

# A Study on the Constitutional Balancing Mechanism Between Freedom of Speech and Content Moderation on Digital Platforms

Yuetong Xu <sup>1,\*</sup>

<sup>1</sup> Hanhong College, Southwest University, Chongqing, 400700, China

\* Correspondence author: YuetongXu2025@163.com

**Abstract:** The ever-expanding power of digital platforms to moderate content has placed the constitutional protection of freedom of speech in a structural predicament. Using constitutional balancing theory as an analytical framework and employing a combination of normative analysis, comparative research, and case studies, this study systematically examines the source of power, operational logic, and impact pathways of digital platform content moderation on freedom of speech. The study finds that the legitimacy of platforms' content moderation powers remains ambiguous. Currently, constitutional norms are insufficient to address speech management by private entities, and there is a lack of stable institutional coordination between platform self-regulatory mechanisms and external oversight. Based on this, this paper proposes a balancing mechanism centered on platform self-regulation—the “moderate-constitutional-horizontal effect extension”—which advocates for institutional arrangements such as improving procedural safeguards, strengthening transparency disclosure obligations, and introducing independent review bodies to promote a dynamic balance between the protection of freedom of speech and platform content governance. The study demonstrates that the effective operation of constitutional balancing mechanisms in the digital age depends on redefining the constitutional status of platforms and establishing institutional solutions that create a structural tension between private law autonomy and public law regulation, thereby providing a normative foundation for the sustained protection of digital expression rights.

**Keywords:** content moderation; constitutional balancing theory; institutional coordination; proportionality-constitutional-horizontal effect extension

## 1. Introduction

In the digital age, digital platforms have become vital venues for public discourse, information dissemination, and the exchange of views. They have broken down the barriers of time and space, enabling information to spread rapidly [1–2]. People can freely express their views on hot-button social issues on various platforms such as Weibo, WeChat, and online forums; whether discussing policies, following public affairs, or evaluating cultural phenomena, their perspectives can be instantly shared with a vast audience [3–6]. This freedom of speech has sparked public enthusiasm for participation and fostered the exchange and collision of ideas. However, the widespread adoption of digital platforms has also given rise to issues such as misinformation, malicious speech, and privacy breaches, making it increasingly important to strike a balance between protecting freedom of speech on digital platforms and content moderation [7–9].

Content moderation refers to technical and administrative measures—such as inspection, filtering, and deletion—applied to information on digital platforms. It is a process of monitoring and managing platform content throughout the information dissemination cycle, aimed at safeguarding national security, protecting the public interest, maintaining social order, and preventing the internet from



---

becoming a breeding ground for illegal activities and crime [10–14]. There exists a delicate relationship between freedom of speech on digital platforms and content moderation, requiring a balance of interests between the two [15]. To protect freedom of speech, the government should allow citizens to freely and lawfully express their views. At the same time, society should adopt an inclusive and open attitude toward differing opinions. However, when speech on digital platforms infringes upon the rights of others, spreads misinformation, or encourages illegal activities, content moderation must be enforced [16–17]. To balance freedom of speech and content moderation, governments and relevant agencies should implement a series of measures, including establishing a robust legal framework that clearly defines prohibited speech and specifies penalties for the dissemination of misinformation and malicious content [18–20]. Secondly, they should strengthen oversight of digital platforms, requiring them to review user-posted information and remove or ban content that violates regulations [21–22].

As a theoretical framework, the theory of balance aligns well with the traditional Chinese ethos of moderation and harmony. Applying it as an analytical tool to examine the issues of freedom of speech and content moderation on China’s digital platforms represents a valuable endeavor that can effectively advance the governance of these platforms. This study employs constitutional balancing theory as its analytical framework and adopts a methodology combining normative analysis, comparative research, and case studies to systematically examine the source of authority, operational logic, and impact pathways of content moderation on freedom of speech in digital platforms. Concurrently, it proposes a balancing mechanism centered on platform self-regulation—the “moderation-constitutional-horizontal effect extension”—and introduces an independent review body to promote a dynamic balance between the protection of freedom of speech and platform content governance.

## 2. Related Concepts

### 2.1. *Equilibrium Theory*

The theory of balance, often referred to as the balance paradigm, represents a new frontier in the field of legal philosophy in recent years. It signifies that legal philosophy, having evolved from a duty-centered approach to a rights-centered one, has now advanced into a new phase: the balance phase. The theory of balance is a novel research paradigm that emerged from the rights-centered approach and incorporates both rights and duties. The theory of balance emphasizes a comprehensive approach to issues. Whether at the level of constructing legal theoretical systems or in legislative and judicial practice, both rights and obligations must be taken into account; neither can be neglected or sidelined, as this would lead to an imbalance. Therefore, in a certain sense, the rule of law is governance based on the balance of rights and obligations. Furthermore, beyond the implication of balancing rights and obligations, the theory of balance has a broader scope, encompassing the balance between individuals and groups, as well as between citizens and the state. This aims to achieve interaction and mutual benefit among multiple parties, ultimately promoting social justice and upholding social order.

