# Research on Public Opinion Hotspot Theme Identification and Guidance Strategies Based on Sentiment Analysis and Semantic Mining

**Bo Wang**
Department of Basic Education, Sichuan Police College, Luzhou 646000, Sichuan, China
Corresponding author: Bo Wang (e-mail: 17778301969@163.com)

**ABSTRACT** Social media environments generate large-scale, rapidly evolving public opinion data, making accurate identification of hotspot themes and effective guidance strategies increasingly important. This study proposes a comprehensive analytical framework integrating fine-grained sentiment analysis and semantic mining to reveal the underlying thematic structures of public opinion evolution. A multi-source data acquisition and preprocessing module is first established to collect heterogeneous social media content. A hybrid sentiment analysis model combining lexicon-based methods and machine learning is then employed to quantify topic-level emotional intensity and sentiment dynamics. Furthermore, semantic network construction and community detection algorithms are utilized to identify latent hotspot clusters and characterize their evolutionary patterns. Based on the multidimensional relationships among topics, sentiment trends, and user roles, a hierarchical guidance strategy generation mechanism is developed. Experimental results demonstrate the effectiveness of the proposed framework in hotspot identification and dynamic trend analysis. The study provides methodological support for large-scale information mining and offers references for information propagation analysis, social sensing systems, and intelligent communication networks.

**INDEX TERMS** sentiment analysis, semantic mining, public opinion identification, information propagation, intelligent communication networks

## I. INTRODUCTION

THE widespread adoption of social media has profoundly reshaped the mechanisms by which public opinion is generated and evolved. Against this backdrop, online public opinion exhibits significant characteristics such as a large volume of information, rapid dissemination, complex and volatile emotions, and interwoven topic structures. This presents a severe challenge for relevant departments to accurately perceive public opinion trends and effectively guide public perception.

This paper aims to construct a comprehensive analytical framework integrating fine-grained sentiment computing and deep semantic network analysis to achieve precise deconstruction and dynamic tracking of the complex thematic structures behind public opinion hotspots. The main contributions of this paper are as follows: First, it proposes a multi-source data collection and preprocessing workflow, laying a high-quality data foundation for subsequent analysis; second, it introduces a sentiment analysis method based on a combination of dictionary and machine learning, decomposing the overall sentiment tendency into multiple discrete dimensions for quantitative scoring, and realizing the calculation of topic-level sentiment intensity, surpassing the limitations of traditional binary sentiment classification; third, it utilizes natural language processing technology to construct a text semantic network, and automatically identifies potential hot topic clusters through community discovery algorithms, ranking and deeply analyzing hot topics from multiple dimensions such as topic intensity and sentiment evolution trends; fourth, based on a multi-dimensional profile of "topic-sentiment-role," it generates a set of hierarchical and categorized precise public opinion guidance strategies, realizing a complete link from "perceiving" public opinion to "cognizing" structure and then to "intervention" guidance.

The structure of this paper is as follows: Part II reviews relevant research on social media and public opinion analysis; Part III elaborates on the research methods proposed in this paper, including data preprocessing, fine-grained sentiment computing, hotspot identification based on semantic networks, and guidance strategy generation models; Part IV verifies the effectiveness and practicality of the proposed methods through case studies in three typical fields, and discusses in depth the characteristics of public opinion evolution in different contexts; Part V summarizes the research results of the entire paper, points out the existing limitations, and provides an outlook on future research directions. This research provides methodological tools for understanding the evolution of public opinion in specific fields such as the market dynamics, and provides an operable path for relevant departments to achieve accurate and efficient public opinion governance on specific issues in the digital age.

