Diffusion Model-Assisted Interactive Scene Generation and Student Participation Prediction for Ideological and Political Education in Textile Majors

Xuan Zhao
Department of Student Affairs, Harbin University of Science and Technology, Harbin 150080, Heilongjiang, China
Corresponding author: Xuan Zhao (e-mail: za14273583@sina.com)

ABSTRACT To address the insufficient integration between textile engineering knowledge and ideological–political education, as well as the lack of real-time assessment of student engagement, this study proposes a diffusion model-assisted framework for interactive scene generation and multimodal participation prediction. A dual-encoder diffusion architecture driven by a Textile Knowledge Graph (TKG) is developed to establish semantic associations between textile process parameters and ethical objectives, enabling the automatic generation of process-aware educational scenarios. Eye-tracking and electromyography signals are jointly modeled using a Long Short-Term Memory network to predict participation status, while a dynamic feedback mechanism adaptively adjusts scene complexity according to learners' cognitive responses. Considering the growing adoption of intelligent sensing and distributed information acquisition technologies in digital education environments, the proposed framework provides a scalable solution for multimodal data fusion and adaptive interaction, offering potential methodological references for future wireless sensing and smart educational infrastructures. Experimental results demonstrate that the proposed method achieves a process parameter trigger rate of 82.1%, an ethical scene coverage of 93.2%, classification accuracies exceeding 85% for highly engaged students, and interaction latency below 50 ms under low-complexity conditions. The framework effectively enhances scene authenticity, participation prediction accuracy, and adaptive instructional capability, providing a reliable technical foundation for intelligent textile education and interdisciplinary human–machine interaction systems.

INDEX TERMS diffusion model, textile knowledge graph, multimodal participation prediction, intelligent sensing, adaptive interactive learning

I. INTRODUCTION

THE ideological and political education of textile majors aims to convey industry values through professional practice, guiding students to balance economic and social responsibilities in technical decision-making, and cultivating awareness of green manufacturing, craftsmanship, and cultural inheritance. However, the current teaching model faces dual challenges closely related to the process complexity and value-guiding characteristics of textile majors: ethical decision-making scenes lack the support of textile process logic, which directly leads to students' difficulty in perceiving the tangible consequences of their technical choices, resulting in low cognitive engagement and passive participation [1,2]. For instance, when students adjust the "yarn density" parameter in a virtual simulation, a scene lacking process logic might only display a generic message such as "This is not sustainable." This fails to demonstrate the concrete, cascading effects—such as increased raw material consumption, higher wastewater biochemical oxygen demand (BOD) levels, or the economic impact on downstream suppliers. Without this causal chain being visually and logically presented, students cannot grasp the real-world significance of their decisions, leading to a disconnection between technical action and ethical consequence. Consequently, the reliance on static evaluation tools, such as after-class questionnaires, which capture only a delayed and superficial snapshot of understanding, fails to address the root cause: the lack of dynamic, process-embedded ethical scenarios that can trigger and sustain active cognitive processing. The root of the problem lies in the fact that existing technologies fail to deeply integrate textile engineering knowledge with ideological and political goals and can neither dynamically generate professional ethical interaction scenes nor quantify students' real-time responses to issues such as sustainable production and intangible cultural heritage protection. This dilemma restricts the cultivation of students' professional ethical judgment and affects their adaptation to the needs of industry green transformation and global responsibility management. In this study, an "ethical decision-making scene" refers to a dynamic, interactive simulation where a student's technical choice triggers an immediate, visual consequence. This scene is not a static image but a parameter-driven event that forces the student to confront the social and environmental implications of their engineering decisions in real time. In recent years, the academic community has focused on research regarding technological empowerment of ideological and political education interaction scenes. Shuhui Wang discussed the current situation and optimization strategies of online ideological and political education for students in vocational colleges, proposing dynamic content reconstruction and multimodal augmented reality (AR) teaching; however, the research lacked course implementation tracking and empirical verification [3]. Siyi Luo's research further expanded the boundaries of technology application and developed a progressive teaching model based on digital twin technology, but the technical implementation path was limited to specific scene adaptation in the video Internet of Things environment [4]. Yixian Chen approached the issue from the perspective of communication, constructing a scene-based ideological and political education communication model based on the "5W" model, and proposed the use of big data and algorithmic technology to achieve precise content push and immersive scene interaction, but did not solve the coupling problem between professional characteristics and ideological and political goals [5]. Existing research has common limitations: the technical tools are not sufficiently integrated with professional courses such as textiles, and process parameters are not linked to knowledge of ideological and political value objectives; multimodal data collection is singular and lacks joint modeling of physiological signals and operational behaviors; the absence of a dynamic feedback mechanism makes it difficult to adaptively match scene complexity and participation status. The root cause is that existing technology has not established a closed-loop mapping between professional characteristics and ideological and political education. Diffusion models have shown great potential in dynamic scene generation for digital education, especially in producing controllable high-resolution content. Ruihan Yang et al.'s end-to-end video generation method based on diffusion models used an autoregressive framework to gradually generate high-quality video frames, improving the perceptual quality and prediction accuracy of dynamic scenes [6]. Other studies have explored multi-view controllable generation and text-driven generation, improving the fidelity and structural controllability of generated content. Heng Yu et al. broke through the bottleneck of dynamic fidelity through four-dimensional scene generation and achieved multi-view controllable high-quality generation [7]. Bing Li et al. further expanded the possibility of text-driven scene generation and generated multi-view video content with dynamic three-dimensional structure [8]. These advances have laid the foundation for educational interactive scenes, but their application in ideological and political education in textile majors still faces challenges: the models lack textile field knowledge embedding, resulting in parameters such as yarn density and dyeing fastness failing to meet physical consistency requirements; the complexity of denoising calculations causes excessive delay in real-time interaction; the existing mechanism fails to visually associate process parameters with ideological and political goals such as "green manufacturing" and "intangible cultural heritage protection," and cannot achieve the mapping of technical adjustment and social responsibility. Existing research focuses on general scene generation and has not yet solved the problem of deep integration of textile professional characteristics and ideological and political values. The framework of this paper provides a feasible solution to this dilemma. To address the need for interactive scene generation in textile education, this paper proposes a dynamic scene generation and multimodal engagement prediction framework based on a diffusion model. The scene generation layer employs a dual-encoder diffusion model driven by a Textile Knowledge Graph (TKG), topologically linking process parameters such as yarn density and colorfastness to ideological and political themes like "sustainable production" and "craftsman spirit." An ethical conflict trigger mechanism is established by incorporating a process constraint module—based on fiber mechanics equations—into the UNet architecture and injecting an ethical decision tree into the cross-attention layer. This allows abnormal parameters such as spindle overspeed or wastewater discharge to be automatically associated with social responsibility indicators, such as carbon emissions and intangible cultural heritage protection. The engagement prediction layer develops a dual-channel eye movement and electromyography (EMG) acquisition system and uses a Long Short-Term Memory (LSTM) network to jointly model gaze heatmaps and EMG signals, establishing a nonlinear mapping between physiological signals and ethical decision-making. A dynamic feedback mechanism adjusts scene complexity in real time based on EMG fatigue thresholds and engagement predictions. This framework bridges the gap between technical logic and ideological education, forming a closed loop from parameter adjustment to value cognition. The proposed system serves as a pedagogical tool to connect technical practice with value awareness in textile education. In this study, "ideological and political education" in textile majors refers to fostering professional ethics and social responsibility through engineering practice. The core idea is to create a direct causal link between technical decisions and social values; for example, increasing yarn density reduces material waste, supporting "sustainable production," and improving dye fastness reflects the "craftsman spirit" through durable quality.

