Available Online: 16 April 2026 DOI: 10.54254/2977-3903/2026.32766

An enhanced BERT-LDA hybrid model for sentiment analysis in AI tool-related academic disputes

Xixi Zhou[1*], Yanjing Qiu[1], Wei Jiang[1]

[1]Shanghai International Studies University, Shanghai, China

*Corresponding Author. Email: zhouxixi_0213@163.com

Abstract. The rapid adoption of Generative Artificial Intelligence (AI) in academia has intensified debates on AI-assisted writing, the reliability of plagiarism detection, and broader concerns over academic integrity. This study examines public opinion on AI-related academic controversies across three major Chinese social media platforms (Weibo, Zhihu, and Xiaohongshu) using data collected from March 2016 to March 2025. An improved BERT-LDA hybrid model is employed to identify topics and analyze sentiment dynamics. The results reveal three primary thematic dimensions: positive perceptions of technology-enabled academic and educational innovation, slightly negative concerns regarding academic norms and integrity systems, and policy-oriented discussions on education management. Temporally, sentiment toward AI-related academic misconduct fluctuates, peaking in early 2016-2017, declining through 2020, and exhibiting a downward trend after 2023. Spatial analysis demonstrates significant regional heterogeneity, with neutral sentiment prevailing in economically developed regions and slightly negative sentiment concentrated in central and western areas, alongside observable spatial clustering effects. Furthermore, three evolutionary patterns of public discourse are identified: steadily improving sentiment on ethical governance, cyclical debates on automated writing misuse, and divergent sentiment on issues such as copyright and model monopolization. These findings highlight the combined influence of technological, policy, and socio-economic factors on the evolution of public opinion.

Keywords: AI academic controversy, BERT-LDA model, social media public opinion, academic integrity

1. Introduction

1.1. Literature review

With the popularization of generative AI tools (such as ChatGPT, DeepSeek and so on) in the academic field, issues like "AI academic misconduct", "AI academic controversy", and "AI dependence" have become core controversial topics in the education and technology sectors. Nevertheless, existing topic modeling studies have not yet formed a systematic analysis framework for this emerging interdisciplinary field. On one hand, most existing studies focus on traditional fields such as climate, health, and consumption (e.g., Lin and Nkhata applied BERT-LDA to sentiment classification of Chinese movie reviews [1,2]) and have not touched upon the topic of AI tool academic controversy, which is both timely and socially valuable. On the other hand, although Chkarka et al. discussed the maintenance of academic integrity among master's students and teachers in the AI

era, pointing out that the abuse of AI may lead to risks such as plagiarism and academic misconduct, their research only conducted qualitative analysis based on interview data, failed to integrate quantitative features of social media public opinion, and did not introduce an analytical logic of "topic-sentiment-spatiotemporal" linkage [3].

Specifically, academic misconduct has long been a concern in various disciplines, with existing research focusing on its forms, prevention, and governance. For example, Li and Kang explored the mining and prevention of implicit academic misconduct in the entire process of academic journal publishing, emphasizing the need for data-driven detection methods [4]; Zou discussed the governance of postgraduate academic misconduct empowered by big data, highlighting its value implications and practical obstacles [5]; Li et al. investigated academic misconduct in medical papers and proposed countermeasures, pointing out the uniqueness of medical research ethics [6]; Feng et al. examined the cognitive differences between authors and editors regarding academic misconduct in academic publishing and suggested prevention measures based on stakeholder perspectives [7]; Zhong analyzed implicit academic misconduct and prevention strategies, calling for improved oversight mechanisms [8]; Yi applied blockchain to prevent data academic misconduct in medical papers, demonstrating the potential of technology in safeguarding integrity [9]; and Long discussed the logic, mechanism, and path of accountability for academic misconduct, critiquing the legitimacy of university governance [10]. These studies collectively underscore the ongoing efforts to address academic integrity issues in traditional contexts, yet they primarily rely on qualitative or policy-oriented approaches without leveraging advanced text mining techniques like topic modeling for large-scale data analysis.

As a core data-driven text mining technology, Topic Modeling has formed a mature application paradigm in multiple disciplines and become a key means to reveal potential semantic connections in massive unstructured texts. Existing studies have verified its effectiveness in various fields. In the field of social media analysis, Dahal et al. conducted topic mining and sentiment analysis on Twitter data related to global climate change using the LDA model. By comparing temporal differences in public opinion across countries through geotagging features, they revealed that the United States paid less attention to climate policy discussions than other countries [11]; In the field of health science, Xiang et al. analyzed the information needs of users in online health communities using the BERT-LDA model, identified 9 core topics including etiology, diagnosis, and treatment, and correlated the distribution of positive and negative sentiments to explore the relationship between health needs and emotional expressions [12]; In the field of finance, Zhou et al. embedded BERT-LDA into financial news analysis [13]. By jointly embedding contextual semantics and thematic narratives, they addressed the issues of insufficient semantic information in traditional LDA and feature sparsity in short texts, generating topic words with higher discriminability; In the field of geography, Chen et al. took "micro-wetlands" as the research object and optimized the topic extraction accuracy of geographical concepts through the BERT-LDA integrated model, overcoming the problem of knowledge integration caused by terminology heterogeneity [14].

Based on these, this study focuses on the specific topic of social media public opinion regarding AI tool academic controversy, using text data from three platforms (Weibo, Zhihu, and Xiaohongshu) from March 2016 to October 2025 as the research object (covering multi-source heterogeneous data types such as short texts and long texts). The core research question is: How to realize the multi-dimensional integrated analysis of "topic identification-sentiment quantification-temporal evolution-spatial heterogeneity" for public opinion on AI tool academic controversy through an improved BERT-LDA model, thereby revealing the triple driving mechanism of technology-policy-geography behind the public opinion.

1.2. Methodology

The sample selection for this study adheres to the following criteria: the release time of the collected social media data ranges from March 1, 2016 to October 31, 2025; the selected content must contain explicit text (excluding content that only includes images or videos without accompanying text descriptions) ; each sample should match geographic tags or IP location information to support subsequent spatial analysis; and the text length of each sample is restricted to 50–500 characters. Specifically, a total of 3, 200 pieces of raw data were collected using the Python Scrapy crawler framework; after undergoing data cleaning processes, including deduplication, removal of meaningless text, and supplementation of missing tags, 1,890 valid samples were retained, which ensures a balanced distribution of samples across different social media platforms and time periods. The overview of the research is shown in Figure 1. Additionally, the sentiment scores of the valid samples are artificially standardized to a range of 0 to 5, where 0 indicates an extremely negative sentiment and 5 indicates an extremely positive sentiment.

