RESEARCH PAPER

Regulation of Social Media Content: Challenges and Opportunities

Dr.Soumya Vishwakarma[1*]

[1*]Assistant Professor Vivekananda Institute of Professional studies (Vips) Affiliated with GGSIPU
Email Id: Soumyavishwakarma1995@gmail.com

Abstract

Regulation of social media content has become a critical issue in digital governance, particularly in relation to misinformation that spreads rapidly across interconnected platforms. The present research examines how content removal is associated with cross-platform dissemination, audience engagement, and the timing of moderation. A quantitative observational approach based on secondary data analysis was employed. The analysis covered 8,104 unique videos linked to dissemination traces across Facebook and Twitter/X, with removal information and engagement indicators used to evaluate moderation patterns.

The findings indicate that regulatory intervention was concentrated among videos with broader circulation and stronger engagement. Content appearing across multiple platforms was considerably more likely to be removed than content with limited spread. Removal rates also increased sharply with dissemination intensity, suggesting that visibility played a decisive role in shaping moderation outcomes. However, the temporal evidence showed that intervention often occurred only after a substantial delay from publication, indicating that moderation was frequently reactive rather than preventive. These results point to a major challenge in social media governance, namely the difficulty of ensuring timely intervention in rapidly evolving information environments. At the same time, they reveal important opportunities for improvement. Cross-platform monitoring, transparency, and dissemination-based prioritization may strengthen content regulation by enabling earlier and more targeted responses. Overall, the analysis demonstrates that social media content regulation is closely linked to amplification dynamics and that effective governance depends not only on what content is regulated, but also on when and how intervention is carried out.

Keywords: social media content regulation, content moderation, misinformation, cross-platform dissemination, platform governance, digital transparency

How to cite this article: Vishwakarma S. Regulation of Social Media Content: Challenges and Opportunities. Int J Drug Deliv Technol. 2026;16(41s): 986-993. DOI: 10.25258/ijddt.16.41s.104

1. Introduction

Social media platforms have become key infrastructures for communication, political discourse, information exchange, and public debate. Simultaneously, they have also been areas of contention regarding what is to be kept visible, what must be blocked and by whom. Regulating social media content is no longer a technical problem hence it is a wider governance problem of platform regulation, user regulation, institutional regulation and harm to the population. Early efforts on social norm enforcement online displayed the potential of online communities to produce strong group responses that can be seen as a form of informal sanctioning and thus demonstrated that content regulation not only arises out of formal moderation policies but also out of social pressures in online settings[1].

A key challenge in this sphere is the fact that the understanding of harmful content does not always have the same perception. What is considered to be severe, offensive, misleading or unacceptable often differs depending on countries, social setting and users. The comparative evidence of the global perceptions of the harmful online content shows that people do not necessarily agree on the severity of the various types of online harm which in turn makes it difficult to establish universally acceptable standards of moderation[2]. This sophistication is enhanced by the difference in perception of moderation systems per se by the users. Studies that compared commercial and volunteer models of moderation provide evidence that user trust is strongly related to the perceived toxicity, fairness, and transparency, meaning that content regulation is not only dependent on the outcomes but also on the validity of the process, according to which these decisions are made[3].

The increased magnitude of these problems has elevated content moderation to the forefront of cross-disciplinary research and policy focus. A recent compilation of the scholarship in content moderation points at the gap which can often exist between documented community rules, practice on the ground, and the recommendations made by researchers themselves, which explains why closer relationships need to be established between the theoretical foundation of platform governance and the practice of moderation itself[4]. Corresponding experimental research on the legitimacy of content removal indicates that the legitimacy of moderation decisions is highly affected by the perceived consensus on the information about community rules, i.e. the effectiveness of the regulation is partially defined by the user perceiving the rules to be normatively justified[5]. Even certain interventions like warning labels and labeling decisions have proven to be a controversial tool

*Author for Correspondence: Soumyavishwakarma1995@gmail.com

of governance at the platform level, particularly in political sensitive situations illustrating how moderation choices have the ability to influence societal trust and the sense of impartiality[6].

