedom of speech includes the freedom of "what to say" and "what not to say", but this does not mean that freedom of speech in the United States has no boundaries. For example, speech and publishing activities that infringe intellectual property rights, threaten the president of the United States, or incite imminent violence are not protected by the first amendment (Volokh, 2008).

The example of the United States shows that freedom of speech must have boundaries, which seems to provide legitimacy for social media to regulate rumors. However, we still need to pay attention to one problem: most of the unprotected speech stipulated by the United States has already belonged to the level of "felony", and its social harmfulness is also obvious. However, in contrast, many rumors regulated by social media are not harmful at all - many are not even "rumors" at all. Moreover, social media tends to discriminate when regulating, that is, they prefer to protect social media accounts with official background, which is clearly reflected in the cases at the beginning of the article and will be discussed in detail later. In short, the "rumor regulation" standard of social media is much higher than the "due standard of freedom of speech" in the general sense, and the object is selected. Therefore, the conflict between regulation and freedom of speech is still worth discussing.

3.3 Social Media as a Public Sphere

Habermas put forward the concept of the "public sphere", that is, a public communication space for organizing, disseminating, discussing, and negotiating the information required for democratic operation in society. This is a virtual structure formed by information that can promote democratic politics. The media is not a natural public sphere, in countries such as North Korea and Saudi Arabia, the rulers exercise absolute power over the media, and any content criticizing the ruling elite can not be seen in the newspapers. Only when the media can act as the "fourth right", that is, exercise the political right of public supervision on behalf of its readers, can the media constitute the "public sphere" in the sense of Habermas (Schulz & Cobley, 2018).

In *A Theory of Justice*, Rawls proposed the principle of justice based on the "veil of ignorance" and regarded them as the realization path of social justice, and the media should act accordingly (Schulz &

Cobley, 2018). Rawls believes that if we know nothing about our situation (that is, the veil of ignorance), we are bound to seek what he calls "justice as fairness"—that is, an egalitarian justice that only tilts towards the most disadvantaged. Because under the veil of ignorance, each of us may worry that we will suffer from all kinds of systematic discrimination. Therefore, only fair and targeted preference can solve the problem. So is the public sphere media or The North Korean media more conducive to the realization of social justice? The answer is obvious. The North Korean media does not have any autonomy, but as a medium in the public sphere, it can speak for the interests of everyone to the greatest extent—including all kinds of vulnerable groups. To sum up, based on the concept of the public sphere and justice theory, the media should strive to become the public sphere to promote the realization of social justice, which is the direction of its ethical legitimacy.

The theoretical framework has been elaborated so far. The following will make a comprehensive analysis of the rumor regulation of social media based on the above theoretical framework.

4. Why is the Legitimacy of Social Media Regulation in Doubt

4.1 Ethical Query on Excessive Regulation

Different from traditional newspapers and other media, the particularity of social media is that almost everyone can speak, and individuals can compete with the largest media in a country. This relative equality enables social media to achieve its mission as a public sphere with the depth and breadth that traditional media do not have, especially in many aspects such as supervision and public opinion feedback. Social media can usually reflect many problems that traditional media can't do with high timeliness and communication speed. But obviously, this public domain attribute is being hurt by excessive regulation, and the degree is gradually deepening.

Social media has established the moral legitimacy of regulation on the grounds of preventing social harm. However, as we have noticed when discussing freedom of speech, "the boundary of freedom of speech" is an excellent excuse, but the regulatory action of social media has far exceeded the necessary boundary. Further, the regulation of social media is fundamentally contrary to "public". Although social media usually lists many rules most of these guidelines are unclear and vague, and the resulting self-examination is more significant. Over time, although social media will not completely lose its uniqueness, it can no longer fulfill its duty—that is, the voice channel of ordinary people. This fundamentally obliterates the legitimacy of social media and is not conducive to the realization of social justice under Rawls' framework. Especially considering the problem based on Rawls' justice principle of "the most unfavorable is the most favorable", ordinary people are undoubtedly in the "most unfavorable" position in the right to speak. The excessive regulation of social media makes them silent and unable to speak even if they have the intention, which is untenable in ethics.

