Bridging Visual Dynamics and Reasoning Evaluation: Multimodal Large Language Models for Short Drama Quality Assessment

Qingyang Liu
Shanghai Jiao Tong University
Shanghai, China
narumimaria@sjtu.edu.cn

Jiangtong Li*
Tongji University
Shanghai, China
jiangtongli@tongji.edu.cn

Zelin Peng
Shanghai Jiao Tong University
Shanghai, China
zelin.peng@sjtu.edu.cn

Shaobo Wang
Shanghai Jiao Tong University
Shanghai, China
shaobowang1009@sjtu.edu.cn

Zhaohe Liao
Shanghai Jiao Tong University
Shanghai, China
zhaoheliao@sjtu.edu.cn

Shuochen Chang
Shanghai Jiao Tong University
Shanghai, China
csc1332741686@sjtu.edu.cn

Bingjie Gao
Shanghai Jiao Tong University
Shanghai, China
whynothaha@sjtu.edu.cn

Haonan Zhao
Shanghai Jiao Tong University
Shanghai, China
2zz-n-24@sjtu.edu.cn

Mu Liu
ByteDance
Beijing, China
liumu.lm@bytedance.com

Jidong Jiang
ByteDance
Beijing, China
jiangjidong@bytedance.com

Li Niu*
Shanghai Jiao Tong University
Shanghai, China
ustcnewly@sjtu.edu.cn

Abstract

Short drama quality assessment is crucial for industrial applications, including procurement decision support and addressing the cold-start problem in recommendation systems. However, existing video quality assessment approaches primarily focus on visual fidelity and often neglect higher-level narrative structure and reasoning logic. Likewise, other video understanding techniques tend to be event-centric, failing to adequately connect narrative elements with visual content. To bridge this gap between visual dynamics and narrative reasoning, we propose a user-centric quality indicator alongside an automated pipeline for constructing a Chain-of-Thought (CoT) dataset. To ensure data quality, this pipeline incorporates a hierarchical filtering mechanism that refines assessment accuracy, logical consistency, and the relevance of the reasoning, thereby steering the Multimodal Large Language Model (MLLM) toward human-aligned short drama assessment. We also develop the first MLLM for this task using a two-stage training framework: a Supervised Fine-Tuning (SFT) stage adapts the model to the assessment task, while a Group Relative Policy Optimization (GRPO) stage, using a customized reward function, further aligns its outputs with human preferences. Experimental results demonstrate that our model shows strong alignment with human preferences in short drama quality assessment and generates coherent explanations. Furthermore, online tests confirm our model boosts the cold-start performance of recommendation systems by improving multiple user engagement metrics. We release our evaluation set at https://github.com/HG-support/Short-Drama-Quality-Assessment.

CCS Concepts

• **Computing methodologies** → *Temporal reasoning*; • **Information systems** → *Multimedia information systems.*

Keywords

Multimodal Large Language Models, Short Drama Quality Assessment, Reasoning Model in Quality Evaluation

ACM Reference Format:
Qingyang Liu, Jiangtong Li, Zelin Peng, Shaobo Wang, Zhaohe Liao, Shuochen Chang, Bingjie Gao, Haonan Zhao, Mu Liu, Jidong Jiang, and Li Niu. 2026. Bridging Visual Dynamics and Reasoning Evaluation: Multimodal Large Language Models for Short Drama Quality Assessment. In *Proceedings of the ACM Web Conference 2026 (WWW '26), April 13–17, 2026, Dubai, United Arab Emirates.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3774904.3792827

1 Introduction

The proliferation of mobile internet and short-video platforms has led to a surge in short dramas, which are characterized by low production costs and fast-paced narratives suited to fragmented viewing habits. This trend motivates the need for automated quality assessment for two key applications: for upstream, identifying high-potential works during content procurement and reducing costs; and for downstream, providing signals that mitigate the cold-start problem for new series in recommender systems. However, this task is challenging, as it requires moving beyond surface-level visual attributes (*e.g.*, cinematography, pacing, acting) to assess

*Corresponding Authors.

WWW '26, Dubai, United Arab Emirates

ACM ISBN 979-8-4007-2307-0/2026/04
https://doi.org/10.1145/3774904.3792827

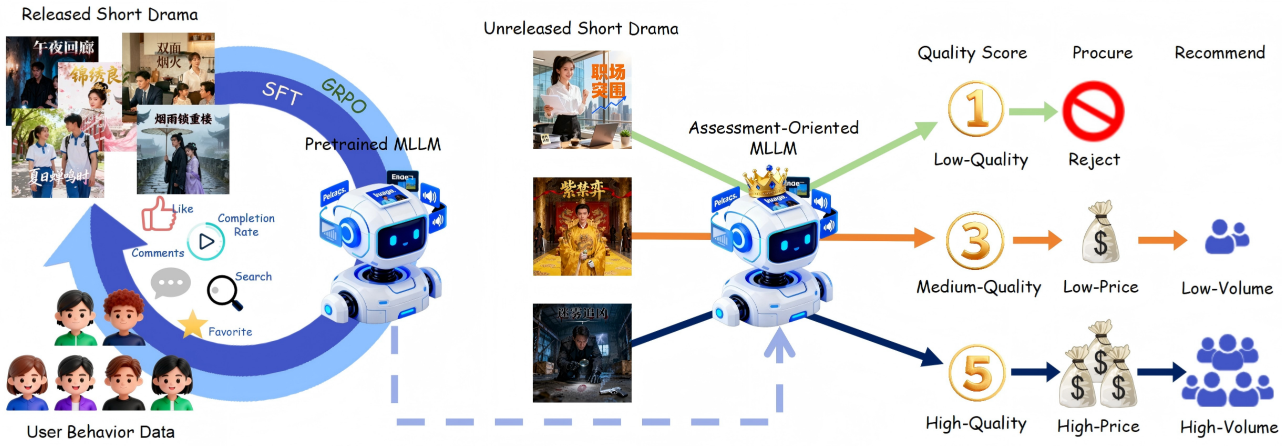

Figure 1: Our quality assessment task utilizes multi-dimensional user behavior data (*e.g.*, likes, completion rates) and a two-stage training framework (SFT and GRPO). This process is used to fine-tune an MLLM specifically for evaluating both visual fidelity and narrative reasoning. For unreleased short dramas, the model generates quality scores that inform procurement and recommendation strategies, thereby providing human-aligned evaluations for decision support.

if the narrative is logically coherent and compelling enough to retain audience engagement, placing high demands on multimodal reasoning capabilities. Furthermore, user preferences are subjective and diverse, making the transformation of implicit feedback into robust supervisory signals a key challenge in model training.

Existing video quality assessment methods primarily focus on the technical fidelity [13, 15, 61] or semantic alignment [21, 27, 37] of generative videos and largely overlook crucial creative elements such as acting performance, cinematographic language, narrative tension, and aesthetic style. Current video understanding methods are largely event-centric [4, 32, 40], with tasks like captioning [12, 45, 56], question answering [3, 33, 42], and retrieval [5, 28]. Therefore, they lack explicit metrics for acting credibility, cinematography, narrative pacing, or the overall coherence and appeal of the storyline. Moreover, existing benchmarks for video quality and understanding [5, 12, 45, 56] generally do not include audience-level preference data, limiting their utility for evaluating quality based on user engagement.

To align our quality metric with user preferences, we define a composite indicator, termed "hot value", derived from post-exposure user signals to quantify the quality of a short drama. For training purposes, we discretize this continuous hot value into a 1–5 quality score. Recent Reinforcement Learning (RL) based methods have demonstrated that MLLMs possess strong reasoning capabilities for video understanding [1, 14, 43]. However, directly applying RL to our task presents two challenges: first, a mismatch between the reasoning process and the final score in some samples, and second, reasoning processes that are often overly brief or semantically shallow. Rewarding the model solely for correctly predicting the score in such cases can prevent it from learning the underlying assessment logic. Therefore, a high-quality dataset with CoT annotations is essential to initially structure the reasoning process and output format of MLLMs. Since conditioning the MLLM on a target score to generate a reasoning process is highly likely to induce hallucinations [17, 17, 34], we design a sample-and-filter pipeline to construct our dataset. In the sampling stage, we provide the short drama video as input to SEED1.6-VL and prompt it to generate a quality score accompanied by a plot-aware reasoning process. In the filtering stage, we again use SEED1.6-VL to discard erroneous or low-quality samples, retaining only instances with the correct score, an accurate analysis, and strong narrative relevance. This pipeline enables the automated construction of a short drama quality assessment dataset featuring high-quality CoT annotations.