## II. RELATED WORK

The rise of social networks has not only accelerated the flow of information, but also spurred a fundamental change in the mechanism of public opinion generation and dissemination. Given the fast speed of information dissemination and wide range of influence of social networks, Lyu et al. [1] proposed a positive public opinion guidance model based on dual learning. Ausat [2] concluded that in the digital age where social media occupies a core position, social media plays an important role in shaping public opinion and thus significantly affects people's economic decisions. Reveilhac et al. [3] found that social media data can effectively supplement traditional surveys, especially in tracking the dynamic changes of public opinion, capturing responses to emergencies, and analyzing discussion networks on specific issues. Bany Mohammed et al. [4] revealed that social media became a key position for digital activism during the Gaza conflict, greatly influencing public opinion at home and abroad. Swastiningsih et al. [5] concluded that traditional media and digital media platforms play different but complementary roles in shaping public opinion.

Kharvi [6] explored the potential and actual negative impacts of deepfake technology on public opinion, political discourse and personal safety. Mudhofi et al. [7] revealed the complex attitudes of the public towards moderate religious ideas through the mining and analysis of social media data. Petrova and Tapsoba [8] explored the complex relationship between information and conflict. Alam et al. [9] explored how sentiment analysis methods that integrate natural language processing and artificial intelligence can profoundly affect the way and depth of our acquisition of public opinion knowledge from social media. Malet [10] proposed whether and how transnational social influence can affect public opinion in a country. This paper aims to integrate fine-grained sentiment computing and semantic network analysis to reveal the deep thematic structure behind public opinion hotspots, in order to provide more scientific methodological support and practical path for precise public opinion guidance in the digital age.

## III. METHODS

### *A. PUBLIC OPINION DATA COLLECTION AND PREPROCESSING*

The goal of this module is to acquire raw, multi-source social media text data and transform it into structured, clean corpus suitable for subsequent analysis. Data collection is primarily conducted through publicly available application programming interfaces (APIs) and social media monitoring tools. Taking domestic social media platforms as an example, real-time crawling and retrospective data collection can be performed targeting preset keyword combinations, specific topic tags, or specified time periods. The collected metadata includes user-posted text content, posting time, unique poster identifier, number of reposts, number of comments, number of likes, and other interaction data.

The preprocessing stage is the key to ensuring the quality of analysis. First, data cleaning is performed, including removing advertising text, duplicate content, and completely meaningless character sequences, and converting traditional Chinese to simplified Chinese. Then, text normalization is performed, such as correcting obvious spelling errors, unifying English capitalization, and converting numbers to uniform symbols. Subsequently, HanLP is used to segment the text, and a custom domain dictionary is loaded to improve the segmentation accuracy of specific nouns (such as technical terms, product models, and organizational abbreviations). Stop word filtering is done in two steps : first, the general stop word list is removed, and then irrelevant high-frequency but low-information words ("Web link", "Image source", "Video loading", etc.) are removed according to the specific analysis scenario [11,12]. Each piece of data after preprocessing becomes a document composed of effective word sequences, and its corresponding metadata is retained to form the initial corpus.