The theory of balance falls within both the epistemological and methodological realms. It profoundly reveals the intrinsic developmental laws of constitutional construction, establishing the principle and essence that the constitution is a system of balanced governance. It enriches and develops the theory and practice of socialist rule of law, offering a new research perspective for the construction of the rule of law and constitutional development as a whole, and providing valuable insights for addressing the practical issues of freedom of speech and content censorship.

### 2.2. *Freedom of speech on the platform*

Speech has been the primary means by which people have communicated their thoughts and emotions throughout human history. Freedom of speech, also known as freedom of expression, is a general term encompassing the freedoms of speech, publication, assembly, and demonstration enshrined in constitutional and legal provisions. The concept of freedom of speech first emerged in the democratic system of ancient Athens, where citizens had the right to deliver speeches and express their opinions at public assemblies. The concept of freedom of speech developed most significantly during the Enlightenment of the 17th and 18th centuries. Enlightenment thinkers John·Locke and Voltaire advocated that the people should enjoy freedom of speech, believing that free expression was key to advancing knowledge, ideas, and social progress. Locke supported absolute freedom of speech, viewing it as an inalienable right, while Voltaire focused more on the practical impact of publications rather than insisting on the abstract concept of freedom of speech. Since then, the theory of freedom of speech has undergone continuous development and refinement, becoming a fundamental political right in democratic nations—namely, the right to freedom of speech.

With the development of the internet and social media, freedom of speech on digital platforms has

---

taken on new characteristics:

First, the global nature of regulation: The internet is not controlled by a single country or institution. Instead, multiple jurisdictions influence the operations of global internet companies. This means that the diverse legal and regulatory systems of different countries exert influence on internet companies operating within their borders, while the users these companies serve are distributed across the globe and come from different nations and cultural backgrounds. Consequently, internet companies cannot uniformly apply a single set of content moderation standards; instead, they must grapple with the differences and conflicts among national legal norms. This creates contradictions and dilemmas, making it extremely difficult for internet companies to navigate cross-border content moderation.

Second, anonymity: The internet provides a relatively private environment that allows individuals to conceal their identities and true appearances to a certain extent, enabling them to express their opinions more freely. At the same time, this increases the likelihood of issues such as online fraud, misinformation, and cyberbullying.

Third, the immediacy of posting: Through technologies such as social media platforms, instant messaging tools, and live streaming, individuals can instantly express their views and opinions and interact with others in real time, which increases the speed and scope of information dissemination and propagation.

Fourth, the complexity of regulation: The sheer volume and rapid evolution of online content pose significant challenges for content moderation, making it difficult for regulatory authorities to comprehensively and promptly monitor online speech.

### *2.3. Platform Content Moderation*

Platform content moderation refers to the practice by online platforms of reviewing user-generated content for legality and taking appropriate measures in response. The scope of review includes information published by users in various formats, such as text, images, videos, and audio. The term “information content review authority” as used in this article refers to the power held by online platforms to review platform information content within the context of online ecosystem governance. This authority is not a public power in the traditional sense of administrative law, nor is it fundamentally equivalent to private power; rather, it exists due to the review obligations imposed on platforms by laws and regulations. The review obligation refers to the legal duty assigned to online platforms to proactively take measures to review whether information posted by platform users is lawful and compliant during online information content governance activities. It also encompasses the obligation to actively take necessary corrective actions upon discovering violations. This concept is sometimes referred to as a regulatory obligation or a governance obligation; this paper does not make a strict distinction between these terms.

This paper argues that the power to review platform content actually derives from the review obligations established by laws and regulations. Although exercised by private entities, this power has, in practice, acquired the attributes of “quasi-public power.” We classify power into that which operates in the private sphere and that which operates in the public sphere; the former constitutes private power, while the latter constitutes public power. Therefore, according to this theory, when the power held by a private entity—such as an online platform—operates within the public sphere, it should be termed “public power exercised by a private entity.” This refers to the power exercised by an online platform as a private entity that possesses the nature of public power. This concept appears in relevant literature under various terms, including private power, platform power, and private governance; for the sake of convenience, this paper does not make specific distinctions between them.

## **3. The Sources of Platform Power and the Process of Its Transfer**

Platforms are the most important business vehicles of the digital age, possessing a multifaceted advantage in terms of information, users, capital, and technology. Based on both legal and practical considerations, platforms have assumed the authority to moderate content. Within the content moderation process, however, platforms have delegated this authority to algorithms.