During training, we employ a two-stage post-training strategy to train the short drama quality assessment MLLM. First, an SFT stage uses our CoT dataset to adapt the MLLM to the desired reasoning structure and output format. Next, to better align the assessments of MLLM with user perspectives on narrative and visual elements and to improve its generalization, we apply GRPO using a custom suite of reward functions. In addition to conventional rewards for answer accuracy and format, we introduce two novel rewards. A caption consistency reward compels the MLLM to ground its quality assessment in the narrative context. A user comment consistency reward encourages the MLLM to evaluate dramas in a unified manner that reflects human integration of visual and narrative aspects. Extensive experiments on our short drama quality assessment dataset and online A/B tests for recommendation cold-start demonstrate that our model effectively evaluates short dramas in line with human preferences, leading to improved downstream performance. Our contributions can be summarized as follows:

- We define a user-centric quality indicator (hot value) and propose an automated pipeline to construct a high-quality CoT dataset, mitigating hallucination in the quality reasoning generation process.
- We design novel reward functions with narrative context and user comments, to guide the MLLM toward human-like assessment that integrates visual dynamics and narrative reasoning.

- We develop the first interpretable MLLM for short drama quality assessment, which is trained on our novel dataset via a two-stage framework to align with user preferences.
- Extensive experiments on our short drama quality assessment dataset and online A/B tests for recommendation cold-start demonstrate that our model effectively evaluates short dramas in line with human preferences.

2 Related work

2.1 Video Quality Assessment

Existing video quality assessment (VQA) methods primarily focus on evaluating the technical perceptual fidelity such as spatial distortion and temporal coherence [9, 29] for streaming media videos or semantic alignment [21, 37, 58] for AI-generated content (AIGC) videos. For perceptual fidelity assessment, traditional full-reference models, such as SSIM [46], VIF [35], and VMAF [25], estimate perceptual fidelity by comparing pristine and distorted signals, while recent no-reference and deep learning-based models leverage spatio-temporal features [20, 49] or transformer architectures [9, 29] to capture complex distortions. For semantic alignment, [37] proposes multi-granularity text–temporal fusion for fine-grained prompt–video consistency, while [27, 58] design holistic AIGC-VQA frameworks with dedicated alignment branches. However, these approaches remain largely limited to low-level fidelity or semantic correctness, showing little sensitivity to audience high-level quality perceptions such as acting performance, cinematographic language, narrative tension, and aesthetic style.

Some video recommendation methods [8, 10] model the user–content correlations from a matching perspective with limited interpretability, failing to present explicit rationale or content-level explanations for the video quality. Some studies [54, 59] infer video quality from user behavior data to enhance recommendation systems. [52] reweights user clicks by their dwell time to better reflect true engagement and improve recommendation accuracy. [54] enhances recommender systems by weighting clicks with dwell time to distinguish casual from engaged user interactions. [60] extracts video quality signals from watch time by removing biases from duration and exposure, yielding recommendations that better reflect true user interest. Nevertheless, these approaches depend on post-exposure indicators and are unable to directly analyze video quality based on visual content alone, limiting their applicability to videos without or with sparse user engagement.

2.2 Multimodal Large Language Models for Video Reasoning

Multimodal large language models (MLLMs) have achieved remarkable progress in video understanding tasks [4, 6, 40, 47], demonstrating strong capabilities in spatiotemporal reasoning [19, 31], cross-modal alignment [24, 48], and semantic comprehension [22, 33, 42]. Recently, Qwen2.5-VL [1] extends MLLMs into a unified vision–language model which supports adaptive perception and strong spatiotemporal reasoning capabilities. InternVL3.5 [43] focuses on reasoning enhancement and efficiency optimization achieving excellent trade-offs between accuracy and latency. DeepSeek-VL2 [51] pays particular attention to robustness under diverse domains and high-resolution document comprehension, making it suitable for OCR-heavy, layout-rich, and chart reasoning tasks. VideoLLaMA2 [7] augments spatio-temporal convolutions with an audio branch to integrate richer multimodal evidence. LongVA [55] extends the language backbone's context window to process substantially longer sequences without specialized video training. VISA [53] couples world knowledge with object tracking to enable knowledge-driven video object segmentation within an MLLM. Recently, reinforcement learning has been widely applied in MLLM based video reasoning methods. VideoChat-R1 [23] uses reinforcement fine-tuning with GRPO to markedly boost spatio-temporal perception and reasoning in video MLLMs. Video-R1 [11] introduces a temporal-aware GRPO variant and mixed image-video reasoning data to achieve strong video reasoning performance. DeepVideo-R1 [33] improves video reasoning by reformulating GRPO into a regression objective and adding difficulty-aware augmentation to stabilize RFT. However, most video understanding approaches and benchmarks have primarily focused on video perception tasks such as identifying relevant objects [30, 62], event detection [16, 41], localizing temporal segments [5, 28] and summarizing the events [12, 45]. Despite these advances in compression, long-context modeling, and knowledge-aware segmentation, hey remain oriented toward perception and event understanding and the development of MLLMs with user perspective video quality reasoning capabilities remains largely unexplored.

3 Method

3.1 Data Construction

To address the need for high-quality training data, we constructed a short drama quality assessment dataset comprising 23,472 dramas with significant user engagement from the Hongguo platform, spanning 26 major genres. This dataset serves a dual purpose: (1) it provides an objective quality metric, the "hot value", which reflects user-perceived popularity; (2) it offers supervisory signals to guide model optimization toward human-aligned evaluation criteria. All prompts used during data construction are formulated in English.

3.1.1 Hot Value. To align our quality metric with user preferences, we define a "hot value" derived from post-exposure user signals to measure the performance of a short drama. This value quantifies overall popularity by combining nine indicators from four categories: consumption, completion, interaction, and search. For the eight non-ratio indicators (*i.e.*, daily average watch time, daily average UV, favorites, likes, six-month cumulative watch time, six-month cumulative UV, comments, and search), we apply a piecewise function to normalize the raw value of each indicator, x_i, to a score, u_i, in the range [5,100+), as defined below:

$$u_i = \begin{cases} 5 + 15\dfrac{x_i - x^i_{min}}{x^i_{p30} - x^i_{min}}, & x_i < x^i_{p30}, \\ 20 + 50\dfrac{x_i - x^i_{p30}}{x^i_{p93} - x^i_{p30}}, & x^i_{p30} \le x_i < x^i_{p93}, \\ 70 + 30\dfrac{x_i - x^i_{p93}}{x^i_{10} - x^i_{p93}}, & x_i \ge x^i_{p93}. \end{cases} \tag{1}$$

Here, x_i, x^i_{min}, x^i_{p30}, x^i_{p93}, and x^i_{10} denote the raw value, minimum value, 30th percentile, 93rd percentile, and 10th-ranked value of the

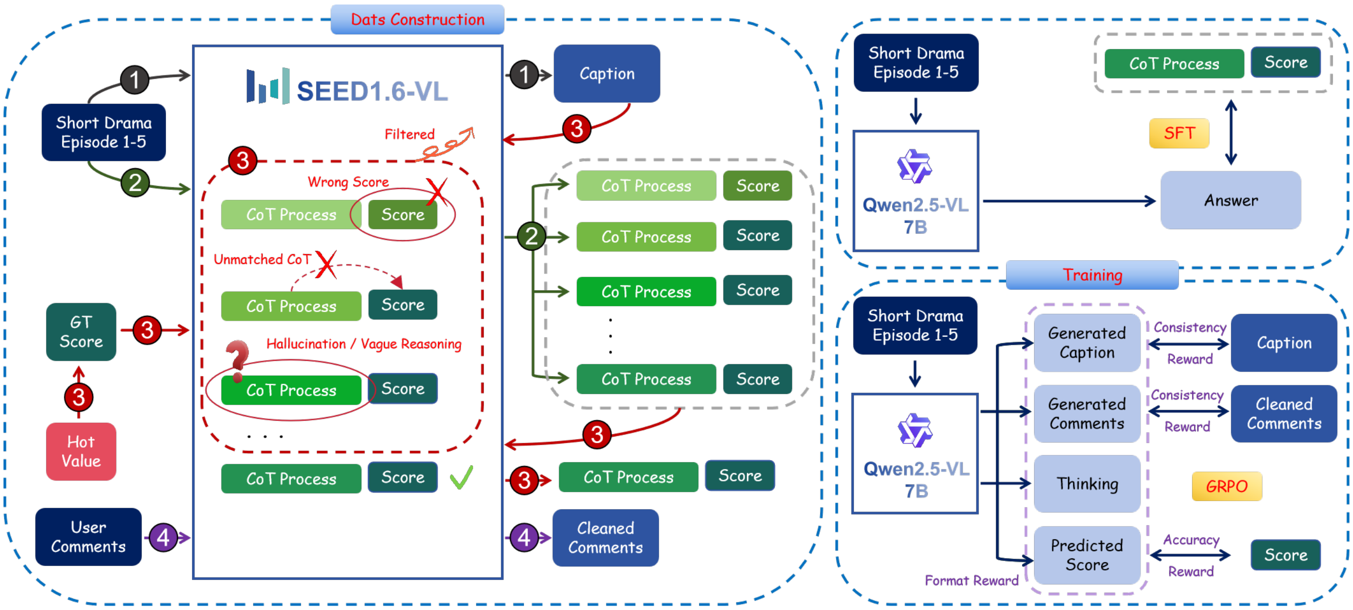

Figure 2: Our framework comprises two stages: data construction and model training. For data construction (left), we utilize SEED1.6-VL to generate high-quality supervisory signals. The data pipeline begins by ① generating a storyline caption for each video, followed by ② multi-round sampling to produce candidate quality scores and reasoning. These candidates are then ③ filtered using three rules based on ground-truth scores derived from our "hot value" metric. In parallel, ④ associated user comments are cleaned for subsequent training. In the training stage (right), we first conduct SFT on the CoT data to adapt the Qwen2.5-VL-7B model to our task's specific output structure. We then apply GRPO with four customized reward functions to encourage assessments that are structured, faithful, and aligned with user preferences.