Those issues have contributed to the increase in policy discussion regarding the way digital platforms should be regulated. The platform-governance framework of UNESCO points to the fact that efficient regulation has to protect the freedom of speech and access to information and has to deal with harmful content by a multi-stakeholder strategy[7]. Simultaneously, the OECD transparency reporting and transparency in the executive branch work is responsive to the disinformation issue, that openness, accountability, and evidence-only policy are key prerequisites to credible digital regulation[8,9]. The relevance of such views, in particular, is due to recent literature that cautions against naive notions of the misinformation harms by stating that the association between misinformation exposure and social harm is more complicated than it is commonly believed[10]. It implies that the regulation should be reasonable, well-calculated, and properly justified by evidence instead of being based on political or other moral panic.

Meanwhile, the judgment on the part of the user is still the key to the regulatory results. Moral judgment and punishment choices in reference to objectionable online content evidence indicates that reporting choices of users are influenced by normative measurements of fault and damage, and this further demonstrates that content regulation is a technical and moral practice [11]. In more recent times, the Digital Services Act has heightened the concern with the risks, transparency and access to the data, bringing new focus to the systematic review of the processes of moderation of platforms[12]. It is against this background that the current study explores the topic of Regulation of Social Media Content: Challenges and Opportunities using an empirical investigation of the circulation and removal of misinformation-related content. The study aims to explain the way in which regulatory action is determined by visibility, amplification, and timing of intervention in modern social media contexts, basing this on the association between cross-platform dissemination, engagements, and timing. The objectives of the study are:

- To examine the extent of cross-platform dissemination of misinformation-related content on social media.
- To analyze the relationship between content visibility, audience engagement, and removal patterns.
- To assess the temporal dynamics of moderation by evaluating the delay between content publication and removal.
- To identify the major challenges and opportunities in the regulation of social media content.

2. Methodology

2.1 Research design

The study adopts a quantitative observational research design that applied the analysis of secondary data to investigate the control of social media contents. The methodology was formed as a case-based study of the misinformation-related video content and its cross-platform distribution. This design was deemed suitable since the control of content is not mainly influenced by the existence of objectionable text, even though content in the form of a network of interconnected platform spaces of circulation and the response of moderation to that circulation. The research design was such that it evaluates three key aspects of regulation namely content removal, cross platform diffusion and timing of intervention. Collectively, these dimensions give a factual foundation on the analysis of the challenges and opportunities that are related to the current social media content governance.

2.2 Data source and scope of the study

The study is based on publicly available secondary data of videos related to misinformation, the date when it was published and deleted, and associated traces of dissemination on Facebook and Twitter/X[13]. The data allowed to consider the moderation results and the cross-platform dissemination and interaction, thus being applicable to investigating the relationships between regulatory intervention and content publicity and amplification.

The study deliberately focuses on a specific type of misinformation to generate a narrow and methodologically feasible analysis. Instead of trying to capture the various policy areas in all social media sites, the research employed a well-articulated misinformation instance to explore larger regulatory trends including selective intervention, delayed enforcement, and the connection between visibility and moderation.

2.3 Unit of analysis

The unit of analysis was the individual video. Every video was presented as one piece of analysis and was connected, when it was possible, with related Facebook and Twitter/X dissemination logs, interaction rates, date of publication, and date of deletion. This allowed investigating whether the content with broader circulation and greater involvement had a higher risk of being dropped and whether this action happened during a short or long period of time.

2.4 Data preparation

The data were first cleaned and standardized to enhance uniformity and eliminate duplication before analysis. Records that represented the same video were grouped together such that only one of these videos was represented in the analytical file. The same Facebook and Twitter/X identifiers were combined into unique post counts and the existing metadata were normalised through the timing, engagement and platform-linkage fields.

Moreover, a number of measures of analysis were calculated to facilitate the empirical analysis. Available removal information was used to make a binary removal indicator, and days to removal calculated between the publication and removal dates wherever both dates were valid. Cross-platform presence and cross-platform post count, as well as Facebook total engagement measures, were also generated. To make comparison easier, measures of dissemination and engagement were clustered into larger intensity bands of lower and greater amplification.

