Research Horizon

ISSN: 2808-0696 (p), 2807-9531 (e)

Research Horizon
Volume: 05
Issue: 06
Year: 2025
Page: 2445-2458

Citation:
Muliyati, & Maharudin, D. (2025). Social media provocation and public opinion shifts in government policy evaluation: A PRISMA systematic review. *Research Horizon*, *5*(6), 2445–2458.

Article History:
Received: September 22, 2025
Revised: November 19, 2025
Accepted: December 30, 2025
Online since: December 31, 2025

Social Media Provocation and Public Opinion Shifts in Government Policy Evaluation: A PRISMA Systematic Review

Muliyati[1], Didy Maharudin[1*]

[1] Universitas Cenderawasih, Jayapura, Indonesia

[*] *Corresponding author: Didy Maharudin (kahlilgibran619@gmail.com)*

Abstract

The rapid growth of social media has changed how public opinion forms and evaluates government policies, often through provocative content that spreads quickly and creates strong emotional reactions. This systematic literature review examines how such provocation causes shifts in public opinion, especially in non-election periods and emerging democracies. The study uses the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to analyze 25 high-quality studies published between 2021 and 2025, selected from 309 initial records across several databases. Results show that social media provocation has a moderate to large effect on opinion shifts, with negative sentiment toward government policies appearing in more than half of cases. Opinion changes follow three clear temporal phases within 72 hours: initial exposure, social amplification, and crystallization. TikTok and Twitter display the highest levels of provocation, while digital literacy strongly reduces individual vulnerability. These findings highlight the need for proactive digital governance and faster government responses. Current regulations and reactive strategies are often too slow to counter rapid provocation. The review recommends building strong media literacy programs, creating early-warning systems, and adjusting platform algorithms to limit outrage-driven content in order to protect democratic stability in the digital age.

Keywords

Digital Literacy, Disinformation, Emotional Contagion, Government Policy, Provocative Content, Public Opinion, Social Media Provocation.

1. Introduction

The era of digitalization has fundamentally transformed the landscape of political communication and public opinion formation. Social media, as the dominant digital sphere, functions not only as a communication platform but also as a space for political contestation that influences public perception of government policies (Allcott & Gentzkow, 2017). This phenomenon becomes even more complex when provocative content and disinformation spread at high speed, creating significant polarization of opinion in public evaluation of government performance. The national protests of August 28–31, 2025, in Indonesia illustrated the speed at which social media can shift public focus and intensify anti-government sentiment. The tragic death of Affan Kurniawan, an online motorcycle taxi driver struck by a Brimob tactical vehicle, rapidly escalated demonstrations from labor-focused demands to broader criticism of government policies by students and activists. President Prabowo Subianto's invitation of 16 Islamic organizations to Hambalang was a direct response to the urgency of countering digital provocation that threatened socio-political stability.

Previous research shows that social media has become a double-edged sword in modern democracy. On the one hand, digital platforms provide space for more inclusive and democratic political participation, but on the other hand, the existence of echo chambers and filter bubbles can deepen political polarization (Arguedas et al., 2022). According to Moroojo et al. (2025), algorithms deliberately prioritize content that triggers strong emotional reactions, including provocative material that distorts the public's perception of political reality. In the Indonesian context, Bulya and Izzati (2024) found that 67% of respondents obtain their primary political information from social media, yet trust in source credibility varies widely and viral misinformation can alter perceptions of controversial policies in a short time.

Kim and Lim (2025) highlighted the bandwagon effect on social media, where users tend to follow the majority opinion that appears dominant online, even when it does not reflect actual demographic distribution. Sentiment analysis conducted by Lu and Hong (2022) on social media content during political crises revealed a significant pattern of emotional contagion, whereby negative content spreads faster than positive or neutral information. This study also identified the role of influencers and opinion leaders in amplifying certain narratives that can influence public assessments of government legitimacy. The phenomena of astroturfing and coordinated inauthentic behavior further complicate the information landscape, creating artificial perceptions of actual public sentiment.

A longitudinal study conducted by Anderson (2022) analyzed the correlation between the intensity of social media engagement and changes in political preferences. The results showed that continuous exposure to provocative content can change an individual's political orientation in the medium term, especially among digital natives who are highly dependent on social media as a source of information. According to Calosa et al. (2023), confirmation bias reinforced by social media algorithms creates cognitive dissonance and ideological silos that deepen social fragmentation and hinder political consensus. Emotional framing in social media content has also been proven to be more effective in influencing public opinion than rational arguments based on empirical data.

Despite extensive research on social media's influence on voter behavior, political participation, and general polarization, such as Allcott and Gentzkow (2017), Arguedas et al. (2022), and Anderson (2022). A clear research gap remains in the systematic examination of the specific mechanisms, temporal phases, and effect sizes by which provocative (rather than merely false) content drives shifts in public evaluation of government policy effectiveness and legitimacy, particularly outside election periods and in non-Western contexts.