indicator, respectively. For the ratio indicator completion rate x_9, we apply a capped piecewise normalization to obtain its score u_9:

$$u_9 = \begin{cases} 100, & x_9 > 0.7, \\ 3 + 12\dfrac{x_9}{0.3}, & 0 \le x_9 < 0.3 \text{ or } x_9 \text{ is null}, \\ 15 + 85\dfrac{x_9 - 0.3}{0.7 - 0.3}, & 0.3 \le x_9 \le 0.7. \end{cases} \tag{2}$$

Here, u_9 is the normalized score for the completion rate x_9 with a range of [3,100]. The score is capped at 100 for rates exceeding 70%, as this is deemed a sufficient indicator of high quality. The hot value H is obtained by a weighted sum of the nine indicator scores, where $\{w_i\}_{i=1}^{9}$ are their corresponding weights:

$$H = \sum_{i=1}^{9} w_i\, u_i, \quad \text{where} \sum_{i=1}^{9} w_i = 1. \tag{3}$$

During the normalization of each non-ratio indicator, the 10th-ranked value serves as a reference baseline, which allows the scores of top-performing items to exceed 100. This unbounded design ensures that the hot value can scale with exceptionally high composite scores, accurately reflecting the popularity of viral dramas without an artificial ceiling. Results from online A/B tests within our recommendation service confirm that the hot value metric is strongly correlated with user preferences and works effectively in real-world scenarios.

3.1.2 Data Construction Pipeline. Based on empirical observations, the popularity of most short dramas peaks within two months of release. We therefore only consider dramas older than two months and use their maximum hot value as a metric for overall quality. Since the distribution of this hot value has multiple modes, we use the K-means clustering method to map each short drama to a quality score from 1 to 5. Our analysis also reveals that while viewer completion rates increase significantly after the first few episodes, the total audience size tends to decline. This highlights the critical importance of a strong start. Therefore, we only use the first five episodes of each drama during dataset construction.

Step ①: Video Caption Preparation. The primary inputs for each data entry are the videos of the first five episodes and their quality score. To create a textual summary for each entry, we input the videos of these five episodes into the SEED1.6-VL model and prompt it to generate a concise plot summary, which serves as the official video caption for the data filtering and quality assessment.

Step ②: CoT Data Sampling. With the video inputs prepared, we generate the core reasoning data using a sample-and-filter method. In the sampling stage, the first five episodes of each drama are provided to SEED1.6-VL. The model is prompted to predict a quality score and provide a detailed reasoning process. This prompt guides the model to evaluate the drama based on several human-centric viewing criteria: thematic novelty, emotional climaxes and twists, audience resonance, cinematography, acting performance, and suspense design in the ending. To capture a wide range of evaluations, we perform multiple sampling rounds for each drama, generating diverse score assignments and reasoning processes.

Step ③: High-Quality CoT Filtering. To ensure the reliability of our training data, the generated CoT samples then move to a filtering stage. In this step, we again use SEED1.6-VL to evaluate and filter out erroneous or low-quality samples, using the ground truth (GT) score and the previously generated video caption as references. We discard any low-quality samples that:

- Predict an incorrect quality score compared to the GT.
- Exhibit inconsistency between the predicted quality score and the reasoning process.
- Generate hallucinated or vacuous reasoning that is not grounded in the actual plot.

This rigorous filtering step produces high-quality CoT annotations that provide reliable supervision for the training.

Step ④: User Comment Curation. To incorporate audience perspectives, we pair each data entry with high-quality user comments. For each short drama, we extract the most-liked comments. In detail, we first collect up to 20 most-liked comments from the first five episodes, and then process them by SEED1.6-VL to remove comments unrelated to the storyline. This step yields a curated set of high-quality user comments that reflect the user-perceived quality of the short drama.

3.2 Training Strategy

We employ Qwen2.5-VL-7B-Instruct [1] as our base MLLM. The training pipeline consists of two sequential stages: a Supervised Fine-Tuning (SFT) stage for cold start, followed by Group Relative Policy Optimization (GRPO) stage for performance enhancement.

3.2.1 Supervised Fine-Tuning Stage. To stabilize the reasoning process and standardize the output format of MLLM, we first introduce a Supervised Fine-Tuning (SFT) stage. Leveraging the short drama CoT dataset we constructed, we use prompts identical to those used for CoT generation. Since SEED1.6-VL is much larger than Qwen2.5-VL-7B, it serves as a teacher model, generating high-quality reasoning traces that embody a fine-grained semantic understanding. By supervising the student model with these cleaned, high-quality CoT annotations, we effectively distill the reasoning ability and format consistency from the large MLLM into a lighter, more efficient model. This CoT-based distillation enables the smaller model to internalize the structured thinking process, yielding a solid foundation for the subsequent reinforcement learning stage to achieve significant performance gains.

3.2.2 Group Relative Policy Optimization Stage. To further bridge the gap between visual dynamics and narrative reasoning from the user perspective and to strengthen the generalization of MLLM across diverse genres, we apply an RL stage using GRPO. While GRPO is effective, it requires explicit reward signals to effectively guide the MLLM in assessing short drama quality. To address this, we define a customized combination of four rewards, which assess key aspects, including narrative consistency, user-centric alignment, scoring accuracy, and structural formatting.

Caption Consistency Reward. To mitigate narrative drift, we prompt the MLLM to generate a plot summary and reward its consistency with the caption generated by a much larger model in the dataset. The caption consistency reward is calculated as the average of ROUGE-1, ROUGE-2, and ROUGE-L scores between the generated plot summary (S) and the ground-truth caption (C):

$$\begin{aligned} R_{\text{caption}} = \text{ROUGE}(S, C) = \tfrac{1}{3}\big(&\text{ROUGE-1}(S, C) \\ &+ \text{ROUGE-2}(S, C) + \text{ROUGE-L}(S, C)\big). \end{aligned} \tag{4}$$

User Comments Consistency Reward. To encourage user-centric quality evaluation, we further prompt the MLLM to generate 10 plausible user comments for each short drama. The user comments consistency reward is calculated as the average of ROUGE-1, ROUGE-2, and ROUGE-L [26] scores between the generated user comments and the real user comments in the dataset. To ensure one-to-one alignment between generated and real user comments, we compute a pairwise ROUGE similarity matrix and apply the Hungarian algorithm [18] to obtain the maximum weight matching. Let $\{g_i\}_{i=1}^{K}$ be the generated comments and $\{r_j\}_{j=1}^{M}$ be the real user comments. We first calculate the ROUGE similarity matrix:

$$M_{ij} = \text{ROUGE}(g_i, r_j), \quad i = 1, \dots, K, \ j = 1, \dots, M. \tag{5}$$

We apply the Hungarian algorithm to find the optimal matching:

$$\mathcal{M} = \text{Hungarian}(M)\,, \tag{6}$$

Finally, the user comment consistency reward is computed as:

$$R_{\text{comments}} = \frac{1}{|\mathcal{M}|} \sum_{(i,j)\in\mathcal{M}} M_{ij}. \tag{7}$$

Accuracy Reward. To measure the correctness of the quality assessment, we apply an accuracy reward. The model is prompted to generate the final quality score (an integer from 1 to 5) within a box, which we extract via regular expressions and verify against the ground truth. The reward is defined as:

$$R_{\text{accuracy}} = \begin{cases} 1, & \text{if quality score = ground truth}\,, \\ 0, & \text{otherwise}\,. \end{cases} \tag{8}$$

Format Reward. To ensure the designed parts are present and correctly ordered, we apply a format reward. We instruct the model to enclose its generated caption, user comments, thinking process, and final score within <caption>···</caption>, <comments>··· </comments>, <think>···</think>, and <answer>···</answer> tags, respectively. The reward is defined as:

$$R_{\text{format}} = \begin{cases} 1, & \text{if the output format is correct}\,, \\ 0, & \text{otherwise}\,. \end{cases} \tag{9}$$

Optimization Process. Given an input x, the algorithm samples N candidate outputs $O = \{o_1, o_2, \dots, o_N\}$ from the old policy $\pi_{\theta_{\text{old}}}$. Each candidate o_j is assigned a composite reward R_j, computed as a weighted sum of the aforementioned rewards:

$$R = \lambda_1 R_{\text{caption}} + \lambda_2 R_{\text{comments}} + \lambda_3 R_{\text{accuracy}} + \lambda_4 R_{\text{format}}, \tag{10}$$

To assess the relative quality, GRPO normalizes rewards using the mean μ_R and standard deviation σ_R across outputs:

$$\tilde{R}_j = \frac{R_j - \mu_R}{\sigma_R}. \tag{11}$$

Table 1: Comparison of model performance on our short drama quality assessment benchmark in Accuracy, MSE, and PCC. Our model achieves the best results across all metrics against open-source and closed-source baselines.

