

Social Media Censorship: Overprotected Platforms and Legal Liability

Zhen Zeng

Yunnan College of Business Management, Kunming, Yunnan, 650304, China

Keywords: Social media, online harms, platform regulation, free speech, censorship, online safety Bill

Abstract: In the wake of sustained efforts to combat abusive behaviour on social media platforms, it has been demanded that social media companies take greater responsibility for the safety of their users. A piece of a piece of legislation (The Online Safety Bill) was announced by the UK government in 2022 with the aim of creating a safer online environment and protecting users from harmful content. This article critically assesses the extent to which social media platforms are overprotected in contemporary society, and whether current regulations are sufficient solutions to the issues around online harms and abhorrent content. A comparative perspective of various media law systems is provided to explore the legal regulations, jurisprudential norms, and societal impacts of different legal systems on the media. The regulation of social media platforms is a complex issue that has sparked debate over the balance between protecting speech rights and preventing abusive behaviour. However, it is not as simple as a binary choice between statutory oversight and unregulated speech, as moderation laws can be too rigid or too vague and may lead to over-censorship or over-protection. There is a need to refine clear boundaries for content governance and liability exemption, while maintaining the original intentions of legislation, aligned with an ethic approach.

1. Introduction

Social media platforms used to receive its positive evaluation of contributing to media democracy, bringing active participating audiences and citizens' expression into public insight, in the time of Web 2.0, a term that "describes the current state of the internet, which has more user-generated content and usability for end-users compared to its earlier incarnation, Web 1.0 world" ^[1]. In recent years, the proliferated online harms cast the spotlight on these platforms that have been accused of acquiesce in abhorrent online materials. A global agreement seems to be reached to hold platforms accountable for their user' contents ^[2].

A critical assessment of US internet censorship will be taken in this article, accompanied by the introduction of the UK'S Online Safety Bill ^[3], an example of recent regulatory approach to tackle online hazards.

2. How the Internet and Platforms are Regulated?

There have been more than two decades since the statutory regulation of public expression on the Internet and most of the laws in the United States were established when few social media platforms existed^[4]. The early laws crafted and passed before the times of the Web 2.0 world, didn't include or anticipate social media, notwithstanding they are now still known as the regulatory framework that most social media platforms inherited^[4]. Gillespie describes the Internet as “a rare example of a true, modern, functional anarchy”^[4] and the nature of it provides simple freedom with no necessity of censorship, especially those with the power of the state. The web was considered as “unregulated and unregulatable”^[4]. Nevertheless, On the contrary to the utopian notion of a free, open community accompanied by democracy and empowered, engaged media users, the actual online environment has been accused of allowing “online harms”^[5]: verbal attacks, harassment, white supremacy, pornography, intrusion into the areas of privacy and copyright. Critical cultural communication scholars note that in the contemporary media sphere, user-generated contents have become the common practice, improving democracy in many ways via transforming users into active participants^[6]. This new media practice by platforms also provoked online harms and the calls for “a duty of care” seems to be a consensus between nations to hold social media platforms liable for the misuse of their services^[7]. Early attempts to fight against online published illicit materials were aimed at intermediaries, the ISPs or service providers, rather than those individual users^[4]. One of the underlying reasons to hold intermediaries liable for infringing contents is that it was difficult, in the 1990s, to directly identify the individual person to blame for, which indicates “liability here was something more like a question of convenience”^[4].

A gathering of laws from diverse aspects, the Communications Decency Act, “the first crafted legislative response to online pornography”, “was passed in 1996 and judged as unconstitutional less than one year later by the Supreme Court”^[4]. The bans on the dissemination of obscene materials were aborted for impeding the right of freedom of speech granted by the First Amendment and Section 230 in this act eventuated to remain^[8].

The U.S. contemporary legislative guidelines are a unique exception compared to the other nation. Unlike the UK's Online Safety Bill, the core concept centred around U.S. telecommunication common law is the management autonomy of intermediaries or content hosts rather than “a duty of care”, emphasizing on the roles of intermediaries only as service providers and content hosts, neither “the publisher” nor “the speaker”^[4]. Section 230 of US telecommunication law, on one hand, “ensures that intermediaries that merely provide access to the Internet or other network services cannot be held liable for the speech of their users”^[4]; on the other hand, protects the intermediaries' right to police: they are protected on user-management, deleting user' contents, removing users, setting up their own service regulations and agreement terms^[4].