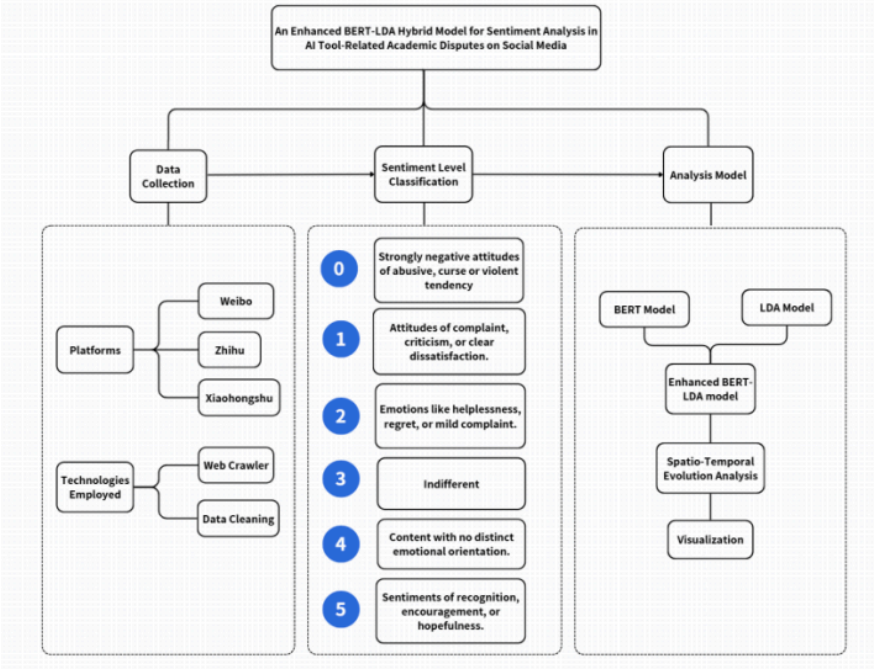


Figure 1. Overview of the research

2. Architecture

2.1. Fine-tuning BERT model

BERT (Bidirectional Encoder Representations from Transformers) is a language representation model built on the Transformer architecture. It adopts the WordPiece method for data preprocessing and undergoes pre-training through two core tasks: Masked Language Model (MLM) and Next Sentence Prediction (NSP). Its overall structure is illustrated in Figure 2 [15].

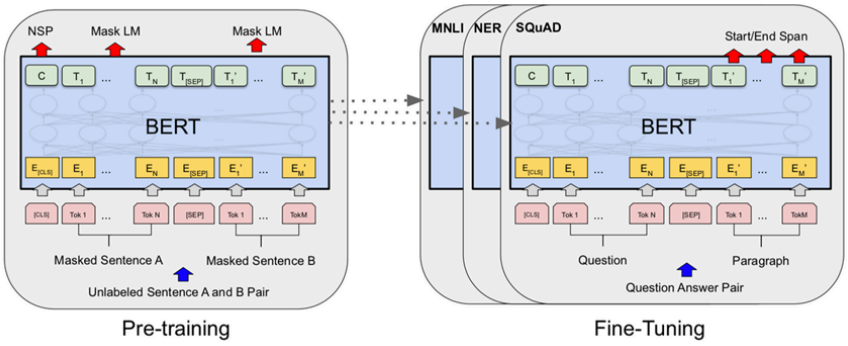


Figure 2. Overall architecture of BERT model

This study retains the basic BERT framework, improves the pre-training tasks, and adds deep pre-training tasks. Furthermore, high-quality text features obtained from the data preprocessing stage are integrated into both the pre-training and fine-tuning processes of BERT. This enables the model to simultaneously learn syntactic, semantic, and classification label-related text features when performing classification tasks.

2.1.1. Transformer architecture

Proposed by Vaswani et al., the Transformer relies on the attention mechanism, which outperforms RNNs in avoiding the forgetting of early words in long sentences (by assigning higher weights to key words) and supports parallel computing to accelerate processing. Its structure (Figure 3) involves converting input into word embeddings, combining with positional encoding, and inputting into multi-layer encoders/decoders. Each encoder layer consists of a multi-head self-attention layer (capturing global sequence information) and a feed-forward neural network (FFNN) layer (implementing non-linear transformations).

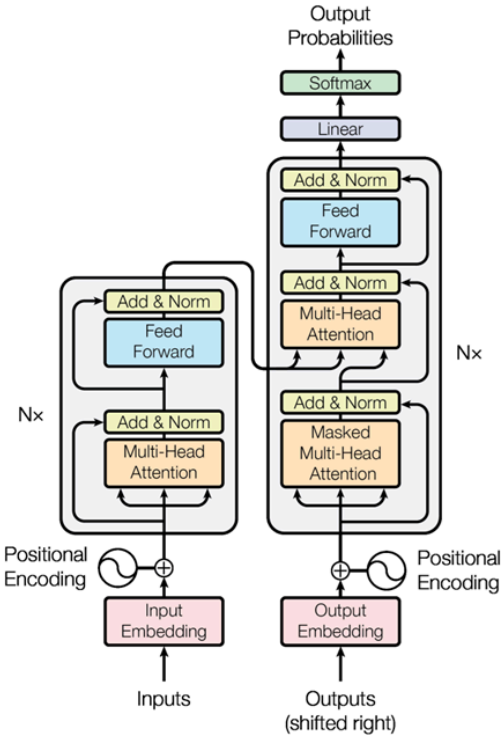


Figure 3. Schematic diagram of the Transformer architecture

2.1.2. MLM (Masked Language Model)

MLM is BERT's pre-training task for large-scale unlabeled text: 15% of input words are randomly masked, and the model predicts them based on context. This bidirectional design overcomes the limitations of unidirectional language models, enhanced by the Transformer encoder's self-attention mechanism. Masked input is processed by multi-layer encoders to generate hidden representations, and masked words are predicted via a classification layer, enabling the model to capture long-range dependencies and rich semantic features.

2.1.3. NSP (Next Sentence Prediction)

Next Sentence Prediction (NSP) is a key learning objective introduced in the BERT model, designed to handle tasks that require judging the relationship between sentence pairs, such as paraphrase detection, natural language inference, and discourse coherence assessment.

BERT uses [CLS] (start of sentence pairs) and [SEP] (separator) tokens for NSP training. Segment embeddings, combined with word and positional embeddings, help distinguish the two sentences. The [CLS] token's output vector from the last layer is used for binary classification (via learned weights), and cross-entropy is adopted to calculate NSP loss.

During training, the output vector of the last layer corresponding to the [CLS] token represents the NSP prediction result. Similar to MLM, a set of learned classification weights $W_{NSP} \in R^{2 \times d_h}$ is used to perform binary classification prediction based on the original [CLS] vector. Finally, cross-entropy is used to calculate the NSP loss for each sentence pair:

$$p = Softmax(W[CLS] + b) \tag{1}$$

2.2. LDA model optimization (for topic identification)

Latent Dirichlet Allocation is a generative topic modeling framework based on probabilistic graphical models, which uses a hierarchical Bayesian structure with latent variables for unsupervised inference of latent semantic structures in a text corpus. Figure 4 presents a schematic diagram of the model's framework structure.