Method	Accuracy ↑	MSE ↓	PCC ↑
Base Model			
Qwen2.5-VL-7B-Instruct [1]	0.254	3.529	0.126
Open-source Models			
LLaVA-NeXT-Video-7B-hf [57]	0.208	3.852	0.038
Keye-VL-1.5-8B [38]	0.248	3.681	0.075
VideoLLaMA3-7B [2]	0.237	3.732	0.059
InternVideo2.5-Chat-8B [44]	0.223	3.694	0.047
VideoChat-R1-7B-caption [23]	0.216	3.829	0.021
LongVU-Qwen2-7B [36]	0.239	3.693	0.068
Video-R1-7B [11]	0.230	3.799	0.049
GLM-4.5V [39]	0.331	2.934	0.213
Closed-source Models			
GPT-4o	0.326	3.353	0.156
Gemini-2.5-Pro	0.384	2.701	0.342
SEED1.6-VL	0.385	2.611	0.335
Ours	**0.438**	**1.962**	**0.502**

The policy is updated by maximizing a KL-regularized objective that balances performance improvement and training stability:

$$\max_{\pi_\theta} \mathbb{E}_{O\sim\pi_{\theta_{\text{old}}}(\cdot|x)} \left[\sum_{j=1}^{M} \frac{\pi_\theta(o_j)}{\pi_{\theta_{\text{old}}}(o_j)} \cdot \tilde{R}_j - \beta\, \mathrm{D}_{\text{KL}}(\pi_\theta \,\|\, \pi_{\text{ref}}) \right], \quad (12)$$

where β is a regularization coefficient, and π_{ref} denotes the reference policy. This formulation enables GRPO to integrate diverse reward signals while maintaining stable optimization.

In contrast to conventional policy optimization methods, GRPO employs a clipping mechanism to mitigate extreme policy deviations and a KL regularization term to preserve alignment with the reference model. Such a design facilitates stable integration of multi-dimensional rewards in the context of short-form drama quality evaluation.

4 Experiment

4.1 Experimental Setup

4.1.1 Implementation Details. We conduct our main experiments using the Qwen2.5-VL-7B-Instruct model [1]. We divide our dataset into a training set of 22,472 samples and an evaluation set of 1,000 samples. We first perform LoRA-based fine-tuning on 10,000 short drama samples using 8 NVIDIA H20 GPUs in the SFT stage. Subsequently, in the GRPO training stage, we train for one epoch on the remaining 12,472 samples using 32 NVIDIA H20 GPUs. For each query, we generate 16 candidate responses to compute group-relative rewards. We preprocess videos from both datasets by first sampling them at 1 FPS and then constraining the frame count and resolution based on video duration. We incorporate the Number Prompt strategy [50] during training and evaluation to overlay absolute timestamps on video frames, which provides precise temporal grounding and improves the MLLM's temporal perception. In the GRPO training stage, we set the reward combination coefficients to $\lambda_1 = 0.2$, $\lambda_2 = 0.2$, $\lambda_3 = 0.5$, and $\lambda_4 = 0.1$.

Table 2: Ablation study of the training strategy, showing that while both SFT and GRPO individually improve performance, the combined SFT→GRPO approach yields the best results.

Variant		Accuracy ↑	MSE ↓	PCC ↑
SFT	GRPO			
-	-	0.254	3.529	0.126
+	-	0.369	2.967	0.267
-	+	0.273	3.431	0.146
+	+	0.438	1.962	0.502

4.1.2 Evaluation Metrics. We formulate the short drama quality assessment task as a five-class classification problem. Accordingly, we first report Accuracy as the primary evaluation metric to measure the proportion of exact matches between predicted and ground-truth quality scores. However, Accuracy only measures exact matches and does not reflect the degree of deviation when predictions are close but not identical. To address this limitation, we incorporate Mean Squared Error (MSE) to quantify the difference between predicted and ground-truth scores, which offers a more fine-grained measure of prediction errors. We also report the Pearson Correlation Coefficient (PCC) to evaluate the linear correlation between predicted and ground-truth scores, which measures if the model's scoring trend aligns with human judgment.

4.2 Main Results

Table 1 presents the quantitative results on our short drama quality assessment benchmark. Our model achieves new state-of-the-art performance with an Accuracy of 0.438, a Mean Squared Error (MSE) of 1.962, and a Pearson Correlation Coefficient (PCC) of 0.502. This performance surpasses all baselines, including strong closed-source models such as SEED1.6-VL (0.385 Acc, 2.611 MSE, 0.335 PCC) and all open-source competitors.

Notably, our method achieves a 24.8% reduction in MSE and a 49.8% improvement in PCC compared to the next-best model. The large decrease in MSE indicates that our model produces more numerically precise quality predictions, while the high PCC score of 0.502 shows a strong positive correlation with human scoring trends, reflecting a better alignment with human judgment. These improvements across discrete (Accuracy) and continuous (MSE, PCC) metrics suggest that our integration of narrative reasoning, visual understanding, and user-centric supervision enables the MLLM to assess short dramas that align with human evaluation.

4.3 Ablation Study

4.3.1 Training Strategy. We compare four variants of our training process: the untrained base model (Qwen2.5-VL-7B-Instruct), an SFT-only model trained on the full dataset, a GRPO-only model

Table 3: Ablation study of GRPO reward components. With other rewards fixed, each new component (caption and comments consistency) individually improves performance, while combining them yields the greatest benefit.

Reward				Acc. ↑	MSE ↓	PCC ↑
Format	Accuracy	Caption	Comments			
+	+	-	-	0.403	2.645	0.327
+	+	+	-	0.423	2.322	0.406
+	+	-	+	0.414	2.465	0.390
+	+	+	+	0.438	1.962	0.502

Table 4: Results of online A/B test on Hongguo platform.

Online Metrics	Relative Improvement
Staytime	+0.101%
Video Consumption Time	+0.094%
Completion Rate	+0.286%
PV Click-through Rate	+0.732%

trained on the full dataset, and the full SFT→GRPO model. As summarized in Table 2, both SFT and GRPO improve performance over the base model. From a principled perspective, SFT distills structured reasoning traces and output conventions from the teacher, thereby constraining the output space to a canonical format and mitigating exposure bias before reinforcement learning, while also providing a more stable initialization for subsequent optimization. In contrast, applying GRPO directly to an untrained policy requires exploring a much larger hypothesis space with sparse and delayed reward signals, which often leads to unstable optimization and limited gains, and can make it difficult for the policy to discover high-quality behaviors reliably. Accordingly, the GRPO-only variant yields only a modest improvement, while the SFT→GRPO approach achieves the best performance. This pattern suggests that SFT provides a crucial warm start and a structured reasoning scaffold, enabling GRPO to refine decision-making more effectively after the policy and output structure have been stabilized.

4.3.2 Reward Design. To isolate the contribution of each new reward component, we conduct an ablation study, fixing the accuracy and format rewards to establish a baseline performance (Accuracy: 0.403, MSE: 2.645, PCC: 0.327). Under this controlled setting with a fixed total reward budget ($\sum_i \lambda_i = 1$), Table 3 shows that individually adding either reward component improves performance. Introducing caption consistency provides a substantial boost, increasing PCC to 0.406, while adding comments consistency also elevates PCC to 0.390. However, combining both yields the largest, synergistic gains, achieving our final superior scores of 0.438 in Accuracy, 1.962 in MSE, and 0.502 in PCC. Mechanistically, the notable jump in correlation from caption consistency anchors reasoning to plot-grounded semantics and reduces narrative drift, while comments consistency effectively injects user-centric priors, and their combination allows the model to align with both narrative structure and audience perception.

4.4 Online A/B Testing

To further validate the effectiveness for cold-start recommendation, we selected 500 newly released short dramas and ran a two-week online A/B test on Hongguo platform. We integrated our method into the existing recommendation workflow for a controlled comparison under identical conditions. Specifically, the treatment group's multi-objective ranker included our model's predicted quality score as a calibrated feature, while the control group used the production ranker without this feature; all other configurations were held constant. We evaluated four online metrics: (1) Stay time: the average per-user time spent within the short-drama context; (2) Video Consumption Time: the average per-user video watching time; (3) Completion Rate: the proportion of videos that users watch to the end; (4) PV Click-through Rate: the ratio of clicks to page views on drama entry points. Across the two-week window, the treatment group showed consistent improvements on all four metrics. This indicates that incorporating our model's quality score into the recommender improves early-stage matching and increases user engagement for newly released short dramas.

