

Application of Interactive Data Visualization Design Based on BERT Sentiment Analysis in Patriotism Education

Yi Cui

Zhengzhou Railway Vocational & Technical College, Zhengzhou 451460, Henan, China

Corresponding author: Yi Cui (e-mail: cuiyi777777@163.com)

ABSTRACT Complex cultural metaphors in patriotism education texts make accurate sentiment analysis difficult, while existing visualization tools provide limited interactivity for instructional support. To address these limitations, this study develops an interactive data visualization framework driven by a domain-enhanced BERT model for high-metaphor and high-compliance educational scenarios. Domain-adaptive pre-training and Fast Gradient Method adversarial training are introduced to strengthen semantic representation learning, enabling the model to identify implicit rhetorical expressions rather than relying on surface lexical features. The optimized DE-BERT model recognizes domain-specific metaphors, euphemisms, and latent value orientations, providing interpretable semantic evidence for educational analysis. Guided by cognitive load theory, a three-level visualization architecture integrating smoothed Bézier sentiment curves, dynamic heatmap arrays, and repulsive bubble labels is constructed to improve information presentation and user interaction. The framework also offers methodological value for semantic perception and intelligent visualization in electromagnetic information processing and human-machine collaborative analysis. Experiments show a classification accuracy of 0.90 and macro F1-score of 0.88 under high-metaphor-density conditions, with response times of 783.4 ms, 186.3 ms, and 874.6 ms across teaching scenarios and a single-operation success rate above 97%.

INDEX TERMS bert, sentiment analysis, interactive visualization, semantic perception, electromagnetic information processing

I. INTRODUCTION

IN contemporary education, text conveys both knowledge and values. In many sectors, sustainability reports encode ethical and compliance values, yet their jargon-laden language hinders accurate assessment [1,2]. Patriotic education texts embed cultural genes through poetic imagery, historical allusions,

and collective memory, relying on metaphor, symbol, and context to transmit values—posing challenges for machine understanding [3-6]. Meanwhile, data visualization is evolving into an interactive interface fostering cognitive engagement [7,8]. Thus, enabling NLP algorithms to grasp the emotional weight of metaphorical expressions is key to integrating technology with humanities education. Current sentiment analysis struggles with semantic decoding and teaching adaptability in patriotic education texts, which use concise language, allusions, and historical context to form complex semantic structures [9]. Phrases like “The Yangtze River and the Yellow River surge endlessly” symbolize national spirit, not natural phenomena; “The red boat

breaks the waves” evokes specific historical memory, with emotion not deducible from word sum [10-12]. General-corpus-trained models lack cultural symbol representation and misjudge emotion, especially in real classroom corpora despite good benchmark performance [13,14]. Visualization also falters: static charts offer dense but inoperable information, hindering dynamic teaching adjustments and meaning negotiation [15]. Teachers cannot trace model decisions [16-18] or guide students in discussing why a sentence is labeled “positive” or “reverent”. Emotional keyword extraction lacks educational logic and fails to adapt to cognitive stages (e.g., elementary figurative vs. high school abstract thinking) [19,20]. The system remains an isolated front end with a closed back end and lacks interactive links between technical output and teaching practice. As a result, it cannot support annotation adjustment, emotional trajectories comparison, or attribution discussions, which limits real-world educational integration. This research builds an interactive sentiment analysis and visualization system for patriotism education, bridging NLP technology and teaching

practice. Centered on pedagogical usability, it contributes in four ways: (1) constructs the first Chinese patriotic corpus annotated with cultural metaphors; (2) proposes DE-BERT, enhancing rhetorical expression recognition via domain-adaptive pre-training and adversarial training; (3) designs a three-level visualization based on cognitive load theory for progressive display of sentiment intensity, keywords, and trends across grades; and (4) develops an interactive mechanism enabling teacher–student collaboration in annotation, feedback, and attribution, turning analysis into a discussion springboard. The innovation lies in integrating language model semantics with pedagogy into a dynamic analysis–presentation–interaction–correction cycle, transforming AI from a black box into a transparent collaborator. The resulting paradigm—deep semantic decoding and pedagogy-driven visualization—is scalable to other domains requiring nuanced ethical interpretation.

II. RELATED WORK

Current research on Chinese sentiment analysis offers a diverse technical landscape. BERT-Base-Chinese excels in semantic modeling across public corpora through bidirectional context encoding. RoBERTa-wwm-ext [21,22] refines training and whole-word masking, improving performance on informal texts like Weibo. The ERNIE series integrates knowledge graphs to resolve ambiguity, especially in texts with proper nouns, and has been used in triage systems and English teaching [23,24]. TextCNN captures n-gram features efficiently for short-text tasks, often applied in extraction and prediction [25,26]. Yet, all these models struggle with allusions, historical analogies, and symbols common in patriotic education. Dictionary-based post-processing or attention-weighted keywords have not resolved core metaphor-understanding gaps. In visualization, Apache Superset [27] offers flexible graphics, while Tableau Public [28,29] and Power BI [30,31] are popular for teaching dashboards—but all focus on static output, lacking mechanisms for real-time label adjustment, judgment tracing, or co-interpretation in class. No existing system bridges semantic analysis and pedagogical interaction in a closed loop. Recent efforts explore domain customization: domain-adaptive pre-training improves BERT’s grasp of professional terminology [32,33]; adversarial training boosts robustness against sarcasm or irony [34,35]; cognitive load theory informs layered visualization in education [36]. However, limitations persist—domain adaptation ignores cultural metaphors; adversarial samples aren’t tailored to educational texts; visual hierarchies overlook developmental differences; and interaction remains unidirectional. This paper proposes a system integrating domain-enhanced BERT with pedagogy-driven visualization. It builds an annotated metaphor corpus, applies adversarial training for better discrimination, and designs a cognitively aligned, adjustable display to enable an interactive, traceable, and negotiable analytical loop—linking deep semantic understanding with classroom dialogue.

III. DOMAIN-ENHANCED BERT AND CONSTRUCTION OF TEACHING VISUALIZATION FRAMEWORK

Figure 1 shows the technical path and application loop of an interactive visualization system for patriotism education based on BERT sentiment analysis. Starting from an annotated patriotic corpus, the system builds an enhanced BERT model with cultural metaphor recognition via domain-adaptive and adversarial training. Model outputs are structured into a three-level visualization aligned with cognitive load theory. Teachers can adjust sentiment labels to trigger model updates; students explore cultural symbols with historical context. All interactions are logged to track cognitive evolution, enabling a full transition from semantic analysis to value internalization. The core innovation integrates NLP precision with pedagogical principles, offering a technically rigorous and educationally applicable solution.

A. CONSTRUCTION AND SEMANTIC ANNOTATION OF PATRIOTISM CORPUS

1) Multi-Source Text Collection and Cultural Metaphor Corpus Screening

The Chinese text collection includes ideological and political textbooks from People’s Education Press, selecting passages on revolutionary traditions and national spirit, emphasizing sentences with historical references, geographical symbols, and classical poetry quotes. Red classics like *Red Rock*, *Song of Youth*, and *Lin Hai Xue Yuan* are included for their narrative and emotional traits. Mainstream media commentary is limited to patriotism editorials from *People’s Daily* and *Guangming Daily* (past three years), focusing on rhetorical units that construct collective identity. Texts are deduplicated, segmented, and cleaned to form the initial corpus. A dual-track screening process involved 10 Master’s-level education researchers independently identifying sentences with cultural metaphors—e.g., historical metonymy, symbolic natural imagery, and allusive emotional projection. Annotators marked metaphor-triggering units and justified judgments. Disagreements were arbitrated by five senior researchers to align semantic boundaries. Finally, 4,872 culturally referential, emotionally discernible sentences were selected as core units, each tagged with metadata (source type, grade level, rhetorical type, paragraph number) for annotation context.

