Analyzing the Dynamics of Hot Searches on TOPBOZZ, Mirco-blog and ZHIHU with the help of LLM: A Multidimensional Exploration of Public Attention Across Different Years

Xiaoxiao Ma

Xi'an International Studies University, Xi'an, Shaanxi, China

2646432529@qq.com

Abstract. Based on a total of 853,656 hot searches data from the three platforms Mirco-blog, ZHIHU and TOPBOZZ from 2021 to 2025, this study systematically analyses the distribution of hot searches on different platforms in terms of category distribution, temporal evolution and differences in public attention with the help of Large Language Model (LLM) for multilevel classification and keyword extraction, and combining with visualization tools like Tree Diagrams, Sanki Diagrams, and Line Charts. It is found that the hot searches on the three platforms show significant differences in content, and periodic events such as college entrance exams, show regular peaks. In addition, the study reveals how the Algorithm - driven hot searches agenda maps social sentiment and public issues, which provides a reference for public opinion monitoring, media optimization and policy making. In the future, it can be further combined with sentiment analysis techniques to deepen the capture and response to public sentiment dynamics.

Keywords: Hot searches, Large Language Model(LLM), Public Attention, TOPBOZZ, Mirco-blog, ZHIHU, Multilevel classification.

1 Introduction

The 55th Statistical Report on the Development of the Internet in China shows that as of December 2024, China's Internet users numbered 1.108 billion, with social network users comprising 99.3% of internet users[1]. The rapid growth of internet technology has established platforms like Microblog as key sources of public information, with social media becoming central to academic research [3]. Within this digital landscape, the concept of hot searches has emerged—lists generated algorithmically based on real-time metrics (e.g., search volume, click rates) that aggregate public attention across the internet. Hot searches provide insights into social sentiment, cultural trends, and public issues, acting as a lens for public opinion. Wu Xiaokun (2020) pointed out that the hot searches are shaped by algorithmic, media power, and economic factors[4]. The public discussion agenda constructed by hot searches not only shapes people's daily behaviors[2], but also influences their perception of the real world. Yu Xupeng (2023)

B. Majoul et al. (eds.), *Proceedings of the 2025 11th International Conference on Humanities and Social Science Research (ICHSSR 2025)*, Advances in Social Science, Education and Humanities Research 945,
https://doi.org/10.2991/978-2-38476-440-2_150

proposed the concept of 'hot search politics', suggesting that the politics of hot searches reflects the dynamic connection between citizens, leadership, and the political system in the digital media era[10]. Hot searches have become an important basis for the public to judge the degree of attention paid to an event, phenomenon, or person, and have become an important carrier and support for the attention of public opinion. Regarding the impact of new media and new technologies such as Large Language Model (LLM) on traditional media, JongRoul et al. (2014) clarified that the emergence of new media will, to a certain extent, contribute to the development and innovation of traditional media [10]. In terms of the impact and inspiration of hot searches on the media industry, Robinson (2011) proposed the concept of 'participatory journalism', which calls for turning the news production process into a dynamic process involving multiple parties and changing the relationship between organizations and their audiences [12]. In Dubois' (2020) article on social media and journalism, it is also mentioned that many journalists now use social media to infer and report on public opinion [13], and Fürst & Oehmer (2021) introduced 'public reaction' as a new news factor[14], suggesting that in the digital era, the influence of hot searches among the public and the public's reaction to them are of value and significance to be explored.

This study investigates hot search data from TOPBOZZ, Microblog, and ZHIHU between 2021 and 2025, aiming to categorize and analyze the evolution of public attention on these platforms. While previous studies have focused on hot search content on media platforms[5][6][7], the temporal evolution and platform differences in public attention remain underexplored. In the research, due to the high labour cost resulted by the large amount of data, this study uses the Large Language Model (LLM) to help the processing and human intervention such as cue engineering. And the study integrates visual tools for deeper insights into public sentiment dynamics, which can provide practical references to the related hot searches or news classification research work.

