

The regulation of content moderation in the DSA: can we trust the due process?

Abstract

The DSA relies heavily on procedural and transparency rules to regulate content moderation carried by the online content providers at the stages of rules-making, rules-applying and rules-appealing. This comprehensive regulation framework provides users with due process rights to contest with the providers and affords the latter with wide discretion. While the progress achievements made by the DSA is remarkable, this essay argues that the transparency rules in rules-making are not substantive enough for users to safeguard their fundamental rights; the wide discretion of the providers in rules-applying might result in inappropriate enforcement; the internal and external contestation in rules-appealing also might fail to offer effective remedies for users on account of their inherent bias and operation issues. Given these flaws of the due process regulation in the DSA, it further advocates that more detailed guidance shall be provided for the providers to safeguard fundamental rights of users in their rules-making and rules-applying, some voluntary standard should be placed to reduce the risk of over-and under enforcement, and a “crow-judging” redress system could better address disputes. In doing so, the DSA could have more checks and balances for the providers to achieve the goal of content moderation regulation.

Keywords: Digital service act; due process obligation; rules-making; rules-applying; rules-appealing

Volume 12 Issue 4 - 2024

Na Jiang, Rong Han

Law school, Beijing Normal University, China

Correspondence: Na Jiang, Law Professor, Beijing Normal University in China, China, Email na.jiang@bnu.edu.cn

Received: September 11, 2024 | **Published:** October 03, 2024

Introduction

The intermediary platforms such as Facebook and TikTok are regarded as online content providers (hereinafter “providers”) since they control over the availability, visibility and accessibility of online information. The providers are so powerful that they almost define the public sphere and the opportunities for citizens both to express themselves and access the information.¹ Driven by commercial interests and regulation, the providers make their own rules to regulate the information on their platforms, so as to remove some information which violate their platform rules or the law. This whole process is called content moderation, regarded as a “privatized hierarchical bureaucracy that applies legislative-style rules drafted by platform policymakers to individual cases and hears appeals from those decisions.”² However, this privatized regulation of information arises many concerns and critics.

Some think the providers and their arbitrary content moderation behaviour are potentially endangering the fundamental rights such as the freedom of speech of users.³ Others concern about the dissemination of disinformation, hate speech, terrorism on platforms at an unprecedented scale and speed.⁴ These issues are pressing and imminent. In this backdrop, the European Union rolled out the Digital Service Act (DSA) in 2023 to hold the providers more accountable. In addition to the precedent rules set by the e-Commerce Directive,⁵

the DSA also set out a series of procedural-oriented obligations for the providers, ranging from clarifying platform rules, receiving notice and taking action to providing internal and external redress measures. It has been believed that the DSA will empower user’s due process rights and safeguard users’ fundamental rights.⁶ There are some doubts about it as well. For example, Berrak⁷ pointed out that the effectiveness of this approach and whether the DSA can fulfill its promise of a new regime will only become apparent once the rules are implemented and put into practice.⁸ Therefore, the advancement made by the DSA is revolutionary, while it is not free from doubt in terms of its effectiveness.

This essay aims to exam the effectiveness of the procedural-oriented regulation of the content moderation in the DSA and explore some possible routes to improve its efficacy. To achieve this goal, the first part will give a full picture of the due process obligations posed to the providers in the DSA; then it will analysis the inadequate regulation of the rules-making power, wide discretion given to private enforcement, limitations of the internal and external redress mechanisms. After that, some constructive suggestions will be further explored and a conclusion is expected in the end.

The due process obligations of the providers in the DSA

The DSA adopts a due process regulation, setting tiered procedures and obligations for different size providers to follow at different stages.

¹Tourkochoriti Ioanna. The Digital Services Act and the EU as the Global Regulator of the Internet. *Chicago Journal of International Law*. 2023;24(1):129–147.

²Evelyn Douek. Content Moderation as Systems Thinking. *Harvard Law Review*. 2022;136(2):526–607.

³Jillian York C. Silicon Values: The Future of Free Speech under Surveillance Capitalism. 2021.

⁴Ben Smith. Inside the Big Facebook Leak. *The New York Times*. 2021.

⁵Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (e-Commerce Directive). OJ L 178. 2000. p. 1–6.

⁶Martin Husovec, Irene Roche Laguna. *Principles of the Digital Services Act*. Oxford University Press. 2023.

⁷Berrak Genc Gelgec. Regulating Digital Platforms: Will the DSA Correct Its Predecessor’s Deficiencies? *Croatian Yearbook of European Law and Policy*. 2022;18:25–60.

⁸Berrak Genc Gelgec. Regulating Digital Platforms: Will the DSA Correct Its Predecessor’s Deficiencies? *Croatian Yearbook of European Law and Policy*. 2022;18:25–60.

Rules-making: public and clarify terms and conditions

The DSA requires all of the providers to make their platform rules, also called terms and conditions (T&Cs), visible and understandable to users. The T&Cs shall use “clear, plain, intelligible, user-friendly and unambiguous language”, disclosing information on the policies, methods, procedures used for the content moderation pursuant to Article 14(1) of the DSA. If the platforms employed algorithmic decision-making to moderate content, this information shall also be provided for users. The Article 14(5) of the DSA adds further obligations to the very large online platforms (VLOPs) and very large online search engines (VLOSEs), requiring them to public a summary of their T&Cs and redress measures. Furthermore, VLOPs and VLOSEs shall conduct risk assessment of their T&Cs at least once every year under Article 34(2) of the DSA.

Overarchingly, the DSA sets out tiered procedural obligations for providers to clarify and public their platform rules. This greatly enhanced the transparency of content moderation, allowing users to know whether and how their content will be moderated on platforms. However, the DSA does not take away the rules-making power of the providers. Instead, it leaves significant discretion for the providers to make their platforms rules, provided that these rules are clarified and conveyed publicly. This means only some grossly disproportionate T&Cs might violate Article 14 of the DSA.