2.5 Analytical approach

The analysis was done in three phases. First, descriptive statistics were applied to present the general characteristics of the sample in the form of frequencies, percentages, medians, means, and percentile distributions. This phase defined the level of cross-platform distribution, interaction, and deletion. Second, the comparative analysis was carried out to determine whether removal was different depending on cross-platform presence, spread intensity, and engagement intensity. The comparison of removed and non-removed content and the identification of whether moderation was evenly spread or clustered around the more widely circulated content were conducted by use of cross-tabulations and summary statistics. Third, cases having valid publication and removal dates underwent a temporal analysis. The timeliness of intervention was assessed with measures of median removal lag, mean lag, percentile distribution, and maximum delay. The time of removal was also compared at the levels of diffusion and engagement to understand whether more salient materials were regulated faster than less salient materials.

2.6 Statistical treatment

Considering the skewed data of digital-trace, particularly with the spread and engagement variables, the analysis used more mediations, grouped rates, and interpretation based on the distribution, compared with the use of means alone. This method was more suitable since few cases which had been highly amplified could affect arithmetic averages disproportionately. Tables and figures have been used to present the findings. The precision of counts, proportions and summary statistics were reported using Tables whereas cross-platform distribution, intensity of removal and temporal distribution of intervention were represented through figures. The strategy facilitated coherence, accuracy, and clear correspondence of the analytical products and the research objectives.

3. Results

3.1 Overall sample characteristics and cross-platform distribution

The number of unique videos that were analyzed was 8,104. The descriptive profile shows that it has a high level of cross-platform dissemination with 8,007 videos (98.8%), being associated with one or more posts on Facebook or Twitter/X. It was found that facebook linkage was present in 7,678 cases (94.7%), and twitter/X linkage was present in 5,202 cases (64.2%). It is worth noting that the quantity of videos shared through Facebook and Twitter/X (4,873 and 60.1, respectively) indicates that misinformation is a multi-platform phenomenon. The removal of content was found in 420 instances (5.2%), and 396 instances had complete publication and removal timestamps that could be used to analyze the time, which is in Table 1.

Table 1. Descriptive profile of the study sample

Metric	Value
Unique videos	8,104
Videos with any linked cross-platform post	8,007 (98.8%)
Videos linked to Facebook posts	7,678 (94.7%)
Videos linked to Twitter/X posts	5,202 (64.2%)
Videos linked to both Twitter/X and Facebook	4,873 (60.1%)
Videos removed	420 (5.2%)
Cases with both publication and removal timestamps	396
Median cross-platform posts per video	5.0
Median Facebook total engagement per video	313.5

The cross-platform presence distribution also indicates that the de-platforming was more focused on videos shared by numerous platforms. Interestingly as Figure 1 demonstrates, the videos shared on both Facebook and Twitter/X had a significantly greater concentration of removal compared to videos shared on a single platform. This trend indicates that the wider the platform inter-visibility the more probable the action by regulatory bodies.

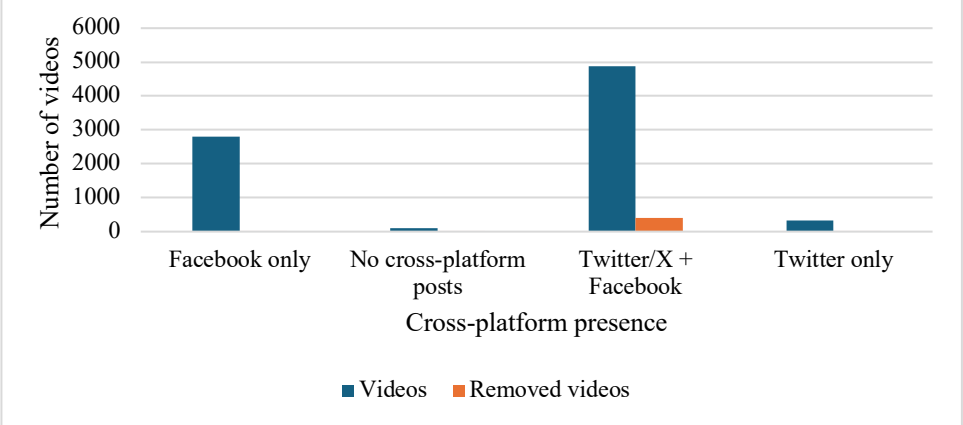

Figure 1. Cross-platform presence and removal pattern

3.2 Removal incidence across cross-platform presence

Removal occurrence was remarkably low depending on the level of cross-platform presence. According to Table 2, the videos on Facebook and Twitter/X had the highest removal rate of 8.17% and Facebook-only content removal rate was 0.21%. Videos linked only to Twitter/X showed a medium rate of removal (4.56%), but those videos about which cross platform dissemination could not be identified were infrequent and showed a low level of removal.