2 Research Design

2.1 Research Data

With the growth of internet usage, hot searches have become a key window into public sentiment. This study focuses on data from Microblog, ZHIHU, and TOPBOZZ, three of China's leading social platforms. These platforms offer diverse user groups, allowing for a broad spectrum of insights, satisfy the diversified needs of the public in obtaining information. Data was collected between April 13, 2021, and January 17, 2025, amounting to 853,656 hot search entries.

2.2 Research Framework and Process

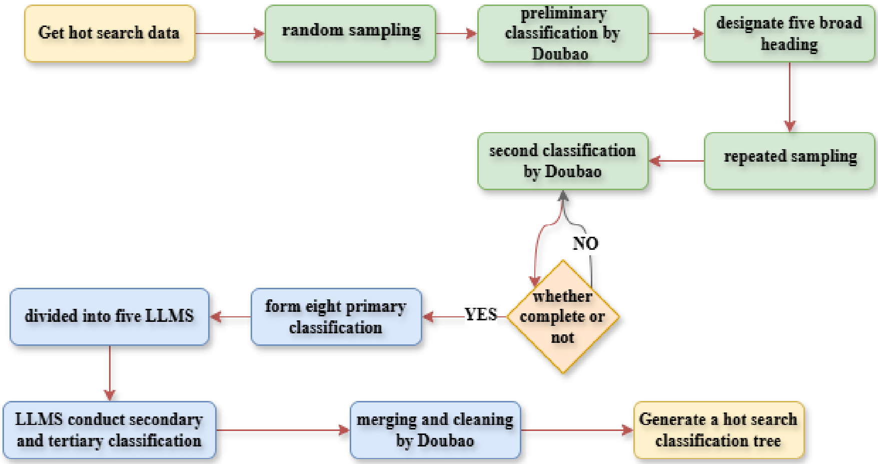

Fig. 1. Hot searches category generation flowchart.

Figure 1 shows the process design of this study. Hot searches from the three platforms were randomly sampled and classified using the Doubao LLM, which incorporated filtered keywords and text analysis. Initially categorized into five broad categories—international affairs, social issues, culture, sports and entertainment, business and technology, medical treatments and public health—these were refined through multiple iterations into eight primary categories: medical treatments and public health, international affairs, business and technology, culture, sports and entertainment, educational information, transportation, and environmental disasters. To enhance classification accuracy, these results were processed using five LLMs (Doubao, ChatGPT, KIMI, Wenxin Yiyan, and Tongyi Thousand Questions). After secondary and tertiary classification, the results were aggregated, and synonyms were merged to ensure each category was distinct.

The accuracy of LLM processing was verified at 90.8%, with manual checks ensuring consistency.

To improve LLM processing, prompts were optimized using the Volcano Engine, which enabled iterative improvements and visual debugging. The specific implementation process is as follows:

Initial Prompt Design: Define the initial instruction template based on the task objectives, and adopt a hierarchical structure design. Task description: read the hot search categories and hot searches data carefully; compare the hot search data with the given hot search categories one by one, and judge the first level of classification that it is most compatible with; continue to classify under the classified categories; extract the keywords within the hot searches; and output the classification results and keywords. Interactive debugging process: Multi-dimensional debugging through the Prompt Studio of the Volcano Engine: fill in the Prompt and fill in the variable contents, which are hot search categories and hot search information, the rich variable data helps the big model to cover a diversity of cases and learn sufficiently, so as to generate specific answers. After filling the variable content, click ‘Generate Model Answer’ on the left side of the debugging interface to get the answer of the big model and debug the effect of Prompt. Finally, add the data (model responses, scores, ideal responses) to the evaluation set to prepare for intelligent optimization. Evaluation System Construction: The evaluation system is divided into three points: content accuracy, content completeness, and format standardization. The categories must accurately match the given hot searches, no new categories added, and no more than 10% of the given hot searches omitted. The keywords extracted can accurately reflect the key information such as locations, people, events and other key information in the hot searches, and the percentage of correct extraction should reach 80% and above. Answers should be presented in the <Results> tag in the format of ‘hot searches, primary category, secondary category, tertiary category, keywords.

2.3 Hot Searches Classification Processing Based on Doubao API

Using the optimized prompt and chain of thought, we call the API of Doubao LLM to classify each hot search and extract the keywords, and after removing the invalid data, Doubao LLM processed a total of 829,233 hot searches. Among them, 145385 hot searches were not classified accordingly. The processed hot searches are then sorted into excel format, which contains hot search information, time, platform, primary classification, secondary classification, tertiary classification and keywords.

