ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Market Applications of AIGC-Generated Content on New Media Short Video Platforms and Their Impact on Brand Communication

Chang Shi[1,*]

[1] Wuxi Vocational College of Science and Technology, Wuxi, Jiangsu, 214028, China

SUMMARY: *In order to solve the problem of insufficient brand communication efficiency and content adaptation in new media short video platform, this paper constructed an AIGC short video generation and market application analysis framework integrating text, visual and audio information. Multi-modal features are extracted from brand short video samples, comment corpus, interaction log and product attributes. After text segmentation, brand semantic cleaning, audio and video noise reduction and principal component compression, a content generation network and a communication effect prediction module under brand constraints are established, which is used to realize the integrated calculation of short video generation, brand recognition and market feedback evaluation. The experimental results show that the semantic consistency of the generated content reaches 0.87, the brand information recognition accuracy is 90.8%, the user interaction composite index is increased from 0.71 to 0.81, and the brand communication effect improvement rate is increased from 12.4% to 21.7%. Although the training time was increased from 126 min to 139 min, the inference speed reached 26.4 frame·s⁻¹, which was still feasible for deployment. The results show that AIGC not only improves the production efficiency of brand short video, but also enhances the controllability of brand expression and communication revenue.*

KEYWORDS: *AIGC; Short video platform; Brand communication; Multi-modal generation*

1 Introduction

The continuous improvement of mobile Internet infrastructure, the algorithmic evolution of short video platform distribution mechanism, and the rapid embedding of generative artificial intelligence technology in content production side have jointly promoted the reorganization of new media communication ecology. Compared with image-text communication, short video has stronger comprehensive expression ability in time compression, emotional mobilization, visual arousal and scene substitution. Therefore, it has gradually become an important medium for brands to reach users, shape cognition and promote transformation. At the same time, the total amount of content supply on the platform has increased exponentially, and brand communication has also shifted from the relatively stable "one-way delivery" in the past to the operation mode of high-frequency iteration, dynamic feedback and multi-round interaction in parallel. In this environment, it is difficult to meet the communication needs of brands in different user segments, different consumption scenarios and different platform mechanisms by relying only on manual script writing, fixed template editing and empirical delivery optimization.

The introduction of AIGC generated content makes short video brand communication

*elaine.shi@163.com
https://doi.org/10.65102/is2026708 Published: 30 April, 2026

further move from "content production" to "content calculation". It not only improves the efficiency of copywriting, image generation, and voice synthesis, but more importantly, it connects natural language processing, computer vision, speech signal processing, and recommendation algorithms into a trainable, iterative, and deployable digital content production chain. According to product attributes, user interest tags, comment emotional distribution, historical interaction behavior and platform hot word changes, the brand side can linkage generate video script, shot organization, subtitle style, background music and virtual character expression, so that the communication content maintains high consistency in semantic layer, visual layer and emotional layer. Therefore, short video is no longer just a media product that is "released", but an intelligent communication unit that can continuously adjust the expression mode according to market feedback.

However, the market application of AIGC in the new media short video platform is not just a problem of technical substitution. Brand communication has clear information constraints, value orientation and identification boundaries. If the generated content deviates from the brand positioning in semantics, weakens the brand symbol in vision, and excessively pursues the platform traffic preference in narrative, the expansion of communication scale may bring brand cognition dilution, authenticity decline and even trust damage. Therefore, how to improve the efficiency of content generation while maintaining the stability of brand expression, how to ensure the identifability of brand key information in the process of multi-modal automatic generation, and how to reverse evaluate the market suitability of AIGC content from user interaction data have become the questions that need to be answered positively in platform marketing research.

Based on this, this paper discusses AIGC short video generation and brand communication evaluation under a unified analysis framework, and does not regard it as a simple creative tool, but defines it as a computational mechanism that integrates text modeling, visual understanding, audio synthesis, communication prediction and market feedback analysis. Focusing on the application scenario of the new media short video platform, this paper attempts to build a research link of "data acquisition - preprocessing - multi-modal feature extraction - content generation - communication effect evaluation", in order to investigate the actual performance of AIGC generated content in brand semantic maintenance, information recognition, user interaction improvement and communication performance optimization. The characteristics of time cost, reasoning efficiency and resource consumption in platform deployment were further analyzed.

1.1 Research Significance

The research significance of this paper is mainly reflected in two levels of theoretical understanding and technical application. From the perspective of communication research, short video brand communication is shifting from the human-centered creative production mode to the content generation mode driven by human-machine collaboration. Brand information is no longer only expressed through artificial experience, but constantly revised through model training, feature learning and feedback update. A systematic review of this change can help to explain the tension between "content scale" and "brand consistency" in platform communication, and also help to enrich the research perspective at the intersection of digital marketing, intelligent communication and brand communication.

From the perspective of computer applications, the value of AIGC in the short video market is not only shown by its fast generation speed, but also by its ability to handle complex multimodal inputs. At the text end, brand core words and emotional tendencies can be extracted through word segmentation, embedding representation and semantic cleaning. The visual side can capture product elements and screen styles by means of object detection, scene

recognition and key frame coding. The audio side can enhance the auditory recognition of content through noise reduction, prosodic analysis and speech synthesis. On this basis, the multi-modal features are fed into the content generation network and the communication effect prediction module to form a closed-loop intelligent system for brand communication tasks. Such research can not only provide more operational content generation and delivery scheme for enterprises, but also provide verifiable technical path for AIGC's engineering realization in advertising marketing, brand operation and platform commercialization.

1.2 Research content and article structure

Focusing on the above problems, this paper takes the brand communication scenario in the new media short video platform as the research object, and focuses on the market application mechanism of AIGC generated content and its communication effect. The structure of the whole paper follows the logical expansion from problem raising to model construction, and then to experimental verification and result discussion. The second part combs the related research of short video marketing, AIGC content production and brand communication evaluation to clarify the shortcomings of existing research in the connection between multi-modal generation and market application. Section III introduces the methodological framework of this paper, including data set construction, text and audio and video preprocessing, multi-modal feature extraction, and AIGC short video generation and market application analysis model for brand communication evaluation. Combined with the experimental results, the fourth part analyzes the semantic consistency of the generated content, the accuracy of brand information recognition, user interaction performance, the improvement of communication effect and model training time. In the discussion section, the influence of reasoning efficiency and resource consumption on the actual deployment is comprehensively explained. Section 5 summarizes the research, summarizes the main findings, and points out the future work in the aspects of authenticity control, platform adaptation and model optimization.

2 Literature Review

Focusing on the issue of brand communication in new media short video platforms, existing research can be roughly summarized into two categories. One category focuses on the marketing mechanism of short video platform itself, focusing on the relationship between content style, interactive form, platform algorithm and consumption response. The other category starts from AI marketing and examines the application paths of natural language processing, machine learning, virtual digital people and generative models in the production of advertising content. The former provides the scene basis for understanding the short video brand communication, while the latter provides technical support for AIGC embedding into the content industry chain. However, there are still few researches linking the two together.

In the research of short video marketing, scholars generally believe that the characteristics of duration compression, emotional reinforcement and algorithm distribution of platform content have changed the way brands reach users. Starting from the return visit behavior of the platform, Cuesta-Valino et al. [1] pointed out that the continuous attraction of short video platform is closely related to user immersion experience, content matching degree and interactive convenience. Deng et al. [2] and Li et al. [3] respectively revealed the influence of short video narrative on consumption judgment from the perspectives of the video of celebrities with goods and the implicit advertisement expression. Barta et al. [4] and Dong et al. [5] further show that humorous expression, contextualized narrative and brand short video

rhythm design can significantly affect user engagement and brand proximity. Rizomyliotis et al. [6] and Alcantara-Pilar et al. [7] conducted research on the consumption tendency of Generation Z and the credibility of Internet celebrities in the TikTok environment, indicating that platform interaction data is no longer just a communication outcome variable, but a computable signal that can enter the optimization process of brand communication in reverse.