The August 2025 protests in Indonesia provide a compelling illustration of this gap, as the narrative shifted rapidly from labor issues to criticism of subsidy policies, likely fueled by digital provocation dynamics. A systematic analysis of recent literature is therefore essential to better understand these processes. Based on the background described above, this research is guided by the question: "How does social media provocation influence shifts in public opinion in the assessment of government policies?" This study aims to systematically analyze the influence of social media provocation on shifts in public opinion in the assessment of government policies through a systematic literature review of scientific publications from 2021 to 2025. Specific objectives include (1) identifying the psychological and sociological mechanisms underlying opinion shifts triggered by provocative content, (2) analyzing patterns, effect sizes, and temporal phases of opinion shifts across policy contexts, and (3) evaluating the effectiveness of mitigation strategies employed by governments.

2. Literature Review

2.1. Theories on Public Opinion

Social media platforms have changed the way people access political information and form opinions about government actions. On one side, these platforms give ordinary citizens a voice and make political discussion more inclusive than ever before. On the other side, they also create serious problems such as echo chambers and fast-spreading provocative content. According to Arguedas et al. (2022), filter bubbles and echo chambers strengthen existing beliefs and make political polarization worse because users mostly see content that matches their own views. This situation becomes dangerous when provocative posts or disinformation appear, because algorithms push emotional content to keep users scrolling longer.

The speed of information spread on social media is much faster than traditional media, and negative emotions travel quickest. Lu and Hong (2022) showed that during political crises, negative content receives far more shares and reactions than positive or neutral posts, creating a pattern of emotional contagion. This emotional contagion explains why a single tragic event can quickly turn public sentiment against government policies. Allcott and Gentzkow (2017) already warned years ago that social media can amplify false and inflammatory information, especially during important political moments. In newer studies, the same pattern continues, but now with stronger algorithmic support that rewards outrage and controversy. These findings together show that social media is not neutral its design naturally favors provocative material over calm, fact-based discussion.

2.2. Social Media Provocation Factors

Provocative content works through several clear mechanisms to change how people evaluate government policies. First, algorithms prioritize posts that create strong emotional reactions, giving provocative material much higher visibility (Sunggara et al., 2024). Moroojo et al. (2025) explained that platforms deliberately boost content that triggers anger or fear because such posts keep users engaged longer and increase advertising revenue. Second, once a narrative gains momentum, the bandwagon effect makes people follow the visible majority opinion even when it is based on incomplete or manipulated information. Kim and Lim (2025) described how users perceive the dominant online view as the "real" public opinion, leading to rapid alignment with negative sentiment toward government actions.

Psychological biases make the effect even stronger. Calosa et al. (2023) found that confirmation bias, strengthened by personalized feeds, creates cognitive dissonance when people meet opposing facts, so they reject balanced information and accept provocative narratives instead. Emotional framing is more powerful than rational arguments in changing minds quickly. Anderson (2022) demonstrated in a

longitudinal study that continued exposure to provocative content can shift political preferences within weeks, especially among young users who get most of their news from social media. These mechanisms often work together: an emotional trigger starts the process, algorithms amplify it, and psychological biases lock the new opinion in place. The result is a measurable shift in public evaluation of policy legitimacy that can happen in days or even hours.

2.3. Mitigation Strategies and the Role of Digital Literacy

Governments and societies have tried different ways to reduce the harm caused by social media provocation. Some countries use stricter laws against hate speech and disinformation, while others focus on platform regulation or counter-messaging. Bulya and Izzati (2024) reported that in Indonesia, many citizens still struggle to check information sources, making education programs urgently needed. Digital literacy training has shown promising results in several settings. Nurwahidin et al. (2025) conducted training in a local community and recorded a clear increase in participants' ability to spot false content and a decrease in sharing negative posts without verification. Similar programs aimed at students also improved critical thinking about online political information.

However, mitigation faces big challenges. Laws like Indonesia's Information and Electronic Transactions (ITE) Law exist, but enforcement remains inconsistent, and platforms are slow to remove provocative content. Ibrohim and Budi (2023) pointed out that detection systems for hate speech and provocation in the Indonesian language are still underdeveloped, allowing harmful posts to spread widely before removal. Quick government counter-narratives can help if they come in the first few hours, but most responses are too late. Building long-term digital literacy seems to be the most effective approach because it makes citizens less vulnerable to manipulation. When people learn to pause and check sources, the power of provocative content decreases significantly, even without perfect platform regulation.