This visible connection between technical actions and their social impact forms the logical basis for effectively integrating values education into professional training.

II. METHODS

A. OVERALL ARCHITECTURE DESIGN OF THE IDEOLOGICAL AND POLITICAL INTERACTION SYSTEM FOR TEXTILE MAJORS

The implementation process of this method revolves around the core needs of "textile process logic support" and "dynamic generation of ethical conflicts." It proposes a dual-module architecture that integrates a TKG-Diffusion model for scene generation and an LSTM-based model for participation prediction, builds a closed-loop system driven by multimodal data, and forms a closed-loop optimization path from parameter adjustment to scene response and then to behavioral feedback. The overall implementation process is illustrated in Figure 1.

The architecture in Figure 1 takes textile process parameters, ethical standards, and multimodal sensor data as input, builds a dual-drive closed-loop system through a process knowledge graph and real-time acquisition module, and realizes the process of "parameter input–scene generation–behavior response–dynamic optimization." The core processing units work together: the diffusion model injects process constraints (noise scheduling function plus fiber mechanics loss function) into the UNet backbone, activates the ethical decision tree through cross-attention weights, and maps anomalies such as spindle speed (SS) overshoot to pixel space; the participation prediction model models the eye movement–electromyography feature time series based on LSTM and outputs the participation probability and decision path. The feedback system adjusts the parameter attenuation factor according to the electromyography fatigue threshold, drives tactile feedback, and generates ethical conflict scenes. The architecture ensures the physical authenticity of the generated scene and the adaptability of ideological and political education in textile majors.

B. DIFFUSION MODEL DESIGN FOR EMBEDDING TEXTILE KNOWLEDGE AND IDEOLOGICAL AND POLITICAL ELEMENTS

The textile knowledge graph's process parameters are aligned with the sustainability principles of GB/T 35611-2017, while their specific operational thresholds are derived from empirical data and industry technical guidelines. Ideological and political education objectives are mapped from the ISO 20400:2017 sustainable procurement guidelines and national curriculum guidelines. Textile engineers and educators verified the correlation between these objectives through expert annotation to ensure the validity and objectivity of the mapping. This study constructs a dual-encoder diffusion model by applying a TKG and process constraint equations [9]. The model inserts a process consistency module into the time-step embedding layer of the UNet backbone network to transform textile process parameters such as yarn breaking strength and dyeing fastness into physical constraints of the diffusion process [10]. In the forward diffusion process, the noise scheduling function α_t is expanded to:

$$\alpha_t' = \alpha_t \cdot \exp\left(-\lambda \cdot \|P_{\text{current}} - P_{\text{standard}}\|_2\right) \tag{1}$$

In Equation 1, λ is the process deviation penalty coefficient, and P_{current} and P_{standard} represent the process parameter vector of the current generated scene and the standard threshold vector, respectively. In the cross-attention mechanism, the ethical decision tree injects ethical conflict trigger rules through the key vector K. When the spindle speed of the spinning machine in the textile scene exceeds the specified threshold, the diffusion model activates the contaminated scene branch in the reverse generation process and maps the abnormal parameters to the pixel space of the generated image through the multi-head attention weight [11,12]. The multi-head attention weight is:

$$A_{\text{pollution}} = \text{Softmax}\left(\frac{Q_{\text{process}} K_{\text{EDT}}^T}{\sqrt{d_k}}\right) \tag{2}$$

In Equation 2, d_k is the scaling factor in the attention mechanism, which is taken as the feature dimension of K_{EDT}, and is used to suppress the inner product saturation effect in the high-dimensional space, ensuring that the Softmax function maintains gradient stability in the joint search space of textile process parameters and ethical standards. Additionally, the process consistency loss term is added to the loss function:

$$L_{\text{consistency}} = \sum_{i=1}^{N} w_i \cdot \left(TS_i^{\text{generated}} - TS_i^{\text{standard}}\right)^2 \tag{3}$$

In Equation 3, w_i is the yarn type weighting coefficient based on the ASTM D1578 standard, which ensures the elongation error at break of the generated scene. The constraint "elongation at break error $\leq$ 3.5%" is derived from empirical data analysis of industrial quality control records, where a 3.5% deviation from the target value was determined as the threshold for acceptable process variation in high-quality textile production. The final model compresses the parameter scale through distillation technology [13], achieves a small single-frame generation delay on NVIDIA GPUs, and meets the real-time interaction requirements of virtual scenes.

