ISSN: 2710-0170

Social Media Weakens Users' Critical Thinking and Requires Strategic Responses

Yuzhuo Xie

University of Glasgow, Glasgow, G12 8QQ, UK

Abstract

As social media becomes more and more popular, it has an increasing impact on the way people think. Many platforms recommend content based on criteria such as the number of likes and view time, a mechanism that makes emotional, one-sided and even misleading information more likely to be seen, while issues that are truly insightful and can help us think are instead buried. By reviewing data and analysing the policies of several countries (e.g., Germany, UK, EU, US, Singapore, China) and major platforms (e.g., Meta, YouTube, X, and TikTok), this study tries to figure out the causes of these problems and what current policies do well and what they don't do well. The study found that while countries and platforms have taken some measures to manage social media content, such as removing false information, protecting youth, and publishing transparency reports, many of the policies still lack coherence and platform governance is not open and transparent enough. At the same time, most of the current policies only focus on 'content', but neglect the understanding and judgement of the users themselves, especially adults, who rarely have the opportunity to receive relevant education or improve their information literacy. This paper suggests that the government, platforms, and society should work together to improve the transparency of platform governance, develop diverse recommendation algorithms, strengthen user education on information literacy, especially for young users, and establish a long-term, open, and monitorable regulatory system. It is hoped that through the cooperation of all parties, social media can become a space conducive to reflection and rational dialogue.

Keywords

Social media; critical thinking; recommend content; respond.

1. Introduction

With the pervasive integration of digital technologies in daily life, social media is deeply embedded in people's daily lives. By 2024, more than half of the world's population will be using social media, a proportion that continues to grow [1, 2], and people spend hours a day browsing, posting, and commenting on a variety of platforms. Behind this seemingly rich and convenient exchange of content lies a widely overlooked problem: social media may be quietly undermining our critical thinking skills, i.e., our ability to identify truth and falsehood, analyse information, and reflect on ideas.

This issue is important for several reasons. Firstly, the technological mechanisms of the platforms unconsciously influence the way we receive information. Meikle [3] points out that social media is not neutral, and behind it lies a highly commercialised logic - distributing information through algorithms to "personalise recommendations " to attract users' attention and prolong their stay. This mechanism results in users being presented with information that is not of their own choosing, but is the result of automated filtering by the platform based on predictions of interest, thus diminishing the possibility of being exposed to a wide range of perspectives and developing critical thinking [4]. This improves the experience of use in the

ISSN: 2710-0170

short term, but in the long term limits information diversity and openness to viewpoints, making us increasingly reluctant to engage with or think about different perspectives [5].

Secondly, social media platforms further interfere with rational thinking by reinforcing an 'emotional climate'. In order to increase the interaction rate, social media platforms often amplify emotions such as anger, prejudice and antagonism through the mechanism of likes, tweets and hot comments, so that people make quick judgements driven by emotions and are unwilling to think deeply [5]. In such an environment, people are more inclined to 'speak according to their feelings' rather than 'think and express', which significantly reduces the space for rational discussion and makes it difficult for critical thinking to be brought into play.

Moreover, The problem of 'cognitive overload' brought about by the excessive use of social media and the information fragmentation is also exacerbating users' mental fatigue [6]. When users are addicted to fragmented information platforms such as Tiktok and Twitter, they are not only easily distracted, but also lose the ability to understand the content in depth. More seriously, the research suggests that long-term social media dependence erodes the 'self-confidence', 'curiosity', and 'cognitive maturity' that are essential to critical thinking [6]. On the other hand, the instant gratification mechanism of social media is also eroding our patience and depth. Users' main motivations for using social media include entertainment, stress relief and getting quick feedback [7]. The thrill of likes, tweets and 'short videos' immerses us in constantly updated stimuli, gradually losing our interest and attention to deeper issues, thus negatively affecting critical thinking.

In addition, psychological research has also shown that people are more likely to question others and agree with themselves in social interaction, habitually neglecting to reflect on their own views, and only activating the 'deep thinking mode'[8] when they encounter obvious conflicts of viewpoints. This 'selective laziness' is even more pronounced on social media, and it is making public discourse increasingly superficial and polarised, preventing the formation of rational and fair public opinions.

While some studies have questioned the exaggeration of certain phenomena, such as Haim et al.'s [9] argument that 'filter bubbles' are not significant on some platforms, mainstream research agrees that the communication structure, platform design, and psychological motivations of social media do limit our ability to actively think, reflect on ideas, and understand complex information on a number of levels. ideas, and the ability to understand complex information.

If the majority of people increasingly rely on algorithms, emotions and rapid feedback to decide 'what to believe', society as a whole risks a decline in public judgement, reduced immunity to disinformation, and the proliferation of extreme views. This paper explores how social media undermines users' critical thinking skills and analyses the shortcomings of current governance, taking into account policy practices in different countries and the responses of major social platforms. Based on this, the paper puts forward a series of actionable policy recommendations to help governments, platforms and educational institutions work together to create a social media environment that is more conducive to rational thinking and information diversity.

It is important to note that this paper focuses mainly on the policy and platform regulation part, but the impact brought by social media cannot really be separated from the behaviour of users themselves. What people usually click on, watch and say on the platform will affect what content the platform pushes. And these choices, which are often subconscious reactions, are difficult to fully articulate in terms of policy or data. So while we are focusing on discussing improvements at the system level, the way users think and their habits of use are an equally important part of the problem.