Table 2. Removal incidence by cross-platform presence

Cross-platform presence	Videos	Removed videos	Removal rate (%)
Facebook only	2,805	6	0.21
No linked cross-platform posts	97	1	1.03
Twitter/X and Facebook	4,873	398	8.17
Twitter/X only	329	15	4.56

The variances in Table 2 indicate that moderation was not consistent with all material. Instead, regulatory intervention was focalized on videos that crossed several platform environments that suggest that higher platform diffusion and visibility could lead to more implementation of enforcement.

3.3 Removal patterns by diffusion intensity

There was an apparent increase in removal between spread-intensity types. According to Table 3, the removal rates rose as low-spread group was 0.51% to very-high-spread group at 36.51%. The median cross-platform post counts and median Facebook engagement increased concurrently with the removal rates which further supports the relationship between amplification and intervention.

Table 3. Removal incidence by spread intensity

Spread band	Videos	Removed videos	Removal rate (%)	Median Facebook engagement	Median cross-platform posts
No spread	97	1	1.03	9	0
Low	4,086	21	0.51	76	2
Medium	2,102	38	1.81	607	11
High	1,096	96	8.76	3,098	47
Very high	723	264	36.51	16,415	209

The same pattern can be observed in Figure 2, in which the removal rate increases exponentially with the spread intensity. The figure shows that multi-post and multi-platform diffusion content with higher content was more prone to being removed.

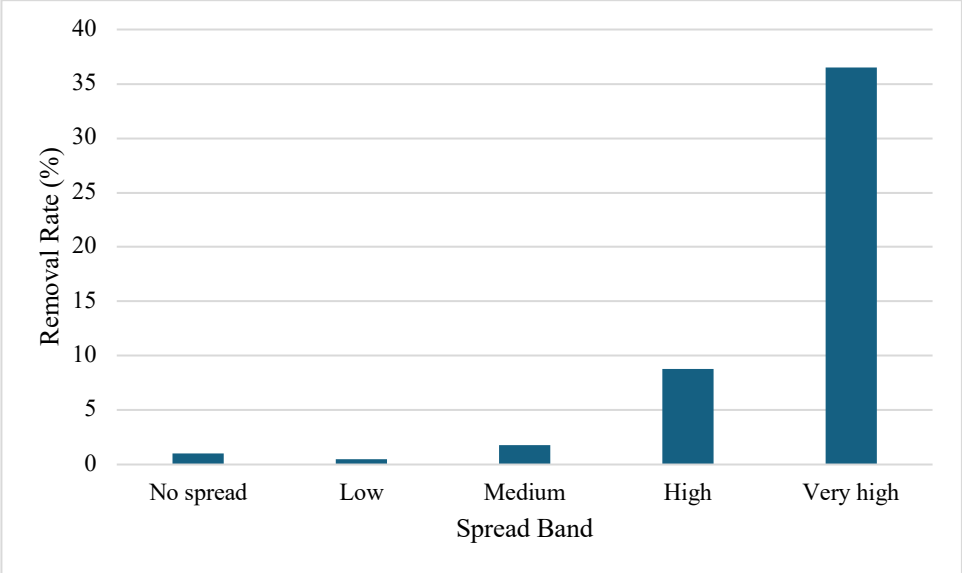

Figure 2. Removal rate by spread intensity

3.4 Comparative profile of removed and non-removed content

A structural comparison of removed and non-removed content shows that there are significant differences in

visibility and engagement. Removed videos, as illustrated in Table 4, had a median of 146.5 cross-platform posts, whereas that of non-removed videos was only 5. The median Facebook total engagement was 8,516.5 with removed videos and 271 with non-removed videos. Similarly, in cases where view-count data was available, videos were removed which had significantly more median view count (51,085) than those that were not removed (15,219.5).

