International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6584-6596

https://doi.org/10.70517/ijhsa463566

Research on Deep Generative Model-Driven Optimization Strategies for Digital Media Content Generation and Dissemination Effects

Wenliang Ji[1,2], Ming Jin[3,*] and Yixin Chen[4]
[1] Institute of Journalism, Communication University of China, Beijing, 100024, China
[2] Institute of Military Management, National Defense University, PLA, Beijing, 100091, China
[3] Institute of Humanities and Law, Yanshan University of China, Qinhuangdao, Hebei, 066004, China
[4] Institute of Journalism and Communication, Henan Academy of Social Sciences, Zhengzhou, Henan, 450000, China
Corresponding authors: (e-mail: tongjikimi@126.com).

Abstract The current digital media have problems such as ambiguous communication orientation, uneven content quality and stereotyped discourse patterns, which not only affect the communication effect of the media, but also limit its sustainable development. This study explores the application strategy of deep generative model in digital media content generation and communication effect optimization. The study proposes a variational digital media content generation model based on adversarial training (VAE-IAT), which combines the advantages of variational self-encoder and generative adversarial network to realize the generation of high-quality digital media content through the collaborative work of three modules, namely, encoder, generator, and discriminator; and at the same time, it constructs three optimization paths to enhance the communication effect, including improving communication precision, deepening content production, and strengthening the interaction with the audience. The experiments are conducted with three datasets: MNIST, SVHN and CelebA, and the results show that the VAE-IAT model exhibits excellent generation ability on all three datasets, and the FID scores are maintained below 6, which is significantly better than that of the control model. The results of the dissemination effect validation experiment show that the experimental group reaches significant differences with P-values of 0.002, 0.007, 0.003, and 0.005 for the four dimensions of media comments, media retweets, media interactions, and media likes, while the control group does not show any significant differences. The results of the study confirm that the digital media content generation technology driven by deep generative model can effectively improve the quality of content, and the communication optimization path constructed based on it can significantly improve the communication effect of digital media, which provides new technical support and theoretical guidance for digital media content creation and communication strategy.

Index Terms deep generative model, variational self-encoder, generative adversarial network, digital media content, dissemination optimization, content generation

I. Introduction

In the digital era, the mode and form of traditional media have changed dramatically, replaced by the rise and development of digital media. Digital media, including new media, social media, mobile applications and other forms, breaks the one-way communication mode and geographical limitations of traditional media, and provides an interactive communication platform with users, who can participate in the creation and dissemination of media content, as well as real-time browsing and access to media content around the globe through cell phones, tablets, and other terminal devices [1]-[4]. In terms of information dissemination, digital media realizes real-time and global dissemination of information, and its customization and personalization features make information dissemination more accurate and targeted, and the media can push relevant content according to users' interests and needs to improve the effect of information dissemination [5]-[7].

With the increasing demand for digital media content, traditional content creation methods can no longer meet the rapidly changing market demand. Instead, digital media content generation relies on advanced computer graphics and signal processing technologies. Through the use of computer software and hardware tools, various operations can be performed on digital media, such as generation, editing, compositing, and rendering, and these techniques are capable of simulating the real world and creating virtual worlds of unlimited imagination [8], [9]. It is an important part of digital art, movie production, advertisement design and other fields, which can bring richer and innovative digital media experiences to users [10]-[12]. However, the increased efficiency of digital media content generation has reduced the quality of generation and customized content has increased homogenization of generated content due to the similarity of generation systems [13], [14]. In terms of dissemination, due to the rapid dissemination characteristics of digital media, platforms are not well regulated, leading to a wider dissemination of false information in media platforms and an exponential increase in false information incidents [15]. By using deep generation models such as variational self-encoders, generative adversarial networks, and autoregressive models, automated content generation can not only significantly improve the production efficiency and provide greater flexibility in personalization and customization, but also carry out false information detection, which contributes a new vitality to optimize the content generation and dissemination effects of digital media [16], [17].

As an important carrier of information dissemination, digital media plays an increasingly important role in today's information society. With the rapid development of Internet technology and artificial intelligence, digital media content generation and dissemination methods are experiencing profound changes. In recent years, breakthroughs in deep learning technology have provided new possibilities for digital media content creation, especially the emergence of deep generative models, which provide powerful technical support for the automated generation and personalized push of digital media content. By learning and analyzing a large amount of data, deep generative models can generate new data with similar characteristics to the original data, a characteristic that makes them show great potential in the field of multimedia content generation such as images, text, and audio. However, while the current digital media is developing rapidly, it still faces problems such as imprecise communication positioning, unstable content quality and unsatisfactory communication effect. These problems not only limit the improvement of digital media communication effect, but also affect the user experience and the realization of media value. How to use deep generation model to improve the quality of digital media content and optimize the communication effect has become an important issue of common concern for academia and industry. Existing research mainly focuses on the technical improvement of the deep generation model itself, or the single-dimension analysis of the communication effect of digital media, lacking systematic research from technology to application. In particular, there is still a lack of research on transforming the technical achievements of deep generative modeling into practical applications in digital media. In addition, there is a relative lack of research on the optimization path of digital media communication effect driven by deep generative model, which needs to be further explored and improved. Based on this, this study focuses on the combination of deep generative model and digital media communication, aims to construct a generative model that can improve the quality of digital media content, and explores the optimization path of digital media communication effect based on this model.

In this study, the basic principles and characteristics of variational self-encoder and generative adversarial network are firstly summarized, based on which the variational digital media content generation model based on adversarial training (VAE-IAT) is proposed. The model extracts short-range and long-range features of media content through an encoder, generates media content sequences based on the fusion of hidden variables using a generator, and performs adversarial training through a discriminator to improve the quality of the generated content. Meanwhile, this study constructs the optimization path of digital media communication effect from three aspects: improving communication accuracy, deepening content production and strengthening interaction with audiences. In the experimental validation part, several public datasets are used to evaluate the performance of the proposed model, and the validity of the communication optimization path is verified by scale testing and statistical analysis. Through this research idea, this paper aims to provide technical support for digital media content generation and theoretical guidance and practical reference for digital media communication effect optimization.

II. Optimization of Media Content Generation and Communication Effect under Deep Generation Modeling

II. A. Depth Generation Model

Deep generative modeling is the process of generating new samples by learning the distribution of a given dataset using deep learning techniques [18]. When training deep generative models, the data used can be in different forms such as images, text, sound, etc. When this process is applied to a set of images, it is called generative image modeling. There are two main categories of generative image modeling techniques: unsupervised techniques and conditional unsupervised techniques. Unsupervised techniques have methods such as variational self-encoders and generative adversarial networks. In this section the main focus is on variational self-encoders and generative adversarial networks.

