Personalized Marketing Content Generation System Combined with Diffusion Model

Jun Feng
Department of Management Engineering, Inner Mongolia Vocational College of Chemical Engineering, Inner Mongolia, Hohhot 010070, China
Corresponding author: Jun Feng (e-mail: nhyfengjun@163.com)

ABSTRACT Current automated marketing content generation systems typically release content based on predefined rules or lightweight models, resulting in limited variation, creativity, and personalization for advanced engineering products, especially those involving electromagnetic shielding materials, wearable antennas, propagation-related devices, or other electromagnetic application scenarios. These systems often fail to account for product-specific visual characteristics, brand distinctions, electromagnetic application contexts, and diverse user preferences. To address this, the proposed system extracts interest and semantic features through multimodal user profiling. Next, it introduces an enhanced Stable Diffusion model to incorporate user semantic vectors during the denoising and generation process for personalized content synthesis. It implements a prompt control module to automatically vary marketing themes, uses the CLIP feature space to control style diversity, and filters high-quality content through a proposed multidimensional quality evaluation network. As a result, the system generates marketing content with improved creativity and personalization for engineering products with advanced electromagnetic application backgrounds. The results indicate that the proposed model outperforms all benchmarks, with an average semantic consistency score of 0.842, a creative diversity score consistently above 0.85, and an emotional fit score of 0.91 for middle- and high-income groups. The overall brand tone match score reaches 0.905, indicating reduced homogeneity and enhanced creative customization.

INDEX TERMS diffusion model, personalized marketing, multimodal semantic fusion, electromagnetic applications, wearable antennas

I. INTRODUCTION

The rapid growth of digital marketing has resulted in a change in content production methods, which leads to a higher reliance by advanced engineering and technology companies on the generation and distribution of marketing content to effectively showcase product innovation, technical features, and application scenarios. Traditional manual creation methods are

constrained by a range of timeliness issues and lack of originality when addressing frequent and differing audience content needs while confronting a variety of content needs [1,2]. As a result, automated marketing content generation systems have become a popular field for research and practice in the market. However, most of the existing or widely used marketing content generation systems rely upon templates or lightweight neural models, and their generation mechanisms depend upon fixed language patterns or easy feature matching algorithms, limiting true original expression (at the semantic level) in the generated outputs [3,4]. The existing systems often preserve a form of logical flow through a piece of content, but the generated outputs lack a deep emotional expression and personality. The results of automated content generation often appear homogenized and do not distinctly represent user preferences or brand differences. With the added modality of user data (e.g., visual, text, emotional signal indicators, and behavioral signals), the traditional systems cannot adequately integrate a range of these modalities [5,6]. Subsequently, the end product cannot substantively represent the contextual characteristics and psychological intentions of users, limiting the exploration of applying marketing automation in higher value contexts [7,8]. Specifically in terms of brand communication and user engagement, companies are under pressure to have a content generation system in order to gain personality-based, creative enrichment, and consistency–capabilities that constitute the primary barrier of traditional algorithm systems. There has been considerable academic and industry research on personalized content generation phenomena. Early research primarily engaged rule-based text generation and recommendation systems producing marketing copy via keyword matching, template swapping, and musings of logical filling. Nevertheless, it did not provide sufficiently natural language or emotional expressivity. Gupta and Mukherjee [9] were the first to construct the conceptual framework for the generative artificial intelligence application in retail product and service information search, setting the stage for future studies. Thomas [10] in particular looked at the accelerated application of generative artificial intelligence in digital marketing to show validity in overall effectiveness and operational value. Finally, Mogaji et al. [11] questioned the applicability of traditional technology acceptance models in the era of generative artificial intelligence, pointing out that the existing theoretical framework needs to be updated to meet the challenges brought by emerging technologies. The current research trend is gradually turning to the use of large-scale pre-trained models, such as the GPT (Generative Pre-trained Transformer) series and the T5 structure, for automatic generation

of marketing copy and topic content expansion. This type of method has achieved certain results in terms of text fluency and logic, but it still has limitations in terms of dynamic response capabilities to individual user characteristics and personalized marketing support [12,13]. Therefore, although a large number of studies have been devoted to improving the semantic expression and content quality of the generation system, the existing methods still have systematic defects in terms of creative expression, user matching, and brand adaptation. The purpose of this study is to create a marketing text generation system that balances personalized expressions with creative generation, specifically addressing the need of advanced engineering and technology industries for content that highlights product differentiation while appealing to individual customer preferences. Furthermore, the methodology aims to overcome the capability limitations of existing user generated content. This study first develops a user profiles system using multimodal data mining methods to extract deep semantic features (e.g., interests, emotions, intonation) to allow for user generated text. A user semantic embedding mechanism is implemented into the improved Stable Diffusion model in order to incorporate user features into the generative process during noise removal and feature recovery. This is also done to ensure the output text is semantically consistent with the user profile. The system also integrates a Prompt control module to automatically modify the generation direction based on differing marketing scenarios, which allows for topic adaptation. Finally, during generation, a specific CLIP feature space distance control mechanism is used to adjust diversity and style variations in the text outputs to improve degree of creative expression. Ultimately, a multi-layered quality assessment network is utilized to filter the produced content and elevate content quality based on semantic consistency, creativity index, and emotional fit. This system design removes any reliance on fixed templates regarding the generation of marketing communications content and operates in a semantically tethered approach based on the diffusion mechanism. This allows for enhanced creative richness, personality fit, and brand adaptability of content produced. The research provides a new technical pathway for generative AI and intelligent marketing to integrate and proposes a unique framework and theoretical support for high-quality automated content production.