The corpus contains D documents, each with N_c words, sharing K pre-specified latent topics, where each topic is a probability distribution over the vocabulary of size V. Two key hyperparameters control the model: α , a K-dimensional vector that controls the sparsity of the topic distribution in a single document. The smaller the value of, the more inclined a document is to focus on a small number of topics. Topic-word prior parameter β : a V-dimensional vector that adjusts the concentration degree of the word distribution within a topic. Reducing the value of β can make a topic focus more on a small number of core words.

For each document, a topic distribution θ is drawn from Dirichlet(α), and for each word in the document, a topic assignment z is drawn from Multinomial(θ). For each topic K, a word distribution φ_K is drawn from Dirichlet(β). Finally, each observed word w is generated from Multinomial(φ_z).

Thus, LDA models documents as mixtures of topics and topics as mixtures of words, with Dirichlet priors inducing sparsity.

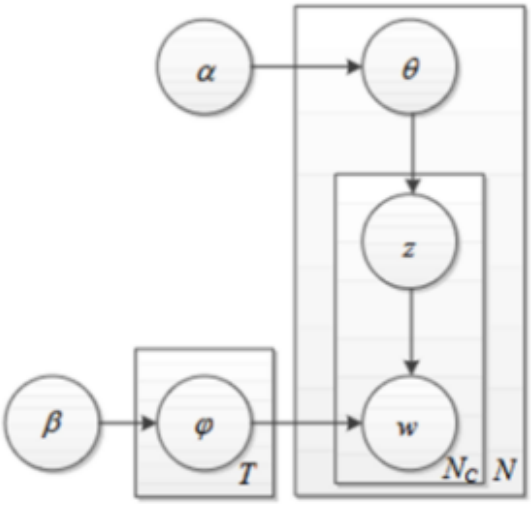


Figure 4. Architecture of LDA model

3. Empirical analysis of social media public opinion

3.1. BERT analysis

3.1.1. Pre-training of the BERT model

When processing text data related to controversies over AI academic tools, traditional BERT models and other existing libraries lack the ability to conduct sentiment analysis for this specific topic. Therefore, customized development and optimization were performed on the pre-trained BERT model to accurately capture emotional information in such texts, as shown in Figure 5.

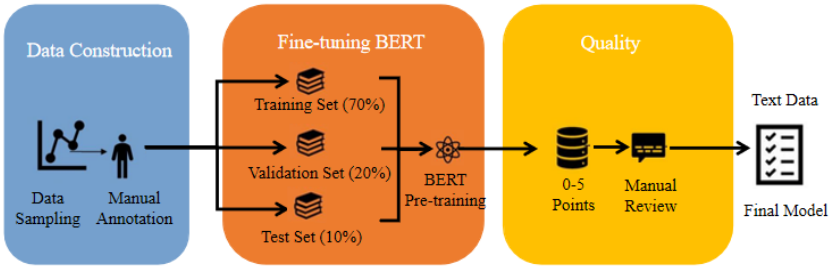


Figure 5. Pre-training of the BERT model

Step 1: Manual Sentiment Annotation

Since sentiment analysis models require supervised learning training, and existing models struggle to accurately capture sentiment intensity in complex Chinese contexts, manual sentiment annotation was first conducted. Texts were scored on a 0-5 scale with the following criteria: ①0 points: Strongly Negative (abuse/curse/violent tendencies) ; ②1 point: Obviously Negative (complaint/criticism/dissatisfaction) ; ③2 points: Slightly Negative (helplessness/regret/mild complaint) ; ④3 points: Neutral (no obvious emotional tendency) ; ⑤4 points: Slightly Positive (recognition/encouragement/hope) ; ⑥5 points: Obviously Positive (praise/joy/strong support). A subset of texts was manually annotated, with the average score of annotations

taken as the final label. The samples were then split into training, validation, and test sets at a ratio of 7: 2: 1, resulting in a fine-tuned BERT model.

Step 2: Environment Setup and Parameter Calibration

To ensure the stability and traceability of model training, parameters such as model paths, data paths, and output paths were configured, and a logging system was initialized. Finally, the tokenizer and a custom multi-task model were loaded from the pre-trained model.

Step 3: Multi-Task Model Architecture Design

Given the complexity of texts on AI academic tool controversies, a custom multi-task model BertForMultiTask was designed, inheriting from BertPreTrainedModel. This model integrates the BERT base model, BertForPreTraining module, and a classification head. The approach fuses the computational logic of Masked Language Model (MLM), Next Sentence Prediction (NSP), and classification tasks, calculating the loss for each task and summing them to obtain the total loss. This multi-task learning architecture enables the model to both learn general linguistic representations of text and effectively acquire topic-specific classification information, enhancing its generalization ability and performance on the target topic.

Step 4: In-Depth Data Preprocessing and Efficient Loading

Texts were cleaned to remove special characters and standardize formats, with titles and main bodies merged. After data loading, necessary columns were validated, texts were merged, cleaned, and invalid data was filtered out.

In the data splitting phase, the dataset was divided into training and validation sets, and class weights were calculated to address data imbalance. To facilitate model training, a SocialMediaDataset class inheriting from Dataset was created to generate data samples containing input IDs, attention masks, MLM labels, NSP labels, and classification labels. Finally, DataLoader was used to encapsulate the dataset into an iterable data loader for efficient data loading.

Step 5: Model Training

The model was trained using the AdamW optimizer and the get_linear_schedule_with_warmup learning rate scheduler for adaptive learning rate adjustment. In each training epoch, the model operated in training mode, performing forward propagation on each batch of data to compute losses, followed by backward propagation to update model parameters.

After each training epoch, the model was switched to evaluation mode, with losses and accuracy calculated on the validation set. The best validation accuracy was recorded, and corresponding model parameters were saved. Through continuous adjustment of training parameters and optimization of model structure, the model was finely tuned to improve its performance in sentiment analysis of AI academic tool controversy texts, avoiding overfitting and underfitting.

Step 6: Model Prediction Results

After training, the saved best model and tokenizer were loaded, and the original data was re-preprocessed. The model made predictions on the data in evaluation mode. Subsequently, the text data with machine-generated scores served as the foundation for the BERT-LDA model, facilitating the mining of thematic structures and emotional tendencies in AI academic tool controversy texts.

3.1.2. Temporal analysis of sentiment polarity based

Spearman's rank correlation coefficient was used to explore the monotonic correlation between machine-generated sentiment scores (0-5 points, ordinal categorical variable) and time. The research data included machine sentiment scores across multiple time points, with time variables containing date information at different granularities (some with specific dates, others precise to hours and minutes).