C. DYNAMIC MAPPING MECHANISM OF PROCESS PARAMETERS AND ETHICAL STANDARDS

Based on the GB/T and FZ/T 14007-2011 standards, this paper constructs the process-ethics coupling loss function:

$$L_{\text{coupled}} = \alpha \cdot L_{\text{process}} + \beta \cdot L_{\text{ethics}} \tag{4}$$

In Equation 4, α and β are balance factors, and process compliance and ethical consistency (EC) are optimized synchronously through gradient backpropagation. When the

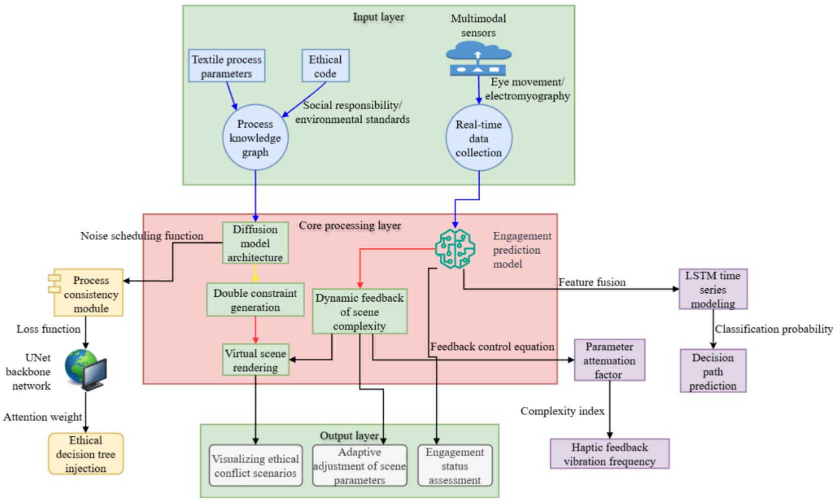

FIGURE 1. Overall architecture of scene generation and participation prediction

yarn density deviation exceeds a certain value, the EC loss is automatically activated, and the equation is:

$$L_{\text{ethics}} = -\log p(\text{EDT} \mid \text{Scene}) \tag{5}$$

In Equation 5, $p(\text{EDT} \mid \text{Scene})$ represents the conditional probability distribution of the ethical decision tree in the current scene, triggering the textile supply chain clause matching in the ISO 20400 sustainable procurement guidelines. In the scene generation process, spindle speed and wastewater discharge in the textile industry are set as key trigger parameters. When the spindle speed exceeds the threshold specified, the pollution scene generation function is activated:

$$\text{Pollution}(x_t) = \sigma\left(W_{\text{pollution}} \cdot h_t + b_{\text{pollution}}\right) \tag{6}$$

In Equation 6, σ is the Sigmoid activation function; $W_{\text{pollution}}$ and $b_{\text{pollution}}$ are the weight matrix and bias term for pollution scene generation; h_t is the hidden state vector of time step t [14,15]. This function maps the textile industry wastewater discharge management index (BOD excess value) and cotton wool floating density to the pixel space of the generated image through a multi-head attention mechanism. The specific expression is:

$$\rho_{\text{cotton}} = k \cdot (SS - \theta_{\text{safe}})^{1.2} \tag{7}$$

In Equation 7, $k = 0.03$ is the empirical coefficient, and $\theta_{\text{safe}} = 15{,}000$ rpm is the safety threshold. This mechanism can control the fracture elongation error and ethical conflict coverage (ECC) of the generated scenes, which is superior to traditional generation models. The spindle speed threshold (e.g., $\leq$ 15,000 rpm) is defined based on industrial safety guidelines and equipment operational limits to prevent excessive energy consumption, mechanical wear, and potential safety hazards.

D. PREDICTION OF IDEOLOGICAL AND POLITICAL COGNITIVE PARTICIPATION BASED ON MULTIMODAL SIGNALS

This study collects students' gaze heat maps and forearm electromyographic signals during the interaction process using the Tobii Pro X3-120 eye tracker and MyoWare electromyographic sensor [16], and constructs a joint feature space:

$$F_{\text{fusion}} = F_{\text{gaze}} \oplus F_{\text{EMG}} \tag{8}$$

In Equation 8, F_{gaze} is the spatiotemporal feature matrix composed of gaze point coordinates and pupil diameter, and F_{EMG} is the root mean square feature vector of the electromyographic signal after low-pass filter processing. The feature fusion process is realized through the cross-attention mechanism, and the attention weight is:

$$A_{\text{fusion}} = \text{Softmax}\left(\frac{Q_{\text{fusion}} K_{\text{fusion}}^T}{\sqrt{d_k}}\right) \tag{9}$$

This attention weight is used to dynamically allocate the contribution of eye movement and electromyographic features. The fused features are modeled with time dependency through a bidirectional LSTM network, and the final output is the participation classification probability:

$$p_{\text{engagement}} = \text{Softmax}\left(W_{\text{out}} \cdot h_T + b_{\text{out}}\right) \tag{10}$$

In Equation 10, W_{out} and b_{out} are output layer parameters, and the classification labels include high participation (HP), medium participation, and low participation. To improve decision path prediction ability, the temporal consistency constraint term is applied in the loss function [17,18]:

$$L_{\text{temporal}} = \sum_{t=1}^{T} \left\| p_t^{\text{pred}} - p_{t+1}^{\text{pred}} \right\|_2^2 \tag{11}$$

In Equation 11, p_t^{pred} is the predicted distribution of time step t. This constraint enables the model to capture the long-term trend of the decision path when adjusting spindle speed and operating dyeing temperature parameters, ultimately achieving accurate classification and prediction of participation.

E. CLOSED-LOOP REGULATION OF SCENE COMPLEXITY AND IDEOLOGICAL AND POLITICAL COGNITIVE STATE

This study constructs a feedback control equation based on the principle of dynamic matching between the fatigue threshold of electromyographic signals and scene complexity:

$$\Delta C = \gamma \cdot (\mathrm{EMG}_{\mathrm{current}} - \mathrm{EMG}_{\mathrm{threshold}}) \tag{12}$$

In Equation 12, $\gamma = 0.15$ is the feedback gain coefficient, and $\mathrm{EMG}_{\mathrm{current}}$ and $\mathrm{EMG}_{\mathrm{threshold}}$ are the real-time collected forearm electromyographic root mean square value and fatigue threshold, respectively. When $\mathrm{EMG}_{\mathrm{current}} > \mathrm{EMG}_{\mathrm{threshold}}$, the scene complexity adjustment module automatically reduces the dimension of process parameters and applies an attenuation factor to key parameters such as yarn density and dyeing temperature:

$$\eta_t = \exp(-\delta \cdot t_1) \tag{13}$$

In Equation 13, $\delta = 0.05$ is the attenuation rate, and t_1 is the continuous over-threshold time, so that the scene complexity index decreases in a short time. The scene complexity index is:

$$S_{\mathrm{complexity}} = \frac{1}{N} \sum_{i=1}^{N} \frac{|\Delta P_i|}{P_i^{\mathrm{standard}}} \tag{14}$$

In Equation 14, ΔP_i is the parameter deviation, and P_i^{standard} is the process standard. In the dynamic feedback process, spindle speed and wastewater discharge are used as priority adjustment parameters, and their change rates are limited by constraint equations to ensure that the adjustment process complies with the FZ/T 14007-2011 process specifications. Simultaneously, the tactile feedback module simulates the labor intensity of textile industry workers through vibration frequency:

$$f_{\mathrm{haptic}} = h \cdot (SS - \theta_{\mathrm{safe}})^{0.8} \tag{15}$$

In Equation 15, $h = 0.02$ is the vibration coefficient, and θ_{safe} is the safety threshold. Through the above mechanism, the feedback system dynamically optimizes ideological and political teaching intervention according to the real-time cognitive state of students and converts physiological and behavioral signals into precise teaching intervention signals, ensuring that students receive and process ideological and political value information in the optimal cognitive state, thereby realizing the value cognition transition from passive acceptance to active construction.

