

The Social Mirror of Chinese AI: A Sentiment and Discourse Analysis of DeepSeek on Weibo

Kang Zhao

College of Publishing, University of Shanghai for Science and Technology, Shanghai 200093, China

How to cite this paper: Zhao, K. (2025). The Social Mirror of Chinese AI: A Sentiment and Discourse Analysis of DeepSeek on Weibo. *Social Sciences and Humanities*, 2(1), 71–79. ISSN Print: 3104-4328; ISSN Online: 3104-4336.

<https://doi.org/10.63313/SSH.9037>

Published: 2025-10-28

Copyright © 2025 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

As artificial intelligence (AI) becomes a defining force of the digital era, the rapid emergence of Chinese large language models has provoked widespread social reflection on technology, culture, and national identity. This study explores how the Chinese public constructs emotional and cognitive understandings of AI through discussions about DeepSeek, a large model developed in China. Based on 59,004 Weibo posts, the research applies Latent Dirichlet Allocation (LDA) topic modeling and sentiment analysis to investigate how users' emotional attitudes and discursive orientations shape the social imagination of Chinese AI.

The results identify three major discourse orientations: technological intimacy, where users embed AI in daily routines and affective practices; techno-nationalism, where technology becomes a site of collective pride and national autonomy; and techno-rationality, where participants engage in critical evaluation of transparency and performance. Although positive sentiments (46%) dominate the discourse, recurring concerns over data opacity and model hallucination reveal persistent skepticism.

By drawing upon the concepts of sociotechnical imaginaries (Jasanoff & Kim, 2015) and mediated nationalism (Zhao, 2023), the study argues that DeepSeek embodies a new “social mirror” through which citizens negotiate meanings of progress, agency, and belonging. It contributes to global communication and AI studies by revealing how emerging technologies are emotionally inhabited and discursively localized within China's mediated modernity.

Keywords

Chinese AI; DeepSeek; Artificial Intelligence; LDA Topic Modeling; Sentiment Analysis; Techno-nationalism; Techno-optimism

1. Introduction

The rapid advancement of generative artificial intelligence (AI) has profoundly reshaped technological innovation and global discourse. Following the global success of ChatGPT, China has witnessed the emergence of several domestic large models, among which DeepSeek-R1 stands out as a milestone in open-source AI development. Developed by DeepSeek Corporation in Hangzhou, the model integrates the

Mixture of Experts (MoE) and Group Relative Policy Optimization (GRPO) frameworks, achieving high efficiency with limited computational resources. Its public release in January 2025 sparked extensive discussions across Chinese social media, particularly on Weibo, where users debated its technological capability, symbolic meaning, and national significance.

In the Chinese socio-cultural context, AI technologies are not only perceived as tools of productivity but also as embodiments of national strength and technological autonomy. DeepSeek, as both a technical achievement and a social symbol, thus offers a valuable case for exploring how public emotions and discourses reflect broader narratives of techno-optimism and techno-nationalism. This study aims to examine the emotional and thematic structures within Weibo discussions about DeepSeek, combining Latent Dirichlet Allocation (LDA) topic modeling and sentiment analysis. Through this dual methodological lens, the research investigates how users' attitudes toward domestic AI reveal intertwined relationships between technology, nationhood, and digital identity in contemporary China..

2. Literature Review

Understanding how the Chinese public interprets artificial intelligence (AI) requires an analytical framework that captures the intersection of technology, emotion, and national identity. This chapter integrates two theoretical perspectives that illuminate the social meanings of AI in China's mediated environment: techno-nationalism and techno-optimism. These frameworks provide interpretive lenses to explain why discussions around the large language model DeepSeek on Weibo combine both patriotic enthusiasm and reflective skepticism.