The calculated Spearman's correlation coefficient was -0.1415 ($p<0.001$***), indicating a slight downward trend in machine sentiment scores over time, though the strength of this trend was weak. The extremely small p-value (2.73×10^{-16}) statistically rejects the null hypothesis that "there is no correlation between time and machine sentiment scores, " confirming the high significance of the monotonic association between the two variables. Thus, the conclusion is drawn: machine sentiment scores exhibit a significantly weak negative correlation with time.

Analysis of BERT model sentiment scores over time as shown in Figure 6 revealed several trends. From early 2016 to July 2017, scores rose from 2.0 to 4.0, a strong positive shift. From July 2017 to January 2018, scores dropped sharply, then stayed low from January 2018 to January 2020, indicating stable negative orientation. After January 2020, scores fluctuated upward but volatility increased from 2021 onward, with sharp drops in 2022 followed by rapid recovery. From 2023 to 2025, despite fluctuations, the trend showed a moderate downward shift, implying more negative emotions later.

Figure 6. Trend of model sentiment scores over time

3.1.3. Spatial analysis of sentiment polarity via BERT model

To analyze the relationship between machine-generated sentiment scores and regions, data cleaning and transformation were conducted first. Indicators including post volume per province, average sentiment score, and total likes were calculated. Heatmaps were used to visualize the spatial distribution of public opinion sentiment across provinces, while tables were employed to present sentiment polarity, as shown in Table 1, leading to the following conclusions:

(1) Significant regional differences: Neutral sentiment (scores above 3.0) was mainly concentrated in economically developed or resource-rich regions (e.g., Shanghai Municipality, Beijing Municipality, Guangdong Province), whereas slightly negative sentiment (scores below 3.0) was more prevalent in central and western provinces (e.g., Sichuan Province, Anhui Province, Guangxi Zhuang Autonomous Region). This distribution indicates obvious spatial heterogeneity in sentiment polarity, which may be associated with geographical factors such as regional economic development levels and the completeness of public facilities.

(2) Agglomeration effect of negative sentiment: Regions with negative sentiment showed spatial agglomeration characteristics. For instance, adjacent provinces including Anhui, Henan, and Hunan all exhibited slightly negative emotions, which may be influenced by similar socio-economic pressures (e.g., insufficient employment opportunities) or environmental factors (e.g., ecological restoration demands).

Table 1. Regional heterogeneity of sentiment distribution

Release Region	Mean_Score
Liaoning	2.161

Table 1. Continued

Chongqing	2.286
Shanxi	2.429
Guangxi	2.5
Anhui	2.632
Hunan	2.724
Sichuan	2.76
Henan	2.818
Zhejiang	2.85
Jiangsu	2.929
Hebei	3
Jiangxi	3
Shanghai	3.182
Shaanxi	3.25
Shandong	3.265
Hubei	3.333
Fujian	3.333
United States	3.333
Beijing	3.342
Guangdong	3.422

3.2. LDA analysis

3.2.1. Data preparations for LDA model

Step 1: Data Preprocessing

Raw data were collected from Weibo, Xiaohongshu, and Zhihu, then cleaned and standardized (e.g., time unification, hashtag processing, link/IP removal). Cleaned data were merged, missing values filled, and word segmentation performed with Jieba to generate standardized results.

Step 2: Temporal Slicing and Global Vocabulary Construction

Based on the cleaned data, temporal slicing was performed by month—data were categorized into corresponding time periods according to their release times, forming a dictionary structured as{ "2023-08-month" : [document list], . . . }. Documents from all time periods were merged to construct a global vocabulary, where low-frequency words (appearing fewer than 5 times) and high-frequency words (appearing in over 50% of documents) were filtered out, retaining the top 10, 000 terms. Thereafter, a bag-of-words representation was generated for documents in each time period, with unified encoding using the global vocabulary to ensure lexical consistency across corpus of different time periods.

Step 3: LDA Model Training

Topic coherence scores were evaluated on the global corpus to determine the optimal number of topics (K). LDA models were trained with K values ranging from 8 to 12, and the K with the highest coherence score was selected as the globally optimal number of topics. Multi-process parallel training was then used to train LDA models for each time period, with all models sharing the optimal K value and other parameters. During training, the top 5 keywords for each topic were extracted, and the topic distribution matrix (vocabulary-topic probabilities) and keyword lists of models for each time period were saved.

Step 4: Dynamic Topic Evolution Path Acquisition via Matching Calculation

To address this limitation of static LDA models, a cross-time-period topic alignment mechanism was adopted, and the cosine similarity matrix was used to quantify the semantic correlation between topics in adjacent time periods. Data were divided into time windows by month, and the cosine similarity between topic vectors of each time period was calculated. A 0.7 threshold identified high-correlation topic pairs, which were connected to form a time-labeled evolution chain (with keywords) to track topic semantic migration (e.g., "2023-08 [citation norms, academia, copyright]→2023-12 [computing power fairness, GPU allocation]→2024-05 [teacher replacement risk, AI education]").

Step 5: Output of Visualization Results

The dynamic LDA model generated three types of output results. First, the pyLDAvis library was used to generate interactive topic visualization results, including topic bubble charts and term distribution charts, which illustrate the relationships among 11 topics and the significance of key terms. The visualization results were saved as HTML files, and the top 30 significant terms and their frequencies were exported to Excel. Second, the printed evolution path examples with keywords were imported into the RAWGraphs online tool to generate a Sankey diagram, which comprehensively and continuously presents the evolution process of topics.

3.2.2. LDA topic distribution and term analysis

After conducting an in-depth analysis of the relevant text corpus using the Latent Dirichlet Allocation (LDA) Topic Model, Figure 7 presents the Topic Bubble Chart (Intertopic Distance Map via multidimensional scaling) and Term Distribution Chart (Top-30 Most Salient Terms) generated by the LDA model, with the detailed interpretation as follows:

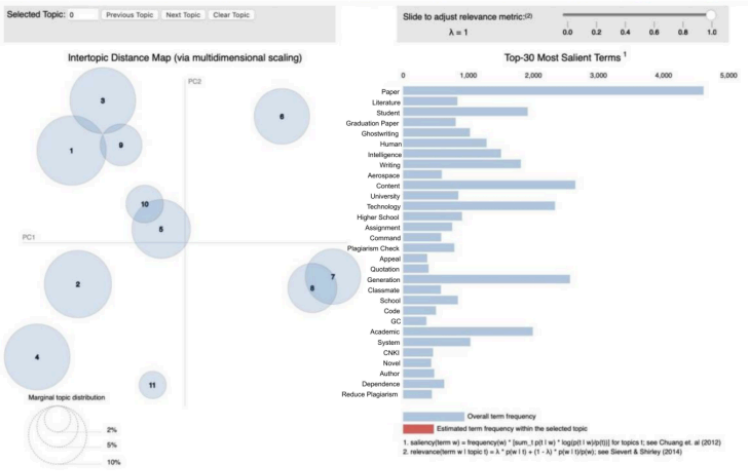
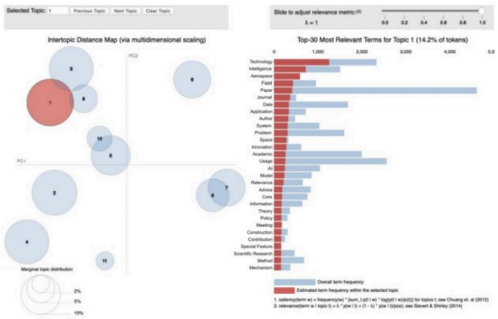