3 Presentation and Analysis of Results

3.1 Tree Diagram of Hot Searches Classification

Figure 2 is a Reingold-Tilford Tree Diagram (a tree layout algorithm) illustrates the classification hierarchy of hot searches. The centre of the graph is the root node, which extends outwards to several major branches, each branch represents different categories and subcategories of hot searches, with a total of secondary classifications. There are a total of eight categories of hot searches, namely, medical treatments and public health, international affairs, business and technology, culture, sports and entertainment, educational information, transportation, and environmental disasters, and there are too many tertiary classifications to be shown here.

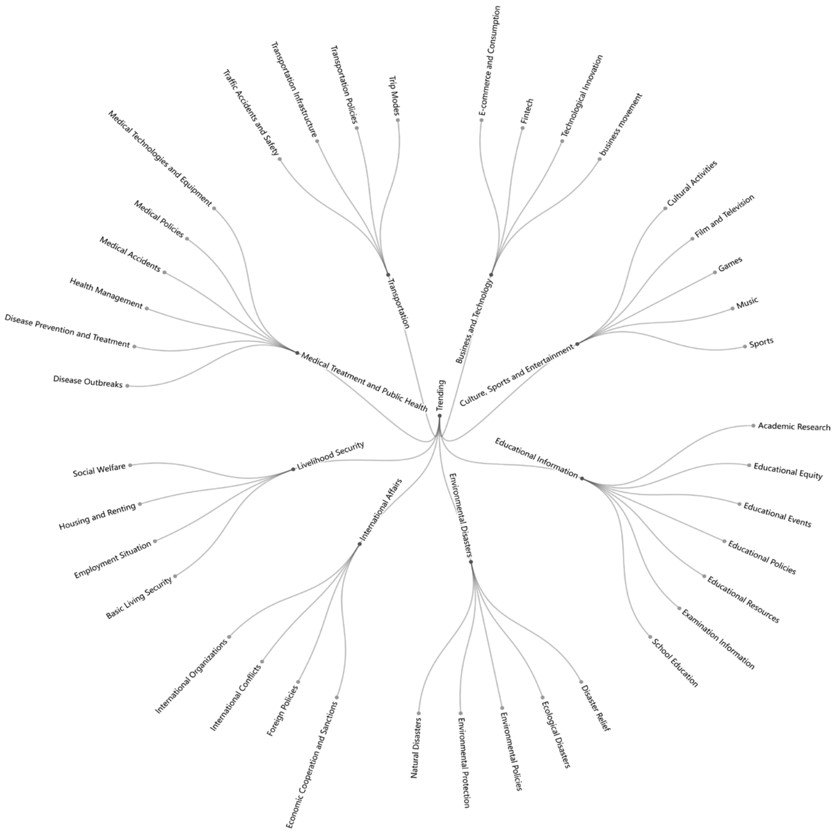

Fig. 2. Tree Diagram of Hot Searches

3.2 Pie Chart of Hot Search Distribution

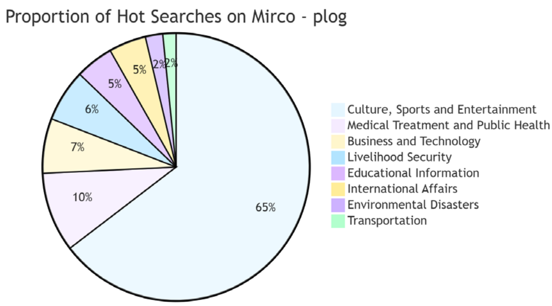

Fig. 3. Mirco-blog hot searches classification ratio

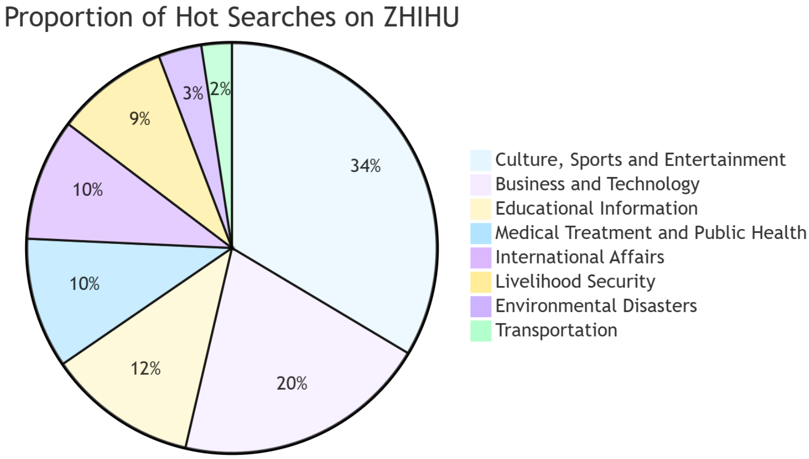

Fig. 4. ZHIHU hot searches classification ratio.

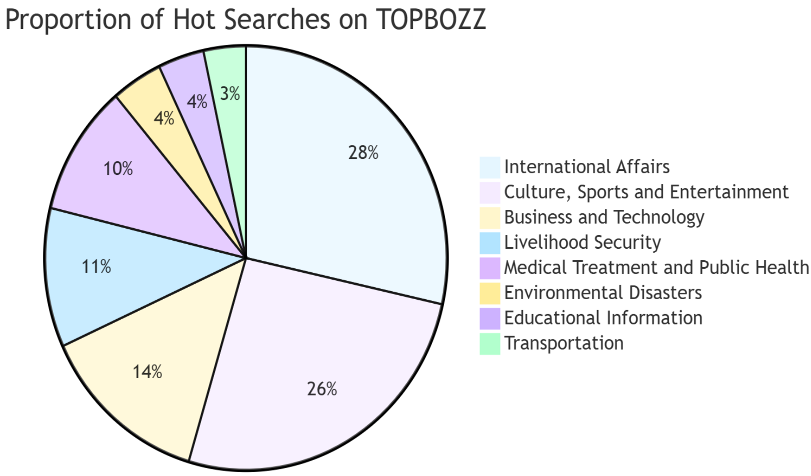

Fig. 5. TOPBOZZ hot searches classification ratio

Figure 3,4,5 show the hot searches classification ratio on three platforms. The pie chart compares the proportion of hot searches across categories for each platform. Microblog shows a strong focus on culture, sports, and entertainment (65%), ZHIHU is more balanced with significant attention to business and technology, medical treatment and public health and so on, reflecting the characteristics of its knowledge-sharing and diversified discussion, while TOPBOZZ places a premium on international affairs (28%) and culture, sports and entertainment (26%), indicating that the platform's users pay more attention to international news and cultural and entertainment information. These differences reflect the different characteristics of user groups, functional positioning and content ecology of each platform.