ISSN: 2710-0170

2. Methodology

In this paper, in order to deeply explore the impact on users' critical thinking in social media information dissemination and to make actionable policy recommendations, A variety of research methods were adopted in this study, including data analysis, comparative analysis and literature review. This multi-dimensional research path not only ensures a multi-faceted presentation of the problem, but also provides a solid evidence base and theoretical support for the policy analyses and recommendations part of this study.

Firstly, In the data analysis section, I refer to global digital reports such as Digital 2024 [1] and Digital 2025 [2]. The latest data shows that by the beginning of 2025, the number of global social media users has grown from 5.07 billion in 2024 to 5.56 billion, accounting for 63.9 per cent of the world's total population, an increase of 1.3 percentage points compared to the previous year. This shows that social media has become an indispensable part of people's lives, not only as a major channel for obtaining information, but also increasingly influencing people's perceptions, topics of discussion, and even political decisions and the direction of social opinion.

At the same time, I also reviewed relevant data reports from organisations such as Statista and Pew Research Center to further understand the business logic and social impacts behind social media. Statista's data for 2023 [10] shows that Meta (the parent company of Facebook and Instagram) generates more than $134 billion in revenue in a year, almost ten times as much as a tech company like Microsoft, which focuses on office software. This suggests that there is a huge profit margin behind social media, so platforms are likely to continue to drive ad revenues and user dwell time by designing content mechanisms that are more appealing to users' attention. Additionally, a survey of teens published by the Pew Research Center in 2025 [11] shows that approximately 48% of teens believe that social media has a negative impact on their peers, an increase of 16 percentage points in three years. This indicates that adolescents are already being affected on a psychological and cognitive level, and this generation is the core citizen of the future society. It is therefore essential to study this issue and respond to it in terms of policy.

Secondly, in the comparative analysis section, in addition to the cross-checking analysis of the changes in data over the years, I have selected some representative countries and regions to understand what policies they have put in place in terms of social media governance and how these policies have responded to the issues of disinformation, platform liability and user rights. On the other hand, I also compared the practices of several major social media platforms (e.g. Meta, YouTube and TikTok, etc.) in terms of content management and information recommendation. By looking at both the government's regulatory measures and the platforms' own management mechanisms, I hope to gain a more comprehensive understanding of how the challenges posed by social media are faced in different contexts, and what problems there may be in the actual implementation of these practices.

Through a systematic literature review, I have sorted out representative research results from the fields of communication, social psychology and science and technology policy as the theoretical foundation for the policy analysis. This literature not only helped me to better understand how social media affects users' mindsets, but also provided me with strong support for my policy recommendations. For example, a study of Cheng et al. [6] found that university students' ability to think deeply may be affected when faced with a large amount of fragmented and emotional information. Some scholars have also put forward the theories of 'filter bubbles' and 'echo chambers, which explain why social media tends to expose users to only a single viewpoint. Theories such as 'affective scaffolding', 'instant gratification' and 'selective laziness of reasoning' explain why many users are more likely to be attracted by emotional and fragmented information rather than actively checking the authenticity of the content. not actively checking the authenticity of the content. These kinds of analyses on the psychological

ISSN: 2710-0170

level of users are difficult for me to see directly through pure data or policy documents in my research, so the literature review fills this part of the gap very well.

It is worth mentioning that although this study balances data, policy and theory in its methodology, it also faces some practical limitations. For example, the limited disclosure of information about the algorithmic logic and content management process of the platforms has led to a blind spot in the assessment of policy effects; at the same time, due to the lack of on-site interviews and user surveys, some data on the public's media literacy and information perception preferences still need to be supplemented.

3. Policy Analysis-National and Regional Policies' Analysis

Against the background of the rapid development of global digitalisation and social media, governments have long begun to realise the increasing impact of the information environment on public perception, especially on critical thinking skills. Emotional content, personalised recommendations and information filtering mechanisms are not only shaping the way users access information, but are also gradually changing people's ability to process information and judge its authenticity. In response to these challenges, different countries and regions have adopted distinctive policy instruments in an attempt to exert governance pressure on social media platforms, improve the information ecosystem and enhance the public's ability to recognise information.

Germany was one of the first countries to make strong policy interventions in this area. The Network Enforcement Act (NetzDG) has since 2018 required platforms to remove clearly illegal content within 24 hours or face fines of up to €50 million [12]. This this has dramatically increased platforms' responsiveness and review processes. However, this mandatory regulation has also sparked a lot of controversy as some platforms may choose to "over-delete" content to avoid legal risks [13], thus undermining the normal space for public discussion and freedom of expression. With a research report [14] pointing out that nearly 99% of deleted comments on Germany,Facebook and YouTube are actually legitimate.

The Digital Services Act (DSA) [15], introduced by the EU in 2022, focuses on the operational mechanisms behind the platforms. It requires large platforms to disclose the principles of their algorithms, publish content audit reports and provide access to users complaints, while encouraging regulators and researchers to participate in monitoring.After the DSA was enacted, TikTok voluntarily shut down the 'Lite Reward' function that encouraged users to swipe short videos [16], and several platforms have also added clearer reporting paths and disclosed some of their algorithmic information. Meanwhile, the EU has initiated a formal investigation process against platforms (e.g. X) that violate the DSA [17]. However, some scholars [18] have found that platforms' performance in enforcing the DSA varies from high levels of compliance to incomplete information. Moreover, the data they submit to the DSA database is sometimes inconsistent with other public sources, suggesting that platforms are not yet sufficiently transparent and that it is difficult for outsiders to accurately judge their actual practices.