Section 230 is rendered as intermediaries, service providers-privileged, “allowing them to have their own free speech and everyone's else too”^[4]. As circulated in Gillespie's work, the law provides a robust safe harbour for intermediaries: “Intermediaries who did nothing were immunized in order to promote freedom of expression and diversity online; intermediaries who were more active in managing user-generated content were immunized in order to enhance their ability to delete or otherwise monitor ‘bad’ content.”^[4]

The law still applies to how social media platforms are regulated although it was born in the age previous to the proliferation of user-generated media. Despite a more complex media landscape nowadays, platforms, such as some tech giants, Meta Platforms, Twitter Inc., are under the protection of the safe harbour. Moderation is a choice, not an obligation and in this framework, social media platforms enjoy massive exempted space from liability. Dated back to the times of Web 1.0, the legislation “ensuring intermediaries the most discretionary power”^[4], were regarded as the facilitators

of technology industries, removing impediments from free competition within the telecommunication market, contributing to the later-on prosperity of user-generated media [8]. The law was born in the academic context in which media democracy was strongly craved and audiences were constructed as active participants^[9]. “The internet could encourage participatory democracy in a revitalised public sphere”^[10] and Section 230 “has been lauded as the most important law protecting internet speech and called perhaps the most influential law to protect the kind of innovation that has allowed the Internet to thrive”^[11]. Despite the critiques at the early stage for infringing individualistic freedom of speech, from a constructionism perspective, the law provided the broad safe harbour for an systematic environment structured by a variety of speakers and perspectives^[4]. It turned out that the websites, platforms are shielded from being aimed at in defamation lawsuits, or other torts committed by users^[11], as long as these service providers “have no actual knowledge” of the contents published or distributed^[4].

3. Overprotected Platforms?

The performance of social media platforms and other interactive services, such as YouTube, has been subject to critical evaluation in recent academic scholarship. Multiple aspects have been criticized, including the datafication of users, the commodification of audiences in political economy research^[6], online manipulation, misinformation, gender-based violence, and the flourishing of terrorism on Twitter and Facebook^[12], as well as the anti-social media campaign in recent years^[13].

These problems, most of which intensively occurred in the last decade, provoke a question concerning public interests: Are the platforms overprotected under the broad immunity and what could be proposed to regulate platforms in the contemporary social context?

The normative assessments of Section 230 contribute to the discussion on freedom of speech and the burgeoning computer interactive services^[12]. Speaking of the law in modern practice, a plenty of lawsuits against Facebook for complete inaction, not to remove harmful contents, defamatory speech, or hate speech for instances, were proven to be plaintiff-unfriendly, with the defence through Section 230 that the harms were caused by the third parties while using Facebook, not by Facebook itself^[12]. In the case of the murder of Godwin vs Facebook^[14], the estate of the victim Robert Godwin Sr. claimed for pain and suffering damages after Godwin was shot and killed with a firearm by Steve Stephens, who post a message with ill intentions on Facebook and broadcast the murder via “Facebook Live” on the same day^[14]. The estate sued Facebook for common law negligence claim: Facebook failure’s to warn the victim of the “dangerous propensity of which Facebook was aware through its data-mining practices”^[14] and “the civil-right-of-recovery claim based on R.C. 2921.22 and 2307.60 stemming from Stephens’s message posted to his social network page “minutes” before Robert Godwin’s tragic and senseless murder”^[14]. The claims were declined with the application of Section 230 to this case, denied the claim that Facebook has a duty to report Stephens for making a terroristic threat^[14].

There have a handful of cases arising from, either Facebook’s action to remove a claimant’s account, suppressing traffic to the users’ pages by pushing the paid advertisements to the top, the terroristic use of the platform that was perceived as Facebook’s direct participation into international terrorism by providing a communication platform and assistance materials through algorithms^[12].

According to Dimitroff’s 2021 study^[12], Facebook has been named as defendant in 34 cases since 1996, when Section 230 was invoked. Only one case has been successful. The Fraley vs Facebook case^[12], with a rare success, did, however, replace Section with a common law that subjects the platform to civil liability. This was because the plaintiffs claimed that Facebook “was actively developing commercial content that violated their legal right of publicity”^[12] instead of alleging its accountability for the harm that occurred on the site.

