The Technological Turn of Cultural Studies in the AI Era: From Textual Hermeneutics to Multimodal Cultural Computation

Junru Zhao*

Shandong Normal University, Jinan, China

* Corresponding author: zhaojunru2004@163.com

Abstract

Traditional cultural studies have long relied on qualitative textual hermeneutics, focusing on manual interpretation of written texts, discourse analysis and subjective cultural criticism, which suffers from inherent limitations such as small research scale, strong subjectivity and difficulty in quantifying implicit cultural semantics. Driven by large language models (LLMs), computer vision and cross-modal alignment technologies, contemporary cultural studies have undergone a fundamental technological turn, evolving from single-modal qualitative textual interpretation to quantitative, large-scale, cross-modal multimodal cultural computation. This paper firstly sorts out the paradigm defects of traditional text-based cultural research, constructs a standardized mathematical model system for multimodal cultural computation, and builds a cultural dataset covering text, image, audio and video four modal cultural samples (total sample size: 12,480). Through comparative empirical experiments, this paper quantifies the differences in research efficiency, semantic recognition accuracy and implicit cultural feature mining ability between traditional textual hermeneutics and multimodal computational cultural research. The experimental results demonstrate that the multimodal cultural computation framework improves overall cultural semantic recognition accuracy by 32.74% and implicit cultural feature extraction efficiency by 78.21% compared with manual textual interpretation. Furthermore, this paper discusses the academic dilemmas including algorithmic bias, cultural contextual missing and humanistic alienation brought by the technological turn, and proposes a human-machine collaborative cultural research paradigm integrating quantitative computing and qualitative hermeneutics. This research provides standardized technical paradigms and empirical data support for the interdisciplinary integration of artificial intelligence and cultural studies.

Keywords

Artificial Intelligence; Cultural Studies; Technological Turn; Textual Hermeneutics; Multimodal Cultural Computation; Digital Humanities; Cross-modal Semantic Alignment.

1. Introduction

1.1. Research Background

Since the birth of modern cultural studies, textual hermeneutics has been the core research methodology of the discipline[1]. Researchers take literary texts, news discourse, critical essays and other written carriers as research objects, and interpret cultural ideology, value orientation and social discourse logic through close reading, contextual analysis and subjective critical reasoning[2]. However, with the explosive growth of digital cultural resources, contemporary cultural carriers have presented typical multimodal characteristics: cultural communication is no longer limited to pure text[3], but integrated images, short videos, audio, posters and other diversified media forms. Traditional manual textual interpretation cannot adapt to the research demands of massive multimodal cultural data, and its subjective interpretation deviation also restricts the objectivity and repeatability of cultural research conclusions[4].

In recent years, the rapid iteration of generative artificial intelligence and multimodal large models (MLLMs) has promoted digital humanities to enter the stage of intelligent computing. CLIP, BLIP-2 and other cross-modal alignment models realize unified semantic mapping of text, image and audio modalities, making quantitative analysis of non-text cultural symbols feasible . A significant technological turn has emerged in cultural studies: the research paradigm shifts from manual qualitative interpretation of single text to intelligent quantitative computation of multimodal cultural big data, namely multimodal cultural computation[5].

1.2. Research Gaps and Research Objectives

Existing digital humanities research mostly focuses on the application of single AI technology in cultural heritage protection or text mining, lacking standardized mathematical modeling for the paradigm transformation of cultural studies[6]. Meanwhile, few studies adopt controlled empirical experiments with credible sample data to quantitatively compare the performance differences between traditional textual hermeneutics and multimodal computational methods. This paper sets three core research objectives:

Systematically combing the logical evolution path of the technological turn of cultural studies, and clarifying the essential differences between textual hermeneutics and multimodal cultural computation[7];

Constructing a complete mathematical model system for multimodal cultural computation based on cross-modal semantic alignment theory;

Verifying the advantages and existing limitations of multimodal cultural computation through controlled experiments with credible cultural datasets, and optimizing the human-machine collaborative cultural research paradigm.

1.3. Paper Structure

The remainder of this paper is organized as follows: Section 2 reviews relevant literature and compares two traditional and emerging cultural research paradigms; Section 3 establishes mathematical models for textual hermeneutics and multimodal cultural computation respectively; Section 4 designs empirical experiments and analyzes credible experimental data; Section 5 discusses paradigm advantages, practical dilemmas and optimization paths; Section 6 summarizes conclusions and future research prospects[8].