Correspondingly, AI marketing research pays more attention to the automatic generation of content, semantic understanding and dissemination prediction. Haleem et al. [8] summarized the application of AI in marketing as personalized recommendation, audience segmentation, content creation and delivery optimization. Sun et al. [9] proved the effectiveness of natural language processing in advertising evaluation and consumer opinion recognition, which provided a methodological basis for brand review mining, emotion recognition and communication feedback modeling. Malthouse and Copulsky[10], He and Zhang[11] proposed that AI touchpoints are reshaping the customer journey, and brand communication no longer relies on a single advertising material, but realizes dynamic communication through multi-node collaboration driven by algorithms. After entering the generative stage, Gao et al. [14] and Chen et al. [16] began to directly discuss the attractiveness, persuasion mechanism and ethical boundaries of AI-generated advertising. Bruns and Meißner[15] remind that although generative content improves efficiency, it may weaken the perception of brand authenticity. Combined with Table 1, it can be found that although the existing artificial intelligence marketing research has involved the automated production of content, most of the results still stay at the level of text advertising, recommendation optimization or consumer attitude, and the discussion of multi-modal generation mechanism in short video scenarios is still insufficient.

In recent years, the research on virtual digital and virtual influencers has further promoted this issue. Kim et al. [17], Koles et al. [18], Looi and Kahlor[19], Liu and Lee[20], and Volles et al. [21] have shown that virtual agent can improve novelty and memory in some product situations, but its communication effect is not stable. Still subject to product type, personification degree and authenticity matching constraints. That is to say, AIGC does not naturally improve the brand communication effect, the key is not "whether it can be generated", but "whether the generated content maintains the brand semantic consistency, whether it is effectively recognized by platform users, and whether it forms positive interaction". According to the comparison results in Table 1, existing research still pays insufficient attention to brand authenticity, recognition accuracy and platform deployment cost, which also means that there is still a large space for deepening the market application research of AIGC short video.

Table 1: Comparison of short video brand communication and AIGC marketing related studies

Research Direction	Representative References	Computer-Related Focus	Main Findings	Existing Limitations
User Behavior and Interaction on Short Video Platforms	[1][2][4][5]	Algorithmic content distribution, interaction data analysis, user profiling	Content style and interaction mechanisms significantly affect engagement and conversion	Discussion of generative content mechanisms remains insufficient
Short Video Advertising and Brand Expression	[3][6][7]	Comment mining, communication path analysis, credibility modeling	Short video narratives and influencer credibility affect brand attitudes	Lacks a multimodal automatic generation perspective
AI Marketing and Communication Touchpoint Optimization	[8][9][10][11][14]	NLP, machine learning, sentiment recognition, predictive modeling	AI can improve advertising personalization and touchpoint coordination efficiency	Most studies focus on text or recommendation, with limited integration of short video scenarios
Generative Content and Virtual Communication Agents	[15][16][17][18][19][20][21]	AIGC generation, virtual human modeling, content matching analysis	AI-generated advertising and virtual influencers can enhance novelty and visibility	Research on brand authenticity, recognition accuracy, and platform resource costs remains insufficient

On the whole, the existing results have explained the short video communication logic and AI marketing method respectively, but there are still several obvious gaps in the market application of AIGC short video in brand communication: the first is the lack of a multi-modal analysis framework that integrates text, picture, audio and interactive feedback; The second is the lack of joint evaluation of brand information recognition accuracy, semantic consistency and communication effect improvement. Third, the engineering implementation issues such as model training time, inference speed and platform resource consumption are less discussed. Based on this, this paper starts from the dual dimensions of content generation and dissemination evaluation, and tries to establish an analysis model closer to the actual operation scenario of the platform.

3 Research Methods

This paper constructs AIGC brand communication analysis method for new media short video platform, and puts content generation and market evaluation in the same computational framework. As shown in Figure 1, the research takes brand short video samples, user reviews, interaction logs and commodity attribute data as input, and completes text segmentation, brand word cleaning, voice noise reduction, key frame extraction and timing alignment in the

preprocessing stage. Then, the model extracts multi-modal features from text, vision and audio, and forms a unified brand semantic representation through the cross-modal fusion layer. At the generating end, AIGC content generation network automatically generates short video script, subtitles, picture combination and dubbing results according to brand positioning, product selling point and platform scene. At the evaluation end, the communication effect prediction module combines click through rate, completion rate, like rate, comment sentiment and conversion feedback to quantitatively analyze the market suitability of the generated content. This method not only pays attention to the quality of generation, but also pays attention to whether the brand information is accurately transmitted, so that the AIGC short video application research has strong engineering interpretation power and communication analysis value.

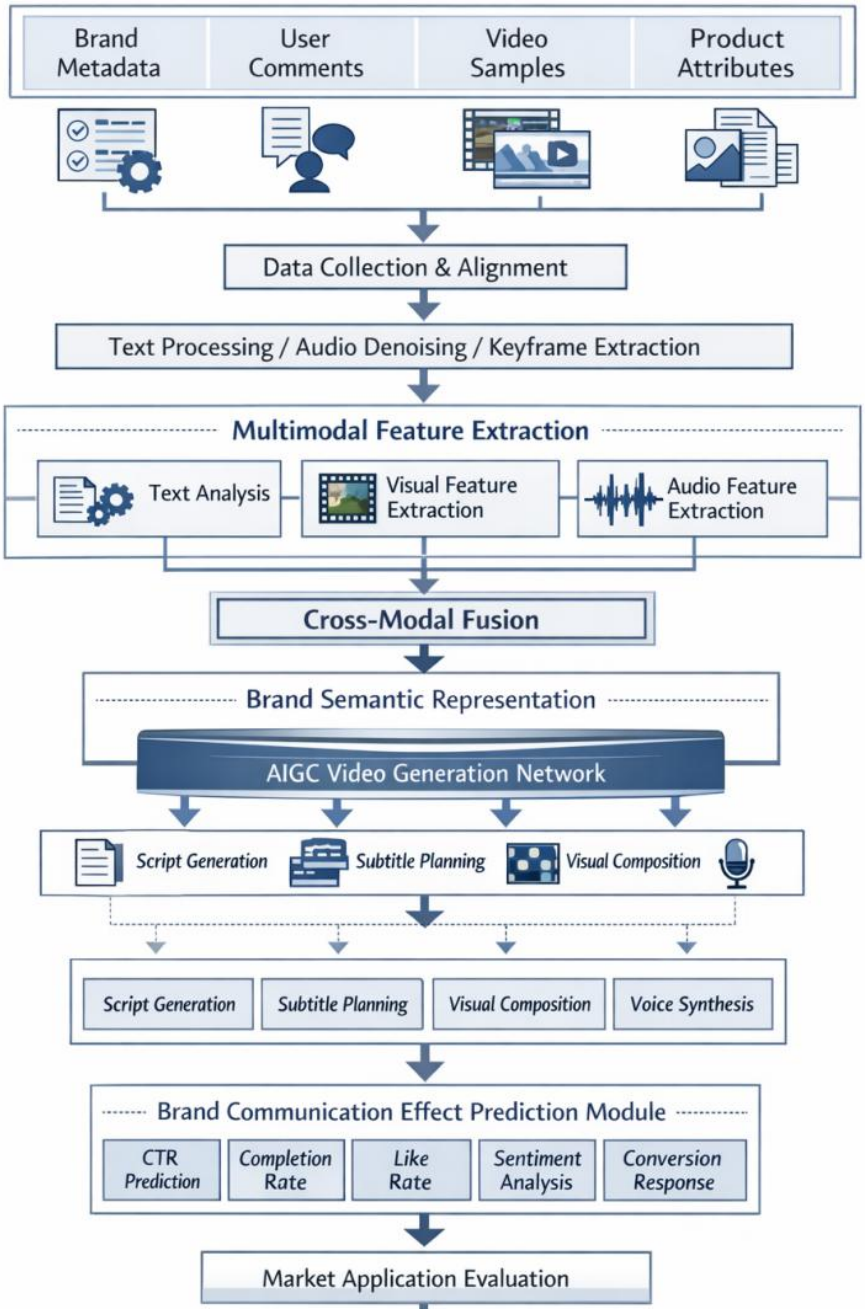

Figure 1: AIGC short video generation and brand communication evaluation method framework