II. A. 1) Variational Auto-Encoder (VAE)

VAE has achieved significant improvements in the representation capabilities of autoencoders [19]. The method provides a probabilistic distribution to describe the vectors of the latent space, the encoder outputs a conditional mean and standard deviation that are responsible for constructing the distribution of the latent space, resulting in a continuous latent space and allowing for random sampling and interpolation of the latent space, whereas the representation of the latent space of the auto-encoder is discrete, and it is this property that has led to the widespread use of the variational auto-encoder in generative tasks. The core of the model design is that if one wishes to model the latent probability distribution of the data, a simple and parametrically known distribution can be predetermined from which new data can be sampled.

Consider a typical Bayesian learning setup for generative modeling, where there is some observable data $x=\{x_1,x_2,\cdots,x_n\}$ that needs to be generated from some latent variables $z=\{z_1,z_2,\cdots,z_m\}$. Mathematically, the goal is to maximize the probability of each x in the observed data under the generation process, as shown in equation (1):

$$p(x)=\int p(x\mid z)p(z)dz \tag{1}$$

The key idea behind the VAE is that in higher dimensions, most z will set $p(x\mid z)$ to zero. Even the conditional probability $p(x\mid z)$ is still more difficult to solve for, so the VAE turns to solving for the posterior probability $p(z\mid x)$, which provides a distribution of latent variables that are more likely to yield observations x. An attempt is made to approximate the posterior probability $q_\phi(x^{(i)}\mid z)$ of the real data using the distribution $q_\phi(x^{(i)}\mid z)$ of the model output and to measure the difference between the two distributions using the Kullback-Leibler (KL) scatter measure, which needs to be minimized by the generating process, i.e., the objective function in training. This is shown in equation (2):

$$\mathrm{KL}(q_\phi(z\mid x^{(i)}),p_\theta(z\mid x^{(i)}))=E_{q_\phi(z\mid x^{(i)})}\log\frac{q_\phi(z\mid x^{(i)})}{p_\theta(z\mid x^{(i)})} \tag{2}$$

where ϕ is the encoder parameter and θ is the decoder parameter.

II. A. 2) Generative Adversarial Networks (GAN)

A generative adversarial network is able to learn from a set of training data and generate new data with similar characteristics to the training data [20]. For example, a generative adversarial network trained on face photographs can generate completely fictitious and realistic faces. A generative adversarial network consists of two neural networks that oppose each other, a generator network and a discriminator network. The generator generates fake data and the discriminator distinguishes the generator's fake data from real examples. Through cyclic training of both, the generator learns to generate more believable examples.

The generator network G receives a random noise z and generates an image that approximates the sample as closely as possible, denoted $G(z)$. Both the generated image from the generator and the real image from the training set are randomly fed into the discriminator network, and the discriminator does not know whether a particular input is from the generator or from the training set. The output $D(x)$ of the discriminator network represents the probability that image x is a real image, if $D(x)=1$ it means that the discriminator network thinks that the input must be a real image, if $D(x)=0$ it means that the discriminator network thinks that the input must be a fake image. During the training process, the two networks fight against each other, eventually forming a dynamic equilibrium.

The training of a generative adversarial network is more complex than that of a standard feedforward neural network; the generator and discriminator are actually two neural networks trained separately, but they also interact with each other, so they cannot be trained completely independently and it is difficult to determine exactly when the generative adversarial network has converged. Since the generator and discriminator have their own loss functions, they must be trained separately. It is possible to train the generator or discriminator separately for one or more iterations and keep the other neural network weights constant.

The adversarial loss function used during the training of the generative adversarial network is shown in equation (3):

$$\min_G V(G,D)=\min_G\max_D E_{x\sim p_{data}}[\log D(x)]+E_{z\sim p_z}[\log(1-D(G(z)))] \tag{3}$$

The generator tries to minimize the output of the above loss function, while the discriminator tries to maximize it. This loss function can act on both the generator and the discriminator. The generator can only minimize the second term in the loss function, since the first term depends only on the discriminator.

To ensure that $V(G,D)$ is maximized, it is common to iteratively train the discriminator k times, followed by 1 iteration of the generator (although in practice it has been found that k is usually sufficient to take 1). When the generator G is fixed, the optimal discriminator D_G^* can be derived from $V(G,D)$, as shown in equation (4):

$$D_G^* = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \tag{4}$$

GAN has been proposed mainly for modeling natural images, and can generate realistic results in image generation tasks. But the biggest problem of the original GAN is that it is difficult to train, and the gradient is easy to disappear. Even for the generation of low-resolution images, the model is difficult to converge, resulting in the generation of unrealistic images, lack of diversity, etc. Since then, researchers have made different degrees of improvements in the model structure, objective function and application direction of generative adversarial networks.

II. B.Adversarial Training Based Media Content Generation Models

In the field of digital media content generation, more and more researchers are now combining VAE and GAN networks, i.e., using VAE as a generative model output to generate digital media content, and based on this, using GAN networks to enhance the performance of VAE model generation. In this subsection, based on VAE and GAN, an adversarial training-based variational media content generation model (VAE-IAT) is proposed to enhance the quality of generated digital media content.

II. B. 1) Model Architecture

The adversarial training-based variational digital media content generation model (VAE-IAT) contains three modules: encoder, generator and discriminator, and the network structure model is shown in Fig. 1.

The VAE-IAT model media content generation process is as follows: train an encoder E_α with parameter α, extract the short-range feature vector f_t^l of the media content and the long-range feature vector f_t^c of the media content. Process the long range feature vector f_t^c and short range feature vector f_t^l through the fully connected layer of the encoder, and output a fusion hidden variable z_t after reparameterization. The fusion hidden variable z_t is combined with real media content data as input to the generator G_θ, and the generator G_θ is trained to generate media content sequences of length T $X_{1:T} = \{x_1, x_2, \cdots, x_t, \cdots, x_T \mid z_t\}$, $t \in (1,T)$, $z_t \in E_\alpha$, and z_t denotes the fused hidden variable sampled by the encoder at moment t. S_t is the sequence of media content generated at moment t $S_t = (x_1, x_2, \cdots, x_t)$, $S_{t'}$ is the sequence of media content complemented by Monte Carlo search $S_{t'} = (x_1, x_2, \cdots, x_t, x_{t+1}, \cdots, x_T)$, and the action a_t is defined as the next word chosen x_{t+1}; a discriminator D_ϕ with parameter ϕ is trained by real data (RealText) and generated data $S_{t'}$ complemented by MCsearch, the discriminator D_ϕ extracts the feature vector f_t^d using CNN as the feature extraction module. Adversarial training of generated data, discriminate the generated sentence generation reward value, and evaluate the generation effect of the sentence. The generator G_θ uses a strategy gradient to interact with the environment to generate the next word x_{t+1}.