Rules-applying: notice, action and justify

In the application of T&Cs, the providers shall undertake several responsive and active procedure obligations. The first one is to respond to reports from users. Specifically, the providers shall design user-friendly electronic mechanisms available to users to report illegal content pursuant to Article 16(1) of the DSA. This mechanism shall confirm the receiving of notification from notifies. Upon receiving the notification, the providers shall conduct an assessment to the reported content, making a decision about whether they are illegal or incompatible with their T&Cs in time. With the requirement of Article 17 of the DSA, the providers also need to justify the reasons and grounds for posing any restrictions of the information to all of the affected recipients of the service. If a user had content which is illegal or in contrary with the T&Cs, the providers have to explain to the user about why and how the content is dealt with by the platform, as well as the redress measures.

Furthermore, there are some active and preventive obligations for the providers. Where the providers have acknowledge of suspicious criminal offence, they shall report to the local law enforcement authorities in time pursuant to Article 18 of the DSA. For VLOPs and VLOSEs, they shall conduct risk assessment of their content moderation systems and their enforcement at least once every year pursuant to Article 34(2) of the DSA. To mitigate the risks related to illegal information and information in contrary with T&Cs, VLOPs and VLOSEs need to adapt their content moderation processes and dedicate resources to content moderation. It is worth mentioning that the voluntary inspection and investigation of the platform content does not cause liability pursuant to Article 7 of the DSA. To sum up, the providers shall fulfill some procedure obligations. Nevertheless, the DSA does not give specific and substantial instructions to the providers about how to conduct or evaluate their enforcement. It seems that the way to comply with the DSA is totally in the hands of the providers. This genre of procedural regulation also extended to the stage of rules-appealing.

Rules-appealing: internal and external contestation

The DSA also provide extra procedural obligations for medium-to-large online platforms. The first one is the establishment of an internal complaint-handling system. The providers shall provide free electronic systems for the recipients of the service to appeal the decision taken by the providers. This internal complaint-handling system shall be easy to access and user-friendly. Furthermore, Article 20(4) of the DSA requires the providers to handle complaints submitted through their internal complaint-handling system in a timely, non-discriminatory, and non-arbitrary manner. Nevertheless, the providers have the power to suspend users who frequently lodge notices or complaints deemed manifestly unfounded under Article 23 of the DSA. Additionally, this type of blunt suspension procedure shall be set out clearly in T&Cs pursuant to Article 23(4) of the DSA.

Another one is the external redress system, which is described as out-of-court dispute settlement under Article 21 of the DSA. The providers shall inform their users the availability of the external dispute settlement certified by EU member states. Likewise, this information shall be easily accessible on their online interface and user-friendly. If the recipients of the service chose to appeal the decision taken by platforms through the external measure, the provider shall engage in good faith to resolve the dispute. Nevertheless, the providers have a right to refuse to participate in the external redress procedure if the dispute had been resolved. It is worth mentioning that the decisions made by the certified out-of-court dispute settlement body is not binding to the providers pursuant to Article 21(2) of the DSA. However, users could always raise the dispute to court without resorting to the above mentioned internal and external redress system.

Interim conclusion

Based on the above analysis, we can conclude that the DSA is built upon a procedural-oriented framework. Most of the regulation rules are procedural in nature.⁹ Disclosing and clarifying rules, explaining decisions and processes, justifying and appealing decisions are the three main due process obligations to the providers.¹⁰ They have responded to some concerned issues of the content moderation practiced by platforms. The disclosure and clarify platform rules increase the transparency of the process of content moderation, which mitigates the information asymmetry between platforms and users. As illustrated by Galantino¹¹ Article 14 of the DSA directly promotes transparency and trust.¹² Furthermore, the notice and action obligation also could better protect the fundamental rights of users since it creates a procedure and time-frame for the report and suspension of content. Additionally, users are provided both internal and external mechanisms to seek redress. In this sense, it has made a huge progress on the protection of due process rights of individuals and creation of a safe, fair digital market.

However, there are just few lenient provisions imposing substantial obligations to providers. Article 14(4) of the DSA merely requires providers to have due regard to the fundamental rights of individuals and enforce their T&Cs in a “diligent, objective and proportionate

⁹Quintais JP, Appelman N, Fathaigh RO. Using Terms and Conditions to apply Fundamental Rights to Content Moderation. *German Law Journal*. 2023;24(5):881–911.

¹⁰Martin Husovec and Irene Roche Laguna. (n 6).

¹¹Sharon Galantino. How Will the EU Digital Services Act Affect the Regulation of Disinformation? *SCRIPted*. 2023;20(1):89–129.

¹²Sharon Galantino. How Will the EU Digital Services Act Affect the Regulation of Disinformation? *SCRIPted*. 2023;20(1):89–129.

manner". This means platforms just need to have due regard to the fundamental rights of users, not necessary to act in this way since the DSA does not impose a direct obligation on them. As worried by Berrak¹³ the effectiveness of DSA is still remain to see since not all of these due diligence obligations will be embraced by platforms.¹⁴ Therefore, it is necessary to examine whether the procedure-oriented rules will actually work and whether there are some potential limitations of the DSA.

