

The Impact of the “Regulations on the Governance of Online Violence Information” on User Behavior in Chinese Social Media Platforms

Ruiqing Zhang¹

¹ South China Normal University, Guangzhou, China

Correspondence: Ruiqing Zhang, South China Normal University, Guangzhou, China.

doi:10.56397/LE.2024.11.04

Abstract

The rise of social media platforms in China, such as Weibo, Douyin, and WeChat, has transformed communication but also led to the proliferation of online violence, including cyberbullying, harassment, and doxxing. These harmful behaviors have serious psychological impacts on victims and disrupt the digital ecosystem by creating toxic environments. In response, the Chinese government implemented the “Regulations on the Governance of Online Violence Information” in 2024. This study examines the effectiveness of these regulations in shaping user behavior and promoting safer online interactions. The regulations impose responsibilities on social media platforms to implement content monitoring, provide reporting mechanisms, and cooperate with law enforcement. Through case studies and user analysis, this paper explores the regulations’ impact on reducing harmful behaviors, fostering self-regulation, and addressing challenges such as technical difficulties in moderation and concerns over freedom of expression. The findings highlight the regulations’ role in improving digital governance and contributing to a healthier online ecosystem.

Keywords: online violence, cyberbullying, social media regulation, content moderation

1. Introduction

The proliferation of social media platforms in China, such as Weibo, Douyin, and WeChat, has fundamentally transformed the ways in which people communicate and share information. These platforms have created dynamic spaces for public discourse, enabling users to engage in conversations, share personal experiences, and access a wealth of information with unprecedented ease. However, alongside these benefits, the rapid growth of social media has also exposed significant societal challenges, one of the most pressing being the rise of online violence.

Online violence encompasses a range of harmful behaviors, including cyberbullying, harassment, doxxing, and hate speech. Cyberbullying involves repetitive and targeted abuse, often designed to humiliate or intimidate individuals. Harassment can take various forms, from persistent messaging to coordinated attacks aimed at overwhelming a target. Doxxing, the public disclosure of private information with the intent to harm, poses severe risks by bridging online violence with real-world consequences. Hate speech, meanwhile, fosters division and incites hostility against specific groups based on race, gender, or other characteristics.

In China, these behaviors have become widespread, leading to toxic environments that deter healthy discourse and inflict serious psychological harm on victims. Numerous high-profile cases have brought the issue to the forefront. For instance, public figures and ordinary users alike have faced waves of coordinated online abuse, sometimes resulting in severe emotional distress, withdrawal from public life, and in extreme cases, self-harm or suicide. The devastating impact of such incidents has spurred widespread public outcry and underscored the urgent need for effective regulatory measures.

In response to these challenges, the Chinese government introduced the “Regulations on the Governance of

Online Violence Information” in 2024. This landmark regulation aims to address the growing issue of online violence by establishing a comprehensive legal framework to govern harmful behaviors on digital platforms. The primary rationale behind the regulation is to protect the dignity, mental well-being, and safety of individuals while ensuring the maintenance of public order in the digital space. By holding both users and platforms accountable for the dissemination and perpetuation of harmful content, the regulation seeks to create a safer and more respectful online environment.

The “Regulations on the Governance of Online Violence Information” outline a series of specific provisions designed to achieve these goals. Social media platforms are mandated to implement robust content monitoring systems capable of detecting and filtering out harmful content in real time. These systems leverage advanced technologies, such as artificial intelligence (AI) and machine learning, to analyze vast amounts of data and identify abusive language or behaviors. Additionally, platforms must provide clear and accessible reporting mechanisms that empower users to flag instances of online violence. Once reported, harmful content must be promptly reviewed and, if necessary, removed to prevent further harm.

Beyond content moderation, the regulation places a strong emphasis on collaboration with law enforcement agencies. Platforms are required to assist in the investigation of severe cases of online violence, such as those involving threats of physical harm or large-scale coordinated attacks. This cooperation ensures that perpetrators of serious online violence can be identified and held accountable under the law.

To enforce compliance, the regulation stipulates a range of penalties for both users and platforms that fail to adhere to its provisions. These penalties include warnings, fines, and in severe cases, suspension of platform services or criminal prosecution. By introducing these deterrents, the regulation aims to foster a culture of accountability and responsibility within the digital ecosystem.

2. Pre-Regulation User Behavior Patterns

Before the implementation of the “Regulations on the Governance of Online Violence Information,” user interactions on Chinese social media platforms were marked by a mix of constructive and harmful dynamics. Social media provided a space for users to share opinions, engage in discussions, and connect with broader communities. However, the rapid pace of online discourse and the often-impulsive nature of user interactions created fertile ground for conflicts and aggressive behaviors. Heated debates on political, social, or entertainment topics frequently escalated into hostile exchanges, overshadowing constructive engagement.