### *3.1. The Source of Platform Moderation Authority*

From a legal perspective, laws and regulations confer upon digital platforms the responsibility and authority to manage and review content. In 2016, the Symposium on Cybersecurity and Informatization stated that “websites should bear primary responsibility for online information management,” thereby establishing the primary responsibility of internet platforms in managing content. In 2018, it was reiterated that “the primary responsibility of internet companies must be reinforced; under no circumstances should the internet be allowed to become a platform for spreading harmful information

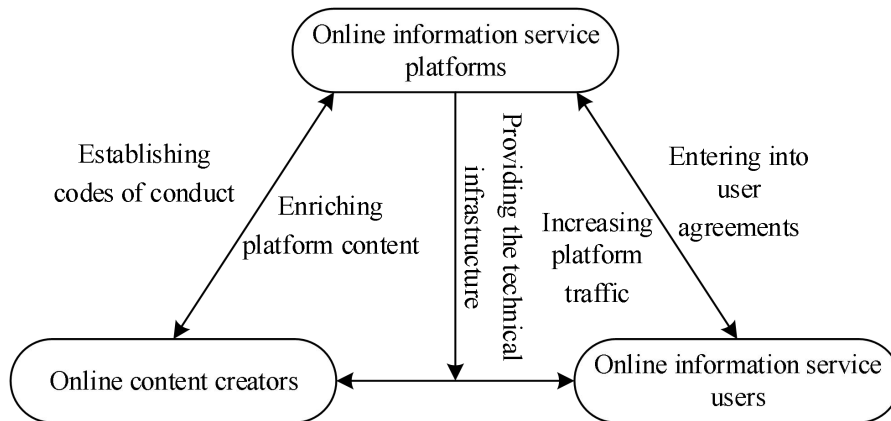
or spreading rumors,” further emphasizing the primary responsibility of internet platforms in information review and management. In 2019, the “Regulations on the Governance of the Online Information Content Ecosystem” issued by the Cyberspace Administration of China stipulated that “online information content service platforms shall fulfill their primary responsibility for information content management,” thereby clarifying the primary responsibility of platforms through regulatory documents.

From a practical perspective, internet platforms are products of technology and must rely on technology for governance. Faced with massive volumes of information, technical means are required to assess and address it, forming a closed-loop system for effective identification, prediction, and prevention. Without technological support, it would be difficult to handle the massive volume of information, and it would be impossible to make rapid judgments and take effective action regarding content. Consequently, direct government regulation of internet content has clearly reached an impasse. In its place, code—with its superior technical capabilities and management efficiency—has emerged as the most powerful weapon in cyberspace.

As can be seen from the above, whether from a legal or practical perspective, internet platforms have assumed the actual power to shape the order of the content ecosystem and regulate information content.

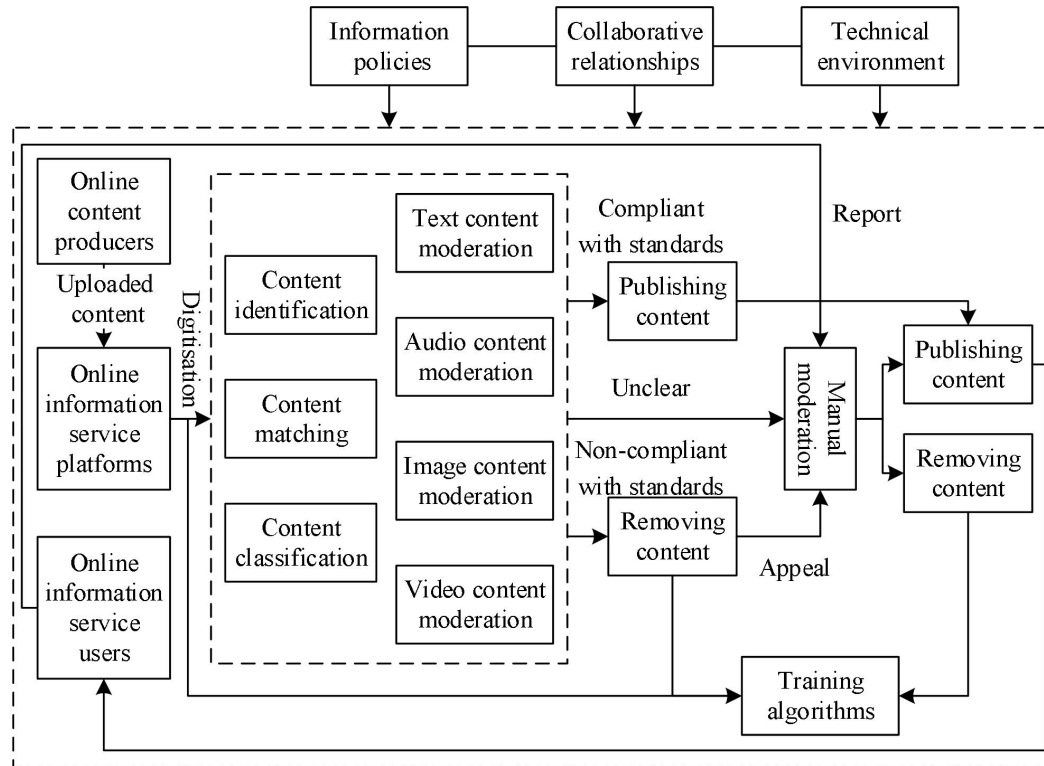
### 3.2. Transfer of Authority During the Review Process