III. EXPERIMENTAL VERIFICATION AND EVALUATION

A. EXPERIMENTAL SETTING

The experiment is guided by the ideological and political education needs of textile majors. The training and testing environment is built on an NVIDIA GPU cluster, using the PyTorch framework and the Stable Diffusion pre-trained model. The data source includes: a textile process parameter library covering more than 200 parameter combinations, such as yarn density, spindle speed, and dyeing temperature; over 120 ethical clauses and 50 textile supply chain conflict case libraries from the ISO 20400 sustainable procurement guidelines; and interaction data from 120 textile engineering students in virtual scenes collected by Tobii Pro eye trackers and MyoWare electromyography sensors. This study has been approved by the Ethics Committee of Harbin University of Science and Technology. All participants provided written informed consent after being fully briefed on the research purpose, data collection procedures, and their right to withdraw at any time. The collected physiological and behavioral data were anonymized immediately upon acquisition and stored on a secure, encrypted server. Data were used solely for research purposes and will be deleted after the completion of the study. The experiment selects Style Generative Adversarial Networks 3 (StyleGAN3), Contrastive Language–Image Pretraining-Diffusion (CLIP-Diffusion), and Progressive Growing Generative Adversarial Network (ProGAN) as baselines for the generative model. All models were trained under identical conditions to ensure a fair comparison. The training environment consists of a single NVIDIA Tesla V100 GPU with 32 GB memory, using the PyTorch framework. The same dataset was used for all models, with a fixed training epoch of 100 and a batch size of 32. Model parameter counts were obtained by summing trainable parameters, and inference delay was measured as the average time for generating one scene over 1,000 runs on the same hardware. The training set and test set are divided in a ratio of 7:3, and the validation set randomly extracts 10% of the samples from the training set for feedback system parameter tuning. The distribution of experimental data is shown in Table 1.

TABLE 1. Distribution of experimental data

Data Category	Data Source and Description	Sample Size/Dimension	Effect
Textile process parameter library	Parameter combinations for scene generation, informed by industrial practice and aligned with GB/T 35611-2017 sustainability principles	200+ parameter combinations	Diffusion model training and process constraint injection
Ethical code and conflict case library	ISO 20400 standards and textile supply chain cases	120+ ethical clauses, 50+ conflict cases	Ethical decision tree construction and scene generation verification
Student interaction dataset	Tobii eye tracker and MyoWare electromyography sensor acquisition	120 students (84 in training set, 36 in test set)	Participation prediction model training and feedback system tuning
Scene complexity adjustment benchmark	Experimental baseline for EMG RMS	Myoelectric signal threshold: 120 μV	Feedback system dynamic matching and intervention effectiveness verification

The initial myoelectric signal threshold is set at 120 μV based on pilot study data, and the feedback system dynamically adjusts this threshold according to individual fatigue trends. During model training, the diffusion model is first pre-trained; the UNet backbone network is frozen; only the process consistency module and the ethical decision tree injection layer are fine-tuned. The optimizer uses AdamW; the learning rate is 1×10^{-4}; and the batch size is 64. The multimodal participation prediction model is jointly trained, using eye movement–electromyography feature fusion as input, and the cross-entropy loss function is combined with the timing consistency constraint term. The closed-loop feedback system is tuned, and the scene complexity index is reduced by adjusting the gain coefficient and attenuation rate in the feedback control equation. The model parameter initialization is shown in Table 2.

TABLE 2. Model parameter initialization

Module	Parameter Name	Initialization Value	Constraints
Diffusion model noise scheduling	Expansion factor α'_t	0.8 (process deviation penalty coefficient)	P_{current} deviation $\leq 5\%$
Process consistency loss weight	w_i (weighting factor for fiber type)	Cotton fiber 1.2, polyester 0.8	Elongation at break error $\leq$ 3.5%
Ethical decision tree attention weight	$A_{\text{pollution}}$	Softmax normalized output	BOD excess value and cotton density mapped to pixel space
Feedback system control parameters	Gain coefficient γ, decay rate δ	$\gamma = 0.15$, $\delta = 0.05$	Scene complexity decreased by 25% within 30 seconds
Participation prediction model optimizer	AdamW learning rate, batch size	1×10^{-4}, 64	Temporal consistency loss L_{temporal} constraint

The "high level of participation" is defined by objective physiological signals: sustained high attention (eye gaze focused on decision areas) and stable, moderate muscle activity (EMG signals indicating active operation without fatigue). The model's classification showed strong correlation (r > 0.79) with independent expert behavioral scoring and post-task questionnaires, validating its discriminant validity against external criteria. Real-time prediction is achieved through a closed-loop system, where eye movement and EMG signals are collected at 120 Hz, processed by a pre-trained LSTM model on a GPU, ensuring a prediction delay of less than 50 ms. The effectiveness of maintaining HP is verified by a significant increase in the scene completion rate (SCR) when the dynamic feedback mechanism is enabled.