### *B. FINE-GRAINED SENTIMENT COMPUTING AND TOPIC INTENSITY QUANTIFICATION*

This module is intended to move beyond the positive/negative basic binary sentiment classification and attempt to perform multi-dimensional sentiment scoring on text, and calculate the overall sentiment intensity value at the topic level. First, it uses a technique that uses a combination of a sentiment dictionary and machine learning models for fine grain sentiment annotation. An expanded sentiment dictionary is built that consists of multiple dimensions of sentiment (such as "joy", "anger", "sadness", "fear", "surprise", and "disgust"). This dictionary not only contains the how much the sentiment words, but it also mines the synonyms and the popular sentiment expressions on the internet using the word vector models. For each preprocessed text, sentiment words are extracted, and the collocation of degree adverbs and negation words is analyzed to calculate its basic score on each sentiment dimension. Simultaneously, to more clearly present the text, a pre-trained Chinese sentiment analysis model (BERT sentiment classification finetuning model) is used to predict the overall sentiment tendency of the text as a calibration reference. To integrate the advantages of the two methods and ensure the robustness of the results, a multi-stage weight fusion mechanism is designed. First, based on a labeled validation set containing 5,000 texts, the prediction accuracy of the dictionary method and the BERT model on each sentiment dimension is calculated separately. For each sentiment dimension, the initial fusion weights are dynamically set according to the F1 scores of both methods on the validation set, giving higher weights to the better-performing method. Second, to avoid errors that may be introduced by simple linear weighting, a "confidence calibration" step is introduced: when the confidence score (softmax output probability) of the BERT model for its prediction results is below 0.6, the confidence of the model's prediction decreases, and the weight of the dictionary method is increased to 0.8; otherwise, the initial weight based on the F1 score is maintained. Ultimately, the sentiment score vector for each text is derived by weighted summation of these two parts, with each score ranging between -1 (extremely negative) and +1 (extremely positive). For example, a critical text might score highly on the "anger" and "disgust" dimensions (e.g., 0.7, 0.6) while scoring poorly on the "joy" dimension (e.g., -0.8) [13,14]. After initial topic identification, topic intensity quantification can be performed. For the identified topic cluster C, its sentiment intensity $E_c$ is defined as the weighted average of the absolute scores of all documents under this topic on a specific sentiment dimension 1, calculated as follows:

$$E_c(d) = \frac{\sum_{i \in C}(|s_{i,d}| \times w_i)}{\sum_{i \in C} w_i} \tag{1}$$

Where $s_{i,d}$ represents the score of document on the sentiment dimension, and $w_i$ is the weight of document $i$, which can be determined based on the document's popularity (e.g., the normalized value of the combined number of reposts, comments, and likes). The total discussion intensity $I_c$ of topic C is obtained by normalizing and summing indicators such as the total number of documents and total interaction within the cluster, calculated as follows:

$$\begin{aligned} I_c = \alpha \cdot \ & \frac{|C|}{\max_{C'} |C'|} \\ & + (1-\alpha) \cdot \frac{\sum_{i \in C} \text{engagement}_i}{\max_{C'} \sum_{i \in C'} \text{engagement}_i} \end{aligned} \tag{2}$$

Where $|C|$ is the total number of documents in the cluster, $\text{engagement}_i$ is the total interaction of the document $i$ (e.g., the sum of reposts, comments, and likes), and $\alpha$ is a coefficient balancing the weights of document quantity and interaction ($0 \leq \alpha \leq 1$, in this paper, $\alpha = 0.5$). By combining sentiment intensity and discussion intensity, we can assess "popular topics with high anger intensity" or "heartwarming topics with high joy intensity," thus providing data support for in-depth analysis.

### C. HOT TOPIC IDENTIFICATION BASED ON SEMANTIC NETWORKS

The goal of this module is to automatically discover and structure the core issue clusters emerging in the public discourse from massive amounts of text. First, a document-term matrix is constructed. The TF-IDF algorithm is used to evaluate the importance of terms in the entire corpus, filtering out terms with excessively low or high TF-IDF values (to remove rare words and generally meaningless high-frequency words), thus forming a feature word space.

Secondly, a semantic co-occurrence network is constructed. Filtered feature words are used as network nodes. A sliding window (e.g., using sentences or a fixed number of words) is set up to count co-occurring word pairs within the window. If the co-occurrence frequency of a word pair exceeds a set threshold, an edge is created between the two nodes, and the weight of the edge represents its co-occurrence frequency. In this way, a large semantic network consisting of words (nodes) and their co-occurrence relationships (edges) is established. Next, the Infomap algorithm is applied to partition the semantic network. This algorithm can gather closely connected nodes into the same community. Each community corresponds to a potential hot topic cluster, and the set of nodes (keywords) in the community constitutes the core vocabulary representation of the topic. For example, a topic cluster about "new energy vehicle battery safety" may include core nodes such as "battery", "spontaneous combustion", "range", "detection", "standard" etc. [15] Finally, trending topics are ranked and labeled. Not all identified topic clusters are trending topics. This study designed a comprehensive popularity index $H_c$ to rank topic clusters. This index comprehensively considers topic cohesion, discussion scale, emotional intensity, and participation in key nodes. Its calculation formula is:

$$\begin{aligned} H_c = \beta \cdot \ & \frac{\sum_{e \in EC} w_e}{\max_{C'} \sum_{e \in EC'} w_e} + \gamma \cdot \frac{|C|}{\max_{C'} |C'|} \\ & + \delta \cdot \frac{EC}{\max_{C'} EC'} + \eta \cdot \frac{KC}{\max_{C'} KC'} \end{aligned} \tag{3}$$

Where, $\sum_{e \in EC} w_e$ represents the sum of the weights of all edges within cluster C, representing topic cohesion; EC is the absolute value of the average emotional intensity of documents within the cluster; KC is the number of documents published by influential users (such as users with more than a certain threshold of followers) within the cluster. The coefficient $\beta$, $\gamma$, $\delta$, $\eta$ represents the weight of each indicator, satisfying $\beta + \gamma + \delta + \eta = 1$.

The system automatically generates concise and accurate topic tags based on the weight and co-occurrence strength of core keywords within a cluster, such as "battery safety standard controversy" and "charging infrastructure anxiety," thereby achieving automated labeling and naming of hot topics.

### D. GUIDING STRATEGY GENERATION MODEL

The generation of guidance strategies is based on the results of the above three-layer analysis (theme, sentiment, and dissemination), forming a three-dimensional profile of "theme-sentiment-role". For each high-profile theme, we analyze its dominant sentiment dimension, sentiment evolution trend (changing over time), key dissemination nodes (such as high-influence accounts and highly interactive posts), and the distribution of core viewpoints.

The guidance strategy model outputs a categorized and graded strategy suggestion library based on a 3D profile. The strategies are mainly divided into three categories: First, information supply strategies, targeting topics where negative emotions arise due to ambiguous or distorted information. Strategy suggestions include “an authoritative institution (such as XX department) should release clarification information through the ZZ platform within the YY time window, focusing on addressing questions A, B, and C.” Second, emotional guidance strategies, targeting topics arising from emotional pent-up feelings or resonance. Strategy suggestions include “organizing live interviews with experts in relevant fields or those who have experienced the events firsthand, telling the story behind D, communicating with empathy, and softening opposing emotions.” Third, agenda-setting strategies, targeting topics that may be maliciously misled or require positive guidance. Strategy suggestions include “proactively setting sub-topics E that are related to core concerns but have a positive perspective, encouraging creators in the F field to produce relevant content, and driving the discussion in that direction.” Each strategy is associated with specific operational guidelines, expected goals, and effectiveness evaluation indicators (such as changes in scores on specific emotional dimensions and the proportion of discussion on related sub-topics).

## IV. RESULTS AND DISCUSSION

### *A. CASE STUDY 1: ANALYSIS OF THE EVOLUTION OF PUBLIC OPINION THEMES AND SENTIMENT IN THE NEW ENERGY VEHICLE INDUSTRY*

This study selected a battery safety controversy involving a well-known new energy vehicle brand as the subject of analysis. After data processing and topic identification, five main topic clusters were extracted and sorted by their comprehensive popularity index, as shown in Table 1.