The relationship between online information service platforms, content producers, and service users is illustrated in Figure 1. Online information service platforms adopt a “take-it-or-leave-it” contract model, entering into agreements with users who utilize their services; once a user violates the platform’s rules, they face corresponding sanctions. Even as online content producers, creators must accept the platform’s code of conduct and create content in accordance with the platform’s requirements; otherwise, they will not be granted access to the platform. Consequently, a platform-centric internet ecosystem has taken shape. The platform is the primary entity responsible for content moderation. Under the current content moderation mechanisms, most platforms employ a combination of human and algorithmic moderation, with algorithmic moderation serving as the primary method, while human moderation is reserved for handling controversial content or user appeals.



**Figure 1.** The relationship among the three parties

The platform’s content moderation process is illustrated in Figure 2. After content producers upload material to the platform, it undergoes a digitization process. The platform then uses algorithmic technology to identify, match, and classify the content, determine whether it meets moderation standards, and automatically flag or remove harmful content. Manual review, along with user appeals and reports, serves only as a supplementary review mechanism and does not extend to the processing of metadata. Consequently, throughout the entire content review process, the platform delegates review authority to algorithms, which effectively hold the power of life and death over information content.



**Figure 2.** Platform content review process

A review of content moderation processes reveals that, whether through external or internal review, the authority to moderate content lies with the platform. In external processes, while information policies, collaborative relationships between stakeholders, and the external technological environment all constrain and influence content moderation, the actual operational authority remains with the platform, which serves as the true “gatekeeper.” In internal processes, although content creators and users have channels for providing feedback, the authority to handle such feedback remains with the platform. The platform not only controls the rules governing algorithms but also establishes the guidelines for manual moderation.

## 4. A Regulatory Examination of “Private Power” in Platform Content Moderation

### 4.1. The Current State of Legislation on “Private Power”

To ensure cybersecurity, the Cyberspace Administration of China has successively issued relevant legal documents based on the “Measures for the Administration of Internet Information Services”, which came into effect in 2000. There are more than 20 laws, administrative regulations, departmental rules, and normative documents concerning the content review obligations of online platforms, specifically including three laws: the “Cybersecurity Law”, the “E-Commerce Law”, and the “Decision on Strengthening the Protection of Online Information”; two administrative regulations: the “Measures for the Administration of Internet Information Services” and the “Measures for the Security Protection of International Computer Information Network Interconnection”; six departmental regulations: the “Provisions on the Administration of Algorithm Recommendations in Internet Information Services,” the “Provisions on the Governance of the Online Information Content Ecosystem,” the “Measures for the Administration of Information Disclosure in E-commerce,” the “Provisions on the Administration of Blockchain Information Services,” the “Provisions on the Administration of Internet News Information Services,” and the “Provisions on the Administration of Online Publishing Services”; and 11 normative documents, including the “Regulations on the Administration of Internet User Public Account Information Services,” the “Regulations on the Administration of Internet Live-Streaming Marketing Information Content Services,” and the “Regulations on the Administration of Online Audio and Video Information Services”.

The legal provisions regarding the content review obligations of online platforms are shown in

Table 1. Among these, the “E-Commerce Law” and the “Cybersecurity Law” stipulate that online platforms are subject to content review obligations. The “Provisions on the Governance of the Online Information Content Ecosystem” are regulations specifically designed for the governance of online information content and serve as the primary basis for administrative agencies to enforce laws regarding information content activities. To govern the online ecosystem more clearly and specifically, administrative agencies have, based on the aforementioned regulations, formulated various normative documents targeting different types of information services. These include the “Regulations on the Administration of Internet Information Search Services” for information search services; the “Regulations on the Administration of Internet Live Streaming Services” and the “Regulations on the Administration of Online Audio and Video Information Services” for information publishing services; and the “Regulations on the Administration of Microblogging Information Services” for social media services. In addition, within China’s internet legislation, the E-Commerce Law, the Measures for the Administration of E-Commerce Information Disclosure, the Regulations on the Management of Algorithm-Based Recommendations in Internet Information Services, the Regulations on the Governance of the Online Information Content Ecosystem, and the Regulations on the Administration of Internet Group Information Services all address the procedural obligations of platforms.

