When Top-ranked Recommendations Fail: Modeling Multi-Granular Negative Feedback for Explainable and Robust Video Recommendation

Siran Chen[1,2,3*], Boyu Chen[1,2,3*], Chenyun Yu[4†], Yi Ouyang [3], Lei Cheng [3], Chengxiang Zhuo [3], Zang Li [3], Yali Wang[1,5†]

[1] Shenzhen Key Laboratory of Computer Vision and Pattern Recognition, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[2] University of Chinese Academy of Science, Beijing, China
[3] Tencent, Shenzhen, China
[4] Shenzhen Campus of Sun Yat-sen University, Shenzhen, China
[5] Shanghai Artificial Intelligence Laboratory, Shanghai, China
chensiran17, chenboyu18@mails.ucas.ac.cn, yuchy35@mail.sysu.edu.cn, yl.wang@siat.ac.cn

Abstract

Existing video recommendation systems, relying mainly on ID-based embedding mapping and collaborative filtering, often fail to capture in-depth video content semantics. Moreover, most struggle to address biased user behaviors (e.g., accidental clicks, fast skips), leading to inaccurate interest modeling and frequent negative feedback in top recommendations with unclear causes. To tackle this issue, we collect real-world user video-watching sequences, annotate the reasons for users' dislikes, and construct a benchmark dataset for personalized explanations. We then introduce the Agentic Explainable Negative Feedback (ENF) framework, which integrates three core components: (1) the Profile Agent, extracting behavioral cues from users' historical data to derive psychological and personality profiles; (2) the Video Agent, performing comprehensive multimodal video analysis; and (3) the Reason Agent, synthesizing information from the other two agents to predict user engagement and generate explanations. Additionally, we propose the S-GRPO algorithm, enabling the model to progressively address complex tasks during reinforcement fine-tuning. Experimental results on the collected dataset show that our method significantly outperforms state-of-the-art baselines in negative feedback prediction and reason explanation. Notably, it achieves an 8.6% improvement over GPT-4o in reason classification. Deployment on the business platform further validates its benefits: increasing average user watch time by 6.2%, reducing the fast-skip rate by 9.4% , and significantly enhancing user satisfaction.

Introduction

With the explosive growth of online multimodal content, short video platforms such as TikTok and Kuaishou have become primary channels for people to entertain themselves, shop, and access key information, exerting a significant impact on our daily life. Traditional recommendation methods (Huang et al. 2015; Zheng et al. 2018; Ying et al. 2018; Yang et al. 2020; Yuan et al. 2020), like collaborative filtering and multimodal methods, have been widely adopted for their simplicity and scalability. However, existing systems inevitably encounter negative user feedback, which involves both explicit forms (*e.g.*, dislikes) and implicit ones (*e.g.*, fast skips). Addressing such negative feedback is crucial for recommendation systems, as it directly reflects users' dissatisfaction and potential misalignment between recommended content and actual preferences. Therefore, we raise a fundamental question: *Why do some top-ranked recommendations consistently trigger negative user feedback?*

As a key topic in recommendation systems, research on user negative feedback faces three core challenges. **First, the scarcity of negative feedback data limits the availability of high-quality datasets for in-depth analysis**. While explicit signals (*e.g.*, dislikes, comments) are highly informative, they are extremely sparse (accounting for approximately 0.3% of all interactions). In contrast, implicit feedback *(e.g.*, watch time, skips) is abundant but low-informative and noisy, making the effective utilization challenging. **Second, the specific reasons behind users' negative feedback are highly unaddressed**. Prior methods typically derive dislike-related features by clustering diverse negative feedback signals, then use these features to suppress similar recommendations (Xie et al. 2021; Wang et al. 2023b; Lai et al. 2025), without understanding the specific reasons behind dislike, these methods may lead to poor generalization across different scenarios. For example, in Fig. 1, if a user dislikes a food-related video, it would be inappropriate to suppress all food-sharing recommendations without understanding the actual cause. **Third, existing LLM-based methods for negative feedback lack detailed evaluation in multimodal scenarios**. Although several LLM-based methods (Bao et al. 2023; Ye et al. 2025; Zhang et al. 2025) can predict user preferences, they largely overlook the complex multimodal content of items. Moreover, the absence of evaluation regarding explainable reasons significantly undermines their credibility.