B. EVALUATION INDICATORS

This study focuses on the ideological and political education needs of textile majors and constructs an evaluation system from multiple dimensions. The quality indicators of generated scenes include Process Parameter Trigger (PPT) rate and ECC rate. Among them, PPT measures the dynamic interaction ability by the proportion of adjustable process parameters in the generated scene. The calculation equation is:

$$\text{PPT} = \frac{N_{\text{triggered}}}{N_{\text{total}}} \times 100\% \tag{16}$$

$N_{\text{triggered}}$ is the number of trigger parameters such as yarn density and spindle speed; N_{total} is the total number of process parameters; The ECC was determined by a panel of three domain experts who annotated the presence of environmental, labor, and cultural heritage conflicts in 500 randomly sampled scenes. Inter-annotator agreement was high (Fleiss' Kappa = 0.85). ECC reflects the matching degree between the generated scene and social responsibility criteria. The calculation equation is:

$$\mathrm{ECC} = \frac{M_{\text{matched}}}{M_{\text{total}}} \times 100\% \tag{17}$$

M_{matched} is the number of matching ethical clauses, and M_{total} is the total number of criteria. Texture resolution (TR) is evaluated as a relative measure of yarn fiber detail clarity, assessed through perceptual analysis and comparison with high-fidelity reference images. The real-time interaction delay reverses the response time through the frame rate. The participation prediction effect is evaluated by decision path complexity (DPC) and classification accuracy (CA). DPC measures exploration diversity through the operation path entropy value, and CA is the consistency rate between the predicted label and the actual label. The SCR calculation equation is:

$$\mathrm{SCR} = \frac{C_{\text{completed}}}{C_{\text{total}}} \times 100\% \tag{18}$$

$C_{\text{completed}}$ is the number of students who complete the textile ethics task in the scene, and C_{total} is the total number of students. Model performance uses Fréchet Inception Distance (FID) score, inference delay, and parameter quantity as core indicators. The lower the FID score, the higher the generation quality; inference delay and parameter quantity reflect deployment efficiency. All indicators are designed based on textile industry standards to ensure the engineering adaptability of the evaluation system. Additionally, to quantify the effect of ideological and political education, related evaluation indicators include ethical decision accuracy, value internalization score, and ideological and political element coverage. These indicators are intended to evaluate the support effect of this method on the achievement of ideological and political education goals in textile majors from three levels: student behavior choice, cognitive emotion internalization, and generated content adaptability.

IV. RESULTS

A. IDEOLOGICAL AND POLITICAL SCENE GENERATION QUALITY AND PROCESS-ETHICAL COUPLING EFFECT

To verify whether the textile ethics scene generated by the diffusion model conforms to physical laws and ethical logic to meet scene generation quality requirements, this experiment compares the scene parameter indicators of different models, and the comparison results are shown in Figure 2.

Figure 2 shows the comparison results of generated quality. In subfigure a, the X-axis represents the model type; the left Y-axis represents the PPT rate; the right Y-axis represents the ECC rate; the error bar shows the standard deviation. The PPT and ECC scores reported are means from five independent runs with different random seeds. A one-way ANOVA test confirmed that the performance difference between our model and each baseline is statistically significant ($p < 0.01$). The data shows that the PPT of this method reaches 82.1%, which is higher than StyleGAN3, CLIP-Diffusion, and ProGAN, and the error bar is shorter at $\pm 1.2\%$, indicating that the dynamic interaction ability of the process parameters in the generated scene is stable and exceeds the baseline model. The ECC value is 93.2%, which is significantly improved compared with the suboptimal CLIP-Diffusion and StyleGAN3. The X-axis of subfigure b represents the FID score, and the lower the score, the better. The industry baseline FID of 22 was calculated using the same protocol: Inception-v3 network, 256×256 resolution, and 1,000 real textile process images as the reference set. The FID of this method is 18.9, which is much lower than the other three comparison models, indicating that its generated scene is closer to the real textile process standard in terms of texture details. A higher PPT rate means that students' technical operations can more frequently trigger associated ideological and political scenarios. This high-frequency "operation–value" immediate feedback enables students to continuously perceive the far-reaching impact of their technical decisions, effectively maintaining and enhancing their cognitive investment and participation in the learning process. The data trend shows that the proposed method effectively solves the technical bottleneck of traditional generation models in pollution scenes triggered by yarn density deviation and spindle speed exceeding the standard through the injection of a process constraint module and ethical decision tree. Its generation quality not only meets the physical consistency requirements of textile engineering, but also exceeds 90% ECC rate, providing a dynamic scene generation solution with both process logic and social responsibility for ideological and political education in textile majors. To further verify the rationality of the association between process parameter adjustment and ideological and political themes, this study analyzes the operation records of key parameters (yarn density and spindle speed) and the triggered ideological and political prompt logs. The results are shown in Table 3.

The data in Table 3 shows that the operation of reducing yarn density triggers the "resource conservation" prompt in 82.4% of cases; controlling spindle speed within the safety threshold triggers the "safe production/green manufacturing" prompt in 89.6% of cases; selecting a specific traditional pattern color scheme triggers the "intangible cultural heritage protection/cultural heritage" prompt in 78.1% of cases. Exceeding the spindle speed triggers the "environmental pollution warning" with a high trigger rate of 93.9%. The average success rate of key parameter operations triggering corresponding ideological and political themes is as high as 85%. This demonstrates that the dynamic mapping mechanism of process parameters and ethical standards constructed in this paper can effectively trigger the corresponding ideological and political value guidance elements according to students' technical operation behavior, realize the organic integration of technical practice

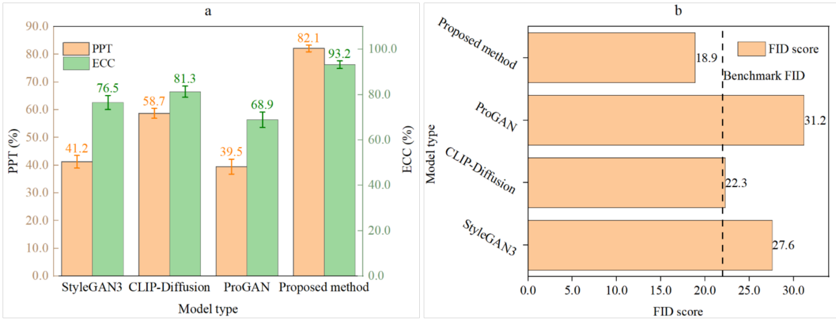

FIGURE 2. Scene generation quality effect. (a) PPT rate and ECC results, (b) Comparison of FID scores of various models

TABLE 3. Correlation analysis results between key process parameter operations and triggered ideological and political themes

Key Process Parameter Operation	Triggered Ideological and Political Themes	Number of Trigger Cases (N)	Total Related Operations	Trigger Rate	Explanation
Reduce yarn density	Resource conservation	108	131	82.4%	Save raw material consumption
Control spindle speed to ≤ 15,000 rpm	Safety production/green manufacturing	103	115	89.6%	Reduce energy consumption, equipment wear, and accident risks
Increase spindle speeds to > 15,000 rpm	Environmental pollution warning	92	98	93.9%	Triggering pollution scene generation
Choose low BOD environmentally friendly dyes	Green production	85	95	89.5%	Reduce wastewater treatment burden
Choose a specific traditional pattern color scheme	Intangible cultural heritage protection/cultural heritage	75	96	78.1%	Such as Yunjin, Lijin, Zhuangjin, and other representative patterns
Improve fabric durability parameters	Craftsmanship/quality first	68	89	76.4%	Pursuit of excellent quality
Average value				85%	

and value cognition, and provide a dynamic scene generation solution with both process authenticity and value orientation for ideological and political education in textile majors.