**TABLE 1.** Public Opinion Topic Identification and Sentiment Analysis of New Energy Vehicle Battery Safety Incidents

| Theme cluster tags | Top 5 core keywords | Emotion-driven dimension (score) | Number of documents within a cluster |
|---|---|---|---|
| T1: Battery Safety Standards and Testing Controversies | Batteries, national standards, testing, spontaneous combustion, regulation | Anger (0.72), Fear (0.65) | 45,820 |
| T2: Evaluation of Corporate Response and Public Relations Strategy | Statement, apology, shifting blame, sincerity, press conference | Disgust (0.68), Anger (0.61) | 38,150 |
| T3: Car Owners' Rights Protection Experiences and Safety Anxiety | The car owner is afraid, charging the car, filing a complaint, and is afraid to drive it. | Fear (0.80), Sadness (0.55) | 32,500 |
| T4: Competitive Product Comparison and Industry Impact Discussion | Brand XX (competitors), industry, trust, sales volume, domestic products. | Surprise (0.60), Disgust (0.45) | 15,200 |
| T5: Technology Popularization and Future Improvement Prospects | Solid-state batteries, thermal management, technology, advancements, expectations. | Surprise (0.50), Joy (0.30) | 8,900 |

Public discussion focused heavily on two core issues: “safety standards” and “corporate responses,” accompanied by strong negative emotions. The T1 theme revealed that the deep-seated concern was not about a single incident, but rather a systemic questioning of industry standards and regulatory effectiveness; its high scores for “anger” and “fear” pointed to a crisis of public trust. The T2 theme showed that the companies’ initial responses failed to effectively quell emotions, instead exacerbating “disgust” due to perceived insincerity. The T3 theme demonstrated the spread of negative emotions from macro-level issues to individual experiences, forming specific safety anxieties. In contrast, the T4 and T5 themes had lower engagement, indicating that rational comparisons and technological perspectives were relatively weak in the early stages of the crisis. Semantic network analysis further revealed numerous key connections between the T1 and T2/T3 theme clusters (such as “detection” and “apology,” “car owner” and “fear”), forming a reinforced narrative chain of “lack of standards -> corporate negligence -> user victimization.” Based on this, the suggestions generated by the guidance strategy model prioritize information supply: it is recommended that regulatory authorities intervene immediately to clarify the specific items of the current battery safety testing standards and the list of authoritative testing institutions, cutting off the chain of speculation about “lack of standards”; at the same time, it is recommended that companies conduct emotional counseling, with top management directly talking to representative car owners to acknowledge their responsibility for safety anxiety, rather than just discussing the matter at hand.

### B. CASE STUDY 2: EMERGENCE AND RESONANCE OF PUBLIC OPINION THEMES IN PUBLIC SAFETY INCIDENTS

The analysis focused on early public opinion surrounding a sudden public safety incident in a certain area. Within 48 hours of the incident, approximately 120,000 discussion data points were rapidly collected and analyzed, identifying four rapidly evolving topic clusters. The changes in their popularity and sentiment over time are shown in Table 2.

TABLE 2. Dynamic Evolution of Public Opinion Topics in the Early Stages of Public Safety Incidents (every 12-hour interval)

| Time phase | Core topic clusters (sorted by popularity at different stages) | Dominant Emotional Dimension | Average emotional intensity | Related themes to the previous stage |
|---|---|---|---|---|
| 0-12 hours | S1: Description of the incident scene and casualties | Fear, sadness | 0.85 | - |
| | S2: Emergency Rescue Progress Report | Surprise, anticipation | 0.40 | - |
| 12-24 hours | S1: Speculation on the cause of the incident and attribution of responsibility | Anger, disgust | 0.78 | Evolved from S1 |
| | S3: Evaluation of Authorities' Notifications and Information Disclosure | Anger, surprise | 0.70 | New |
| | S2: Touching details of the rescue and prayers | Sadness, joy | 0.35 (Joy comes from affirmation of the rescue efforts) | continued |
| 24-48 hours | S3: Comparison and accountability of similar historical events | Anger, disgust | 0.75 | Continue and deepen |
| | S4: Safety Hazard Investigation and System Reflection | Fear, Expectation | 0.60 | Derived from S1 and S3 |
| | S1/S2: Heat decay | - | - | - |