2) Implementation of Fine-Grained Sentiment Annotation and Word-Level Metaphor Annotation

Sentiment annotation employs a five-level scheme: “Extremely Positive” (high respect/strong agreement), “Positive” (affirmation without praise), “Neutral” (objective/ritualistic), “Negative” (critical but not rejecting core values), and “Extremely Negative” (highly adverse expressions). Five annotators—trained in ideological-political education or linguistics—undergo standardized instruction using 20 typical and 10 controversial examples. Annotation begins only after inter-annotator Kappa reaches 0.83. Each sentence is doubly annotated; consistent labels enter the

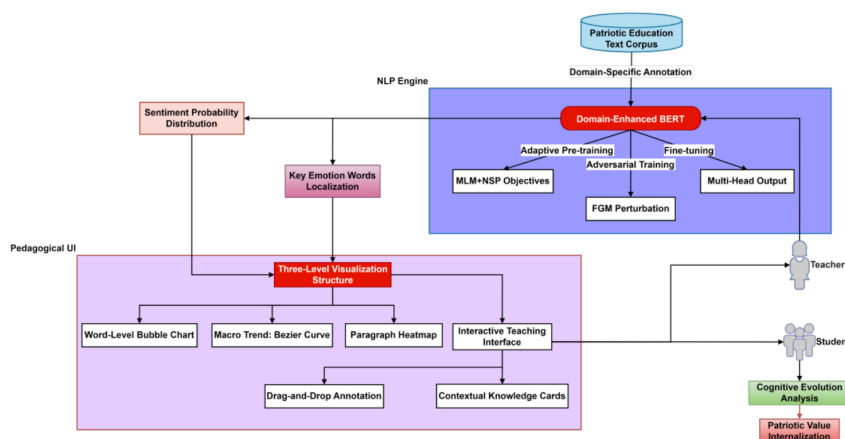


FIGURE 1. Interactive visualization system based on BERT sentiment analysis

training set, while disagreements are resolved by expert review. Word-level annotation then marks metaphorical units (e.g., “backbone”, “hot soil”) with roles such as symbolic subject, value carrier, and historical mapping. A custom web tool supports highlighting and label selection with real-time saving. A 10% periodic audit is conducted, and annotation continues only after three consecutive rounds with an error rate below 2%. The final corpus includes 4,872 sentence-level sentiment labels and 12,631 metaphorical units (avg. 2.59 per sentence). After desensitization, data is split 7:2:1 into train/validation/test sets. All records are version-logged for traceability and directly fed into the model input layer to guide the attention mechanism toward cultural keywords. Table 1 summarizes the distribution of word-level metaphor annotations in the patriotic text corpus, covering three main sources: school textbooks, red classics, and mainstream media commentary. Annotation types are categorized as symbolic subject, value carrier, and historical mapping. The data shows that textbooks contain the highest number of metaphors (5,442), with the largest proportion of symbolic subject annotations (2,541); red classics follow closely (4,548), with significant value carrier annotations (1,458); mainstream media commentary has a lower metaphor density (2,641), with only 647 historical mappings. Overall, symbolic subject annotations are the core form of cultural metaphor (5,842 in total), reflecting the prominent role of symbolic expression in patriotic texts. The data is double-blindly annotated and verified, with an error rate of less than 2%, supporting the domain-adaptive training of subsequent models.

Figure 2 systematically compares the emotional distribution of three cultural metaphor types (historical symbols, geographical metaphors, and classic quotations) in patriotic education texts across different grades (elementary school to university). The five sub-figures show the proportions of five emotional levels: extremely positive, positive, neutral, negative, and extremely negative. The data shows that extremely positive emotions associated with historical symbols increase significantly with grade level (from 15%

in elementary school to 50% in university), reflecting a deepening of value identification in higher grades. The proportion of extremely positive emotions associated with classic quotations surges to 60% in university, reflecting the strong emotional orientation of political theory texts. Neutral emotions are highest in elementary school (reaching 50% for geographical metaphors), consistent with the figurative nature of lower grades. Negative emotions generally do not exceed 15%, and extremely negative emotions are strictly controlled at 1%, confirming the filtering mechanism of educational texts against extremely negative expressions. The data reveal two key patterns: first, the emotional load of metaphors varies significantly across different types, with historical symbols having the most substantial positive emotional reinforcement effect; second, increasing grade level is positively correlated with growing metaphor complexity.

B. DESIGNING DOMAIN-ADAPTIVE PRETRAINING STRATEGY TO OPTIMIZE THE BERT MODEL

1) Domain-Adaptive Pretraining and Parameter Update Mechanism

The model initializes with Chinese BERT-wwm, freezing all layers except the embedding layer, and undergoes continuous pretraining on a patriotic corpus using joint MLM and NSP objectives. Input length is 512; dynamic masking enhances contextual adaptability. Batch size is 32; learning rate starts at $2e-5$ with linear warmup and cosine decay over 80 epochs to avoid overfitting. The corpus features historical and political idioms; during pretraining, the model refines attention weights to better capture high-frequency cultural keywords like “red flag” and “Long March”. Special focus is placed on [CLS] token cohesion: a backpropagation path strengthens its link to cultural keywords, and Frobenius-norm regularization constrains [CLS]–entity attention deviations, stabilizing sentiment anchoring. Checkpoints are saved each epoch; the version with best MLM accuracy and lowest NSP loss on validation is selected. The final model shows robust context perception in unlabeled texts, maintaining coherence on dense rhetorical phrases like “iron

TABLE 1. Distribution of word-level metaphor annotation types

Source Type	Symbolic Subject	Value Carrier	Historical Mapping	Total
School Textbooks	2,541	1,872	1,029	5,442
Red Classics	2,103	1,458	987	4,548
Media Commentary	1,198	796	647	2,641
Total	5,842	4,126	2,663	12,631

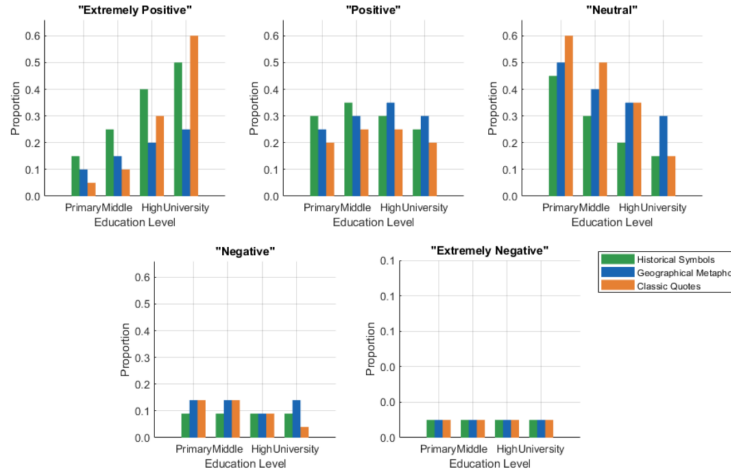


FIGURE 2. Emotional distribution of cultural metaphor types in different grades

shoulders bear morality,” avoiding breaks from unconventional collocations. In the continuous pre-training stage, the model adopts masked language modeling and next sentence prediction joint optimization objectives, and its composite loss function is defined as:

$$L_{\text{pretrain}} = \alpha \cdot L_{\text{MLM}} + (1 - \alpha) \cdot L_{\text{NSP}} \quad (1)$$

L_{MLM} represents the masked language modeling loss (negative log-likelihood over masked position), and L_{NSP} denotes the binary cross-entropy loss for next sentence prediction. $\alpha \in [0, 1]$ is a weighting coefficient (set to 0.7 in the experiment) that emphasizes the language modeling objective to strengthen semantic representation learning. This loss drives parameter updates, enabling the model to more accurately reconstruct cultural keywords in the context of patriotism.