**Table 1.** The current legislative status of "private power" in platform content review

<b>NO.</b>	<b>File name</b>	<b>Legal hierarchy</b>	<b>Implementation time</b>
1	E-commerce Law	Law	2019.01.01
2	Cybersecurity Law	Law	2017.06.01
3	Decision on Strengthening the Protection of Network Information	Law	2012.12.28
4	Administrative Measures for Internet Information Services	Administrative regulations	2011.01.08 (Revised)
5	Measures for the Administration of Security Protection of International Networking of Computer Information Networks	Administrative regulations	2011.01.08 (Revised)
6	Regulations on the Administration of Algorithmic Recommendation for Internet Information Services	Departmental regulations	2022.03.01
7	Regulations on the Governance of the Network Information Content Ecosystem	Departmental regulations	2020.03.01
8	Administrative Measures for the Publicity of E-commerce Information	Departmental regulations	2020.02.12
9	Regulations on the Administration of Blockchain Information Services	Departmental regulations	2019.02.15
10	Regulations on the Administration of Internet News Information Services	Departmental regulations	2017.06.01
11	Regulations on the Administration of Online Publishing Services	Departmental regulations	2016.03.01
12	Regulations on the Administration of Internet User Public Account Information Services	Normative documents	2021.02.22
13	Regulations on the Administration of Internet Live-streaming Marketing Information Content Services	Normative documents	2020.11.13 (Publish)
14	Regulations on the Administration of Network Audio and Video Information Services	Normative documents	2020.01.01
15	Regulations on the Administration of Microblog Information Services	Normative documents	2018.03.20
16	Interim Provisions on the Pre-installation and Distribution Management of Application Software for Mobile Intelligent Terminals	Normative documents	2017.07.01
17	Regulations on the Security Assessment and Management of New Technologies and Applications in Internet News Information Services	Normative documents	2017.12.01
18	Regulations on the Administration of Internet Group Information Services	Normative documents	2017.10.08
19	Regulations on the Administration of Internet Comment Services	Normative documents	2017.10.01
20	Regulations on the Administration of Internet Forum Community Services	Normative documents	2017.10.01
21	Regulations on the Administration of Internet Information Search Services	Normative documents	2016.08.01
22	Regulations on the Administration of Internet Live	Normative	2016.12.01

## 4.2. Case Study

This article identifies administrative penalty decisions related to review obligations by searching the official websites and WeChat official accounts of internet regulatory authorities. According to publicly available information on administrative penalties from the Beijing Municipal Bureau of Culture and Tourism, from January 2021 to January 2022, a total of 23 internet companies violated Article 24 of the “Regulations on the Administration of Internet Audiovisual Program Services,” Article 28 of the “Interim Provisions on the Administration of Internet Culture,” and Article 24 of the “Regulations on the Administration of Online Publishing Services” by producing, disseminating, replicating, or publishing the prohibited content listed in the provisions. As a result, they were subject to warnings, confiscation of illegal gains, or fines. Although various departments have jointly launched special campaigns such as “Sword Net” and “Blade” to combat illegal information, and have jointly issued normative documents regarding the review of illegal content, various types of illegal online information continue to emerge. Six cases of violations of Article 24 of the “Regulations on the Administration of Internet Audiovisual Program Services” are shown in Table 2; nine cases of violations of Article 52 of the “Regulations on the Administration of Online Publishing Services” are shown in Table 3; and eight cases of violations of Article 28 of the “Interim Provisions on the Administration of Internet Culture” are shown in Table 4.

**Table 2.** Six typical cases

Grounds for punishment	Administrative counterpart	Penalty content	Penalty time
The audio-visual programs disseminated contain prohibited content	Youku Information Technology Co., LTD	Warning; A fine of 10,000 RMB will be imposed	2021-03-23
	Beijing Byte Dance Technology Co., LTD	Warning; A fine of 10,000 RMB will be imposed	2021-03-29
	Beijing Kuaishou Technology Co., LTD	Warning; A fine of 10,000 RMB will be imposed	2021-05-10
	Sohu Internet Information Service Co., LTD	Warning; A fine of 10,000 RMB will be imposed	2021-03-29
	Beijing iQIYI Technology Co., LTD	Warning; A fine of 30,000 RMB will be imposed	2021-03-12
	Beijing Kuwo Technology Co., LTD	Warning; A fine of 10,000 RMB will be imposed	2021-03-23