Overall, on the one hand, excessive regulation hurts the public domain attribute of social media itself, on the other hand, it almost strangles the voice right of ordinary people, which are two reasons why the current social media regulation should be criticized ethically.

4

4.2 Regulator's Lies

We can go back to the case to further think about the problem of "regulation" itself. As mentioned earlier, the current regulation of social media is selective. Individuals are in the most disadvantageous position in the regulation, and various official—or "mainstream" accounts will be protected. There were many loopholes in the false information in the case, but many official accounts turned a blind eye to it. It was not until a few days after many netizens questioned it that the false information was deleted by some official accounts.

This is enough to make us doubt what the rumor regulation of social media is for. As we all know, in many cases, social platforms prevent and control rumors of some negative news quite quickly. For example, social media will respond within an hour when some negative social events begin to spread. However, although the rumors in the case at the beginning of this paper spread widely, they were not regulated by social media. As of December 14, 2021, this false news is still kept in the weibo of various multimedia such as the northern morning post. In ordinary times, many "rumors" will be deleted and even further investigated for responsibility without factual verification. However, it turns out that this information is not a rumor. Some official organizations will choose to declare that the real news is a rumor, and their behavior is a strict rumor. For example, in the dismissal of nurses from Xianyang Maternal and Child Health Hospital, the hospital directly claims that this "false report" is related to "hostile forces" without showing any evidence. Similarly, the hospital of the incident, like many official media in the case at the beginning of this article, was not punished, although they made quite severe and unfounded charges against the whistleblowers of the incident.

The above events jointly prove a view already put forward above, that is, the "rumor regulation" is not as legitimate as they claim. The text of the rules and regulations is vague and "dumb" when encountering specific objects. This is enough to prove that the ostensible reason for rumor regulation, that is, the platform statute and its declared "social interests", are just excuses. On the contrary, regulation is more often just to serve the "mainstream" accounts and the people behind them, which is the true face of "rumor regulation". It is undeniable that many regulated rumors may be rumors, but why do unregulated rumors slip through the net? Why are ordinary people likely to stop their accounts and even be punished by law when they spread rumors, and these "mainstream" accounts are safe? Let alone those accounts that claim to be "refuting rumors" but are actually "fabricating rumors", why should they not be responsible for their actions at all? These are questions that the current social media rumor regulation is unwilling to answer and impossible to answer.

With the gatekeeper theory, this is perfectly normal. Why do they conduct self-censorship because of regulation? It is because they are aware of the possibility of punishment for their statements, not because of the so-called "social interest" in censoring whether their statements are "rumors" that are being gated. In other words, they are aware that social media gate-keeping is not essential for the social good, nor is it really about fighting rumors, but rather is a pretext for some other purpose. From this point of view, they may even be able to exercise different levels of self-censorship on different social

5

media. If social media were to regulate rumors purely in the social interest, why would they need to "tailor" their self-censorship standards for different social media?

Meanwhile, the impure motivation of regulation can also explain why these "mainstream" accounts can be exempted from punishment. Otherwise, these "mainstream" accounts cannot go unpunished. But in fact, regulation exists for these "mainstream" accounts, and "mainstream" accounts can even be said to be the actual controllers of regulation to a considerable extent, so they can "spread rumors" almost irresponsibly.

4.3 Personal Responsibility

So far, the negative impact of social media regulation and the impure purpose of regulation have been obvious. But some people may question it. After all, rumor regulation has indeed played a role in combating rumors. Don't individuals have to be responsible for their rumor-spreading behavior at all? Of course, we do not think that all rumors should not be held responsible. After all, freedom of speech can not be unlimited. But in many cases, asking for personal responsibility is just empty talk. Unlike professional news media, it is almost impossible for individuals to fulfill the obligation of fact verification when disseminating information, which is determined by personal limitations. If an individual has to check the facts carefully, the only appropriate result is to keep them quiet. If you are strict about "not spreading rumors", you can't even be the microphone of the mainstream media. After all, even the mainstream media may make mistakes. Therefore, all punishments beyond "deletion" are inappropriate unless the parties deliberately and try to harm society with rumors.