4.5 Case Study

To further illustrate our model's effectiveness, we present case studies for a low-quality and a high-quality drama in Figure 3. We compare the responses of the base model, the strongest baseline SEED1.6-VL, and our model. The untrained base model tends to produce generic and superficial descriptions, failing to connect visual or narrative cues to the overall quality of the short drama. In contrast, SEED1.6-VL, while capable of offering more detailed analyses and basic quality judgments, often fails to identify key narrative signals or the specific aspects that engage users. By comparison, our model demonstrates a more accurate capability in short-drama quality assessment. In contrast, our model provides a more nuanced assessment. For the low-quality drama, it identifies significant logical flaws that undermine the viewer experience (highlighted in yellow) and a forced plot development where villagers unquestioningly trust the antagonist (highlighted in green). For the high-quality drama, our model recognizes that the mutual admiration theme is a user-preferred narrative (highlighted in yellow) and identifies the male lead's appearance as a key factor in sustaining viewer engagement (highlighted in green). These findings, which align closely with real user preferences, show that our model evaluates short dramas from a user-centric perspective. This enhanced video understanding allows our model to assign both high and low scores with greater confidence and justification than the baseline models.

5 Conclusion

In this paper, we propose the first interpretable Multimodal Large Language Model (MLLM) for short drama quality assessment, which holistically bridges visual dynamics and narrative reasoning with user preference alignment. Our introduction of a user-centric hot value and a high-quality Chain-of-Thought (CoT) dataset, built via a rigorous sample-and-filter pipeline, provides the model with a robust foundation of explicit reasoning paths, which directly mitigates the tendency for hallucination and supports more transparent, evidence-grounded evaluations. Furthermore, our two-stage post-training strategy is crucial: Supervised Fine-Tuning (SFT) initially

GT Quality Score: 1

Qwen2.5-VL-7B-Instruct
<think>The short drama begins with a promising setup, introducing a family in a domestic setting, which quickly transitions into a more intense scenario involving medical and financial concerns. The characters' interactions are dynamic, with clear emotional stakes and a sense of urgency conveyed through their dialogue and body language. However, the narrative pacing is somewhat uneven, with some scenes feeling rushed while others drag slightly. The emotional climax is impactful, but the resolution seems abrupt, leaving some questions unanswered. The acting is generally solid, with the characters displaying a range of emotions, though there are moments where the expressions feel slightly forced. The use of close-ups and quick cuts enhances the dramatic effect, but the overall story could benefit from more depth in character development and a clearer resolution.</think><answer>4</answer>

SEED1.6-VL
<think>Strengths: The opening quickly establishes a high-stakes conflict with the protagonist needing 60,000 yuan for his mother's medical treatment by 6 PM, creating immediate tension. The ending suspense is maintained through multiple 'To Be Continued' intertitles and unresolved crises, such as the protagonist's collapse after a blood transaction and the failed rice sale, encouraging viewers to continue watching. Weaknesses: The narrative suffers from abrupt and illogical progression. For example, the doctor's unprofessional behavior (e.g., touching the female character's thigh, refusing treatment without payment) lacks plausible motivation, and the protagonist's decision to sell blood to a shady figure (307-315 seconds) is underdeveloped. Acting is stiff, with characters delivering lines monotonously—e.g., the female lead's flat reaction to the protagonist's collapse (448-459 seconds). Visuals are cheap and repetitive, relying on plain gray backdrops and minimal set design (e.g., the hospital room and street scenes). Thematic expression is superficial, reducing complex issues like medical debt and ethical dilemmas to melodramatic conflicts without emotional depth.</think><answer>2</answer>

Ours
<think>Strengths: The suspenseful ending design left unresolved tensions: the mother's medical bills were still unpaid, and the rice sales problem remained unsolved; in the final shot, multiple characters were left in an unresolved state, effectively pushing viewers to continue watching. Weaknesses: The lead actor's acting was stiff and his emotional expression was flat; for example, in the scene where he falls after selling blood, there was a lack of pain and weakness on his face. The plot is severely unrealistic: selling 800cc of blood and surviving medically is not feasible, which weakens credibility. There is insufficient logical groundwork: villagers blindly trusted "Stone" about the buyer refusing to accept the rice without evidence or reasonable motivation. The narrative pace is slow and disjointed: there are abrupt transitions such as the doctor's unprofessional behavior, and the scenery is repetitive, reducing immersion.</think><answer>1</answer>

GT Quality Score: 5

Qwen2.5-VL-7B-Instruct
<think>The short drama begins with a promising opening scene that captures the audience's attention through its use of a school setting and a romantic moment between two characters. However, the subsequent scenes introduce a series of conversations and interactions that feel somewhat disjointed and lack depth in terms of character development and emotional engagement. The dialogue-heavy scenes, while aiming for realism, often come across as overly scripted and lack the spontaneity needed to fully immerse the viewer. The ending, while leaving some questions unanswered, does not provide enough closure to satisfy the audience, which might lead to a sense of incompleteness. The overall pacing is uneven, with moments of high tension interspersed with quieter, less impactful scenes.</think><answer>3</answer>

SEED1.6-VL
<think>Strengths: The opening uses a dream sequence (18-54 seconds) with intimate scenes between the female lead and Pei Siyu, creating initial intrigue about their relationship. The plot twist of Song Jiajia pretending to be her best friend on a blind date (155-162 seconds) adds comedic and dramatic tension, as her exaggerated 'ugly' disguise (big glasses, messy braids) contrasts humorously with her internal panic. The emotional climax in the café, where Pei Siyu recognizes her necklace (299-303 seconds) and recalls their high school encounter (306-310 seconds), effectively connects past and present, evoking empathy for Jiajia's decade-long secret love (149 seconds). The ending suspense is maintained when Pei Siyu asks for her name again (337-339 seconds), leaving uncertainty about whether he truly remembers her.Weaknesses: The narrative pacing is uneven, with abrupt transitions between the dream, reality, and flashbacks (e.g., 17-18 seconds jump from school to blindfolded intimacy). Character motivations are underdeveloped: Pei Siyu's sudden interest in 'Ye Ningning' (245-247 seconds) lacks foreshadowing, and Jiajia's decision to sabotage the date feels forced. Emotional delivery is stiff in key scenes, such as Jiajia's reaction to the blind date mix-up (223-224 seconds), where her lines ('What bad luck!' feel overly scripted. Set design is repetitive, with the café (164 seconds) and bedroom (55 seconds) appearing generic and lacking distinct visual texture. Thematic expression of unrequited love is superficial, relying on voiceovers ('Ten years of secret love,' 149 seconds) rather than nuanced interactions.</think><answer>3</answer>

Ours
<think>Strengths: The opening scene, where the female lead hands a letter to the male lead in high school, quickly establishes romantic tension. The "mutual admiration" setup is effective, creating a nostalgic and romantic atmosphere. The male lead is good-looking, which enhances audience appeal. The emotional climax—an unexpected reunion during a blind date—smoothly connects with the main storyline, maintaining viewer engagement. Audience immersion is strengthened by realistic set designs, such as the cozy bedroom and elegant cafe shop, which align with the plot's romantic tone. The ending is clever, and the bar encounter builds a romantic atmosphere that leaves viewers curious about what will happen next. Weaknesses: Some dialogue feels awkward; for example, Ye Ningning's request to "help me go on a blind date and ruin it" comes across as overly scripted.</think><answer>5</answer>

Figure 3: Case study on one low-quality and one high-quality short drama.

aligns the model with the structure of quality assessment, after which Group Reward Policy Optimization (GRPO) with our designed rewards fine-tunes its judgment to capture the nuanced signals of a human-centered perspective. This staged optimization enables the model to internalize standardized assessment behaviors before being further aligned with user-centric criteria. As validated by extensive experiments and online A/B tests, this comprehensive approach ensures our model delivers evaluations that are not only accurate but also interpretable and aligned with user preferences.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (Grant No. 62471287, 62402341). We thank all engineers and annotation staff at Fanqie and Hongguo for their technical support and contributions to data annotation.

References

[1] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. 2025. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923* (2025).

[2] Zesen Cheng Zhiqiang Hu Yuqian Yuan Guanzheng Chen Sicong Leng Yuming Jiang Hang Zhang Xin Li Peng Jin Wenqi Zhang Fan Wang Lidong Bing Deli Zhao Boqiang Zhang, Kehan Li. 2025. VideoLLaMA 3: Frontier Multimodal Foundation Models for Image and Video Understanding. *arXiv preprint arXiv:2501.13106* (2025). https://arxiv.org/abs/2501.13106

[3] Xinlong Chen, Yuanxing Zhang, Yushuo Guan, Bohan Zeng, Yang Shi, Sihan Yang, Pengfei Wan, Qiang Liu, Liang Wang, and Tieniu Tan. 2025. VersaVid-R1: A Versatile Video Understanding and Reasoning Model from Question Answering to Captioning Tasks. *arXiv preprint arXiv:2506.09079* (2025).

[4] Yi Chen, Yuying Ge, Rui Wang, Yixiao Ge, Junhao Cheng, Ying Shan, and Xihui Liu. 2025. GRPO-CARE: Consistency-Aware Reinforcement Learning for Multimodal Reasoning. *arXiv preprint arXiv:2506.16141* (2025).