Undoubtedly, the cases and the verdicts explain well why Facebook declined to delete fake, untrustworthy news that deluged on the platform during the 2016 US election campaign ^[15]. Section 230 has progressively become the weapon to defend almost every kind of civile claim for the harms coming from the user-generated contents, even though interpretations of Section 230 in courts might have deviated from the original intent, only to broadly provide platforms with immunity to almost “all state statutory and common law causes of action” ^[12]. Failing to prove the platforms’ proximate causation of the harms, individuals are effectively blocked from using the private causes of action to recover against social media platforms. Therefore, Section 230 is criticized for “enabling civil rights violation” ^[12].

It is undeniable that social media platforms, though framing themselves as “open, impartial and noninterventionist” ^[4], take advantages of the protection part of Section 230, through advantageous management mechanism sticking to their agenda. It has been argued that the privileged rights of providers override public interests. It might be inappropriate to subscribing to the idea that the law should be blamed for all existing problems of the platforms, ranging from monetising users’ data to social media abuse, as the world cannot be interpreted by distinct, identified and isolated forces ^[16]. The adoption of Section in lawsuits, however, encourages platforms to pursue commercial gains without being socially responsible.

3.1 Call for a More Stringent Regulation

Legislative regulation of social media platforms and other online services has been an international consensus, in addition to the U.S., as the Internet are increasingly integrated into people’s life and platforms has exerted greater influences on the public landscapes. The UK’s Online Safety Bill is one example of current statutory approach to tackle online hazards, setting out a new regulatory framework for Internet services ^[16].

Most of major social media platforms, based in the United States, have covered their user-generated media services in many nations around the world, UK included [4]. On 12 May 2021, the Draft Online Safety Bill was tabled, on the basis of the Government’s “Online Harm White Paper” in 2019, and then the Online Safety Bill was published on 17 March 2022 ^[16].

Now the central thematic is the “duty of care”, as the Bill will hold platforms responsible for the content they host and the purpose of the bill is “to create a new regulatory regime to address illegal and harmful content online” ^[2]. A general duty in Clause 12 in the Bill is applicable to all online services to “have regard to protecting users right to freedom of expression and protecting users from unwarranted infringements of privacy ^[17] and the “user-to user” services (interact with each other online, such as the manner in which users typically interact on platforms like Facebook and Twitter ^[2] and search engine services that links in UK are included in the scope of the Bill. “Once enacted the legislation will require Ofcom to issue codes of practice which will outline the systems and processes that companies need to adopt to fulfil their duty of care” ^[17]. As an online safety regulator, “It will have the power to fine companies up to £18 million, or 10 per cent of annual global turnover, whichever is higher, if they are failing in their duty of care”^[17].

3.2 Evaluation of Recent Regulatory Efforts

The Bill and its other international equivalents, for instance, Digital Services Act of the European Union, are the product of official legislative responses to the abhorrent phenomena in user-generated media, in compliance with the pressing demands on moderation of platforms. The roles of service providers are extended from the focus in Section 230, to the necessity of being socially responsible in an ethical view, when they ground massive users and “enable complex networks of relationships between users, and in doing so create social spaces” ^[5], more than host content. “Regulation as

outlined in the bill is a significant step towards ensuring the transparency and accountability of tech companies “[8].

Meanwhile, some features of the Bill have triggered a heated discussion among researchers. It is criticized for its wide duties, vague demarcation and general censorship. The Bill requires social media platforms to remove all illegal content in the first place, prevent adult and children from seeing it [3], but it also includes “the legal but harmful” materials in its scope of duty. The definition of harm is ambiguous: “The provider has reasonable grounds to believe that the nature of the content is such that there is a material risk of the content having, or indirectly having, a significant adverse physical or psychological impact on a child (adult)”[17]. The added duty overburden providers: practically speaking, “moderation requires a great deal of labour and resources” [4], especially with an implicit definition, the platforms possibly examine users’ content through minimum standards, from a utilitarian perspective.

The wider duty risks a tendency to wider censorship, a reversal of the broad immunity ensured by Section 230, and in consideration of the power entitled to Ofcom and the Minister of State, the Bill is seen as a possible democratic deficit. The power of the Minister to “direct the Codes of Practice”, implies that “they are effectively granted the authority to determine how user-to-user services control speech on their platforms” [2], and this turns out as the same as Section 230 in modern practice, deviating from the initial vision.

4. Conclusion: What we should do?

It is not a binary formulation of statutory oversight against the guilt of unregulated speech and platforms’ discretion. The problematic parts rest on the “overly prescriptive and rigid, or conversely, overly vague and sweeping” moderation laws [7]. It is the matter of interpretation of the laws that give rise to over-censorship or over-protection. Some scholars argue that there is a necessity of refining clear boundaries for content governance and liability exemption, adhering to the original intentions of legislation.