2) Applying Adversarial Perturbation and Sentiment Localization Joint Output on the Fine-Tuning Stage

Entering the supervised learning stage, end-to-end fine-tuning is performed using a labeled fine-grained dataset. The task goal is a dual-output structure of multi-class sentiment discrimination and key sentiment word localization. The classification head is connected to the final hidden state of the [CLS] token; the sentiment classification task outputs a five-category probability distribution. Let $h[\text{CLS}]$ denote the final hidden vector at the [CLS] position. The classification layer is parameterized by W_c and b_c , and then the sentiment logits are computed as:

$$z_c = W_c h[\text{CLS}] + b_c \quad (2)$$

The sentiment probability vector is obtained by softmax normalization and used for the final decision. The localization head generates word-level sentiment saliency scores by linearly transforming the hidden vectors at each position of the sequence. The binary tag sequence (whether it is a sentiment-driven word) is used to calculate BCELoss. The two losses are weighted and merged at a fixed ratio. The adversarial sample generation mechanism is applied during training, and the fast gradient method (FGM) is used to construct the small perturbation direction. After forward calculation to obtain the initial loss, the maximum ascending path in the gradient direction of the embedding layer is calculated, and a perturbation vector with an amplitude of $\epsilon = 1e-3$ is added along this direction. The model is re-entered for a second forward propagation, and the parameters are updated. This process simulates possible semantic changes in the text (e.g., replacing “unyielding fighting spirit” with “immortal fighting spirit”), forcing the model to rely on deep semantics rather than surface-level word co-occurrence when extracting key features. To improve the robustness of the model to semantic perturbations, the fast gradient method is applied to generate adversarial samples. Let the input embedding be E and the loss be $L(\theta, E)$. The gradient with respect to the embedding is:

$$g = \nabla_E L(\theta, E) \quad (3)$$

The adversarial perturbation δ is then constructed as:

$$\delta = \epsilon \cdot \frac{g}{\|g\|_2} \quad (4)$$

Here, $\epsilon = 1 \times 10^{-3}$ controls the perturbation amplitude to prevent deviation from the original semantic space. The updated embedding $E' = E + \delta$ is used as the new input to recalculate the loss and update the model parameters, enhancing the stability to synonymous substitution and rhetorical variation. Adversarial training rounds are performed simultaneously with regular fine-tuning, and a disturbance injection is performed on each batch of data to ensure that the model maintains consistency in sentiment judgment when facing slight variations such as synonym substitution and word order adjustment.

C. DESIGN OF VISUALIZATION HIERARCHICAL STRUCTURE BASED ON COGNITIVE LOAD THEORY

1) Construction of Three-Level Hierarchical Architecture and Information Density Control

The visualization framework's effectiveness hinges on the model's stable decoding of cultural metaphors. Standard models like BERT or TextCNN suffer from superficial word-frequency bias in high-metaphor texts (e.g., misreading "Red Boat" as neutral), distorting heatmaps and bubble labels, disrupting the whole-part-detail cognitive flow, and increasing interpretive load. DE-BERT, via domain adaptation and FGM-based adversarial training, ensures smooth, credible emotional trends with high alignment between keyword attribution and expert annotations, supporting reliable hierarchical visualization. The structure follows whole-part-detail logic across three independently activated levels. Level 1 shows overall sentiment trend: the horizontal axis marks chapter reading order, the vertical axis maps continuous sentiment polarity, and Bézier curves smooth fluctuations to prevent visual misinterpretation. Let P_0 and P_3 be the the start and end points, and P_1 and P_2 be the control points. The cubic Bézier curve is defined as:

$$B(t) = (1-t)^3 P_0 + 3(1-t)^2 t P_1 + 3(1-t) t^2 P_2 + t^3 P_3, \quad t \in [0, 1] \quad (5)$$

This expression ensures a smooth transition of the trend line, suppresses the visual jump caused by local noise, and improves reading coherence. The color system employs a blue-white-red CIELAB gradient: cold hues for negative sentiment, warm for positive, and gray-white for neutral, ensuring accessibility for color vision deficiency. It shows only the main trend, omitting auxiliary marks to preserve interface clarity and reduce cognitive load. The second level reveals paragraph-level sentiment fluctuations. Tapping any primary interval loads a sentence-level heatmap: each cell represents one sentence, with color depth indicating sentiment confidence and enhanced edges for boundary clarity. Fixed spacing creates visual rhythm; consecutive same-direction sentiments trigger a semi-transparent emotional accumulation zone overlay. Horizontal scrolling and gesture

zooming ensure mobile usability. The third level exposes keyword sentiment contributions. Long-pressing a sentence expands it to the word level: the text is semantically aligned, with semi-transparent bars beneath words showing attribution weights (the height indicating significance). High-impact words also represented by bubbles, where the diameter indicates sentiment strength and the color indicates sentiment polarity. Bubbles repel dynamically to avoid overlap. This layer activates only on demand and auto-reverts upon exit, forming an explore-return loop. Transitions use slow-in/slow-out animations under 300 ms to maintain attention continuity.

2) Dynamic Information Presentation Mechanism and Interactive Visual Control

The interface uses a responsive layout: the main view occupies 70% of screen width, with the rest for controls and instructions. Non-core elements are low-opacity by default, becoming fully opaque on hover or selection, enabling dynamic visual prioritization. Color contrast meets WCAG 2.1 AA ($\geq 4.5:1$), using a sans-serif font with size and weight reflecting hierarchy—headlines bold, key numbers monospaced. An intelligent focus mechanism activates after 800 ms of dwell, highlighting related data (e.g., showing a paragraph's position in the overall trend). In multi-document mode, layouts switch to side-by-side; sentiment curves use distinct line styles with a shared color base for cross-document tracking. Users drag anchor points to redefine segments, triggering real-time heat map updates. Clicking bubble markers opens pop-ups with etymological analysis, co-occurrence networks, and collocation patterns. All graphics are vector-based for lossless scaling. The modular, collapsible legend saves space. Night mode inverts colors on a dark background to reduce eye strain. Event-driven state management ensures lag-free feedback. The design offers layered information density, revealing analytical depth on demand while maintaining cognitive fluency in teaching.

D. IMPLEMENTING TWO-WAY INTERACTIVE MECHANISM FOR TEACHERS AND STUDENTS

1) Dynamic Sentiment Label Correction and Model Response Mechanism

The front-end embeds a drop-down icon in each sentence's heatmap cell. Clicking it reveals five emotion options—"Extremely Positive" to "Extremely Negative"—with the model's original prediction highlighted. Selecting a new label sends the sentence and correction via an encrypted channel to the back-end. The server locks the paragraph's context window, reruns inference using the fine-tuned domain-enhanced BERT model, and updates sentiment probabilities and keyword weights. Results return in real time, refreshing the heat map color, trend curve, and word-level bubbles, with a pulsing halo marking the change. The model's response mechanism only re-executes forward inference on sentences with modified labels during interactions, without updating model parameters. The system leverages the pre-

trained DE-BERT model to recalculate sentiment distribution and keyword weights for newly annotated contexts, with real-time updates to visualization results. All model weights remain frozen after deployment to ensure system responsiveness and stability. Each correction operation generates a structured record containing the operation timestamp, user role identifier, original label, new label, contextual text snippet, and model response latency. This record is written to a distributed log queue for subsequent analysis. Teachers can modify multiple sentence labels continuously, and the system processes them one by one in sequence. Batch submission and single-step undo are supported. Undo actions also trigger view rollback and model state restoration, ensuring reversibility. The right panel of the interface provides a “Modification History” timeline, which lists all annotation adjustment events in chronological order and supports filtering by sentence position or sentiment type, helping teachers review their decision paths.