3.3 Sankey Diagram of Hot Search Flows

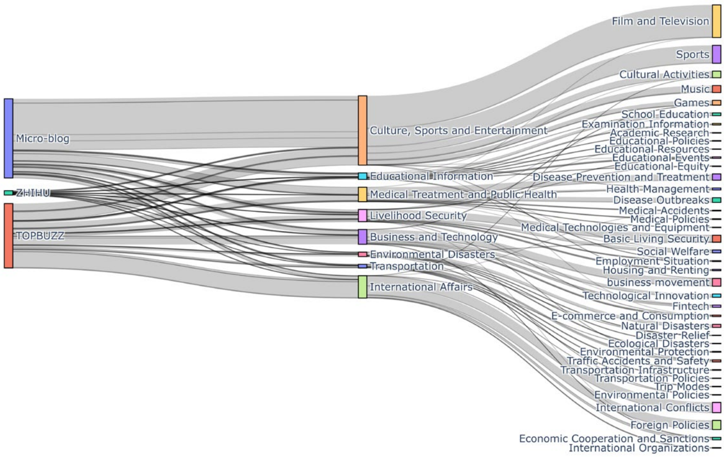

Fig. 6. Sankey diagram of three platforms' hot search

Figure 6 is a Sankey diagram that presents the flow relationship of hot search information between different platforms and content areas. The starting node shows three information source platforms, Mirco-blog, ZHIHU and TOPBOZZ, the second node represents the eight first-level categories of hot searches, and the terminating node represents the second-level categories of the hot search classification, which demonstrates the flow direction of hot search information between the first-level and second-level classifications of the three platforms. The thickness of the lines reflects the size of the information flow from each platform to the corresponding field, the thicker it is, the larger the flow is, which intuitively demonstrates the attention and information distribution of each platform to different content areas, presenting a diversified diversification of information. The chart shows that culture, sports and entertainment have the largest data traffic among the three platforms, which is an important area of traffic convergence, and are internally subdivided into multiple directions, indicating that the public pays great attention to this area, and the points of interest are more dispersed, covering film and television, sports, music and other aspects. ZHIHU shows a strong focus on educational information, such as academic research, education policy, and there is also some traffic distribution in the categories of medical and public health, business and technology, while TOPBOZZ emphasizes international affairs and sports. Overall, this Sankey diagram reflects the tendency of users on different platforms to pay attention to various types of hot searches, as well as the importance and segmentation of different areas in information dissemination.

3.4 Hot Searches Line Charts

Specifically in this study, in order to show the evolution characteristics of various types of hot searches in different periods within five years in a more detailed way, the timeframe of the hot search data was divided into 47 nodes according to the unit of month, and three line charts were produced by platforms.