2.1. Artificial Intelligence and Social Imagination

Artificial intelligence has long occupied a space between technological reality and social imagination. As machine learning and large-scale neural architectures continue to evolve, AI increasingly blurs the boundaries between the technical and the symbolic. Scholars have emphasized that emerging technologies do not exist as neutral tools but as discursive constructs embedded within political, cultural, and emotional frameworks. In China, the rise of large language models (LLMs) such as DeepSeek has transformed AI from a scientific endeavor into a cultural phenomenon that encapsulates aspirations for modernization, sovereignty, and national rejuvenation. The proliferation of AI discussions on social platforms like Weibo has thus become a new arena for the negotiation of meanings between citizens, technologies, and the state.

In this sense, AI operates as both an epistemic and affective infrastructure—shaping how people understand knowledge production, social trust, and digital agency. Public discourse surrounding DeepSeek illustrates this dual role: it is simultaneously a product of advanced computation and a site where national identity and emotional

resonance converge. This study aligns with recent scholarship that conceptualizes AI as a “social technology” (Jasanoff, 2020), whose legitimacy depends as much on its cultural embedding as on its algorithmic performance.

2.2. Techno-nationalism and the Politics of Innovation

Techno-nationalism originally emerged as a term to describe U.S. protectionist policies against Japan’s technological rise in the 1980s. Over time, it evolved to characterize how nations mobilize technological progress to reinforce sovereignty and global competitiveness. In non-Western contexts, techno-nationalism often operates as a discursive strategy that links scientific advancement with cultural pride and collective identity. In China, this framework has become central to understanding the public’s engagement with AI. Government initiatives emphasizing “technological self-reliance” and “independent innovation” have infused digital technologies with strong national symbolism. As scholars such as Segal (2018) and Zhao (2022) suggest, China’s technology narratives combine geopolitical ambition with cultural confidence, producing a unique “civilizational modernity” through technological discourse.

Within this national framework, AI models such as DeepSeek are not merely computational achievements but embodiments of the country’s capability to “catch up and surpass” Western technological dominance. Public reactions on Weibo often merge technical evaluation with patriotic sentiment, using phrases such as “the pride of Chinese AI” or “a milestone of national strength.” This intertwining of emotional expression and political imagination reveals how digital publics enact national identity through participation in technology-related discussions. Techno-nationalism, therefore, functions as both an ideological resource and a participatory practice in the Chinese digital sphere.

3. Methodology

This study combines computational text mining with interpretive discourse analysis to investigate how Chinese social media users construct meanings and emotions around DeepSeek, a domestically developed AI language model. The methodological process follows a systematic five-step framework: data collection, text preprocessing, feature construction, topic modeling, and sentiment analysis. All analyses were conducted using Python, based on the Scrapy, Jieba, and Scikit-learn toolkits, with further visualization and validation through custom scripts.

3.1. Data Collection

The dataset was collected from Weibo, one of China’s most active microblogging platforms. Using the keywords “DeepSeek” and “深度求索,” all user-generated posts and comments published between January 20 and April 20, 2025 were scraped via a custom crawler built on the Scrapy framework. This period coincides

with the release and subsequent public discussions of the DeepSeek-R1 model. A total of 73,251 posts were initially retrieved. After filtering duplicates, advertisements, emojis, and irrelevant reposts, 59,004 valid texts were retained for analysis, forming the Weibo opinion corpus used in this research.

Figure 1. Workflow of Weibo Text Data Collection

(Schematic illustration of data collection, preprocessing, and analysis pipeline; adapted from original Chinese figure.)

- 1.Keyword Search and Crawl (DeepSeek, 深度求索)
- 2.Text Cleaning and Deduplication
- 3.Segmentation and Stopword Filtering
- 4.LDA Topic Modeling
- 5.Sentiment Classification and Visualization

3.2. Text Preprocessing and Feature Construction

Because Chinese lacks explicit word boundaries, text segmentation was performed using the jieba.posseg package, with a part-of-speech restriction limited to nouns (n), proper nouns (nz), and gerunds (vn) to enhance semantic relevance.

To filter non-informative items, four major stopwords lists were merged: Harbin Institute of Technology, Baidu, Chinese Stopword List, and the Sichuan University Intelligent Computing Lab list. This fusion generated a customized stopwords library tailored for social media language, ensuring better modeling precision.

