WEP Warwick Evans Publishing

International Journal of Social Sciences and Public Administration

ISSN: 3005-9836 (Print), ISSN: 3005-9585 (Online) | Volume 10, Number 2, Year 2026
DOI: https://doi.org/10.62051/ijsspa.v10n2.04
Journal homepage: https://ijsspa.org

Governing Cyberviolence in China: Limitations of the Cyberviolence Information Governance Provisions

Yuxuan Zhao

School of Maritime Law and Humanities, Dalian Ocean University, Dalian, Liaoning 116000, China

ABSTRACT

Cyberviolence, as a side effect of increasing internet penetration, is now prevalent worldwide. To combat the risks and acts of cyberviolence, in 2024, the Chinese government implemented its special regulation over 'cyberviolence information'. These new 'Provisions' not only constitute an important part of the Chinese cybersecurity framework but also advance it from a reactive model to a proactive model, from governing *behaviours* to governing *information*. More notably, it greatly expands platforms' role in cyber governance and empowers them to carry out content moderation and manage user accounts. However, 'cyberviolence information' is too broad and inaccurate of a legal concept to be used for censoring online speech and over-criminalizing legitimate communication. The strategy of 'privatization of regulation' and the empowered 'digital Leviathan' also pose severe threats to future Chinese development of the rule of law and democracy.

KEYWORDS

Cyberviolence; Cyberviolence Information; Online Speech; Overcriminalization; Privatization of Regulation.

1. INTRODUCTION

The rapid development of internet technology has significantly extended cyberspace, establishing it as a critical platform for public communication and interaction. However, the anonymity and openness inherent to this digital environment have unintentionally brought about the rise of cyberviolence-a multifaceted societal challenge with extensive repercussions. It typically refers to a range of online malicious behaviour conducted via internet platforms with the intent to inflict psychological, emotional, or social harm on individuals or groups. Cyberviolence takes various forms, including cyberharassment (repeatedly sending unwanted or aggressive messages), cyberbullying (intimidation or threats through digital platforms), doxxing (publicly releasing personal information without consent), rumour dissemination (intentionally spreading false information to damage reputations), and hate speech (insulting or discriminatory remarks based on characteristics such as race, religion, or gender). These methods and forms can also evolve continuously with the development of cyberspace and social networks.

The Chinese internet user base expanded from 620,000 in 1997 to 1.108 billion in 2024, with the internet penetration rate reaching 78.6%. According to a survey conducted by the Social Survey Center of China Youth Daily, 65.3% of the surveyed young people reported that they or those around them had suffered from cyberviolence. Furthermore, cyberviolence has emerged as a catalyst for multiple incidents that have attracted significant public attention. A notable case is that of Liu Xuezhou, a Hebei Province teenager who sought to reunite with his biological family. Following his successful reunion and public disclosure of chat records indicating that his birth mother had blocked him, he endured malicious misinterpretation and verbal abuse. Certain netizens assigned derogatory

labels to him, including ‘ingrate’, ‘attention-seeker’, ‘vain’, and ‘calculating’. On January 24, 2022, Liu chose to end his life at the age of fifteen years .

In another notable case, Zheng Linghua, who shared a video of presenting her graduate school acceptance letter to her grandfather in the hospital, faced online verbal abuse solely because of her pink-dyed hair. Subsequently, on January 23, 2023, 23-year-old Zheng ended her own life, calling forth public attention to the issue of cyberviolence. Moreover, following the tragic death of a primary school student in Wuhan, who was fatally struck by a teacher’s vehicle on campus, netizens harshly scrutinized and criticized the immediate emotional responses of the victim’s mother. Amid her grief over her child’s loss, she endured cyberviolence attacks and baseless accusations, which ultimately led to her suicide .

These damaging effects on individuals’ psychological health and social order raise public sentiment that ‘the internet is not a free-for-all for trolls’. Stricter regulation is needed, especially on behalf of victims and vulnerable groups such as children and women. A key milestone came in September 2023 when the Ministry of Public Security, the Supreme People’s Court (SSC), and the Supreme People’s Procuratorate (SSP) jointly issued the *Guiding Opinions on Punishing Cyberviolence Violations and Crimes in Accordance with Law* (‘Guiding Opinions’). This document marked a shift towards treating cyberviolence as a serious societal harm. Moreover, in June 2024, the Cyberspace Administration of China (CAC), along with other agencies, released the *Provisions on the Governance of Cyberviolence Information* (‘Provisions’), effective from August 1, 2024.

Although the ‘Provisions’ provide swift enforcement, stricter punishments and tech-empowered prevention measures, negative ramifications may also be generated from this mode of cyberviolence governance, especially because of the vague definition of online ‘violence’ and the privatized surveillance power entrusted to platforms. The existing literature has focused mainly on the constructive role of the ‘Provisions’ in combatting cyberviolence; however, it lacks a critical reflection on the shortfalls and adverse impacts of the ‘Provisions’ on China’s netizens’ civil liberties. This paper aims to address this gap by providing a comprehensive analysis of the definition and scope of regulation concerning cyberviolence and the interplay between the specific ‘Provisions’ and other laws on cybersecurity to identify and critically evaluate the unique features of the Chinese legal framework of cyberviolence governance.

2. CYBERVIOLENCE: AN EMERGING SOCIAL CHALLENGE

As a social phenomenon that has emerged with the development of the internet, the understanding of cyberviolence has improved with the increase in regulations and relevant research. Early research took cyberbullying as the main form of cyberviolence, while Onesemus Awiria et al. summarize its three constitutive elements: intentional harm, repetition, and power imbalance.[1] Similarly, Peter Smith et al. (2008: 376) define it as ‘an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend themselves’.[2] Ersilia Menesini and Annalaura Nocentini [3] contend that ‘repetition’ is not key to constituting cyberviolence because a single act (such as privacy leakage) can inflict sustained harm on victims because of the uncontrollable spread of online content. In response, Ira-Katharina Peter and Franz Petermann propose a reconciliatory framework that defines repetition as ‘technology-enabled virtual repetition’, namely, a single anonymous attack may generate repeated harm through potential dissemination.[4] This framework thus captures the power dynamics enabled by digital anonymity and can integrate power imbalance, direct/indirect forms and victim perception.