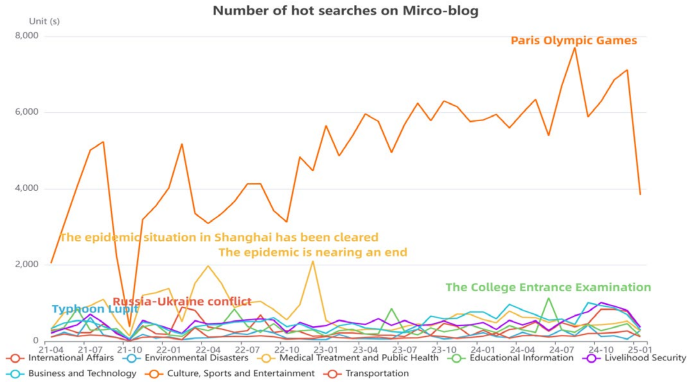

Fig. 7. Mirco-blog hot search line chart

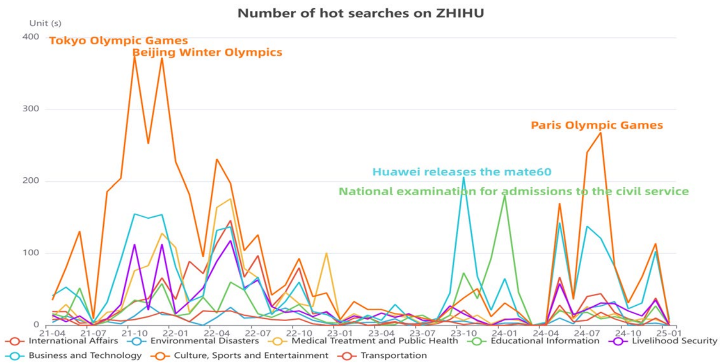

Fig. 8. ZHIHU hot search line chart

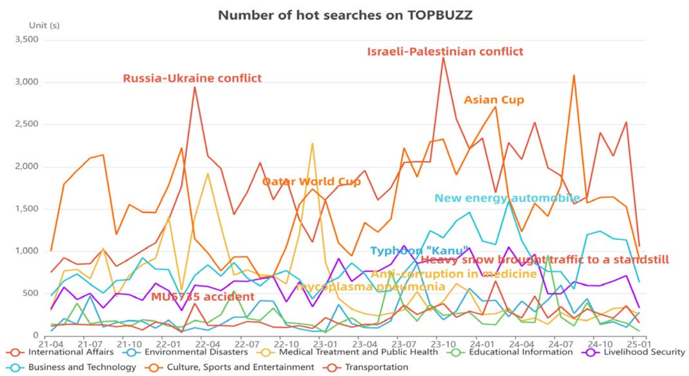

Fig. 9. TOPBOZZ hot search line chart

Figure 7,8,9 show the hot searches line charts on three platforms. Analyzing the above line graph, it can be seen that the three platforms Mirco-blog, ZHIHU and TOPBOZZ possessed common trends, which were the high heat of sports events, the attention paid to sudden events and the attention paid to periodic regular events. Among the three platforms, the number of hot searches in the category of culture, sports and entertainment showed significant peaks during the organization of major sports events such as the Olympic Games (Tokyo Olympics, Beijing Winter Olympics, Paris Olympics), the Asian Cup and the Qatar World Cup. This shows that sports events have a strong appeal and can trigger the attention and discussion of a large number of users. And some international conflict events, such as the Russia-Ukraine conflict and the Palestine-Israel conflict, triggered a high number of hot searches on the TOPBOZZ platform, showing that such major international events are one of the focuses of public attention. Meanwhile, on the Mirco-blog platform, public health-related emergencies such as the Shanghai epidemic clearing up and the epidemic nearing its end also led to a significant rise in the number of hot searches in the medical treatment and public health category. Some events with cyclical patterns receive much attention at the corresponding time of the year. For example, the hot searches in the educational information category peak in June every year due to the domestic high school and college entrance exams, which are reflected in every platform. From the overall hot searches trend changes of the three line charts, the number of hot searches in different categories in different time dimensions shows obvious fluctuations in each platform. This reflects the impact of the occurrence and development of various events on public attention over time. For example, at specific time periods in different years, the number of hot searches in corresponding categories fluctuates due to the concentrated outbreak or continuous fermentation of related events. Meanwhile, as time advances and develops, some emerging hotspots continue to emerge, such as the release of Huawei Mate60, which triggered a high degree of heat on both ZHIHU and Mircoblog platforms, reflecting the new focus brought by technological development. Meanwhile, the number of hot searches for new energy vehicles on the TOPBOZZ platform has gradually risen, reflecting the impact of industry development on public topics.