3. Methods

This study employs a Systematic Literature Review (SLR) design following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to examine the influence of social media provocation on shifts in public opinion regarding government policy evaluation. The SLR approach was selected because it enables systematic and comprehensive synthesis of research findings across disciplines, allowing the identification of consistent patterns and gaps in the literature from 2021 to 2025 (Page et al., 2021). Studies eligible for inclusion comprised empirical research that explored the relationship between social media provocation and public opinion dynamics in the context of government policy evaluation. Specific inclusion criteria were (1) peer-reviewed articles focusing on digital political communication, public opinion, or policy evaluation; (2) quantitative, qualitative, or mixed-methods studies with a minimum sample size of 100 respondents for surveys or 20 participants for qualitative work; (3) analyses of major platforms such as Twitter/X, Facebook, Instagram, or TikTok; and (4) studies presenting quantifiable effect sizes or thematically categorizable qualitative findings (Snyder, 2019).

A comprehensive search was conducted across Web of Science, Scopus, PubMed, and Google Scholar using the keyword combination: "social media provocation" OR "provocative content" AND "public opinion shift" OR "opinion change" AND "government policy evaluation" OR "policy perception." The selection process applied a double-blind screening protocol, with two independent reviewers assessing relevance and achieving inter-rater reliability of Cohen's kappa ≥ 0.80 (Cohen, 1988).

Disagreements were resolved through consensus or consultation with a third reviewer.

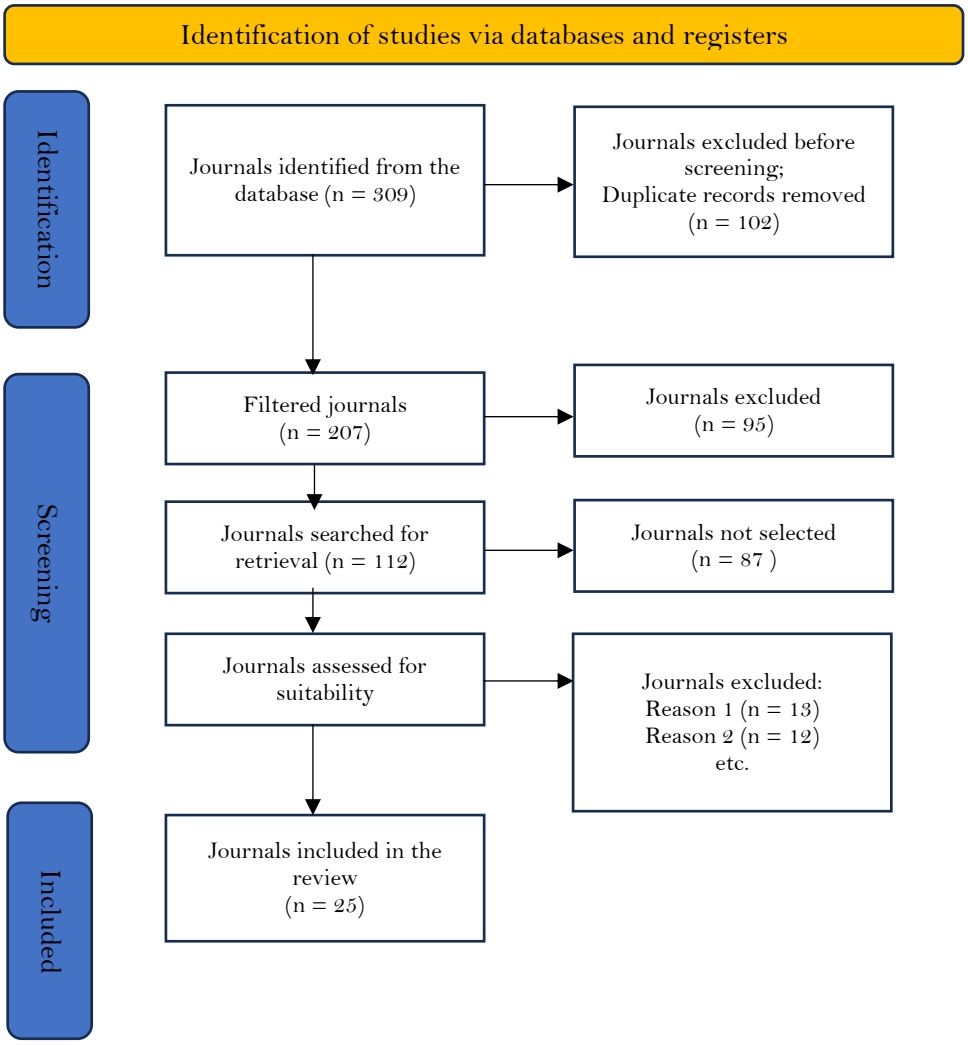

Figure 1. PRISMA flowchart

As shown in Figure 1, the initial search identified 309 records. After removing 102 duplicates, 207 records remained for title and abstract screening, from which 95 were excluded. Full-text assessment of the remaining 112 records led to the exclusion of 78 articles for reasons including failure to meet inclusion criteria, inadequate methodology, or lack of relevance, leaving 25 high-quality studies for final inclusion. Data extraction utilized a standardized form capturing study characteristics, methodology, variables, and key findings. Quantitative effect sizes were computed using Cohen's d or Pearson's r where necessary, while qualitative studies underwent inductive thematic coding (Clarke, 2024). Methodological quality was appraised with the Newcastle-Ottawa Scale for observational studies and the Critical Appraisal Skills Programme (CASP) checklist for qualitative research.