An examination of the effectiveness of the due process regulation in the DSA

Opinions are divergent about the effectiveness of the due process regulation of content moderation. For instance, Martin Husovec¹⁵ has a positive attitude towards the DSA, believing that it promotes individual agency by giving users due process rights and tools to against providers.¹⁶ Others also believed that the transparency requirements may indirectly contribute to addressing illegal content, as advertising systems pose a risk of disseminating or financially rewarding harmful or illegal content.¹⁷ In contrast, Zalnieriute¹⁸ criticized the procedural fetishism regulation in the high tech sector because of incomplete disclosure, box-ticking compliance and legitimization of market power of the providers.¹⁹ Obviously, the due process regulation has its merits by empower users, while its limitations should not be neglected given the enormous market power of platform. This sector thus will examine the effectiveness of the procedure-oriented rules in the DSA.

Rules-making: what is visible is not always reliable

The merits of the publicity and clarity of platform T&Cs in the DSA are self-evident. By reading detailed T&Cs, users and authorities could know the policy purpose, procedures, methods and tools used by the providers for content moderation. The T&Cs also provides legal certainty. Once the providers have made and published T&Cs on their website or other channels, it means they will be the rule standard for the afterwards content moderation disputes. The providers shall not filter or suspend content on their platform arbitrarily and inconsistently, and users are also given a tool to against providers. Overall, the transparency and certainty provided by the rules-making in T&Cs lays a foundation for the regulation of content moderation. Nevertheless, what is visible is not always reliable. This essay argues that the mandatory transparency and clarity rules in the DSA are not substantive enough to deliver its effectiveness in reality. The reasons and inadequacies will be illustrated blow.

The first deficit lies in the very fact that the publicity and clarify required by the DSA do not constrain the providers' power over rule-making. The providers act as private rules-maker, dominating the contractual relationship with users through T&Cs and the latter have no say on the formation of T&Cs. Consequently, users have to accept all the clauses in the T&Cs if they wanted to use the service of the provider. This makes no change to the imbalance structural

conditions between platforms and users.²⁰ Moreover, as mentioned above, the DSA just requires the providers to have "due regard" to the fundamental rights of users in their T&Cs, which technically is not binding to them at all. Without the engagement of the users and little interference of the regulation, the unreasonable and unfair T&Cs de facto remain uncontestable.

In addition to giving considerable rules-making power to the providers, the transparency rules in the DSA are insufficient. It does not adequately tackle the transparency of human content moderation concerning disinformation, such as fact-checking organizations and internal security teams, which still maintains the lack of transparency in recommender systems, contextualization tools, and the regulation of coordinated inauthentic behavior.²¹ For instance, questioned that Article 14 of the DSA does not indicate that whether such types of content moderation like ranking, recommending and demonetizing content shall be informed to users.²² Additionally, the way the providers to interpret their T&Cs remains unknown to the users. Article 14(4) DSA does not provide guidance on how to weight the different interest and rights of different stakeholders in content moderation. This greatly undermines the certainty of rules and fails to supply meaningful transparency for users.

The last shortfall is that the transparency itself is not necessarily equal to the protection of rights or a substantial change in practice. First and foremost, disclosing T&Cs may fail to deliver what they have promised outcomes and produce unintended consequences. For example, Birchall²³ indicated that as more information becomes accessible, citizens may perceive a decrease in their agency and ultimately feel less informed.²⁴ Even worse, users are not really looking at these T&Cs; even though they do, they still lack the necessary legal and technology skills to actually understand the meaning of the piles of clauses.²⁵ In this case, the declaration of T&Cs does not empower users in a meaningful way, which in turn provides a justification for the content moderation by the providers. For this reason, it has been criticized that the publicity and transparency disguised the substantive fairness and legitimacy of content moderation with procedural rules.²⁶ As Alloa²⁷ discovered, transparency has evolved into an uncontested principle because it claims to regulate only the manner, not the substance, of social interactions.²⁸ Therefore, the disclose of T&Cs is not enough to make a difference to safeguard the fundamental rights of users. Overall, although the DSA has improved the transparency and publicity in the process of rules-making, we can conclude that (1) the providers' rules-making power is barely constrained; (2) there are still some opaqueness of human content moderation and the interpretation of T&Cs; (3) the transparency itself does not lead to any more substantial changes. These inevitably have a negative influence on the application and enforcement of T&Cs in practice.

²⁰Quintais JP, Appelman N et al. 2023;908(8).

²¹Zalnieriute Monika. (n 14).

²²Quintais JP, Appelman N et al. 892(8).

²³Birchall C. Introduction to 'Secrecy and Transparency: The Politics of Opacity and Openness. Theory, Culture & Society. 2011;28(7-8):7-25.

²⁴Birchall C. Introduction to 'Secrecy and Transparency: The Politics of Opacity and Openness. Theory, Culture & Society. 2011;28(7-8):7-25.

²⁵Chris Reed, Laura Edgar. Consumer Protection in the Cloud. In Christopher Millard (editor). Cloud Computing Law. 2nd edn. Oxford Academic. 2021.

²⁶Zalnieriute Monika. (n 14).

²⁷Emmanuel Alloa, Dieter Thoma. Transparency, Society, Subjectivity: Critical Perspectives. London: Palgrave Macmillan. 2018.

²⁸Emmanuel Alloa, Dieter Thoma. Transparency, Society, Subjectivity: Critical Perspectives. London: Palgrave Macmillan. 2018.

¹³Berrak Genc Gelgec. (n 7).

¹⁴Berrak Genc Gelgec. (n 7).

¹⁵Martin Husovec, Irene Roche Laguna. (n 6).

¹⁶Martin Husovec, Irene Roche Laguna. (n 6).

¹⁷Berrak Genc Gelgec. 48(7).

¹⁸Zalnieriute Monika. Transparency-Washing" in the Digital Age: A Corporate Agenda of Procedural Fetishism. UNSW Law Research Paper No. 21-33. 2021;8(1):39-53.

¹⁹Zalnieriute Monika. Transparency-Washing" in the Digital Age: A Corporate Agenda of Procedural Fetishism. UNSW Law Research Paper No. 21-33. 2021;8(1):39-53.