We can use a hypothesis similar to the "veil of ignorance" to provide another basis for the weakening of personal responsibility. If we do not directly prohibit the public from speaking but only strictly investigate the responsibility for public rumors, everyone will be afraid. After all, individuals cannot carefully verify the facts. It is difficult for you to know whether the information you reprint is true. In other words, when we reprint the news, each of us is actually in a small "veil of ignorance" so we are bound to ask for a lighter punishment for our possible rumors. If not, one day everyone will tacitly conduct strict self-examination and even shut up, thus falling into the strange scene of "no voice is prohibited, but no voice".

In addition, social media should defend the voice of almost everyone. Most traditional media have their positions and prejudices on many issues such as politics, some algorithms have been able to identify this bias (Hamborg et al., 2019). However, no matter what kind of prejudice they have, the opinions of the mainstream media are mostly within a certain range, that is, "Established Pluralism". This is not to say that there is only one voice in the mainstream media, but most of these voices still have their direction, that is, they are helpers of the authorities, not complete critics (Lee, 2002). We should note that if the media want to fulfill their mission as the public sphere, they should be independent of power (Schulz & Cobley, 2018). However, few media are completely independent of power in the world. This does not mean that the media is controlled by the government, nor does it mean that the media has reached cooperation agreements with some powerful figures. After all, the audience has their positions

and preferences, they will tend to some power figures, and the media is bound to cater to the audience. Of course, the media can certainly have their position and shape public opinion based on it, but the public that these media can not represent is in a state of complete aphasia. At this time, the particularity of social media is revealed: social media provides a voice platform for "outsiders", allowing them to be seen by the world; This determines that social media, as a public sphere, must be more representative and popular than traditional media, and also makes its role in supervising power more significant.

Finally, let's turn our attention to the case and the problem that has been repeatedly emphasized: it is absurd not to require the responsibility of "mainstream" accounts while requiring personal responsibility. After all, they have much more influence than ordinary people. However, there is another difficulty in holding "mainstream" accounts accountable. After all, directly disabling the "mainstream" accounts will cause considerable inconvenience and may even lead to legal disputes. However, at least we can be sure that the punishment of "deletion" should be applied to this obvious rumor. But until today, this rumor of many "mainstream" accounts has not been deleted. Therefore, since the official account has not suffered even the lightest punishment, why can we assert that it is legitimate for individuals to be disabled and even bear legal responsibility?

To sum up, even those who spread rumors should not be required to bear much responsibility if they do not have great harm and subjective malignancy. If not, social media itself will lose its meaning - at least in the public sphere. Although some people may be happy to see such a scene, it does not comply with the due principles of social justice of social media and will make social media lose its "the fourth estate" attribute.

5. Lack of Legitimacy of Social Media Rumor Regulation

The excessive regulation of communication media shows a lack of legitimacy, that is, it is not "a value whereby something or someone is recognized and accepted as right and proper" (Chen, 2016). Although defenders of regulatory behavior will claim that "deplatforming" (i.e., disabling accounts) can effectively prevent the spread of false information (Chandrasekharan, 2017), it is doubtful whether this measure is "necessary" for ordinary social media users. The particularity of social media is that most people use it not to play the role of "media" in the traditional sense, but more to obtain emotional support (Hamburg et al., 2019), establish people to win respect (Chua, 2016), and other needs, which means that "deplatforming" not only curbs users from spreading rumors, It also infringes other rights and interests of users - especially those that have little to do with news communication. Therefore, for ordinary people, the legitimacy of direct "deplatforming" is not widely recognized, which is a matter of course. As mentioned above, the excessive regulation of social media and the self-censorship caused by it does great harm to social justice, even from a non-utilitarian point of view. Even from the perspective of deontology, "deplatforming" also hurts the natural human right of freedom of speech. Although freedom of speech does not mean that pseudoscientists have the freedom of speech in academic journals (i.e., the issue of "time, place and way of freedom of speech"), social media is different. It is

7

closer to the nature of "traditional public forum", so it should only be regulated by "content-neutral time, place and way restrictions" Specifically, such regulatory measures should "leave sufficient channels for information exchange" (O'neill, 1999). Depriving users of their freedom to express other opinions because they publish some content beyond the bottom line is a kind of "prior review", but this is not justified.

