DOI: 10.54254/2753-7048/2025.NS27539

# *The Psychological Mechanisms and Legal Regulation of Information Manipulation on Social Media*

**Zijun Nie**

*Jinling High School Hexi Campus, Nanjing, China*
*44013818@qq.com*

***Abstract.*** In the algorithm-driven landscape of social media, platform manipulation of user cognition and behavior has become increasingly prominent, shaping public opinion and social perception as a critical influencing factor. Based on psychology and law, the paper defines information manipulation, outlines its primary types and psychological mechanisms, and explores how cognitive bias, emotional drive, and social influence jointly contribute to its systemic influence on individual cognition, public opinion, and information security. By integrating literature review with case analysis, it uncovers key regulatory challenges and puts forward a layered governance model that emphasizes platform accountability, technical oversight, and psychological intervention. The results reveal that information manipulation on social media platforms operates through the interplay of algorithmic filtering, emotional amplification, and social influence, consistently altering user cognition and behavior while posing major challenges to public discourse, emotional autonomy, and information security.

***Keywords:*** Information Manipulation, Social Media, Cognitive Bias, Legal Governance, Psychological Intervention

## 1. Introduction

In the digital age, social media platforms have shifted from neutral intermediaries to active agents of information manipulation. Through algorithmic recommendations, content filtering, and ranking mechanisms, they not only reshape the flow of information but greatly influence users' cognitive processes and decision-making behavior. Despite promises of improved user experience, concerns over manipulation continue to rise. Existing research has investigated the impact of algorithms on user behavior or discussed governance strategies, yet comprehensive research on the psychological mechanisms underlying platform manipulation and corresponding legal responses remains limited. This paper aims to analyze the potential impact of information manipulation on users' judgment and behavior from the perspectives of cognitive biases, emotional drivers, and social influence, and further examines issues such as platform responsibility definition, feasibility of legal intervention, and difficulties in regulatory enforcement. Through literature analysis and case studies, this study offers interdisciplinary insights from psychology and law, aiming to propose practical interventions and governance strategies. Furthermore, it seeks to deepen the understanding of manipulation in digital environments and provide theoretical support for policy-making and the enhancement of media literacy.

## 2. Overview of information manipulation behaviors on social media platforms

### 2.1. Core concept and operational mechanisms

In algorithm-driven media ecosystems, information manipulation refers to the deliberate shaping of content exposure and interpretation via selection, framing, and contextual control. By reordering truthful content, omitting contextual information, or amplifying emotional cues, it subtly influences perception and behavior, without resorting to blatant falsehoods [1-3]. This form of manipulation is structurally embedded within platform architectures, most notably through algorithmic curation. These algorithms filter and rank content based on user-specific data profiles, thereby constructing personalized yet epistemically constrained information environments. Consequently, the visibility of content is determined less by its relevance or veracity and more by behavioral tendencies and predicted engagement metrics [4,5]. Moreover, emotional optimization acts as a central mechanism in this process, as platforms promote emotionally charged content that evokes anger, fear, or affirms group identity in order to maximize user retention. By prioritizing emotional salience, these systems shape public discourse around affective polarization and reduce the space for deliberative reasoning [2]. In the long term, algorithmic selection and emotional reinforcement converge to produce echo chambers, which are closed loops of confirmatory content that systematically filter out dissent. This dynamic constrains cognitive adaptability, amplifies group polarization, and erodes the foundations of collective epistemic coherence [6]. In essence, modern information manipulation operates not through outright falsehoods but through the systematic shaping of attention, emotion, and exposure, reflecting a shift from deliberate deception to ambient influence driven by platform design.

### 2.2. Potential risks and social consequences

The manipulation of information on social media platforms presents significant risks both at the individual and societal levels. At the individual level, prolonged exposure to algorithmically filtered and emotionally driven content impairs users’ ability to critically assess the authenticity and source of information. This fosters cognitive dependence on the platform, where users begin to passively accept information without questioning its validity, thus resulting in the erosion of media literacy and diminished critical thinking skills. In addition, continuous exposure to polarized or emotionally charged content fosters negative emotions such as anxiety and anger, thereby leading to emotional dysregulation and behavioral shifts [7]. At the societal level, information manipulation reshapes the processes through which public opinion is constructed and disseminated. In particular, emotionally charged content further intensifies polarization, making it increasingly difficult for individuals with opposing views to reach mutual understanding, which fractures public dialogue and fuels divisive attitudes [8]. By amplifying these dynamics, platforms amplify their impact on opinion formation, reducing users’ capacity for critical reflection and increasing vulnerability to external manipulation. This enables malicious actors to manipulate algorithmic systems to advance partisan narratives, undermine collective cognition, and influence group decision-making [9]. Moreover, this raises serious concerns about privacy, as platforms collect personal data like browsing history, emotional responses, and social interactions for algorithmic targeting. In the absence of adequate oversight, such data may be used for surveillance, political manipulation, and commercial exploitation [10,11].