Table 4. Comparative profile of removed and non-removed content

Status	Videos	Median cross-platform posts	Median Facebook total engagement	Median view count (available cases)	Rows with view count available
Not removed	7,684	5	271	15,219.5	514
Removed	420	146.5	8,516.5	51,085	420

Table 4 shows that the content that was removed was in a much higher place of public visibility compared to material that was not removed. The finding indicated that the decisions made on moderation were strongly linked to content reach and participation and not equally spread on all problematic content.

3.5 Temporal dynamics of removal

The 396 cases that had valid publication and removal timestamps were temporally analyzed. The findings showed a median of 38.5 days and average of 45.0 days of removal lag, which implies that it was common to moderate well after the content was published. As shown in Table 5, a quarter of removals were made within 16 days, and three quarters within 61.2 days. The data was highly skewed to the right with a highest removal delay of 1,213 days.

Table 5. Distribution of time to removal

Statistic	Value
Valid timed removals	396
Median days to removal	38.5
Mean days to removal	45.0
25th percentile	16.0
75th percentile	61.2
90th percentile	82.0
95th percentile	93.2
Maximum	1,213.0

Figure 3 depicts the time distribution and indicates a clustering of removals in the earlier stages of the post-publication variable, and then a long right tail of the distribution, that is, longer delays in intervention.

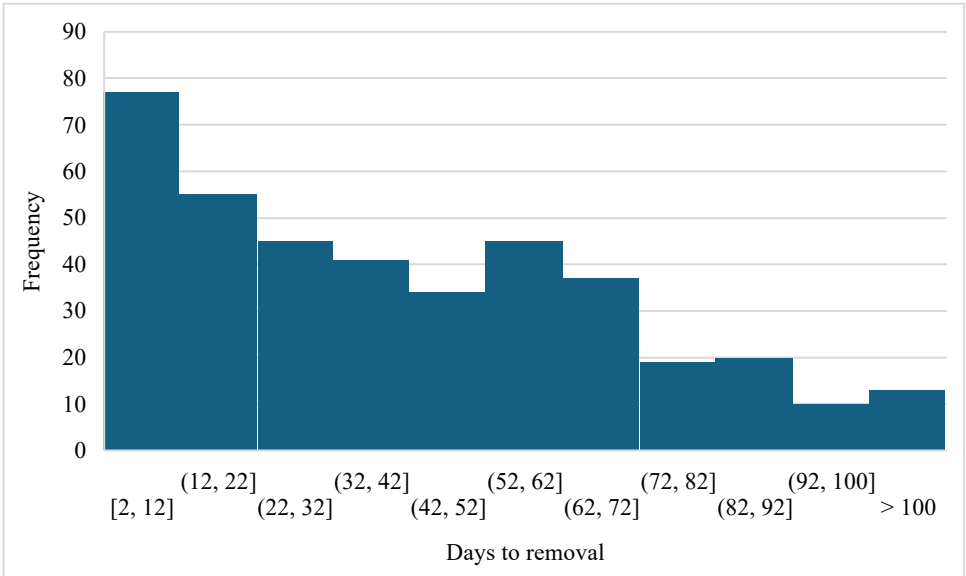

Figure 3. Distribution of days to removal

Moderation was frequently delayed when compared to the original publication as indicated in Table 5 and Figure 3. Though most of the removals were done during the initial few weeks, a significantly high percentage still

continued staying on the internet, much longer. This observation is one of the key issues with content regulation: potentially harmful content might remain viral before appropriate actions are implemented.

3.6 Removal timing across spread and engagement groups

Removal lag was also determined in spread category and engagement category to determine whether the most visible content is removed quicker or slower compared to the less visible content. Table 6 demonstrates the median removal lag to be 29 days with low-spread content, 58 days with medium-spread content, 46 days with high-spread content, and 29 days with very-high-spread content. In the category of engagement groups, within the groups of low, medium, high, and very-high engagement, there was a median removal lag of 46 days, 51.5 days, 46 days, and 26 days respectively.