In legislative texts, France’s Penal Code prioritizes ‘victim perception’ by establishing the ‘deterioration of mental health’ as a constitutive element of cyberviolence (Article 222-33-2-3), while U.S. federal and state legislation uniformly require ‘substantial harm’ as a constituent element of cyberbullying. Comparatively speaking, cyberviolence in the Chinese legal system combines both

‘objective acts’ (such as verbal abuse, defamation, slander, and privacy violations) and ‘victim perception’ (such as ‘social death’, mental health disorders, or even suicide).

Because cyberviolence is carried out indirectly through the internet and is aimed at causing the ‘social death’ of the victims, Xie Dengke (2023: 93) classifies it as ‘psychological violence’ or ‘soft violence’[5], where there is no direct physical contact between the perpetrator and the victim other than verbal intimidation and insults. Cyberviolence usually uses moral labelling narratives and amplifies their aggressive efficacy through sophisticated semantic manipulation and discursive strategies. For example, Xu Youping analyses cyberbullying language that targets adults in Chinese social media through the Wang Fengya incident, revealing prominent features of aggressive cyberviolence discourses. He [6] finds that the linguistic negation of cyberviolence operates lexically (stigmatizing labels such as ‘dog beast [gou chusheng]’ or ‘vicious mother [hendu de ma]’), grammatically (rhetorical questions with exclamation marks), and discursively (assertive statements and cursed ‘blessings’).

In addition to group-based verbal abuse, Adejoke Adediran’s research[7] reveals another main type of cyberviolence in Nigeria, specifically, ‘information abuse’, such as ‘outing’ (publicizing private information without consent and leveraging social media for humiliation diffusion) and ‘trickery’ (obtaining private content through inducement or coercion followed by threats of disclosure or misuse). Lovelock Michael examines the phenomenon of ‘catfishing’.[8] In digital contexts, ‘catfishing’ refers to creating false social networking profiles for fraudulent purposes, often to establish deceptive romantic relationships.

The emergence of cyberviolence worldwide can be attributed to multiple reasons, such as individual, social, and technological problems. Zhao Hong (2023: 7) notes that reliance on smartphones and social media blurs the boundaries among individuals, machines, and algorithms, causing people to gradually lose their awareness of complex emotions and attention to the situation of others.[9] In addition, the dark side of group communication and deliberation can push biased opinions towards extreme views. Research by Peter and Petermann also highlights how anonymity in online interactions lowers perpetrators’ accountability, often resulting in extreme insults and metaphorical threats.[4]

As a reflection of real-life conflicts, potential victims of cyberviolence can vary from country to country. Smith et al. conduct two studies on UK secondary school pupils (aged 11–16) and reveals a significant overlap between cyberbullying and traditional bullying:75% of cybervictims are also victims of traditional bullying.[2] This finding captures the interconnectedness of online and offline aggression, with cyberbullying often occurring alongside traditional forms of bullying such as verbal or physical harassment. Research has further indicated that older students (14–16 years) face greater risks of cyberbullying, which is attributed to their ‘more complex online social ecosystems’ . Fatma Mohamed Hassan et al. (2020: 4) identifies distinct victimization patterns among Egyptian women, highlighting that the prevalent forms are ‘receiving images or symbols with sexual content’ (41.2%), ‘insulting e-mails or messages’ (26.4%), and ‘offensive/humiliating posts/comments’ (25.7%).[10] The authors thus critically framed cyberviolence as ‘a digital extension of traditional patriarchal structures’. In other words, cyberviolence perpetuates gender-based discrimination and control in online spaces.

The global proliferation of cyberviolence has generated preliminary academic consensus on the necessity of legal regulation. Nevertheless, marked divergence persists within academia regarding the optimal legal regulatory approaches, especially in terms of whether specialized legislation is required to address cyberviolence. This debate reflects the underlying tension between legal deterrence and the protection of civil liberties. Proponents, grounded in practical judicial challenges, explicitly argue that existing legal frameworks fail to adequately identify and regulate the digital-specific harms of cyberviolence. Conversely, opponents emphasize the potential systemic risks inherent in specialized

legislation: lacking a precise definition of behavioural thresholds, cyberviolence legislation risks triggering arbitrary enforcement overreach, causing unintended damage to free speech protection.

Cyberviolence thrives in a sociotechnical environment where anonymity intersects with weak law enforcement deterrence. The limitations of existing legal frameworks stem from their failure to anticipate the digital-specific attributes of cyberviolence, which enables offenders to systematically exploit regulatory blind spots. For example, the European Court of Human Rights' landmark ruling in Volodina v. Russia [11] exposed the systemic inadequacies in contemporary legal frameworks that govern cyberviolence. In this case, the applicant, Ms. Volodina, endured prolonged exposure to hybrid violence that combined the digital and physical dimensions, encompassing nonconsensual dissemination of intimate imagery alongside sustained threats and stalking. The adjudicating panel asserted that the existing legal framework demonstrates 'structural deficiencies in its judicial system which cause delays' to protect victims of gender-based cyberviolence . However, Jack Balkin worries that specialized cyberviolence regulation risks conflating harmful speech with legitimate expression through definitional overreach.[12] Vague statutory language, such as criminalizing 'offensive' or 'harmful' content, could suppress dissent. Cybersecurity legislation can become a tool to 'blur reasonable criticism and illegal cyber attacks under the guise of cyber violence governance'.