II. SYSTEM DESIGN AND IMPLEMENTATION

A. MULTIMODAL USER PROFILE CONSTRUCTION

Data Preprocessing and Feature Representation

In order to provide tailored support in the process of generating and designing marketing content, this study first samples and normalizes users' multimodal behavioral data evenly. The data types include users' browsing history, click behavior logs, purchases, and text content from social media. All raw data is tagged together in a time-series fashion and mapped to behavioral events using unique IDs to ensure consistent relationships among data types. Text data undergo language cleaning, stopword filtering, and word segmentation normalization before being fed into a pre-trained BERT (Bidirectional Encoder Representations from Transformers) model. The model extracts a contextual semantic vector $E_t \in \mathbb{R}^{768}$ for each sentence. This vector represents the deep semantic distribution of the text and reflects the user's underlying interests and emotional preferences expressed in the language. Numerical behavioral data (such as click count, dwell time, and conversion rate) is normalized to the [0, 1] range using Min-Max normalization, and short-term and long-term interest trends are calculated using time window aggregation. In the feature fusion stage, the system uses a multilayer perceptron (MLP) to perform nonlinear mapping on the numerical features to generate a behavioral embedding vector E_b. Visual modality data (such as advertisements or product images viewed by users) is encoded using the CLIP image encoder to obtain a feature vector E_v. The three types of features are aligned and then concatenated into a comprehensive feature matrix, as shown in Formula (1):

$$U = [E_t \oplus E_b \oplus E_v] \tag{1}$$

The symbol $\oplus$ represents the vector concatenation operation, and the resulting matrix U serves as the input for subsequent semantic modeling and personalized generation. To reduce the interference of redundant features, the system uses a dimensionality reduction module based on principal component analysis (PCA) after concatenation, retaining 95% of the variance explained, ensuring a compact feature

space and high semantic differentiation. This process eliminates cross-modal feature scale differences and improves the separability and robustness of user profiles. Although BERT and CLIP features are highly compressed, the concatenated dimensions still reach 1536 (768+512+256), which can easily lead to redundancy and computational burden when directly input into subsequent networks. PCA configurations that retained 90%, 95%, and 99% variance through ablation experiments are compared. It is found that a 95% variance retention rate achieves the best balance in personalized tasks: the semantic consistency score only decreases by 0.012, but the inference speed increases by 23%, and the user clustering NMI index even slightly improves. This indicates that moderate dimensionality reduction helps suppress noisy features and highlight discriminative semantic signals. Therefore, the 95% variance retention strategy is a validated and reasonable choice. To mitigate the scale and semantic gap between numerical behavioral features and high-dimensional semantic vectors, an independent MLP encoder (with LayerNorm) is applied to the behavioral vectors before concatenation, mapping them to a semantic subspace similar to text/visual features. Furthermore, in the subsequent Transformer fusion layer, a modality gating mechanism is introduced to dynamically adjust the contribution weights of each modality, preventing low-dimensional behavioral features from being overwhelmed by high-dimensional semantic vectors.

Semantic Embedding and User Profile Modeling

After completing the feature representation, the system enters the user profile modeling stage. This stage realizes the abstract expression of multi-dimensional user features through semantic aggregation and sentiment layer modeling. First, the fused comprehensive vector U is input into the bidirectional Transformer encoding layer to perform global context modeling on the semantic sequence to capture the implicit dependencies between different modalities. The encoding output is average pooled and weighted by attention to generate a global feature representation Z, which is used to reflect the user's comprehensive semantic state. Then, the system uses a clustering algorithm based on K-Means++ to divide the semantic distribution in the feature space, map the user behavior features to the interest category cluster, and obtain the user interest center vector Ci. At the same time, the social text semantic features are input into the RoBERTa sentiment

classifier to calculate the sentiment tendency score $s_e \in [-1, 1]$, and use it as the sentiment dimension feature in the user profile. For intonation-related features, the system extracts audio intonation parameters (if voice data exists) through the Prosody2Vec model and adds them to the final portrait in vectorized form. After combining the above steps, the user profile can be expressed as Formula (2):

$$P_u = f(Z, C_i, s_e) \tag{2}$$

The function f(·) represents a multi-layer fusion mapping function, which integrates semantic, interest, and emotional features to generate the final multidimensional personality representation vector. It should be noted that audio prosodic features are only introduced as an optional modality in scenarios where users actively authorize and provide voice interaction data (such as intelligent customer service dialogue recordings and brand voice assistant interaction logs). In most conventional marketing scenarios, the system defaults to using only text, behavioral, and visual data to construct user profiles. The paper mentions Prosody2Vec to demonstrate the theoretical completeness of the system in multimodal expansion, rather than forcibly relying on voice data. All processing involving personal voice data complies with GDPR and local data privacy regulations, and anonymization and desensitization are performed with the user's explicit consent. To facilitate the profiles' dynamic adaptability, the system incorporates a sliding window update scheme over temporal axis. New semantic distribution weights are computed, along with interest centers, based on new data, yielding a time-variant user feature trajectory. This allows the user profile to continuously reflect either interest shifts and/or behavioral changes, and the generative model can therefore provide semantic guidance based on the users' latest state when generating content. Figure 1 demonstrates the process of pooling the multimodal user profile construction results into visualizations and analysis. In Figure 1 (a), the horizontal and vertical axes represent the first and second principal components after applying principal component analysis, and present the user distribution across the feature space, which illustrates user characteristics after integrating text semantics, behavioral characteristics, and visual features as part of multimodal analysis. The clustering results indicate that these different user groups fall into different partitions in the feature space, and where the center and grouping

are the aggregation center of each group of users. The direction of the first principal component is contributed to primarily by the semantic features and behavioral features, while the direction of the second principal component reflects a change to visual features, demonstrating the multimodal model's ability to distinguish user interests and content features as interactions with the brand. Figure 1 (b) shows the correlation heat matrix across the multimodal features, which serves to illustrate the strength of relationships between the specific multimodal features. The horizontal and vertical axes correspond to the eight dimensions of features that are used in the modality, while the color depth illustrates the magnitude of a given feature being positively or negatively correlated. The emergence of a positive correlation between the text semantic features and behavioral features indicates consistency between language expression tendencies and interactive behaviors associated with the brand.

The correlation coefficient between the sentiment of the texts and the behavioral engagement of the users (CTR, time on site, conversion, etc.) is 0.84, suggesting that the user sentiment to show emotions is substantially correlated to their true behavioral engagement. Simply put, the greater the user's emotional affiliation with the brand or content, the more positive the user engagement behavior is. Finally, the user profile is saved in a high-dimensional embedding database, which the diffusion generation network can refer to during modeling. The diffusion model reads the user profile vectors that refer to the tendencies and interactive behaviors associated with the brand.