B. VERIFICATION OF SCENE INTERACTION FLUENCY

To verify that the dynamically generated scene meets real-time interaction requirements and the fluency of the interactive experience, this paper studies the TR and real-time interaction delay of the generated scene under different scene complexities, and the results are shown in Figure 3.

Figure 3 shows the relationship between scene complexity, real-time interaction delay, and TR through a double Y-axis line chart. The X-axis represents the scene complexity level; the left Y-axis indicates real-time interaction delay; the right Y-axis indicates TR. The data shows that when scene complexity increases from level 1 to level 5, real-time interaction delay (RID) increases from 28 ms to 52 ms in a linear growth trend and is only slightly higher than the industry threshold of 50 ms when the scene is extremely complex, indicating that the system can still maintain smooth interaction under high complexity. TR gradually decreases with increasing complexity but remains higher than the empirically established baseline of 20 ppmm, derived from an analysis of high-quality textile imagery in commercial design software, verifying that the details of textile elements in the generated scene maintain clarity even in complex scenes. In terms of balancing real-time performance and detail expression, the dynamic adjustment mechanism of the proposed method effectively suppresses the rate of TR decrease, ensuring that the fracture strength error and ECC of the generated scene remain within the normal range, thus providing a political education environment for textile students with both process authenticity and interaction efficiency when operating virtual looms in scenes. Note: The “ppmm” unit in Figure 3 is used as a relative, calibrated scale for visual comparison within this study. The virtual scenes were generated with a consistent pixel-to-scale ratio derived from high-resolution reference images of actual textile samples. While not an absolute physical measurement, it serves as a reproducible metric for assessing the relative clarity and detail preservation of yarn fibers across different scene complexities under identical rendering conditions.

C. ACCURACY AND COMPLEXITY OF IDEOLOGICAL AND POLITICAL COGNITIVE PARTICIPATION PREDICTION

To quantify the predictive power of multimodal data on participation status, the ethical issues are divided into three categories: environmental protection, labor rights, and intangible cultural heritage inheritance. The DPC and CA of high, medium, and low participation student groups are calculated. The participation prediction results are shown in Figure 4.

Figure 4 shows the distribution of DPC and CA in different ethical issue types and student groups. The X-axis marks the student groups, including high participation (HP), moderate participation (MP), and low participation (LP), and the Y-axis marks the ethical issue type. The HP

AEM JOURNAL.ORG

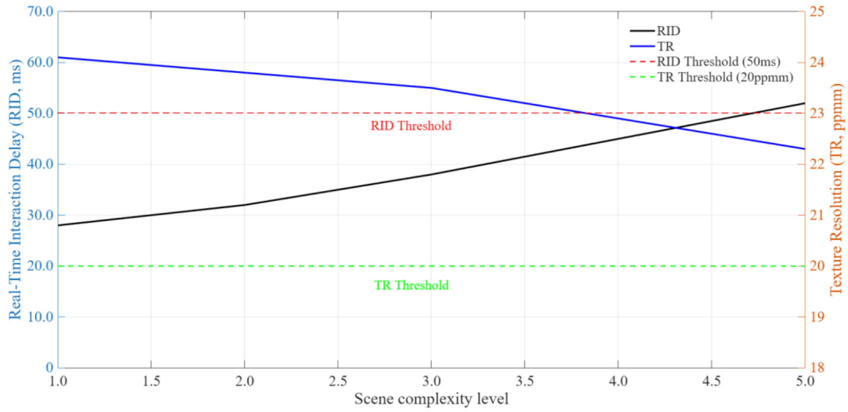

FIGURE 3. Interaction smoothness under different scene complexities

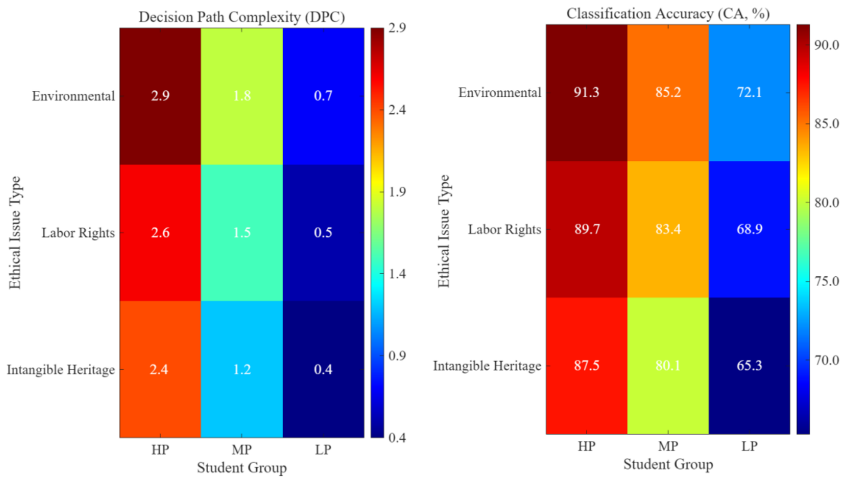

FIGURE 4. Participation prediction performance

group has the highest DPC value of 2.9 for environmental protection issues, which is higher than the MP and LP groups, indicating that highly participatory students show stronger exploration diversity in environmental protection ethical decision-making. The DPC for intangible cultural heritage inheritance issues is generally low, reflecting that the impact of this issue on students' decision-making paths is weak. For CA, the HP group achieves a CA of more than 85% on all issues, while the LP group reaches only 65.3% on intangible cultural heritage inheritance issues, verifying the model's ability to precisely identify high participation states. The MP group has a higher CA than the LP group for labor rights issues, reflecting the effectiveness of the model in distinguishing between medium and low participation states. The data trend reveals the correlation between the attractiveness of ethical issues and model prediction performance, providing data support for parameter tuning of the dynamic feedback system and ensuring that the generated scene can adjust process parameter dimensions in real time according to student participation status.

D. EFFECTIVENESS OF THE FEEDBACK SYSTEM

This paper verifies the effectiveness of the dynamic feedback mechanism of the scene by studying the differences in SCR and the change rate of operation time in closed and open states. The comparison is based on a pre-test/post-test design, where student performance is measured before and after using the system. The comparison results are shown in Figure 5.