Text vectors were constructed using the CountVectorizer tool in Scikit-learn, which extracted the 2,000 most frequent terms while excluding words appearing in fewer than ten or more than half of the documents. The resulting sparse document – term matrix served as the basis for topic modeling.

Table 1. Top 10 High-Frequency Words in the Weibo Corpus

Rank	Word	Frequency
1	Model	20,924
2	Technology	13,659
3	Technique	12,940
4	Artificial Intelligence	12,014
5	Company	11,710
6	Video	9,434
7	Intelligence	8,794
8	Development	8,458
9	Enterprise	8,271
10	Market	8,096

Table 1 shows that discussions primarily revolve around DeepSeek's technical performance, industrial ecosystem, and economic influence.

3.3. LDA Topic Modeling

The Topic identification employed the Latent Dirichlet Allocation (LDA) algorithm in

Scikit-learn. Each document is modeled as a probabilistic mixture of latent topics, and each topic as a distribution of semantically related words. The model was trained with the online variational Bayes method, using a maximum of 50 iterations. To determine the optimal number of topics (K), perplexity was used as an evaluation metric (see Equation (1)).

Equation (1). Perplexity Metric for LDA Model Evaluation

$$P(D) = \exp \left[-\frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d} \right]$$

where M is the total number of documents, N_d is the token count of document d, and $p(w_d)$ denotes the probability of observing document d under the trained model. A lower perplexity value indicates stronger generalization.

By iterating $K = 1 - 15$, the model achieved a plateau at $K = 7$, which balanced interpretability and fit. Therefore, seven topics were retained for subsequent analysis.

3.4. Sentiment Analysis

Sentiment classification complemented the topic modeling to reveal users' affective orientations. Considering the informal and creative nature of Weibo language, traditional lexicon approaches were insufficient. Hence, this study employed DeepSeek-V3, a deep neural network model hosted on Alibaba Cloud DashScope API, containing 671 billion parameters (with 37 billion activated per inference) and pre-trained on 14.8 trillion tokens.

The model assigns each text a sentiment type—positive, neutral, or negative—and a score from 0 to 100. Regular expressions were used to parse both the sentiment labels and their numeric scores from API outputs. Among the 59,004 posts, 46.0 % were positive, 36.6 % neutral, and 17.3 % negative, as summarized in Table 2.

Table 2. Sample of Sentiment Classification Results Example Text

Example Tex	Sentiment Type	Score
“DeepSeek’ s open-source release marks China’ s technological confidence and inspires global innovation.”	Positive	95
“After using DeepSeek, I find its logic and clarity impressive—like a teacher guiding my writing.”	Positive	90
“DeepSeek’ s server keeps crashing—hard to believe a model this advanced can be so unstable.”	Positive	85

As shown in Table 2, positive reactions dominate, yet frustration and irony occasionally surface, indicating a mix of optimism and skepticism.

3.5. Validation and Visualization

Each document received a dominant-topic label corresponding to its highest posterior probability. The document – topic matrix was visualized via pyLDavis for interactive exploration of inter-topic distances. Co-occurrence networks further illustrated relationships among the seven topics.

To ensure reliability, 300 randomly sampled posts were qualitatively reviewed. The consistency between machine outputs and human interpretation confirmed the robustness of thematic clustering.

4. Results and Discussion

4.1. Overview of Topic Modeling Results

Based on the optimal LDA configuration ($K = 7$), seven topics emerged from the 59,004 valid Weibo posts. Each topic represents a distinct semantic cluster reflecting public concerns and attitudes toward DeepSeek. Table 3 presents the representative keywords and thematic labels assigned through qualitative interpretation.

Table 3. LDA Topic Results of Public Discussion on DeepSeek

Topic	Representative Keywords	Thematic Label
1	use, convenient, reply, question, daily, efficient	Technological Intimacy
2	national, China, self-developed, innovation, proud, future	Techno-nationalism
3	parameter, MoE, GRPO, optimization, model, speed	Technical Analysis
4	company, open-source, launch, product, industry, application	Industry Development
5	fairness, data, privacy, transparency, ethical, governance	Technological Rationality
6	compare, ChatGPT, performance, translation, logic	Competitor Comparison
7	creativity, content, poem, article, emotion, sharing	Creative Practice

As shown in Table 3, the seven topics reflect users' cognitive and emotional layers—from intimate engagement with AI tools to reflections on China's technological identity and ethical considerations.