## 3. The psychological mechanism of information manipulation on social media platforms

### 3.1. Cognitive bias and information manipulation

In social media environments, information is not passively received but actively interpreted through users' cognitive biases. Platforms exploit this by subtly shaping perception via mechanisms such as the anchoring effect. Emotional headlines, provocative visuals, and suggestive cues are used to prompt immediate judgments. For example, trending topics often feature emotionally charged terms like outrage or injustice, reinforced by highly upvoted comments to entrench intuitive responses [12]. On this basis, confirmation bias is deliberately activated. Users' clicks and dwell time expose their preferences, enabling algorithms to repeatedly feed similar viewpoints that reinforce existing beliefs and exclude opposing perspectives. This gradually constructs an echo chamber, effectively creating a closed informational cocoon [13]. In the process, subjective judgment is amplified and eventually mistaken for objective truth. At the same time, to ease cognitive load, users often rely on heuristic strategies for quick decision-making. The representativeness heuristic prompts them to make judgments based on typical features. For example, perceiving a well-dressed individual as a credible speaker [14]. For example, repeated exposure to short videos portraying "rising crime" in a particular area, even without statistical evidence, can cause users to overestimate the perceived risk [15]. Increased exposure to symbolic and emotionally charged content allows platforms to reinforce existing impressions and amplify underlying biases. In addition to cognitive mechanisms, platforms exploit social conformity to influence user judgments. By highlighting metrics like like-counts and pinned comments, platforms manufacture perceived consensus and prompt conformity, leveraging cognitive and social biases through selective content presentation [16].

### 3.2. Emotion-driven responses and information control

In contrast to the subconscious pathways activated by cognitive biases, social media platforms more frequently leverage emotional triggers as a key strategy for manipulating information. By triggering instinctive reactions, emotional stimuli bypass rational thought and accelerate content dissemination. In particular, emotional arousal affects attention, judgment, and behavioral tendencies, especially in response to negative emotions such as anger, anxiety, and fear. These emotional responses reduce critical thinking and prompt users to engage with and spread content more readily [17]. To provoke such intense emotional responses, platforms often use sensational headlines, polarizing language, and emotionally charged narratives that evoke unease, anger, or anxiety. This type of content often attracts large volumes of clicks, comments, and shares in a short period, quickly turning into viral high-engagement posts. Algorithms then prioritize these emotionally charged messages, further increasing their visibility. This cycle contributes to what is known as the emotion amplifier effect, in which stronger emotions lead to wider dissemination and a higher likelihood of similar content being recommended. This process is reinforced by physiological and psychological mechanisms [18]. On the physiological level, high-arousal emotions like anger and fear activate the sympathetic nervous system, placing individuals in a state of heightened alertness while suppressing executive functions needed for rational thinking. As such, users become more prone to making rapid, intuitive judgments [19]. Psychologically, emotional arousal boosts memory formation, thus reinforcing the impact of related content. Moreover, negative emotions are more contagious within social networks, triggering group emotions and responses [20]. At the technical level, platforms amplify emotional dynamics through algorithmic design. Mechanisms like trending topics, popularity rankings, and personalized recommendations consistently prioritize emotionally charged content, keeping users immersed in a highly emotional information environment. This feedback loop polarizes information, deepens divisions, weakens rationality, and fuels rumors and fragmentation.