2) Student-Side Interactive Response and Knowledge-Linked Feedback Design

Label editing is disabled for students, but emotional hotspot detection remains. Hovering or tapping a hotspot triggers detection of emotional extremes, identifies the highest-attribution words, and sends a request (word, context, emotion category) to the backend. The server queries a pre-built cultural semantic knowledge base—covering historical, policy, and revolutionary texts—and returns a floating card with definition, historical reference, symbolic meaning, and examples, tailored to learning levels. The card appears below the hotspot, avoiding text obstruction, and closes via tap or swipe.

The classroom Q&A module resides in the right toolbar. Teachers enable “question mode”, select a sentence, and apply open-ended templates such as “Why is emotional intensity high here?” or “Why is this word the emotional core?” Published questions sync to students. Upon submission, answers are compared to the original sentiment attribution using a pre-trained semantic matching model; cosine similarity quantifies alignment with the standard interpretation. Assuming the student answer encoding vector v_s and the reference attribution vector v_r , the matching degree is:

$$\text{sim} = \frac{v_s^T v_r}{\|v_s\|_2 \cdot \|v_r\|_2} \quad (6)$$

Similarity above 0.75 indicates cognitive alignment; 0.50–0.75 indicates partial understanding; and below 0.50 indicates understanding deviation, which is used to generate aggregated views of teaching feedback. Responses above the threshold are labeled “cognitive alignment”; whereas those below are labeled “understanding deviation”. Aggregated anonymously, these results show class-wide cognitive hotspots on

the teacher’s dashboard. All interactions—clicks, dwell times, Q&A—are logged in JSON with device ID,

session number, and role tags for persistent storage, enabling offline analysis of common misunderstandings. A teaching replay feature lets teachers review session recordings, replay interactions, and trace student paths to refine instruction. Figure 3 shows the interface.

IV. SYSTEM PERFORMANCE VERIFICATION

A. EVOLUTION OF ATTENTION WEIGHTS FOR CULTURAL KEYWORDS

During pre-training, domain-enhanced BERT’s attention distribution is analyzed using 12 patriotic keywords as observation variables. Every 10 rounds, validation samples containing these keywords undergo a forward pass; mean attention weights from the last Transformer layer’s [CLS] token at keyword positions are extracted. Original text structures are preserved to avoid bias, attention scores are normalized across samples, and a time-keyword weight matrix is recorded. Figure 4 displays the evolution of [CLS] token attention weights for 12 patriotic keywords during domain-adaptive pre-training, computed from validation set forward passes. Average attention scores between the [CLS] query and each keyword’s key vector are shown, with training rounds on the vertical axis, keywords on the horizontal, and color intensity indicating attention strength. Attention to “Red Boat” and “Long March” rises to 0.46 and 0.44; “Revival” and “Unity” also grow steadily. All peak at round 50, confirming enhanced focus on core semantic units. The top-to-bottom color gradient forms an enhancement band, demonstrating synchronous, cumulative representation learning and validating the domain-data-driven semantic cohesion mechanism.

B. CHANGES IN SENTIMENT CLASSIFICATION ACCURACY UNDER COMPLEX SEMANTIC PERTURBATIONS

To evaluate the model’s stability under semantic perturbations, a test matrix is constructed using the control variable method. A perturbed text set is generated using the dual-axis parameters of lexical substitution rate and sentence structure variation. The original text is selected from sentences with high metaphor density in the patriotic corpus (defined as containing ≥ 4 word-level metaphorical units per 100 words). Keyword substitutions are performed using a synonym library, and structural adjustments are made using syntactic transformation rules. Fifty perturbed samples are generated for each combination. Both models perform forward inference on the same data, and the sentiment classification accuracy is recorded. The data is normalized and interpolated into a continuous surface for visual comparison. Figure 5 uses a hyperbolic surface to show sentiment classification accuracy before and after FGM adversarial training under compound semantic perturbations. Lexical substitution (0–30%) replaces terms like “red boat” with “ship”; sentence variation (0–50%) adjusts word order or rhetoric (e.g., passivation). Without adversarial training, accuracy plummets below 0.1 at 30% substitution and 50%

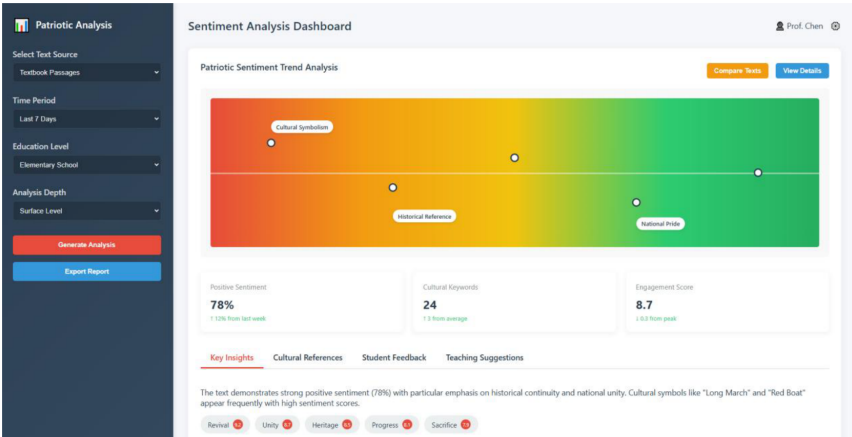


FIGURE 3. System operation interface

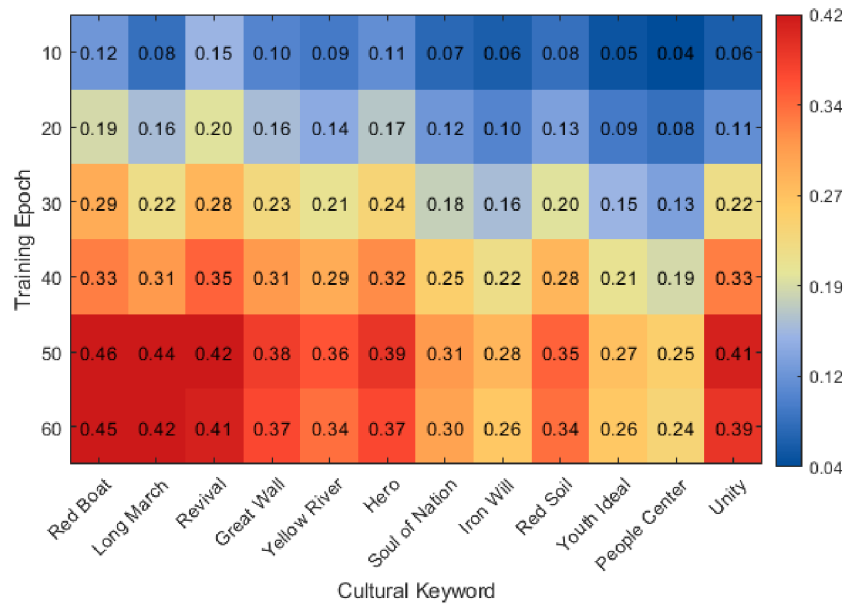


FIGURE 4. Evolution of attention weights for cultural keywords in the [CLS] token

variation, revealing overreliance on surface features and failure under synonym shifts. With FGM, accuracy stays above 0.5 under the same conditions, indicating learning of deep, invariant cultural metaphor representations by forcing the model to prioritize contextual value and rhetoric over keywords.