3.1 Dataset Construction

The data set used in this paper is constructed for the research link of "new media short video platform -- AIGC content generation -- brand communication effect evaluation". Instead of using a single platform sample, short video content with clear brand identity, commodity link or marketing label is selected from Douyin, TikTok and Instagram Reels. Meanwhile, the title text, topic tags, subtitle transcription, cover frames, audio track segments, comment content, and interactive records such as likes, comments, forwarding, and completion are synchronously captured. On this basis, the attribute information of the brand, such as industry, product category, launch scenario and release time, is supplemented to form a structured sample set that can be used for multimodal analysis. After deduplication, outlier elimination and field alignment, 18,420 valid samples were finally retained, including 16 core fields. The visual dimension records the key frame style, product frequency and scene characteristics. The audio dimension covers speech rate, energy distribution, background music rhythm and speech clarity. The behavioral dimension corresponds to click rate, like rate, comment rate, share rate and conversion response. In order to facilitate subsequent modeling, this paper divides the brand communication effect into three categories of labels: high, medium and low, and uses the brand information recognition accuracy and user interaction intensity as auxiliary evaluation variables. This data set can completely present the content characteristics and market feedback of AIGC short videos in the transmission chain, but there are also some limitations. Although the sample covers multiple mainstream platforms, it may still be affected by industry distribution, regional preference and platform recommendation mechanism differences.

3.2 Data preprocessing

Brand communication data sources in short video platforms are complex, and text, picture and sound often appear synchronously, but they do not naturally have a consistent computational structure. If the original data is directly fed into the generation model and communication evaluation module, it is easy to weaken the stability of brand semantics due to colloquialization, label redundancy, caption mismatch and noise interference. Therefore, in this paper, a unified data preprocessing link is set up before modeling to transform the discrete multimodal content into an input representation that can be aligned, encoded, and comparable. The process revolves around the brand communication task, which not only deals with the dirty data in the general sense, but also pays more attention to the semantic correspondence between brand names, product selling points, scene words and user ratings. After preprocessing, the expression boundary of text information is clearer, and the semantics of video captions and comments can be more accurately mapped to the brand theme, which also provides a more stable input basis for subsequent multi-modal fusion and communication effect prediction.

3.2.1 Text segmentation and brand semantic cleaning

In the process of text processing, the video title, topic tag, subtitle transcription, comment content and product description are incorporated into the scope of semantic cleaning to construct a text corpus for brand communication analysis. Considering that the short video platform text has the characteristics of strong spoken language, frequent abbreviations, intensive emoticons and high repetition of marketing words, this paper combined word segmentation algorithm with brand dictionary to complete the preliminary segmentation, and specially labeled brand names, series names, product models, core selling points and activity terms to maintain a high recognition priority in the subsequent coding process. For stop words,

repeated hashtags, stacked emotional symbols and irrelevant drainage words that have no actual propagation meaning, they are jointly removed by rule filtering and context matching to reduce the interference of noise on semantic representation.

On this basis, this paper further introduces the brand semantic cleaning strategy to normalize synonymous variants, spelling errors, cross-language mixing and common abbreviations on the platform. For example, the nickname, abbreviation and official name of the same brand in the comment section are mapped to a unified entity, and their semantic attribution is corrected by word vector similarity and context window. After this process, the text data is transformed from loose natural expression to semantic sequence with clearer structure, which not only improves the accuracy of brand information extraction, but also enhances the consistency of AIGC generated content in brand positioning, product expression and communication theme.

3.2.2 Short video audio and video noise suppression and frame sequence normalization

In the new media short video platform, brand communication content is often accompanied by the characteristics of fast transition, handheld shooting, filter superposition, background music coverage and dynamic floating subtitles. Although these expressions can strengthen sensory stimulation, they will bring significant disturbance to the content understanding on the computer side. For the research of AIGC brand short video, if the original audio and video data are directly entered into the multi-modal modeling stage without processing, the model is easy to misjudge compression artifacts, environmental noise, camera jitter and invalid transition as effective propagation signals, and then weaken the stability of brand identity extraction, scene semantic recognition and user perception prediction. Therefore, in this paper, audio and video noise suppression and frame sequence normalization are set as the key steps in the preprocessing. The purpose is not only to improve the clarity of the material, but also to provide a time-consistent and structurally comparable input basis for subsequent brand semantic generation and communication effect evaluation.

On the video side, this paper uniformly regularize the frame sequence to solve the problems of jitter, brightness mutation and transition fragmentation common in short video samples. Let the original video sequence be $V = \{f_1, f_2, \dots, f_T\}$, where f_T represents the t-th frame image. In order to weaken the influence of high-frequency noise on image boundary and product contour recognition, Gaussian smoothing is used to locally filter a single frame, and its expression is as follows:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{1}$$

where σ is the smoothing strength parameter, and x and y represent the position of the pixel relative to the center of the convolution kernel. After filtering, the structural similarity between adjacent frames and the optical flow amplitude are combined to identify the shot transition region, and the duplicate frames, blurred frames and transition frames with very low information are eliminated. For the reserved frames, this paper completes the time axis reordering according to uniform sampling interval, so that different video clips have better comparability in input length. The normalized frame index can be expressed as follows:

$$\hat{f}_k = f_{\lfloor k \cdot \Delta \rfloor}, \quad k = 1,2, \dots, K \tag{2}$$

where Δ is the resampling step and K is the number of frames after warping. In this way, the brand Logo, product body, character appearance and usage scene in the video sequence can be

extracted in a more stable temporal structure, which is conducive to the subsequent visual coding network to learn the continuous pattern related to the brand, rather than being distracted by irrelevant clips.

On the audio side, short video materials often have problems such as strong soundtrack coverage, aliasing of human voice and environmental sound, and large differences in radio equipment. Considering that the narration, spoken-word selling points and emotional timbre in brand communication content have a direct impact on audience understanding, this paper uses Wiener filtering to denoise the audio signal. Let the observed signal be $y(t) = s(t) + n(t)$, where $s(t)$ is the original valid speech and n(t) is the additive noise. In the frequency domain, the Wiener filter is written as follows:

$$H(\omega) = \frac{S_{ss}(\omega)}{S_{ss}(\omega) + S_{nn}(\omega)} \tag{3}$$

Here, $S_{ss}(\omega)$ represents the effective signal power spectral density and $S_{nn}(\omega)$ represents the noise power spectral density. Aiming at minimizing the mean square error, this method can suppress the background noise while retaining the speech backbone information as much as possible. After denoising, this paper further performs energy normalization and silence segment clipping on audio segments to reduce the influence of recording loudness differences between different samples on model judgment. For the problem of audio and video synchronization, the frame and audio segments are aligned according to the caption timestamp and voice activity detection results, so that the brand broadcast content, product display shots and caption information can be matched in a similar time window.

3.3 Multi-modal feature extraction

After the completion of text segmentation, brand semantic cleaning, audio and video denoising and frame sequence normalization, the input data has good structural consistency, but this does not mean that it can be directly used for AIGC generation and propagation effect evaluation. Short video brand content itself has the characteristics of modal parallelism, information heterogeneity and unstable expression rhythm. The text carries brand discourse, product selling point and emotional orientation, the vision contains Logo, product appearance, character action and scene background, and the audio further superposes the mouth broadcast information, music rhythm and emotional intensity. Without an effective feature extraction mechanism, the model is easy to amplify local disturbances in high-dimensional noise, and it is difficult to accurately capture the core signal that is truly meaningful for brand communication. Therefore, this paper introduces a unified feature extraction strategy in the multi-modal modeling stage, maps the text, visual and audio coding results into a shared representation space, and combines principal component analysis to complete dimensionality reduction and redundancy compression, so as to improve the computational efficiency and semantic stability of the subsequent generation and prediction modules.