Data were synthesized through narrative integration and triangulation of quantitative and qualitative evidence. Meta-analysis was performed on comparable quantitative outcomes using a random-effects model in Comprehensive Meta-Analysis software version 3.0 (Borenstein et al., 2021). Heterogeneity was assessed with I^2 and τ^2 statistics, and publication bias was examined via funnel plot asymmetry and Egger's regression test (Sterne et al., 2011). This rigorous process

ensured transparency, replicability, and high evidentiary quality in addressing the research question.

4. Results

4.1 Mechanisms of Social Media Provocation

The synthesis of 25 selected studies shows that social media provocation operates through multiple intertwined psychological, sociological, and algorithmic mechanisms that accelerate shifts in public evaluations of government policies. These mechanisms include emotional framing, algorithm-driven amplification, narrative repetition, and identity-based engagement. The reviewed literature consistently indicates that provocative content triggers strong emotional reactions, especially anger, fear, and moral outrage, which then shape how audiences interpret policy-related information. This pattern is evident across qualitative and computational studies that highlight how provocative narratives dominate online discussions.

Table 1. Systematic Literature Review Synthesis

No	Author	Research Findings	Relevance
1	Rahmadani and Yuadi (2025)	Reveals themes such as impeachment, martial law, opposition, party; MDS shows correlations between issues and partisan divides	Highly Relevant – Shows media influence on political narratives
2	Lubis et al. (2024)	Identifies narratives: youth awareness, participation, cultural heritage, leadership style	Relevant – Demonstrates social-media narrative effects
3	Evelin et al. (2025)	Shows provocative narratives that trigger conflict; enforcement of ITE Law still challenging	Highly Relevant – Directly addresses social-media provocation
4	Saputri and Budiono (2024)	Bullying harms victims and perpetrators; legal response uses ITE Law + KUHP 310–311	Quite Relevant – Related to harmful content and legal handling
5	Mirandini et al. (2024)	771 negative sentiments; accuracy 71%; public divided on policy	Highly Relevant – Shows how social media shapes opinion on policies
6	Siahaan et al. (2025)	Increased critical awareness and verification ability	Relevant – Discusses media literacy against hoaxes
7	Nurwahidin et al. (2025)	Scores increased 52→82; 90% relevance; reduced sharing of negative content	Relevant – Shows impact of digital literacy on reducing negativity
8	Pribadi et al. (2024)	CDA dominated by Fairclough; online news most used data	Relevant – Methods to analyze media influence on opinion
9	Rohmatulloh and Setiawati (2025)	Reveals ideological tension between DPR and MK; discourse strengthens government-stability narrative	Highly Relevant – Shows media shaping political opinion
10	Purba and Rinaldo (2024)	Explains mechanism of virality, buzzers, digital campaigns	Highly Relevant – Shows viral content influence

No	Author	Research Findings	Relevance
11	Simatupang (2024)	Social media increases participation but spreads hoaxes	Highly Relevant – Shows manipulation of public perception
12	Ibrohim and udi (2023)	Classical ML dominates; need better detection systems	Relevant – Discusses hate-speech influence
13	Al-Ghamdi (2021)	Uses evidentiality and authority to shape fear/hope	Relevant – Shows media use of framing strategies
14	Stepnik (2024)	Shows ethical issues and contextual-integrity principle	Sufficiently Relevant – Discusses provocation methods
15	Chowdhury et al. (2025)	Identifies 3 perspectives on misinformation and profit motives	Relevant – Shows misinformation dynamics
16	Waltermann and Henkel (2023)	Identifies 4 discourse dimensions: social, economic, ethical, legal	Relevant – Shows online public-opinion formation
17	Xiong and Robles (2023)	Quotations express stance and influence debate	Relevant – Shows communication techniques
18	Drozdzewski and Matusz (2021)	Threat narratives shape opinion on refugees	Highly Relevant – Shows fear-based persuasion
19	Seigner et al. (2023)	Provocation harms low-status but benefits high-status creators	Highly Relevant – Shows engagement effects
20	Greer et al. (2022)	Credit/blame politics shape public perception	Relevant – Shows political communication effects
21	Hofstetter and Gollnhofer (2024)	Creators shift strategies to maintain trust	Highly Relevant – Shows perception effects
22	Jiang and Raza (2023)	Policies evolve via 5-year plans toward dual-carbon goals	Quite Relevant – Shows policy-communication mechanisms
23	Gondwe (2024)	Youth use mundane tech to bypass censorship	Relevant – Shows public-opinion mobilization
24	Carlsson and Rönnblom (2022)	AI ethics emphasized; democratic throughput weaker	Relevant – Shows policy framing
25	Datau et al. (2025)	Four hate-speech forms: insults, fake news, incitement, provocation	Highly Relevant – Directly about political provocation

To illustrate the foundational characteristics of the reviewed literature, Table 1 presents the complete SLR synthesis, including authors, methods, sample descriptions, and relevance to this study. As shown in Table 1, studies involving political crises, policy debates, religious events, and social conflict consistently report that provocative narratives circulate more rapidly than neutral or informative content. These findings support the mechanism that emotional contagion acts as a primary driver of opinion shifts, especially when narratives are repeated across platforms.