To address these gaps, we first construct **TVNF**, a multimodal video recommendation dataset containing diverse negative feedback from the business scenario (*i.e.*, Tencent News). It includes basic user profiles and multi-grained interaction data, such as watch time, dislikes, and actual

*Equal contribution.
†Corresponding author.

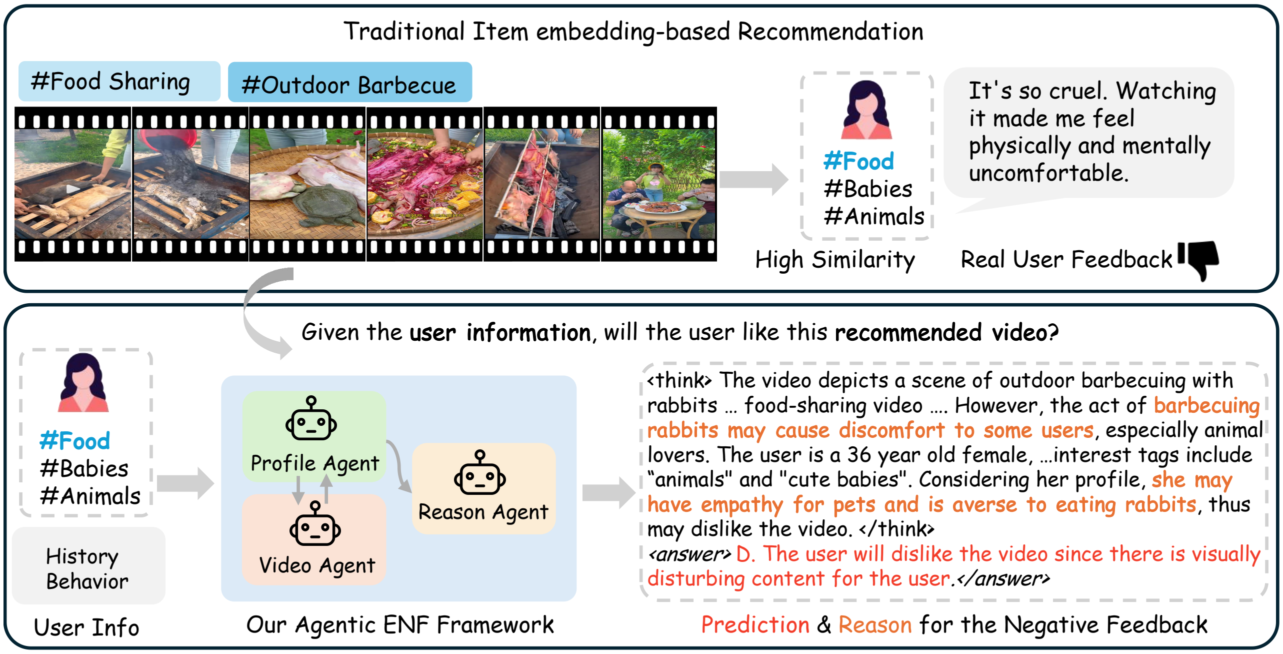

Figure 1: **User negative feedback in a real recommendation scenario.** Traditional method recommends a food-sharing video to a user who loves food-related themes based on high embedding similarity, while this triggers strong negative feedback. Our ENF framework successfully predicts the reason for user's negative feedback, avoiding similar recommendations in the future.