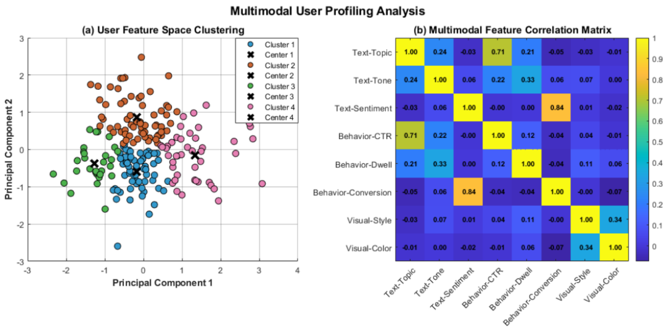

FIGURE 1. Analysis of Multimodal User Profile Construction. (a) User Feature Spatial Cluster Distribution; (b) Multimodal Feature Correlation Matrix

corresponding user profile in the generation phase, and it is able to intervene with semantic constraint signals running through the reverse denoising phase of the diffusion model allowing for personalized control of the generative functions.

B. SEMANTICALLY GUIDED DIFFUSION GENERATIVE NETWORK DESIGN

Although the diffusion model is initially designed for continuous signals, this paper employs a discrete diffusion model for text generation. Specifically, the word sequence after BPE segmentation is used as the initial state, and words with [MASK] are progressively replaced during forward diffusion. The reverse denoising process uses a conditional U-Net to predict the original word distribution at the masked positions. To maintain syntactic integrity, a pre-trained RoBERTa context constraint head is embedded into the denoising network, allowing only candidate words that conform to the language model's probability distribution to be generated, thus maintaining syntactic validity during the diffusion process.

Model Structure Construction and Semantic Condition Injection

For the development of the generative model, this research uses an enhanced variant of Stable Diffusion to perform semantically conditional generation of personalized marketing content. The overall model comprises a forward diffusion process, and a reverse denoising generation process. The noise scheduling function and feature fusion module are restructured to include user semantic vectors as conditional constraints. During training, distinct multi-step Gaussian noise perturbations are first applied to the original content samples (text and visual content) to create a series of noise sequences xt, whose distribution can be stated in (3):

$$q(x_t \mid x_0) = \mathcal{N}\left(x_t; \sqrt{\alpha_t} x_0, (1-\alpha_t)I\right) \tag{3}$$

αt represents the noise preservation factor at time step t. This process simulates the degradation trajectory of content under random perturbations, providing a modeling target for subsequent reverse generation. The reverse generation stage utilizes an improved denoising U-Net network structure. A cross-attention mechanism for the user semantic vector Pu is introduced at each denoising layer, embedding individual features into the feature map of the diffusion model. Before entering the model, the semantic vector is resized to the same dimensional space as the image or text features through a linear projection layer. At each denoising step t, it is dynamically injected via a conditional residual block, enabling semantically guided content reconstruction. During the cross-modal fusion process, the system uses the self-attention mechanism in the Transformer structure to capture the dependencies between semantic and visual features. The specific implementation is: joint attention calculation is performed on the user semantic vector, content encoding vector and time step embedding, and a semantic weight matrix is generated through query-key-value mapping, so that the model automatically focuses on the semantic area related to the user's interest during each denoising. This mechanism ensures that the generated content is continuously controlled by individual characteristics during the structural recovery process, avoiding the generation results from deviating from the target user portrait. In order to further stabilize the training process, the system uses a weighted combination of the noise prediction loss function L_{noise} and the semantic consistency constraint loss L_{semantic} for optimization, as shown in Formula (4):

$$L_{\text{total}} = L_{\text{noise}} + \lambda L_{\text{semantic}} \tag{4}$$

λ is a balancing coefficient that adjusts the weight between generation quality and semantic matching. Through this design, the model maintains the stability of diffuse generation while strengthening the inward constraints of semantics, achieving directional control of content personalization. Figure 2 shows the core training and feature change patterns of the semantic-guided diffusion generation model. In Figure 2(a), the horizontal axis is the number of training rounds; the vertical axis is the loss value (logarithmic scale);

the three curves represent the noise prediction loss, semantic consistency loss, and total loss, respectively. As training progresses, the total loss curve decreases steadily, indicating that the model gradually converges in both semantics and noise reconstruction. Figure 2(b) shows the evolution trajectory of the feature space during the diffusion process. The horizontal and vertical axes represent the two main feature dimensions, and the distribution points of different colors correspond to different diffusion steps.

Semantic Consistency Control and Content Personalization Reconstruction

During the generation phase, the system inputs the user portrait vector Pu and the corresponding marketing theme prompt text, and generates a conditional embedding vector Ec through the text encoder. The semantic control module fuses Ec and Pu and injects them into each layer of the back-diffusion process, ensuring that the semantic control maintains continuity in the time step dimension. To avoid semantic drift, the model calculates the similarity constraint between the generated content and the user semantic embedding after each denoising, using the cosine similarity function, as shown in Formula (5):

$$S = \frac{E_g \cdot P_u}{\|E_g\|\|P_u\|} \tag{5}$$

E_g stands for the intermediate feature representation of the generated content. The similarity score is used to adjust the denoising strength of the diffusion process dynamically. When the semantic deviation increases, the system increases the attention weight with respect to the condition, thereby strengthening the role of individual features in guiding and thus ensuring that the generated content is consistent semantically and stylistically with the user profile.

C. MARKETING THEME ADAPTIVE CONTROL MODULE

Prompt Control Mechanism and Semantic Scheduling Strategy

In the semantic-driven phase of personalized content generation, this research develops a topic-adaptive mechanism using a prompt controller to ensure that generated outputs align with particular marketing themes. The primary step for this module is to convert marketing objective information into semantic control signals that can be understood by the diffusion model, allowing dynamic topic-related scheduling behaviors during the generation. To begin, it builds a marketing topic knowledge base for some scenarios, such as promotional events, holiday events, marketing communications, and product promotions. Each topic contains a series of keywords, style tags, and ordinary semantic patterns that occur with that topic. The topic text is semantically encoded using a combination of TF-IDF (Term Frequency–Inverse Document Frequency) and BERT embedding to obtain a topic vector representation.