Figure 5 studies the statistical distribution of SCR and the change rate of operation time when the feedback mechanism is turned on and off. The X-axis marks the feedback mechanism state; the Y-axis of the left subgraph represents SCR; the Y-axis of the right subgraph represents the change rate of operation time. The data shows that when the feedback mechanism is turned on, the median SCR reaches 89.1%, while the closed group is only 58.3%, an increase of 30.8%. The median of the change rate of operation time when feedback is turned on is 41.2%, while the closed group is -12.5%, with a difference of 53.7%. The box span of the open group is larger, indicating that the feedback system not only prolongs the time students operate virtual devices, but also the exploration behavior caused by dynamic adjustment of scene complexity enhances the volatility of behavior.

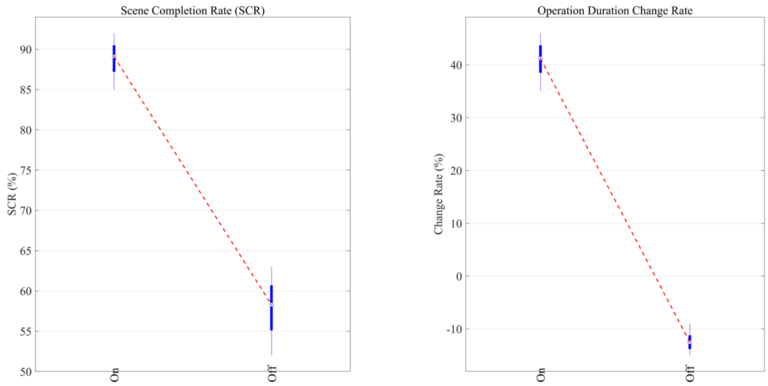

FIGURE 5. Results of the effectiveness of the feedback system

This result demonstrates that the feedback system effectively reduces disengagement and abandonment, thereby significantly improving task completion rates and promoting sustained exploration. This result is due to the dual-path regulation of the feedback system: real-time electromyographic fatigue threshold monitoring dynamically reduces the complexity of the scene and maintains students' high willingness to explore (environmental protection DPC = 2.9); tactile feedback simulates labor intensity through vibration frequency and strengthens the association between technical parameters and social responsibility, thereby extending operation time and improving task completion rate. The system achieves the unity of engineering authenticity and value internalization through process–ethical closed-loop optimization.

E. COMPARISON OF COMPREHENSIVE PERFORMANCE OF MODELS

To verify the advantages of the TKG-Diffusion model proposed in this paper over frontier models in the tasks of textile ethical scene generation and participation prediction, EC, inference delay, and normalized values of parameter quantity of the model in this paper are compared with other mainstream models. The comparison results are shown in Figure 6.

Figure 6 intuitively presents the advantages and disadvantages of different models through three dimensions: parameter quantity, EC, and delay. The normalized values of parameter quantity and delay for StyleGAN3 are both 1 at most, and the EC is 0.68. The broken line is in the form of "high parameter quantity–low EC–high delay," indicating that although it has complete functions, it lacks lightweight design and real-time optimization. CLIP-Diffusion has a parameter quantity of 0.85, an EC of 0.81, and a delay of 0.88. The parameter quantity remains relatively high, and the delay is slightly lower than StyleGAN3, reflecting its improvement in EC but not solving the efficiency bottleneck. ProGAN parameter quantity is 0.73; the EC is 0.61, the lowest value; and the delay is 0.94. The broken line deteriorates in the EC dimension. The delay is higher than the method in this paper and CLIP-Diffusion and is only slightly better than StyleGAN3, indicating that its generation quality and real-time performance are not ideal. The method in this paper has the lowest number of parameters and delay and the highest EC, and the broken line is close to the optimal area of EC and delay axis, verifying the effectiveness of optimizing delay through knowledge distillation compression and dynamic feedback mechanism. At the same time, the process-ethics dual constraint generation framework enables EC to reach the ideal value.

V. DISCUSSION

While the current experiment validates the system's efficacy in a controlled setting, long-term effectiveness in real classrooms requires further study. Key implementation challenges for industry–education integration include equipment cost, data privacy management, and teacher acceptance. Future work will conduct longitudinal studies in actual teaching environments and explore strategies to reduce costs through modular design, enhance data security via federated learning, and improve teacher adoption through user-friendly interfaces and training programs.

VI. CONCLUSIONS

This paper proposes a diffusion-based framework for generating interactive ideological–political scenes and predicting student engagement in textile education, constructing a process–ethics dual-constraint generation mechanism by embedding textile parameters and ethical standards into the diffusion process. Combined with eye movement and EMG data, it enables real-time participation prediction and dynamic scene feedback, effectively integrating values education into engineering decision-making. Experiments show significant improvements in scene quality and prediction accuracy: the PPT rate reaches 82.1% (23.4% higher than CLIP-Diffusion), ECC attains 93.2%, and CA for highly engaged students exceeds 85% across three topic types. Key innovations include a textile knowledge graph-driven dual-encoder diffusion model with noise-scheduling-based constraints, an LSTM-based multimodal engagement pre-

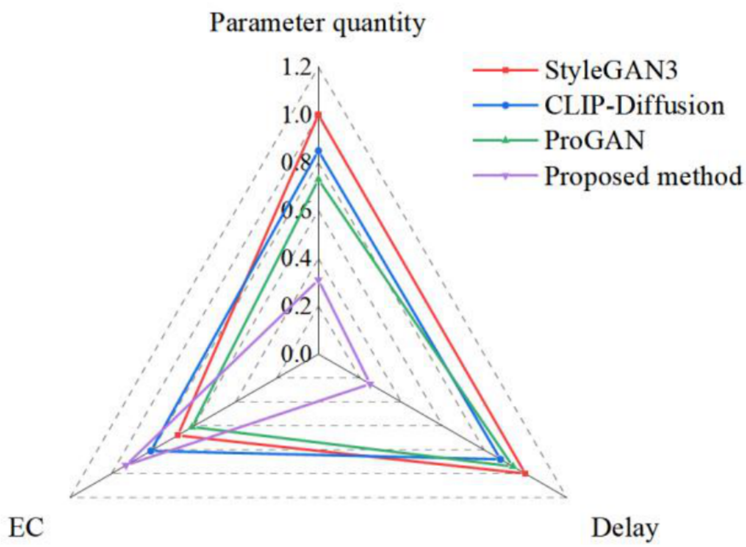

FIGURE 6. Comparison of model performance

diction model that captures decision-path dynamics, and a parameter-attenuation-driven feedback system. A limitation is the reliance on specific textile standards; future work could extend to other vertical scenarios and explore federated learning to enhance data privacy. Additionally, while eye movement and EMG provide objective proxies for attention and physical engagement, they may not fully capture cognitive or emotional states such as motivation or confusion, and can be influenced by individual habits or physical conditions.