feedback reasons, enabling comprehensive analysis of user behaviors. Based on specific user feedback contents, we categorize dislike reasons into following types: negative events, vulgar or conflicting values, boring plots, and visually disturbing elements. Further, we propose an MLLM-based agentic framework designed to simulate personalized user perspectives, aiming to understand videos and generate **E**xplainable diagnostics for **N**egative **F**eedback (**ENF**). Specifically, the ENF framework comprises three hierarchically structured agents: (1) The **Profile Agent** constructs dynamically updated user profiles by analyzing demographic data and historical viewing patterns, generating additional psychographic features (*e.g.*,"passion for fantastic plots"). (2) The **Video Agent** leverages the multimodal capabilities of MLLMs to decompose video content, providing content descriptions and value analysis to the Profile Agent for cross-modal validation. (3) The **Reason Agent** evaluates videos from the user's perspective using the updated profile from the Profile Agent, ultimately predicting user preference likelihood and generating interpretable explanations. In addition, to ensure the ENF pipeline to simulate complex human behaviors, we propose **S-GRPO**, a reinforcement learning paradigm which employs a stepwise reward mechanism to address challenging tasks in a progressive manner. Specifically, it includes three sequential rewards: a binary judgment reward, a multi-choice selection reward, and an interpretability-oriented reward. Notably, the latter reward is only triggered when the output of the previous step is correct, which stands in contrast to earlier GRPO tasks (Liu et al. 2025; Li et al. 2025) that rely on a single reward signal per iteration. Through this stepwise reward design, the model can incrementally learn and infer human viewing patterns, thereby achieving more accurate predictions.

Experimental results validate the effectiveness of our framework: compared with MLLMs and state-of-the-art methods that directly utilize reinforcement learning, our approach achieves higher accuracy in both fast-skip prediction and explainable reason classification, effectively bridging the gap between system rankings and actual user preferences. Our contributions can be summarized as follows:

- To the best of our knowledge, we are the first to identify, explain, and evaluate implicit negative feedback using LLMs. We construct TVNF, a practical multimodal video recommendation benchmark with multi-granularity explicit and implicit feedback. It integrates multimodal content, annotated explainable reasons, and labels, enabling evaluation for explainable negative feedback.
- We propose the Agentic ENF framework, which leverages collaborative MLLM-based agents to effectively simulate user behaviors. In addition, we introduce the S-GRPO training strategy, a progressive reinforcement learning paradigm that ensures explainable and personalized video recommendations.
- Extensive experiments demonstrated the improvements by our ENF method in negative feedback prediction and reason explanation. Additionally, we evaluate ENF in real-world recommendation scenario and observe 6.2% improvement in average play rate and 9.4% decline for fast-skip rate, significantly enhancing user satisfaction.

Related Work

LLM as User Simulator. Considering the powerful semantic understanding and reasoning capabilities of LLMs, numerous studies have leveraged them to facilitate user inference simulations (Ma et al. 2022; Wang et al. 2023a; Zhang et al. 2024b; Corecco et al. 2024; Zhang et al. 2024a; Feng et al. 2025b). For instance, iEvaLM (Wang et al. 2023a) explores two interaction types within a conversational recommendation benchmark: attribute-based question answering and free-form chit-chat using ChatGPT (Achiam et al. 2023). To simulate user search behavior, USimAgent (Zhang et al. 2024b) prompts an LLM-based agent to construct complete search sessions, including querying, clicking, and stopping behaviors, according to specific search tasks. Agent4Rec (Zhang et al. 2024a) initializes LLMs as agents with unique user profiles that encompass tastes and social traits to simulate more realistic user behaviors. Additionally, LLM_Simulator (Zhang et al. 2025) simulates user preferences by matching the positive and negative attributes of items with LLM-generated user preferences to determine whether a user would like an item. However, prior LLM-based user simulation approaches have relied on frozen LLMs, and using them solely through prompting would risk discrepancies with real user behavior and potential hallucinations (Zhang et al. 2024a).