[5] Yi Chen, Yuying Ge, Rui Wang, Yixiao Ge, Lu Qiu, Ying Shan, and Xihui Liu. 2025. Exploring the effect of reinforcement learning on video understanding: Insights from seed-bench-r1. *arXiv preprint arXiv:2503.24376* (2025).

[6] Yukang Chen, Wei Huang, Baifeng Shi, Qinghao Hu, Hanrong Ye, Ligeng Zhu, Zhijian Liu, Pavlo Molchanov, Jan Kautz, Xiaojuan Qi, et al. 2025. Scaling rl to long videos. *arXiv preprint arXiv:2507.07966* (2025).

[7] Zesen Cheng, Sicong Leng, Hang Zhang, Yifei Xin, Xin Li, Guanzheng Chen, Yongxin Zhu, Wenqi Zhang, Ziyang Luo, Deli Zhao, et al. 2024. Videollama 2: Advancing spatial-temporal modeling and audio understanding in video-llms. *arXiv preprint arXiv:2406.07476* (2024).

[8] Lulu Dong, Guoxiu He, and Aixin Sun. 2024. Not All Videos Become Outdated: Short-Video Recommendation by Learning to Deconfound Release Interval Bias. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 179–188.

[9] Huiyu Duan, Qiang Hu, Jiarui Wang, Liu Yang, Zitong Xu, Lu Liu, Xiongkuo Min, Chunlei Cai, Tianxiao Ye, Xiaoyun Zhang, et al. 2025. Finevq: Fine-grained user generated content video quality assessment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 3206–3217.

[10] Andrii Dzhoha, Katya Mirylenka, Egor Malykh, Marco-Andrea Buchmann, and Francesca Catino. 2025. Short-Form Video Recommendations with Multimodal Embeddings: Addressing Cold-Start and Bias Challenges. *arXiv preprint arXiv:2507.19346* (2025).

[11] Kaituo Feng, Kaixiong Gong, Bohao Li, Zonghao Guo, Yibing Wang, Tianshuo Peng, Junfei Wu, Xiaoying Zhang, Benyou Wang, and Xiangyu Yue. [n. d.]. Video-r1: Reinforcing video reasoning in mllms, 2025. *URL https://arxiv.org/abs/2503.21776* ([n. d.]).

[12] Kaituo Feng, Kaixiong Gong, Bohao Li, Zonghao Guo, Yibing Wang, Tianshuo Peng, Junfei Wu, Xiaoying Zhang, Benyou Wang, and Xiangyu Yue. 2025. Video-r1: Reinforcing video reasoning in mllms. *arXiv preprint arXiv:2503.21776* (2025).

[13] Qihang Ge, Wei Sun, Yu Zhang, Yunhao Li, Zhongpeng Ji, Fengyu Sun, Shangling Jui, Xiongkuo Min, and Guangtao Zhai. 2025. LMM-VQA: Advancing video quality assessment with large multimodal models. *IEEE Transactions on Circuits and Systems for Video Technology* (2025).

[14] Dong Guo, Faming Wu, Feida Zhu, Fuxing Leng, Guang Shi, Haobin Chen, Haoqi Fan, Jian Wang, Jianyu Jiang, Jiawei Wang, et al. 2025. Seed1. 5-vl technical report. *arXiv preprint arXiv:2505.07062* (2025).

[15] Chenlong He, Qi Zheng, Ruoxi Zhu, Xiaoyang Zeng, Yibo Fan, and Zhengzhong Tu. 2024. COVER: A Comprehensive Video Quality Evaluator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 5799–5809.

[16] Rongpei Hong, Jian Lang, Jin Xu, Zhangtao Cheng, Ting Zhong, and Fan Zhou. 2025. Following clues, approaching the truth: Explainable micro-video rumor detection via chain-of-thought reasoning. In *Proceedings of the ACM on Web Conference 2025*. 4684–4698.

[17] Zhangqi Jiang, Junkai Chen, Beier Zhu, Tingjin Luo, Yankun Shen, and Xu Yang. 2025. Devils in middle layers of large vision-language models: Interpreting, detecting and mitigating object hallucinations via attention lens. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 25004–25014.

[18] Harold W Kuhn. 1955. The Hungarian method for the assignment problem. *Naval research logistics quarterly* 2, 1-2 (1955), 83–97.

[19] Hosu Lee, Junho Kim, Hyunjun Kim, and Yong Man Ro. 2025. ReFoCUS: Reinforcement-guided Frame Optimization for Contextual Understanding. *arXiv preprint arXiv:2506.01274* (2025).

[20] Dingquan Li, Tingting Jiang, and Ming Jiang. 2019. Quality assessment of in-the-wild videos. In *Proceedings of the 27th ACM international conference on multimedia*. 2351–2359.

[21] Jiaze Li, Haoran Xu, Shiding Zhu, Junwei He, and Haozhao Wang. 2025. Multilevel semantic-aware model for ai-generated video quality assessment. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1–5.

[22] Xinhao Li, Ziang Yan, Desen Meng, Lu Dong, Xiangyu Zeng, Yinan He, Yali Wang, Yu Qiao, Yi Wang, and Limin Wang. 2025. Videochat-r1: Enhancing spatio-temporal perception via reinforcement fine-tuning. *arXiv preprint arXiv:2504.06958* (2025).

[23] Xinhao Li, Ziang Yan, Desen Meng, Lu Dong, Xiangyu Zeng, Yinan He, Yali Wang, Yu Qiao, Yi Wang, and Limin Wang. 2025. VideoChat-R1: Enhancing Spatio-Temporal Perception via Reinforcement Fine-Tuning. *arXiv preprint arXiv:2504.06958* (2025).

[24] Yunxin Li, Xinyu Chen, Zitao Li, Zhenyu Liu, Longyue Wang, Wenhan Luo, Baotian Hu, and Min Zhang. 2025. VerIPO: Cultivating Long Reasoning in Video-LLMs via Verifier-Gudied Iterative Policy Optimization. *arXiv preprint arXiv:2505.19000* (2025).

[25] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Krishna Moorthy, and Megha Manohara. 2016. Toward a Practical Perceptual Video Quality Metric. Netflix Tech Blog. Available at https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652.

[26] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*. 74–81.

[27] Yiting Lu, Xin Li, Bingchen Li, Zihao Yu, Fengbin Guan, Xinrui Wang, Ruling Liao, Yan Ye, and Zhibo Chen. 2024. Aigc-vqa: A holistic perception metric for aigc video quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6384–6394.

[28] Fuwen Luo, Shengfeng Lou, Chi Chen, Ziyue Wang, Chenliang Li, Weizhou Shen, Jiyue Guo, Peng Li, Ming Yan, Ji Zhang, et al. 2025. MUSEG: Reinforcing Video Temporal Understanding via Timestamp-Aware Multi-Segment Grounding. *arXiv preprint arXiv:2505.20715* (2025).

[29] Yachun Mi, Yu Li, Weicheng Meng, Chaofeng Chen, Chen Hui, and Shaohui Liu. 2025. MVQA: Mamba with Unified Sampling for Efficient Video Quality Assessment. *arXiv preprint arXiv:2504.16003* (2025).

[30] Kun Ouyang. 2025. Spatial-r1: Enhancing mllms in video spatial reasoning. *arXiv e-prints* (2025), arXiv–2504.

[31] Kun Ouyang, Yuanxin Liu, Haoning Wu, Yi Liu, Hao Zhou, Jie Zhou, Fandong Meng, and Xu Sun. 2025. SpaceR: Reinforcing MLLMs in Video Spatial Reasoning. *arXiv preprint arXiv:2504.01805* (2025).

[32] Runqi Ouyang, Haoyun Li, Zhenyuan Zhang, Xiaofeng Wang, Zheng Zhu, Guan Huang, and Xingang Wang. 2025. Motion-R1: Chain-of-Thought Reasoning and Reinforcement Learning for Human Motion Generation. *arXiv preprint arXiv:2506.10353* (2025).

[33] Jinyoung Park, Jeehye Na, Jinyoung Kim, and Hyunwoo J Kim. 2025. DeepVideo-R1: Video Reinforcement Fine-Tuning via Difficulty-aware Regressive GRPO. *arXiv preprint arXiv:2506.07464* (2025).

[34] Ashish Seth, Dinesh Manocha, and Chirag Agarwal. 2024. Towards a Systematic Evaluation of Hallucinations in Large-Vision Language Models. *arXiv preprint arXiv:2412.20622* (2024).

[35] Hamid R Sheikh and Alan C Bovik. 2006. Image information and visual quality. *IEEE Transactions on image processing* 15, 2 (2006), 430–444.

[36] Xiaoqian Shen, Yunyang Xiong, Changsheng Zhao, Lemeng Wu, Jun Chen, Chenchen Zhu, Zechun Liu, Fanyi Xiao, Balakrishnan Varadarajan, Florian Bordes, Zhuang Liu, Hu Xu, Hyunwoo J. Kim, Bilge Soran, Raghuraman Krishnamoorthi, Mohamed Elhoseiny, and Vikas Chandra. 2024. LongVU: Spatiotemporal Adaptive Compression for Long Video-Language Understanding. *arXiv:2410.17434* (2024).