Scene Recognition and Theme Adaptive Optimization

To allow the system to automatically adapt to different marketing scenarios, the marketing topic adaptive control module includes a scenario recognition subsystem prior to generation. This subsystem is responsible for analyzing the input task text description, the user's current behavior, and brand context to identify the type of marketing scenario being discussed. This recognition is done using a class model BERT multi-class classification model. The probability distribution that is output by the multi-classification model is used to identify the type of topic category that best matches the semantics features of the present state. After identifying the marketing scenario, the appropriate topic template is retrieved from the topic knowledge base, and the core semantic features are extracted from the topic template to create the Prompt initialization vector, which is placed into a diffusion generation network. To further enhance the adaptive capability and generation stability for differing topics, the system employs a dynamic optimization mechanism for topic parameters. In real time throughout the generation process, the controller checks for semantic drift from the model output and employs feedback signals to update its distribution of the weights in the Prompt. The incremental approach to weight change is handled by a gradient feedback module, where the input is the vector difference of derived semantics and topic semantics,

and the output is the corrective magnitude for each word vector in the Prompt. Through this mechanism, the Prompt topic template word vectors can be adapted. Figure 3 illustrates the capabilities of the topic adaptive control module in relation to semantic consistency and optimizing parameters. Figure 3(a) plots the generation step on the X-axis against semantic similarity on the Y-axis, which illustrates the continuous shifts in semantics when switching to different marketing topics including "Holiday Promotion" and "Product Launch". The model restores semantic coherence effectively after switching the topic, preserving the logical flow of the generated copy. Figure 3(b) plots the adaptive parameter optimization where the number of optimization iterations is on the X-axis and the parameter value is on the Y-axis. The three curves represent the values captured for the learning rate, attention weight, and semantic strength. The learning rate decreases toward the end of the process from about 1.0 to a value of around 0.22, while the attention weight has an increase from 0.3 to around 0.80. The curve for semantic strength has a decrease prior to the end which appears stable at the end of the process.

D. CONTENT DIVERSITY AND STYLE CONTROL MECHANISM

Feature Space Mapping and Style Difference Control

In order to improve the stylistic richness and creative variety of generated content, this research adapts the CLIP feature space distance control algorithm in the reverse generation process of the diffusion model. This mechanism gradually modulates the noise level during the diffusion

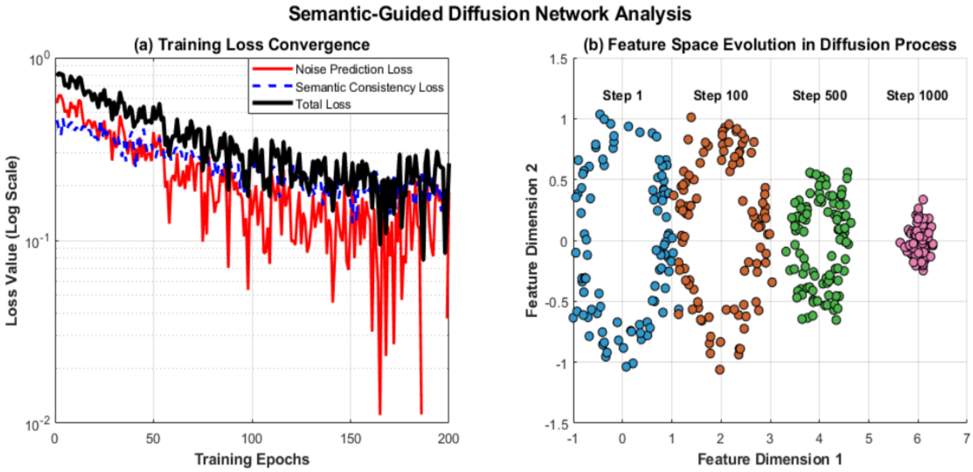

FIGURE 2. Semantically guided diffusion network analysis. (a) Training loss convergence curve; (b) Feature space evolution during diffusion

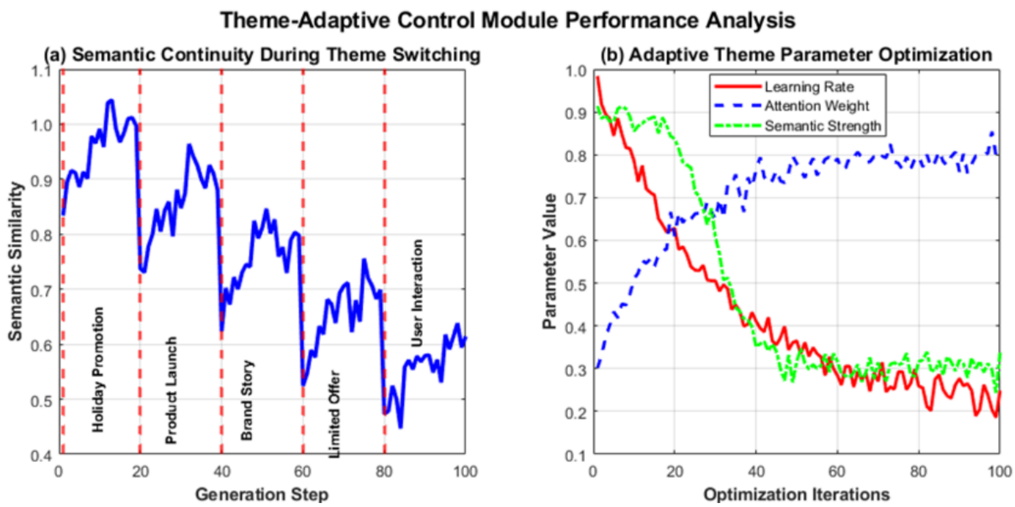

FIGURE 3. Performance Analysis of the Theme Adaptive Control Module. (a) Semantic Continuity Changes During Topic Switching; (b) Adaptive Optimization Trajectory of Topic Parameters

process and encourages the model to maintain a diverse trend while denoising and generating, by limiting the repetition of outputs. At every time step, the diffusion network computes how the distance between the current sample and the reference sample is changing, and uses this to provide feedback for style control. If multiple steps exhibit not divergence in distances, the system concludes that they are converging, and the perturbation weight is increased. To maintain stability in the generation process, it operates within a set allowable range of noise which prevents excessive perturbation from disrupting the economy of meanings. The approach balances content stability and diversity in a dynamic, content distance-based feedback control manner during the creative generation process, producing diverse outputs while keeping semantic consistency.