Trengove [2] brings forward an ethical design approach to deal with the concerns The Online Safety Bill attempts to address, rather than the Code of Practices model. Different from the newly published law, Section 230 has been in use for more than 25 years in spite of recent calls for reform. When it comes to resolve controversies and problems arising from the law, it might be better to inspect them within the internal context of the U.S. As stated in Dimitroff’ study [12], “the law itself is not an ideological proposition” or “representative of a set of values”, but “an affirmative defense to litigation that has expanded beyond its intended scope through decades of judicial interpretation”.

References

- [1] Kenton, W. (2022) *What Is Web 2.0? Definition, Impact, and Examples*. Available at: <https://www.investopedia.com/terms/w/web-20.asp> (accessed 31 December 2022).
- [2] Trengove, M., Kazim, E., Almeida, D., Hilliard, A., Zannone, S. and Lomas, E. (2022) *A critical review of the Online Safety Bill*. Available at: <https://doi.org/10.1016/j.patter.2022.100544> (accessed 7 January 2023).
- [3] *A guide to the Online Safety Bill* (2022) Available at: https://www.gov.uk/guidance/a-guide-to-the-online-safety-bill?gclid=CjwKCAiA5Y6eBhAbEiwA_2ZWISCAxiWQRQwnF6UbaMefeQITMOLUk5qv3qQSKvxJNaJmkVn77X_CtBoCEYoQAvD_BwE (accessed 12 January 2023).
- [4] Gillespie, T. (2018) *Custodians of the Internet*. New Haven: Yale University Press.
- [5] Price, L. (2021) ‘Platform responsibility for online harms: towards a duty of care for online hazards’, *The Journal of media law*, 13(2), 238–261.
- [6] Jennes, I., Pierson, J. and Van den Broeck, W. (2014) ‘User Empowerment and Audience Commodification in a Commercial Television Context’, *The Journal of Media Innovations*, 1(1), 70–85.
- [7] Burns, H. (2021) *Online Abuse: Why Management Liability isn’t The Answer?* Available at: <https://www.openrightsgroup.org/blog/online-abuse-why-management-liability-isnt-the-answer/> (accessed 8 January 2023).

- [8] Canadian Centre for Child Protection Supports UK's Online Safety Bill (2021) Available at: <https://protectchildren.ca/en/press-and-media/news-releases/2021/uk-online-safety-bill> (accessed 10 January 2023).
- [9] Carpentier, N., Schroder, K.C., and Hallett, L. (2013) *Audience Transformations: Shifting Audience Positions in Late Modernity*. New York: Routledge.
- [10] Thornton, A.L. (2001) *Does the Internet create democracy?* *Ecquid Novi*, 22(2), 126-147.
- [11] *Harvard Law Review* (2018) 'Section 230 As First Amendment Rule', *Harvard law review*, 131(7), 2027–2048.
- [12] Dimitroff, K. (2021) 'Mark Zuckerberg, Joe Manchin, and ISIS: What Facebook's international terrorism lawsuits can teach us about the future of section 230 reform', *Texas law review*, 100(1), 153–188.
- [13] Lawrence, K. (2021) ALL RED ALL EQUAL Man Utd launch huge anti-social media abuse campaign 'See Red' and urge use of online reporting system to combat it. Available at: <https://www.thesun.co.uk/sport/football/14529260/man-utd-see-red-anti-social-media-abuse-campaign/> (accessed 11 January 2023).
- [14] Goldman, E. (2020) *Facebook Doesn't Have a Duty to Prevent a Murder—Godwin v. Facebook*. Available at: <https://blog.ericgoldman.org/archives/2020/10/facebook-doesnt-have-duty-to-prevent-a-murder-godwin-v-facebook.htm> (accessed 10 January 2023).
- [15] *Platform Responsibility Reform: Observation on Amendments to Section 230 of the Communications Decency Act of the United States* (2020) Available at: <https://www.secrss.com/articles/26591> (accessed 9 January 2023).
- [16] Johnson, L. (2022) *Online Safety Bill, Remaining Stages, House of Commons, 5 December 2022*. Available at: <https://www.local.gov.uk/parliament/briefings-and-responses/online-safety-bill-remaining-stages-house-commons-5-december> (accessed 6 January 2023).
- [17] Coe, P. (2021) 'Misinformation, disinformation, the Online Safety Bill, and its insidious implications for free speech', *Communications law (Haywards Heath, England)*, 26(3), 127-129.