Among the most prevalent forms of online violence were cyberbullying, harassment, and doxxing. Cyberbullying involved repetitive and targeted abuse, often aimed at individuals expressing unpopular or controversial opinions. Public figures, including celebrities and influencers, were frequently targeted, with users banding together to flood their posts with insults and derogatory comments. Harassment extended to private messages, where users could send threatening or abusive content directly. Doxxing, the practice of publicly revealing personal information to intimidate or harm individuals, emerged as a particularly invasive form of online violence, causing real-world consequences such as job loss or physical threats.

A significant factor enabling these harmful behaviors was the anonymity afforded by social media platforms. Users could hide behind pseudonyms or anonymous accounts, reducing the accountability for their actions. This anonymity often emboldened individuals to engage in aggressive behavior they might otherwise avoid in face-to-face interactions. Platform features such as public comment sections, trending topic algorithms, and the ability to amplify posts through shares or reposts further exacerbated the problem. These features, while designed to enhance user engagement, inadvertently provided tools for amplifying harmful content and mob-like behavior. Consequently, without robust regulatory oversight, these dynamics created a challenging online environment, where harmful behavior flourished with limited deterrents.

3. Implementation of the Regulations

The implementation of the “Regulations on the Governance of Online Violence Information” marked a significant shift in the responsibilities of social media platforms in China. Under the new regulatory framework, platforms such as Weibo and Douyin were mandated to take proactive measures to combat online violence. These responsibilities included establishing robust content moderation systems, providing clear and accessible reporting channels for users, and ensuring timely removal of harmful content. Platforms were also required to conduct regular audits of their content to identify and address patterns of abusive behavior. Failure to comply with these responsibilities could result in penalties, including fines, suspension of platform operations, or even criminal liability for platform operators in severe cases.

To detect online violence, platforms implemented advanced technological solutions, including artificial intelligence (AI) and machine learning algorithms. These systems were designed to monitor content in real-time, flagging potentially harmful posts based on keywords, patterns of abusive language, and user reports. Additionally, platforms introduced community guidelines to educate users about acceptable behavior and the

consequences of violating these rules. Reporting mechanisms were enhanced, allowing users to easily report instances of online violence, with options to categorize the type of abuse and provide evidence such as screenshots or links.

Enforcement practices on major platforms like Weibo and Douyin provide clear examples of how the regulations have been operationalized. On Weibo, content flagged as abusive is reviewed by a combination of automated systems and human moderators. In cases of confirmed violations, the offending content is removed, and the user may receive warnings, temporary bans, or permanent account suspension depending on the severity and recurrence of the offense. Douyin, known for its video-sharing format, employs similar measures, including real-time comment filtering and user-reporting options. Additionally, Douyin launched initiatives like “positive energy campaigns” to promote constructive content and reduce the visibility of harmful behaviors.

These enforcement practices illustrate a dual approach of prevention and reaction, aiming not only to penalize offenders but also to create an online culture less conducive to violence. While the regulations have placed significant compliance burdens on platforms, their implementation has been pivotal in addressing the pervasive issue of online violence and fostering a safer digital environment.

4. Changes in User Behavior Post-Regulation

The introduction of the “Regulations on the Governance of Online Violence Information” has led to notable shifts in user behavior on Chinese social media platforms. One of the most immediate changes observed was in user posting habits and engagement levels. Users became more cautious about the content they shared, particularly when discussing sensitive or controversial topics. The fear of being flagged for violating the regulations or facing potential penalties, such as account suspension, resulted in a decline in openly hostile or inflammatory posts. This shift contributed to a more moderated tone in online discussions, with users increasingly focusing on constructive or neutral engagement.

An important aspect of this behavioral change was the rise in self-regulation among users. Many began to exercise greater restraint in their online interactions, avoiding language or actions that could be perceived as abusive. This self-regulation extended to content creation, with users reviewing their posts more carefully before publication to ensure compliance with platform guidelines. Additionally, social media influencers and public figures, who often wield significant influence over their audiences, started emphasizing responsible online behavior, further promoting a culture of self-regulation.

The regulations also spurred platforms to enhance their content moderation efforts, creating a more structured and transparent environment for users. For instance, automated moderation tools became more efficient in detecting and filtering harmful content before it reached a wider audience. This, combined with human moderation, created a robust safety net, reducing the prevalence of online violence.

Case studies highlight these behavioral shifts in action. One prominent example is a series of high-profile online harassment incidents involving public figures on Weibo. Prior to the regulations, such incidents often escalated quickly, with little intervention. However, after the regulations were enforced, platforms responded swiftly by removing harmful content, penalizing offenders, and publicly reinforcing their commitment to maintaining a respectful online space. In another instance, Douyin implemented real-time comment filtering during live streams, significantly reducing abusive interactions and fostering a more supportive environment for content creators.

These changes illustrate the regulations’ effectiveness in curbing harmful behaviors and encouraging a more respectful online culture. While some users have expressed concerns about potential over-regulation, the overall impact has been a positive step toward mitigating online violence and promoting healthier digital interactions.