In public safety emergencies, public opinion exhibits a clear evolutionary path of "fact description -> attribution inquiry -> institutional reflection," with the emotional tone rapidly shifting from initial fear and sadness to intense anger and skepticism. In the first 0-12 hours, the core demand in the public sphere is information, with theme S1 dominating. From 12-24 hours, as more fragmented information floods in, theme S1 quickly evolves into speculation about the causes. Simultaneously, the newly emerging theme S3 (evaluation of official reports) is directly related to S1, exhibiting high emotional intensity, indicating that insufficient or untimely initial information supply will immediately give rise to skepticism. From 24-48 hours, the discussion enters a deeper phase, with theme S4 emerging, shifting the focus from a single event to systemic security, forming a resonant narrative of "historical comparison - current accountability - future prevention." Semantic network analysis reveals that the path connecting "rescue" (S2) and "accountability" (S3) is far weaker than the path connecting "casualties" (S1) and "accountability" (S3), suggesting that emotional sentiment cannot offset the rational demand for accountability. Based on this, the guidance strategy model suggests adopting a "rhythm response" strategy: In the first 0-12 hours, the strategy focuses on saturating information supply (S1, S2), continuously releasing verifiable information from the scene to suppress rumors; in the last 12-24 hours, the strategy must shift to proactively setting the agenda, guiding public attention to S4 (institutional reflection), and channeling anger towards constructive reflection through interpretation by authoritative experts and the release of investigation plans, thus preventing emotions from accumulating on the S3 (accountability) theme.

### C. CASE STUDY 3: LONG-TERM THEMATIC EMOTIONAL TRACKING OF CONTROVERSIAL TOPICS IN THE ENTERTAINMENT INDUSTRY

This paper selected a controversial event in the entertainment industry that lasted for more than a month as the subject of our analysis and conducted full-cycle thematic sentiment tracking. Each month was divided into an analysis period , for a total of three periods. Key data comparisons are shown in Table 3.

**TABLE 3.** Multi-period theme-sentiment comparison of long-term controversial events

| Analysis cycle | Core topic clusters (top three in terms of popularity within the period) | Core emotional dimension (intensity) | Average user profile characteristics | Key propagation node types |
|---|---|---|---|---|
| First cycle (outbreak period) (Week 1) | C1: List of the causes and facts of the incident | Surprise (0.88), Anger (0.70) | Entertainment for the masses, diverse | Entertainment influencers, mainstream media |
| | C2: Retrospective analysis of the historical behavior of both parties involved | Disgust (0.75), Anger (0.65) | | |
| | C3: Fan group conflict and infighting | Anger (0.80), Disgust (0.70) | | |
| Second cycle (standoff period) (Weeks 2-3) | C3: Fan group conflict and infighting | Anger (0.85), Disgust (0.75) | Core fans, haters, and those who enjoy entertainment. | Fan leader, marketing account |
| | C4: Legal and ethical analysis | Surprise (0.40), Neutral | Those interested in law and education | Professional blogger |
| | C2: Heat has decreased but still exists | Dislike (0.60) | | |
| Third cycle (sedimentation period) | C5: Industry Impact and Reflections | Neutral, Surprised (0.30) | Industry observer, media professional | Industry self-media, institutional media |
| (Week 4 and beyond) | C4: Legal and ethical analysis | neutral | | |
| | C3: Residual layered opposition reduced. | Anger (0.50) but range | extreme fans | Small circle opinion leaders |