**Table 3.** Nine typical cases

Grounds for punishment	Administrative counterpart	Penalty content	Penalty time
Publish and disseminate non-compliant online publications	Beijing Baidu Netcom Science and Technology Co., LTD	A fine of 20,000 RMB is imposed	2021-3-29
	Beijing Qihoo Technology Co., LTD	A fine of 3,000 RMB is imposed	2021-6-2
	Beijing Fantasy Zongheng Network Technology Co., LTD	Confiscate the illegal gains in 95.28 RMB; A fine of 10,000 RMB will be imposed	2021-5-7
Publish and disseminate online publications containing the prohibited content as described in Article 24	Beijing Baidu Netcom Science and Technology Co., LTD	Confiscate the illegal gains of RMB 239.79. A fine of 20,000 RMB will be imposed	2021-3-12
	Beijing Sina Internet Information Service Co., LTD	A fine of 10,000 RMB will be imposed	2021-12-9
	Beijing Qihoo Technology Co., LTD	A fine of 2,000 yuan is imposed	2021-12-9
	Beijing Sogou Information Service Co., LTD	A fine of 5,000 yuan will be imposed	2021-12-8
	Beijing Weimeng Chuangke Network Technology Co., LTD	A fine of 20,000 RMB will be imposed	2021-12-8

Beijing Sohu Internet Information Service Co., LTD	A fine of 20,000 RMB will be imposed	2021-11-9
--	--------------------------------------	-----------

*Table 4.* Eight typical cases

Grounds for punishment	Administrative counterpart	Penalty content	Penalty time
Business-oriented Internet cultural units provide Internet cultural products that contain elements that harm social morality or the fine cultural traditions of the nation Business-oriented Internet cultural units provide Internet cultural products that endanger national unity, sovereignty and territorial integrity	Beijing Sohu Internet Information Service Co., LTD	A fine of 10,000 RMB is imposed	2021-5-17
	Youku Information Technology (Beijing) Co., LTD	A fine of 10,000 RMB is imposed	2021-5-8
	Beijing Kuwo Technology Co., LTD	A fine of 10,000 RMB is imposed	2021-5-10
	Beijing Baidu Netcom Science and Technology Co., LTD	A fine of 10,000 RMB is imposed	2021-12-13
	Beijing iQIYI Technology Co., LTD	A fine of 10,000 RMB is imposed	2021-8-20
Business-oriented Internet cultural units provide Internet cultural products containing prohibited content	Beijing Kuaishou Technology Co., LTD.	Confiscate the illegal gains of RMB 102.75. A fine of 10,000 RMB will be imposed	2021-7-09
	Kuaikan World (Beijing) Technology Co., LTD	A fine of 10,000 RMB is imposed	2021-5-10
	Beijing iQIYI Technology Co., LTD	A fine of 15,000 RMB is imposed	2021-5-7

Research has found that in January 2021, in response to repeated instances of information prohibited by laws and regulations being published or transmitted on platforms such as Sina Weibo, Douban, Baidu Tieba, and Zhihu, the Cyberspace Administration of China (CAC) summoned the heads of these platforms in accordance with the law, issued warnings, and ordered them to immediately rectify their repeated violations of content review requirements. From January to November 2021, Douban was subject to a total of 20 disciplinary actions, while Sina Weibo faced a total of 44 disciplinary actions.

Based on the above data and penalty cases, this paper argues that the legitimacy of current platform content review authority remains ambiguous. Constitutional norms are insufficient for addressing speech management issues involving private entities, and there is a lack of stable institutional coordination between platform self-regulation mechanisms and external oversight. A prominent contradiction exists between the current outdated regulatory mechanisms and the diverse review obligations of various internet service providers—the regulatory mechanisms cannot fully meet the urgent need for diverse forms of oversight required to comprehensively implement public law review obligations. Given the diversity of business models in the internet industry and the distinct characteristics of various online services, the specific content and targets of public law review obligations borne by different internet service providers vary significantly. Furthermore, the fulfillment of these review obligations is highly opaque and involves complex web of interests.

## 5. A Constitutional Balancing Mechanism Based on Platform Self-Regulation

In light of the above issues, this paper proposes a balancing mechanism centered on platform self-regulation—the “moderation-constitutional-horizontal effects extension”—which advocates for institutional arrangements such as improving procedural safeguards, strengthening transparency disclosure obligations, and introducing independent review bodies to promote a dynamic balance between the protection of freedom of speech and platform content governance. The effective operation

---

of constitutional balancing mechanisms in the digital age depends on redefining the constitutional status of platforms and establishing institutional solutions that create a structural tension between private-law autonomy and public-law regulation, thereby providing a normative foundation for the ongoing protection of digital expression rights.