Style Parameter Adaptation and Creativity Enhancement Mechanism

In managing the feature differential control, the system also designs a style parameter adaptation module to facilitate the purposefully controlled content style adjustment. The style parameter adaptation module can statistically model the style features in the training sample set distribution to create a style encoding matrix that can include, but not be limited to, the dimensions of color saturation, tone intensity, language rhythm, etc. In the content generation process, this style baseline vector is established, determined by the target user profile and the topic prompt. The style baseline vector is mapped to the CLIP feature space, and a comparative style distance is analyzed with the feature vector representation of the generated content. If the style distance exceeds a set range, the system can automatically adjust the mean and variance parameters of the adaptive normalization layer during the diffusion step, thus modifying the feature distribution for style reconstruction.

E. PERSONALIZED CONTENT QUALITY OPTIMIZATION AND SCREENING

Multi-Dimensional Quality Evaluation Network Construction and Scoring Mechanism

To ensure that generated content meets personalized marketing needs in terms of semantic expression, creative differentiation, and emotional fit, this study designs a multi-dimensional quality assessment network to perform multi-level scoring and screening on the diffusion model output. The system first feeds the generated text and image into a multimodal encoder, extracting semantic and visual feature representations using the CLIP architecture. The encoded results are fed into three independent sub-networks to calculate the semantic consistency score, creativity index, and emotional fit, respectively. The semantic consistency sub-network calculates similarity between the BERT embedding and the user profile semantic vector to measure the degree of fit between the generated content and the user's characteristics. The creativity index sub-network incorporates a convolutional feature transformation module to explore the differences in features between the generated and existing samples to assess the degree of innovation and differences in style. The emotional fit sub-network employs the RoBERTa sentiment classification layer to detect the content's emotional direction and performs cosine alignment with the user's historical emotional feature vector to calculate the emotional fit strength.

Adaptive Screening and Brand Tone Alignment Mechanism

Once the multi-dimensional quality evaluation is done, the system goes to the content screening phase. To ensure that the output matches the brand tone and user style, the system has developed an adaptive screening module to provide dynamic optimization and filtering of candidate content. The module first constructs an embedding space for brand tone, then utilizes a pre-trained language model to extract tone feature vectors from the brand copy corpora to constitute a style baseline representation. The generated content is encoded by the tone recognition model as well as the brand baseline vector to calculate the tone departure. If the departure exceeds a cutoff, the system triggers a tone correction submodule operation, which can adjust generated text in word vector distribution and grammatical structures, changing expression style until it converges back to the brand tone. In the final output process, the system constructs a high-quality result set from within the pool of candidate samples. The system retains only the top 10% of generated samples for each marketing task and captures

the semantic features for each sample to record to a database to serve as feature prior update data for future generation tasks, thus creating a closed-loop optimization mechanism for content quality. Figure 4 demonstrates the results of the personalized content creation module with respect to quality improvement and brand tone monitoring. In Figure 4(a), tone dimension 1 (formality) is represented on the horizontal axis, and tone dimension 2 (emotional intensity) is shown on the vertical axis. The original content is represented by red dots; the optimized content is represented by blue squares; the brand tone benchmark is represented by black triangles. Each of the optimized samples overall moves toward the brand tone, indicating that the model is able to effectively adjust the tone of the content at the semantic and emotional level. Figure 4(b) displays the number of optimization rounds on the horizontal axis and the quality score on the vertical axis. The vertical axes denote the trend for semantic consistency, creativity index, emotional alignment, and brand alignment, as each of the optimization rounds increases. Each of the indicators shows a steady increase in quality over the rounds of optimization, demonstrating that the model is able to incrementally improve the overall quality of content through repeated iterations.

III. EVALUATION OF CREATIVE CONTENT GENERATION SYSTEMS

The study's evaluation experiments leverage a multi-source, diverse dataset. The data sources primarily include a public marketing text dataset (AdDataset) and a brand corpus built by a company. The public dataset provides some sampled marketing content in various contexts, including promotional advertisements, holiday events, and brand promotional text. The company-built corpus, constructed from historical brand marketing content and user interaction comments, is used to create user interest and emotional preferences and brand tone characteristics. All data are manually cleaned and annotated hierarchical, creating approximately 120,000 training and validation samples with multi-dimensional information such as a brand label, topics category, sentiment polarity, and interaction records. Although the system supports multimodal output (including fabric texture images and accompanying text), this paper focuses on evaluating text generation capabilities, as language remains the core carrier of marketing information. The visual generation component is published separately as a subsequent work. Therefore, all evaluation metrics revolve around text quality, but the generation process is consistently guided by CLIP visual semantics to ensure semantic alignment between the text and the visual features of the fabric.

A. SEMANTIC CONSISTENCY ASSESSMENT

During the evaluation phase, the generated content, user interest descriptions, and topic prompts are first vectorized and encoded using the BERT model. The similarity between the generated text and the topic text in semantic space is then calculated, and high-frequency keywords are extracted and matched against the topic keyword library to detect

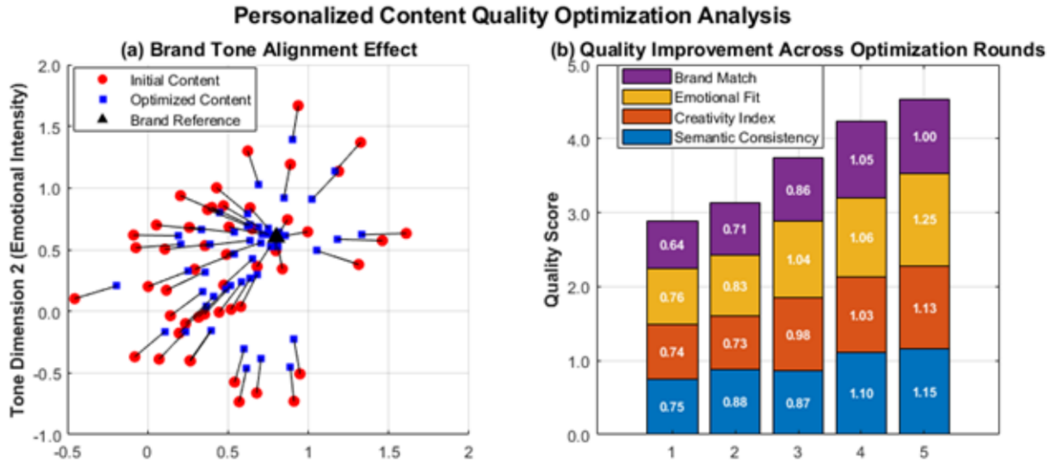

FIGURE 4. Personalized Content Quality Optimization Analysis. (a) Brand Voice Alignment Results; (b) Quality Improvement During Multiple Rounds of Optimization