The United Kingdom has taken a 'step-by-step' approach to social media regulation. In the earliest days, it adopted the Online Harms White Paper [19], which set out directions for setting up a regulator, encouraging platforms to self-regulate, and protecting young people. Despite the early lack of legal constraints, this policy laid the groundwork for later formal legislation, and in 2023 the Online Safety Act [20] was passed, making it the first law in the UK to specifically regulate the content of online platforms. It requires platforms to take responsibility for the safety of their users, especially children, and must remove illegal or harmful content and provide tools such as reporting and blocking. Regulator Ofcom has also been given greater powers to fine offending platforms or restrict services. By the end of 2024, the policy had begun to be implemented, with large social media platforms such as Meta and TikTok required to

ISSN: 2710-0170

submit assessments of 'illegal content' and 'risk to children' [21], and Ofcom began proactively enforcing the policy in 2025. Ofcom began proactively enforcing the Illegal Content Review in early 2025 and promoting engagement with smaller platforms [22], and within six months,Ofcom had carried out at least nine enforcement investigations [23], although its focus was on children and types of sites, and the number of children and children's organisations that had been subject to the review was low. Public information on social media remains limited. Meanwhile, in 2024, the Science, Innovation and Technology Committee also launched an investigation into "social media disinformation and harmful algorithms". Although the UK has not yet legislated in this regard, it has proposed directions to enhance the transparency of algorithms and external supervision. It indicates that the UK may take substantive steps in platform algorithm governance in the future [24].

The United States has shown a complex and ambivalent approach to social media governance. On the one hand, Section 230 of the Communications Regulation Act guarantees that platforms are not responsible for the speech of their users, making it difficult for the government to intervene directly in content [25]; on the other hand, in the face of the proliferation of disinformation during elections and epidemics, the government has had to intervene. For example, the Department of Homeland Security (DHS) and its subsidiary Customs and Border Protection (CBP) have established the Social Media Monitoring and Situational Awareness System, which is a system for monitoring and monitoring the situation in the country. Awareness System" to conduct keyword searches and real-time tracking of content on public platforms to identify potential security threats [26]. In 2024, the U.S. passed the Protecting Americans from Foreign Adversary Controlled Applications Act [27], which requires TikTok to complete its divestiture or exit the U.S. market by early 2025 or else it will face a ban. This action is partly a reflection of the government's heightened vigilance over the potential for social media to influence users' thinking and social perceptions. Meanwhile, in recent years, a number of congressional proposals have sought to amend Section 230 to hold social media platforms more accountable for their algorithmic recommendations and advertising content [28]. It is worth noting that a bill to limit platforms' addictive recommendation algorithms has been proposed in the US state of Washington, which has not yet been legislated but shows an active attempt by local governments to protect youth and regulate algorithms [29].

In Asia, Singapore has adopted a more centralised and efficient response model. Its Prevention of Online Falsehoods and Manipulation Act (POFMA), introduced in 2019, allows the government to compel platforms, media outlets, and even individuals to issue corrections, remove information, or block accounts [30]. As of August 2024, 152 orders have been issued under POFMA, mainly focusing on public health and order, with more than half of the orders involving disinformation related to the New Crown epidemic [31]. This policy has dramatically increased the efficiency of the government's response to disinformation and has forced platforms to operate with a high degree of vigilance and cooperation. However, this highly centralised mechanism has also led to criticisms of political manipulation and suppression of speech, especially when the power to determine what is false is concentrated in the hands of the government, and the lack of an independent judicial review mechanism is difficult to convince the public.

Social media governance in China adopts a hybrid approach of 'government-led, platform-coordinated'. Under the guidance of relevant policies, such as the Regulations on the Ecological Governance of Network Information Content [32] and the Regulations on the Management of Internet User Account Information [33], the government, through the Office of the Central Cyberspace Affairs Commission and other departments, strictly regulates social platforms, including content publishing, account management, comment review, and even recommendation algorithms. These platforms have also created a large team of content reviewers in line with government policy, prioritising positive, officially sanctioned content to

ISSN: 2710-0170

the general public. In addition, international platforms such as Facebook and Instagram are not available in China, and users mainly rely on domestic platforms for information [34]. This governance mechanism excels in terms of implementation efficiency and public opinion control, but has some limitations in terms of the diversity of public opinion discussions and the openness and global nature of information.

From a comprehensive perspective, different countries have taken their own unique approaches to controlling social media, with their own successes and problems. The European Union promotes more openness and transparency on platforms, which helps to improve the quality of information, but implementation is slow and platforms' cooperation varies. Germany focuses on 'fast-tracking', requiring platforms to remove illegal content within a short period of time, which improves the speed of response, but also makes it easy to mistakenly delete normal speech. The UK has moved from encouraging self-regulation to legislative regulation, but specific measures are still being implemented. The United States is difficult to directly intervene in the content due to the protection of freedom of speech, but it has indirectly made efforts by blocking TikTok and pushing for modification of Article 230, and the effectiveness is still under observation. Singapore adopts centralised management, using laws to force platforms to correct or remove false information, which is highly efficient in governance but tends to raise concerns about too much government power. The Chinese model is the most centralised, with the government strictly controlling content and platforms co-operating in enforcement, effectively maintaining information order, but also limiting the diversity of information and the space for critical thinking by users. Overall, Europe and the United States favour transparency and rules, while some Asian countries focus more on efficiency and order, but how to balance security, freedom and diversity is a common challenge for all countries.