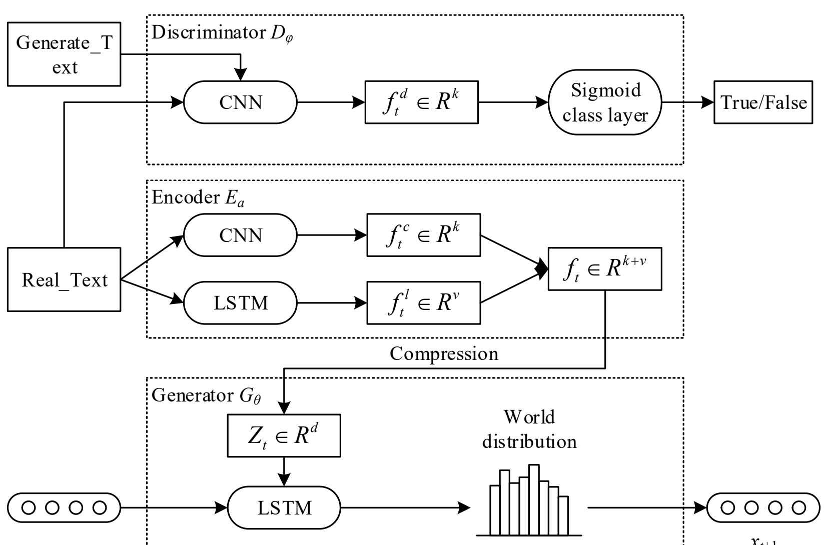


Figure 1 Network structure model

II. B. 2) Encoders and Generators

In the generation process of VAE, the generation of hidden variable z requires the computation of the posterior distribution $p(z|x)$, but the posterior distribution $p(z|x)$ is difficult to be computed directly, so $q(z|x)$ is constructed to approximate the posterior distribution $p(z|x)$, and Kullback-Leibler (KL) scattering is used to measure the similarity between $q(z|x)$ and $p(z|x)$. degree, the loss function of VAE is shown in equation (5):

$$l_{VAE} = l_{rec} + D_{KL}\left(q(z|x) \| p(z|x)\right) \tag{5}$$

where l_{rec} is the reconstruction loss of the generator and D_{KL} measures the similarity of $q(z|x)$ to the true posterior distribution $p(z|x)$.

The encoder mainly contains feature extraction module and feature processing module. In the feature extraction stage, in order to ensure that the model can fully and effectively extract the semantic and sequential information of the media content in the feature processing stage, the model in this paper adopts the encoder short-range feature extraction module (LSTM) to extract the feature information of the media content data, and constructs the short-range feature vector f_t^l; The long-range feature extraction module (CNN) extracts the feature information of the media content data and constructs the long-range feature vector f_t^c, as shown in equation (6):

$$f_t^l, f_t^c = E_\alpha(x_t) \tag{6}$$

where E_α denotes the encoder, and f_t^l and f_t^c denote the short-range feature vector and long-range feature vector, respectively.

The f_t^l and f_t^c are input to the feature processing module of the encoder, and the f_t^l is set to be the media content mean feature μ output by the encoder after fully-connected compression coding; the f_t^c is set to be the media content variance feature ε output by the encoder after fully-connected compression coding. To facilitate the computation of the encoder gradient estimator, the noise σ is introduced, σ obeys a normal distribution, and the fusion hidden variable z is output using the reparameterization trick. The fused hidden variables are computed as shown below:

$$\mu, \varepsilon = E_\alpha(f_t^l, f_t^c) \tag{7}$$

$$Z = \mu + \varepsilon \times \sigma \quad \sigma \sim N(0,1) \tag{8}$$

where μ and ε are the mean and variance of the media content output by the encoder after processing the input features, z is the fused hidden variable after reparameterization, and $z \sim (\mu, \varepsilon)$. The samples generated by the generator are sampled and the KL scatter is used to compute the encoder loss function as shown in equation (9):

$$l_{E_\alpha} = D_{KL}\left(q(z|x) \| p(z|x)\right) = -\frac{1}{2}\sum_{j=1}^{J}\left[1 + \varepsilon_j^{(i)} - (\mu_j^{(i)})^2 - e^{\varepsilon_j^{(i)}}\right] \tag{9}$$

In the generator, the real media content data sequences are preprocessed to get the word vectors of the input data, and the generator model uses LSTM network to generate the media content sequences. In the generator process, the fused hidden variable z_t output from the encoder and the current input media content sequence x_t are used as common inputs to the generator model, which is combined with the hidden layer state matrix h_{t-1}^θ output from the LSTM network at the moment of $t-1$, to output the hidden layer state matrix h_t^θ at the moment of t, and to sample the t moment of the generated sample x_{t+1}, computed as shown below:

$$h_t^\theta = G(h_{t-1}^\theta, x_t; z_t) \tag{10}$$

$$G_\theta(x_{t+1} | s_t) = soft\max(h_t^\theta) \tag{11}$$

The negative log likelihood function (NLL) is used as the loss function of the generator as shown in equation (12):

$$l_{G_\theta} = G_\theta\left(x_t | x_{t-1}, z_t\right) = \frac{1}{N}\sum_{n=1}^{1}\ln\left(1 + e^{-y_n W^T x_n}\right) \tag{12}$$

II. B. 3) Identifiers

The discriminator D_ϕ in the generative adversarial network is a classification model. The model is adversarial trained with real media content data and generated media content data, which makes the discriminator unable to discriminate the class of the data input to the discriminator, and thus improves the generative power of the generative model. Its optimization objective is to minimize the negative log-likelihood value as shown in equation (13):

$$l_{D_\phi} = \min_{\phi} - E_{Y-data}[\log D_\phi(X)] - E_{Y-G_\theta}[\log(1-G_\theta(Z))] \tag{13}$$

As the quality of the generated media content is low in the process of media content generation by the variational self-coder, in order to improve the reasonableness of the generated media content, the model in this paper combines the reward function and the generator reconstruction loss in reinforcement learning to realize the reinforcement of the model for adversarial training. The sequence of generated media content, $x_{1:t-1}$, is defined as the state of the generator model at the moment $t-1$, and the generator generates the next word, x_t, and then gets a reward value, which indicates the expected reward of the generator for generating x_t at the moment t in the state of the generator at the moment $t-1$. The reconstruction loss of the generator for combining reward functions, i.e., the reconstruction term l_{rec} is defined as shown in equation (14):

$$l_{rec} = E_{x_{1:t-1} \sim G_\theta}\left[\sum_{x_t \in X} G_\theta(x_t \mid x_{t-1}, z_t) \cdot Q_{D_\phi}^{G_\theta}(a = x_t, s = x_{1:t-1})\right] \tag{14}$$

where, z_t tables the fusion hidden variables of the encoder output at moment t, and the reward value $Q_{D_\phi}^{G_\theta}(a = x_t, s = x_{1:t-1})$ is defined as the expected reward in the current state $s_t = (x_1, x_2, \cdots, x_t)$. $D_\phi(s_t)$ is defined as the reward value that the discriminator feeds back to the generator. In order to evaluate the generator's media content generation effect in the intermediate state $(t < T)$, the MC method is used to complement the media content sequence as shown in equation (15):

$$Q_{D_\phi}^{G_\theta}(a = x_t, s = x_{1:t-1}) = \begin{cases} \frac{1}{n}\sum_{n=1}^{n} D_\phi(x_{t:T}), x_{t:T} \in MC^{G_\theta}(x_{t:T}; n) & for(t < T) \\ D_\phi(x_{t:T}) & for(t = T) \end{cases} \tag{15}$$

If the generator has generated a complete media content sequence, the media content sequence is used as an input to the discriminator, and the output of the discriminator is the reward value of the current time step; if the generator has not generated a complete media content sequence, Monte Carlo search is used to complement the media content sequence. In the complementary process, the generator samples multiple complete media content sequences. These media content sequences are used as inputs to the discriminator, and the average of the output result matrix of the discriminator is the reward value in the current state.