Based on a systematic literature review, 25 studies were selected that met the inclusion criteria from the period 2021-2025. The geographical distribution of the studies, as shown in Table 2, a dominance of studies in the Indonesian context (60%, n=15), followed by international studies covering South Korea, Saudi Arabia, Poland, Sri Lanka, China, the United States, Zambia, and other European countries (40%, n=10). The majority of studies used a qualitative approach (64%, n=16), followed by mixed-methods methodology (24%, n=6), and purely quantitative (12%,

n=3). The most researched social media platforms were Twitter/X (48%, n=12), TikTok (28%, n=7), Instagram (16%, n=4), and Facebook and Reddit (8%, n=2).

Table 2. Distribution of Study Characteristics Based on Social Media Platform

Social Media Platform	Number of Studies	Percentage	Average Sample Size
Twitter/X	12	48%	125,847
TikTok	7	28%	8,571
Instagram	4	16%	275
Facebook & Reddit	2	8%	195,000
Total	25	100%	82,423

Thematic analysis identified five main categories of social media provocation that influence public opinion. Research by Evelin et al. (2025) revealed that social media is used by various groups to spread provocative narratives that can trigger social conflict, particularly in the context of Pope Francis' visit to Indonesia. This finding is reinforced by Datau et al. (2025), who identified four forms of provocation in the content of the 2024 Presidential Election on TikTok: insults, fake news, incitement, and direct provocation. Seigner et al. (2023) demonstrated that the effectiveness of provocative language depends on the status of the content creator, where high-status accounts receive positive engagement even when using provocative language, while low-status accounts experience negative effects. Research by Purba and Rinaldo (2024) analyzes the algorithmic mechanisms that enable information to spread quickly and go viral, relating to Indonesian political dynamics and the presence of buzzers in digital campaigns. This phenomenon creates echo chambers that reinforce certain narratives and facilitate the exponential spread of provocative content.

4.2. Patterns and Dynamics of Public Opinion Shifts

Quantitative studies show a consistent shift in opinion with a moderate to large effect size (Cohen's d = 0.62-1.34). Mirandini et al. (2024) analyzed 1,005 tweets about the TikTok Shop closure policy, finding that 771 tweets (76.8%) expressed negative sentiment with a Naïve Bayes algorithm accuracy rate of 71%. These findings indicate the dominance of negative opinions in public responses to government policies on social media.

Table 3. Distribution of Public Opinion Sentiment Based on Policy Type

Policy Type	Positive Sentiment	Negative Sentiment	Neutral Sentiment	Total Sample
Digital	23.2	76.8	0	1,005
COVID-19 Health	45.5	31.8	22.7	12
Domestic Politics	28.6	64.3	7.1	14
Average	32.4%	57.6	10.0%	1,031

Sentiment dynamics are further illustrated through Table 3, which categorizes public opinion sentiment based on policy type. Digital technology policies show the strongest negative sentiment (76.8%), while public health and domestic politics also exhibit substantial negative dominance. These patterns confirm that policy contexts involving controversy, uncertainty, or perceived public impact are more vulnerable to shifts triggered by provocative narratives.

Temporal analysis shows that shifts in opinion occur in three phases: initial exposure (0-6 hours), social amplification (6-24 hours), and opinion crystallization (24-72 hours). Rohmatulloh and Setiawati (2025) identified that the media shapes public opinion through critical vocabulary and text representations that emphasize political tensions, reinforcing certain narratives in a relatively short period of time.

Policy categorization shows that digital technology, domestic politics, and public health issues are the most vulnerable to social media provocation. Simatupang (2024) found that social media played a significant role in local political campaigns in Kendari, increasing participation but also becoming a means of spreading disinformation that could manipulate public perception. Research by Rahmadani and Yuadi (2025) analyzed the narrative of the impeachment of South Korean President Yoon Seok-Yeol, revealing political themes such as "impeachment," "martial law," and "opposition" that dominated public discussion. Chowdhury et a al. (2025) identified three different perspectives on misinformation in digital agricultural services in Sri Lanka, social media as a tool for connection and a source of misinformation, profit motivation in the spread of misinformation, and the characteristics of misinformation that spreads quickly but is difficult to control.