semantic deviation. The system sets a similarity threshold and marks samples exceeding the threshold as consistent content. To verify the stability of the evaluation, samples are randomly selected for manual review. The BERT layer output weights are adjusted based on the difference between the manually annotated semantic relevance results and the model-calculated values. Finally, the consistency score results are summarized, and the semantic accuracy of the overall generation task is statistically analyzed to verify the semantic stability of the model under topic guidance. Figure 5 shows a comprehensive comparison of four text generation models in terms of semantic consistency. In Figure 5(a), the horizontal axis shows the semantic similarity score, and the vertical axis shows the

probability density, reflecting the degree of closeness between the generated content of different models and the reference semantics. The distribution curves show that the proposed model (green) has the highest density in the high-similarity range, with a mean of approximately 0.842. This significantly outperforms the rule-based template method (mean approximately 0.657), the GPT marketing model (approximately 0.768), and the BERT fine-tuned model (approximately 0.810), demonstrating that its generated text is more stable in terms of semantic preservation. In Figure 5(b), the horizontal axis shows different marketing topics, and the vertical axis shows the semantic consistency score. It can be observed that the semantic consistency scores for product launch events and corporate social responsibility (CSR) campaigns reach 0.92 and 0.91, respectively. The proposed model maintains a score above 0.85 for all topics, performing particularly well for product launch content, demonstrating its robustness and generalization ability in multi-context content generation.

B. CREATIVE DIVERSITY ASSESSMENT

Initially, the system extracts the syntactic tree and CLIP feature embeddings of all the generated content, and compares the feature distance distribution across the content. It then aggregates the overall feature distance for each content group, which is the magnitude of the samples' dispersion in the semantic space. If the

average distance between samples converges or shows a clear convergence trend, the system recognizes that the generated results are converging, and optimizes the diffusion noise perturbation coefficient to regenerate the results. To ensure no incidental data drift occurs, the evaluation is performed over multiple batches of the generation task and their variance change trend can be tested. Finally, by standard deviation the average distance of samples in each batch the information can be tabulated based on diversity scores to assess the model's creative generation ability involving multiple scenarios and to ensure the output content remains stylistically different meeting an innovative expression involving a single unified theme. Figure 6 shows the comprehensive evaluation results of creative diversity and semantic consistency in this paper's multi-scenario personalized marketing generation task. Figure 6(a) shows different marketing themes (such as Holiday Sale, Product Launch, etc.) on the horizontal axis and creative diversity scores on the vertical axis. The proposed model consistently scores above 0.85 in all scenarios, significantly outperforming the GPT marketing model and the rule-based template method, demonstrating its ability to more effectively maintain creative diffusion and semantic differentiation in content generation. Figure 6(b) shows the creative diversity score on the horizontal axis and the semantic consistency score on the vertical axis, with the ideal region located in the upper right corner. The generated results of this method are concentrated in this region, demonstrating that it maintains high

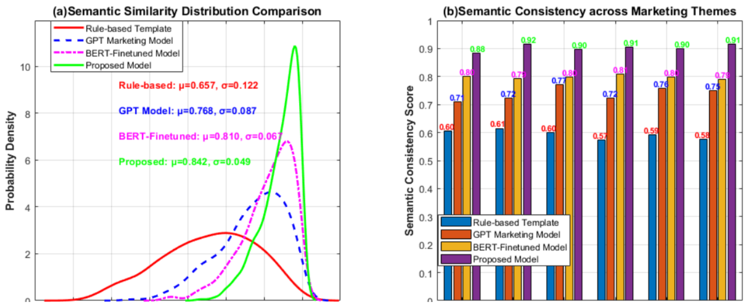

FIGURE 5. Semantic consistency evaluation results analysis. (a) Comparison of semantic similarity distribution; (b) Semantic consistency performance across different marketing themes

semantic coherence and topic relevance while maintaining the innovativeness of the generated content.

To enhance the reliability of the evaluation results, in addition to the automated indicators, a double-blind human evaluation experiment is conducted. Twelve professional reviewers with marketing copywriting experience are invited to score 100 generated pieces of content each from the proposed model, the GPT marketing model, and the rule-based template method on a 5-point scale across three dimensions: "novelty", "uniqueness of expression", and "thematic extension ability". The results show that the proposed model achieves an average score of 4.12 (standard deviation 0.31) in the human evaluation, significantly higher than the GPT model's 3.65 and the template method's 2.87 ($p < 0.01$, paired t-test). This verifies that the output of the creative diversity evaluation network is highly consistent with human subjective judgment, enhancing the credibility of the "creative superiority" conclusion.

C. EMOTIONAL COMPATIBILITY ASSESSMENT

The system employs the RoBERTa sentiment recognition model to group the produced text and generate a sentiment distribution vector. In the meantime, it retrieves sentiment preference vectors based on historical user interaction data and evaluation data. The difference in similarity is calculated based on two criteria: sentiment type consistency and intensity closeness. For robustness, the system concatenates samples with disparate intonations and contexts to test. The average deviation value is developed for the grouping, where high-deviation samples are tagged and analyzed as possibly semantically imbalanced. Afterwards, the system produces a sentiment consistency curve graph from the matched results and compares the sentiment stability of each generated batch in an effort to verify the sentiment personalization capability of the model. As demonstrated in Figure 7, the emotional fit of personalized marketing content is evaluated based on the diffusion model. The x-axis of Figure 7(a) displays six measures of emotional fit (emotional intensity, emotional persistence, emotional authenticity, emotional consistency, emotional adaptiveness, and emotional appeal), while the y-axis captures eight core emotion categories. From the heat map, it is clear that the emotions associated with "trust" and "joy" align highly with all the measures of emotional fit, suggesting that the model exhibits a high degree of naturalness and consistency in generating content and marketing expressions related to positive emotions. Conversely, negative emotions such as "fear" and "disgust" score relatively low, reasoning that the model tends to favor creating marketing expressions

centered on positive emotions. Figure 7(b) displays the emotional intensity consistency associated with the different user segments. Interestingly, the high-income user segment has the strongest consistency with a score of 0.91 on the strong emotional dimensions. This indicates that the content generated by the model is stable and consistently displayed with this segment. In contrast, the consistency of the middle-age family segment on the complex emotional dimension is only found to be 0.69, emphasizing how the model has decreased adaptability towards generating combinations of multidimensional emotional content. Reading multidimensional emotional traits and examining different user segments, the proposed method is capable of producing highly fit content that is also stable in terms of emotional intensity across most emotional types and important user segments.