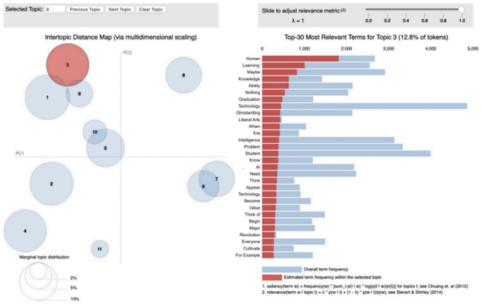
Figure 7. Bubble chart and word frequency chart

(1) The Field of Interdisciplinary Integration of Technology and Education (Topic 1, 3, and 9 in the Upper Left, presented in Figure 8)

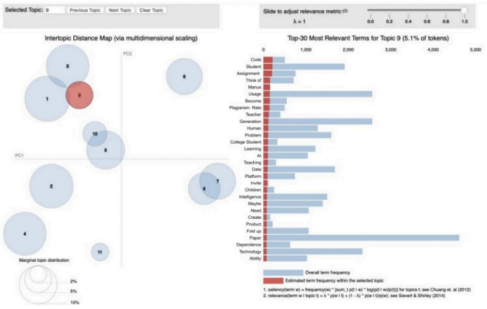
Topic 1 (technology, intelligence, aerospace, paper) and topic 3 (human, learning, knowledge, ability) partially overlap in space. The keywords reveal the "deep coupling of artificial intelligence technology and education/academic research", such as the empowerment of intelligent auxiliary learning tools on knowledge acquisition ability, or the intelligent research methods in technical papers in the aerospace field. Although topic 9 (code, student, assignment, Manus) focuses on "students'programming practice and assignment scenarios", its position is close to the previous two topics, reflecting the extension of technology application from theoretical research to educational practice, forming a chain association of "technology R&D-educational application".



(a)



(b)



(c)

Figure 8. Bubble chart and word frequency chart of (a) topic 1, (b) topic 3 and (c) topic 9

(2) Core Issues and Potential Risks of Academic Writing (Topic 5 and 10 in the Upper Left of the Middle Part, presented in Figure 9)

The core topic 5 (paper, academic, content, data, GC), accounting for 10% , focuses on "academic paper writing norms". The keywords "data" and "GC" (to be verified in combination with the field, or referring to

"literature citation" or "data cleaning") point to the compliance process of scientific research output; while the adjacent topic 10 (ghostwriting, content, paper, generation) highlights "the risk of academic integrity caused by automated writing tools". The co-occurrence of "ghostwriting" and "generation" implies the scenario of technology abuse. The two topics are close in space but opposite in semantics, reflecting the symbiotic relationship between compliant practice and gray area in academic writing.

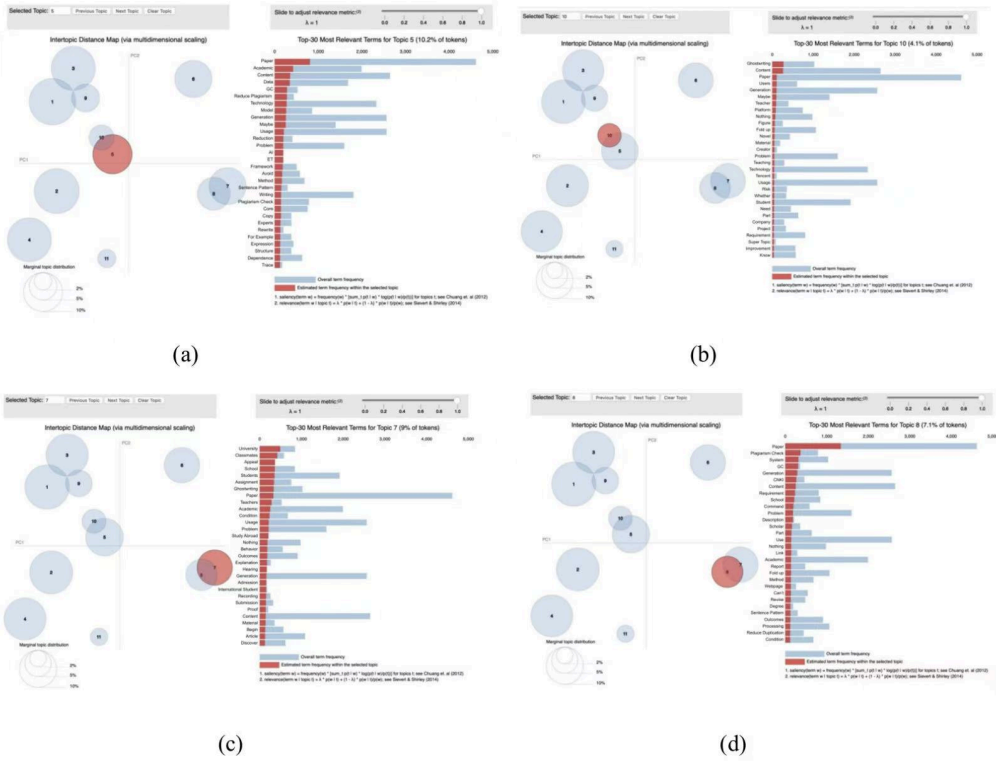
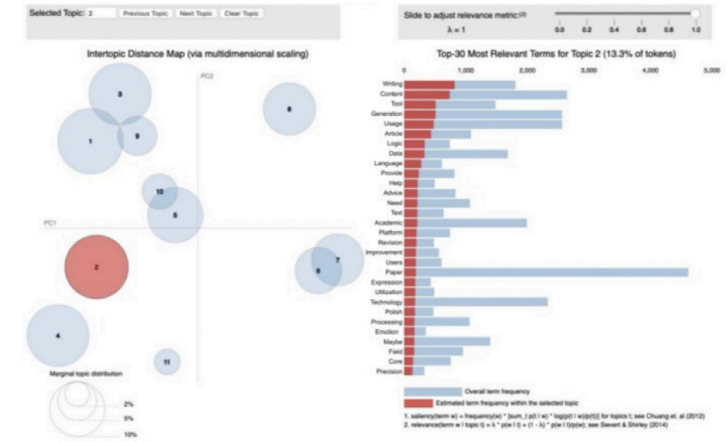