Another question of legitimacy of regulation is why "rumors" are "rumors". If we judge a piece of information as a rumor directly based on the statements of stakeholders without combining other evidence, this behavior is not justified at all. Interestingly, this is the most common situation on social media in some regions. This does not mean that stakeholders cannot refute rumors about themselves, but they need evidence to refute rumors. Of course, not all rumors can produce evidence. For example, if someone forges a document and claims that it is a top-secret document of a company, it is difficult for the company to prove it is false information. In this case, it is still necessary to review the authenticity and support of the information itself—but similarly, this operation should be cautious.

Generally speaking, the legitimacy of rumor regulation lies in the injustice of regulation means on the one hand, and the injustice of regulation content on the other hand. As long as these two problems are solved, its legitimacy is bound to be significantly improved. Finally, the mission of social media as the public sphere can be achieved only on the premise of appropriate supervision, and the goal of social justice can be achieved only in this way.

6. Conclusion

We conclude that: social media claim that they regulate rumors for social interests, but the facts have proved that this is not entirely true. If regulation can be based on social interests, the problems of its legitimacy and justice will no longer exist to a great extent. "Public sphere" is always one of the most precious attributes of social media. The abused "deplatforming" and unreasonable "rumor regulation" will only hurt this attribute, but can not play its social role. The content discussed in this paper is not just rumors, but all social media regulations. Today, I think this issue deserves our attention.

References

Bar-Tal, D. (2017). Self-Censorship as a Socio-Political-Psychological Phenomenon: Conception and Research. *Political Psychology*, *38*, 37-65. https://doi.org/10.1111/pops.12391

Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., & Gilbert, E. (2017). You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech. *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 1-22. https://doi.org/10.1145/3134666

Chen, J. (2016). *Useful Complaints: How Petitions Assist Decentralized Authoritarianism in China*. Rowman & Littlefield.

Chua, T. H. H., & Chang, L. (2016). Follow Me and Like My Beautiful Selfies: Singapore Teenage

8

Girls' Engagement in Self-Presentation and Peer Comparison on Social Media. *Computers in Human Behavior*, 55, 190-197. https://doi.org/10.1016/j.chb.2015.09.011

Gray, G. C., & BISHOP KENDZIA, V. (2009). Organizational Self-Censorship: Corporate Sponsorship, Nonprofit Funding, and the Educational Experience. *Canadian Review of Sociology/Revue canadienne de sociologie*, *46*(2), 161-177. https://doi.org/10.1111/j.1755-618X.2009.01209.x

Hamborg, F., Donnay, K., & Gipp, B. (2019). Automated Identification of Media Bias in News Articles: an Interdisciplinary Literature Review. *International Journal on Digital Libraries*, *20*(4), 391-415. https://doi.org/10.1007/s00799-018-0261-y

Lee, C. C. (2002). Established Pluralism: US Elite Media Discourse About China Policy. *Journalism Studies*, *3*(3), 343-357. https://doi.org/10.1080/14616700220145588

Lewin, K. (1947). Frontiers in Group Dynamics: II. Channels of Group Life; Social Planning and Action Research. *Human Relations*, *1*(2), 143-153. https://doi.org/10.1177/001872674700100201

Morahan-Martin, J., & Schumacher, P. (2003). Loneliness and Social Uses of the Internet. *Computers in Human Behavior*, *6*(19), 659-671. https://doi.org/10.1016/S0747-5632(03)00040-2

O'neill, K. F. (1999). Disentangling the Law of Public Protest. *Loy. L. Rev.*, *45*, 411.

Robinson, D., & Tannenberg, M. (2018). Self-Censorship in Authoritarian States: Response bias in measures of popular support in China. *V-Dem Working Paper*, *66*. https://doi.org/10.2139/ssrn.3161915

Schulz, P., & Cobley, P. (Eds.). (2018). *Handbooks of Communication Science*. De Gruyter Mouton. https://doi.org/10.1515/9783110481129-001

Shen, X., & Truex, R. (2021). In Search of Self-Censorship. *British Journal of Political Science*, *51*(4), 1672-1684. https://doi.org/10.1017/S0007123419000735

Volokh, E. (2008). *The First Amendment and Related Statutes: Problems, Cases and Policy Arguments*. Foundation Press.

9