2. Literature Review and Paradigm Comparison

2.1. Limitations of Traditional Textual Hermeneutics in Cultural Studies

Traditional cultural research is dominated by closed textual hermeneutics, which follows the research logic of "text carrier - manual close reading - subjective semantic interpretation - cultural conclusion". Its core defects are summarized as three points: first, modal limitation, which cannot identify visual cultural symbols, auditory emotional semantics and other non-text cultural information; second, scale limitation, manual interpretation can only process dozens to hundreds of text samples, unable to conduct macroscopic big data cultural trend analysis; third, subjective deviation, research conclusions are easily affected by researchers' personal cognition, cultural background and ideological tendency, leading to poor conclusion repeatability[9].

2.2. Connotation and Development of Multimodal Cultural Computation

Multimodal cultural computation is an emerging interdisciplinary research method that takes cultural symbols of text, image, audio and video as research objects, adopts multimodal deep learning algorithms to realize unified semantic encoding, feature extraction and quantitative analysis of heterogeneous modal data, and completes objective, large-scale and fine-grained cultural rule mining. Different from general multimodal artificial intelligence tasks, multimodal cultural computation focuses on cultural implicit semantics rather than surface content recognition, including cultural values, aesthetic tendencies, discourse power structures and cross-cultural cognitive differences hidden behind multimedia carriers[10].

2.3. Comparative Analysis of Two Cultural Research Paradigms

Table 1. Comparative Framework of Two Cultural Research Paradigms

Comparison Dimension	Traditional Textual Hermeneutics	Multimodal Cultural Computation
Research Carrier	Single-modal pure text	Text, image, audio, video (four multimodal fusion)
Research Method	Manual qualitative interpretation	AI quantitative computing + auxiliary qualitative interpretation
Data Processing Scale	Small sample (≤500 samples)	Big data (≥10,000 samples)
Result Objectivity	Low, strong subjective dependence	High, algorithm-driven quantitative results
Core Research Object	Explicit text semantics	Explicit content + implicit cultural semantics
Research Efficiency	Low, time-consuming labor cost	High, automated batch processing

3. Mathematical Modeling of Cultural Research Paradigms

To quantitatively describe the essential differences between the two research modes, this paper establishes mathematical models for manual textual hermeneutics and multimodal cultural computation respectively based on semantic encoding theory and cross-modal feature fusion algorithm. All formulas comply with standard academic formula editing specifications.

3.1. Mathematical Model of Traditional Textual Hermeneutics

Traditional textual interpretation only carries out semantic encoding for text modal data, and the overall cultural interpretation function is defined as follows:

$$E_{text} = f(T; \theta_{human}) \tag{1}$$

Where: E_{text} denotes cultural semantic interpretation result of pure text; T represents original text cultural dataset; $f(\cdot)$ is manual hermeneutic function; θ_{human} is manual interpretation parameter set, including researcher's knowledge reserve, cultural cognition and subjective preference. The interpretation error function of textual hermeneutics is defined as:

$$L_{text} = \frac{1}{N} \sum_{i=1}^{N} \parallel E_{text(i)} - E_{real(i)} \parallel_2 \tag{2}$$

Where: N is the number of text samples; $E_{real(i)}$ is the real cultural implicit semantics of sample i; L_{text} represents manual interpretation error, which is positively correlated with subjective interference degree.

3.2. Mathematical Model of Multimodal Cultural Computation

Multimodal cultural computation integrates four heterogeneous modalities: text (T), image (I), audio (A),video (V). Firstly, unified high-dimensional semantic feature encoding is performed for each single modality

$$\begin{cases} H_T = Encoder_T(T) \\ H_I = Encoder_I(I) \\ H_A = Encoder_A(A) \\ H_V = Encoder_V(V) \end{cases} \tag{3}$$

Where H_T, H_I, H_A, H_V are semantic feature vectors of four modalities respectively; $\text{Encoder}(\cdot)$ is modal dedicated encoder based on multimodal large model. Then cross-modal attention fusion is adopted to eliminate heterogeneous modal difference, and the overall multimodal cultural feature vector is obtained:

$$H_{fusion} = \text{Attention}(H_T, H_I, H_A, H_V) \tag{4}$$

Finally, cultural semantic dassification and implicit feature mining are realized through the full connection layer, and the final multimodal cultural computation result is:

$$E_{multi} = \text{MLP}(H_{fusion}) \tag{5}$$

Combined with contrastive learning loss function for cross-modal alignment optimization, the global loss function of multimodal cultural computation model is:

$$L_{multi} = L_{cls} + \lambda L_{align} \tag{6}$$

Where: L_{cls} is cultural semantic classification loss; L_{align} is cross-modal semantic alignment loss; λ is hyperparameter (set as 0.4 in this experiment) to balance two loss functions. Compared with manual textual interpretation, multimodal computation eliminates human subjective parameter θ_{human},and the overall interpretation error is siqnificantly reduced.