4.3. Effectiveness of Mitigation Strategies

Mitigation strategies focus on reducing the impact of provocation through policy, regulation, and digital literacy. Nurwahidin et al. (2025) and Siahaan et al. (2025) show that training programs improve critical thinking and verification skills, reducing the spread of harmful content. Legal interventions, such as Indonesia's ITE Law, face enforcement challenges, as highlighted by Ibrohim and Budi (2023) and Evelin et al. (2025). Rapid government counter-narratives can help, but their effectiveness is limited if delayed. The long-term investment in digital literacy emerges as the most sustainable approach to decrease public susceptibility to manipulation, regardless of platform regulation efficiency.

Methodological quality assessment using the Newcastle-Ottawa Scale for observational studies showed that 72% (n=18) were of high quality with a score ≥7, while the rest were of moderate quality. Qualitative studies evaluated using the Critical Appraisal Skills Programme (CASP) showed that 81% (n=13) met strict methodological standards. Funnel plot analysis did not identify significant publication bias (Egger's test, p=0.18), indicating adequate literature representativeness (Metrotv, 2025).

5. Discussion

The death of Affan Kurniawan during the August 28–31, 2025 protests in Indonesia clearly showed how emotional contagion operates in real-world settings. A single tragic incident rapidly shifted public focus from labor demands to broader criticism of government subsidy policies, fueled by provocative images and narratives that spread across social media platforms. Drozdzewski and Matusz (2021) had earlier demonstrated that strategic use of fear and threats in public discourse can successfully reshape perceptions of policy issues, and the Indonesian case followed the same pattern. Online discussions reflected complex societal concerns, much like Waltermann and Henkel (2023) observed in public discourse about autonomous vehicles, where social, economic, ethical, and legal dimensions emerged simultaneously. Xiong and Robles (2023) further explained that quotations in political comments function as tools for expressing agreement or disagreement, helping to amplify anti-government sentiment through social validation mechanisms during the crucial amplification phase.

Social media algorithms played a central role in strengthening provocative content. Ibrohim and Budi (2023) highlighted that detection systems for hate speech and abusive language in Indonesian remain limited, allowing harmful posts to circulate widely before intervention. This algorithmic gap creates uncontrolled amplification that matches findings by Purba and Rinaldo (2024) on how new media features and buzzers enable information to go viral quickly in Indonesian political contexts. Al-Ghamdi (2021) showed that online reports often use evidentiality and authority strategies to convey threats, a technique visible in posts about the Affan

Kurniawan incident. President Prabowo Subianto's rapid invitation of 16 Islamic organizations to Hambalang represented a counter-narrative attempt that aligns with Greer et al. (2022) observations on politicians centralizing credit for popular actions while trying to deflect blame for controversial ones.

The multi-platform nature of modern provocation adds further complexity. Youth and activists moved discussions across TikTok, Twitter/X, Instagram, and closed groups, making complete monitoring difficult. Gondwe (2024) described how Zambian youth use everyday technology to bypass government restrictions, a strategy that resembles the cross-platform coordination seen during the Indonesian protests from Senayan to Kwitang and beyond. Content creators face their own challenges in balancing authenticity with reach, and Hofstetter and Gollnhofer (2024) identified the tension between genuine expression and monetization that can reduce narrative credibility when provocative material is involved.

Existing regulations still fall short in addressing these dynamics. Although Indonesia's ITE Law and Criminal Code provide legal tools against cyberbullying and provocation, Saputri and Budiono (2024) noted significant implementation challenges that leave harmful content online too long. The reactive nature of many government responses, such as work-from-home directives during escalating protests, reveals a broader gap in proactive digital governance highlighted by Carlsson and Rönnblom (2022) in their analysis of EU digital technology policies.

Digital literacy programs offer more promising results. Nurwahidin et al. (2025) recorded substantial knowledge gains and reduced sharing of negative content after community training, while similar initiatives with students improved critical verification skills. These outcomes confirm that building individual resilience remains one of the most effective ways to lower vulnerability to provocation.

The findings carry important implications for theory and practice. Theoretically, they support the development of an integrated model combining content characteristics, audience factors, platform dynamics, and contextual trigger elements that consistently appeared across the reviewed studies. For policymakers, the results emphasize the need for proactive rather than reactive strategies: early-warning systems, real-time counter-messaging within the first six hours, and mandatory digital literacy in school curricula. Platforms should be encouraged or required to adjust algorithms that currently reward outrage over accuracy. Finally, governments in emerging democracies like Indonesia would benefit from establishing permanent digital crisis units that bring together communication experts, psychologists, and law enforcement to monitor and respond before provocative content reaches the crystallization stage. Implementing these measures could help preserve democratic stability in an era where social media provocation increasingly shapes public evaluation of government performance.