In summary, the VAE-IAT model draws on the training process of SeqGAN in its implementation, and divides the whole training process into two phases, with the first phase being the pre-training phase, which mainly trains the encoder and generator. In the pre-training stage, the loss function of the model is shown in equation (16):

$$l_{VSGAN}^{1} = l_{G_\theta} + l_{E_\alpha} \tag{16}$$

where l_{G_θ} and l_{E_α} are the losses of the generator and encoder, respectively.

The second stage is the intensive adversarial training stage. In the stage of intensive adversarial training mainly need to train the parameters of the generator and the discriminator, the loss function of the model during the training process is shown in equation (17):

$$l_{VSGAN}^{2} = l_{rec} + l_{D_\phi} \tag{17}$$

where l_{rec} is the reconstruction loss combined with the reward function and l_{D_ϕ} is the loss function of the discriminator.

The l_{rec} of the generator is updated by means of the strategy gradient. The strategy gradient update uses the most rapid descent method, and the generator parameters are updated as shown in equation (18):

$$\theta_{new} = \theta_{old} + \lambda \nabla_\theta G(\theta) \tag{18}$$

where θ is the generator parameter and λ is the learning efficiency.

II. C.Paths to optimize communication effectiveness

In the rapid development of digital media, there are problems such as ambiguous communication positioning, uneven content quality, and stereotyped discourse mode, which not only affect the communication effect of digital media methods, but also restrict its sustainable development. Aiming at such problems, three communication effect optimization paths are proposed based on the perspective of deep generation model.

II. C. 1) Improving communication accuracy

When workers produce digital media content, the first step is to verticalize and personalize it. This means that digital media need to pinpoint the audience groups and dig deeper into the material in the field through the deep generation model to meet the specific needs of the audience. At the same time, digital media workers also need to understand and analyze the user's needs, think about the topics that the audience cares about and is most interested in from their point of view, and accurately push the required content for the users with the help of the deep generation model.

II. C. 2) Addressing content production

In the process of deepening the content, the integration of innovative elements is also essential. Novel narrative methods, unique filming techniques, personalized editing styles, and new technological means such as virtual reality can bring a brand new audiovisual experience to short video news products. The use of in-depth generative modeling to create an immersive experience environment, these high-tech means can greatly broaden the collection channels and scope, so that the media's perspective is more diversified and the information is more comprehensive. On this basis, workers can process the collected media content with the help of in-depth generation model, presenting the news facts in a more vivid and graphic way, and enhancing the audience's intuitive experience of digital media content.

II. C. 3) Enhanced interaction with the audience

In terms of strengthening interaction with the audience, using the deep generation model to build a digital media platform, workers should make full use of the functions of liking, commenting and forwarding on the digital media platform, respond to the audience's questions in a timely and positive manner, and take the initiative to provide positive guidance for those comments that may have a negative impact. For those contents that may trigger the audience to understand difficulties or questions, journalists should do a good job of information disclosure, and strive to create high-quality digital media works that meet the needs of the audience.

III. Exploring Digital Media Content Generation and Distribution

III. A. Validation Analysis of Digital Media Content Generation Models

III. A. 1) Setting up the experimental environment

The experiments in this vignette are run in a GPU server environment, with a GPU equipped with NVIDA Tesla P40, an operating system selected from Ubuntu20, and a coding implementation using both the Python 3.8 coding language as well as the open-source framework for deep learning, Pytorch, and a graphics card driver version of NVIDA450.66.

III. A. 2) Data sets

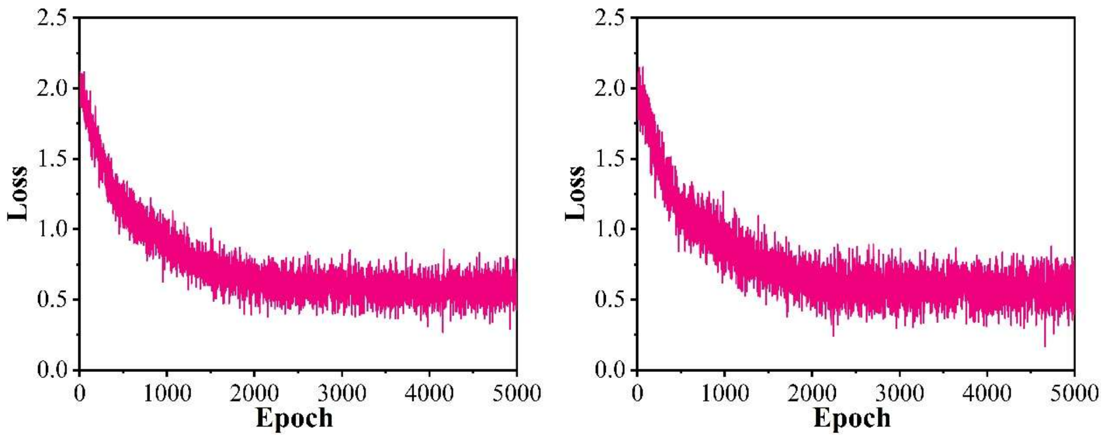
The experimental datasets are the MINST dataset, the SVHN dataset, and the CelebA dataset. The experimental dataset is the MINST dataset, the MINST dataset is a handwritten digit dataset, which is organized by the National Institute of Standards and Technology (NIST), a total of 250 handwritten digit images from 250 different people were counted, of which 50% were high school students and 50% were from the staff of the Census Bureau, of which there are 60,000 handwritten digit images with a size of 28×28 pixels in the training set and 10,000 images in the test set, each image is labeled with 0-9 pixels. 28 pixels in black and white, and the test set has 10,000 images, each labeled with one of the ten numbers 0-9.The SVHN dataset is derived from Google Street View door numbers, and each image in the dataset contains border information as well as a random number of 0-9 digits, i.e., the door number digits, for a total of 70,000 RGB images with a size of 32×32 pixels.The Celeb A dataset is the face attribute dataset, which includes 202599 images of 10177 celebrity identities, while each image is preprocessed with good feature markers, face bbox labeled boxes, 5 facial feature point coordinates, and 40 attribute markers, and each image is an RGB image, while each image is set to be 64×64 pixels in size.