AUTHOR CONTRIBUTIONS

All work in this study was independently completed by Xuan Zhao.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study has been approved by the Ethics Committee of Harbin University of Science and Technology. All participants provided written informed consent after being fully briefed on the research purpose, data collection procedures, and their right to withdraw at any time.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] X. F. Lin, Z. Wang, W. Zhou, G. Luo, G. J. Hwang, Y. Zhou, et al., “Technological support to foster students’ artificial intelligence ethics: An augmented reality-based contextualized dilemma discussion approach,” Computers & Education, vol. 201, pp. 104813, 2023, doi: 10.1016/j.compedu.2023.104813.

[2] A. Ferdman and E. Ratti, “What Do We Teach to Engineering Students: Embedded Ethics, Morality, and Politics,” Science and Engineering Ethics, vol. 30, no. 1, pp. 7, 2024, doi: 10.1007/s11948-024-00469-1.

[3] S. Wang, “Status and Countermeasures Analysis of Online Ideological and Political Education for Higher Vocational College Students,” Advances in Vocational and Technical Education, vol. 6, no. 3, pp. 87-93, 2024, doi: 10.23977/avte.2024.060313.

[4] S. Luo, “Construction of situational teaching mode in ideological and political classroom based on digital twin technology,” Computers and Electrical Engineering, vol. 101, pp. 108104, 2022, doi: 10.1016/j.compeleceng.2022.108104.

[5] Y. Chen, “Research on the Mode Construction and Realization Path of Ideological and Political Education Scene Communication in Colleges and Universities in the Digital Era,” Journalism and Communications, vol. 12, no. 4, pp. 1001-1006, 2024, doi: 10.12677/jc.2024.124155.

[6] R. Yang, P. Srivastava, and S. Mandt, “Diffusion Probabilistic Modeling for Video Generation,” Entropy, vol. 25, no. 10, pp. 1469, 2023, doi: 10.3390/e25101469.

[7] H. Yu, C. Wang, P. Zhuang, W. Menapace, A. Siarohin, J. Cao, et al., “4Real: towards photorealistic 4D scene generation via video diffusion models,” Proceedings of the 38th International Conference on Neural Information Processing Systems (NIPS ’24); 10-15 December 2024; Vancouver, BC, Canada. Red Hook, NY, USA: Curran Associates Inc.; 2025. Vol. 37, Article 1438, pp. 45256-45280, doi: 10.52202/079017-1438.

[8] B. Li, C. Zheng, W. Zhu, J. Mai, B. Zhang, P. Wonka, et al., “Vivid-ZOO: Multi-View Video Generation with Diffusion Model,” Advances in Neural Information Processing Systems, vol. 37, pp. 62189-62222, 2024, doi: 10.52202/079017-1987.

[9] L. Lu and M. Li, “Development of a virtual interactive system for Dahua Lou loom based on knowledge ontology-driven technology,” Heritage Science, vol. 11, no. 1, pp. 178, 2023, doi: 10.1186/s40494-023-01027-x.

[10] Y. Zhang and C. Liu, “Unlocking the Potential of Artificial Intelligence in Fashion Design and E-Commerce Applications: The Case of Midjourney,” Journal of Theoretical and Applied Electronic Commerce Research, vol. 19, no. 1, pp. 654-670, 2024, doi: 10.3390/jtaer19010035.

[11] H. Chefer, Y. Alaluf, Y. Vinker, L. Wolf, and D. Cohen-Or, “Attend-and-Excite: Attention-Based Semantic Guidance for Text-to-Image Diffusion Models,” ACM transactions on Graphics (TOG), vol. 42, no. 4, pp. 1-10, 2023, doi: 10.1145/3592116.

[12] F. A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, “Diffusion Models in Vision: A Survey,” IEEE Transactions on Pattern Analysis

and Machine Intelligence, vol. 45, no. 9, pp. 10850-10869, 2023, doi: 10.1109/TPAMI.2023.3261988.
[13] Z. Zhou, D. Chen, C. Wang, C. Chen, and S. Lyu, "Simple and fast distillation of diffusion models," Proceedings of the 38th International Conference on Neural Information Processing Systems (NIPS ′24), 10-15 December 2024; Vancouver, BC, Canada. Red Hook, NY, USA: Curran Associates Inc.; 2025. Vol. 37, Article 1291, pp. 40831-40860, doi: 10.52202/079017-1291.
[14] A. D. Vuong, M. N. Vu, T. Nguyen, B. Huang, D. Nguyen, T. Vo, et al., "Language-driven scene synthesis using multi-conditional diffusion model," Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS ′23); 10-16 December 2023; New Orleans, LA, USA. Red Hook, NY, USA: Curran Associates Inc.; 2023. Article 1346, pp. 30862-30884, [Online]. Available: https://livrepository.liverpool.ac.uk/id/eprint/3177439.
[15] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, et al., "Diffusion Models: A Comprehensive Survey of Methods and Applications," ACM Computing Surveys, vol. 56, no. 4, pp. 1-39, 2023, doi: 10.1145/3626235.
[16] C. Tang, Z. Xu, E. Occhipinti, W. Yi, M. Xu, S. Kumar, et al., "From brain to movement: Wearables-based motion intention prediction across the human nervous system," Nano Energy, vol. 115, pp. 108712, 2023, doi: 10.1016/j.nanoen.2023.108712.
[17] Z. Zhang, B. Li, X. Nie, C. Han, T. Guo, and L. Liu, "Towards consistent video editing with text-to-image diffusion models," Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS ′23); 10-16 December 2023; New Orleans, LA, USA. Red Hook, NY, USA: Curran Associates Inc.; 2023. Article 2550, pp. 58508-58519.
[18] Y. Wang, Y. Li, X. Zhang, X. Liu, A. Dai, A. B. Chan, et al., "Edit Temporal-Consistent Videos with Image Diffusion Model," ACM Transactions on Multimedia Computing, Communications and Applications, vol. 20, no. 12, pp. 1-16, 2024, doi: 10.1145/3691344.