[37] Shangkun Sun, Xiaoyu Liang, Bowen Qu, and Wei Gao. 2025. Content-rich aigc video quality assessment via intricate text alignment and motion-aware consistency. *arXiv preprint arXiv:2502.04076* (2025).

[38] Kwai Keye Team. 2025. Kwai Keye-VL Technical Report. arXiv:2507.01949 [cs.CV] https://arxiv.org/abs/2507.01949

[39] V Team, Wenyi Hong, Wenmeng Yu, Xiaotao Gu, Guo Wang, Guobing Gan, Haomiao Tang, Jiale Cheng, Ji Qi, Junhui Ji, Lihang Pan, Shuaiqi Duan, Weihan Wang, Yan Wang, Yean Cheng, Zehai He, Zhe Su, Zhen Yang, Ziyang Pan, Aohan Zeng, Baoxu Wang, Bin Chen, Boyan Shi, Changyu Pang, Chenhui Zhang, Da Yin, Fan Yang, Guoqing Chen, Jiazheng Xu, Jiale Zhu, Jiali Chen, Jing Chen, Jinhao Chen, Jinghao Lin, Jinjiang Wang, Junjie Chen, Leqi Lei, Letian Gong, Leyi Pan, Mingdao Liu, Mingde Xu, Mingzhi Zhang, Qinkai Zheng, Sheng Yang, Shi Zhong, Shiyu Huang, Shuyuan Zhao, Siyan Xue, Shangqin Tu, Shengbiao Meng, Tianshu Zhang, Tianwei Luo, Tianxiang Hao, Tianyu Tong, Wenkai Li, Wei Jia, Xiao Liu, Xiaohan Zhang, Xin Lyu, Xinyue Fan, Xuancheng Huang, Yanling Wang, Yadong Xue, Yanfeng Wang, Yanzi Wang, Yifan An, Yifan Du, Yiming Shi, Yiheng Huang, Yilin Niu, Yuan Wang, Yuanchang Yue, Yuchen Li, Yutao Zhang, Yuting Wang, Yu Wang, Yuxuan Zhang, Zhao Xue, Zhenyu Hou, Zhengxiao Du, Zihan Wang, Peng Zhang, Debing Liu, Bin Xu, Juanzi Li, Minlie Huang, Yuxiao Dong, and Jie Tang. 2025. GLM-4.5V and GLM-4.1V-Thinking: Towards Versatile Multimodal Reasoning with Scalable Reinforcement Learning. arXiv:2507.01006 [cs.CV] https://arxiv.org/abs/2507.01006

[40] Shulin Tian, Ruiqi Wang, Hongming Guo, Penghao Wu, Yuhao Dong, Xiuying Wang, Jingkang Yang, Hao Zhang, Hongyuan Zhu, and Ziwei Liu. 2025. Ego-R1: Chain-of-Tool-Thought for Ultra-Long Egocentric Video Reasoning. *arXiv preprint arXiv:2506.13654* (2025).

[41] Khoa-Dang Tran. 2025. Explainable Manipulated Videos Detection Using Multimodal Large Language Models. In *Companion Proceedings of the ACM on Web Conference 2025*. 725–728.

[42] Ashwin Vinod, Shrey Pandit, Aditya Vavre, and Linshen Liu. 2025. EgoVLM: Policy Optimization for Egocentric Video Understanding. *arXiv preprint arXiv:2506.03097* (2025).

[43] Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, Xingguang Wei, Zhaoyang Liu, Linglin Jing, Shenglong Ye, Jie Shao, et al. 2025. Internvl3. 5: Advancing open-source multimodal models in versatility, reasoning, and efficiency. *arXiv preprint arXiv:2508.18265* (2025).

[44] Yi Wang, Xinhao Li, Ziang Yan, Yinan He, Jiashuo Yu, Xiangyu Zeng, Chenting Wang, Changlian Ma, Haian Huang, Jianfei Gao, Min Dou, Kai Chen, Wenhai Wang, Yu Qiao, Yali Wang, and Limin Wang. 2025. InternVideo2.5: Empowering Video MLLMs with Long and Rich Context Modeling. *arXiv preprint arXiv:2501.12386* (2025).

[45] Ye Wang, Boshen Xu, Zihao Yue, Zihan Xiao, Ziheng Wang, Liang Zhang, Dingyi Yang, Wenxuan Wang, and Qin Jin. 2025. Timezero: Temporal video grounding with reasoning-guided lvlm. *arXiv e-prints* (2025), arXiv–2503.

[46] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[47] Ziyang Wang, Jaehong Yoon, Shoubin Yu, Md Mohaiminul Islam, Gedas Bertasius, and Mohit Bansal. 2025. Video-RTS: Rethinking Reinforcement Learning and Test-Time Scaling for Efficient and Enhanced Video Reasoning. *arXiv preprint arXiv:2507.06485* (2025).

[48] Diankun Wu, Fangfu Liu, Yi-Hsin Hung, and Yueqi Duan. 2025. Spatial-mllm: Boosting mllm capabilities in visual-based spatial intelligence. *arXiv preprint arXiv:2505.23747* (2025).

[49] Haoning Wu, Erli Zhang, Liang Liao, Chaofeng Chen, Jingwen Hou, Annan Wang, Wenxiu Sun, Qiong Yan, and Weisi Lin. 2023. Exploring video quality assessment on user generated contents from aesthetic and technical perspectives. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 20144–20154.

[50] Yongliang Wu, Xinting Hu, Yuyang Sun, Yizhou Zhou, Wenbo Zhu, Fengyun Rao, Bernt Schiele, and Xu Yang. 2025. Number it: Temporal grounding videos like flipping manga. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 13754–13765.

[51] Zhiyu Wu, Xiaokang Chen, Zizheng Pan, Xingchao Liu, Wen Liu, Damai Dai, Huazuo Gao, Yiyang Ma, Chengyue Wu, Bingxuan Wang, et al. 2024. Deepseek-vl2: Mixture-of-experts vision-language models for advanced multimodal understanding. *arXiv preprint arXiv:2412.10302* (2024).

[52] Ruobing Xie, Lin Ma, Shaoliang Zhang, Feng Xia, and Leyu Lin. 2023. Reweighting clicks with dwell time in recommendation. In *Companion Proceedings of the ACM Web Conference 2023*. 341–345.

[53] Cilin Yan, Haochen Wang, Shilin Yan, Xiaolong Jiang, Yao Hu, Guoliang Kang, Weidi Xie, and Efstratios Gavves. 2024. Visa: Reasoning video object segmentation via large language models. In *European Conference on Computer Vision*. Springer, 98–115.

[54] Changshuo Zhang, Zihan Lin, Shukai Liu, Yongqi Liu, and Han Li. 2025. Comment Staytime Prediction with LLM-enhanced Comment Understanding. In *Companion Proceedings of the ACM on Web Conference 2025*. 586–595.

[55] Peiyuan Zhang, Kaichen Zhang, Bo Li, Guangtao Zeng, Jingkang Yang, Yuanhan Zhang, Ziyue Wang, Haoran Tan, Chunyuan Li, and Ziwei Liu. 2024. Long context transfer from language to vision. *arXiv preprint arXiv:2406.16852* (2024).

[56] Xingjian Zhang, Siwei Wen, Wenjun Wu, and Lei Huang. 2025. Tinyllava-video-r1: Towards smaller lmms for video reasoning. *arXiv preprint arXiv:2504.09641* (2025).

[57] Yuanhan Zhang, Bo Li, haotian Liu, Yong jae Lee, Liangke Gui, Di Fu, Jiashi Feng, Ziwei Liu, and Chunyuan Li. 2024. LLaVA-NeXT: A Strong Zero-shot Video Understanding Model. https://llava-vl.github.io/blog/2024-04-30-llava-next-video/

[58] Zhichao Zhang, Wei Sun, Li Xinyue, Jun Jia, Xiongkuo Min, Zicheng Zhang, Chunyi Li, Zijian Chen, Wang Puyi, Sun Fengyu, et al. 2025. Benchmarking Multi-dimensional AIGC Video Quality Assessment: A Dataset and Unified Model. *ACM Transactions on Multimedia Computing, Communications and Applications* 21, 9 (2025), 1–24.

[59] Haiyuan Zhao, Guohao Cai, Jieming Zhu, Zhenhua Dong, Jun Xu, and Ji-Rong Wen. 2024. Counteracting duration bias in video recommendation via counterfactual watch time. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4455–4466.

[60] Haiyuan Zhao, Lei Zhang, Jun Xu, Guohao Cai, Zhenhua Dong, and Ji-Rong Wen. 2023. Uncovering user interest from biased and noised watch time in video recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 528–539.

[61] Qi Zheng, Li-Heng Chen, Chenlong He, Neil Berkbeck, Yilin Wang, Balu Adsumilli, Alan C Bovik, Yibo Fan, and Zhengzhong Tu. 2025. Subjective and Objective Quality Assessment of Banding Artifacts on Compressed Videos. *arXiv preprint arXiv:2508.08700* (2025).