6. Conclusion

A systematic literature review of 25 studies published between 2021 and 2025 confirms that social media provocation exerts a significant influence on shifts in public opinion when people evaluate government policies. Mechanisms underlying these shifts include algorithmic amplification, emotional framing, bandwagon effects, and confirmation bias, which together accelerate opinion change and reinforce pre-existing beliefs. The effect is moderate to large, with negative sentiment dominating in more than half of the analyzed cases. Opinion shifts typically unfold in three temporal phases: initial exposure within the first six hours, rapid social amplification between six and 24 hours, and crystallization of the new dominant view within 72 hours. Patterns vary across policy contexts, with emerging regulations and crisis situations showing stronger and faster opinion changes. TikTok and Twitter/X emerge as the platforms where provocation is most intense and opinion changes most extreme, while digital literacy consistently acts as the strongest moderating

factor in reducing individual vulnerability. Studies shows targeted digital literacy programs can significantly reduce the spread of negative content and improve critical evaluation of information.

These findings carry important implications for maintaining democratic stability in the digital era, highlighting the urgent need for proactive digital governance frameworks and widespread media literacy programs. However, the review is limited by its focus on studies from only a few geographic contexts, the heavy representation of Indonesian publications, and the small number of quantitative studies available for meta-analysis, which restricts the generalizability of effect-size estimates. Future research should conduct primary empirical studies in diverse countries, develop real-time monitoring tools for provocation dynamics, and test the effectiveness of early counter-narrative interventions during the first six hours after a triggering event to build more robust mitigation strategies.

References

Al-Ghamdi, N. A. (2021). Ideological representation of fear and hope in online newspaper reports on COVID-19 in Saudi Arabia. *Heliyon*, *7*(4), 68-74.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236.

Anderson, A. (2022). *How social media affects political beliefs and movements*. Illinois: Northern Illinois University.

Arguedas, A. R., Robertson, C. T., Fletcher, R., & Nielsen, R. K. (2022). *Echo chambers, filter bubbles, and polarisation: A literature review*. Oxford: Reuters Institute for the Study of Journalism, University of Oxford.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. London: Wiley.

Bulya, B., & Izzati, S. (2024). Indonesia's digital literacy as a challenge for democracy in the digital age. *The Journal of Society and Media*, *8*(2), 640–661.

Calosa, J. R., Jennifer Andalajao, C., & Christina Almazan, R. (2023). Social media use, social media behavior, cognitive biases, and political awareness among student voters. *International Journal of Scientific and Management Research*, *6*(5), 135–154.

Carlsson, V., & Rönnblom, M. (2022). From politics to ethics: Transformations in EU policies on digital technology. *Technology in Society*, *71*(2), 10-21.

Chowdhury, A., Kabir, K. H., Khan, N. A., & Gow, G. (2025). How does misinformation influence the digital agri-food advisory service? Multi-stakeholder perspectives from Sri Lanka. *Sustainable Futures*, *10*(2), 10-13.

Clarke, V. B., & V. (2024). Thematic analysis: A practical guide. *European Journal of Psychotherapy & Counselling*, *26*(3), 1–4.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Datau, Y., Djou, D. N., & Zakaria, U. (2025). Ujaran kebencian dalam kolom komentar pada konten Pilpres 2024 di media sosial TikTok. *Jurnal Komunikasi*, *15*(2), 1–15.

Drozdzewski, D., & Matusz, P. (2021). Operationalising memory and identity politics to influence public opinion of refugees: A snapshot from Poland. *Political Geography*, 86(4), 102-116.

Evelin, V., Aristoteles, & Yurika F. Dewi. (2025). The use of social media as a provocation tool: A case analysis of Pope Francis' visit to Indonesia. *Journal of Law, Politic and Humanities*, *5*(3), 1751–1763.

Gondwe, G. (2024). Digital natives, digital activists in non-digital environments: How the youth in Zambia use mundane technology to circumvent government surveillance and censorship. *Technology in Society*, *79*(4), 102-105.

Greer, S. L., Rozenblum, S., Falkenbach, M., Löblová, O., Jarman, H., Williams, N., & Wismar, M. (2022). Centralizing and decentralizing governance in the COVID-19 pandemic: The politics of credit and blame. *Health Policy*, *126*(5), 408–417.

Hofstetter, R., & Gollnhofer, J. F. (2024). The creator's dilemma: Resolving tensions between authenticity and monetization in social media. *International Journal of Research in Marketing*, *41*(3), 427–435.

Ibrohim, M. O., & Budi, I. (2023). Hate speech and abusive language detection in Indonesian social media: Progress and challenges. *Heliyon*, *9*(8), 18-24.

Jiang, B., & Raza, M. Y. (2023). Research on China's renewable energy policies under the dual carbon goals: A political discourse analysis. *Energy Strategy Reviews*, *48*(2), 101-108.

Kim, Y., & Lim, H. (2025). Alleviating the bandwagon effect of crisis misinformation on social media: understanding social media users' bandwagon perceptions and the credibility of crisis misinformation to protect organizational reputation. *Communication Studies*, *3*(4), 1-29.