Semantic alignment is cosine similarity between student and standard response vectors from domain- enhanced BERT. Group consistency is the variance contribution of the first PCA component of class response vectors—higher values mean convergent understanding. Core meaning deviation uses word-frequency- weighted edit distance, penalizing misinterpretations of key symbols. Focus shift is tracked via symbolic vocabulary proportion changes, dynamically weighted over time to capture cognitive progression from concrete to abstract. All metrics are normalized. Figure 6 systematically presents the cognitive evolution of students during interactive patriotism instruction. This data originates

from a supervised classroom intervention experiment involving 212 first-year high school students (approximately 1:1 male-to-female ratio) from six secondary schools. All participants used the same patriotic education materials (excerpts from Red Rock) and standardized teaching scripts. Under identical hardware conditions, students completed a 45-minute interactive task using the system, with teachers providing only guided prompts without personalized instruction. Semantic consistency was assessed through cosine similarity calculations between student responses and expert-standardized explanations (using DE- BERT encoding). The entire evaluation process was fully automated, eliminating subjective human grading to ensure consistent task execution and measurement. Figure 6a shows students' evolving semantic alignment (solid line) with the teacher's standard interpretation, rising from 0.45 to 0.85, and group response consistency (dashed line), both indicating reduced cognitive bias and convergent understanding through interactive feed-

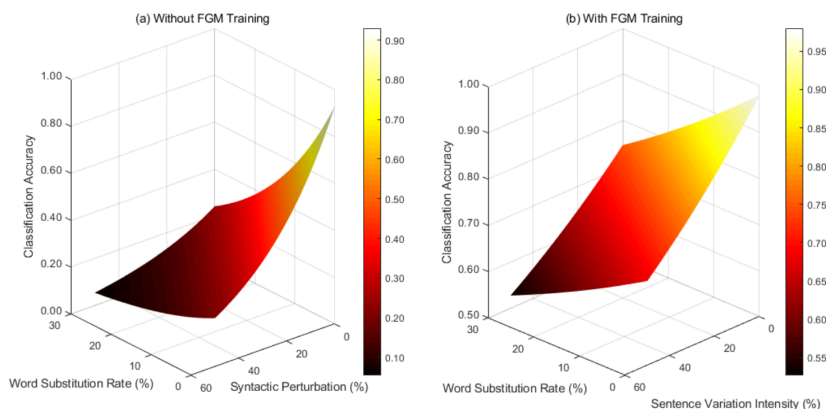


FIGURE 5. Changes in sentiment classification accuracy under compound semantic perturbations

back. Figure 6b reveals a structural shift: deviation from core meaning drops from 0.31 to 0.12 (solid line), suppressing superficial interpretations, while interpretive focus (orange triangles) shifts non-linearly—dipping at the 25th minute due to cognitive conflict but rebounding to 0.36, signaling a transition from concrete to abstract understanding. This mechanism enhances comprehension and internalization of values, offering dynamic evidence for technology-enabled ideological education.

C. SENTIMENT CLASSIFICATION PERFORMANCE

The test set is split into three cultural metaphor density levels: low (≤ 1 metaphors per 100 words), medium (2–3), and high (≥ 4), ensuring balanced samples across grades. After preprocessing, each text is input to five models—DE-BERT, BERT-Base-Chinese, RoBERTa-wwm-ext, TextCNN, and ERNIE—for independent sentiment predictions. Accuracy and macro-F1 (mean F1 per class, mitigating class imbalance) are computed by comparing predictions with true labels. Evaluation occurs in a closed environment with frozen parameters, no online learning, reflecting native model performance. Metrics are grouped by density, enabling dual-dimensional comparison across models and semantic complexity. Figure 7 shows sentiment classification performance across five models under varying metaphor densities. DE-BERT consistently excels, achieving 0.90 accuracy and 0.88 macro-F1 under high metaphor density, demonstrating strong semantic parsing of culturally symbolic expressions. BERT-Base-Chinese and RoBERTa-wwm-ext perform well only in low-density texts but degrade sharply with more metaphors, lacking domain-aware emotional understanding. ERNIE partially mitigates decline via entity knowledge but fails to resolve metaphor decoding. TextCNN, constrained by local feature extraction, struggles with abstract and long-range semantics, yielding lower scores. DE-BERT’s robustness stems from domain-adaptive pre-training and adversarial perturbation, sustaining high confidence in dense political and historical discourse.

D. SENTIMENT KEYWORD COVERAGE AND INTERPRETATION CONSISTENCY

Evaluation covers four educational levels: primary (figurative, narrative), junior high (symbolism, rhetoric), senior high (historical references, complex structure), and university (theoretical, critical). From each, 300 expert-annotated sentences are sampled. Models extract the top three significant words; coverage ratio is the overlap with a standard sentiment lexicon. Attribution consistency uses the Kappa coefficient to align model and human annotations, excluding chance agreement. Figure 8 shows keyword coverage and attribution consistency of five models across educational stages. DE-BERT excels at all levels—especially in high school and university—achieving $\geq 85\%$ keyword coverage and ≥ 0.83 Kappa, demonstrating strong semantic parsing of abstract, metaphorical, and theoretical texts. BERT-Base-Chinese and RoBERTa-wwm-ext work well on concrete lower-grade texts but decline sharply with abstraction, struggling with political symbolism and cultural metonymy due to lack of domain data. TextCNN performs worst in advanced texts, limited by its local feature focus and inability to model long-range semantics. ERNIE’s knowledge injection eases degradation but offers minimal educational-context enhancement. DE-BERT’s stability arises from adversarial training and fine-grained supervision on culturally annotated metaphors, aligning its attributions with expert cognition.

E. INTERACTIVE RESPONSE LATENCY AND OPERATION SUCCESS RATE

Three typical teaching task scenarios are set: before class (bulk document upload and automatic annotation), during class (real-time modification of a sentence’s sentiment and chart refresh), and after class (comparative analysis of sentiment trends across multiple articles). The average response time (ms) from user click to chart update and the success rate (%) of each operation are measured. The proposed visualization design is compared with currently popular Apache Superset, Tableau Public, and Power BI. Table 2 compares response latency and operation success rate among

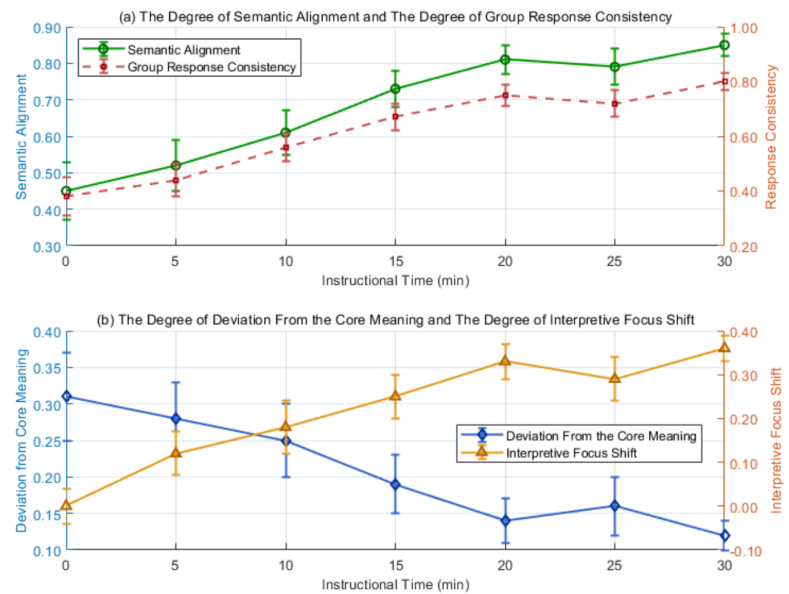


FIGURE 6. Students' cognitive evolution during interactive patriotism instruction