It is worth noting that the model scored low (mean <0.55) on the fit of negative emotion categories such as "fear" and "disgust". This behavior primarily results from the scarcity of negative-emotion marketing samples in the training data–in the brand corpus and AdDataset, negative emotion-oriented marketing texts account for less than 3.2%. This results in the diffusion model lacking sufficient learning of negative emotion semantic vectors during the denoising process. This reflects the current system's limitation in its ability to control the entire emotion spectrum, and future improvements can be made by introducing emotion-enhanced datasets and controllable emotion guidance mechanisms.

D. BRAND TONE MATCH ASSESSMENT

In the evaluation stage, a benchmark corpus of brand voice is created, and brand style vectors are developed using a pretrained language model. The model is then used to innovativeness of the generated content.

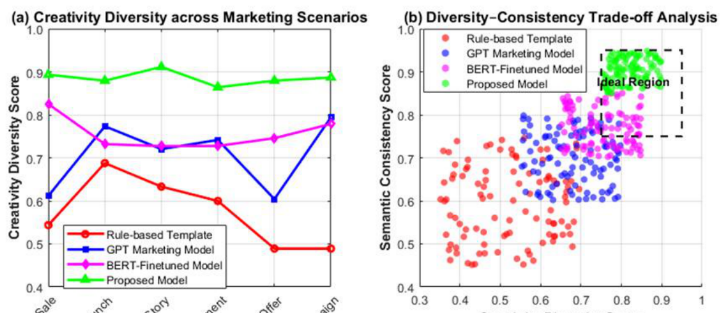

FIGURE 6. Creative Diversity Evaluation Results Analysis. (a) Creative Diversity Distribution in Different Marketing Scenarios; (b) Diversity-Consistency Balance Analysis

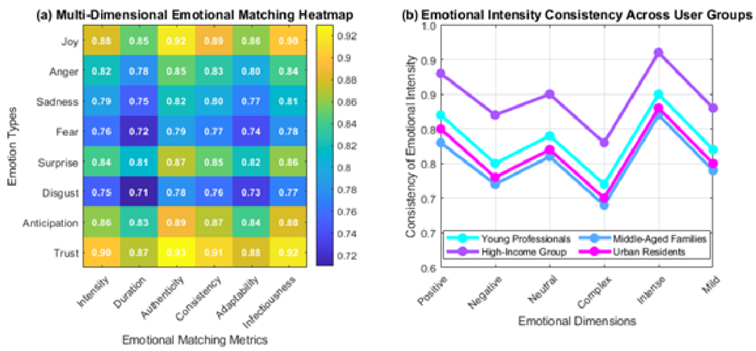

FIGURE 7. Emotional Compatibility Assessment Results. (a) Heatmap of Matching Across Multiple Emotional Dimensions; (b) Emotional Consistency Curves for Multiple User Groups

encode the generated text and the offset or distance between the two text representations in vector space is calculated. Text samples with smaller offsets are labeled as a match in style. As a follow-up, in order to ascertain accuracy of a match, the system includes a grammatical pattern comparison algorithm which computes quantitative measures of sentence complexity, tone structure, and rhetorical use. Retrospective analysis on deviating samples identifies common types of imbalances in intonation in order to fine-tune the model as necessary in subsequent steps. Finally, the model produces an overall brand voice consistency report based on the mean match scores from the sample text in order to gauge how well the output from the model aligned with the brand context. To ascertain how well the generated text aligned with the brand voice, the study conducts a multiple-dimensional measure across six selected samples on brand style vector similarity, grammatical complexity, tone consistency and rhetorical match rate where an overall score is generated and presented in Table 1.

Table 1 presents the core results of the brand tone match assessment, which measures the degree of alignment between the system-generated content and the brand benchmark corpus in terms of tone, sentence structure, and rhetoric. Sample S004 achieves the highest overall match score (0.94), with a brand style vector similarity of 0.961 and a grammatical complexity deviation of only 0.10, indicating that the text

closely matches the brand context in terms of tone control and language structure. Sample S003, on the other hand, scores relatively low (0.86), primarily due to a low style vector similarity (0.882) and a large grammatical deviation (0.19), reflecting that some sentence structures deviate from brand expression conventions. Overall, the system achieves an average brand similarity of 0.927 and an average grammatical deviation of 0.14 across the six samples, resulting in an overall match of 0.905, demonstrating the model's strong stability and consistency in brand tone control.

E. CONTENT QUALITY AND DISSEMINATION POTENTIAL ASSESSMENT

Evaluations of content quality and dissemination potential evaluate overall performance of generated outputs with respect to readability, interest and dissemination potential. The first step is to score fluency in language through a structured evaluation of both grammaticality and structure, and then to score the naturalness of language based on the variety of sentences and balance of the frequency distribution of words. A simulated user click model is then introduced in order to project the dissemination potential of generated

TABLE 1. Brand tone matching evaluation results

Sample ID	Brand Style Vector Similarity (Cosine)	Grammatical Deviation	Complexity	Tonal Consistency Score	Rhetorical Match Rate (%)	Overall Matching Score
S001	0.945	0.12		0.93	92.5	0.93
S002	0.918	0.15		0.89	90.1	0.89
S003	0.882	0.19		0.85	87.6	0.86
S004	0.961	0.10		0.94	94.8	0.94
S005	0.901	0.17		0.87	89.3	0.88
S006	0.953	0.11		0.92	93.5	0.93
Average	0.927	0.14		0.90	91.3	0.905