### *5.1. Platform Self-Regulation*

In modern society, as governance concepts have evolved, an increasing number of stakeholders have become involved in the process of national governance. Governments, industry organizations, and citizens all participate and interact with one another; consequently, national governance is no longer solely the responsibility of the government. Under the concept of collaborative governance, platforms and the government hold equal standing, and government intervention in platforms should be characterized by restraint. Generally speaking, the government does not intervene in platform self-regulation; platforms may design their own self-regulatory rules within the legal framework. However, this restraint is not absolute; when platforms violate relevant laws and regulations, the relevant administrative agencies will still intervene.

Consequently, in platform governance relationships, the government must particularly respect and guide the industry and enterprises in exercising self-governance in accordance with the law. It should encourage internet industry organizations and enterprises to leverage their professional expertise in the development of soft law for the internet, promote the formation of a soft law system with accurate content and a scientific structure across various sub-sectors of internet services, and constrain the behavior of internet platforms through self-regulation and mutual regulation. This approach aims to reduce the incidence of illegal activities by internet platforms, thereby making soft law the first legal line of defense against such platforms. At the same time, platform self-regulation must adhere to the principle of strictly protecting user rights, as the likelihood of user rights being infringed upon by platforms increases significantly under a self-regulatory framework. This requires platforms, on the one hand, to comprehensively and equitably consider rights protection across the entire platform, balancing the rights of specific entities to be protected against the rights of other entities affected by such governance measures; on the other hand, platforms must provide immediate protection of rights, such as promptly addressing user-posted content related to terrorism or violence.

### *5.2. Constitutional checks and balances*

Based on balance theory, platform content moderation is constantly influenced by internal and external environments, and thus this process is in a state of dynamic evolution. If an organization loses its original functional attributes due to environmental influences, it must reconstruct its internal structure. Therefore, the initial structure of the organization is crucial and must be robust and effective.

This paper argues that effectively safeguarding the public's right to freedom of speech on platforms and maintaining a constitutional balance between such rights and content moderation requires not only that platforms strictly and self-disciplinedly comply with China's current laws and regulations but also that they remain constantly attentive to changes in China's legislative landscape to ensure their organizational structure aligns with legal provisions. However, judicial practice reveals that China's constitution often lags behind the evolution of social realities. Consequently, to ensure that the structure of administrative organizations keeps pace with or even anticipates changes in the internal and external environment, national and local legislative bodies must enhance the effectiveness of legislation and strengthen its real-time regulatory capacity. This will compel administrative organizations, in the performance of their statutory duties, to ensure that their administrative actions are legally bound, have a legal basis, and are protected by law. To this end, this paper proposes a balancing mechanism centered on platform self-regulation, namely the "moderate-constitutional-horizontal effect extension".

First, regarding the private rights involved in platform content moderation, it is essential to ensure that decisions are made in a scientific, reasonable, and lawful manner. For specific administrative actions, laws and regulations may adopt either an enumerative or a general approach; however, it must be ensured that every administrative action is based on a specific legal citation. Only in this way can government and platform oversight be effectively strengthened in response to administrative actions.

Second, regarding the overlap and duplication of administrative functions, the organizational structure of agencies must be further optimized. Where potential overlaps in functions exist, laws and regulations should provide clear definitions. If a specific distinction cannot be made, when handling administrative matters, any department should proactively accept requests from the public and then coordinate with other administrative bodies through inter-agency communication to handle the matter.

Third, laws and regulations should standardize the specific procedural processes of administrative actions, effectively constrain the detailed conduct of administrative personnel, effectively correct

---

instances of inaction caused by loopholes in behavioral norms before or during the process, and effectively prevent administrative inaction from occurring at any stage of the administrative process.

Fourth, since the majority of actors involved in administrative inaction are individuals, the focus must be on holding individuals accountable. Improvements to legislation governing administrative organizations should effectively distinguish between individual and organizational responsibilities. In accountability, individual responsibility must be kept separate from organizational responsibility. Only in this way can individuals be prevented from evading responsibility under the guise of the organization. Accountability measures should prioritize prevention and be combined with appropriate disciplinary mechanisms.