### 3.3. Social influence and information manipulation

Information manipulation on social media is not confined to individual-level cognition or emotional stimuli; it is systematically reinforced by social interactions and algorithmic systems, ultimately manifesting as a deeply embedded sociocultural phenomenon. This process is driven by emotional resonance, which aggregates scattered individual emotions into collective sentiment, solidifying into polarized public opinion and entrenched social attitudes that significantly influence collective judgment and behavior. In the initial stages, platforms rapidly gather emotional feedback through likes, comments, and shares. These digital cues, such as emojis and popular reactions, act as subtle social signals that foster emotional convergence among users. It has been shown that high-arousal negative emotions like anger and anxiety exhibit strong contagion on social platforms, enabling large-scale emotional transmission across networks [21]. As a result, fragmented emotions coalesce into a collective emotional atmosphere. By capitalizing on users' emotional alignment, platforms systematically shape affective responses into structured emotional communities through content tagging, sentiment-based recommendations, and algorithmic homophily. Users are algorithmically sorted into into opposing roles, such as supporters or opponents, victims or perpetrators, thereby intensifying group identity and social polarization. This categorization mechanism marginalizes moderate voices and diverse viewpoints, driving public discourse toward heightened polarization. Research shows that algorithms consistently guide users into emotionally uniform environments, reinforcing echo chambers and ultimately creating tightly knit opinion communities [22]. As this emotion-driven, algorithmically reinforced group influence deepens, the logic of public opinion formation undergoes a fundamental shift. While social platforms appear to offer a space for free expression, their underlying manipulation relies on emotional resonance to shape attention patterns and steer the flow of discourse. Users under strong emotional influence often abandon independent judgment, conform to group sentiment, and fuel polarization while suppressing rational debate [23]. In this way, platforms subtly govern public emotion and opinion through non-coercive soft control.

## 4. Legal regulation of information manipulation behaviors on social media platforms

### 4.1. Platform responsibility and legal intervention

Though the subtle manipulation techniques adopted by social media platforms often remain within legal boundaries, they still shape cognition, emotion, and public discourse. However, prevailing regulatory frameworks focus primarily on overt harms such as misinformation, defamation, or illicit content, and fall short in addressing the more covert psychological effects driven by algorithmic design. This regulatory gap reduces accountability and allows platforms to present themselves as neutral intermediaries. Nevertheless, as such practices gain traction, increasing recognition of their structural risks has led to a move from self-regulatory norms toward mandatory oversight. For example, the European Union's Digital Services Act (DSA) exemplifies this shift by integrating major platforms into a formal risk governance regime, requiring algorithmic transparency, societal impact assessments, and mitigation strategies [24]. By codifying platforms' influence over emotion and public discourse, the law grants regulatory bodies enhanced authority to intervene. In addition, Germany's Network Enforcement Act (NetzDG) establishes a regulatory framework for platform communication, whereas France's Anti-Fake News Law targets the democratic threats posed by information manipulation [25]. As platforms exert growing influence over the architecture of digital information, traditional accountability models prove insufficient. In response, regulatory approaches must scale with platform influence, ensuring transparency, independent auditing, and accountability across content curation, interface architecture, and user engagement systems. Besides, laws should require platforms to conduct “emotional impact assessments” before deploying new features or algorithms, evaluating their potential effects on public emotional stability and social polarization, and enforcing minimum risk mitigation measures. To safeguard cognitive integrity and emotional autonomy, platforms must bear legal responsibility for psychological influence, moving beyond reliance on voluntary ethics. This shift creates a robust institutional basis for the health and diversity of the information ecosystem.

### 4.2. Technical regulation and legal enforcement

Information manipulation on social media is embedded in complex algorithms and extensive data infrastructures, making it difficult for traditional manual reviews or user reports to detect subtle and systemic risks. To bridge the gap between legal mandates and advisory norms, technical regulation plays a crucial role in ensuring actionable governance. A multi-tiered framework that integrates legal enforcement with algorithmic oversight is needed. Platforms, as primary responsible entities, should implement internal self-assessment mechanisms to detect risks like cognitive bias, emotional amplification, and content filtering. Furthermore, external auditors or designated regulators must conduct regular and targeted reviews of recommendation systems, data practices, and emotional design patterns to ensure alignment with legal standards and public interest [26]. Besides, platforms should be legally required to publish transparency reports regularly, disclosing algorithmic changes, risk assessment outcomes, and user emotional feedback data to provide factual bases for supervision. State regulatory agencies must be granted lawful access and audit rights to strengthen external oversight capabilities, including reviewing core algorithmic documentation, interface call logs, and internal risk control records for independent verification and real-time evaluation. Additionally, efforts should be made to establish a “social emotional risk early warning system” utilizing natural language processing and emotion recognition technologies to detect abnormal emotional clustering, frequent intense discourse, or signs of information manipulation early on, complemented by human judgment for dynamic intervention [27]. Throughout this process, a careful balance between data privacy and audit access is essential to ensure the protection of users’ rights. Meanwhile, clear legal liabilities and penalties should be established for platforms that fail to disclose information, obstruct oversight, or pose manipulation risks. Such measures may include fines, functional restrictions, service removal, or suspension orders, increasing violation costs and enhancing deterrence. Despite ongoing efforts, challenges remain in boosting emotion recognition, ensuring transparent algorithm audits, and balancing data access with privacy. For governance to be effective, legal enforcement must be strengthened, technical tools refined, and social consensus deepened.