Let the preprocessed text, visual and audio features be denoted as $X_t \in \mathbb{R}^{n \times d_t}$, $X_v \in \mathbb{R}^{n \times d_v}$ and $X_a \in \mathbb{R}^{n \times d_a}$, respectively, where n is the number of samples and d_t,d_v and d_a represent the original feature dimensions of the three types of modalities, respectively. The text side mainly extracts brand entity distribution, keyword weight, sentiment polarity and context embedding. Visual emphasis encodes product boundaries, color styles, scene semantics and character appearance features in key frames. The audio side retains information such as MEL spectrum, energy fluctuation, speech rate variation and prosodic contour. In order to avoid the bias caused by the inconsistency of different modal dimensions, the feature matrix is first standardized:

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{4}$$

Here, μ_j and σ_j represent the mean and standard deviation of the JTH dimension feature, respectively. After normalization, each modal feature is concatenated into a joint representation matrix $X = [X_t, X_v, X_a]$, which is used for subsequent dimension reduction and key component extraction.

Considering the high dimension of the concatenated joint features and the correlation between different modalities, this paper uses principal component analysis to compress the multi-modal features. Let the covariance matrix of the joint features be:

$$C = \frac{1}{n-1} X^T X \tag{5}$$

The matrix C is eigen decomposed to obtain the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ and the corresponding eigenvectors $u_1, u_2, \ldots, u_m$. Among them, the larger the eigenvalue is, the more information of the sample variance is retained in that direction. The KTH principal component is then constructed as follows:

$$z_k = Xu_k \tag{6}$$

When the first r principal components are retained, the multimodal low-dimensional representation can be written as follows:

$$Z = XU_r \tag{7}$$

where $U_r = [u_1, u_2, \ldots, u_r]$. In order to ensure that enough brand communication information can still be retained after dimension reduction, this paper determines the number of principal components according to the cumulative explained variance ratio, which is calculated as follows:

$$\eta(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{m} \lambda_i} \tag{8}$$

When $\eta(r)$ reaches the preset threshold, it stops expanding the dimension. The processed feature representation not only retains the main change directions in brand semantics, visual recognition and audio rhythm, but also effectively weakens the interference of repetitive features, weak correlation features and platform noise on model training.

It should be noted that the use of PCA in this paper is not simply to reduce the computational burden, but more importantly to refine the propagation features with more explanatory power through linear projection. For example, in the brand short video samples, some principal components often correspond to the compound change of "increased brand word density, increased product main body appearance, and enhanced oral broadcast rhythm" at the same time. This kind of structured information can better reflect the real strength of brand communication content than single modal features. After multi-modal feature extraction, the propagation signal originally scattered in text, picture and sound is compressed into a relatively compact low-dimensional representation, which provides a more stable and efficient input basis for the subsequent AIGC short video generation network and brand communication effect prediction module.

3.4 AIGC short video generation and market application analysis model for brand communication evaluation

3.4.1 AIGC short video content generation network

In order to make the generated results not only conform to the communication rhythm of the short video platform, but also do not deviate from the original semantic boundary of the brand, this paper constructs an AIGC short video content generation network constrained by brand. As shown in Figure 2, the network takes the multi-modal fusion feature Z, brand knowledge vector B and platform scene encoding P as joint inputs, and completes script generation, shot planning, caption arrangement, speech synthesis and visual segment organization in the shared semantic space. Different from the general model that simply pursues "generating visible content", this paper emphasizes the correspondence between content output and brand communication goals, that is, the generated results should not only have narrative integrity, but also be stable in the presentation of product selling points, the salience of brand identity and the direction of emotional expression.

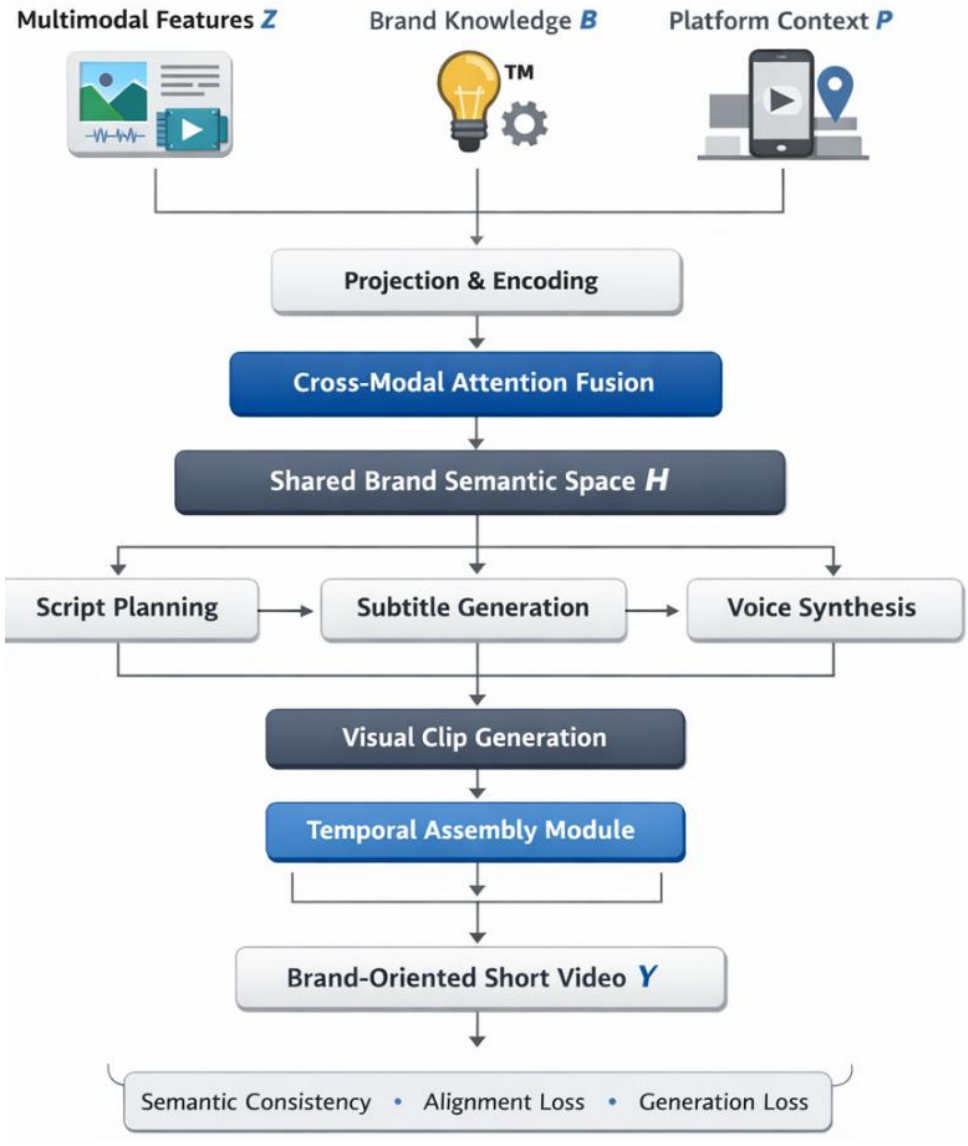

Figure 2: AIGC short-video content generation network for brand communication

In the encoding stage, text semantics, keyframe visual features and audio prosodic features are first mapped into a unified representation space, and then weighted fusion is completed by cross-modal attention mechanism, whose expression is:

$$H = \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d}}\right)V \tag{9}$$

Here, Q, K and V are obtained by linear transformation of Z, B and P respectively, d represents the feature dimension, and H is the shared semantic representation under the brand condition constraint. The representation not only retains the brand discourse and product selling points in the text, but also absorbs the picture style, the character appearance way and the rhythm preference in the platform content environment, so as to provide more propagation-targeted conditional information for the subsequent generation module.