5. Public Perception and User Reactions

The “Regulations on the Governance of Online Violence Information” have elicited varied responses from the public, shaped largely by their awareness and understanding of the new legal framework. Many users on platforms like Weibo and Douyin exhibit a moderate to high level of awareness regarding the regulations, often discussing their implications in online forums. However, understanding of the specifics varies, with some users displaying a thorough grasp of the rules and their objectives, while others possess only a general sense of their existence and purpose. Awareness campaigns run by platforms, including pop-up notifications and tutorials, have played a key role in educating users about their responsibilities and the mechanisms available for reporting abuse.

Public opinion on the regulations tends to revolve around the delicate balance between effective governance and the preservation of freedom of expression. A significant portion of users appreciates the regulations as a necessary measure to combat the rampant online violence that plagued social media prior to their implementation. They view the regulations as a step towards fostering a more respectful and inclusive digital

environment. However, concerns have also emerged regarding the potential for overreach. Some users fear that the broad scope of the regulations might suppress legitimate criticism or dissenting opinions, thereby stifling open discourse.

Trust in the fairness and effectiveness of enforcement has been another critical dimension of public perception. Many users have expressed confidence in the regulations' ability to curb harmful behavior, citing the visible decline in abusive content and swifter responses from platforms. Nevertheless, skepticism persists among certain groups who perceive enforcement as inconsistent or biased. For instance, there are allegations that influential users or entities receive lenient treatment compared to ordinary users, raising questions about the impartiality of moderation practices.

Overall, while the regulations are generally seen as a positive development, ongoing efforts are needed to ensure their enforcement is perceived as both fair and transparent. Strengthening public trust will require continuous engagement with users, clear communication of enforcement outcomes, and regular reviews of the regulations to address evolving public concerns and maintain a balance between governance and freedom of expression.

6. Challenges in Enforcement

The enforcement of the "Regulations on the Governance of Online Violence Information" has encountered several significant challenges, particularly in the areas of content identification, regulatory overreach, and user resistance. These challenges highlight the complexities involved in maintaining a safe online environment while respecting user rights.

(1) Technical Difficulties in Identifying and Moderating Harmful Content

One of the primary challenges lies in the technical difficulties associated with identifying and moderating harmful content. Although platforms like Weibo and Douyin employ advanced artificial intelligence (AI) and machine learning algorithms to detect abusive language and behaviors, these systems are not foolproof. Harmful content often evolves in ways that make it difficult for automated systems to keep up. For example, users may use coded language, slang, or emojis to bypass moderation filters. Additionally, context plays a crucial role in determining whether certain content is harmful, and AI systems frequently struggle to interpret nuanced interactions. This can lead to false positives, where benign content is flagged, and false negatives, where harmful content goes undetected.

(2) Potential for Misuse or Overreach in Content Regulation

Another significant challenge is the potential for misuse or overreach in content regulation. While the regulations aim to protect users from online violence, the broad scope of the rules can sometimes result in the suppression of legitimate expression. Content that critiques government policies or addresses sensitive social issues may be removed under the guise of preventing online violence, raising concerns about censorship. This risk of overreach has led to public debates about the fine line between safeguarding users and maintaining freedom of expression. Additionally, there is the potential for misuse by individuals or groups who may exploit reporting mechanisms to silence dissenting voices or target specific users unjustly.

(3) Resistance from Users and Limitations in Addressing Subtle Forms of Online Violence

Resistance from users also poses a challenge to enforcement. Some users view the regulations as overly restrictive and express frustration at what they perceive as the loss of their ability to engage freely on social media platforms. This resistance can manifest in non-compliance, such as creating new accounts after being banned, or in finding creative ways to continue harmful behaviors without triggering moderation systems. Furthermore, subtle forms of online violence, such as passive-aggressive comments, exclusionary tactics, and indirect harassment, are particularly difficult to address. These behaviors often fall into grey areas that are harder to categorize and penalize, allowing them to persist despite regulatory efforts.

References

- Chen, X., & Li, J., (2020). Online behavior and regulatory challenges: The rise of cyber violence in China. *Asian Journal of Law and Society*, 7(2), 134-152.
- Feng, Y., & Zhao, H., (2019). Governance in the digital age: Examining the role of legal frameworks in curbing online violence in China. *China Law Review*, 16(1), 89-102.
- Huang, L., (2021). The intersection of censorship and digital freedom: An analysis of China's regulatory approach to social media. *Journal of Digital Policy and Regulation*, 9(4), 45-67.
- Huang, Y., Qin, Y., & Wang, Y., (2023). Public opinion misconduct and online gender violence: Exploring China's social media crisis beneath the surface. *Chinese Human Rights Review*, 19.
- Lee, J., (2013). Regulating blogging and microblogging in China. *Journal of Internet Law*.

- Noh, K., & Wang, B., (2020). A study on the control and improvement of cyber violence in China. *East Asia Law Journal*, 32, 237.
- Ong, R., (2017). Confronting online social aggression in Hong Kong: A wake-up call. *International Journal of Law and Psychiatry*, 54, 19-30.
- Tang, Z., (2022). Social media moderation and the impact of new governance laws: A case study of Douyin. *Global Media Journal*, 14(3), 203-219.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).