The three charts again highlight distinct content trends across the platforms. On Microblog, culture, sports, and entertainment dominate, especially during events like the Paris Olympics, which saw significantly higher search volumes than other categories. While topics such as education (e.g., college entrance exams) and environmental disasters (e.g., typhoons) also attract attention, the platform's focus remains on entertainment. ZHIHU, with its knowledge-sharing focus, sees higher interest in educational content like civil service exams and career development. TOPBOZZ, in contrast, features a broader range of interests, with notable attention to international affairs, sports, traffic events (e.g., the MU5735 accident), environmental disasters, and business technology (e.g., new energy vehicles), reflecting its diverse user interests.

4 Inspiration and Recommendations

4.1 Practical Significance of Hot Searches in Society

By analyzing the dynamics of hot searches and public attention across the three platforms, it becomes clear that hot searches, driven by algorithms, effectively capture public focus and sentiment in real time, and are also a convenient window for the public to express their views freely, forming a public sphere for discussing[8]. For instance, the emergence of the "Shanghai epidemic" topic on Microblog significantly increased healthcare-related searches, reflecting the public's concerns and expectations regarding public health events. Due to their timeliness and broad influence, hot searches not only mirror public attention but also guide the trajectory of public opinion, which can, in turn, pressure authorities and prompt offline actions. For example, the hot searches of ‘MU5735 accident’ triggered large-scale concern in the current time period, prompting airlines to revise safety policies. Similarly, discussions surrounding international affairs have influenced diplomatic discourse, becoming a key component of civic diplomacy. Public opinion, foundational to democracy, also represents the social construction of the public[15] and influences social cohesion [16]. In turn, from the government's perspective, media, via hot searches, is the lens through which citizens collect political information about the performance of political institutions and adjust their attitudes accordingly [17]. This makes hot searches a vital tool for government to monitor social sentiment, enabling preemptive and responsive policy actions to foster social stability and progress. The impact of social media platforms is far-reaching and needs to be carefully considered by policy makers, educators and business leaders [18]. The broken circle effect brought by hot searches is also very valuable to explore, due to the different platform content preferences, the public can learn about different circles of information in different platforms, for example, in the same time period, combining tables 8 and 9, we can see that in October 2023, the most discussed event in ZHIHU was ‘Huawei launching mate60’, and the most discussed topic in TOPBOZZ was ‘Israeli-Palestinian conflict’, the difference in platforms allows the public with different preferences to learn about cultural exchanges in different fields, which is not only a microcosm of the cultural conflict, but also provides a data basis for cross-circle dialogue.