4. Platform Policies' Analysis

In social media governance, in addition to national policies, platforms' own content recommendations, user management and business models are also key to influencing user perceptions. More and more studies have found that platforms are not neutral tools, but actively influence users' emotions and thinking. Therefore, analysing the platforms' own policy responses and shortcomings is an important step towards more effective governance.

Meta's Facebook and Instagram have been optimising their governance of misleading and emotional content in recent years. According to their latest platform implementation reports [35], the platforms seek to reduce the impact of misinformation without outright deletion by flagging, restricting distribution, and even deleting suspicious content through a 'tiered processing' mechanism. The platforms also introduces a 'community notes' feature that allows users to add contextual information and provide multiple perspectives on the content [36]. Meanwhile, Meta has introduced an age-appropriate content management policy for different age groups, with special protection for teenage users [37]. The platform automatically restricts the visibility of certain types of content based on a user's age, such as videos and posts involving violence, nudity or other adult themes. The platforms under the Meta umbrella also adopt different operational strategies by adapting content management and review standards to the legal and regulatory requirements of each country [38]. Meanwhile, Meta regularly publishes processing data in its Transparency Centre [39], including the number of removals, the type of violation, and the outcome of appeals in an effort to increase public trust. However, the logic of the platform's algorithmic recommendation remains undisclosed and is easily influenced by subjective judgement. In addition, the platforms' criteria for 'downgrading' information are not transparent, making it difficult for users to know whether the content they see has been modified or restricted.

ISSN: 2710-0170

YouTube, one of the world's most popular video platforms, has established a 'three-tier warning' system through its Community Guidelines, whereby false, hateful or violent content is graded and, in severe cases, banned or blocked [40]. Meanwhile, YouTube also uses AI technology to automatically identify high-risk content and reduce the frequency of its recommendations. For sensitive topics (e.g. vaccines, wars, elections, etc.), the platform also directs users to authoritative sources of information [41], in an attempt to maintain a balance between free speech and public safety. In terms of youth protection, YouTube has adopted a 'dual-track' content management strategy. For children under the age of 13, the platform has launched a separate app, YouTube Kids [42], which provides a safer viewing environment for kids through age rating, parental control and content review. Parents can set viewing hours, block specific channels, and restrict search functions. For teens, YouTube also offers a 'supervised experience' that allows parents to choose the level of content they want to see based on their child's developmental stage [43], as well as control ads, comments, and recommendation settings. This mechanism helps young people to gradually access a wider range of content within a controlled range. However, there are still loopholes in YouTube's auditing, with an over-reliance on parental supervision and a lack of guidance for critical thinking among young people. Although the platform has improved transparency through policy enforcement [44], the recommendation algorithm and content distribution mechanism still need to be improved. At the same time, because it relies on creators to generate revenue, YouTube tends to be more conservative in dealing with non-compliant accounts, which affects the rigour of governance.

After its acquisition by Musk in 2022, X (formerly Twitter) dramatically adjusted its content governance strategy, dismantling its original censorship team and fact-checking partnerships in favour of an emphasis on "absolute freedom of expression" [45]. The platform now relies heavily on the user-added Community Notes annotation mechanism [46], which is controversial for its slow review and limited coverage. Although X continues to ban manipulation, hate speech and misleading content in its rules [47], the lack of a professional review team and independent oversight has increased the visibility of misleading or polarising content on the platform. Some studies have pointed out [48] that the frequency of hate speech and conspiracy theories spreading on X has increased significantly since the governance mechanisms were adjusted, suggesting that user autonomy and automated systems alone are still difficult to effectively support the governance of complex information environments. It also calls into question whether its platform ecology is still conducive to critical thinking.

TikTok, a platform originating in China and operating globally, has localised its content management strategy to suit different countries,and has set strict "authenticity and trustworthiness" standards that prohibit the manipulation of the platform's algorithms, the dissemination of false information or the erosion of community trust [49]. The platform explicitly requires uploaded content to be authentic and original, and ensures traceability and transparency of information. TikTok's latest community rules emphasise a balance between preventing harm and safeguarding freedom of expression, and employ a parallel mechanism of automated detection and manual review, whereby allegedly unlawful content is first blocked, and then further examined or deleted by the review team [50]. Although TikTok has invested heavily in AI review (e.g., significant cuts in content reviewers in Malaysia and a shift to AI), the elimination of manual review may pose risks of review blind spots and bias [51].In addition, for teenage users, TikTok restricts the use of TikTok to users under the age of 13 in its Terms of Service [52], and provides 'family pairing,' 'time limits,' and 'content filtering' and other parental monitoring tools is to ensure that teens are exposed to the platform's content in a safer environment. At the same time, the platform's rule that videos from creators under 16 cannot appear on the 'Recommended Page' reduces potentially misleading influences at the source. While TikTok has done a good job of content vetting and youth protection, its recommendation

ISSN: 2710-0170

mechanism places a higher value on interaction and viewing time, which makes it easier to push short, eye-catching content, and makes it harder for in-depth content to be seen.