In the generation phase, the network adopts a three-level structure of "semantic planning-fragment generation-temporal assembly". The semantic planning layer generates short video script skeleton and shot label sequence according to H. The segment generation layer outputs the subtitle text, dubging vector and visual segment candidates respectively. The timing assembly layer completes the content splicing according to the camera transition rules. Its conditional generation process can be expressed as follows:

$$Y = G(H, r) = G(Z, B, P, r) \tag{10}$$

where, G(·) represents the content generation network, r is the random disturbance term, which is used to control the diversity of picture details and expression styles, and Y represents the final generated brand short video content. In order to avoid the model overcatering to the platform traffic preference and weakening the brand core information, this paper introduces the brand semantic consistency loss in the training process to constrain the deviation between the generated text and the brand knowledge representation:

$$L_{\text{brand}} = 1 - \cos(E(Y), E(B)) \tag{11}$$

Here, E(·) is the semantic encoding function, and $\cos(\cdot)$ represents the cosine similarity. The smaller the item is, the closer the generated content is to the original semantic of the brand. Furthermore, the total model loss function is written as follows:

$$L = \lambda_1 L_{\text{gen}} + \lambda_2 L_{\text{brand}} + \lambda_3 L_{\text{align}} \tag{12}$$

Here, L_{gen} is the base generation loss, L_{align} is the multi-modal alignment loss, and $\lambda_1, \lambda_2, \lambda_3$ are the weight coefficients. This design enables the network to continuously correct the direction of brand expression while ensuring the fluency of generation, and reduces the problem of "correct content generation but out-of-focus brand communication". As a whole, the network is not simply reusing the general AIGC structure, but embedding the identification, coherence and platform adaptability in brand communication into the generation process, so that the output content is more suitable for the actual marketing materials on the new media short video platform.

3.4.2 Brand communication effect prediction and market application evaluation module

After the AIGC short video is generated, the communication value cannot be judged only by the surface quality of the content, but it also needs to be further tested whether it really

improves the brand recognition, user interaction and market conversion. Based on this consideration, this paper sets up the brand communication effect prediction and market application evaluation module after generating the network, as shown in Figure 3. The module takes the multi-modal representation Y of the generated content, the brand knowledge vector B, the historical behavior characteristics U of the platform and the parameters P of the launch scene as the joint input, extracts the propagation state representation through the shared coding layer, and then feeds it into the three task branches of brand identification, interactive response and conversion tendency respectively to realize the synchronous prediction of AIGC short video market performance. The purpose of this design is to avoid relying on a single click metric to make a one-sided judgment of the communication effect, but to "be seen", "be understood" and "be responded" in the same calculation framework. Let the shared propagation be denoted by h, then its generation process can be written as follows:

$$h = \phi(W_y Y + W_b B + W_u U + W_p P + b) \tag{13}$$

where W_y, W_b, W_u and W_p are the mapping weights for different inputs, b is the bias term, and $\phi(\cdot)$ is the nonlinear activation function. On this basis, the brand information recognition probability, user interaction probability and conversion propensity score are respectively expressed as follows:

$$\hat{y}_r = \sigma(W_r h), \quad \hat{y}_e = \sigma(W_e h), \quad \hat{y}_c = \sigma(W_c h) \tag{14}$$

In the formula, $\hat{y}_r$ represents the user's recognition degree of the brand's core information, $\hat{y}_e$ represents the comprehensive probability of interactive response such as liking, commenting, sharing and completion, and $\hat{y}_c$ corresponds to the predicted value of purchase, click, favorite or subsequent conversion behavior. In order to facilitate the comparison of the practical application potential of different AIGC contents, this paper further constructs a comprehensive market evaluation score:

$$S = \alpha \hat{y}_r + \beta \hat{y}_e + \gamma \hat{y}_c - \delta C \tag{15}$$

where C is the content deployment cost term, including inference delay, computing power occupancy and material generation time, α, β, γ, δ are weight coefficients. This equation means that communication effectiveness is not simply to maximize interaction, but to seek a more robust balance between brand recognition, user feedback and platform cost.

In the training process, the module adopts a multi-task joint optimization strategy to sum the recognition loss, interaction prediction loss and transformation prediction loss, so as to improve the information sharing ability between different tasks. After this process, the system can not only judge whether a short AIGC video is popular, but also explain why it is effective or invalid. For example, when the interaction rate is high but the brand recognition score is low, it can be determined that the content is more inclined to platform traffic expression than brand precipitation. If the brand recognition and interaction performance are better at the same time, it means that the generated content has higher delivery value at the market application level. Therefore, this module plays a key role in connecting content generation and commercial evaluation in the full-text method system, making AIGC short video research no longer stay in the generation quality comparison, but can further enter the empirical level of brand communication performance analysis.

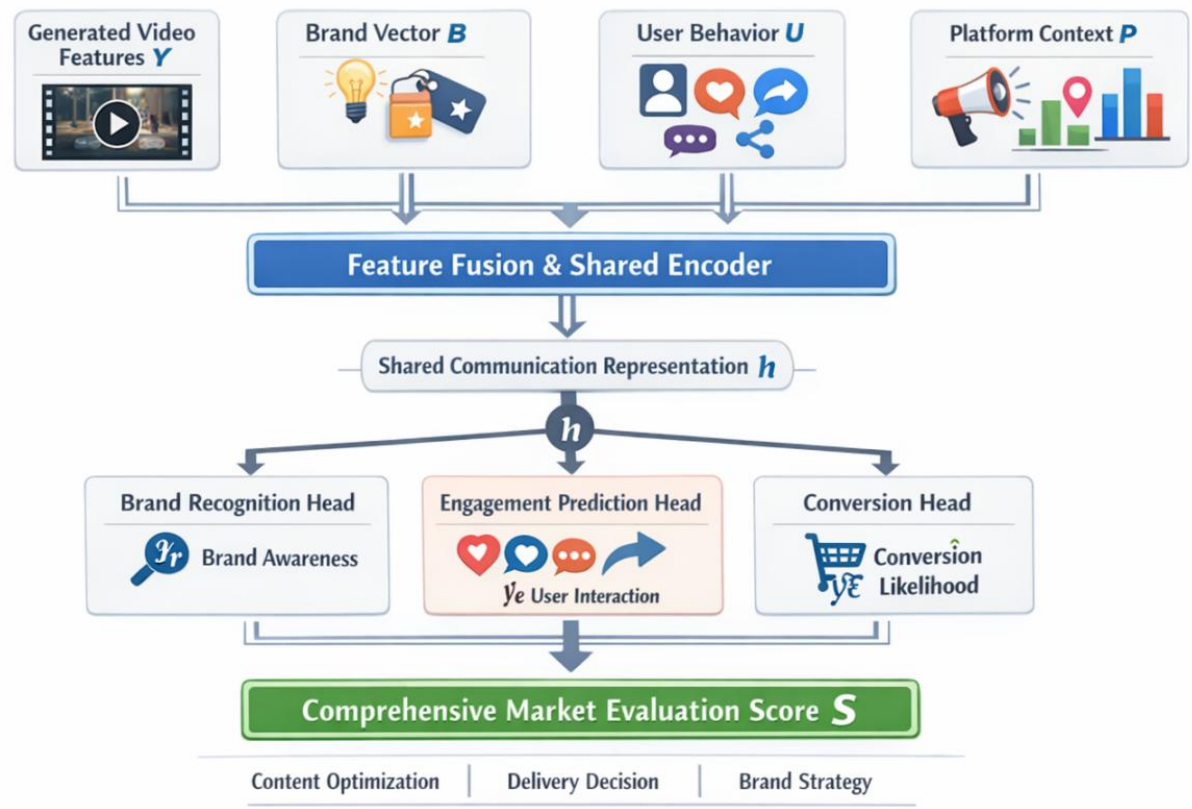

Figure 3: Brand communication effect prediction and market application evaluation module

4 Results and discussion

In order to verify the actual performance of the model constructed in this paper in the generation and propagation evaluation of AIGC brand short video, the model in this paper is compared with the basic multi-modal generation model in the unified data set, the same training rounds and the consistent hardware environment. The experimental platform uses Windows 10 operating system, Intel Core i7 processor, 32 GB memory, and the development environment is Python 3.10 and PyTorch. The evaluation index is not limited to the generation quality itself, but also includes the brand semantic consistency, brand information recognition accuracy, user interaction performance, communication effect improvement, training time, reasoning speed and platform resource consumption, in order to more completely describe the comprehensive value of AIGC short video in the market application.