## 5. Psychological defense and legal governance of social media information manipulation

### 5.1. Psychological mechanisms for individual resistance to information manipulation

The persistent threat of information manipulation arises from technological means and the strategic exploitation of psychological tendencies such as cognitive biases, emotional responses, and social conformity. Although legal regulation establishes structural protections, it remains insufficient in addressing manipulation risks rooted in individual psychological vulnerabilities. Thus, enhancing public psychological resilience is vital to reinforcing external governance and curbing manipulation at its source. To counter manipulation effectively, individuals must develop the ability to recognize and interpret cognitive cues. Rather than presenting information transparently, many manipulative strategies deliberately bypass logic by triggering emotional responses, encouraging black-and-white thinking, and spotlighting partial truths. In response, developing metacognitive abilities, such as detecting framing devices and scrutinizing intuitive reactions, becomes crucial. These skills can be nurtured through targeted educational initiatives, including media literacy programs and case-based training in manipulation recognition, implemented across schools, communities, and digital learning platforms. In addition, strengthening emotional self-regulation is crucial for mitigating vulnerability to manipulative cues. By deliberately provoking high-arousal emotions such as fear, anger, or moral outrage, manipulation efforts often prompt impulsive responses and widespread sharing. Enhancing emotional literacy, including the ability to separate subjective emotional reactions from the factual reliability of content, can serve as a protective buffer. By offering opportunities for reflection and shared learning, tools such as digital self-assessments and community workshops help reinforce emotionally conscious behavior in online contexts. In addition, reflective media habits are equally crucial. Many users engage passively with digital platforms, unaware of algorithmic influence. While institutional reform is essential, individuals must take responsibility for their media habits by critically assessing content, avoiding impulsive sharing, and recognizing how algorithms shape information flows, which is crucial to building a more informed and resilient digital environment. In short, psychological intervention counters manipulation by fostering critical thinking, emotional control, and mindful media use, reinforcing resilience beyond legal measures.

### 5.2. Legal frameworks for structural governance of information manipulation

Information manipulation reflects not only individual vulnerability but also systemic flaws in digital infrastructures. Thus, psychological resilience must be reinforced by proactive legal frameworks that define platform responsibilities and protect the integrity of the public information sphere. To enable effective oversight and close regulatory gaps, legal reforms must begin by clearly defining manipulation-related practices such as subtle framing, algorithmic influence, and covert persuasion. In tandem, codifying platform responsibilities in areas like algorithmic transparency, manipulation risk assessment, and procedural responses to harmful content can shift reliance from informal industry norms to binding legal standards. These obligations may further entail mandatory reporting on high-risk content flows, clear labeling of emotionally charged material, and public disclosure of content moderation policies [28]. Besides, the enhancement of judicial and regulatory mechanisms is critical, given the limitations of traditional fact-based methods in identifying subtle and pervasive manipulation. Therefore, the integration of legal, psychological, and computational disciplines into regulatory bodies facilitates the detection of manipulation and evaluation of its cognitive impact. These expert bodies support the formulation of evidentiary standards, risk assessment models, and procedures for auditing platform behavior. Moreover, regulators should work closely with academic institutions to create databases of manipulation patterns, implement forensic tracking tools, and design audit protocols that enhance the traceability and accountability of platform operations [29]. To effectively govern manipulation in the global digital landscape, the transnational nature of platforms calls for national legislation to be complemented by coordinated international regulation. Harmonized legal frameworks, unified standards, and joint enforcement mechanisms are essential to addressing cross-border governance challenges. Building on existing initiatives such as the EU's DSA, which mandates systemic risk audits for large platforms, other regions including the U.S. and

Australia have also begun exploring statutory frameworks for platform transparency and influence moderation [30]. Besides, advancing cross-border cooperation and joint enforcement is essential for a resilient global information environment.