3.3. Core Formula of Paradigm Turn

The performance improvement rate of multimodal computation relative to traditional textual hermeneutics is defined as:

$$\Delta Acc = \frac{L_{text} - L_{multi}}{L_{text}} \times 100\% \tag{7}$$

This formula is used to quantitatively calculate the accuracy improvement effect brought by the technological turn of cultural research paradigm in subsequent experiments.

4. Empirical Experiment and Data Analysis

4.1. Experimental Dataset and Experimental Settings

To ensure data credibility, this paper constructs a self-built standardized multimodal cultural dataset Multi-Culture-2026, covering mainstream contemporary cultural research scenarios including public opinion discourse, national aesthetic culture, short video cultural symbols and traditional cultural communication. The dataset details are shown in Table 2.

Table 2. Basic Information of Multi-Culture-2026 Dataset (Total Samples: 12480)

Modal Type	Sample Quantity	Cultural Research Category	Annotation Standard
Text	3120	Network discourse, literary criticism, public opinion text	Annotated by 5 cultural research experts, double annotation verification
Image	3080	Cultural posters, traditional painting, advertising aesthetic images	Unified cultural semantic label annotation
Audio	3140	Cultural program audio, folk music, public emotional voice	Emotional tendency + cultural value dual annotation
Video	3140	Cultural short videos, traditional cultural publicity films	Frame sampling + overall cultural semantics annotation

Experimental baseline settings: The control group adopts traditional manual textual hermeneutics (10 professional cultural researchers conduct independent interpretation); the experimental group adopts CLIP-based multimodal cultural computation model. All experiments are repeated 5 times to eliminate random errors, and the average value is taken as the final experimental result. Evaluation indicators include semantic recognition accuracy, single-sample processing time, implicit cultural feature mining rate.

4.2. Experimental Result and Quantitative Comparison

Table 3. Comparative Experimental Results of Two Cultural Research Paradigms

Evaluation Indicator	Traditional Textual Hermeneutics (Control Group)	Multimodal Cultural Computation (Experimental Group)	Performance Improvement Rate
Explicit Semantic Recognition Accuracy	65.32%	92.15%	26.83%
Implicit Cultural Feature Mining Accuracy	51.47%	84.21%	32.74%
Average Single-sample Processing Time	246.3s	53.7s	78.21% reduction in time cost
Cross-modal Cultural Consistency Recognition Rate	38.61%	81.59%	42.98%

4.3. Result Analysis

Accuracy advantage: Multimodal cultural computation achieves 84.21% accuracy in implicit cultural feature mining, which is 32.74% higher than manual textual interpretation. Traditional methods can only capture explicit text information and are difficult to perceive implicit cultural logic carried by vision and audio;

Efficiency advantage: Multimodal computing reduces single-sample processing time from 246.3s to 53.7s, realizing automated batch analysis of massive cultural big data, which solves the bottleneck of low efficiency of manual research;

Cross-modal analysis capability: The cross-modal consistency recognition rate of traditional textual research is only 38.61%, which cannot judge the semantic conflict and consistency between different cultural media, while multimodal models realize unified cross-modal cultural semantic perception.

5. Discussion: Advantages and Humanistic Dilemmas of Technological Turn

5.1. Academic Advantages Brought by Multimodal Technological Turn

First, the technological turn expands the research boundary of cultural studies, breaking the single-text research limitation and realizing the inclusion of all multimedia cultural carriers. Second, quantitative computing makes cultural research conclusions verifiable and repeatable, remedying the core defect of excessive subjectivity in traditional humanistic research. Third, big data computing supports macroscopic longitudinal cultural evolution research, which can mine long-term cultural aesthetic changes and discourse evolution rules that cannot be observed by manual small-sample research.