From the overall results, the model in this paper performs more stable in communication related indicators, and the generated content not only maintains high semantic coherence, but also shows stronger advantages in brand identity and interactive feedback. Table 2 presents the results of the core experiments. It can be seen that the semantic consistency score of the proposed model reaches 0.87, which is 0.09 higher than that of the baseline model. The accuracy of brand information recognition increased from 84.6% to 90.8%; The comprehensive index of user interaction and the improvement rate of brand communication effect also increased by 10 percentage points and 9.3 percentage points respectively. At the same time, although the model training time and resource consumption are slightly increased, the inference speed is maintained in the acceptable interval, which indicates that the proposed method achieves a relatively balanced result among generation quality, propagation effect and deployment cost. Based on this result, the following subsections will carry out specific

analysis from the dimensions of semantics, recognition, interaction, efficiency and resources.

Table 2: Comparison of the overall experimental results of the model

Indicator	Baseline Model	Proposed Model
Semantic Consistency of Generated Content	0.78	0.87
Brand Information Recognition Accuracy (%)	84.6	90.8
Comprehensive User Interaction Index	0.71	0.81
Improvement Rate of Brand Communication Effectiveness (%)	12.4	21.7
Model Training Time (min)	126	139
Content Generation Inference Speed (frames·s^{-1})	22.8	26.4
Platform GPU Memory Usage (GB)	4.9	5.4

4.1 Generate content semantic consistency analysis

The semantic consistency of the generated content determines whether the AIGC short video can maintain the stable boundary of brand expression while pursuing the platform communication efficiency. In this paper, a comprehensive consistency index is constructed by brand theme matching, product selling point coverage and subtitle-screen semantic correspondence, and the proposed model is compared with the baseline model. The results show that the semantic consistency score of the proposed model on the test set reaches 0.87, which is higher than 0.78 of the baseline model, indicating that its performance is more stable in maintaining the core semantic of the brand. As shown in Figure 4, with the increase of training rounds, the consistency curve of the proposed model continued to rise and leveled off after the 40th round. Although the baseline model has also improved, the fluctuation is more obvious in the middle and late stages, indicating that its collaborative constraint on multi-modal information is still not sufficient.

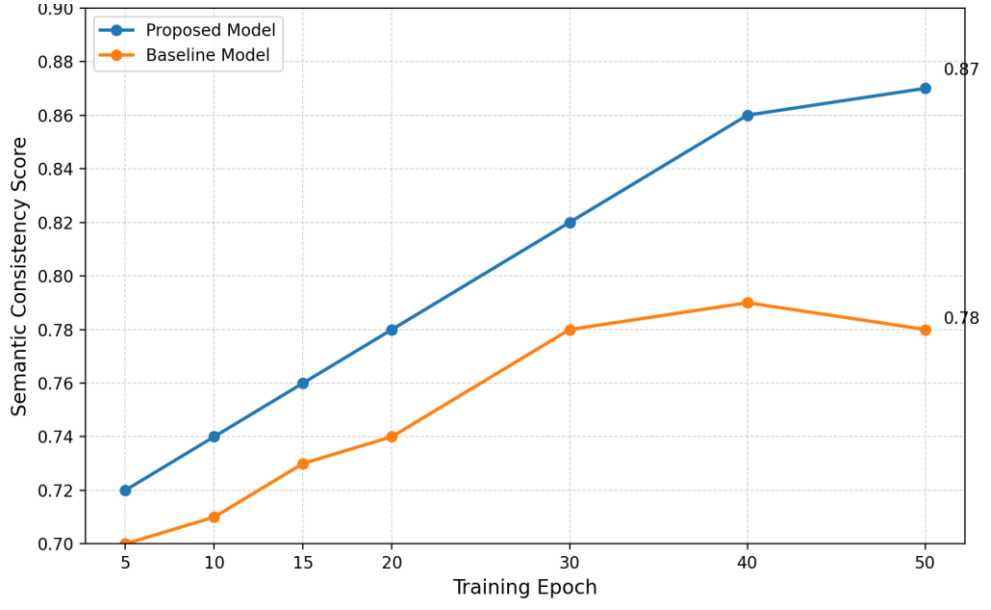

Figure 4: Generated content semantic consistency change curve

From the generation results, the proposed model can better maintain the correspondence between brand names, product function words and context descriptions, and reduce the phenomenon of "image attraction but semantic drift". The reason is that the model introduces

brand knowledge constraints into the shared semantic space, so that text generation, shot arrangement and voice expression are no longer separated from each other, but complete collaborative output under a unified semantic goal. Therefore, this result not only shows that the proposed method improves the quality of content generation, but also shows that AIGC short videos have stronger semantic controllability in brand communication scenarios.

4.2 Accuracy analysis of brand information recognition

The accuracy of brand information recognition reflects whether the AIGC short video can be clearly identified as a specific brand content by users in the dissemination process, and is also a key indicator to judge whether the generated results have commercial usability. In this paper, brand name recognition, product category recognition, core selling point recognition and visual identity matching are jointly included in the evaluation scope, and the classification accuracy is used as the main measure. It is calculated as follows:

$$\mathrm{Acc} = \frac{\mathrm{N}_{\mathrm{correct}}}{\mathrm{N}_{\mathrm{total}}} \tag{16}$$

Here, $\mathrm{N}_{\mathrm{correct}}$ represents the number of correctly identified samples and $\mathrm{N}_{\mathrm{total}}$ represents the total number of test samples.

The experimental results show that the brand information recognition accuracy of the proposed model reaches 90.8%, which is significantly higher than the 84.6% of the baseline model. As shown in Figure 5, the gap between the two types of models is relatively limited in the two dimensions of brand name and product category. However, in terms of core selling point identification and visual logo matching, the advantages of the proposed model are more obvious. This shows that after the constraint of brand knowledge and multi-modal collaborative modeling, the generated content is not only close to the brand expression in the surface text, but also can maintain strong consistency between the picture structure, the subtitle prompt and the voice broadcast. In other words, what users recognize is not just "a short video", but "a clear expression of a certain brand in a specific context".