## 6. Conclusion

This study reveals that information manipulation on social media substantially influences individual cognition, emotional regulation, and the formation of public opinion by leveraging cognitive biases, emotional stimuli, and social dynamics. Effective governance is still constrained by structural flaws, including the lack of clear platform accountability, limited technical oversight, and obstacles to legal enforcement. In response, it advocates a dual strategy integrating psychological interventions with legal governance to boost public psychological resilience and build robust systemic safeguards. However, Literature review and legal text analysis serve as the primary methodological basis, yet they lack empirical validation through platform data or user behavior, hence creating a gap between theoretical reasoning and practical evidence. Accordingly, addressing information manipulation requires coordinated, long-term governance through interdisciplinary cooperation and international regulation to ensure a transparent and trustworthy digital environment.

## References

[1] Thorson, E., & Wells, C. (2016). How gatekeeping still matters: Understanding the role of editors in the digital news era. Journalism Studies, 17(5), 509-527.
[2] Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. Digital Journalism, 6(2), 154-175.
[3] Zhang, G. (2020). Cognitive mechanisms of manipulative communication. Journalism & Communication, 25-34.
[4] Sunstein, C.R. (2017). #Republic: Divided democracy in the age of social media. Princeton University Press.
[5] Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. New York: Penguin Press.
[6] Flaxman, S., Goel, S., & Rao, J.M. (2016). Filter bubbles, echo chambers, and online news consumption. Public Opinion Quarterly, 80(S1), 298-320.
[7] Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188, 39-50.
[8] Tucker, J.A., Guess, A., Barberá, P., et al. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. Political Science Quarterly, 133(4), 655-688.
[9] Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. Colorado Technology Law Journal, 13(203), 203-218.
[10] Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. New York: PublicAffairs.
[11] Andrejevic, M. (2014). Surveillance and alienation in the online economy. Surveillance & Society, 12(3), 381-397.
[12] Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.
[13] Nickerson, R.S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2(2), 175-220.
[14] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.
[15] Schwarz, N., et al. (1991). Ease of retrieval as information: Another look at the availability heuristic. Journal of Personality and Social Psychology, 61(2), 195-202.
[16] Bond, R., & Smith, P.B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. Psychological Bulletin, 119(1), 111-137.
[17] Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. Psychological Science, 14(2), 144-150.
[18] Brady, W.J., Wills, J.A., Jost, J.T., et al. (2017). Emotion shapes the diffusion of moralized content in social networks. Proceedings of the National Academy of Sciences, 114(28), 7313-7318.
[19] Kensinger, E.A., & Schacter, D.L. (2006). Processing emotional pictures and words: Effects of valence and arousal. Cognitive, Affective, & Behavioral Neuroscience, 6(2), 110-126.

[20] Goldenberg, A., Gross, J.J., & Gal, Y. (2020). Digital emotion contagion. Trends in Cognitive Sciences, 24(4), 316-328.

[21] Kramer, A.D.I., Guillory, J.E., & Hancock, J.T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788-8790.

[22] Cinelli, M., Morales, G.D.F., Galeazzi, A., et al. (2021). The echo chamber effect on social media. Proceedings of the National Academy of Sciences, 118(9), e2023301118.

[23] Papacharissi, Z. (2015). Affective publics: Sentiment, technology, and politics. Oxford University Press.

[24] European Commission. (2022). The Digital Services Act (DSA). https: //ec.europa.eu/commission/presscorner/detail/en/ip_22_2545

[25] Richards, N.M., & Raso, F.A. (2023). The future of data governance: Algorithmic accountability and platform regulation. Harvard Journal of Law & Technology, 36(1), 1-45.

[26] Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.

[27] [Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C.R., et al. (2022). How behavioural sciences can promote truth, autonomy and democratic discourse online. Nature Human Behaviour, 6(2), 156-165.

[28] Helberger, N., Pierson, J., & Poell, T. (2020). Governing online platforms: From contested to cooperative responsibility. The Information Society, 36(1), 1-14.

[29] Susser, D., Roessler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. Georgetown Law Technology Review, 4(1), 1-45.1·

[30] Kaye, D. (2021). Social media regulation: Problems and solutions. European Journal of International Law, 32(2), 583-598.