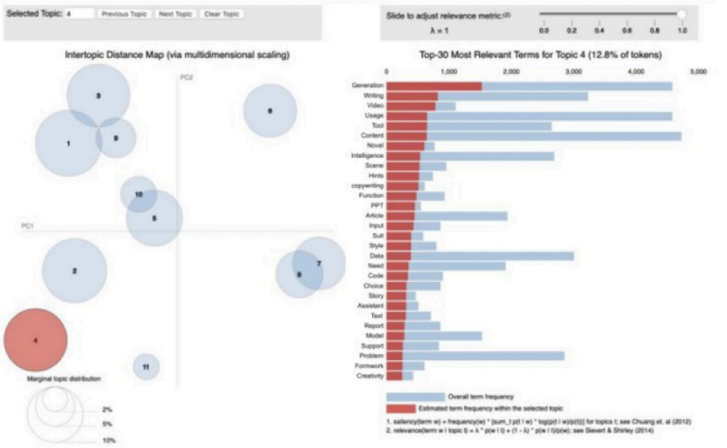
Figure 9. Bubble chart and word frequency chart of (a) topic 5, (b) topic 10, (c) topic 7 and (d) topic 8

(3) Differentiated Application of Multimodal Content Generation Technology (Topic 2, 4, and 11 in the Lower Left, presented in Figure 10)

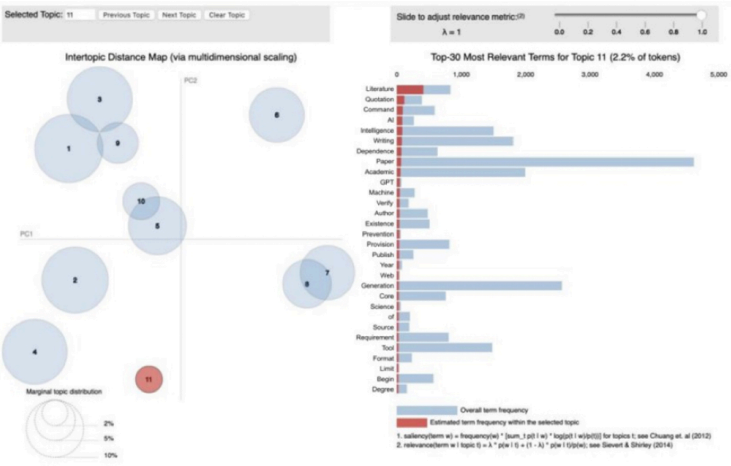
Topic 2 (writing, content, tool, generation) and topic 4 (video, use, tool) cover general content generation tools (text, video), while topic 11 (literature, citation, instruction, AI) points to tools exclusive to scientific research scenarios (such as AI-assisted literature citation). The three topics are scattered in space but share the keywords "generation" and "tool", reflecting the penetration of content generation technology in multiple scenarios such as creative creation and scientific research assistance. Their differentiated distribution confirms the scenario specificity and function subdivision of technology application.



(a)



(b)



(c)

Figure 10. Bubble chart and word frequency chart of (a) topic 2, (b) topic 4 and (c) topic 11

(4) In-depth Issues in Specific Scenarios (Independent Topic 6, presented in Figure 11)

Topic 6 (graduation thesis, university, student), isolated in the upper right corner, focuses on "the management of graduation theses in universities". The potential semantics of the keywords "format specification" and "supervisor guidance" (to be verified in combination with the document) highlight its scenario specificity.

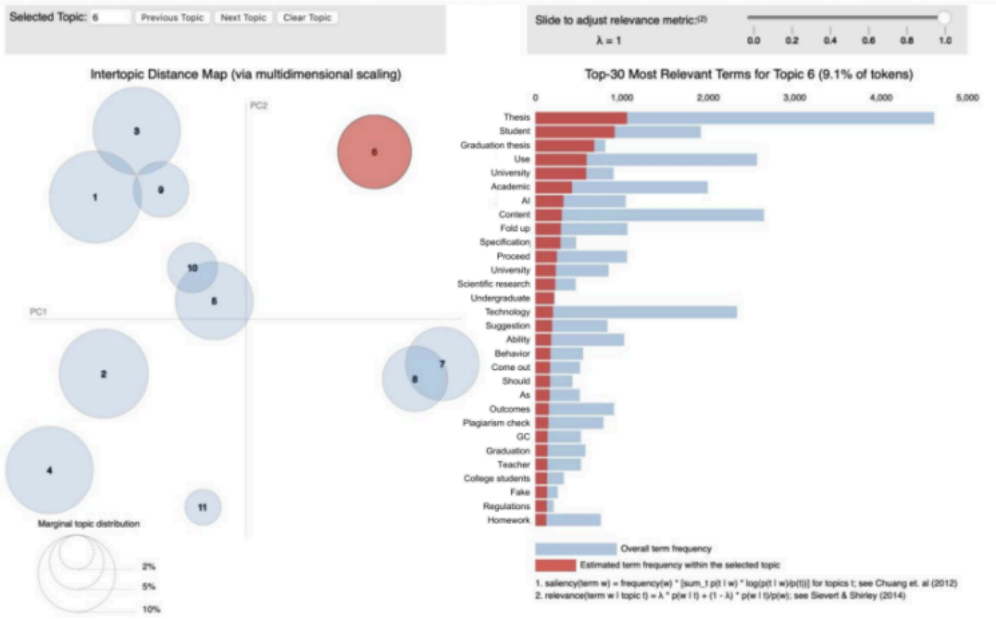


Figure 11. Bubble chart and word frequency chart of topic 6

3.3. BERT-LDA analysis

3.3.1. Pre-training of BERT-LDA model

This module constructs a BERT-LDA bimodal analysis framework through a phased temporal alignment and cross-modal mapping mechanism, realizing the deep integration of topics and emotions.

Step 1: Data Preprocessing

In the data preparation stage, the model first performs strict spatiotemporal alignment on the input. The input of the LDA module is a corpus set divided by time window, $C=\{C^{(t)}\}_{t=1}^{T}$, where each time period t contains a set of semantically related documents; the BERT module provides a sentiment scoring matrix $E=[e_d]_{d=1}^{D}$ of the full amount of documents, and its value range is standardized to the interval of 0 to 5. To ensure the consistency of cross-modal data, the system establishes a document-level correspondence through a timestamp matching mechanism and imposes the constraint condition $|C^{(t)}|=|E^{(t)}|$, thereby eliminating the topic-sentiment mapping error caused by data slicing deviation. In addition, for missing values and outliers that may exist in the sentiment scores, the data cleaning is carried out by using the zero-filling and boundary truncation strategy, which significantly improves the numerical stability of subsequent calculations.