4.2 Analyzing the Development of the News and Media Industry

From the perspective of the news and media industry, the era of integrated media requires a shift from traditional news production as a discrete product to a dynamic, collaborative process involving journalists, audiences, platforms, and other stakeholders, etc.[12]. Newspaper publishers and journalists can respond to the digital media environment by optimizing writing and headlines for search engines and social media [19], and during an epidemic, for instance, media organizations can create and disseminate relevant content through keyword searches on digital platforms[20], with public reactions becoming a key element in news value [14].While new media, emerging with the rise of the Internet, may challenge traditional outlets, platform-driven news often prioritizes traffic over public interest, limiting its contribution to meaningful news coverage[21].

The arrival of smart media also has a synergistic effect on the use of traditional media[11], with trending topics increasing user engagement with popular news posts within each news outlet [22]. Journalists today are increasingly using social media data to infer and report on public opinion by citing social media posts, identifying trending topics and reporting on overall sentiment [13]. News for the ‘hot searches’ to provide material sources of information, and ‘hot searches’ in turn promote the dissemination and development of news, a news event can become a hot spot in a short period of time, from the perspective of the workers who publish the news, repeated publication of related events can enhance the publisher's or organization’s attention and engagement[22], while from the platform side, similar events will be frequently recommended to the public for browsing, increasing the heat of the event, making up for the weaknesses of traditional reporting such as insufficient participation of social forces and insufficient richness of perspectives. Analyzing news production through the lens of Large Language Models (LLMs) reveals both challenges and opportunities to media workers. LLMs can assist journalists in better understanding public preferences and refining content to match audience interests. While AI may complicate fact-checking and source reliability, its benefits are expected to outweigh the drawbacks [23]. By maintaining journalistic ethics, media professionals can use LLMs to optimize resource allocation, quickly identify trending topics, and engage more effectively with audiences, thus overcoming traditional media's limitations and adding significant value to the industry in the AI era.

4.3 Research Limitations and Future Expansion

This study primarily examines the classification of hot searches over time across different platforms but does not systematically capture the dynamic evolution of netizen sentiment. The strong link between sentiment and topic content makes it challenging to identify complex emotions, often leading to one-sided analyses. Therefore, developing sentiment lexicons tailored to specific types of crises is crucial for more accurate understanding and responses [24]. Hot searches on social media platforms are more inclined to topics that trigger emotional resonance and involve individual or group rights [9]. In the future, it is anticipated that Large Language Models (LLMs) will not only classify the emotional tone of comments but also enhance their ability to recognize emotions. By integrating time-series analysis, LLMs could track fluctuations in netizen sentiment, offering deeper insights into social emotions and enabling more timely, appropriate responses to emerging issues, ultimately fostering a harmonious and stable social environment.

5 Conclusion

This study takes the network hot search data of Mircoblog, ZHIHU and TOPBOZZ as the research samples from 2021 to 2025, classifies the hot search data and extracts the keywords with the help of Large Language Model (LLMM), conducts in-depth excavation of the public's attention of hot searches in different platforms at different times by analyzing the classification results, and illustrates the evolution of the hot searches by using the visualization analysis. The study finds that China's Internet hot searches are mainly classified into eight categories: international affairs, sports and entertainment, business and technology, medical and health, livelihood protection, traffic and travelling, environmental disasters, and education information, and that the pattern of the public's attention to each category of Internet hot searches under the chronological order can be a revelation for the construction of a harmonious society, and for the optimization of the direction of the work of the mass media to improve its efficiency. This study also has some limitations, such as the lack of a systematic capture of the dynamic evolution of public sentiment under hot searches, which limits further analysis and multi-angle refinement based on themes. In the subsequent research, we will try to combine multiple sentiment text classification techniques for further screening to optimize the response posture about the public, and further combine it with the Intelligent Large Language Model (LLM).