## 6. Conclusion

Innovations in digital technology and the emergence of the online public sphere have made “everyone as a media outlet” a reality, transforming the traditional “individual-state” binary framework into a “individual-platform-state” triadic framework. In this context, citizens’ freedom of expression has been greatly expanded. At the same time, a series of online misconduct issues have followed in quick succession. Given the chaos on digital platforms, platform operators—as the creators of the online public sphere—have been assigned the duty of speech moderation. Research has found that the excessive censorship obligations imposed on platforms are closely linked to the platforms themselves, public authorities, and industry organizations: public authorities are the primary drivers behind the imposition of additional obligations on platforms, leading to excessive burdens; dual considerations of risk and economics drive platforms toward self-regulation; and as intermediary organizations connecting public authorities and platforms, their deviation from their dual roles has also contributed to the ambiguity of censorship obligations. The combination of these various factors collectively contributes to the ambiguity surrounding platforms’ censorship obligations. To address this ambiguity, this paper proposes a balancing mechanism centered on platform self-regulation, termed “Moderation-Constitutional-Horizontal Effect Extension.” It advocates for institutional arrangements, improved procedural safeguards, strengthened transparency and disclosure obligations, and the introduction of independent review bodies to foster a dynamic balance between the protection of free speech and platform content governance.

## References

1. Bondi, M., & Cacchiani, S. (2021). Knowledge communication and knowledge dissemination in a digital world. *Journal of Pragmatics*, 186, 117-123.
2. Branda, F., Ciccozzi, M., & Scarpa, F. (2025). Managing epidemics in the digital age: the crucial role of social media in information dissemination. *Pathogens and Global Health*, 119(5-6), 166-172.
3. Aguiar Filho, A. S. D., Silva, A. M. D., & Carvalho, A. C. M. D. (2025). The use of digital platforms to share information and knowledge. *Coletânea Luso-brasileira XIV: Information management, education and technological inclusion*.
4. Serpa, S., Ferreira, C. M., Sá, M. J., & Santos, A. I. (2020). Dissemination of knowledge in the digital society. *Digital Society and Social Dynamics*, 2-16.
5. Alkiviadou, N. (2025). Platform liability, hate speech and the fundamental right to free speech. *Information & Communications Technology Law*, 34(2), 207-217.
6. Arisanty, M., Wiradharma, G., & Fiani, I. (2020). Optimizing Social media platforms as information dissemination media. *Jurnal Aspikom*, 5(2), 266-279.
7. Buchanan, T. (2020). Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *Plos one*, 15(10), e0239666.
8. Cho, J. H., Rager, S., O’Donovan, J., Adali, S., & Horne, B. D. (2019). Uncertainty-based false information propagation in social networks. *ACM Transactions on Social Computing*, 2(2), 1-34.
9. Ali, S., Islam, N., Rauf, A., Din, I. U., Guizani, M., & Rodrigues, J. J. (2018). Privacy and security issues in online social networks. *Future Internet*, 10(12), 114.

- 
10. Haimson, O. L., Delmonaco, D., Nie, P., & Wegner, A. (2021). Disproportionate removals and differing content moderation experiences for conservative, transgender, and black social media users: Marginalization and moderation gray areas. *Proceedings of the ACM on Human-computer Interaction*, 5(CSCW2), 1-35.
  11. Wilson, R. A., & Land, M. K. (2020). Hate speech on social media: Content moderation in context. *Conn. L. Rev.*, 52, 1029.
  12. Young, G. K. (2022). How much is too much: the difficulties of social media content moderation. *Information & Communications Technology Law*, 31(1), 1-16.
  13. Rozenshtein, A. Z. (2023). Moderating the fediverse: Content moderation on distributed social media. *J. Free Speech L.*, 3, 217.
  14. Roberts, S. T. (2018). Digital detritus: 'Error' and the logic of opacity in social media content moderation. *First Monday*.
  15. Sander, B. (2019). Freedom of expression in the age of online platforms: The promise and pitfalls of a human rights-based approach to content moderation. *Fordham Int'l LJ*, 43, 939.
  16. Dias Oliva, T. (2020). Content moderation technologies: Applying human rights standards to protect freedom of expression. *Human Rights Law Review*, 20(4), 607-640.
  17. Wu, Y. (2022). Privacy, free speech and content moderation: A literature review and constitutional framework analysis. *Innovation in Science and Technology*, 1(4), 30-39.
  18. Pasquale, F. (2016). Platform neutrality: Enhancing freedom of expression in spheres of private power. *Theoretical Inquiries in Law*, 17(2), 487-513.
  19. Koltay, A. (2021). The protection of freedom of expression from social media platforms. *Mercer L. Rev.*, 73, 523.
  20. Karanicolas, M. (2019). Squaring the circle between freedom of expression and platform law. *Pitt. J. Tech. L. & Pol'y*, 20, 177.
  21. Li, Y. (2025). Market for Speech and Freedom of Expression: Rethinking Antitrust Responses to Platform Content Control in the Age of AI. Available at SSRN 5618812.
  22. Haupt, C. E. (2021). Regulating speech online: free speech values in constitutional frames. *Wash. UL Rev.*, 99, 751.