Table 6. Median removal lag across spread and engagement groups

Group	Timed removals	Median days to removal
Low spread	19	29
Medium spread	35	58
High spread	91	46
Very high spread	250	29
Low engagement	3	46
Medium engagement	42	51.5
High engagement	163	46
Very high engagement	188	26

According to the trend that has been reported in Table 6, moderation did not maintain a linear trend. Whereas content of very high level of engagement was removed comparatively fast, medium level groups frequently had longer delays. This implies that intervention can only be more responsive once the content attains particularly salient amounts of public exposure, and the moderately diffused misinformation exists on the Internet longer.

4. Discussion

The present findings depict that the regulation of the social media content operates in a highly selective and reactive manner. The removals were focused in the videos that got high cross-platform distribution and significant public attention, and the time analysis showed that the removal tended to happen when the piece of content was left in the Internet long enough. These tendencies indicate that the content governance depends not only on the presence of misinformation, but also on the promulgation and propagation speed of misinformation. In this regard, the research supports the claim that misinformation on social media must be perceived as a governance issue with tangible effects in society, instead of being a matter of information quality. Denniss and Lindberg made it clear that misinformation diffuses in contagion-like ways and may produce severe public-health effects, which is why the issue of delayed moderation in high-visibility misinformation settings is of particular concern[14].

One of the key conclusions of the findings is that the concept of visibility seems to be an implicit enforcement trigger. The content with a greater intensity of dissemination and stronger engagement was significantly more prone to removal as compared to the content with the limited spread. This tendency leads to the idea that moderation can be influenced not so much by the mere presence of problematic content as by its possible size of influence. This finding is consistent with wider literature which indicates that misinformation is not an objective or predetermined category, but it is a category which is relative, which is dependent on context, evidence and interpretation, and judgment of experts[15]. Defining misinformation is sensitive to both evidentiary and contextual meaning, both of which, as Vraga and Bode argue, need consideration, which can in part explain why platforms do not ban all dubious material equally[15]. Rather, intervention seems to escalate as misinformation becomes more apparent, more destructive in the extent of its path and more challenging to neglect.

The findings also highlight the significance of cross platform dynamics in content regulation. Videos shared on both facebook and twitter/X had the greatest rates of removal, which indicates that the dissemination of misinformation on several platforms can be the subject of more stringent regulation. This is a critical discovery since it shifts the discussion away of platform specific moderation and to a networked conception of digital content governance. The regulation of social media cannot be understood solely with the single platform analysis approach as the problematic content tends to become transmitted by a broader online ecosystem and then it can be moderated. The findings in this respect can help to promote more integrated regulatory thinking, in which dissemination across platforms is viewed as an important risk sign. Roberts et al. pointed out the usefulness of large scale open monitoring infrastructure in tracking the flow of information throughout the digital landscape[16]. Their work particularly applies to this context since the current results suggest that governance is made more significant when it is informed by cross-platform visibility and not platform events in isolation[16].

The temporal insight is another important contribution of the study. The median publication to removal lag appears to indicate that moderation often takes place when the misinformation has already had a good opportunity to spread. This lag represents one of the most enduring problems of social media regulation:

harmful or misleading information can gain massive publicity before an effective response can be mandated. Even though certain high-profile material may have been taken off fairly fast, the general trend was heavily to the right, with a significant number of situations showing long delays. This observation can be related to the wider issue of the fact that existing moderation structures are usually reactive, as opposed to proactive[14]. In the event of delayed intervention, the practical effect of removal can be less than otherwise, since exposure, reposting and audience interaction could have occurred. This implies that, in regulatory terms, the effectiveness cannot be measured by the mere fact that the content has been finally removed, but rather in the speed with which the latter happens.