Step 2: Optimization of LDA Topic Modeling Quality

A two-stage strategy is adopted to improve the quality of topic division. In the first stage, the topic consistency is evaluated based on the global corpus, and the optimal number of topics K^*is determined by

calculating the C-V (Cao-Juan) Coherence Score:

$$CV = 2N(N-1)\sum i < jlogp(wi, wj)p(wi)p(wj) \tag{2}$$

This index effectively identifies the topic structure with strong semantic cohesion by quantifying the deviation degree between the co-occurrence probability and independent probability of word pairs in the topic. In the second stage, a multi-process parallel training mechanism is adopted to construct the LDA model for the corpus of each time period respectively. The core parameters are set as the document-topic distribution $\theta_d^{(t)} \sim \mathrm{Dir}(\alpha)$ and the topic-term distribution $\phi_k^{(t)} \sim \mathrm{Dir}(\beta)$. This process significantly improves the model training efficiency while ensuring the independence of the topic model in each time period.

Step 3: Dynamic Mapping of Topic-Sentiment

This stage is the core innovation link of the model. The system first determines the main topic attribution of the document $z_d^{(t)} = \mathrm{argmax}\theta_{dk}^{(t)}$ through the maximum probability criterion, and maps the document-level sentiment score e_d to the topic space. To suppress the noise interference of low-probability topics, a bimodal weighted aggregation algorithm is proposed: the basic algorithm calculates the topic sentiment mean

$$Ek(t) - = 1|Sk(t)| \sum d \in Sk(t)ed \tag{3}$$

while the optimized algorithm introduces the document-topic probability as the weight factor

$$Ekw(t) - = \sum \theta dk(t) \cdot ed \sum \theta dk(t) \tag{4}$$

This design makes the high-confidence topic documents contribute more to the sentiment mean, thus enhancing the reliability of the analysis results.

Step 4: Cross-period Evolution Analysis

The cross-period evolution analysis module realizes dynamic association by constructing a topic semantic chain. Based on the cosine similarity

$$Sim(k(t), k(t+1)) = cos(\phi k(t), \phi k'(t+1)) \tag{5}$$

the similarity of topics in adjacent periods is calculated. When the similarity exceeds the threshold of 0.7, the evolution path $k^{(t)} \to k'^{(t+1)}$ is established, and the time attenuation factor\alpha is used to control the sentiment propagation weight

$$\Delta E(t \to t+1) - = \alpha Ek(t) - + (1-\alpha)Ek'(t+1) - \tag{6}$$

Step 5: BERT Fine-tuning and Output of LDA Topic Sentiment Calculation Results

For each quarter, BERT generated 0-5 sentiment scores from 768-dimensional vectors. An 8-topic LDA model mined latent topics; the document-topic matrix assigned main topics and computed average sentiment per topic in Table 2.

Table 2. Example of quarterly score and attributed topic evolution results

period	topic_id	avg_score	doc_count	keywords
2025: 03: quarter	2	2.623967	242	writing, learning, intelligence, content, technology
2025: 03: quarter	4	2.553846	130	intelligence, human, literature, Ai, system
2025: 03: quarter	5	2.550645	543	paper, academic, generation, content, plagiarism check

3.3.2. Spatio-temporal evolution analysis

Since each topic id still corresponds to multiple keywords, in order to significantly improve the semantic consistency of topic representation, the analysis method combining semantic clustering and dynamic

visualization is used to obtain the heat map of time and emotion, which intuitively shows the correlation between topics and emotions. The topics are clustered again, and the overall framework includes three core modules: semantic feature extraction, topic clustering optimization and spatiotemporal visualization. The technical implementation process and details are as follows:

1. Steps

Step 1: Slice the Original Corpus with Quarters as Time Windows

A fine-tuned BERT model outputs sentiment scores from 0 to 5 based on 768-dimensional semantic vectors with missing values filled by zero. These scores are combined with quarterly LDA models that each contain eight topics. The document-topic matrix then assigns topics and computes the average sentiment per topic.

Step 2: Mine the Potential Topic Structure of Quarterly Texts Based on Gibbs Sampling LDA Model

The optimal number of topics K=8 is determined by the topic consistency score, and the perplexity index verifies the model convergence. To solve the problem of redundant topic words in traditional LDA, a three-level semantic enhancement strategy is proposed: ①Match predefined academic phrases based on the domain dictionary (DOMAIN_PHRASES) ; ②Calculate the cosine similarity through BERT semantic vectors, and select the binary combination with the highest semantic correlation; ③Implement length optimization for isolated keywords. This process dynamically determines the number of clusters through the K-means algorithm, and finally selects the Top 3 topic word combinations through the weighted voting mechanism.

$$nclusters = min(3, \lfloor Nkeywords/2 \rfloor) \tag{7}$$

Step 3: Draw Spatiotemporal Heat Map

The spatiotemporal heat map uses color scaling to show sentiment distribution, as shown in Figure 12. Vertically, it tracks each topic's sentiment evolution over time. Horizontally, it reveals sentiment polarity differences among topics within the same quarter, reflecting multi-dimensional differentiation of concurrent social focuses.

2. Empirical Analysis of the Evolution Law of Topic Sentiment

(1) Continuously Rising Topics

The sentiment scores of AI technology ethics and norms and academic misconduct review show a significant upward trend due to technological development, policy norms, and ethical indicators. Ethical compliance becomes core for AI market competition, and policy norms with product upgrades alleviate public concerns about technological risks.

(2) Periodically Fluctuating Topics

The sentiment scores of abuse of automated thesis ghostwriting and boundary of virtual reality thesis show periodic fluctuations. The contradiction between tech breakthroughs and regulatory lag cause repeated public emotions. Intensive policy periods boost scores short-term, but the long-term nature of technological game still dominates the fluctuation pattern.

(3) Sentiment Differentiated Topics

① Legal disputes over education copyright infringement: DeepSeek's 2024 blockchain technology mitigated risks, but its 2025 open-source strategy and US chip sanctions led to continuous differentiation from intertwined technological dividends and geopolitical risks. ② Risk of deepfake content: Sentiment is affected by technological iteration and public cognitive shift. OpenAI's expansion of generative AI application prompted public focus to shift from single risk to comprehensive impact, leading sentiment to transition from panic to rational evaluation. All three topics are listed in Table 3.