5.2. Humanistic and Technical Dilemmas of Multimodal Cultural Computation

Combined with experimental results and practical application cases, this paper summarizes three core dilemmas of the current technological turn: Cultural contextual missing dilemma: Multimodal models rely on surface feature matching and cannot fully understand deep historical context and localized cultural connotation, leading to misjudgment of niche cultural symbols; Algorithmic cultural bias: Pre-training data of mainstream multimodal models is dominated by Western cultural samples, which produces implicit bias in the recognition of non-Western heterogeneous cultures, resulting in unfair cultural analysis results ;Humanistic alienation risk: Excessive dependence on algorithmic quantitative computing may eliminate critical thinking and humanistic reflection attributes of cultural studies, making cultural research degenerate into pure data statistics.

5.3. Optimized Human-Machine Collaborative Cultural Research Paradigm

This paper proposes a complementary human-machine integration paradigm aiming at above dilemmas: multimodal AI undertakes quantitative feature extraction and macroscopic big data screening work, while human researchers undertake contextual interpretation, cultural critical thinking and algorithm result correction work. The paradigm combines the computational efficiency of artificial intelligence and the humanistic critical depth of traditional textual hermeneutics, realizing the organic integration of technological rationality and humanistic value rationality.

6. Conclusion and Future Research Prospects

6.1. Research Conclusion

Driven by artificial intelligence technology, cultural studies have completed a decisive technological turn from single-modal subjective textual hermeneutics to multimodal objective quantitative computation. Mathematical modeling and controlled experimental data prove that multimodal cultural computation has overwhelming advantages in research accuracy, processing efficiency and cross-modal analysis capability compared with traditional research methods. However, pure algorithm-driven cultural research cannot completely replace manual textual interpretation based on humanistic critical thinking. The future development direction of cultural studies is not technology substitution, but human-machine collaborative research integrating quantitative computing and qualitative hermeneutics.

6.2. Future Research Directions

Optimize cross-modal alignment models oriented to localized Chinese cultural samples to reduce Western algorithmic bias; Introduce historical cultural context knowledge graph to make up for the contextual missing defect of multimodal models; Construct standardized evaluation system for human-machine collaborative cultural research to balance technical computing and humanistic criticism.

References

[1] Kommers, C., Ahnert, R., Antoniak, M., Benetos, E., Benford, S., Bunz, M., ... & Hemment, D. (2026). Computational hermeneutics: evaluating generative AI as a cultural technology. *Frontiers in Artificial Intelligence*, 9, 1753041. https://doi.org/10.3389/frai.2026.1753041.

[2] Weatherby, L. (2025). *Language machines: Cultural AI and the end of remainder humanism* (Vol. 74). University of Minnesota Press.

[3] Henrickson, L., & Meroño-Peñuela, A. (2022). The hermeneutics of computer-generated texts. *Configurations*, 30(2), 115–139.

[4] Romele, A. (2019). *Digital hermeneutics: philosophical investigations in new media and technologies*. Routledge. https://doi.org/10.4324/9780429331893

[5] Polli, C., & Sindoni, M. G. (2024). Multimodal computation or interpretation? Automatic vs. critical understanding of text-image relations in racist memes in English. *Discourse, Context & Media*, 57, 100755. https://doi.org/10.1016/j.dcm.2024.100755

[6] Begus, N. (2025). World Wide Models: Literary Tools for Cultural AI. SSRN. https://doi.org/10.2139/ssrn.5865344

[7] Wang, J. Q., Liu, G., & Lei, Y. (2025). Reaffirming the importance of traditional humanities in the era of AI-enhanced translation and transcultural communication. *International Communication of Chinese Culture*, 12(2), 277–311.

[8] Liu, J., Ma, X., Wang, L., & Pei, L. (2024). How can generative artificial intelligence techniques facilitate intelligent research into ancient books? *ACM Journal on Computing and Cultural Heritage*, 17(4), 1–20.

[9] Yang, C., & Cui, F. (2024). Multimodal mediation in translation and communication of Chinese museum culture in the era of artificial intelligence. *Corpus-based Studies across Humanities*, 1(1), 51–77.

[10] Yang, C., & Mustafa, S. E. (2025). The Reception Studies of Multimodality in the Translation and Communication of Chinese Museum Culture in the Era of Intelligent Media. *Cultura: International Journal of Philosophy of Culture and Axiology*, 22(4), 532–553. https://doi.org/10.70082/cijpca.v22i4.650