[62] Liyun Zhu, Qixiang Chen, Xi Shen, and Xiaodong Cun. 2025. VAU-R1: Advancing Video Anomaly Understanding via Reinforcement Fine-Tuning. *arXiv preprint arXiv:2505.23504* (2025).

A Prompt Template

Our data construction pipeline includes a caption generation process, a CoT data sampling process, a CoT data filtering process, and a user comments filtering process. The prompts used in each stage are illustrated in Figures 4, Figures 6, Figures 5, and Figures 7, respectively. In addition, the prompt shown in Figures 6 is also used in the SFT stage of training and in the inference stage of all methods. In the GRPO training stage, we build on the prompt used in the SFT training stage, revising the output format and specifying requirements for the caption and comment outputs, as shown in Figure 8.

Prompt Template For Caption Generation Process

You are given the first five episodes of a shortdrama. Please generate a detailed and objective plot summary that accurately describes the main storyline, key characters, their relationships, motivations, and the central conflict.
- The summary should cover the background, majorevents, emotional turning points, and unresolved tensions by the end of the fifth episode.
- Focus on narrative coherence and character development rather than visual descriptions.

Figure 4: Prompt template for caption generation process.

Prompt Template For CoT Data Filtering Process

You are given three inputs:
1. The ground-truth (GT) quality score of a short drama. The GT score range from 1 to 5, where 1 represents the lowest quality and 5 represents the highest quality.
2. The predicted quality score and its Chain-of-Thought (CoT) reasoning process.
3. The video caption summarizing the drama's plot.
You are required to evaluate whether this CoT sample is valid for training.

<instructions>
Discard the sample if it meets any of the following conditions:
- The predicted quality score does not match the GT score.
- The reasoning process is inconsistent with the predicted score. i.e. the reasoning lists many strengths but few weaknesses yet assigns a low score, or conversely, lists many weaknesses but few strengths yet assigns a high score.
- The reasoning contains hallucinations or vacuous statements that are not grounded in the video caption or actual plot. You need to determine whether the mentioned plot actually exists based on the provided caption.

Your response must follow the JSON format below:
<formatting_example>
{
"reason": "Brief explanation of why the sample is valid or invalid",
"valid": true/false
}
</formatting_example>
</instructions>

Figure 5: Prompt template for CoT data filtering process.

Prompt Template For CoT Data Sampling Process, SFT Training Stage And Inference Stage

You are a professional short-drama evaluation expert. You will watch the first five episodes of a short drama and perform a comprehensive quality analysis. You are required to assign a score to the drama based on the following criteria:

<instructions>
The short drama should be rated on an integer scale from 1 to 5, where 1 represents the lowest quality and 5 represents the highest quality.
During the evaluation, ensure that the score distribution across different dramas covers the entire range (1–5).
When assessing each drama, you must reference specific scenes, dialogues, character actions, or cinematic techniques from the episodes as evidence.

You should analyze the strengths and weaknesses as follows:

Strengths:
- Opening Engagement:
  - The opening effectively captures the audience's attention
  - Contains unconventional storytelling (e.g., flashback, sudden conflict)
- Emotional Climax and Plot Twist:
  - Includes key conflict scenes with strong tension and convincing acting
  - Plot twists are both logical and unexpected
- Audience Immersion:
  - Storyline easily evokes empathy and emotional resonance
  - Acting details or set design are realistic and consistent with the plot
- Ending Suspense Design:
  - The final shot leaves unresolved tension
  - Important clues remain open-ended

Weaknesses:
- Slow-paced narrative and flat storytelling
- Abrupt plot progression lacking logical causality or proper foreshadowing; unclear character motivations and inconsistent emotional shifts
- Crying or quarrel scenes feature stiff expressions and lack emotional impact; dialogue delivery feels memorized and lacks emotional layering
- Overused sets, monotonous spaces, or cheap visual texture
- Superficial thematic expression lacking emotional resonance

Your response must follow the JSON format below:
<formatting_example>
{
  "think": "Your analysis of the strengths and weaknesses",
  "answer": "Your predicted score"
}
</formatting_example>
</instructions>

Figure 6: Prompt template for CoT data sampling process, SFT training stage and inference stage.

Prompt Template For User Comments Filtering Process

You will receive a list of short-drama comments, where each element is a string and represents one comment. You are required to perform a short-drama user-comment understanding task. Your goal is to exclude comments that are unrelated to the drama's quality from a given list of comments.

<instructions>
The following comments are considered unrelated to plot/quality; assign label 0:
1. Comments irrelevant to the drama content, e.g., "first," "check-in," "urge update," "No.1," "sign-in," etc.
2. Comments that are meaningless symbols, e.g., "NULL," "null," etc.
3. Personal associations unrelated to the plot, e.g., "I have the same dress as the heroine," "I saw this café yesterday," etc.

The following comments are considered related to drama quality; assign label 1:
1. Plot summaries or reflections, e.g., "The male lead should treat the female lead harshly because she likes it," "It takes two to tango," etc.
2. Emotional reactions to the plot, e.g., "I'm crying," "so touching," "hahaha," "LOL," etc.
3. Remarks about characters/actors, e.g., "The male lead is handsome," "I hate the villain," etc.
4. Short but plot-related remarks, e.g., "surprising ending," "great plot twist," etc.

Now begin a detailed understanding for each comment:
<per_comment_reasoning>
Carefully analyze each comment's meaning and determine whether it relates to the drama's quality.
</per_comment_reasoning>

After completing the per-comment reasoning, filter out comments unrelated to the short drama's quality.

<structured_information>
Provide the final structured output. For each item, if it is related to drama quality, assign label 1 (integer); if it is unrelated, assign label 0 (integer). The output must be JSON.
</structured_information>

Strictly follow the steps above and output only the result of <structured_information> in JSON format. Keep every item. Use the comment text itself as the JSON key; no numbering.

Your response must follow the JSON format below:
<formatting_example>
{
  "Comment 1 content": {"label": X},
  "Comment 2 content": {"label": X}
  ...
}
</formatting_example>
</instructions>

Figure 7: Prompt template for user comments filtering process.

Prompt Template For User Comments Filtering Process

You are a professional short-drama evaluation expert. You will watch the first five episodes of a short drama, summarize the plot, output 10 possible user comments, and perform a comprehensive quality analysis. You are required to assign a score to the drama based on the following criteria:

<instructions>
Please strictly follow the steps below to complete the task:

<plot summary>
Carefully review the provided images and text, and summarize the short drama's core storyline, character relationships, background/setting, primary conflicts, and overall emotional tone.
</plot summary>

<comments generation>
Based on the image–text content, generate comments that viewers might leave after watching. You may assess strengths and weaknesses from the following perspectives:
- Whether the plot is engaging or conflict-driven;
- Whether the plot development is reasonable;
- Whether the story evokes emotional resonance in the audience;
- Whether the narrative is novel and contains unexpected ideas;
- The actors' presentation, including whether appearance fits the roles and whether acting aligns with the plot;
- Whether the visuals possess aesthetic value or expressive power;
- Whether the dialogue is authentic and impactful;

Output 10 comments, listed as bullet points in a single list.
</comments generation>

<quality analysis>
The short drama should be rated on an integer scale from 1 to 5, where 1 represents the lowest quality and 5 represents the highest quality.
During the evaluation, ensure that the score distribution across different dramas covers the entire range (1–5).
When assessing each drama, you must reference specific scenes, dialogues, character actions, or cinematic techniques from the episodes as evidence.

You should analyze the strengths and weaknesses as follows:

Strengths:
- Opening Engagement:
  - The opening effectively captures the audience's attention
  - Contains unconventional storytelling (e.g., flashback, sudden conflict)
- Emotional Climax and Plot Twist:
  - Includes key conflict scenes with strong tension and convincing acting
  - Plot twists are both logical and unexpected
- Audience Immersion:
  - Storyline easily evokes empathy and emotional resonance
  - Acting details or set design are realistic and consistent with the plot
- Ending Suspense Design:
  - The final shot leaves unresolved tension
  - Important clues remain open-ended

Weaknesses:
- Slow-paced narrative and flat storytelling
- Abrupt plot progression lacking logical causality or proper foreshadowing; unclear character motivations and inconsistent emotional shifts
- Crying or quarrel scenes feature stiff expressions and lack emotional impact; dialogue delivery feels memorized and lacks emotional layering
- Overused sets, monotonous spaces, or cheap visual texture
- Superficial thematic expression lacking emotional resonance

</quality analysis>

<quality scoring>
Based on your previous analysis, assign a score to the drama from the perspective of audience. The score must be an integer between 1 and 5.
<quality scoring>

<output_requirements>
Please strictly follow the format below (four parts in total):
<caption>
Plot summary of the video.
</caption>
<comments>
["Comment 1", "Comment 2", ..., "Comment 10"]
</comments>
<think>
Quality analysis from the perspective of audience.
</think>
<answer>
\boxed{Quality score (integer 1–5)}
</answer>
</output_requirements>
</instructions>

Figure 8: Prompt template for GRPO training stage.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009