III. A. 3) Hyperparameterization

The model uses Adam optimizer, the loss function is updated by the back propagation algorithm, the initial learning rate is set to 0.00001, the size of each batch is set to 128, and 100 batches are trained in each round, and the hyper-parameters of the model, and the encoder parameters are also determined.

III. A. 4) Loss function training

The generation loss g_loss and discrimination loss d_loss of VAE-IAT are monitored simultaneously in the experiment, and their loss curves are shown in Fig. 2, where (a)~(b) are g_loss and d_loss, respectively. The generation loss decreases gradually with the increase of the number of iterations, and it tends to converge at about 4.5k iterations, and the discrimination loss fluctuates relatively more, but it shows a decreasing trend in the whole. The gradient drop of VAE-IAT is stable, and there is no gradient disappearance during the experiment.



(a)g_loss (b)d_loss

Figure 2: Loss function training

III. A. 5) Evaluation indicators

In order to quantitatively assess the classification error rate of the proposed model of this paper, this paper adopts the error rate as the evaluation criterion, the lower the error rate, the better the effect, the specific formula is shown in equation (19):

$$errorrate = \frac{FP}{TP + FP} \tag{19}$$

where, TP denotes the number of positive examples judged as positive class by the model, and FP denotes the number of negative example samples judged as positive class by the model.

In order to better assess the diversity and quality of images in generated digital media content, this paper uses FID, which is a metric for evaluating the quality of generated images and better calculates the similarity between generated and real images. Using the output from the activation function to generalize each image, the score is the FID score and the lower the score the better, the specific formula is shown in equation (20):

$$FID(x, x_G) = \left\| \mu_x - \mu_{x_G} \right\|_2^2 + Tr\left(\sum\nolimits_x + \sum\nolimits_{x_G} - 2\left(\sum\nolimits_x \Sigma_{x_G} \right)^{\frac{1}{2}} \right) \tag{20}$$

III. A. 6) Control models

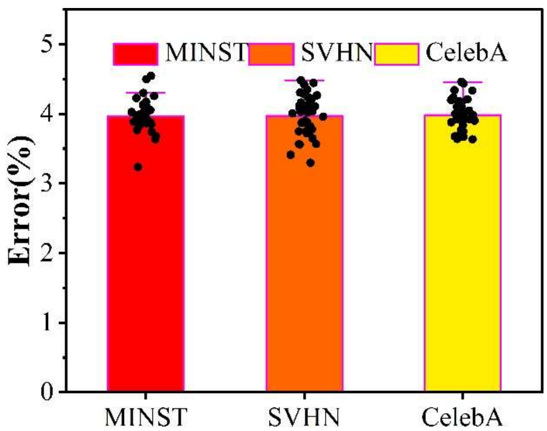
(1) TripleGAN: an improved semi-supervised deep generative model consisting of three components.

(2) SSVAE: Processing the semi-supervised learning problem by building two models, M1 model provides embedding or representation features of the data, and M2 model generates the data through hidden variables and labeling information, M1 is essentially a variational self-encoder model, which is used to improve the capability of M2 model.

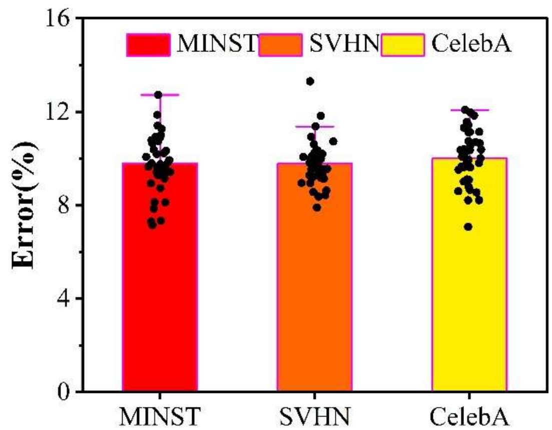
(3) ADGM: Based on the improvement of the variational self-encoder, constraints are added to the model encoder to establish a more expressive distribution, which enables the model to generate the corresponding samples according to the given labeling information and further improves the predictive performance of the model.

III. A. 7) Data analysis

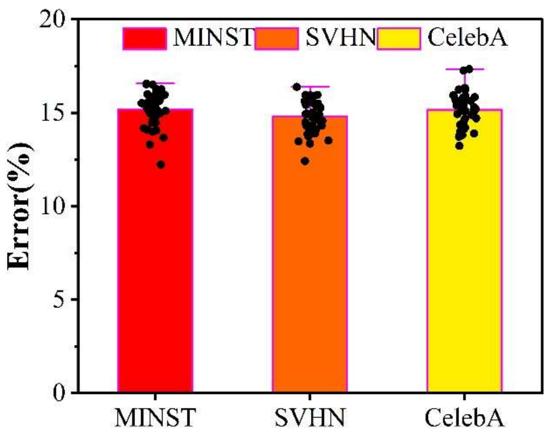
In order to further analyze whether the model in this paper meets the expected requirements, this paper compares the classification error rate of the VAE-IAT model with the TripleGAN model, the SSVAE model, and the ADGM model on the MINST dataset, the SVHN dataset, and the CelebA dataset, respectively The lower the classification error rate is, the better the classification ability of the model is, and the classification error rate of the VAE-IAT model is the same as that of the TripleGAN model, the SSVAE model, and the ADGM model, respectively. The classification error rate comparison is shown in Figure 3, where (a) to (d) are the VAE-IAT model with TripleGAN model, SSVAE model, and ADGM model, respectively. The results show that, no matter in MINST dataset or SVHN dataset or CelebA dataset, the VAE-IAT model demonstrates an ultra-low classification error rate, which is much better than the other three types of models (TripleGAN model, SSVAE model, ADGM model), which is attributed to the fact that the VAE-IAT model, by means of a posteriori distributions, is able to learn better high-dimensional spatial distribution, and the posterior distribution provides better implicit information about class labels for the generative model, which makes the model's classification ability secure.