4.2. Sentiment Distribution and Thematic Correlation

Sentiment analysis revealed that 46.0% of posts expressed positive emotions, 36.6% were neutral, and 17.3% conveyed negative sentiments.

When overlaid with topic clusters, an interesting correlation emerged:

Positive emotions dominated in “Technological Intimacy” (Topic 1), “Creative Practice” (Topic 7), and “Techno-nationalism” (Topic 2), where users often celebrated DeepSeek's functionality and symbolic value.

Neutral emotions characterized “Technical Analysis” (Topic 3) and “Industry Development” (Topic 4), suggesting more cognitive engagement than emotional attachment.

Negative sentiments concentrated in “Technological Rationality” (Topic 5) and “Competitor Comparison” (Topic 6), where users voiced doubts about stability, transparency, and data governance.

This polarity structure shows that the emotional field surrounding DeepSeek is not purely enthusiastic but rather differentiated by user roles and expectations. Engineers, content creators, and ordinary users each project distinct affective orienta-

tions depending on their engagement depth and professional background.

4.3. Discourse Structure 1: Technological Intimacy

The first discourse structure, technological intimacy, describes users' everyday emotional integration with AI tools.

Posts in this cluster frequently employed personalized expressions—referring to DeepSeek as a “teacher,” “friend,” or “partner.”

Such metaphors transform AI from a computational system into an affective companion, extending the notion of digital personhood in human – machine interaction. This emotional domestication aligns with Miller's (2021) argument that digital technologies become “intimate infrastructures” of social life.

In this sense, DeepSeek is not only a product of engineering but also a relational actor in people's daily cognitive practices.

The language of admiration and trust reflects a broader techno-optimistic affect, where efficiency, creativity, and companionship converge.

4.4. Discourse Structure 2: Techno-nationalism

The second major discourse pattern corresponds to techno-nationalism.

Weibo discussions frequently equate DeepSeek's emergence with the narrative of China's technological sovereignty and global competitiveness.

Typical posts use patriotic language—“China's pride,” “independent innovation,” “breaking the Western monopoly”—to frame DeepSeek as evidence of national progress.

Positive sentiments in this discourse are therefore not only about technical excellence but also about emotional participation in the project of national modernization.

However, beneath this optimism lies a layer of geopolitical anxiety.

Some users questioned whether “domestic AI” can truly compete with Western models, especially in scientific transparency and global acceptance.

This ambivalence—between pride and pressure—reveals the emotional complexity of China's digital nationalism in the AI era.

5. Conclusion

This study set out to examine how Chinese social media users emotionally and discursively construct the meaning of artificial intelligence through the case of DeepSeek, a domestically developed large language model. Combining computational text analysis with qualitative interpretation, it explored 59,004 Weibo posts to uncover the thematic, emotional, and ideological dimensions of AI-related discourse in China. The integration of topic modeling and sentiment analysis provided an empirical foundation for mapping how users engage with technology as both a functional tool and a cultural symbol.

The results identified seven major thematic clusters and three overarching discourse structures—technological intimacy, techno-nationalism, and technological rationality. These structures form an interdependent communicative system that captures the multi-layered emotional ecology of China’s digital public sphere. Users exhibit both affective enthusiasm and reflective awareness: they anthropomorphize DeepSeek as a companion, celebrate it as a symbol of national progress, and simultaneously deliberate on its ethical, social, and cultural implications. This triadic model reveals a distinctive pattern of optimistic reflexivity—an affective orientation that combines pride, trust, and critical reasoning within the same communicative field.