From a different perspective, Alexey Ilin [18](2022: 83-85) argues that China and Russia do not need to enact new laws, as existing civil, administrative, and criminal laws, through amendments and interpretations, are already sufficient to address cyberviolence.[13] The key lies in removing procedural barriers in litigation, such as cumbersome notarization procedures. In sum, proponents for specialized legislation overstate the inability of static legal frameworks to cope with technological challenges such as generative artificial intelligence, whereas opponents dismiss the law's function in anchoring normative values, ultimately rendering the debate trapped in a false dichotomy between 'deterrence futility' and 'regulatory laissez-faire'.

The literature has thoroughly discussed the definitions, features, causes and victim composition of cyberviolence, providing strong justifications for cyberviolence governance. However, perspectives on legitimate governance models still diverge due to tensions between regulatory necessity and civil liberties. In this case, evaluating the effects and effectiveness of China's 'Provisions' requires both contextualization efforts and critical reflection.

3. THE EVOLVING HISTORY OF CHINESE CYBERVIOLENCE GOVERNANCE

(1) The early stage: Regulating 'defamation' with a general law

China's journey in regulating cyberviolence has been a dynamic and evolving process closely intertwined with the rapid growth of computer device usage and the expanding netizen population. In 1997, the State Council took the first step in internet governance by promulgating *Measures for Security Protection Administration of the International Networking of Computer Information Networks*. These pioneering regulations set the tone by prohibiting content that 'damages the reputation' of state organs or 'spreads rumours' (Article 5). In addition, online defamation can be prosecuted under the Chinese 1979 Criminal Law (Article 246) if serious consequences are caused by online insult or defamation. At this early stage of internet development, the legal response to cyberviolence was rather rudimentary. When the phenomenon of 'human flash search' (namely, doxxing) emerged in the 2000s, particularly after the 2006 'cat abuse incident', where the abuser's privacy was callously exposed either to voice criticism or out of sheer curiosity, the existing legal system struggled to keep pace. Courts resorted to applying tort provisions related to reputation damage (General Principles of the Civil Law of the PRC, 1986, Article 101) to address the personal harm inflicted by such incidents.

The Decision of the Standing Committee of the National People's Congress on Preserving Computer Network Security came out on December 28, 2000 and represented an enormous achievement in cyberviolence legislation. Article 4 of this vital document establishes a connection between online insults and defamation that constitutes one or more criminal offenses. It successfully overcomes the restrictions of traditional laws that were centred primarily around physical spaces. Article 6 also clarifies that noncriminal cases of cyberviolence are regulated by the *Public Security Administration Punishment Law*, which empowers public security organizations to carry out investigations. In 2009, Criminal Law (Amendment VII) introduced the crime of 'illegally obtaining citizens' personal information' (Article 253-1). This addition was a significant change. By criminalizing the illegal acquisition of personal information, this law directly addresses the root cause of cyberviolence incidents such as doxxing.

(2) The second stage: The birth of online-specific rules

In the 2010s, China's internet entered a booming period, during which social media penetration began to rise. Against this background, initial countermeasures against cyberviolence also emerged. On July 1, 2010, the *Tort Liability Law* was promulgated and implemented. Article 36 clearly stipulates that civil liability should be introduced for various types of online harm. In addition, the relevant platforms may be held accountable if they do not take timely action to remove the illegal content after being notified. In 2011, the State Internet Information Office (SIIO) introduced a real-name registration policy on Weibo. The policy was a direct response to the abuses that were rampant on the internet at the time, including rumour-spreading and harassment. Although this policy is not a formal legal provision, it has been effectively implemented with the active cooperation of various platforms.

In 2013, the SPC and the SPP issued the *Interpretation of the Supreme People's Court and the Supreme People's Procuratorate on Several Issues concerning the Specific Application of Law in the Handling of Defamation through Information Networks and Other Criminal Cases*

('Interpretation'), which aims to further clarify how criminal law should be applied in relation to cyberviolence. According to this 'Interpretation', if a post has more than 5,000 views or has been retweeted more than 500 times, then the online defamation involved in the post can be considered a criminal act (Article 2). In this way, there is a quantifiable measure of the scope of cyberviolence and the degree of disruption that it causes.

(3) The third stage: Embedding cyberviolence governance in the cybersecurity framework

Since the mid-2010s, cyberviolence regulation has been incorporated into a broader and more comprehensive cybersecurity framework, with platforms playing a central and proactive role in enforcement. The *Cybersecurity Law*, which went into effect on June 1, 2017, was a landmark piece of legislation. Article 12 of this law prohibites online content that 'spreads rumours, disrupts social order, or harms others' reputations'. Article 47 mandates that platforms actively monitor and remove illegal content. Noncompliance with these provisions could result in fines of up to 500,000 CNY. Although the term 'cyberviolence' was not explicitly mentioned, it was clearly included in the category of 'harmful content'. The *Provisions on Ecological Governance of Network Information Content*, which took effect on March 1, 2020, further broadens the definition of 'illegal content'. It includes 'negative information' that could cause psychological harm or social unrest (Article 7), and Article 6 explicitly lists defamation and rumour-spreading as prohibited activities. This expansion of the legal definition indicates the growing recognition of the diverse forms and impacts of cyberviolence. In response to these regulations, companies such as Baidu and ByteDance have made substantial investments in artificial intelligence technologies for instant harassment detection. Moreover, since October 1, 2017, the CAC issued the *Notice of the Cyberspace Administration of China on Issuing the Provisions on the Administration of Internet Comments Posting Services*, which has required platforms to conduct prepublication reviews of user comments to prevent the spread of offensive and malicious content, which demonstrates a shift towards more proactive and preventive measures (Article 5-3).