In summary, although major platforms have made some management measures on the surface, such as setting up fact-checking, community rules and content review tools, their governance tends to be fragmented and inconsistent, and is still more at the discretion of the platforms themselves, with the main aim of attracting traffic and making money. Many policies lack real external oversight and are not very transparent. These problems make it difficult for platforms to create an environment of pluralistic information, rational discussion and open-mindedness, which is precisely the key to cultivating critical thinking among users. In fact, the platform itself both creates a lot of information problems and has the ability to be the party that solves them. What is really difficult is how to make these platforms assume their due social responsibility, rather than just focusing on serving algorithms and commercial interests.

While this part of the analysis focuses on the policy and platform management perspective, we cannot ignore the impact of users' own behaviour. Each person's choices to click, retweet, and comment on a platform affect what content the platform pushes. The psychological motivations behind these behaviours are complex and difficult to fully reflect with data or regulations. Therefore, to truly improve the impact of social media on ways of thinking, in addition to strengthening regulation, we also need to focus on users' habits and judgement.

5. Analysis of existing policies' issues

Although governments and major social platforms have introduced a number of governance policies against disinformation, extreme content, and algorithmic recommendations, from the perspective of critical thinking cultivation and the health of the information ecosystem, many structural problems are still exposed in the implementation of existing policies. These problems not only constrain the realisation of policy effects, but also exacerbate public distrust of the information environment, especially as the cognitive space of young user groups on social platforms is increasingly closed and emotional, and the problem awaits a systematic response.

Firstly, there is a lack of coordination between national policies. For example, the EU and Germany have passed legislation to force platforms to increase transparency and remove illegal content, but these measures sometimes overly restrict freedom of expression. On the contrary, although the United States attaches importance to free expression, regulation lacks uniform legal support, and platforms rely on voluntary cooperation for implementation, making it difficult to form substantive constraints. Other countries, such as Singapore and China, have adopted strong interventionist measures, which are highly efficient but can easily lead to concerns about 'overregulation'. Globally, standards are not uniform across countries, and co-operation mechanisms are not sound, resulting in platforms implementing different rules in different regions, and user experience and information quality vary greatly.

Secondly, the self-regulatory mechanisms of platforms seem to be many, but the implementation is quite problematic. For example, Meta and YouTube will mark and downgrade some content, but the judgement criteria behind these operations are not transparent, making it difficult for users to know why the content is 'unseen', and Platform X even puts the responsibility of auditing on the user community, which encourages participation, but due to the lack of professional gatekeeping, many false or extreme comments are still being made. Although TikTok has done a good job of protecting young people, the content it recommends still favours entertainment and emotional stimulation, making it difficult for serious, rational content to be pushed to the front.

In addition, most policies currently focus on 'content management', with little attention paid to users' own understanding and critical thinking skills. In reality, users' own psychology and behaviour are also crucial, for example, many people are more easily attracted by emotional

ISSN: 2710-0170

and simplified information, and less willing to think about complex or opposing views. Without good education or guidance, when faced with information that is difficult to distinguish between truth and falsehood and with obvious bias, users often can only judge by feelings, which is not only easy to misbelieve, but also not conducive to the development of critical thinking [53]. So far, neither countries nor platforms have long-term plans to improve the information judgement of social media users.

Finally, the policies and rules themselves are not transparent enough. Platforms do not disclose their recommendation algorithms, nor do they know what kind of content is downgraded or deleted; and there are few channels for users to understand or participate in the government's regulatory process. This leaves users in a state of 'not knowing who is in charge, how and why', with a lack of trust and a lack of space for feedback.

Overall, although the existing policy has some positive effects, there is still a big gap between it and building a healthy and open information environment that is conducive to the development of critical thinking. This is precisely where we need to start in making further recommendations for improvement.

6. Conclusion and Policy Recommendations

In an era where social media shapes how we consume and interpret information, its negative impact on users' critical thinking cannot be ignored. While countries and platforms have introduced various governance measures, most remain fragmented, lacking transparency and long-term strategy. Equally important, user behavior and media literacy—especially among adults—are often overlooked. To counter the erosion of rational discourse, a collaborative approach is essential. Governments, platforms, and society must co-develop transparent systems, promote diverse content exposure, and invest in media literacy education. Only by addressing both systemic design and human factors can we build an information environment that truly supports critical thinking.

6.1. Government Enables Multi-Party Collaboration

In order to more effectively address the impact of social media on the public's critical thinking, we recommend a 'multi-stakeholder' policy direction that emphasises collaboration between the government, platforms and the public. First, the government should take stronger and more detailed measures to regulate platforms, especially in enhancing their transparency. Currently, the recommendation algorithms and content moderation mechanisms of many platforms are still not open to the public, making it difficult for users to understand why they are seeing certain content, or which posts have been 'downvoted' and made inaccessible. In this regard, governments could look to the EU's Digital Services Act (DSA) and legislate to require large platforms to publish regular transparency reports detailing their content moderation, flagging, recommendation and blocking mechanisms.But it's not just enough to 'require publication'.