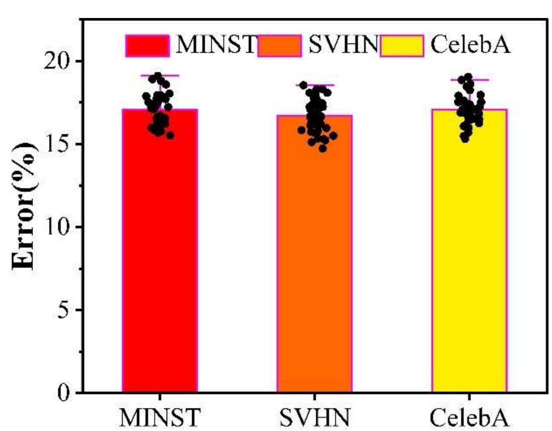
(a)VAE-IAT model



(b)TripleGAN model



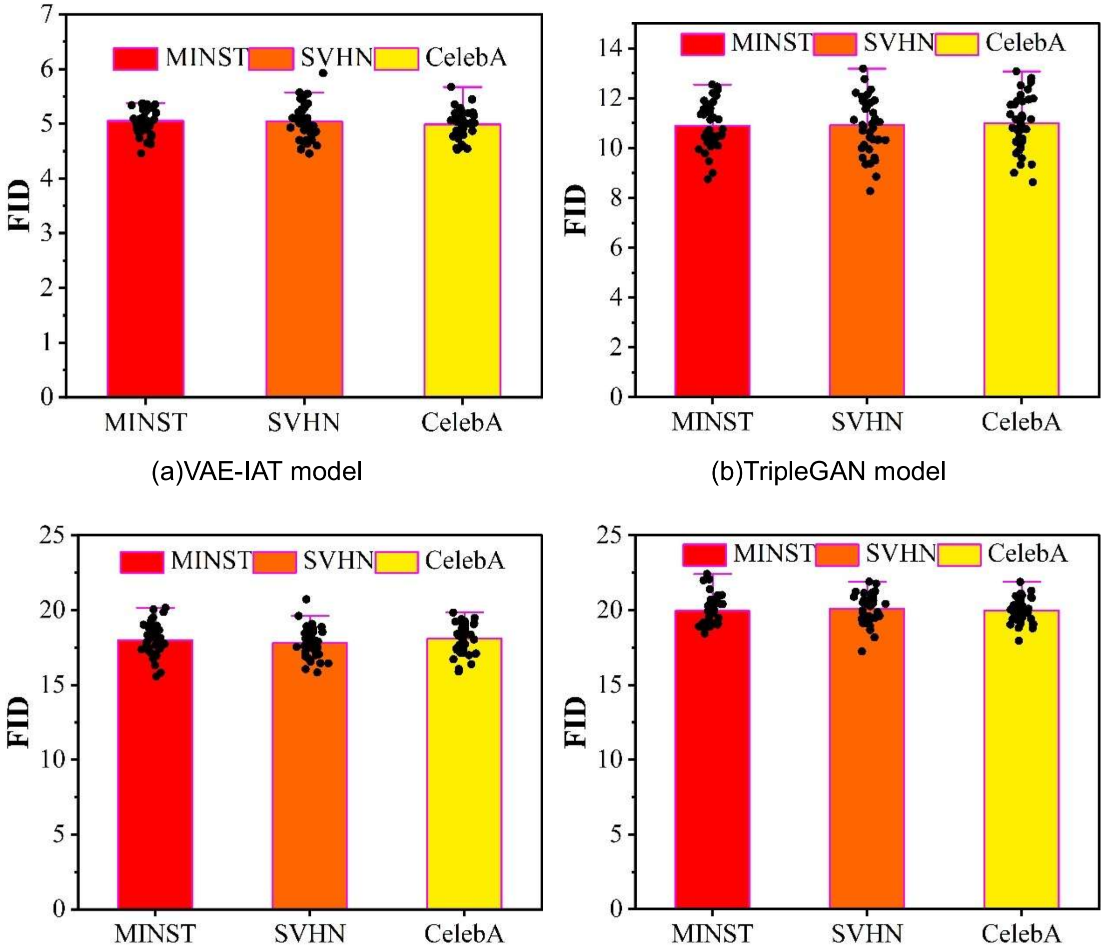
(c)SSVAE model



(d)ADGM model

Figure 3: Based on the comparison of classification error rates on different datasets

In order to further quantitatively assess the sample quality, the FID score is used to evaluate the model generation ability, FID is the standard score for generating the model, the lower the FID score, the better the quality of the samples generated by the model, based on the comparison of FID on different datasets as shown in Fig. 4. The VAE-IAT model constructed in this paper shows good sample generation ability compared to other models (TripleGAN model, SSVAE model, ADGM model) on the three datasets (the FID value is maintained below 6), especially on the CelebA dataset the results have been significantly improved, which indicates that the VAE-IAT model is able to better learn the data and its categories of joint probability distributions, which can effectively improve the quality of digital media content and better serve digital media workers.



(a)VAE-IAT model (b)TripleGAN model

(c)SSVAE model (d)ADGM model

Figure 4: Based on the FID comparison on different datasets

III. B. Communication Optimization Path Validation Analysis

III. B. 1) Research program

(1) Research Purpose

This paper hopes to confirm the effectiveness of the digital media communication optimization path that integrates the deep generation model by means of scale testing and statistical analysis.

(2) Research Objectives

The deep generative model is used to construct digital media communication paths to promote the effect and quality of digital media communication, so that it can better serve the user community. In this regard, 40 digital media users are selected as the object of this study, and according to the random sampling method, the research object is divided into two groups, respectively, the experimental group and the control group, the experimental group adopts the digital media communication path based on the depth generation model, while the control group adopts the traditional digital media communication path.

(3) Research Tools

(a) Scale

With reference to the relevant References, a digital media communication effect test scale was developed, which has four dimensions, media comments, media retweets, media interactions, media likes, and each dimension has 10 questions, and the scale adopts the Likert scale scoring method, with 5 points for strongly agree, 4 points for agree, 3 points for reluctant to agree, 2 points for disagree, and 1 point for strongly disagree. After completing the development of the scale, it is also necessary to carry out reliability and validity tests, and it was concluded that the scale has a good reliability and validity performance, which fully meets the standard requirements of this study.

(b) Statistical analysis method

Through the scale test, the initial data required for this study were obtained, and the SPSS statistical analysis software was used to carry out independent sample t-test on the initial data to confirm the actual communication effect of the digital media communication optimization path in this paper.

(4) Implementation process

The implementation steps of this experiment mainly include the following four steps:

In the first step, the research subjects are asked to fill in the digital media communication effect test scale to understand whether the research subjects selected in this paper are homogeneous or not, in addition to which no information about the experiment is disclosed to the subjects before the experiment is carried out.