Since the 2020s, regulations concerning cyberviolence have been entangled with struggles over data protection. Besides the *Provisions on Ecological Governence of Network Information content,* Another major legislative piece is the *Personal Information Protection Law*(PIPL), which came into effect on November 1, 2021. It focuses on privacy infringements such as doxxing and includes Article 49, which allows victims to claim compensation, and Article 66, which imposes massive fines on companies convicted of data breaches. *The Data Security Law* (DSL) combats cyberviolence through data protection and compels companies to protect data against leaks that may result in being used to facilitate cyberviolence (Article 27). In November 2022, the CAC issued the *Notice by the Secretary Bureau of the Office of the Central Cyberspace Affairs Commission of Effectively Strengthening the Governance of Cyber Violence* ('Notice'), which marked the official shift from 'passive handling' to the 'technology-based proactive prevention and control' of cyberviolence. Platforms are required to establish a 'one-click protection' function

(4) The fourth stage: Specialized cyberviolence regulation

China entered the specialized cyberviolence regulation stage. *Regulation on the Protection of Minors in Cyberspace was* promulgated on January 1, 2024. It was China's first specialized administrative regulation on safeguarding minors online. It establishes a national coordination mechanism led by a cyberspace administration and imposes primary obligations on network service providers, such as digital literacy enhancement, content governance, personal information protection, and internet addiction prevention. On August 1, the 'Provisions' came into effect with a particular focus on 'cyberviolence information' and require platforms to establish sophisticated early warning systems, stringent verification processes and integrated content blacklists. Failure to do so may attract fines of up to 2 million CNY (Article 30). The latest *Network Information Key Data Security Management Regulation*, effective from January 1, 2025, also addresses data-fuelled cyberviolence, although it does not specifically target it.

From the intricate history of cyberviolence regulation in China, several distinctive features of its regulatory model regarding cyberviolence are revealed. First, the Chinese legal framework for cyberviolence is deeply rooted in practical experience, even without a thorough theoretical basis. It has evolved in a strategic manner, responding to real-world incidents and needs. The formal definition of cyberviolence emerged relatively late in the regulatory timeline as an inaccurate summary of the complex, diverse and ever-changing forms of online harm. Although this approach has allowed for a more targeted and effective response to specific challenges as they arose, it also leaves a significant vacuum in regulating or criminalizing cyberviolence.Clarifying the concept and classifying behaviours should be the first step in achieving effective preventive regulation of cyberviolence. A clear classification system not only facilitates early detection and appropriate response but also reflects the requirements of the rule of law and human rights protection.

Second, China's regulatory approaches to cyberviolence have become increasingly diverse. Cyberviolence governance has evolved from a primary reliance on criminal law to a wide range of legal mechanisms, including civil law, administrative law, departmental rules, and platform-enacted rules. This multilayered framework can provide comprehensive and flexible means to address cyberviolence by enabling different types of legal responses that vary in accordance with the nature and severity of incidents. However, controversies will probably arise when conflicts between different layers and sources of norms conflict with one another. Most worryingly, the whole regulatory framework of cyberviolence governance may be founded without fully considering the Chinese constitutional restraints related to citizens' civil liberties because of China's lack of effective mechanisms of constitutional adjudication and review.

Third, in the new regulatory landscape, platforms have become key actors in the struggle against cyberviolence, which is often described as the 'privatization of regulation', and this has placed substantial powers and obligations on platforms. Therefore, crucial questions have been raised about the distribution of power and responsibility among the government, platforms and internet users. Jing

lijia contends that the platforms' obligations ought to be situation-specific. Industries have their own idiosyncratic technology landscape and user base.[14] A small, regional streaming platform, for example, might not have the same resources as a global social media behemoth. Thus, considering the nature and size of the industry, the obligations imposed on network service providers should be reasonable and feasible. Zhu Xiaoyan recommends that platforms should establish a content supervision mechanism for their early warning systems according to their own content ecology.[15] In the case of visual-sharing platforms, inappropriate or violent visual components should be detected by monitoring systems, and text-based platforms should focus on upgrading natural language processing algorithms to recognize violent or harassing linguistic patterns.

Finally, the focus of regulation has gradually expanded from individual violent behaviours to the entire information ecosystem. Along with datafication, regulations now target mainly the spread and impact of cyberviolence-related information, leveraging data-driven approaches to detect and prevent potential harm. This shift represents a fundamental change in the regulatory mindset as it moves from a reactive approach to a more proactive and preventive approach. Datafication also enables the embeddedness of cyberviolence governance in the general framework of cybersecurity and serves the purpose of maintaining social harmony and national security. Against this background, the one-sided support for stricter regulation and criminalization of cyberviolence is understandable, but public sentiment concerning the over-censorship of online speech cannot be heard or put in an equal position under the constitutional principle of proportionality. Therefore, cyberviolence governance can cause adverse ramifications, which we examine in the following section.

4. SHORTFALLS OF CHINESE CYBERVIOLENCE GOVERNANCE

(1) Overcriminalization of 'cyberviolence'

As we examined, China's regulatory framework for cyberviolence has been an evolving process whose shifts are largely shaped by the distinct logic of risk criminal law theory. According to Ulrich Beck's insights into risk society, risks are embedded within the very institutional frameworks that generate them. Scientific development is both the cause of risk and the medium through which risk is defined, which generates uncertainties that exceed the control of its own institutional frameworks.[16] Following this logic, the risk of cyberviolence is derived from the iterative development of internet information technology and the continuous expansion of internet users and online communitie, while the 'institutionalized security commitments' in the traditional industrial era are deconstructed by digitalization. For example, deepfake technology, which relies on generative adversarial networks (GANs), generates unreal audio-visual content with a high level of fidelity. This amplifies the degree of harm that false information causes to traditional legal interests, such as national security, social order, and civil rights, and poses significant or even structural challenges to the traditional criminal law framework [17].

Risk criminal law theory was developed to respond to institutional dilemmas by transforming the role of criminal law from a tool for 'legal remedies' into a 'risk control system'. Since information technology amplifies traditional harm through digital media's fissile dissemination, legal regulation is compelled to move from the classical 'ex post accountability' model towards ex ante information interdiction frameworks.