Governments should not only set transparency reporting requirements based on local laws with a uniform format and minimum content standards, but also establish an independent body with the capacity to actually review platform reports, and bring in civil society organisations or academic teams to participate in the review process in order to prevent platforms from 'doing things superficially'.In addition, the transparency report should cover key data such as the logic of the platform's internal policy adjustments, the rate of erroneous deletions, and the dispute handling process, and open up feedback channels to the public. The government should also set up a punitive mechanism. If a platform is found to have intentionally concealed or falsely reported, it should be warned, fined, or have its operating privileges restricted in accordance with the law. Only when platforms know that their reports will be subject to real scrutiny and

ISSN: 2710-0170

face real consequences will transparency cease to be formalism and become an important tool that can actually enhance public trust and platform accountability.

6.2. Platforms promote social responsibility

Platforms themselves should seriously reflect on whether their content recommendation mechanisms are quietly influencing the way users think. Currently, most social media platforms still use interaction data as their main recommendation criteria. Platforms should balance commercial interests and social responsibility, try to optimise their algorithmic logic, gradually introduce the principle of 'diversity recommendation', and take more dimensions into account when recommending content, such as the credibility of the source of information, whether it covers a wide range of viewpoints, and whether it has an emotional or misleading tendency. For example, when recommending videos on social issues, the system can appropriately intersperse interpretations from different positions rather than content with a single tendency, so as to create a more balanced atmosphere for discussion.

Meanwhile, mechanisms such as Meta's 'community notes' or X's 'user annotations' are encouraging explorations that can stimulate user participation, help add background information and alert others to content biases. However, without professional audit support, such mechanisms are prone to be too subjective and misleading. It is recommended that platforms introduce a certain percentage of third-party experts or a team of trained editors to conduct secondary audits of highly controversial content to ensure the accuracy and neutrality of community labelling. In addition, platforms should establish a more open and transparent feedback mechanism. Currently, many platforms do not disclose the operational criteria for 'downgrading' or 'marking' content, so users often do not know why their content is restricted and have no way to complain. In order to enhance public trust, platforms should regularly publish real impact assessment reports on content governance and recommendation mechanisms, and invite independent organisations and academic research teams to conduct reviews and public assessments, and accept social supervision. This will not only help improve the quality of governance, but also encourage platforms to give more consideration to their social responsibility towards the public discussion environment beyond commercial gains.

6.3. Information Literacy Enhancement for Youth

It is also important that platforms and governments should work together to promote information literacy education, especially among youth. The current policy focuses more on 'content control' and ignores whether users have the ability to recognise the authenticity of information and maintain independent thinking. As a result, many users, especially teenagers, can only rely on their instincts or emotions to make judgements in the face of fast, fragmented and emotional information on social media, and lack the ability to make in-depth analyses and think from multiple perspectives. This not only undermines their critical thinking, but may also affect the quality of public discourse in society as a whole in the long run. Therefore, the Government can formally incorporate information literacy into the primary and secondary education system as a basic competency, so as to cultivate in students from an early age the ability to identify misleading information, understand differences in opinions and avoid the 'echo chamber effect' in the online environment. The curriculum can incorporate real-life social media cases, so that students can learn basic skills such as recognising headline-grabbers, identifying biased language and using authoritative information sources. At the same time, the government can encourage teacher training and resource development, and support schools in integrating these elements into their daily teaching, rather than just as an ad hoc supplementary programme.

Social media platforms themselves should also take the initiative to assume responsibility for education. Currently, many platforms still rely heavily on 'parental control' mechanisms to protect young people (YouTube Kids, TikTok Family Pairing, etc.), but these mechanisms often

ISSN: 2710-0170

lack consistent guidance in actual use and are highly dependent on parental digital literacy, which makes them difficult to be effective in the long term. Platforms can consider developing more interactive and guiding educational tools. For example, when users browse sensitive or controversial topics, 'relevant background information' or 'reminder of diversity of opinions' will pop up automatically; 'misleading information identification exercise' module will be provided to encourage users to participate in marking or judging the authenticity of the information and give feedback; and 'social media civic literacy programme' can even be launched in cooperation with schools or public organisations so that young people will not only be the recipients of information, but also become the participants of rational communication.

More importantly, governments and platforms need to work together. The two sides can set up a 'digital literacy co-operation mechanism', with the participation of the education sector, representatives of various platforms and independent research institutes, to formulate a unified standard for information literacy development, and integrate online platform resources with offline education practices, so as to ensure that every young person can gradually build up the judgement and discernment skills needed to deal with complex information in the process of growing up. This will not only help to reduce social media misinformation, but will also help to reduce the number of young people being misled by social media. This will not only help reduce the impact of social media misinformation, but also improve the public's adaptability and resistance to the information environment.

6.4. Information Literacy Enhancement for Adult Users

In addition to the protection of young people, the enhancement of information literacy among adult users is equally important but often overlooked. Many adults spend a lot of time on social media every day, but lack the awareness and ability to identify true and false information and analyse multiple viewpoints. Especially when it comes to sensitive issues such as politics and health, adults are more likely to fall into the 'echo chamber' of information recommended by their circle of acquaintances or algorithms, and are only exposed to content that supports their own stance, which further exacerbates emotional judgement and social division. Adult users' misinformation is often more difficult to correct than that of teenagers, as they are less likely to engage in formal education and are less willing to change their perceptions.