Application and enforcement: the risk of process manipulation

After making T&Cs, the providers also need to establish a notice and action mechanism to apply and enforce its platform rules. Basically, the providers are acting like an administrative body that enforces its regulation within a specified area, but limited by the principle of legitimacy and appropriateness. Looking at the bright side, the DSA gives users due process rights since the providers are not allowed to take down their contents abruptly without a legitimate explanation and proper procedure. But the thing is that, the DSA does not offer detail guidance as to how to enforce the private power, and it also fails to limit the enforcement of the private power with adequate balances and checks from outsiders. This could result in the risk of process manipulation, posing the fundamental rights of users at a great stake.

First, the DSA gives wide discretion to the providers in rules-applying, which could lead to disproportional enforcement and fundamental rights degradation. The DSA does not give a guidance on what is legal or illegal on platform, nor indicate how to strike a weigh the fundamental rights and interests of stakeholders. Instead, it barely examines the content of the T&Cs. Without substantial constraint or guidance from the DSA, the providers rely on their own criteria to examine the claimed illegal or improper content on platforms, and decide the intensity of their enforcement. This affords them considerable enforcement discretion and decision-making risk.²⁹ One of the negative consequences is the over-enforcement. For example, Bassini³⁰ pointed out that the exercise of private powers by the providers could be incentivised to remove rather than keep more content, limiting the free speech from a board view.³¹ Similarly, the DSA also empowers entities within the EU to report content deemed illegal under their country's laws, compelling platforms to promptly remove such content.³² This approach may encourage platforms to remove illegal content that might be protected in other countries.³³ Consequently, the fundamental rights of users would be severely encroached by the private enforcement.

This is closely related to another danger, which is putting the fundamental rights of users secondary to the business decisions. The providers, more often than not, would regard their business interests as a priority when they are carrying out content moderation, enabling the protection of fundamental rights conditional upon business considerations.³⁴ From an economic standpoint, it is rational for platform operators to accept complaints at face value and promptly remove content.³⁵ This action minimizes costs and mitigates the legal risks of litigation, in which the platform operator, lacking first-hand knowledge, is inherently disadvantaged.³⁶ In this sense, some criticized that governance under private control may create ceremonial structures that lend an appearance of legitimacy to corporate actions

but do little to advance the public values at stake.³⁷ It may be argued that Article 14 of the DSA clearly requires platforms to have “due regard” to the fundamental rights of the users. Nevertheless, the fundamental rights are generally unenforceable horizontally against the private firms at the frontlines of enforcement.³⁸ Therefore, the “due regard” obligation might not be strong enough to hold platforms accountable for users' fundamental rights.

The third one is the lack of external balances and checks to the private enforcement. In the notice and action mechanism, the providers are just obliged to respond and take action. However, the DSA does not grant users the right to issue a counter-notice to safeguard their fundamental rights.³⁹ As described above, users are granted the right to file their claim only following a decision made by an intermediary. Likewise, the civil society organizations are only referenced in relation to codes of conduct addressing systemic risks, online advertising codes, and crisis protocols pursuant to Article 35(2), Article 36(1) and Article 37 (3) of the DSA.⁴⁰ Similarly, the Recitals of the DSA just mention the voluntary character of the governmental intervention in analyzing the performance of the provider's enforcement; and the Commission's role is quite weak based on the terms describing their intervention such as “facilitate”, “invite” and “aim to ensure”.⁴¹ Consequently, the enforcement of the providers is rarely checked and balanced by outsiders such as users, civil societies and governments.

To sum up, the DSA delegates the enforcement power to the providers without substantial constraints. It is problematic in the following aspects: (1) the over-enforcement of rules, such as taking down legitimate contents of users; (2) the degradation of fundamental rights for the sake of economic interests; (3) the lack of checks and balances from outsiders. Consequently, the whole enforcement process might be manipulated by the providers, without serious engagement in safeguarding fundamental rights and reducing harms. This in turn will affect its effectiveness and lead to more disputes.

Contestation: justice is more than due process

The notice and action framework designed by the DSA enables the providers to be a quasi-judicial body. The benefits of the private redress have been well elaborated by Martin Husovec et al.,⁴² who argue that it is more efficient and economical than the overloaded court, though they are still available as the last resort in any case.⁴³ Also, the external independent dispute resolution bodies can offer valuable insights into the effectiveness of a platform's content moderation systems, and the standards of clarity, implementation, and enforcement of disinformation.⁴⁴ These arguments are well founded given the sheer scale of the disputes over the removal or suspension of content on platforms. While these positive changes of the contestation are undisputed, this essay also recognizes several aspects to be desired.

First, the internal redress mechanism is questioned on the basis of impartiality and delayed justice. In the internal appeal system, the

²⁹Jörg Wimmers. The Out-of-Court Dispute Settlement Mechanism in the Digital Services Act: A Disservice to Its Own Goals. *Jipitec*. 2022;12(5):381, 392.

³⁰Marco Bassini. Fundamental rights and private enforcement in the digital age. *Eur Law J*. 2019;25(2):182–197.

³¹Marco Bassini. Fundamental rights and private enforcement in the digital age. *Eur Law J*. 2019;25(2):182–197.

³²Dawn Carla Nunziato. The Digital Services Act and the Brussels Effect on Platform Content Moderation. *Chicago Journal of International Law*. 2023;24(1):115–128.

³³*Ibid*.

³⁴Bassini M. 191(23).

³⁵Jörg Wimmers. 387(22).

³⁶*Ibid*.

³⁷Lauren B Edelman. Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law. *American Journal of Sociology*. 1992;97(6):1531–1542.

³⁸Sharon Galantino. 108(10).