Risk criminal law theory is reflected in Chinese legislation and adjudication. The 2013 'Interpretation' transcended the traditional reliance on individual legal interest violations by establishing quantitative thresholds ('5,000 clicks') as the objective criteria for public prosecution. This marked the official recognition of cyber order as an independent legal interest. When defamatory acts intersect with large-scale dissemination in cyberspace, their harm is no longer confined to individual reputation infringement but is abstracted into systemic risk, which undermines societal trust mechanisms and thereby shifts the evaluative focus of defamation from private legal interests to public order. The

Criminal Law Amendment IX further introduces the offense of ‘fabricating and intentionally disseminating false information [bianzao guyi chuanbo xujia xinxi zui] (Article 291); this decouples criminal liability with actual harm and redefines ‘false information + dissemination’ as an abstract danger. This legislative approach thus advances criminal law intervention to the stage of risk creation, reflecting a functionalist turn in risk control.

Chinese judicial practice also shows this paradigm shift. In the Hangzhou Courier Defamation Case (Guiding Cases of the Supreme peoples Procuratorate No. 137), the procuratorate departed from the traditional evidentiary requirement of proving ‘concrete reputational damage’ to construct a framework of criminal illegality grounded in cyberspace-specific data metrics, namely, the information spread across 110 WeChat groups (with a total of 26,000 members), accumulated 20,000 views, and generated 1,000 pageviews on one website. Drawing on this quantitative analysis, the prosecution concluded that the defendants not only severely infringed the victim’s right to personal dignity but also generated public unease and eroded societal perceptions of security.[18]By correlating statistical indicators of information dissemination with the probabilistic harm to public order, the judiciary transformed ‘social reputation damage’ into quantifiable ‘information loss-of-control risk’ and shifted the standard of criminal wrongfulness from the concrete infringement of individual rights to abstract predictive threats to cyber order. One example is the Qvod case[(2016) Beijing 01 Criminal Appeal No. 592]. The court determined that although the internet service provider (ISP) Qvod Company did not directly upload or disseminate obscene videos, it constructed a P2P network platform through its QSI resource server program and Qvod player. These mechanisms enabled users to publish video links, while cache servers automatically stored popular videos including obscene videos based on click volumes, thereby accelerating dissemination. The court held that Qvod’s cache server technology was not merely a ‘neutral tool’ but that it substantially intervened in the dissemination of obscene videos. By emphasizing that Qvod evaded regulation through fragmented storage, which is the core element of this crime, the judiciary showed a proactive attitude towards network technical risk. Such proactive intervention in potential risks closely aligns with risk criminal law theory in ‘replacing retribution with prevention’.[19]A defamation case[(2022) Guangdong 1971 Criminal First Instance No. 2188] offers another example. The facts that the defendant created were widely reposted and discussed online (75,608 comments, 31,485 reposts, and over 470 million views), leading to extremely adverse social impacts(Supreme People's Court of the People's Republic of China, 2023).

Risk criminal law theory identifies actors’ unlawful acts as risk variables within public order and transforms the elements of a crime into algorithmic matrices in cyberspace. This judicial technique dissolves the exculpatory space for neutral technical acts to achieve an elastic expansion of criminal liability. The concrete manifestation of risk criminal law theory in Chinese judicial practice reveals a dual dynamic of normative reconstruction and institutional alienation driven by technology. Those metrics, such as click-through rates (CTR), now serve as normative proxies for abstract collective interests of ‘information ecosystem security’. However, the idea of ‘data as justice’ has the potential to subordinate the objective foundation of criminalization to the instrumental rationality of technological governance.

China’s cyberviolence regulation thus reveals a profound tension between the demands of cybergovernance and normative rationality. Although the normative reconstruction responds to anxieties about the uncontrolled proliferation of illegal information, it precipitates a crisis in the clarity of the constituent elements of criminalization. As Chinese criminal legal scholar Zhang Mingkai [20]argues, if criminal law were to prohibit all conduct involving risks, then social development would be impossible. Unlawful risks such as cyberviolence are indeed byproducts of cyberspace development, but criminal law must tolerate ‘permissible dangers’ even in a risk society to avoid stifling social vitality through excessive risk control. The task of criminal law is to regulate unlawful human conduct rather than eliminate all risks, and its application must be based on a rational balance between the harmfulness of conduct and its social necessity.

(2) Freedom of speech is endangered by cyberviolence governance

Cyberviolence, as a soft, indirect type of violence, is distinguished from the traditional form of violence by its virtual exertion and is thereby usually categorized as a form of verbal abuse. This feature inevitably connects cyberviolence governance to speech regulation and problematizes whether cyberviolence will curtail freedom of expression in China. This issue is seldom discussed, probably because there is not mature free-speech jurisprudence in China. In the following sections, we examine this threat to free speech from three perspectives.

A. Lack of classified cyberviolence information

According to Article 32 of the "Provisions," cyber violence information refers to illegal and harmful content collectively published online against individuals through text, images, audio, video, and other means, including insulting remarks, defamation, privacy violations, as well as severe psychological harm caused by moral coercion, derogatory discrimination, and malicious speculation. Based on this definition, cyber violence information in Chinese law is broad in scope, most similar to "harmful speech" or "hate speech" in American jurisprudence, yet it lacks detailed standards for the core and subcategories of such speech.

In American legal theory, freedom of speech is one of the fundamental rights and cornerstones of constitutional principles. Both former U.S. Supreme Court Justice John Paul Stevens and renowned scholar Cass Sunstein have highlighted the key role of freedom of speech in democratic deliberation due to its tolerance of intolerant expression.[21] Freedom of speech represents recognition of human dignity, autonomy, and rationality, promotes the discovery of truth, enhances civic awareness, reduces social instability, and fosters government accountability.[22] Not all speech should be protected to the same degree.[23] Derogatory speech targeting groups based on race, gender, religion, or other characteristics constitutes hate speech and is an exception to free speech protection. Such speech has become increasingly prevalent on online platforms due to the anonymity and accessibility of internet services, potentially intimidating users, fueling radicalism, and inciting violence.[23] The internet has fostered a "global racist subculture," amplifying hate speech, harassment, and discrimination, making hate speech subject to regulation. Hate speech should be subject to varying degrees of restriction, depending on factors such as the number of victims and the reasons behind it.[24] While typical cyber violence cases in China target specific individuals or small groups, falling into a regulable category, the preventive measures taken by the Chinese government based on statistical indicators are too broad to support precise distinctions that safeguard democracy-enhancing speech.[25] Without clear classification, the concepts of "hate speech" or "cyber violence information" remain ambiguous.