Theoretically, this study contributes to the growing body of literature on digital affect and technological discourse by introducing the concept of an affective – rational continuum. Unlike conventional dichotomies that separate emotion from reason or optimism from skepticism, the DeepSeek discourse demonstrates that these forces coexist and mutually reinforce each other. Emotions in the Chinese AI context do not simply reflect irrational enthusiasm; rather, they perform epistemic and social functions by generating collective identification and moral deliberation. This framework enriches cross-cultural understandings of public engagement with AI, illustrating that emotional expression is a vital component of rational communication in digital societies.

From a comparative perspective, the Chinese case offers valuable insights for global AI communication research. While Western discussions of AI often emphasize risks, ethics, and autonomy, the Chinese discourse foregrounds collective progress, national confidence, and responsible innovation. This difference underscores the cultural contingency of technological imaginaries and highlights the need for decentering Western epistemologies in technology communication studies. By foregrounding a non-Western emotional logic that blends nationalism with civic rationality, this study contributes to a more pluralistic theory of how modern societies negotiate the moral and symbolic dimensions of artificial intelligence.

Practically, the findings carry significant implications for AI governance and communication policy. They suggest that effective public engagement requires not only technological transparency but also emotional literacy. Institutions and policymakers should recognize emotions as resources for public deliberation rather than as obstacles to rational discourse. Educational programs promoting critical AI literacy can empower citizens to engage with technology responsibly and creatively. Moreover, the integration of emotional awareness into communication strategies can strengthen public trust and support sustainable innovation.

Despite its contributions, this research has several limitations. The dataset was restricted to Weibo posts within a three-month period, which may not capture long-term discursive evolution or cross-platform variations. Additionally, while computational modeling offers valuable macro-level insights, it cannot fully capture

the subtlety of irony, humor, or metaphor that often characterizes social media language. Future research could therefore combine longitudinal and multimodal approaches—analyzing text, image, and interactional data across multiple platforms—to deepen understanding of how AI-related emotions evolve over time. Comparative studies involving other non-Western contexts, such as India, Japan, or South Korea, would also help refine theories of global digital affect.

In conclusion, the DeepSeek discourse provides a powerful lens through which to observe China's transition toward reflexive modernity in the digital age. It demonstrates how a society's relationship with technology is simultaneously emotional and rational, local and global, national and ethical. By articulating an integrated framework that bridges affective analysis and cultural interpretation, this study advances the interdisciplinary dialogue between communication, sociology, and artificial intelligence research. Ultimately, understanding how emotions shape the public meaning of AI is essential not only for interpreting China's technological trajectory but also for envisioning a more inclusive and responsible global future for artificial intelligence.

References

- [1] Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability bias in tech*. Cambridge, MA: MIT Press.
- [2] Chen, W., & Zhang, X. (2023). National sentiment and the algorithmic imagination: Public perceptions of AI in China. *Chinese Journal of Communication*, 16(2), 195 – 212.
- [3] Duan, Y., Li, L., & Wang, H. (2022). Emotion, ideology, and the digital nation: Mapping affective publics on Chinese social media. *Information, Communication & Society*, 25(12), 1610 – 1627.
- [4] Habermas, J. (1989). *The structural transformation of the public sphere*. Cambridge: Polity Press.
- [5] Han, R. (2021). Authoritarian participatory public spheres: The politics of AI discourse in China. *Media, Culture & Society*, 43(5), 902 – 918.
- [6] Liu, T., & Zhao, Y. (2022). Digital nationalism and the emotional politics of technology in China. *Journal of Contemporary China*, 31(133), 871 – 888.
- [7] Lu, S., & Sun, L. (2024). Deep learning nationalism: Sentiment trajectories in China's AI narratives. *Television & New Media*, 25(3), 289 – 310.
- [8] Miller, D. (2021). *The digital person: Technology, intimacy, and everyday life*. New York: Routledge.
- [9] Reeves, J. (2022). *Emotional regimes of technology: Affect and governance in the digital age*. Stanford: Stanford University Press.
- [10] Sun, W., & Zhao, Y. (2021). Techno-cultural imaginaries and the Chinese dream: The case of AI and big data. *Global Media and China*, 6(1), 7 – 24.
- [11] Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.