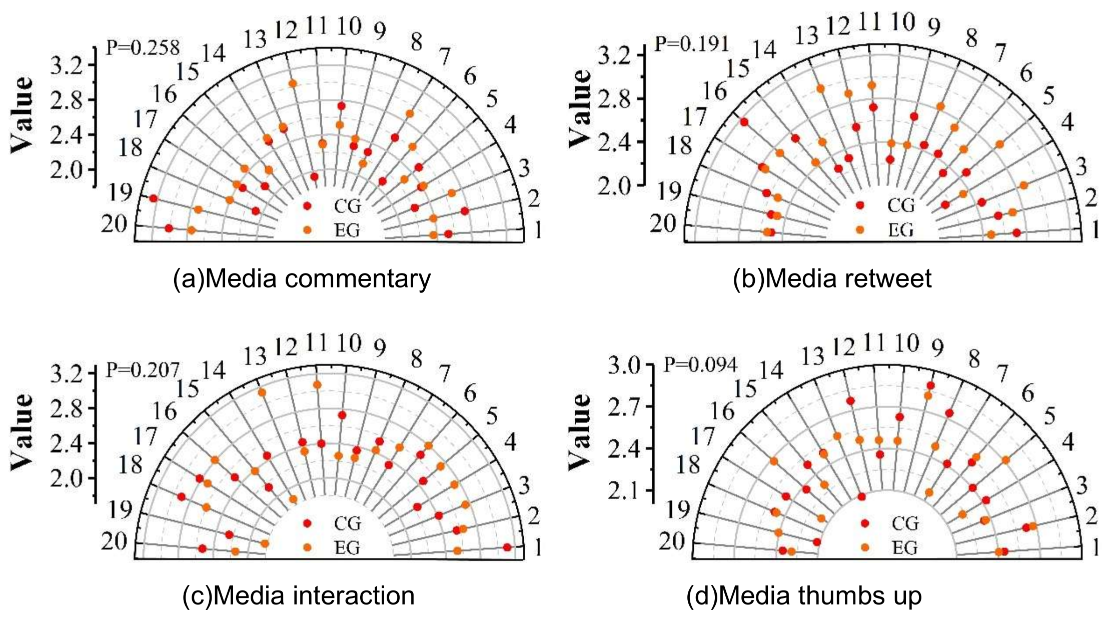
In the second step, before the beginning of the experiment, the participants were given a brief description of the experiment and how to use the experimental materials, and after the experiment was carried out without giving more hints and restrictions, the two groups of subjects used different methods to carry out digital media communication, and the experimental time was stipulated.

In the third step, all subjects were asked to complete the scale test on the spot when the experiment was due.

In the fourth step, after a two-week interval, the experimental subjects were asked to take the scale test again, focusing on the differences in media comments, media retweets, media interactions, and media likes among the different groups.

III. B. 2) Homogeneity test

On the basis of the scale test, with the help of the independent sample t-test method in SPSS statistical analysis software, the homogeneity test was conducted on the research subjects of the experimental group and the control group, and the results of the homogeneity test are shown in Fig. 5, in which EG and CG represent the experimental group and the control group, respectively, in which (a) to (d) are media comments, media retweets, media interactions, and media likes, respectively. The P-value in the figure shows that before the experimental intervention, the experimental group and the control group do not have significant differences in media comments, media retweets, media interactions, and media likes, which satisfies P<0.05, indicating that the selected research subjects satisfy the requirement of homogeneity to carry out further research work.



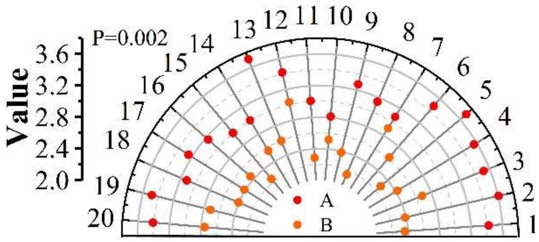
(a)Media commentary (b)Media retweet

(c)Media interaction (d)Media thumbs up

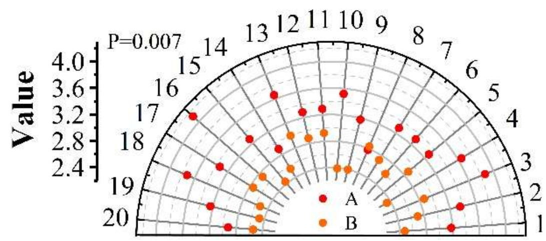
Figure 5: Homogeneity test results

III. B. 3) Comparative analysis within groups

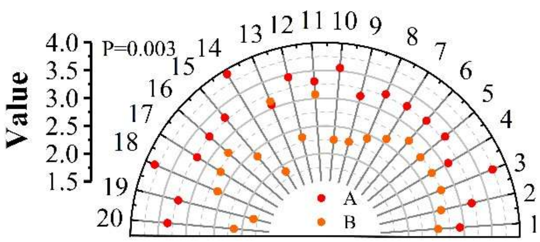
Using the same method as above, the difference analysis was conducted before and after the intervention of the experimental group and before and after the intervention of the control group, respectively, and the comparative analysis of the experimental group before and after the intervention is shown in Figure 6, and the comparative analysis of the control group before and after the intervention is shown in Figure 7. The synthesis of Fig. 6 and Fig. 7 shows that before and after the intervention of the experimental group, there are significant differences in media comments, media retweets, media interactions, and media likes, and their p-values are 0.002, 0.007, 0.003, and 0.005, respectively. On the contrary the control group before and after the intervention did not show a significant difference in all four dimensions of communication effect, with p-values of 0.227, 0.108, 0.061, and 0.079 for each dimension. In summary, compared with the traditional digital media communication path, the deep generative model-driven digital media communication path performs better, which verifies the effectiveness of the digital media communication path in this paper, and also provides guidance for the current development of digital media communication in society.



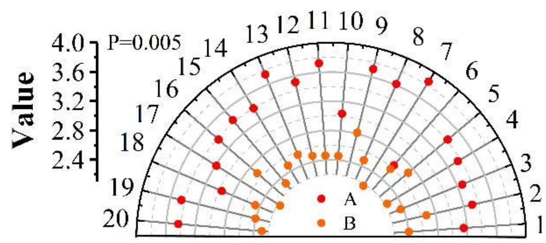
(a)Media commentary



(b)Media retweet

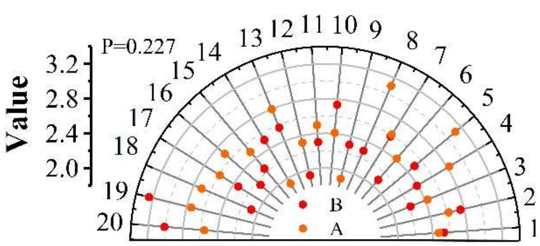


(c)Media interaction

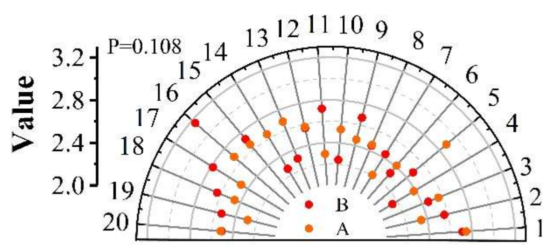


(d)Media thumbs up

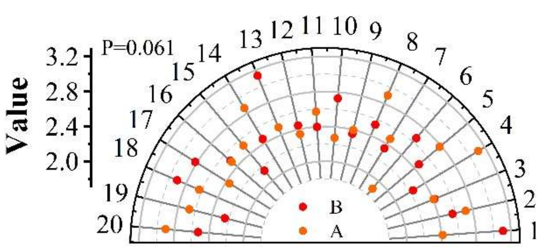
Figure 6: Comparative analysis of the EG before and after the intervention