B. Lack of classification of platform types

The 2022 ‘Notice’ requires strengthening the management of comments, key topic groups and sections, and live streams and short videos. After that, because of the processing obligation of this information, the 2023 ‘Guiding Opinions’ prescribe that network service providers must be convicted and punished for the crime of refusing to perform the information network security management obligation in accordance with Article 286-1 of the Criminal Law. The 2024 ‘Provisions’ officially put platforms at the centre of cyberviolence information governance with a full list of powers and obligations (Article 7), such as content moderation, user registration (Article 8), account management (Article 9), the protection of personal information, and the review of information release, early warning, identification and disposal (Articles 17, 18, 21 and 22). Article 19 of the provisions is particularly targeted at network forum communities and network groups. When there are posts involving cyberviolence in these forums, internet services for these communities will be severely restricted. These measures thus greatly empower platforms and replace the traditional dyadic relationship between the state and speakers with a triadic model of speech regulation.

However, in China, there are no equivalent accounts that address platform types with varying protection of users' speech rights. In 2024, several prominent Chinese platforms published their annual reports on cyber-ecology, revealing effects of their governance strategies and policies. TikTok, as the largest video sharing platform, claimed to have deleted 45.2 million pieces of rumour information and blocked 34 million pieces of rumour information. It also ceased the transmission of 1,700 million pieces of cyberviolence information and imposed punitive measures on 4.5 million user accounts. The microblog (Weibo) disposed of 138,608 rumour messages, closed 1,122 user accounts and published 6,399 rumour refutation statements . Another platform, Zhihu, an aggregated information site that highly relies on collective contributions, disposed of more than 270.2 million 'unfriendly' pieces of information and 96,000 'unfriendly' accounts while claimed to have protected 394,406 users from strangers' 'intrusion [qinrao)'. It also dealt with more than 17,000 instances of misinformation in 2024 . Although these reports comply with Article 11 of the 'Provisions', they do not publicly demonstrate how they fulfil the requirements of Article 12, which imposes on platforms the obligation to specify the standard rules for identifying and classifying this information based on the feature database and sample database of typical cases. Therefore, we have no idea as to whether the 'unfriendly information [bu youshan xinxi)' categorized by Zhihu is equivalent to cyberviolence information and what effects will be generated to this aggregated information site if strangers are blocked from contacting users. Without reporting their standards for classifying this information, we cannot analyse platforms' prioritization and calculate their error rates, not to mention whether there is any effective complaint system for over-censorship. This situation indicates that in governing cyberviolence information, freedom of expression is seldom discussed as trumping over-surveillance. Future research demands a context-sensitive approach to explore the types of Chinese social platforms, their modes of information transmission and their extent of social harm to specify the categories of Chinese social media and real-life threats that arise from these platforms.

C. Lack of detailed accountability mechanisms for platforms

As the new governor, platforms have a series of content moderation and account management powers. Article 14 allows them to apply China's social credit system to reduce the credit rating of the user accounts involved in cyberviolence or blacklist them in future services. Article 15 gives service providers the ability to cease transmission; take disposal measures such as deleting, shielding, and disconnecting links when they find illegal information involving cyberviolence or find key links that easily attract the attention of users. Although Article 26 requires cyber information service providers to consciously accept public scrutiny, publish the handling process and report the handling results, it aims to empower victims in seeking remedies rather than addressing users' complaints about over-censorship. Therefore, these obligations to comply with the law are significantly one-sided; there are no legal norms concerning how these platform standards and measures should be reviewed in accordance with higher law to protect the legitimate rights of users. The fundamental flaws of lacking a clear definition of cyberviolence and lacking transparency in case handling make Article 33 of the 'Provisions' and Article 1 of the 'Notice' redundant, as there is no clear distinction between cyberviolence and public opinion supervision or benign criticism.

Evelyn Douek (2021: 784) proposes that in comtemporary network society, a new model of systemic balancing, driven by proportionality and probability should be applied in speech protection to replace the 'posts-as-trumps' model. The proportionality principle requires balancing the harm of misinformation and hate speech against speech rights, while the probability principle evaluates acceptable platform error rates [23]. As long as human and AI moderators cannot review all content accurately, transparency about error rates becomes critical to maintain platforms' accountability. Platforms must clearly articulate the rules and their purposes, openly report error rates and develop reliable mechanisms to challenge and rectify moderation errors. That is, it is necessary that platforms adopt a 'technological due process' [24]to ensure transparency, notice, and fair hearings for moderation decisions. In the future, redress and remedies should be provided legally by imposing transparency and participation requirements on platform-enacted rules and by providing effective

channels of appeal [25]. Rather than banning harmful speech, allowing counter-speech as a redress for cyberviolence should also be encouraged to welcome civil society inputs and more voices.

(3) Privatization of regulation is creating a 'digital Leviathan'

In the online public sphere, the rise of algorithmic regulation has driven a structural transfer of regulatory power. The Chinese government has delegated part of its public functions for cyberviolence governance to platforms, enabling them to evolve from mere technical service providers into governance entities that integrate 'quasi-legislative, quasi-enforcement, and quasi-judicial' functions.