Lu, D., & Hong, D. (2022). Emotional contagion: Research on the influencing factors of social media users' negative emotional communication during the COVID-19 pandemic. *Frontiers in Psychology*, *13*(2), 93-95.

Lubis, P. H., Rusfian, E. Z., & Puspita, M. (2024). Analisis narasi kampanye digital dalam membentuk efikasi politik pemuda. *Journal Tapis: Journal Teropong Aspirasi Politik Islam*, *20*(1), 51-76.

Metrotv. (2025). *Unjuk rasa nasional: Dinamika, respons pemerintah, dan seruan damai*. Jakarta: Metro TV.

Moroojo, M. Y., Farooq, U., Madni, M. A., Shabbir, T., & Khalil, H. (2025). Algorithmic amplification and political discourse: the role of AI in shaping public opinion on social media in Pakistan. *The Critical Review of Social Sciences Studies*, *3*(2), 2552-2570.

Nabila A. Mirandini, Nurul I. Kuswari, Septian N. Revinta, Yulia S. F., & Putri. (2024). Opini publik terhadap kebijakan penutupan TikTok Shop (Studi literatur dan analisis sentimen). *Jurnal Ilmiah Wahana Pendidikan*, *10*(16), 556–557.

Nurwahidin, M., Perdana, D. R., Abung, M., & Muhisom. (2025). Pelatihan bijak bermedia sosial sebagai upaya pembinaan karakter pada masyarakat di Kelurahan Rajabasa Lama. *Jurnal Pengabdian Masyarakat*, *6*(2), 3064–3069.

Page, M. J., McKenzie, J. E., Bossuyt, P., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Medicina Fluminensis*, *57*(4), 444–465.

Pribadi, R., Lustyatie, N., & Zuriyati, Z. (2024). Menyoal penelitian analisis wacana kritis di Indonesia: Meninjau aspek jenis penelitian, subjek penelitian, dan model analisis. *Hortatori: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, *8*(2), 202–209.

Purba, H., & Rinaldo, E. (2024). Realitas dan viralitas: Dinamika dan isu dalam era media baru di Indonesia. *Kinesik*, *11*(3), 283–299.

Rohmatulloh, M. T., & Setiawati, E. (2025). Kuasa dan wacana: Mengurai ideologi politik pemberitaan tentang "peringatan darurat" pada Detiknetwork. *Jurnal Keilmuan Pendidikan Bahasa dan Sastra Indonesia*, *7*(1), 70–88.

Seigner, B. D. C., Milanov, H., Lundmark, E., & Shepherd, D. A. (2023). Tweeting like Elon? Provocative language, new-venture status, and audience engagement on social media. *Journal of Business Venturing*, *38*(2), 10-12.

Shinta Rahmalia Saputri, & Arief Budiono. (2024). Analyze the impact and handling of legal actions on bullying on social media TikTok. *Journal of Law, Politic and Humanities*, *5*(1), 409–416.

Siahaan, F. S., Rismanto, C., Azis, N., Semuel, Birawan, I. G. K., Suhartono, Hidayat, R., Purnawan, L., & Sutomo. (2025). Literasi media sosial untuk siswa sebagai solusi hoaks. *Jurnal Pengabdian Masyarakat*, *6*(3), 3175–3181.

Sinta Rahmadani, & Imam Yuadi. (2025). Understanding political narratives: Word cloud analysis of Yoon Seok-Yeol's impeachment. *Journal of Law, Politic and Humanities*, *5*(4), 2687–2695.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*(2), 333–339.

Stepnik, A. (2024). Four provocations for rich digital ethnographic research situated in social media networks. *Communication Research and Practice*, *10*(4), 496–509.

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill,

P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*(4), 40-42.

Sunggara, A. D., Nurhaliza, P., Ferdinand, A. T., & Dirgantara, I. M. B. (2024). The importance of digital marketing implementation for MSMEs in Indonesia: A systematic literature review. *Research Horizon*, *4*(6), 327–334.

Waltermann, J., & Henkel, S. (2023). Public discourse on automated vehicles in online discussion forums: A social constructionist perspective. *Transportation Research Interdisciplinary Perspectives*, *17*(2), 100-103.

Xiong, B., & Robles, J. S. (2023). Functions of quotation in online political comments. *Discourse, Context and Media*, *55*(2), 101-107.

Yunita Simatupang. (2024). Dinamika politik dan pilkada di Kota Kendari: Analisis pengaruh media sosial dalam kampanye politik lokal. *Journal Publicuho*, *7*(1), 439–447.

Acknowledgment
We gratefully acknowledge the contributions of individuals who supported the completion of this article.

Funding Information
This research did not receive any funding.

Conflict of Interest Statement
The authors declare that there is no conflict of interest.

Ethical Approval and Originality Statement
Ethical approval was obtained for this study. The manuscript represents original work and has not been previously published, nor is it under consideration by another journal.

Data Disclosure Statement
The data that support the findings of this study are available from the corresponding author upon reasonable request.