³⁹Berrak Gene-Gelgec. 60(7).

⁴⁰Cauffman Caroline, Goanta Catalina. A New Order: The Digital Services Act and Consumer Protection. *European Journal of Risk Regulation*. 2021;12(4):758–774.

⁴¹*Ibid*.

⁴²Martin Husovec and Irene Roche Laguna (n 6).

⁴³Martin Husovec and Irene Roche Laguna (n 6).

⁴⁴Sharon Galantino. 116(10).

providers are unlikely to repel their decisions based on the T&Cs that they've made in the first place unless some manifest fault automatic-making decisions. Put differently, the impartiality is at stake since the providers are both the athletes and referees in this procedure, which could discourage the confidence of users to appeal. Using the German Network Enforcement Act (NetzDG) as an example, it employs comparable 'notice and action' complaint procedures to the DSA, but there are fewer requests for takedowns complaints than anticipated by policymakers.⁴⁵ Another criticism is the shortage of ex ante content moderation redress. As Balkin⁴⁶ acknowledged, digital filtering systems may function as "prior restraints" on speech, inhibiting individuals from speaking rather than penalizing them afterward.⁴⁷ In this case, users are unable to resort to appeal redress in terms of the 'prior restraints', which means the justice coming too late.

Second, the external redress system is not immune from imperfections. The out-of-court dispute settlement are equipped with outsider experts, but its decision is not mandatory to the providers. Even with the decision, the providers could change their T&Cs immediately if they are not satisfied with the result. This consequently enables the whole contestation to be in vain. Another problem comes from its operation. The out-of-court settlement bodies mainly funded by the fees charged from the providers (users) and maybe supported by the member states. Some worry that the operation cost might encourage out-of-court settlement bodies make decisions in favour of the applicants to attract as many cases as possible.⁴⁸ Additionally, the out-of-court settlement bodies might cause further fragmentation and legal uncertainty. As mentioned before, the DSA refrains to offer criteria or standards for the intricate factual and legal assessments and the balancing of rights concerning online speech. This could lead to the decisions varying from bodies to bodies among member states, which in turn attracts users to go to the certified body that favors their claim.⁴⁹ Furthermore, it has been questioned whether it is necessary to have an out-of-court redress mechanism given the fact that the easily accessible and cost-free internal appeal system is barely initiated by users.⁵⁰ Therefore, the external dispute settlement system has some flaws in terms of its non-binding natural, sustainable and consistent operation.

Third, both the internal and external redress systems are designed for individual cases. It has been questioned that the individual decisions are inadequate remedies for users.⁵¹ The outcome is not substantive change, but rather "Process Theater."⁵² While it can be argued that the risk management prescribed by the DSA could prevent massive errors,⁵³ whether the general rules actually work is still a question. Paradoxically, the risk assessment obligations are focused solely on the system's structure rather than its application in specific cases, which also undermines its effectiveness at the individual level.⁵⁴ In this sense, censoring colossal information and communications with automatic systems could potentially suppress speech at a sheer scale,

⁴⁵Counter Extremism Project. ICYMI: New Report on Germany's NetzDG Online Hate Speech Law Shows No Threat of Over-Blocking. 2018.

⁴⁶Jack Balkin M. Old-School/New-School Speech Regulation. *Harv L Rev.* 2014;127(8):2296–2318.

⁴⁷Jack Balkin M. Old-School/New-School Speech Regulation. *Harv L Rev.* 2014;127(8):2296–2318.

⁴⁸Jörg Wimmers(n 22), 399.

⁴⁹Jörg Wimmers(n 22), 395.

⁵⁰Jörg Wimmers(n 22), 401.

⁵¹Kate Klonek, 'Of Systems Thinking and Straw Men' (2023) 136(6) *Harvard Law Review* 339, 358.

⁵²Evelyn Douek. Content Moderation as Administration. *Harv L Rev.* 2022;136(2):526–607.

⁵³Martin Husovec and Irene Roche Laguna.

⁵⁴Quintais JP and Appelman N et al. 906(8).

which also arises the legitimacy issues.⁵⁵ To address this issue, the DSA requires online platforms to ensure that qualified personnel are involved in the decision-making process to prevent solely automated decisions. However, whether platforms are willing to employ enough qualified staff to comply this rule is still a problem.⁵⁶ While Article 23 of the DSA mandates platforms to promptly disclose their initial content moderation decisions and provide reasons, this does not encompass the entirety of content moderation practices, which include complaints stemming from those decisions and subsequent involvement with redress mechanisms.⁵⁷ Therefore, the redress mechanisms may find difficulty in solving systematic issues and they neither could be able to provide complete redress measures for users.

Lastly, the shortcomings of the internal and external contestation could be hardly ameliorated by the traditional court. As underlined above, the DSA put excessive emphasis on the form constraints of platform power over content moderation, which do not provide users with a new substantive right or right of action as a judicial remedy before the court.⁵⁸ A consequent of this is that users may not be able to utilize Article 12 of the DSA in court as an effective means of safeguarding their fundamental rights against the providers. Whether users can directly or indirectly appeal to their fundamental rights in relation to an intermediary's content moderation decisions is a contentious issue.⁵⁹ Up to date, there is no precedent from the ECtHR or CJEU regarding a user initiating legal action against a platform for content removal. Some national courts are invoking fundamental rights in decisions regarding platforms' content removal due to their significant influence over online public discourse.⁶⁰ However, it is not feasible to directly apply the free speech standards to platforms in other member states since the rights enshrined in the Charter are directed towards the institutions and bodies of the Union and do not directly regulate the horizontal relationship between private parties.⁶¹ As a result, the court could not amend the flaws of the contestation provided by the DSA. Over-achingly, justice is more than due process. The internal redress measure has the inherent impartiality problem, while the external redress measure is lacking of binding effects. Meanwhile, both of them focus on the ex post contestation, which fails to provide complete remedy for ex ante content moderation. Their shortcomings are unlikely to be mitigated by court given the lack of substantial regulation of the DSA.