The results reveal the "decentralization" and "stratification" characteristics of the long-term evolution of controversial public opinion. During the initial outbreak phase, the themes were diverse, emotions were intense, participants were widespread, and key nodes were mostly accounts with broad influence. Entering the confrontation phase, the popularity of public topics (C1, C2) declined, and the core battleground shrank to the highly polarized topic of fan infighting (C3), with emotional intensity reaching its peak. At this point, the participating user profiles became highly specialized, and the dissemination nodes became the "leaders" within specific strata, forming an emotional resonance within an information cocoon. Although rational discussion topics (C4) emerged, their popularity and emotional intensity were far less than C3, making it difficult to penetrate the strata. In the settling phase, public attention shifted, the overall popularity of the public opinion field declined, and the remaining discussions diverged into two directions: one was rational reflection directed at industry observers (C5), with emotions tending towards neutrality; the other was solidified opposition within a small area (remnants of C3), with reduced intensity but extremely stubborn stances. Semantic network analysis reveals that the core community structure of the network has undergone dramatic changes in different cycles, from a multi-center, tightly interconnected structure during the outbreak phase, to a few dense and opposing clusters during the confrontation phase, and then to sparseness during the consolidation phase. The guidance strategy model suggests a "tiered and segmented" strategy for such events: during the outbreak phase, focus on intervening in factual topics (C1), promoting authoritative sources to set the tone, and preventing the establishment of rumors; during the confrontation phase, identify and restrict the spread of highly inflammatory key nodes in extremely opposing topics (C3), while amplifying the spread of rational voices (C4) to provide traffic support for their "breakthrough"; during the consolidation phase, focus on and promote constructive topics (C5), transforming them into valuable public discussions.

## V. CONCLUSION

This research overcomes the limitations of existing social media sentiment analysis methods—such as shallow topic detection and ambiguous assessment—by providing a specialized framework for deciphering the complex technical and ethical discourse surrounding material innovations and production standards. By replacing generalized strategies with industry-specific insights, the study enables precise and targeted guidance of public opinion regarding the material science and environmental impact of modern materials. It systematically builds a comprehensive analytical framework combining fine-grained sentiment computation and deep semantic network mining and tests its feasibility and effectiveness by multi-case empirical analysis. The study finds that trending public opinion is not a monolithic entity but rather a complex ecosystem made up of multiple interlinked sub-topics with different tones of emotion attached to them. Fine-grained sentiment analysis can accurately quantify the public's emotional states in different dimensions such as "anger," "fear," "disgust," and "expectation" and therefore goes beyond "support/oppose" dichotomy and understands the specific motivations behind these emotions. Semantic network community discovery can automatically and structure extract these sub-topic clusters in a large amount of texts to make their relationship clear and response visually shows that internal structure of the public opinion field.

There are also some limitations in this study. First, the analysis being conducted relies heavily on text data and its capability of extracting sentiment and themes from multimodal data such as images, videos and live streams is inadequate. Second, the construction of the semantic networks relies on word co-occurrence; and its presentation of the ability to deal with complex phenomena in linguistic life such as implicit semantic connections and irony has room for improvement. Future research can explore more deeply the multimodal information fusion, more advanced context-aware semantic representations (like deep semantic understanding based on large language models), as well as the automated generation and effect simulation and prediction of guidance strategies. This will further strengthen the intelligence and practical ability of the public opinion analysis system, offering a more powerful technological toolbox to construct a clean cyberspace and ensure the healthy development of industry reputation through more effective social governance. By aligning advanced sentiment monitoring with the specific nuances of material ethics and industrial standards, this system provides the necessary precision to maintain a balanced and scientific public discourse.

## AUTHOR CONTRIBUTIONS

Bo Wang designed, collected and analyzed the data, and drafted the manuscript. Bo Wang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Bo Wang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## CONFLICTS OF INTEREST

The author declares no conflict of interest.

## FUNDING

This research was supported by Intelligent Policing and National Security Risk Management Laboratory, Sichuan Police College, 2026 Annual Independent Project "Research on Collaborative Mechanism of AI-Driven Cyber Language Risk Perception, Governance, and Prevention" (Grant No. ZHZZYB2602).

The Program for Scientific Research Start-up Funds of Sichuan Police College (Grant No. 2025YJBSQD30).

## ACKNOWLEDGEMENTS

Not applicable.

## REFERENCES

[1] B. Lyu, Y. Du, J. Jiang, et al., "Positive public opinion guidance model based on dual learning in social network," Intelligent Data Analysis, vol. 27, no. 3, pp. 833-853, 2023, doi: 10.3233/IDA-226602.