TABLE 2. Comprehensive evaluation table of content quality and dissemination potential

Sample ID	Fluency Score	Structural Integrity	Linguistic Diversity	Predicted CTR (%)	Actual CTR (%)	Quality Index
C001	0.93	0.91	0.88	5.6	5.3	0.90
C002	0.89	0.87	0.84	4.8	4.5	0.85
C003	0.95	0.93	0.91	6.2	6.0	0.93
C004	0.87	0.85	0.82	4.4	4.1	0.83
C005	0.91	0.90	0.86	5.2	5.0	0.88
C006	0.96	0.94	0.92	6.5	6.3	0.94

content by calculating the interaction rate regimen for different channels. The system compares generated content across numerous iterations, tracking the difference between click predictions to the user experience, in order to variably weight calculation for model evaluations. The evaluation data ultimately outputs a composite quality score, and the commonalities among the characteristics of the six sample high-quality items generated provides recommendations for improving content quality based on linguistic characteristics. This ultimately directs the model retraining of the system, and iteratively increases the performance of the marketing generation system. Table 2 evaluates the quality and dissemination potential of generated content across dimensions like fluency, structure, diversity, and click-through rates (CTR). Sample C006 excels with a fluency score of 0.96, structural integrity of 0.94, a predicted CTR of 6.5%, and an actual CTR of 6.3%, resulting in the highest quality index of 0.94. In contrast, C004 scores 0.83, with a CTR of 4.4%, highlighting areas for improvement. The average quality index across samples is 0.89, with a minimal variance between predicted and actual CTR (±0.3%), indicating strong stability and predictive accuracy for marketing optimization.

IV. CONCLUSION

This paper identifies the problems of extreme homogeneity, lack of originality, and user impersonality in traditional marketing content generation systems, and proposes a personalized marketing content generation system that includes a diffusion model. The personalized system involves the integration of multimodal user profiles, a semantically directed diffusion generation network, a marketing theme adaptive control module, a content diversity and style regulation, and a multidimensional quality optimization filtering network, which altogether enables a closed-loop user feature extraction to personalized creative content generation process. During the content generation process, the system incorporates user-semantics vectors, along with brand informed-tone, within a topic-directed diffusion generation framework. The system also mitigates content homogeneity and maximizes relevance using feature space distance and a quality assessment framework. Research findings indicate that the system designs and generates marketing content relevant to consumer personalization and self-efficacy in articulation, while producing some creative expression, addressing the issues of the previous generation of intelligent marketing content. The system is scalable and creative for automated intelligent marketing content generation, allowing the textile industry to effectively communicate complex product stories across various platforms.

AUTHOR CONTRIBUTIONS

Jun Feng designed, collected and analyzed the data, and drafted the manuscript. Jun Feng conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Jun Feng 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] G. H. Lee, K. J. Lee, B. Jeong, and T. Kim, "Developing personalized marketing service using generative AI," IEEE Access, vol. 12, pp. 22394–22402, 2024, doi: 10.1109/ACCESS.2024.3361946.

[2] V. Soni, "Adopting generative AI in digital marketing campaigns: An empirical study of drivers and barriers," Sage Science Review of Applied Machine Learning, vol. 6, no. 8, pp. 1–15, 2023.

[3] S. Dimitrieska, "Generative artificial intelligence and advertising," Trends in Economics, Finance and Management Journal, vol. 6, no. 1, pp. 23–34, 2024, doi: 10.69648/EYZI2281.

[4] C. V. Lim, Y. P. Zhu, M. Omar, and H. W. Park, “Decoding the relationship of artificial intelligence, advertising, and generative models,” Digital, vol. 4, no. 1, pp. 244–270, 2024, doi: 10.3390/digital4010013.
[5] P. Cillo and G. Rubera, “Generative AI in innovation and marketing processes: A roadmap of research opportunities,” Journal of the Academy of Marketing Science, vol. 53, no. 3, pp. 684–701, 2025, doi: 10.1007/s11747-024-01044-7.
[6] V. Kumar, P. Kotler, S. Gupta, and B. Rajan, “Generative AI in marketing: Promises, perils, and public policy implications,” Journal of Public Policy & Marketing, vol. 44, no. 3, pp. 309–331, 2025, doi: 10.1177/07439156241286499.
[7] M. Heitmann, “Generative AI for marketing content creation: New rules for an old game,” NIM Marketing Intelligence Review, vol. 16, no. 1, pp. 10–17, 2024, doi: 10.2478/nimmir-2024-0002.
[8] K. B. Ooi, A. Koohang, E. C. X. Aw, T. H. Cham, C. Cobanoglu, C. Dennis, et al., “Unveiling the potential of generative artificial intelligence: A multidimensional journey into the future,” Industrial Management & Data Systems, vol. 125, no. 2, pp. 417–432, 2025, doi: 10.1108/IMDS-10-2023-0703.
[9] A. S. Gupta and J. Mukherjee, “Framework for adoption of generative AI for information search of retail products and services,” International Journal of Retail & Distribution Management, vol. 53, no. 2, pp. 165–181, 2025, doi: 10.1108/ijrdm-05-2024-0203.
[10] I. Thomas, “Using generative AI to turbocharge digital marketing,” Applied Marketing Analytics, vol. 9, no. 3, pp. 270–280, 2023, doi: 10.69554/DXFN2668.
[11] E. Mogaji, G. Viglia, P. Srivastava, and Y. K. Dwivedi, “Is it the end of the technology acceptance model in the era of generative artificial intelligence?” International Journal of Contemporary Hospitality Management, vol. 36, no. 10, pp. 3324–3339, 2024, doi: 10.1108/IJCHM-08-2023-1271.
[12] B. McCarthy, “Can generative artificial intelligence help or hinder sustainable marketing? An overview of its applications, limitations and ethical considerations,” Journal of Resilient Economies, vol. 4, no. 2, pp. 18–33, 2024, doi: 10.25120/jre.4.2.2024.4153.
[13] J. Kim and G. Choi, “Assessing the impact of generative artificial intelligence on customer engagement in business-to-customer scenarios,” Asia-Pacific Journal of Convergent Research Interchange, vol. 10, no. 2, pp. 89–104, 2024, doi: 10.47116/apjcri.2024.02.09.