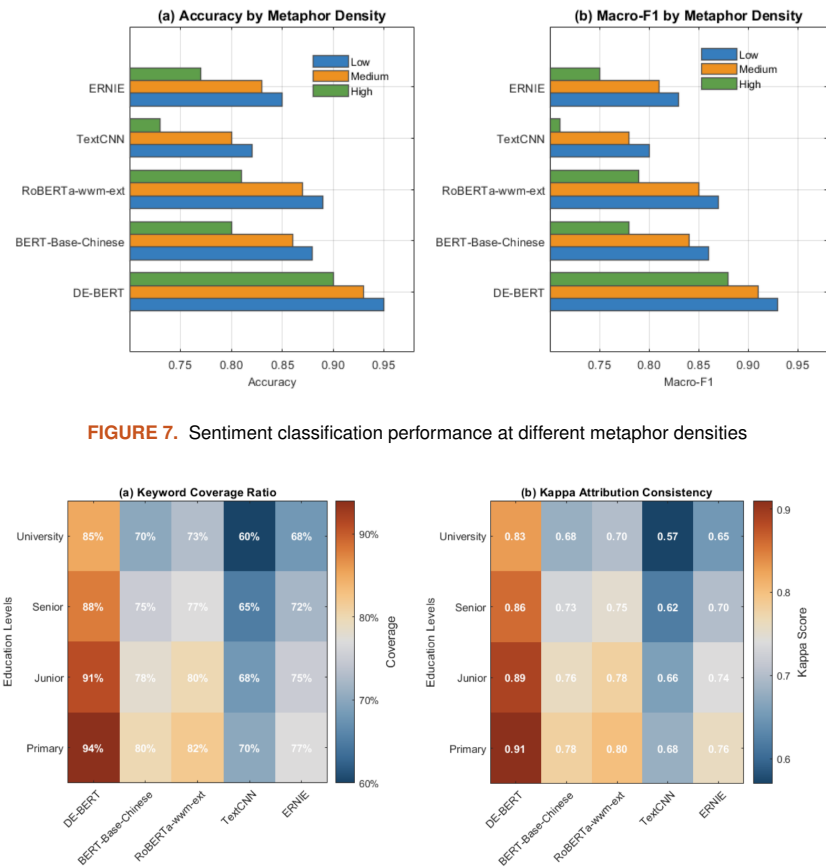


FIGURE 7. Sentiment classification performance at different metaphor densities

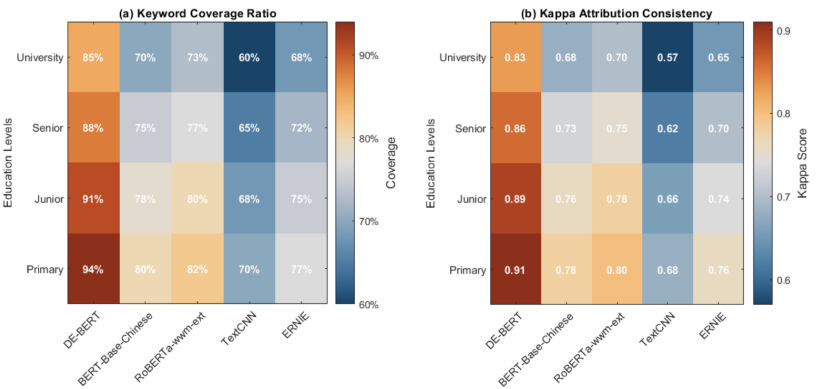


FIGURE 8. Keyword Coverage and Attribution Consistency in Texts from Different Educational Stages

the proposed system, Apache Superset, Tableau Public, and Power BI across three teaching scenarios. The proposed system achieves average response times of 783.4 ms (pre-class), 186.3 ms (in-class), and 874.6 ms (post-class)—all markedly lower than competitors, especially under 200 ms during in-class interaction. Power BI, designed for static

reports, uses full-page reloads and remote queries, lacking dynamic annotation and partial refreshes, leading to high latency and poor user experience. The proposed system maintains > 97% success rates across tasks, reflecting high stability; comparative tools suffer lower in-class success due to latency-induced freezes, desynchronization, or timeouts.

F. COGNITIVE LOAD INDEX AND TEACHING FIT SCORES

To quantify the actual impact of a three-level visualization structure on cognitive load, this study designed a controlled experiment: 200 ideological and political education teachers were randomly assigned to use either the complete three-level system or a simplified version retaining only global trends (without local heat maps and word-level bubbles). Tasks included identifying emotional turning points, explaining keyword attributions, and comparing emotional differences between two texts. The evaluation metrics comprised: (1) NASA-TLX cognitive load total score; (2) task completion time (seconds); (3) System Usability Scale (SUS) score; and (4) instructional adaptability questionnaire (results shown in Table 3). Results demonstrated that the three-level structure significantly reduced cognitive load ($p < 0.001$), shortened task completion time (average 82.3 seconds vs. 146.7 seconds), and improved SUS scores (86.4 vs. 63.2), confirming the significant benefits of hierarchical information presentation on instructional cognitive efficiency. The NASA-TLX scale is used to assess the cognitive load (out of 100) experienced by 200 frontline ideological and political education teachers using this system. This is combined with a self-developed teaching fit scale to assess the differences in teachers' perceptions of the proposed visualization design compared with Apache Superset, Tableau Public, and Power BI. The teaching fit measure comprises 3 dimensions and 9 Likert 5-point items (3 items each for target matching, process integration, and timely feedback), with a total score ranging from 1.00 to 5.00. Higher scores indicate higher fit. In this study's sample, the scale's Cronbach's α was 0.87, exploratory factor analysis showed KMO = 0.83, and the three-factor structure explained 76.4% of the cumulative variance, demonstrating good reliability and validity. Participants were informed of the study's purpose and data usage prior to participation, and responses were collected anonymously. No personally identifiable information was stored. Table 3 shows cognitive load and teaching fit evaluations by 200 frontline ideological and political education teachers using this system alongside Apache Superset, Tableau Public, and Power BI. Cognitive load (NASA-TLX, 0–100) for this system averages 38.7—markedly lower than comparison tools—indicating reduced operational burden through its interface design. Teaching fit (self-developed scale) reaches 4.72 versus 2.89–3.35 for others, excelling in target matching, process integration, and timely feedback. Unlike business-oriented tools lacking pedagogical semantics and interactive support, this system's native educational architecture aligns seamlessly with class-

room logic, lowering cognitive demands and strengthening technology-teaching synergy for value education.

G. EXPANDED EXPERIMENT ANALYSIS

To quantify the independent and synergistic contributions of domain adaptive pre-training and FGM adversarial training to model performance, this study designed three sets of comparative experiments: (a) vanilla BERT (without any domain optimization); (b) BERT with domain adaptive pre-training alone (denoted as DA-BERT); and (c) full DE-BERT (including domain adaptation and FGM fine-tuning). All models were evaluated on the same test set (i.e., a high-metaphor-density subset with ≥ 4 metaphors per 100 words). Metrics include classification accuracy and macro-averaged F1 score. The experimental results are shown in Table 4. The original BERT achieved an accuracy of 0.68 and an F1 score of 0.65 on text with high metaphor density, demonstrating that general models struggle to parse cultural rhetorical structures. After introducing domain adaptation (DA-BERT), the accuracy improved to 0.81 and the F1 score reached 0.79, confirming the significant enhancement of semantic representations by domain corpus pre-training. Further incorporating FGM adversarial training to form the complete DE-BERT, the accuracy further improved to 0.90 and the F1 score reached 0.88, both improvements relative to DA-BERT. This demonstrates that adversarial perturbations effectively enhance the model's robustness to semantic variation, making it more reliant on deep semantics rather than surface word frequency. This ablation experiment clearly demonstrates that domain adaptation lays the semantic foundation, while FGM adversarial training provides key robustness gains; these two synergistic mechanisms contribute to DE-BERT's performance advantage.