Meanwhile, the findings are not consistent with an entirely negative picture. The high correlation of removal, spread, and engagement also gives key governance opportunities. To the extent that widely distributed content has a higher likelihood of eliciting moderation, then dissemination signals can be employed more systematically as warning signals. This leaves room to more active models of governance where platforms focus on reviewing on a cross-platform amplification, intensity of interaction and velocity of diffusion. Doshi and Schmidt explain transparency as a soft style of governing platforms that can influence their behavior without necessarily having them directly coerced[17]. Their argument is very much applicable in this case: when visibility patterns and the decisions regarding moderation are more transparent, they would be able to enhance accountability, allow external checks, and facilitate earlier intervention[17].

This relates to the policy relevance of transparency. The conclusion is that the ability to enhance public understanding of how and when moderation takes place is one of the most promising prospects of social media regulation. MacCarthy has contended that transparency is a key to good regulation under the pretext that it enables the regulators, researcher, and the citizens to assess whether platforms are acting reasonably and in good faith[18]. The current evidence justifies that stance. In the absence of open reporting on the pattern of removals, when and whether to whom, it is hard to tell whether platforms are acting in response to actual harm, under public pressure, reputational risk, or other institutional pressure. Increased transparency would not only enhance accountability, but also, contribute to pinpointing areas in which delays and inconsistencies in enforcement still pose the biggest concerns.

The research has implications on normative governance arguments as well. The UNESCO guidelines focus on the importance of balancing governance of digital platforms to ensure that they do not embody a content moderation policy and freedom of expression, right to access information and accountability by multiple stakeholders[19]. This equilibrium is particularly significant in the case of misinformation regulation, in which excessive intervention can pose a threat of legitimate speech, whereas inadequate enforcement can give birth to harmful content. The existing evidence indicates that the contemporary regulation is somewhat selective in the acts, such as banning of highly amplified material but not total suppression which can signify an effort- intentional or pragmatic- to regulate this conflict. Nevertheless, the noted delays also suggest that selective moderation is not necessarily enough when intervention is introduced when high rates of dissemination have been already attained[19].

In general, the discussion yields a twofold conclusion. The main difficulty is that the timing and selective concentration of intervention is reactive, and thus before it is removed, misinformation may become widespread. The main opportunity will be to utilize the transparency, cross-platform surveillance, and dissemination-based prioritization to construct more timely and responsible systems of content control. By so doing, the paper contributes to a larger conceptualization of social media regulation as a dynamic process that is informed by visibility, amplification, timing, and governance design instead of removal decisions.

5. Limitations

This paper lies within a well-defined empirical framework. It prioritizes video based contents on misinformation and as such does not apply to all types of problematic online content, hate speech, cyberbullying or copyright infringements. Moreover, it is analyzed on the basis of platform-related dissemination traces recorded in key social media settings, and broader circulation trends might be broader than the ones considered in this study. The time evaluation of deletions is also restricted to those with valid dates of publication and deletion. These considerations must be interpreted as the conditions of the scope and not as the weaknesses. They assist in putting the findings into perspective without obstructing the overall contribution of the study, which is a specific analysis of how the content visibility and cross-platform dissemination and the timing of interventions influence the social media content regulation.

6. Conclusion

The current study has analyzed how social media content is controlled by considering the example of misinformation-related video distribution, cross-platform distribution, and time of moderation. The results reveal that not only is content regulation nondispersive, but it is evenly distributed. Rather, it is regulatory intervention that is focused on the content that attains increased visibility, more engagement, and wider cross-platform dissemination. The videos shared on both Facebook and Twitter/X had a significantly higher chance to be removed than the content with little diffusion, which proves that amplification is a much more important factor in causing moderation responses. The study also proves that timing is one of the key problems of content governance. Even with highly circulated content, removal was usually noted with a significant delay period post-publication, although this was frequently the case. This implies that the intervention is often reactive and not preventive and

problematic content can have a larger reach before enforcement is effected. This kind of trend shows a glaring shortcoming in the existing moderation infrastructure especially in the fast paced digital worlds where exposure can spread very quickly across interconnected spaces. Simultaneously, the findings indicate significant prospects of regulation betterment. The close correlation between removal, dissemination intensity, and a more vivid engagement implies that quantifiable signs of spread are capable of assisting in preventive detection, prioritization that is more selective, and moderation practices that are more responsible. Transparency, cross-platform surveillance, and in-time intervention turn out to be particularly useful governance instruments in enhancing the effectiveness of regulation without necessarily having to do away with covering content wholesale. In general, the analysis demonstrates that the regulation of social media content is influenced by the characteristics of problematic content, as well as the speed, extent, and the number of platforms of its distribution. The paper offers more subtle insights into the issues and the opportunities that shape the modern content governance of social media by connecting patterns of removal with the process of amplification.