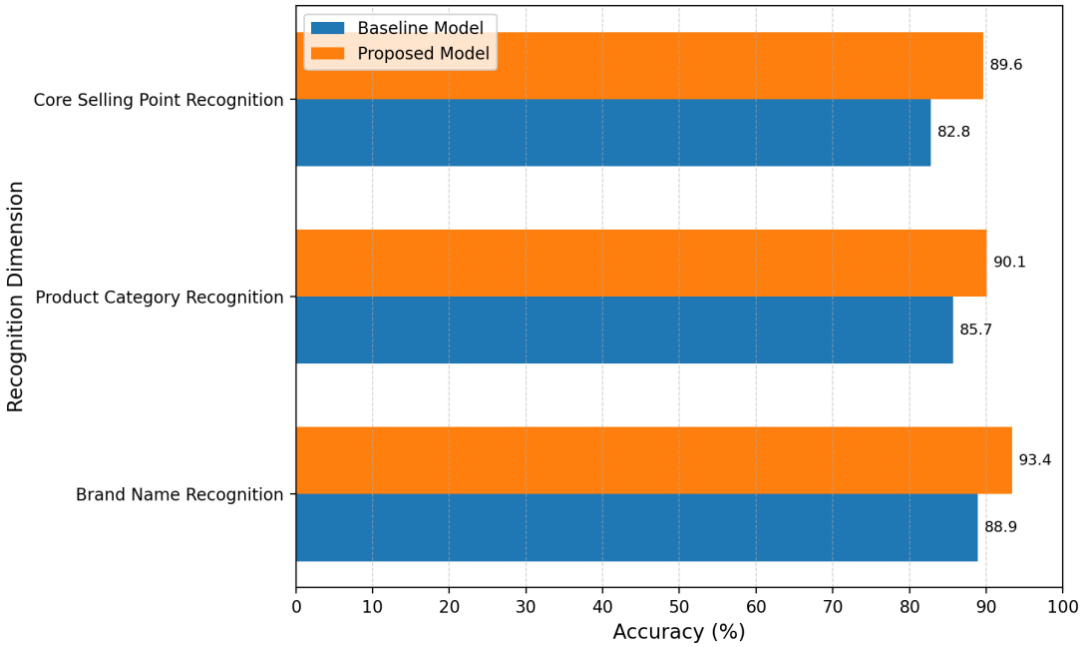

Figure 5: Comparison of brand information recognition accuracy

This result indicates that AIGC content generation can significantly reduce the recognition bias caused by information drift if it can be jointly trained with brand semantic rules, visual element detection and communication task labels, so as to lay a more stable communication foundation for subsequent interaction growth and transformation promotion:

$$I = \omega_1 L + \omega_2 C + \omega_3 S + \omega_4 F \tag{17}$$

Here, L represents the like rate, C represents the comment rate, S represents the share rate, F represents the completion rate, and ω_i is the corresponding weight. It can be seen from Table 3 that the interaction composite index of the proposed model reaches 0.81, which is higher than 0.71 of the baseline model, indicating that the generated content has more advantages in attracting stay, stimulating expression and promoting secondary communication. In terms of sub-items, the improvement of comment rate and share rate of the model in this paper is more obvious, which indicates that the coupling between brand information and content context is higher, and users are not only willing to watch, but also more willing to participate in the discussion and diffusion of content.

Table 3: Comparison of user interaction performance

Indicator	Baseline Model (%)	Proposed Model (%)
Like Rate	12.8	15.6
Comment Rate	4.9	7.1
Share Rate	3.7	5.8
Video Completion Rate	62.4	69.3
Comprehensive Interaction Index	0.71	0.81

4.3 Analysis on the improvement of brand communication effect

The improvement of brand communication effect is not only reflected in the increase of short-term interaction, but also reflected in the synchronous improvement of brand cognition, attitude change and subsequent transformation intention. To this end, this paper incorporates brand memory improvement rate, favorability improvement rate and purchase intention improvement rate into the unified analysis, and constructs a comprehensive communication gain index:

$$E = \alpha R + \beta A + \gamma P \tag{18}$$

Here, R represents the improvement rate of brand memory, A represents the improvement rate of brand favorability, P represents the improvement rate of purchase intention, and α, β, γ are the weight coefficients. Experimental results show that the performance of the proposed model is significantly better than the baseline model in terms of comprehensive propagation gain, with the overall improvement rate reaching 21.7%, while the baseline model is 12.4%. As shown in Figure 6, the improvement rates of the proposed model in the three dimensions of brand memory, brand goodwill and purchase intention reach 22.8%, 19.7% and 22.6%, respectively, which are higher than 13.1%, 11.6% and 12.5% of the baseline model, and the improvement of brand memory and purchase intention is the most obvious.

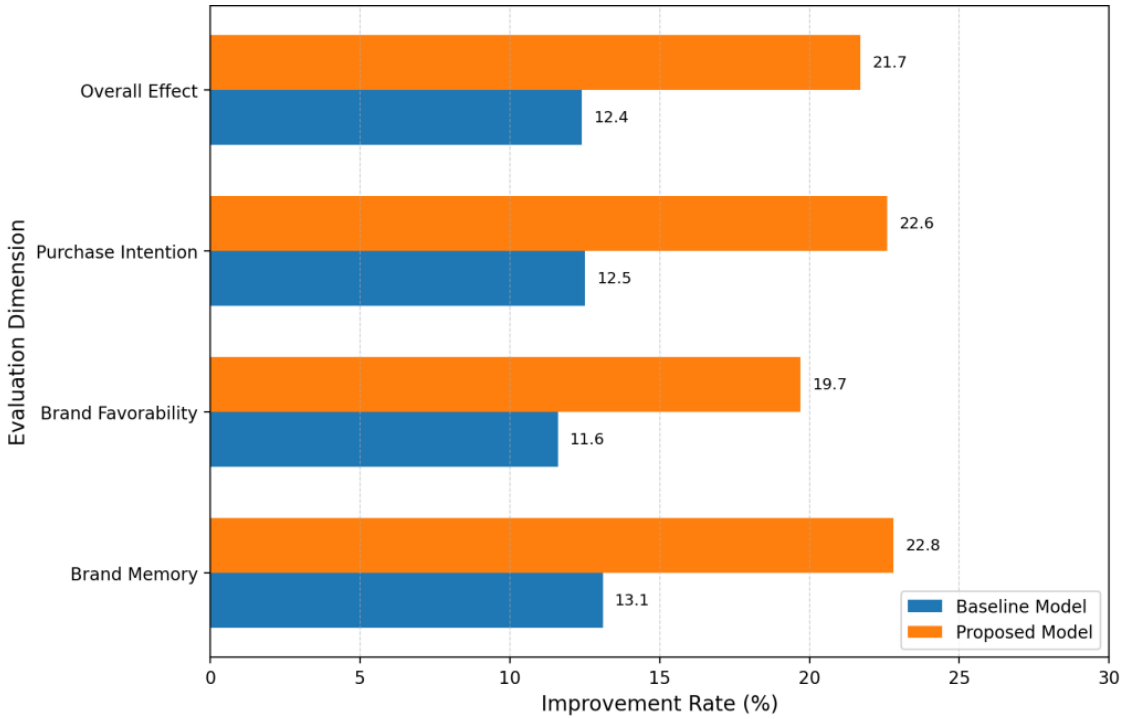

Figure 6: Comparison of brand communication effect improvement

This result shows that after brand knowledge constraint and multi-modal generation optimization, AIGC short video not only improves the content visibility, but also enhances the efficiency of brand information entering the user cognitive system. In other words, after finishing watching, users are not only more likely to remember the brand, but also more likely to form positive evaluations and generate subsequent consumption tendencies. It can be seen that the market application value of AIGC in the new media short video platform is not limited to the production speed, but lies in its ability to transform the content generation ability into more observable brand communication benefits.

4.4 Model training time analysis

The model training time reflects the engineering feasibility of AIGC short video system in practical deployment. For brand communication tasks, although the model can improve the generation quality, it needs too long training period, which will weaken its application value in high-frequency content update scenarios. In this paper, the training time of the baseline model and the proposed model are compared under the condition of the same data size, the same batch size and the consistent hardware environment. The results show that the total training time of the proposed model is 139 minutes, which is higher than 126 minutes of the baseline model, with an increase of 10.3%. This difference mainly comes from the extra overhead caused by the brand knowledge constraint layer, the multi-task propagation prediction branch, and the cross-modal alignment calculation.

However, from the perspective of the training convergence process, although the time consumption of a single round of the proposed model is slightly higher, the effective convergence rounds are more concentrated, the fluctuations in the middle and late stages are smaller, and the parameter update is more stable. Table 4 shows that the loss of the proposed model decreases significantly faster than that of the baseline model after the first 20 rounds, indicating that after the introduction of brand semantic constraints, the model search direction is clearer and the invalid shock is reduced. In other words, the moderate increase in training

time gains the synchronous improvement in semantic consistency, recognition accuracy and communication effect, and this cost is still in the acceptable range in the commercial application scenario of brand short videos.