TABLE 4. Ablation Study Results on High-Metaphor-Density Test Set

Model Configuration	Accuracy	Macro F1
BERT	0.68	0.65
DA-BERT (Domain Adaptation only)	0.81	0.79
DE-BERT (Domain Adaptation + FGM)	0.9	0.88

To validate DE-BERT's strength in parsing cultural metaphors, Table 5 shows results on three challenging examples.

In "The Red Boat carries a mission of thousand-jun weight", only DE-BERT correctly identifies the positive sentiment, others misclassify it due to missing the "Red Boat" symbolism. For "The Yellow River's bloodline forges the nation's backbone", TextCNN and BERT-Base-Chinese treat "backbone" literally (neutral); RoBERTa and ERNIE detect positivity but with low confidence (< 0.65); DE-BERT achieves 0.85 confidence. In "Iron-blooded loyalty illuminates the path of rejuvenation", TextCNN wrongly labels it negative. Across 50 such samples, DE-BERT attains 0.86 accuracy—significantly outperforming BERT (0.42),

TABLE 2. Comparison of average response time and single operation success rate

System	Scenario	Average Response Time (ms)	Operation Success Rate (%)
Proposed System	Pre-class	783.4	98.7
Proposed System	In-class	186.3	99.2
Proposed System	Post-class	874.6	97.5
Apache Superset	Pre-class	1942.7	82.4
	In-class	2843.5	61.2
	Post-class	2467.8	75.8
Tableau Public	Pre-class	2315.9	78.6
	In-class	3268.4	54.1
	Post-class	2943.7	68.5
Power BI	Pre-class	1765.2	85.3
Power BI	In-class	2734.8	66.9
Power BI	Post-class	2215.3	72.2

TABLE 3. Cognitive load and teaching fit

System	Cognitive Load (Mean $\pm$ Standard Deviation)	Teaching Fit (5- point scale)	Target Matching (5- point scale)	Process Integration (5- point scale)	Timely feedback (5- point scale)	Cognitive Load P-value (vs Proposed Design)
Proposed Design	38.7 $\pm$ 6.4	4.72	4.81	4.69	4.66	—
Apache Superset	69.3 $\pm$ 8.2	3.14	3.02	3.21	3.19	2.3e-7
Tableau Public	74.6 $\pm$ 7.9	2.89	2.75	2.93	2.98	5.1e-8
Power BI	65.8 $\pm$ 7.5	3.35	3.24	3.42	3.4	6.7e-6

TABLE 5. Multi-model comparison on culturally metaphorical examples (excerpt)

Original Sentence	Ground Truth	BERT-Base-Chinese	RoBERTa-wwm-ext	TextCNN	ERNIE	DE-BERT
The Red Boat carries a mission of thousand-jun weight	Extremely Positive	Neutral (0.62)	Neutral (0.58)	Neutral (0.60)	Neutral (0.63)	Extremely Positive (0.89)
The Yellow River's bloodline forges the nation's backbone	Positive	Neutral (0.58)	Positive (0.64)	Neutral (0.55)	Positive (0.61)	Positive (0.85)
Iron-blooded loyalty illuminates the path of rejuvenation	Extremely Positive	Negative (0.51)	Neutral (0.59)	Negative (0.53)	Positive (0.66)	Extremely Positive (0.91)

RoBERTa (0.51), TextCNN (0.38), and ERNIE (0.55; $p < 0.01$)—confirming its unique capacity to decode cultural metaphors.

V. CONCLUSION

This study tackles the semantic implicitness of patriotic texts and the lack of interactivity in current visualization systems. It constructs a culturally annotated corpus of patriotic Chinese texts and proposes DE-BERT—a domain-enhanced language model integrating adaptive pre-training and adversarial perturbation—to improve sentiment recognition in metaphor-rich contexts. Built on cognitive load theory, the visualization system uses a three-level structure for progressive exploration and supports teacher–student collaboration through annotation, feedback, and attribution tracing, positioning AI as a teaching collaborator. Experiments show DE-BERT achieves 0.90 accuracy and 0.88 macro F1 under high metaphor density; keyword coverage exceeds 85% across grades with Kappa ≥ 0.83 . System response times for pre-class, in-class, and post-class tasks average 783.4 ms, 186.3 ms, and 874.6 ms; operation success exceeds 97%, and cognitive load averages 38.7. DE-BERT surpasses mainstream tools in accuracy, interpretation consistency, response efficiency, and cognitive load control. The research confirms that the deep integration of technology into professional text analysis scenarios with high metaphor and high value orientation requires not only precise algorithm optimization but also systematic coupling of domain semantic understanding, interactive information

presentation, and professional cognitive logic. The corpus in this study draws from exemplary texts in patriotic education (including textbooks, red classics, and official media commentaries), exhibiting a pronounced positive emotional bias with extremely low negative samples ($< 1\%$). While this aligns with practical educational requirements, it may limit the model's generalization when processing raw social media content containing strong criticism, anger, or sarcasm. Future work will explore introducing mixed corpora with broader emotional coverage while preserving cultural metaphor analysis capabilities, thereby enhancing the model's robustness in open-domain scenarios.