References

1. Center, C. I. N. I. (2025). The 55th statistical report on China's internet development. *Beijing2025*.
2. Auxier, B., & Anderson, M. (2021). Social media use in 2021. *Pew Research Center*, *1*(1), 1-4.
3. Garrote-Quintana, A., Urquía-Grande, E., & Delgado Jalón, M. L. (2025). Trending Topics in E-government, E-participation, and social media in the Crisis Era. *Administration & Society*, *57*(2), 254-280.
4. Wu, X. K. (2020). The underlying logic of hot search and the adjustment of social responsibility. *People's Tribune, *(28), 107–109.
5. Yu, Q., Weng, W., Zhang, K., Lei, K., & Xu, K. (2014). Hot topic analysis and content mining in social media. In *2014 IEEE 33rd International performance computing and communications conference (IPCCC)* (pp. 1-8). IEEE.
6. Huang, M. C., Chen, K., Xiang, S. B., Zhu, X. T., Xu, H., & Wang, G. Q. (2021). Mining and analysis of hot topics in social media. In *2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM)* (pp. 155-159). IEEE.
7. Shen, X., & Yang, K. L. (2024). Topic mining, evolution, and implications of online public opinion in emergencies based on Weibo hot search data. *Information Technology and Management Applications*,3(06), 98-112.doi: 10.20186/j.cnki.hustitama2022.2024.06.09.
8. Habermas, J. (1962). *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. MIT Press.
9. Pew Research Center's Project for Excellence in Journalism. (2010). *New media old media: How blogs and social media agendas relate and differ from the traditional press*. Pew Research Center.
10. Yu, X. P. (2023). "Hot Search Politics": Political Forms in the Digital Media Era. Ningxia Social Sciences, (6), 69–78.
11. Woo, J., Choi, J. Y., Shin, J., & Lee, J. (2014). The effect of new media on consumer media usage: An empirical study in South Korea. *Technological Forecasting and Social Change*, *89*, 3-11.
12. Sue Robinson. (2011). "Journalism as Process": The Organizational Implications of Participatory Online News. Journalism & Mass Communication Monographs,13(3),137-210.

13. University of Ottawa, Ottawa, Ontario, Canada, Ryerson University, Toronto, Ontario, Canada & Ryerson University, Toronto, Ontario, Canada. (2020). Journalists' Use of Social Media to Infer Public Opinion: The Citizens' Perspective. Social Science Computer Review,38(1),57-74.
14. Fürst Silke & Oehmer Franziska. (2021). Attention for Attention Hotspots: Exploring the Newsworthiness of Public Response in the Metric Society. Journalism Studies,22(6),799-819.
15. The University of Utah, USA. (2019). Social media as public opinion: How journalists use social media to represent public opinion. Journalism,20(8),1070-1086.
16. González-Bailón, S., & Lelkes, Y. (2023). Do social media undermine social cohesion? A critical review. *Social Issues and Policy Review*, *17*(1), 155-180.
17. Ceron, A. (2015). Internet, news, and political trust: The difference between social media and online media outlets. *Journal of computer-mediated communication*, *20*(5), 487-503.
18. Rachmad, Y. E. (2023). Social Media Impact Theory.
19. Athey, S., & Mobius, M. (2012). The impact of news aggregators on internet news consumption: The case of localization.
20. Showkat, N., & Gull, M. (2024). The study of Google search trends for an effective communication during COVID-19 pandemic. *Authorea Preprints*.
21. Yin, Q., Fu, Z., & Zheng, S. (2024). Meso news-space in China: peripheral news production of platform journalism. *Digital Journalism*, *12*(5), 680-699.
22. Yang, T., & Peng, Y. (2022). The importance of trending topics in the gatekeeping of social media news engagement: a natural experiment on Weibo. *Communication Research*, *49*(7), 994-1015.
23. Dijkstra, A. M., de Jong, A., & Boscolo, M. (2024). Quality of science journalism in the age of Artificial Intelligence explored with a mixed methodology. *Plos one*, *19*(6), e0303367.
24. J. Dai et al. (2024).Sentiment-topic dynamic collaborative analysis-based public opinion mapping in aviation disaster management: A case study of the MU5735 air crash.Int. J. Disaster Risk Reduct.