[2] A. Ausat A M, "The role of social media in shaping public opinion and its influence on economic decisions," Technology and Society Perspectives (TACIT), vol. 1, no. 1, pp. 35-44, 2023, doi: 10.61100/tacit.v1i1.37.

[3] M. Reveilhac, S. Steinmetz, and D. Morselli, "A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion," Multimedia tools and applications, vol. 81, no. 7, pp. 10107-10142, 2022, doi: 10.1007/s11042-022-12101-0.

[4] A. Bany Mohammed, M. Al-Okaily, D. Qasim, et al., "Digital activism and public opinion: understanding the role of social media during the Gaza Conflict," Journal of Islamic Marketing, vol. 16, no. 11, pp. 3366-3393, 2025, doi: 10.1108/JIMA-03-2024-0101.

[5] S. Swastiningsih, A. Aziz, and Y. Dharta, "The role of social media in shaping public opinion: a comparative analysis of traditional vs," digital media platforms. The Journal of Academic Science, vol. 1, no. 6, pp. 620-626, 2024, doi: 10.59613/fm1dpm66.

[6] L. Kharvi P, "Understanding the impact of AI-generated deepfakes on public opinion, political discourse, and personal security in social media," IEEE Security & Privacy, vol. 22, no. 4, pp. 115-122, 2024, doi: 10.1109/MSEC.2024.3405963.

[7] M. Mudhofi, I. Supena, A. Karim, et al., "Public opinion analysis for moderate religious: Social media data mining approach," Jurnal Ilmu Dakwah, vol. 43, no. 1, pp. 1-27, 2023, doi: 10.21580/jid.v43.1.16101.

[8] M. Petrova and A. Tapsoba, "Information and conflict: From the role of (social) media and public opinion to big data and forecasting," Economic Policy, vol. 40, no. 124, pp. 845-878, 2025, doi: 10.1093/epolic/eiaf012.

[9] S. Alam M, H. Mrida M S, and A. Rahman M, "Sentiment analysis in social media: How data science impacts public opinion knowledge integrates natural language processing (NLP) with artificial intelligence (AI)," American Journal of Scholarly Research and Innovation, vol. 4, no. 01, pp. 63-100, 2025, doi: 10.63125/r3sq6p80.

[10] G. Malet, "Cross-national social influence: How foreign votes can affect domestic public opinion," Comparative Political Studies, vol. 55, no. 14, pp. 2416-2446, 2022, doi: 10.1177/00104140221088846.

[11] M. Hasanuddin, S. Khodijah, and A. Rizki C, "Analysis of the Role of Social Media in Shaping Public Opinion on Social Issues," Journal of Computer Science Artificial Intelligence and Communications, vol. 1, no. 1, pp. 7-11, 2024, doi: 10.64803/jocsaic.v1i1.1.

[12] T. Li, Z. Zeng, J. Sun, et al., "Using data mining technology to analyse the spatiotemporal public opinion of COVID-19 vaccine on social media," The Electronic Library, vol. 40, no. 4, pp. 435-452, 2022, doi: 10.1108/EL-03-2022-0062.

[13] C. Liu, "Who supports expanding surveillance? Exploring public opinion of Chinese social credit systems," International Sociology, vol. 37, no. 3, pp. 391-412, 2022, doi: 10.1177/02685809221084446.

[14] A. Suryawati I G A, W. Wahyuandari, A. Halmina, et al., "The Influence of Social Media on Public Opinion: An Analysis from Literary and Cultural Perspectives," The Journal of Academic Science, vol. 1, no. 8, pp. 1171-1178, 2024, doi: 10.59613/xfr4rv11.

[15] S. Nikitina A, V. Ruchkin A, and S. Isacov A, "The Concept of Digital Social Control: Results of Public Opinion," International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies, vol. 13, no. 4, pp. 1-11, 2022, doi: 10.14456/ITJEMAST.2022.68.