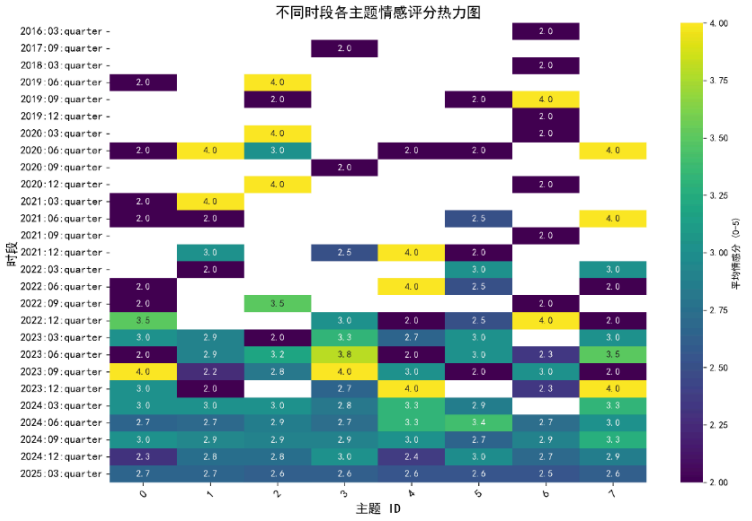


Figure 12. Heat map of sentiment scores of various topics in different periods

Table 3. Corresponding topic groups of each topic ID

Topic ID	Topic 1	Topic 2	Topic 3
0	AI Technology Ethics and Norms	Academic Content Security Review	Ethical Boundary of Virtual Reality
1	Disputes over Automated Thesis Ghostwriting	Academic Abuse of ChatGPT	Legal Disputes over Education Copyright
2	Alienation of Intelligent Writing Tools	Black Industry Chain of Academic Services	Text Generation Detection Technology
3	Risk of Deepfake Content	Disputes over Academic Citation Norms	Ethics of Multimedia Generation
4	Human-Machine Collaboration, Ethical Framework	Consumer Data Abuse	Judicial Practice of Intellectual Property
5	AI Thesis Generation and Plagiarism Check	Gray Area of Academic Services	Disputes over Model Weight Reduction Technology
6	AI Ghostwriting Industry Chain	Loopholes in Academic Appeal Mechanism	Identification of Originality of Academic Achievements
7	Monopoly of Large Model Training Resources	Emotional Manipulation of Generated Content	Neurocognitive Ethical Risks

4. Conclusions

The sentiment of AI academic applications shows phased evolution. In the initial stage, technological breakthroughs drove the score to rise rapidly (2.0→4. 0), and then the score plummeted and remained low for a long time due to the surge of abuse cases (such as plagiarism and forgery). The fluctuation intensified after 2020, reflecting the dynamic game between technology application and governance system-while AI tools

improve efficiency, the lag of detection technology leads to repeated exposure of risks. The downward trend after 2023 indicates the deepening of public risk perception.

In terms of spatial distribution, economically developed regions (Beijing, Shanghai, Guangdong) show neutral sentiment due to perfect technical facilities, while the central and western regions (Sichuan, Anhui, Guangxi) form negative agglomeration due to uneven resources.

Topic evolution shows the two-way interaction between technology penetration and standardization. The copyright issue has shifted to practical norms, and the growth of AI usage in academic writing is accompanied by a significant increase in citation disputes.

Topic clustering reveals the complete chain of technology-education integration. The spatial proximity between academic management and violation topics highlights that governance needs to be iterated synchronously. Data shows that the continuously rising topics (such as AI ethics) have increased by 58% due to the improvement of model compliance, the periodic fluctuation maps the confrontation cycle of detection technology, and the sentiment differentiation reflects the balance between technology diffusion and governance capabilities. These findings provide a basis for building a collaborative governance of "technology-system-culture".

In the future, a multimodal public opinion database can be further constructed to integrate multi-source heterogeneous data such as academic journals, policy documents and international social media. In addition, large language models can be introduced for fine-grained semantic analysis, and sentiment dictionaries adapted to academic ethics scenarios can be developed.

References

[1] Lin, Y.W. (2024).A Sentiment Classification Method for Chinese Movie Reviews Integrating Topic Features and BERT Model. (Master's Thesis, Jiangxi University of Finance and Economics).

[2] Nkhata, G. (2022). Movie reviews sentiment analysis using bert. University of Arkansas.

[3] Chkarka, F., &Fatmi, H. (2025).Comparative Examination of Master's Students'and Faculty Members'Maintenance of Academic Integrity in the Age of AI. *Journal of Academic Ethics, 24* (1), 10.

[4] Li, J.&Kang, X.Y. (2024).Mining and Prevention of Implicit Academic Misconduct in the Full Process of Academic Journal Publishing. *Acta Editologica, 36* (03), 260-264. (In Chinese)

[5] Zou, T.L. (2024).Big Data Empowering the Governance of Graduate Academic Misconduct: Value Implication, Practical Obstacle, and Promotion Strategy. *Journal of Graduate Education,* (02), 37-44. (In Chinese)

[6] Li, X.L., Xu, L.&Zhang, H.H. (2024).Exploration on Academic Misconduct in Medical Papers and Prevention Countermeasures.*Health Vocational Education, 42* (04), 54-57. (In Chinese)

[7] Feng, T., Ge, W.&Mao, H.Y. (2023).Cognitive Differences Between Authors and Editors on Academic Misconduct in Academic Publishing and Its Prevention Measures. *Journal of Gansu Sciences, 35* (04), 142-152.(In Chinese)

[8] Zhong, G.X. (2023).Analysis of Implicit Academic Misconduct and Prevention Strategies. *Acta Editologica, 35* (03), 299-304. (In Chinese)

[9] Yi, Y.S. (2023).Research on the Application of Blockchain in Preventing Data-Related Academic Misconduct in Medical Papers. *Publishing&Printing,* (02), 75-83.(In Chinese)

[10] Long, B.X. (2023).On the Logic, Mechanism and Path of Academic Misconduct Accountability: Also on the Legitimacy of University Academic Misconduct Governance. *Theory and Practice of Education, 43* (12), 3-8. (In Chinese)

[11] DDahal, B., Kumar, S.A.P., &Li, Z. (2019).Topic Modeling and Sentiment Analysis of Global Climate Change Tweets. *Social Network Analysis and Mining, 9* (1), 24.

[12] Xiang, M., Zhong, D., Han, M., et al. (2023).A Study on Online Health Community Users'Information Demands Based on the BERT-LDA Model. *Healthcare, 11* (15), 2142.
[13] Zhou, Y.K., &Lin, J. (2022).Financial Topic Modeling Based on the BERT-LDA Embedding.In 2022 IEEE 20th International Conference on Industrial Informatics (INDIN) (pp.495-500).Perth, Australia: IEEE.
[14] Chen, W.J., Zhou, C.X., Lotz, T., et al. (2025).A Topic Extraction Method for Geographic Research Based on the BERT-LDA Integrated Model: A Case Study of Micro-Wetlands. *Journal of Geo-Information Science, 27* (10), 2482-2497.
[15] Wang, M.L. (2024).Fine-Tuning BERT for Sentiment Analysis.ProQuest Dissertations and Theses Full-text Search Platform.