Therefore, it is recommended that the government and platforms work together to develop information literacy programmes for adult users, for example, by organising short courses or workshops on 'recognising information on the internet' through public channels such as local communities, libraries and news platforms, or by incorporating simple, practical and interactive tools into social media platforms such as Simple and practical interactive tools can also be added to social media platforms, such as the 'Look at it from another angle' button, which allows users to see different opinions or positions on the same news, guiding them to proactively expand their sources of information, and helping them to practise multi-perspective thinking.

In addition, platforms can introduce a 'usage alert' mechanism, for example, when users continuously forward unverified content multiple times, the system automatically reminds users to check the source, or recommend relevant fact-checking pages. This approach will not infringe on freedom of expression, but also gradually influence the user's habits, so that rational thinking becomes a subtle social norm. Although it is difficult to directly change the behaviour and psychology of adults through a single policy, the public's power of discernment and judgement can still be enhanced on a wider scale through the long-term building of information literacy, coupled with the optimisation of platform mechanisms.

Overall, the solution to the impact of social media on critical thinking should not rely solely on deleting or blocking content, but rather on building a transparent, rational and responsible information environment at the source. This requires the government to formulate stronger

ISSN: 2710-0170

rules and stricter censorship procedures, platforms to improve the algorithmic mechanism for recommending content, and guidance for people to learn to distinguish information and think independently. At the same time, the government, platforms and social organisations should work together to improve the information literacy of users, and develop the ability to think critically from the ground up. Only through the efforts of the government, platforms and society can we truly build a cyberspace that is conducive to serious thinking and free discussion.

References

[1] Kemp, S. (2024). Digital 2024: Global Overview Report. https://datareportal.com/reports/digital-2024-global-overview-report

[2] Kemp, S. (2025). Digital 2025: Global Overview Report. https://datareportal.com/reports/digital-2025-global-overview-report

[3] Meikle, G. (2024). Communication. In G. Meikle (Ed.), Social media: the convergence of public and personal communication (2nd ed., pp. 15-34). Routledge.

[4] Pariser , E. (2011). The race for relevance. In E. Pariser (Ed.), The Filter Bubble: What the internet is hiding from you. The penguin Press.

[5] Steinert, S., Marin, L., & Roeser, S. (2025). Feeling and thinking on social media: emotions, affective scaffolding, and critical thinking. Inquiry, 68(1), 114-141.

[6] Cheng, L., Fang, G., Zhang, X., Lv, Y., & Liu, L. (2022). Impact of social media use on critical thinking ability of university students. Library Hi Tech, 42(2), 642-669.

[7] Quan-Haase, A., & Young, A. L. (2010). Uses and gratifications of social media: A comparison of Facebook and instant messaging. Bulletin of science, technology & society, 30(5), 350-361.

[8] Trouche, E., Johansson, P., Hall, L., & Mercier, H. (2016). The selective laziness of reasoning. Cognitive Science, 40(8), 2122-2136.

[9] Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. Digital journalism, 6(3), 330-343.

[10]Stacy, J. D. (2023). Social Media Statistics. https://www.statista.com/topics/1164/social-networks/.

[11]Faverio, M., Anderson, M., & Park, E. (2025). Teens, Social Media and Mental Health. https:// www.pewresearch.org/internet/2025/04/22/teens-social-media-and-mental-health/

[12]Wikipedia. (2022). Network Enforcement Act. https:// en.wikipedia.org/ wiki/ Network_ Enforcement_Act

[13]Human Rights Watch. (2018). Germany: Flawed Social Media Law. https://www. hrw.org/news/ 2018/02/14/germany-flawed-social-media-law

[14]The Future of Free Speech. (2024). Preventing 'Torrents of Hate' or Stifling Free Expression Online? https://futurefreespeech.org/preventing-torrents-of-hate-or-stifling-free-expression-online/

[15]European Commission. (2022). The EU's Digital Services Act. https://commission. europa.eu/ strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

[16]European Commission. (2024). TikTok commits to permanently withdraw TikTok Lite Rewards programme from the EU to comply with the Digital Services Act. https:// ec.europa.eu/ commission/presscorner/detail/en/ip_24_4161

[17]Reuters. (2025). EU says it is probing corporate structure of Musk's X months after xAI deal. https://www.reuters.com/business/eu-probes-musks-xai-buyout-x-bloomberg-news-reports-2025-06-19/

[18]Trujillo, A., Fagni, T., & Cresci, S. (2025). The DSA Transparency Database: Auditing self-reported moderation actions by social media. Proceedings of the ACM on Human-Computer Interaction, 9(2), 1-28.