Interim conclusion

The DSA grants users several due process rights to against the providers. These are huge progress to empower users and regulate content moderation. Nevertheless, the DSA has transferred the responsibility of the ex-ante and ex-post filtering, blocking and removing illegal and harmful online information to the providers.⁶² This allows the providers to enjoy wide discretion given the majority obligation posed on them are procedural-oriented. Meanwhile, the ambiguity and procedural focus regulation of the DSA might prompt

⁵⁵Machado CCV, Aguiar TH. Emerging Regulations on Content Moderation and Misinformation Policies of Online Media Platforms: Accommodating the Duty of Care into Intermediary Liability Models. *Business and Human Rights Journal.* 2023;8(2):244–251.

⁵⁶Berrak Genc Gelgec. 52(7).

⁵⁷Sharon Galantino. 121(10).

⁵⁸Bengi Zeybek, Joris van Hoboken, Ilaria Buri. Redressing Infringements of Individuals' Rights Under the Digital Services Act. *DSA observatory.* 2022.

⁵⁹Matthias Kettemann C, Anna Sophia Tiedeke. Back up: Can users sue platforms to reinstate deleted content? *Internet Policy Review.* 2020;9(2).

⁶⁰Ibid.

⁶¹Jörg Wimmers. 382(22).

⁶²Marco Bassini. (23).

symbolic responses from the providers, which may not effectively address the issue at hand. Overall, although this essay acknowledges the DSA itself as a progress to the regulation of content moderation, it argues that procedural-oriented regulation in the DSA alone is insufficient for enabling substantial change. This essay, therefore, calls for more balances and checks of the due process from outsiders. Accordingly, the following sections will discuss some potential improvements.

More balance and checks to the due process in the DSA

Having fully realized the deficits of procedural-guided regulation in the DSA, this section of this essay will discuss some potential reforms to have more balanced and checks to the due process.

Rules-making: more substantial guidance and transparency

The rules-making is the start point of content moderation, which requires more regulation, democracy and legitimacy. EU council has noted that we should not automatically assume that online intermediaries are the most qualified to determine the legality or illegality of content.⁶³ The definition work necessitates certain competencies, independence, and representativeness, substantive and political expertise.⁶⁴ Therefore, the formation of T&Cs shall not be exclusive to the providers, which calls for intervention of law and other stakeholders.

First, the regulator could give the providers some primary guidance in rules-making. Some suggest to develop criteria for determining what constitutes hateful, violent, dangerous, offensive, or defamatory expression, to complement the DSA regulation.⁶⁵ For example, the NetzDG mandates social media platforms to remove content deemed 'clearly illegal,' such as the dissemination of specific propaganda, commission of forgery, and incitement to crime—all delineated in distinct statutes.⁶⁶ Similarly, a non-exclusive list of the primary sources or categories of harms could be rolled out by the government.⁶⁷

Globally, internet activists, participants in international forums, and advocates for Internet freedom have urged the adoption of an Internet Bill of Rights—a global agreement that would bind both public and private entities to safeguard individuals' freedoms and rights.⁶⁸ The international human rights bodies and civil society organizations even have advocated for platforms to "explicitly integrate" principles of fundamental rights law into their T&Cs to enhance the protection of freedom of expression on the internet.⁶⁹

This essay align with the above suggestions since the above domestic and international rules and guidance are substantial limits to the rules-making power of the providers. The providers should

⁶³Guidance note on content moderation. Adopted by the Steering Committee for Media and Information Society (CDMSI). 2021.

⁶⁴Deirdre Mulligan K, Kenneth Bamberger A. Allocating Responsibility in Content Moderation: A Functional Framework. *Berkeley Technology Law Journal*. 2021;36(3):1091–1126.

⁶⁵Ioanna Tourkochorit. The Digital Services Act and the EU as the Global Regulator of the Internet. *Chicago Journal of International Law*. 2023;24(1):129–133.

⁶⁶Sharon Galantino. 106(10).

⁶⁷Lorna Woods. Online harm reduction—a statutory duty of care and regulator. 2019.

⁶⁸Redeker D, Gill L, Grasser U. Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights. *International Communication Gazette*. 2015;80(4):302.

⁶⁹Quintais JP, Appelman N et al. 881(8).

not be the only rules-maker and interpreter of the standard of content moderation, the legitimacy of the T&Cs should be challenged by regulators and users. This could prevent the misuse of T&Cs by the providers to harm the fundamental rights of users. It might be problematic on account of the voluntary characteristic of these guidance, but the providers have no reason to do not follow them to build up their image and reputation.

In addition to adding guidance to the formation and interpretation of T&Cs, the publicity and transparency rules could be further strengthened. The primary aspect is to disclose ex ante censorship. Content moderation entails more than just addressing individual posts; it involves a complex blend of both preemptive and reactive content assessment, which necessitates navigating challenging trade-offs.⁷⁰ For this reason, platforms ought to be more transparent about how, when, and why they deploy ex ante automated and human screening of user-generated content.⁷¹ Rory⁷² suggested that a regulator might opt to include provisions that would make automated ex ante content screening less inscrutable, either by providing for government audits, facilitating independent research, or by requiring disclosure.⁷³ Therefore, the transparency rules could be improved by adding ex ante content moderation disclosure.

To conclude, some substantial guidance could be provided for the providers to make their T&Cs both at the domestic and global level, and the transparency rules shall be extended to ex ante content moderation.