Table 4: Comparison of model training time

Indicator	Baseline Model	Proposed Model
Total Training Time (min)	126	139
Average Training Time per Epoch (min)	2.52	2.78
Loss Value at Epoch 20	0.412	0.356
Loss Value at Epoch 40	0.287	0.241
Convergence Epoch	47	43

4.5 Discussion

Combined with the above experimental results, it can be found that the market application of AIGC generated content in the new media short video platform does not rely solely on generation efficiency to obtain communication advantages, and its real effect is reflected in the linkage between brand semantic control, multi-modal information collaboration and platform feedback closed loop. The overall improvement in semantic consistency, brand recognition accuracy and user interaction performance of the proposed model indicates that when text generation, visual organization, voice expression and brand knowledge constraints are incorporated into the same computing framework, short video content can more stably convey brand core information while maintaining the communication rhythm of the platform. In other words, AIGC did not weaken the control of brand communication, but enhanced the degree of coupling between content production and communication goals under certain conditions. However, this result does not mean that the application boundary of AIGC short video has been eliminated. The increase of training time and resource occupancy in the experiment shows that the joint modeling of multi-modal generation and propagation prediction still has high computational cost, especially when brands need to update materials frequently and quickly respond to hot scenes, the real-time performance and computing power adaptation ability of the system still need to be optimized. At the same time, brand communication is not completely equivalent to interaction growth. If the model excessively seeks platform indicators such as comments, forwarding and completion, it may also lead to content style tilting to traffic preference, thereby weakening the long-term stability of brand expression. From this point of view, the effective application of AIGC in brand communication should be based on the balance of "generation quality, communication effect and deployment cost". Future research needs to further expand the scope of industry samples, and introduce more fine-grained user cognition and conversion data to improve the generalization ability of the model in different platforms, different categories and different communication stages.

5 Conclusion

This paper focuses on the content market application of AIGC in the new media short video platform, constructs an analysis framework that integrates data preprocessing, multi-modal feature extraction, short video content generation, brand communication effect prediction and market application evaluation, and investigates its actual influence on brand communication from the perspective of computer modeling. The research shows that AIGC short video not only improves the efficiency in the content production side, but more importantly, it can form

a computable closed-loop relationship between brand semantic expression, platform communication rhythm and user feedback analysis through the collaborative modeling of text, visual and audio information. Experimental results show that the proposed model is superior to the baseline model in many core indicators. The semantic consistency of the generated content is increased from 0.78 to 0.87, the accuracy of brand information recognition is increased from 84.6% to 90.8%, the user interaction composite index is increased from 0.71 to 0.81, and the improvement rate of brand communication effect is increased from 12.4% to 21.7%. This shows that under the joint effect of brand knowledge constraints and multi-modal fusion mechanism, the content generated by AIGC can more accurately carry the brand name, product selling point and scene semantics, and it is also easier to trigger comments, sharing and subsequent dissemination in the platform environment. Meanwhile, the training time of the proposed model was 139 minutes, which was higher than the baseline model's 126 minutes, but the content generation and inference speed reached 26.4 frame·s^{-1}, which still maintained good engineering deployability. In general, the market application of AIGC in the new media short video platform has gradually shifted from "assistant creation tool" to "brand communication computing system". Its value is not only reflected in the speed of production, but also reflected in the measurable communication effect, controllable content logic and optimized brand expression. The research results of this paper provide an empirical basis for brands to understand the commercial application boundaries of AIGC short videos, and also lay a methodological foundation for subsequent intelligent communication research with higher precision and stronger real-time performance.

References

[1] Cuesta-Valiño P, Gutiérrez-Rodríguez P, Durán-Álamo P. Why do people return to video platforms? Millennials and centennials on TikTok[J]. Media and Communication, 2022, 10(1): 198-207.

[2] Deng D S, Seo S, Li Z, et al. What people TikTok (Douyin) about influencer-endorsed short videos on wine? An exploration of gender and generational differences[J]. Journal of Hospitality and Tourism Technology, 2022, 13(4): 683-698.

[3] Li D, Chow U T, Cheong C Y M. A critical genre analysis of covert advertising through short-videos in Douyin: The Chinese version of Tik-Tok[J]. Sage Open, 2022, 12(4): 21582440221134608.

[4] Barta S, Belanche D, Fernández A, et al. Influencer marketing on TikTok: The effectiveness of humor and followers' hedonic experience[J]. Journal of Retailing and Consumer Services, 2023, 70: 103149.

[5] Dong X, Liu H, Xi N, et al. Short video marketing: what, when and how short-branded videos facilitate consumer engagement[J]. Internet Research, 2024, 34(3): 1104-1128.

[6] Rizomyliotis I, Lin C L, Konstantoulaki K, et al. TikTok short video marketing and Gen Z's purchase intention: evidence from the cosmetics industry in Singapore[J]. Journal of Asia business studies, 2024, 18(4): 930-945.

[7] Alcántara-Pilar J M, Rodriguez-López M E, Kalinić Z, et al. From likes to loyalty: Exploring the impact of influencer credibility on purchase intentions in TikTok[J]. Journal of Retailing and Consumer Services, 2024, 78: 103709.

[8] Haleem A, Javaid M, Qadri M A, et al. Artificial intelligence (AI) applications for marketing: A literature-based study[J]. International Journal of intelligent networks, 2022, 3: 119-132.

[9] Sun H, Zafar M Z, Hasan N. Employing natural language processing as artificial intelligence for analyzing consumer opinion toward advertisement[J]. Frontiers in psychology, 2022, 13: 856663.

[10] Malthouse E, Copulsky J. Artificial intelligence ecosystems for marketing communications[J]. International Journal of Advertising, 2023, 42(1): 128-140.

[11] He A Z, Zhang Y. AI-powered touch points in the customer journey: a systematic literature review and research agenda[J]. Journal of Research in Interactive Marketing, 2023, 17(4): 620-639.

[12] Mariani M M, Hashemi N, Wirtz J. Artificial intelligence empowered conversational agents: A systematic literature review and research agenda[J]. Journal of Business Research, 2023, 161: 113838.

[13] Oc Y, Plangger K, Sands S, et al. Luxury is what you say: Analyzing electronic word-of-mouth marketing of luxury products using artificial intelligence and machine learning[J]. Psychology & Marketing, 2023, 40(9): 1704-1719.

[14] Gao B, Wang Y, Xie H, et al. Artificial intelligence in advertising: advancements, challenges, and ethical considerations in targeting, personalization, content creation, and ad optimization[J]. Sage Open, 2023, 13(4): 21582440231210759.

[15] Brüns J D, Meißner M. Do you create your content yourself? Using generative artificial intelligence for social media content creation diminishes perceived brand authenticity[J]. Journal of Retailing and Consumer Services, 2024, 79: 103790.

[16] Chen Y, Wang H, Hill S R, et al. Consumer attitudes toward AI-generated ads: Appeal types, self-efficacy and AI's social role[J]. Journal of Business Research, 2024, 185: 114867.

[17] Kim E, Kim D, E Z, et al. The next hype in social media advertising: Examining virtual influencers' brand endorsement effectiveness[J]. Frontiers in Psychology, 2023, 14: 1089051.

[18] Koles B, Audrezet A, Moulard J G, et al. The authentic virtual influencer: Authenticity manifestations in the metaverse[J]. Journal of Business Research, 2024, 170: 114325.

[19] Looi J, Kahlor L A. Artificial intelligence in influencer marketing: A mixed-method comparison of human and virtual influencers on Instagram[J]. Journal of Interactive Advertising, 2024, 24(2): 107-126.

[20] Liu F, Lee Y H. Virtually authentic: examining the match-up hypothesis between human vs virtual influencers and product types[J]. Journal of Product & Brand Management, 2024, 33(2): 287-299.

[21] Volles B K, Park J, Van Kerckhove A, et al. How and when do virtual influencers

positively affect consumer responses to endorsed brands?[J]. Journal of Business Research, 2024, 183: 114863.