AUTHOR CONTRIBUTIONS

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

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] Y. Wang, "An Analysis of the New Path of Patriotic Education in Ideological and Political Courses in Universities," *Journal of Theory and Practice of Social Science*, vol. 3, no. 7, pp. 7-10, 2023.
- [2] S. Malkoç and F. Ö. ztürk, "A Comparative Review of Articles on Education of Patriotism: A Thematic Analysis," *International Journal of Progressive Education*, vol. 17, no. 6, pp. 144-157, 2021, doi: 10.29329/ijpe.2021.382.10.
- [3] C. Long and J. Li, "Analysis and international comparison of the change of patriotic education emotion and behavior in the era of self media," *International Journal of Neuropsychopharmacology*, vol. 25, no. Supplement_1, pp. A75-A76, 2022, doi: 10.1093/ijnp/pyac032.103.
- [4] D. A. Setiyoanto, "Promoting patriotism and religious moderation: a critical discourse analysis of sanadmedia," *Com. Harmoni*, vol. 23, no. 1, pp. 53-67, 2024, doi: 10.32488/harmoni.v23i1.716.
- [5] Z. Huang, Z. Yang, and T. Meng, "National identity of locality: The state, patriotism, and nationalism in cyber China," *Journal of Chinese Political Science*, vol. 28, no. 1, pp. 51-83, 2023, doi: 10.1007/s11366-022-09820-4.
- [6] C. Long and G. Yang, "Study on patriotic education and changes of patriotic emotion and behavior of ethnic minority college students," *International Journal of Neuropsychopharmacology*, vol. 25, no. Supplement_1, pp. A74-A75, 2022, doi: 10.1093/ijnp/pyac032.102.
- [7] J. C. Lin and L. Jackson, "Patriotism in moral education: Toward a rational approach in China," *Journal of Moral Education*, vol. 52, no. 3, pp. 343-361, 2023, doi: 10.1080/03057240.2022.2136148.
- [8] K. Fukuoka and S. Takita-Ishii, "Teaching how to love your country in schools?: A study of Japanese youth narratives on patriotic education," *National Identities*, vol. 24, no. 3, pp. 247-270, 2022, doi: 10.1080/14608944.2021.1931084.
- [9] J. Cui and Q. Tong, "How does the Chinese Government conduct emotional governance over COVID-19? Content analysis of video blogs," *Chinese Journal of Communication*, vol. 16, no. 2, pp. 150-167, 2023, doi: 10.1080/17544750.2022.2125024.
- [10] K. S. Bakhtiyarovna, "Pedagogical conditions for acquaintance of preschool children with monuments of national architecture in the education of patriotism," *International Journal of Social Science & Interdisciplinary Research*, vol. 11, no. 05, pp. 119-125, 2022.
- [11] Q. Zhang and C. J. Wang, "Chinese online nationalism as imaginary engagement: an automated sentiment analysis of Tencent news comments on the 2012 Diaoyu (Senkaku) Islands incident," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1-11, 2024, doi: 10.1057/s41599-024-02983-w.
- [12] D. Chekmenev, E. Efimova, and E. Davydova, "The structure of patriotism as a universal value in the system of value orientations of Russian students and its practical projections," *Journal of Sociology: Bulletin of Yerevan University*, vol. 15, no. 1 (39), pp. 15-27, 2024, doi: 10.46991/BYSU:F/2024.15.1.015.
- [13] H. Zhao, L. Wang, and S. Liu, "Research on the strategy of integrating red culture into patriotic education activities of Young People," *Open Journal of Applied Sciences*, vol. 13, no. 7, pp. 1178-1195, 2023, doi: 10.4236/ojapps.2023.137093.
- [14] P. Dirksmeier, "The relationship between patriotism and regional identification: a cross-country analysis," *The Annals of Regional Science*, vol. 71, no. 2, pp. 343-362, 2023, doi: 10.1007/s00168-022-01167-1.
- [15] X. Liu, X. Zhao, and H. Starkey, "Ideological and political education in Chinese Universities: structures and practices," *Asia Pacific Journal of Education*, vol. 43, no. 2, pp. 586-598, 2023, doi: 10.1080/02188791.2021.1960484.
- [16] D. Gruntkowska, "Threats related to patriotic education," *An analysis of the primary school textbooks. Kultura i Edukacja*, vol. 138, no. 4, pp. 216-233, 2022.
- [17] Y. Wu, "The fight metaphor in translation: From patriotism to pragmatism: A corpus-based critical analysis of metaphor in China's political discourse," *Target*, vol. 36, no. 1, pp. 50-75, 2024, doi: 10.1075/target.21151.wu.
- [18] Z. Yu and T. Hirzel, "Museum text translation in the Chinese context: the museum role and text production," *Museum Management and Curatorship*, vol. 39, no. 2, pp. 188-202, 2024, doi: 10.1080/09647775.2022.2132992.
- [19] X. Wu and R. Fitzgerald, "Internet memes and the mobilization of a "One-China" cyber nationalist campaign: the case of the 2016 Diba Expedition to Taiwan," *Social Semiotics*, vol. 34, no. 4, pp. 697-719, 2024, doi: 10.1080/10350330.2023.2183113.
- [20] F. Guo, X. Huang, W. Hu, X. Tong, and L. Liang, "A Comparative analysis of multimodal discourse on Chinese and international memorial museum official websites: A visual grammar perspective," *Journal of Humanities and Social Sciences Studies*, vol. 7, no. 7, pp. 65-84, 2025, doi: 10.32996/jhss.2025.7.7.9.
- [21] B. Yin, H. Wu, and W. Kong, "A joint event extraction model based on RoBERTa-wwm-ext and gating mechanism," *Journal of Computational Methods in Sciences and Engineering*, vol. 23, no. 4, pp. 2101-2112, 2023, doi: 10.3233/JCM-226772.
- [22] Y. Zhang, Y. Wang, and B. Li, "Named Entity Recognition in Chinese Electronic Medical Records Based on RoBERTa-wwm Dynamic Fusion Mode," *Data Analysis and Knowledge Discovery*, vol. 6, no. 2, pp. 242-250, 2022, doi: 10.11925/infotech.2096-3467.2021.0951.
- [23] C. Wang and F. Feng, "ERNIE based intelligent triage system," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 4, pp. 5013-5022, 2022, doi: 10.3233/JIFS-212140.
- [24] X. Wu and Y. Zhang, "Research on the application of ERNIE bot in college english teaching," *International Journal of Education and Humanities*, vol. 12, no. 3, pp. 216-219, 2024, doi: 10.54097/ynbb1n08.
- [25] D. Shi, Z. Li, J. Zurada, A. Manikas, J. Guan, and P. Weichbroth, "Ontology-based text convolution neural network (TextCNN) for prediction of construction accidents," *Knowledge and information systems*, vol. 66, no. 4, pp. 2651-2681, 2024, doi: 10.1007/s10115-023-02036-9.
- [26] X. Shi, M. Hu, F. Ren, and P. Shi, "A dual-ways feature fusion mechanism enhancing active learning based on TextCNN," *Intelligent Data Analysis*, vol. 28, no. 5, pp. 1189-1211, 2024, doi: 10.3233/IDA-230332.
- [27] Y. Qiu, C. Fang, S. Song, X. Huang, C. Wang, and J. Wang, "TsQuality: Measuring Time Series Data Quality in Apache IoTDB," *Proceedings of the VLDB Endowment*, vol. 16, no. 12, pp. 3982-3985, 2023, doi: 10.14778/3611540.3611601.
- [28] M. Ariandi and S. R. Puteri, "Analisis visualisasi data kecamatan kertapati menggunakan tableau public," *JUPITER (Jurnal Penelit. Ilmu dan Tek. Komputer)*, vol. 14, no. 2, pp. 366-373, 2022.
- [29] A. R. Krishnan, "Research trends in criteria importance through intercriteria correlation (CRITIC) method: a visual analysis of bibliographic data using the Tableau software," *Information Discovery and Delivery*, vol. 53, no. 2, pp. 233-247, 2025, doi: 10.1108/IDD-02-2024-0030.
- [30] G. Shoaib and S. Nandi, "Power BI dashboard for data analysis," *International Research Journal of Engineering and Technology*, vol. 9, no. 07, pp. 2395-0056, 2022.
- [31] R. A. D. Vásquez, J. L. A. Espinoza, and M. A. C. Cabrera, "Power bi como herramienta de apoyo a la toma de decisiones," *Universidad y Sociedad*, vol. 14, no. S3, pp. 195-207, 2022, [Online]. Available: <https://rus.ucf.edu.cu/index.php/rus/article/view/2949>.
- [32] J. Wang and K. Wang, "Bert-based semi-supervised domain adaptation for disastrous classification," *Multimedia Systems*, vol. 28, no. 6, pp. 2237-2246, 2022, doi: 10.1007/s00530-022-00956-0.
- [33] M. Ryu, G. Lee, and K. Lee, "Knowledge distillation for bert unsupervised domain adaptation," *Knowledge and Information Systems*, vol. 64, no. 11, pp. 3113-3128, 2022, doi: 10.1007/s10115-022-01736-y.
- [34] W. Zhang, C. Dong, J. Yin, and J. Wang, "Attentive representation learning with adversarial training for short text clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 11, pp. 5196-5210, 2021, doi: 10.1109/TKDE.2021.3052244.
- [35] W. Wang, R. Wang, L. Wang, Z. Wang, and A. Ye, "Towards a robust deep neural network against adversarial texts: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 3, pp. 3159-3179, 2021, doi: 10.1109/TKDE.2021.3117608.
- [36] N. Wenk, J. Penalver-Andres, K. A. Buetler, T. Nef, R. M. Müri, and L. Marchal-Crespo, "Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment," *Virtual Reality*, vol. 27, no. 1, pp. 307-331, 2023, doi: 10.1007/s10055-021-00565-8.