A. Quasi-legislative power: Privatized rule-making

As previously mentioned, Article 12 of the 'Provisions' requires internet service providers to implement a classification-based governance mechanism for information. This implies that platforms must first refine classification criteria and then achieve precise identification of new forms of cyberviolence through updating feature databases and sample databases of typical cases. Some platforms choose indirect governance through community convention rules. Taking the Weibo Community Convention [weibo shequ gongyue] (2021) as an example, although it does not directly define cyberviolence, Article 36 includes behaviours such as malicious marketing, hate propaganda, and fan-circle violations under the category of 'harmful information'. Nevertheless, the platforms execute their obligations mainly through keyword detection and data flow control, which indicates a replacement of legal rules with technical standards. As Lawrence Lessig puts it, 'Code is Law', and technological architectures have evolved into normative systems with coercive force.[26]

In addition, when users seek to join specific online communities or spaces, they must sign user agreements, 'voluntarily' ceding part of their behavioural freedom to become 'virtual citizens' within these digital realms.44[27] In this process, although internet service providers' terms of service ostensibly define the rights and obligations of contracting parties under the guise of service provisions, they essentially constitute 'pseudo-legal contracts'.[27] Although community regulatory norms must adhere to supraordinate national laws and refine rules to adapt to specific commercial contexts, their practical implementation remains highly dependent on the technical spaces autonomously constructed by platforms.

As Nancy Kim and Jeremy Telman indicated, 'Internet giants create virtual spaces that consumers inhabit. When consumers do so, they enter a world in which private companies are both service providers and regulatory bodies that govern their own and their users' conduct'.[28] Within these virtual spaces enabled by private technologies, quasi-legislation that mirrors real-world laws has emerged.

B. Quasi-enforcement power: Privatized execution

Generality is the essential feature of legal norms. Traditional administrative enforcement relies on law enforcement officers' professional knowledge and specific judgment about individual case contexts. Similarly, the regulatory practices of internet service providers over cyberviolence also requires transforming abstract legal concepts into technical implementation approaches. 'When human beings use an object, there arises a "technologically mediated intentionality"-a relation between human beings and the world mediated by a technological artifact'[29]. As a technical intermediary, algorithms leverage natural language processing to analyse normative elements in legal texts, such as constitutive elements of crime, and they then rely on machine learning to construct data models recognizable by algorithms, such as semantic feature vectors and behavioural pattern weights. This process thus develops a technical parameter system and establishes a fully automated regulatory system that includes classification standards, feature databases, and case sample repositories.

This technology-mediated regulatory model effectively grants platforms quasi-enforcement powers analogous to 'private administration'. Through 'normative translation', platforms practically transform abstract legal obligations into operational technical execution logic. However, platforms

are driven by technological capital and do not possess a public welfare nature, nor is it strictly constrained by public law procedures and lacks democratic legitimacy, which renders it difficult to effectively achieve the original intention of maintaining network order and safeguarding public interests.

C. Quasi-judicial power: Privatized adjudication

As Petter Törnberg observes, platforms leverage their technical advantages to assume the role of dispute adjudication traditionally held by public institutions.[30] According to the research of Jamila Venturini et al., in terms of judicial remedies, as many as 95% of the sample platforms stipulate mandatory arbitration in their terms of service and explicitly require users to waive class actions[31] Moreover, 83% of the platforms unilaterally specify a particular jurisdiction to handle judicial disputes. Venturini et al.[31]further note that this unilateral control enables platforms to define the dispute resolution mechanism independently, such as by constructing internal arbitration processes; on the flip side, this greatly restricts the channels through which users can seek remedies via the conventional judicial system.

Furthermore, when handling disputes through automated systems, online platforms often make judgments based solely on data characteristics such as keyword density and interaction frequency while ignoring causal relationships, users' subjective intent, and the specific context of speech. Account suspension is an example. Platforms are not required to fulfil due process obligations such as notifying reasons or hearing pleadings, and they directly enforce penalties and handle disputes through algorithmic decisions. Therefore, asymmetry in power and obligations is produced: while the relationship between platforms and users is defined as a civil relationship in Chinese law, platforms nevertheless assume the role of 'governors of public spaces'.

Accordingly, the governance of cyberviolence in China involves a 'public-private collusion' between governmental regulation and platform technological forces. This collusion is actualized through legislative authorization, technological dependency, and data embedding, giving rise to a digital Leviathan that restructures the power dynamics in cyberspace. The state retains formal norm-making authority while delegating the technical parameters for rule refinement and enforcement powers to platforms, enabling platforms to fill legal gaps through technological capabilities. Malcolm Langford criticizes this power transfer as 'digital welfare states masking neoliberal agendas'.[32] The combination of regulatory outsourcing and the publicization of private power enables the government to retreat behind the veil of 'technological neutrality' and shift accountability risks to platforms on the one hand while platforms employ 'standard-form contracts' and 'trade secrets' as shields on the other hand. This transforms algorithmic black boxes into unaccountable power barriers and results in mutual responsibility avoidance.

5. CONCLUSION

Cyberviolence, as the dark side of internet penetration in the contemporary world, is certainly in need of governance. However, this paper uses China's cyberviolence governance as a window to observe the transformation of public space, public interests and public powers, and emphasizes the tensions between the necessity of governance and the civil liberties of Chinese netizens. Through the lens of cyberviolence governance, this paper argues that state authorization, platforms' technological monopolies and the digital Leviathan engendered by their collaboration have undermined the principles of the separation of powers, transparency and proportionality. Platforms, as the new governor, are not held accountable through traditional constitutional constraints but derive legitimacy from technological efficacy.