Enforcement: more participation of outsiders

The enforcement of content moderation also calls for more checks and balances. The first one is to have a broad interpretation of Article 14 of the DSA by the authorities. Quintais et al proposed that Article 14(4) could be read as translating such human rights recommendations into a legal obligation on platforms to apply fundamental rights law under their T&Cs not just through general obligations, like risk assessments, but also in concrete decisions.⁷⁴ This means Article 14 will make it possible, in principle, to establish the indirect horizontal effect of fundamental rights in the relationship between online platforms and their users. In other words, Article 14 must be operationalized within the framework of the international and European fundamental rights standards for the application and enforcement of T&Cs to take due regard of fundamental rights.⁷⁵ In doing so, the enforcement of content moderation by providers could be more serious to the protection of the fundamental rights of users.

Second, it has been proposed that the government, civil societies and other non-governmental organizations could make guidelines on content moderation automated systems, including accuracy and error thresholds could be adopted to complement the DSA.⁷⁶ Accordingly, further guidance on human intervention and what entails human review should be provided. AI-based filtering tools should be

⁷⁰Kate Klonick. Of Systems Thinking and Straw Men. *Harvard Law Review*. 2023;136(6):339–354.

⁷¹Hannah Bloch Wehba. Automation in Moderation. *Cornell International Law Journal*. 2020;53:41–96.

⁷²Rory Van Loo. Regulatory Monitors: Policing Firms in the Compliance Era. *Colum L Rev*. 2019;119(2):369–424.

⁷³Rory Van Loo. Regulatory Monitors: Policing Firms in the Compliance Era. *Colum L Rev*. 2019;119(2):369–424.

⁷⁴Quintais JP, Appelman N et al. 896(8).

⁷⁵Quintais JP, Appelman N et al. 881(8).

⁷⁶Maria Barral Martinez. Platform Regulation, Content Moderation, and AI-Based Filtering Tools: Some Reflections from the European Union. *J Intell Prop Info Tech & Elec Com*. 2023;14(1):211–225.

designed, developed, and deployed when they meet certain safety and quality performance criteria. Accuracy standards and error rate thresholds must be established to ensure predictability and most importantly, a clear role responsibility.⁷⁷ For example, the Department of Telecommunications in India, through its AI Standardization Committee, solicited feedback to develop a standard Artificial Intelligence Stack that Indian companies can adopt as a template.⁷⁸ Regulators could take the initiative to develop standardised stacks in a consultative manner with participation from industry and civil society stakeholders.⁷⁹ This is also in line with Article 34 of the DSA which supports the formation of voluntary industry standards.

Interestingly, it has been proposed to establish multi-stakeholders Social Media Councils (SMCs), which are also soft-law institutions and relied on voluntary adherence. As its name suggested, SMCs would encompass civil society organizations, platforms, users, and governments, advising them on content moderation issues that extend beyond mere takedowns, as well as providing guidance on overarching issues that impact multiple platforms or the broader social media sector within the scope of regional, national and even global.⁸⁰ In doing so, it could mitigate the spill effects resulted from the enforcement disparities of multiple platform rules.

Therefore, the wide discretion enforcement power of the providers could be constrained by compliance with a more detailed standard at the domestic and even global sphere. It might be argued that the intervention of the government may infringe freedom of speech given state actors continue to employ false or misleading information as a weapon.⁸¹ On account of this, this essay favour the standard developed by the civil societies and non-governmental institutions. That said, developing an error-free standard is far from easy, but this standard could be gradually improved in practice.

Another way is to strengthen user censorship. Tourkochoriti proposed the platforms could use the paraphrasing technology to allow users to edit the hateful and insulting content that reaches them while protecting the speaker's expression at the same time.⁸² This is similar to the report on combating online harms through innovation of the American Federal Trade Commission, which listed user tools in its recommendations to tackle harmful content.⁸³ These tools could help users to control what content they see on the internet, shifting the content moderation effort from private platforms towards users. Nevertheless, the use of paraphrasing technology might out of control to distort/alter the meaning of the text of the writer, and the interests and values of freedom of speech could also be compromised.⁸⁴ User could also participate in content moderation through the contractual relationship with the providers. This means the content moderation conducted by the providers shall satisfy the the contractual expectations of the networks' members and advance their common goal. In this sense, contract law could empower users by offering a decentralized and diversified check over the platforms'

⁷⁷Ibid.

⁷⁸Varun Ramdas, Identifying an Actionable Algorithmic Transparency Framework: A Comparative Analysis of Initiatives to Enhance Accountability of Social Media Platforms. *Nat'l LU Delhi Stud LJ*. 2022;74(88).

⁷⁹Varun Ramdas. 90(67).

⁸⁰Molly Land K. Against Privatized Censorship: Proposals for Responsible Delegation. *Virginia journal of international law*. 2019;60(2):365–432.

⁸¹Commission. Action Plan Against Disinformation. JOIN/2018/36 final. 2018.

⁸²Ioanna Tourkochoriti. 2023;142(55).

⁸³Federal Trade Commission Report to Congress. Combatting Online Harms through Innovation. 2022.

⁸⁴Ioanna Tourkochoriti. 142(55).

content moderation practices.⁸⁵ If users could successfully bring contractual claims against platforms and hold them responsible for arbitrary, biased, or unfair content removal decisions, they could compel platforms to align their content moderation policies with the collective interests of the user community. This approach to contract interpretation may facilitate a bottom-up check on content moderation via private ordering, thus increasing platforms' accountability.

In conclusion, given the high risk of private enforcement, this essay calls for more participation of the outsiders. The authorities could have a broad interpretation of Article 14 of the DSA to enable the providers be more responsible to the fundamental rights of users. The civil societies and other non-governmental institutions could help to develop a universal voluntary standard for content moderation to give more substantial guidance in enforcement. Additionally, users could also participate in this process by giving them contractual rights and other technique tools.