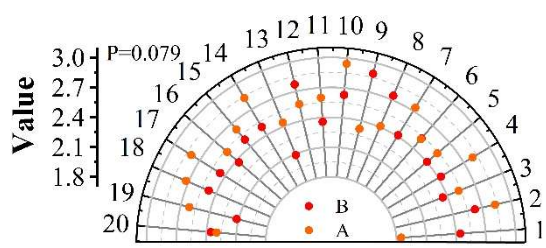
(a)Media commentary



(b)Media retweet



(c)Media interaction



(d)Media thumbs up

Figure 7: Comparative analysis of the CG before and after the intervention

IV. Conclusion

In this study, we have proposed a variable-aggregate digital media content generation model (VAE-IAT) based on adversarial training, and constructed a corresponding optimization path for propagation effect, which was verified and analyzed through a series of experiments. The main research conclusions are as follows:

The VAE-IAT model designed in this study significantly improves the quality of digital media content generation by extracting short-range features and long-range features of media content through the encoder and combining the adversarial training of the generator and the discriminator. The experimental results show that the VAE-IAT model exhibits excellent classification performance and generation capability on all three datasets, MNIST, SVHN and CelebA, with the FID scores remaining below 6. In particular, it significantly outperforms the control model on the CelebA dataset. This demonstrates that the VAE-IAT model is able to better learn the joint probability distribution of the data and its categories, effectively improving the quality of digital media content.

The three communication effect optimization paths constructed based on the VAE-IAT model - improving communication accuracy, deepening content production and strengthening interaction with the audience - have shown significant effect improvements in experimental verification. Through scale tests and statistical analysis, it was found that the experimental group had significant improvements in the four dimensions of media comments, media reposts, media interactions, and media likes after the intervention. The P values were 0.002, 0.007, 0.003, and 0.005 respectively, which were much lower than the significance level of 0.05. The P values of the control group in these four dimensions were 0.227, 0.108, 0.061 and 0.079 respectively, none of which reached the level of significant difference.

In addition, during the model training process, both the generation loss and discrimination loss of the VAE-IAT model converge after about 4.5k iterations, and the gradient decline is stable without gradient disappearance, indicating that the model has good training stability and convergence performance.

In summary, the adversarial training-based variational digital media content generation model and its propagation effect optimization path proposed in this study not only theoretically enriches the application of deep generation models in the field of digital media, but also provides feasible technical support and strategic guidance for digital media content generation and propagation practice, which is of great significance for improving the quality and propagation effect of digital media content.

References

[1] Ohme, J. (2019). Updating citizenship? The effects of digital media use on citizenship understanding and political participation. Information, Communication & Society, 22(13), 1903-1928.

[2] Ferdinandus, D. D., & Alvin, S. (2023). Snackable content creation in the digital age: a case study of social media content production at Net TV. International Journal of Economic, Business, Accounting, Agriculture Management and Sharia Administration (IJEBAS), 3(3), 669-680.

[3] Oyighan, D., & Okwu, E. (2024). Social media for information dissemination in the digital era. RAY: International Journal of Multidisciplinary Studies, 10(1), 1-21.

[4] Hallur, G. G., Aslekar, A., & Prabhu, S. G. (2023). Digital solution for entertainment: An overview of over the top (OTT) and digital media. Digital entertainment as next evolution in service sector: Emerging digital solutions in reshaping different industries, 35-53.

[5] Coleman, R. (2020). Making, managing and experiencing 'the now': Digital media and the compression and pacing of 'real-time'. New media & society, 22(9), 1680-1698.

[6] Kapoor, S., & Singh, J. (2024). Leveraging digital media for effective healthcare information dissemination: A study of reach, accessibility, and impact. LWATI: A Journal of Contemporary Research, 21(3), 38-57.

[7] Reyna, J., Hanham, J., & Meier, P. (2018). The Internet explosion, digital media principles and implications to communicate effectively in the digital space. E-learning and Digital Media, 15(1), 36-52.

[8] Wei, Z., Xiao, M., & Rong, R. (2021). Network size and content generation on social media platforms. Production and Operations Management, 30(5), 1406-1426.

[9] Wu, Y., & Liu, D. (2024). Research on the Application of Image Processing and Image Recognition Algorithms in Digital Media in Content Editing and Production. J. Electrical Systems, 2024, 1126-1138.

[10] Darda, K. M., Estrada Gonzalez, V., Christensen, A. P., Bobrow, I., Krimm, A., Nasim, Z., ... & Chatterjee, A. (2025). A comparison of art engagement in museums and through digital media. Scientific Reports, 15(1), 8972.

[11] Chen, Z., Zhu, Y., Xu, P., & Liu, R. (2019, November). A brief analysis of the application research of digital media technology--taking digital film and television as an example. In Journal of Physics: Conference Series (Vol. 1345, No. 5, p. 052058). IOP Publishing.

[12] Mao, Y. (2017, November). On the valUes of Digital Media Design in Dynamic Advertisement Designing. In 2017 International Conference on Information, Communication and Engineering (ICICE) (pp. 219-220). IEEE.

[13] Zhu, X., Lang, M., & Dietl, H. M. (2023). Content quality assurance on media platforms with user-generated content. Journal of Theoretical and Applied Electronic Commerce Research, 18(3), 1660-1686.

[14] Roy, P. (2024). An Investigation into How Global Convergence of Media is Influencing Cultural Exchange, Diversity and Homogenization of Media Content in Different Regions. Global Media and Cultural Exchange, 1.

[15] Buchanan, T. (2020). Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. Plos one, 15(10), e0239666.

[16] Ramagundam, S., & Karne, N. (2024, September). The New Frontier in Media: AI-Driven Content Creation for Ad-Supported TV using Generative Adversarial Network. In 2024 7th International Conference of Computer and Informatics Engineering (IC2IE) (pp. 1-6). IEEE.

[17] Guo, Y., Hu, S., Li, Y., Di, C., & Liu, J. (2025). Dmvae: a dual-stream multi-modal variational autoencoder for multi-task fake news detection. Pattern Analysis and Applications, 28(2), 65.

[18] Razieh Rastgoo,Nima Amjady,Syed Islam,Innocent Kamwa & S.M. Muyeen. (2025).Extreme outage prediction in power systems using a new deep generative Informer model.International Journal of Electrical Power and Energy Systems,167,110627-110627.

[19] Mendes Tomás,Cardoso Pedro J. S.,Monteiro Jânio & Raposo João. (2022). Anomaly Detection of Consumption in Hotel Units: A Case Study Comparing Isolation Forest and Variational Autoencoder Algorithms. Applied Sciences,13(1),314-314.

[20] Zhang Mi & Hu Xiangyun. (2017). Translation-aware semantic segmentation via conditional least-square generative adversarial networks. Journal of Applied Remote Sensing,11(4),1-1.