For future improvement, the Chinese legal regulation of cyberviolence must develop substantive laws or judicial interpretations to clearly distinguish permissible risks from impermissible risks. When imposing criminal liability, defendants must be granted the right to present counter-evidence to

activate the proviso function of Article 13 of the Criminal Law, such as algorithm logs and user behaviour data, to prove that their actions did not create an imminent risk or actual possibility of legal interest infringement. Interdisciplinary expert committees must be introduced to conduct systematic audits of platform technical architectures and to disclose technical error rates in carrying out content moderation. Finally, the key to addressing the digital Leviathan lies in reconstructing the constitutional order of the digital age through transparency checks and participatory empowerment. Legislative frameworks for algorithmic transparency must mandate platforms to reveal core parameter sets and provide appeal and correction mechanisms for algorithmic decisions. This can ensure that regulatory agencies and independent technical auditors are able to penetrate the 'technological black box' to guarantee the alignment between algorithmic logic and public interests. Platforms should hold public hearings when formulating their terms of service and widely engage user representatives and legal experts in their rule-making discussions. In the future, a multi-stakeholder 'anti-digital Leviathan' governance coalition should be implemented to integrate government officers, users, and independent institutions to build a regular consultative mechanism for checking platform powers.

REFERENCES

[1] Awiria O, Olweus D and Byrne B. Bullying at school-What we know and what we can do [J]. British Journal of Educational Studies, 1994, 42(4): 403-406.

[2] Smith P K, Mahdavi J, Carvalho M, et al. Cyberbullying: Its nature and impact in secondary school pupils [J]. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 2008, 49(4): 376-385.

[3] Menesini E, Nocentini A. Cyberbullying definition and measurement: Some critical considerations [J]. Zeitschrift für Psychologie/Journal of Psychology, 2009, 217(4): 230-232.

[4] Peter I K, Petermann F. Cyberbullying: A concept analysis of defining attributes and additional influencing factors [J]. Computers in Human Behavior, 2018, 86: 350-366.

[5] Xie Dengke. Public-private collaborative governance model for cyber violence crimes [J]. Science of Law (Journal of Northwest University of Political Science and Law), 2023, 41(5): 92-101.

[6] Xu Youping. The invisible aggressive fist: Features of cyberbullying language in China [J]. International Journal for the Semiotics of Law-Revue internationale de Semiotique juridique, 2021, 34(4): 1041-1064.

[7] Adediran A O. Cyberbullying in Nigeria: Examining the adequacy of legal responses [J]. International Journal for the Semiotics of Law-Revue internationale juridique, 2021, 34(4): 965-984.

[8] Lovelock M. Catching a catfish: Constructing the good social media user in reality television [J]. Television& New Media, 2016, 17(4): 203-217.

[9] Zhao Hong. The connection between civil execution responsibility and integration in cyber violence cases [J]. Northern Legal Science, 2023, 17(5): 5-20.

[10] Hassan F M, Khamis A, Ewis A. Cyber violence pattern and related factors: Online survey of females in Egypt [J]. Egyptian Journal of Forensic Sciences, 2020, 10(6): 1-7.

[11] Volodina v. Russia(No. 2), No. 40419/19, ECHR. (2021).

[12] Balkin J M. Free speech in the algorithmic society: big data, private governance, and new school speech regulation [J]. University of California, Davis Law Review, 2018, 51: 1149-1210.

[13] Ilin A. A new cyberbullying law? Extension of legal interpretations in China and Russia [J]. International Journal of Law, Ethics and Technology, 2022, 2: 56-86.

[14] Jing Lijia. The development of the system of network service providers' cyber violence governace obligations [J]. Northern Legal Science, 2023, 17(5): 37-50.

[15] Zhu Xiaoyan. Towards media justice: Platform role and law implementation of cyber violence information governance [J]. Journal of Central South University (Social Sciences), 2024, 30(1): 50-62.

[16] Beck U. Risk society: Towards a New Modernity [M]. Ritter M Trans. London: SAGE Publications, 1992.

[17] Jiang Ying. The direction and limits of criminal law regulation on the risks of artificial intelligence deepfake technology [J]. Nanjing Journal of Social Sciences, 2021, 9: 101-109.

[18] Supreme People's Procuratorate of the People's Republic of China. The 34th Batch of Guiding Cases of the Supreme People's Procuratorate [EB/OL]. (2022).Available at: https://www.spp.gov.cn/spp/jczdal/202202/t20220221_545125.shtml(accessed26 August 2025).

[19] The First Intermediate People's Court of Beijing Municipality. Criminal judgment: Case No. (2016) Beijing 01 Criminal Appeal No. 592.

[20] Zhang M K. Reflection on several theoretical issues of criminal law in the 'risk society' [J]. Studies in Law and Business, 2011, 28(5): 83-94.

[21] Sunstein C R. Free speech now [J]. The University of Chicago Law Review, 1992, 59: 255-316.

[22] Stevens J P. The freedom of speech [J]. The Yale Law Journal, 1993, 102: 1293-1313.

[23] Douek E. Governing online speech: From 'post-as-trumps' to proportionality and probability [J]. Columbia Law Review, 2021, 121(3): 759-833.

[24] Klonick K. The new governors: The people, rules and processes governing online speech [J]. Harvard Law Review, 2018, 131: 1598-1670.

[25] Dvoskin B. Expert Governance of Online Speech [J]. Harvard International Law Journal, 2023, 64(1): 85-135.

[26] Lessig L. Code: Version 2.0 [M]. New York: Basic Books, 2006.

[27] Van Dijk J, Poell T, De Waal M. The Platform Society: Public Values in a Connective World [M]. Oxford: Oxford University Press, 2018.

[28] Kim N S, Telman D A. Internet giants as quasi-governmental actors and the limits of contractual consent [J]. Missouri Law Review, 2015, 80: 723-770.

[29] Verbeek P P. What Things Do: Philosophical Reflections on Technology, Agency, and Design [M]. University Park, PA: Penn State Press, 2005.

[30] Tomberg P. How platforms govern: Social regulation in digital capitalism [J]. Big Data & Society, 2023, 10(1): 1-13.

[31] Venturini J, Bruno F, Dantas M T, et al. Terms of Service and Human Rights: An Analysis of Online Platform Contracts [M]. 2nd ed. Rio de Janeiro: Editora Revan, 2016.

[32] Langford M. Taming the digital leviathan: Automated decision-making and international human rights [J]. American Journal of International Law Unbound, 2020, 114: 141-146.