Contestation: "crowd-judging" and digitization of court

Given the shortfalls of the internal and external dispute settlement system, some proposals for reforms will be analyzed in this section. This essay proposes to establish a crowd-judging system, which could resolve the impartiality issue of the current internal redress system, operation and other issues of the external redress system. Additionally, as the last resort, the court system should be more digital-proof to be accessible and affordable to users.

First of all, given the legitimacy, democratic and systematic flaws of internal and external redress mechanism, the providers could establish a "crowd-judging" system. The concept of crowd-judging was initially tested in 2008 by eBay India, which introduced a community court where eBay buyers and sellers could crowd source adjudication for disputes related to online feedback.⁸⁶ The Alibaba in China further developed this mechanism to resolve a broad range of disputes and to vote on transactional rules and regulations for its e-commerce platforms--Taobao and T-mall. This system proves to be very an efficient and economical mechanism to address platform disputes.⁸⁷ In this sense, this essay argues that the providers could also introduce this "crowd-judging" system to solve content moderation disputes. There are two reasons supporting this opinion. The first is that the crowd-jurors could overcome the limited human resources of the internal and external redress mechanism. The users themselves could be a judge to deal with the disputes at minimal operation cost. Also, the wisdom of the crowd gives it more credit and legitimacy. In other words, the decision made by the crowd is more trusted and accepted to users who might also be a juror. This solved the concerns for impartiality of the internal settlement system, as well as the non-binding enforcement issue of the external address system.

That said, Crowd-judging may face two potential drawbacks, which are variability and bias.⁸⁸ However, this could be mitigated by proper allocating the cases among jurors. Specifically, the inexperience jurors are in charge of the clear-cut and straightforward disputes of content moderation, such as manifestly illegal content or legal content, while the sophisticated jurors could deal with the more

⁸⁵Niva Elkin Koren, Giovanni De Gregorio, Maayan Perel. Social Media as Contractual Networks: A Bottom up Check on Content Moderation. *107 Iowa L Rev* 987, 987. *Iowa law review*. 2022;107(3).

⁸⁶Alan Kwan, Alex Yang S, Angela Huyue Zhang. Crowd-judging on Two-sided Platforms: An Analysis of In-group Bias. *Management Science*. 2023;70(4).

⁸⁷Ibid.

⁸⁸Ibid.

complicated cases. Additionally, platforms, cooperated with civil societies, governments and institutions could make efforts to improve media literacy, equipping citizens with the judgment skills and ability to identify harmful and unlawful contents.⁸⁹ The policy also improves the participation of inexperienced jurors, thus nurturing a more sustainable pool of crowd-jurors.⁹⁰ Therefore, this essay argues that the crowd-judging system could be a better contestation mechanism.

Second, the court could be more digital-proof to be the last resort for users. Cauffman (2021) argued that the facilitation of access to court proceedings through the introduction of harmonised rules limiting the costs of such proceedings might be more appropriate than the promotion of out-of-court dispute settlement, especially when it comes to the cases where fundamental rights such as the free speech is at stake.⁹¹ Likewise, Jörg Wimmers(2022) believes that the existing court system infrastructure could be modernized to satisfy the needs of users to pursue their rights.⁹² For example, online court hearings could be introduced to facilitate the trial and non-profit organizations could be delegated by users to bring claims, all of which could reduce the legal cost and improve efficiencies.⁹³ While this essay agree with the digital-proof of the court, it also put equal weight on the introduction of the out-of-court dispute system or the above-mentioned crowd-judging system. It can not be denied about the authority and superiority of the court system, which gives users the highest guarantee of justice. Additionally, the decision of the court is binding to the providers, which could make a difference to both the practice and law. However, the court is not the sole solution to safeguard fundamental rights. The out-of-court dispute system or the crowd-judging system have their unparalleled benefits of easy access and swiftness. Also, the large amount of disputes also requires the decentralize of the justice to decrease the legal cost. Therefore, the digitization of traditional court could be in tandem with the establishment of alternative dispute settlement system. To sum up, comparing the internal and external

address system designed by the DSA, the “crow-judging” system is a better option for the sake of impartiality, cost and enforcement. We should also bear in mind that the traditional court could be more digital-proof to act as the last resort.

Conclusion

The aim of this essay is to examine the effectiveness of the procedural-oriented regulation of content moderation in the DSA. The first part identifies the main three due process obligations posed to the providers from the perspective of rules-making, rules-applying and rules-appealing. It admits that there are merits of the procedure-oriented regulation of the DSA, including but not limited to the increase of transparency and empower users. Nevertheless, the DSA also gives considerable discretion to the providers, without much substantial constrains to the power of the private regulation of content moderation carried out by them. Subsequently, the second part analyzed the transparency rules in rules-making are not enough to hold the providers be responsible for the protection of users’ fundamental rights. Additionally, the privatization of enforcement could be excessive and lack of surveillance from the authorities and users. The internal and external redress mechanism have their inherent bias and maintenance issues, which are not effective for users to defend their rights. In light of these deficiencies of due process regulation, this essay suggested that more substantial guidance and transparency should be given to rules-making, more outsider’s participation in the enforcement, establishment of a “crowd-judging” redress system and digitization of court. In doing so, the DSA could have more balances and checks to the providers and better achieve its regulation goal.

Acknowledgments

None.

Conflicts of interest

The author declares there is no conflict of interest.

⁸⁹Sharon Galantino. 105(10).

⁹⁰Alan Kwan, Alex Yang S et al. (75).

⁹¹Cauffman Caroline and Goanta Catalina. (32).

⁹²Jörg Wimmers. 401(22).

⁹³Ibid.