[19]GOV.UK. (2020). Online Harms White Paper: Full government response to the consultation. https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response

ISSN: 2710-0170

[20]GOV.UK. (2023). Online Safety Act: Explainer. https://www.gov. uk/government/ publications /online- safety-act-explainer/online-safety-act-explainer

[21]Reuters. (2024). Britain sets first codes of practice for tech firms in online safety regime. https://www.reuters.com/world/uk/britain-sets-first-codes-practice-tech-firms-online-safety-regime-2024-12-16/

[22]Ofcom. (2025). Enforcing the Online Safety Act: Platforms must start tackling illegal material from today. https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/enforcing-the-online- safety- act-platforms-must-start-tackling-illegal-material-from-today

[23]Ofcom. (2025). Enforcing the Online Safety Act: Ofcom opens nine new investigations. https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/enforcing-the-online-safety-act-ofcom-opens-9-new-investigations

[24]UK Parliament. (2024). Social media, misinformation and harmful algorithms. https:// committees.parliament.uk/work/8641/social-media-misinformation-and-harmful-algorithms

[25]CONGRESS. GOV. (2024). Section 230: an Overview. https://www.congress.gov/crs-product/R46751

[26]U.S. Department of Homeland Security. (2019). DHS/CBP/PIA-058 Publicly Available Social Media Monitoring and Situational Awareness Initiative. https://www.dhs.gov/publication/dhscbppia-058-publicly-available-social-media-monitoring-and-situational-awareness

[27]CONGRESS.GOV. (2024). H.R.7521 - Protecting Americans from Foreign Adversary Controlled Applications Act. https://www.congress.gov/bill/118th-congress/house-bill/7521.

[28]Musquera, A., & Brennen, J. S. (2025). What Has Congress Been Doing on Section 230? https://www.lawfaremedia.org/article/what-has-congress-been-doing-on-section-230

[29]Transparency Coalition. (2025). A House committee in Olympia is stalling this bill to protect kids from addictive apps. https://www.transparencycoalition.ai/news/house-committee-in-olympia-is-stalling-this-bill-to-protect-kids-from-addictive-apps

[30]Singapore Statutes Online. (2019). Protection from Online Falsehoods and Manipulation Act 2019. https://sso.agc.gov.sg/Act/POFMA2019

[31]Goh, Y. H. (2024). Five years of Pofma: How has the law been used to combat fake news? https://www.straitstimes.com/singapore/five-years-of-pofma-how-has-the-law-been-used-to-combat-fake-news

[32]Office of the Central Cyberspace Affairs Commission. (2019). Regulations on the Ecological Governance of Network Information Content. https://www.cac.gov.cn/2019-12/20/c_1578375159509309.htm

[33]Office of the Central Cyberspace Affairs Commission. (2022). Regulations on the Management of Internet User Account Information. https://www.cac.gov.cn/2022-06/26/c_ 1657868775042841.htm

[34][Zucchi, K. (2021). Why Facebook Is Banned in China & How to Access It. https:// www.investopedia.com/articles/investing/042915/why-facebook-banned-china.asp

[35]Meta. (2022). Taking action. https://transparency.meta.com/zh-cn/enforcement/taking-action/

[36]Meta. (2025). Community Notes: A New Way to Add Context to Posts. https:// transparency. meta.com/en-gb/features/community-notes/.

[37]Meta. (2024). Helping Teens See Age-Appropriate Content. https://transparency.meta.com/zh-cn/policies/age-appropriate-content/

[38]Wikipedia. (2019). Censorship of Facebook. https:// en.wikipedia. org/wiki/ Censor ship_ of_Facebook.

[39]Meta. (n.d.). Transparency reports. https://transparency.meta.com/reports/

[40]YouTube. (n.d.). Community Guidelines. https:// www.youtube. com/intl/ ALL_uk/ howyoutubeworks/ policies/community-guidelines/#taking-action-on-violations

[41]YouTube. (n.d.). Our Policies. https://www.youtube.com/howyoutubeworks/our-policies/

[42]Popular Timelines. (2019). History of YouTube Kids in Timeline. https:// populartimelines. com/timeline/YouTube-Kids/full

ISSN: 2710-0170

[43]YouTube. (n.d.). Kid & Teens. https://www.youtube.com/howyoutubeworks/kids-and-teens/

[44]Google Transparency Report. (2024). Disposal of content that violates the YouTube Community Guidelinest. https://transparencyreport.google.com/youtube-policy/removals

[45]Desmarais, A. (2025). Hate speech on X up 50% under Musk, new study finds. https://www.euronews.com/next/2025/02/13/hate-speech-on-x-now-50-higher-under-elon-musks-leadership-new-study-finds

[46]X. (2021). About Community Notes on X. https://help.x.com/en/using-x/community-notes

[47]X. (n.d.). The X rules. https://help.x.com/en/rules-and-policies/x-rules

[48]Chuai, Y., Tian, H., Pröllochs, N., & Lenzini, G. (2024). Did the roll-out of community notes reduce engagement with misinformation on X/Twitter? Proceedings of the ACM on Human-Computer Interaction, 8(CSCW2), 1-52.

[49]Tiktok. (2024). Overview. https://www.tiktok.com/community-guidelines/en/overview

[50]TikTok. (n.d.). Content violations and bans. https://support.tiktok.com/en/safety-hc/account-and-user-safety/content-violations-and-bans

[51]Latiff, R. (2024). ByteDance's TikTok cuts hundreds of jobs in shift towards AI content moderation. https://www.reuters.com/technology/bytedance-cuts-over-700-jobs-malaysia-shift-towards-ai-moderation-sources-say-2024-10-11/

[52]TikTok. (2025). Safety Resources for Parents, Guardians, and Caregivers. https://www.tiktok.com/safety/en/guardians-guide/

[53]Polanco-Levicán, K., & Salvo-Garrido, S. (2022). Understanding social media literacy: A systematic review of the concept and its competences. International journal of environmental research and public health, 19(14), 8807.