References

1. Rost K, Stahel L, Frey BS. Digital social norm enforcement: Online firestorms in social media. PLoS one. 2016 Jun 17;11(6):e0155923.
2. Jiang JA, Scheuerman MK, Fiesler C, Brubaker JR. Understanding international perceptions of the severity of harmful content online. PloS one. 2021 Aug 27;16(8):e0256762.
3. Cook CL, Patel A, Wohn DY. Commercial versus volunteer: Comparing user perceptions of toxicity and transparency in content moderation across social media platforms. Frontiers in Human Dynamics. 2021 Feb 19;3:626409.
4. Singhal M, Ling C, Paudel P, Thota P, Kumarswamy N, Stringhini G, Nilizadeh S. SoK: Content moderation in social media, from guidelines to enforcement, and research to practice. In2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P) 2023 Jul 3 (pp. 868-895). IEEE.
5. Aguerri JC, Miró-Llinares F, Gómez-Bellvís AB. Consensus on community guidelines: An experimental study on the legitimacy of content removal in social media. Humanities and Social Sciences Communications. 2023 Jul 15;10(1):416.
6. Bradshaw S, Grossman S, McCain M. An investigation of social media labeling decisions preceding the 2020 US election. Plos one. 2023 Nov 15;18(11):e0289683.
7. Jelassi T. Towards an Internet of Trust—UNESCO's Guidelines for the Governance of Digital Platforms. InLMDE Conference 2023 Jun 19 (pp. 3-12). Cham: Springer Nature Switzerland.
8. Llanos J. Transparency reporting: considerations for the review of the privacy guidelines. OECD Digital Economy Papers. 2021;(309). Paris: OECD Publishing. Available from: https://doi.org/10.1787/e90c11b6-en
9. Matasick C, Alfonsi C, Bellantoni A. Governance responses to disinformation: how open government principles can inform policy options. OECD Working Papers on Public Governance. 2020;(39). Paris: OECD Publishing. Available from: https://doi.org/10.1787/d6237c85-en
10. Budak C, Nyhan B, Rothschild DM, Thorson E, Watts DJ. Misunderstanding the harms of online misinformation. Nature. 2024 Jun 6;630(8015):45-53.
11. Vahed S, Goanta C, Ortolani P, Sanfey AG. Moral judgment of objectionable online content: Reporting decisions and punishment preferences on social media. Plos one. 2024 Mar 25;19(3):e0300960.
12. Sekwenz MT, Gsenger R. The Digital Services Act: Online Risks, Transparency and Data Access. Digital Decade. 2025 Jun 24:115-40.
13. Knuutila A. A dataset of Covid-related misinformation videos and their spread on social media [dataset]. Zenodo; 2021. Available from: https://doi.org/10.5281/zenodo.4557828
14. Denniss E, Lindberg R. Social media and the spread of misinformation: infectious and a threat to public health. Health promotion international. 2025 Apr;40(2):daaf023.
15. Vraga EK, Bode L. Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. Political Communication. 2020 Jan 2;37(1):136-44.
16. Roberts H, Bhargava R, Valiukas L, Jen D, Malik MM, Bishop CS, Ndulue EB, Dave A, Clark J, Etling B, Faris R. Media cloud: Massive open source collection of global news on the open web. InProceedings of the International AAAI Conference on Web and Social Media 2021 May 22 (Vol. 15, pp. 1034-1045).
17. Doshi AR, Schmidt W. Soft governance across digital platforms using transparency. Strategy Science. 2024 Jun;9(2):185-204.
18. MacCarthy M. Transparency is essential for effective social media regulation. Brookings Institution. November. 2022 Nov;1:2022.
19. UNESCO. Guidelines for the governance of digital platforms: Safeguarding freedom of expression and access to information through a multi-stakeholder approach. UNESCO; 2023.