Implicit Negative User Feedback. Implicit feedbacks are ubiquitously generated during user browsing behavior, such as video watch rates and impression click-through rates. They are noisy with subtle signals but still beneficial for recommendation systems (Frolov and Oseledets 2016; Lai et al. 2024; Seo et al. 2022; Cena, Console, and Vernero 2023; Yang et al. 2025b), how to effectively leverage such pervasive yet inconspicuous implicit feedback to enhance user understanding persists a critical challenge. Some works (Pan et al. 2016) aim to establish relationships between explicit and implicit feedback through feature mapping or transfer learning. DFN (Xie et al. 2021) utilizes both internal and external feedback interactions to learn users' unbiased preferences for Click-Through Rate prediction. CDR (Chen et al. 2021) employs users' explicit dislike signals as a metric to evaluate the relative significance of different behavioral sequences. SINE (Pan et al. 2023) models passive-negative feedback as a mismatch of specific sub-interests. However, all previous methods focus solely on item relevance rather than exploring the specific causes of negative feedback, making it difficult to generalize to new items.

Method

Dataset Construction

We constructed a dataset, named TVNF, from real-world multi-modal video recommendation scenarios on Tencent News. It comprises approximately 10,000 users, 20,000 videos and 320,000 interaction behaviors over seven consecutive days. For each user, we gathered basic demographic information (age, gender, occupation), interest tags, and detailed viewing behaviors, including video titles, durations and watch times. All personally identifiable information was anonymized for privacy protection. To ensure data quality, we applied a filtering criterion to ensure that users have at least 15 recorded viewing instances. Each of the 20,000 unique videos is accompanied by its original URL and 16 uniformly sampled frames for visual content analysis. Additionally, a distinctive advantage of our interpretable benchmark is that we have collected approximately 1k specific instances of users' negative feedback (such data is extremely scarce and difficult to collect in practice). In these instances, users explicitly state their reasons for disliking specific videos, which serve as important references for understanding preference mismatches. By analyzing real user comments, we categorize the reasons for negative feedback into the following types: i) the video contains negative events, vulgar content, or conflicting values for the user; ii) the video content lacks sufficient appel and fails to arouse the user's interest; iii) the video contains disturbing visual elements that cause discomfort to the user. For the remaining vast volume of data, we treat cases where the user's viewing rate is below 0.3 as implicit negative feedback. We instruct GPT-4o (Hurst et al. 2024) to label each implicit negative feedback instance according to the aforementioned categories, followed by manual verification to ensure more reliable classification. Content comparison with previous datasets (Hou et al. 2024; Asghar 2016; Xie et al. 2021; Gao et al. 2022; Harper and Konstan 2015; Ni et al. 2023) are presented in Tab. 1.

Dataset	Multimodal Video Data	Explicit Feedback	Implicit Feedback	Real User Dislike Reason
Amazon	✗	✓	✗	✓
Yelp	✗	✓	✗	✓
MultiFeed	✗	✓	✓	✗
KuaiRand	✗	✓	✓	✗
MovieLens	✗	✓	✗	✗
MircoLens	✓	✗	✗	✗
TVNF	✓	✓	✓	✓

Table 1: **Content comparison with previous datasets.**

Our Agent-based ENF Pipeline

In contrast to objective questions with standard answers, user behaviors are highly subjective and individual, rendering the simulation of specific user behaviors complex. Consequently, we propose a multigrained agent-based framework to collaboratively perform prediction and causal analysis of user behaviors, as shown in Fig. 2.

First, the Profile Agent infers users' psychological profiles from their behavioral patterns, with the objective of addressing key limitation of traditional recommendation systems that rely solely on interest tag embeddings. Traditional systems, for instance, often neglect nuanced psychological tendencies of users: a celebrity fan may strongly prefer positive content about their idol while rejecting critical narratives; a food enthusiast might react negatively to videos with overly graphic depictions of ingredient preparation. These examples highlight that user preferences extend beyond surface-level interest tags, being rooted instead

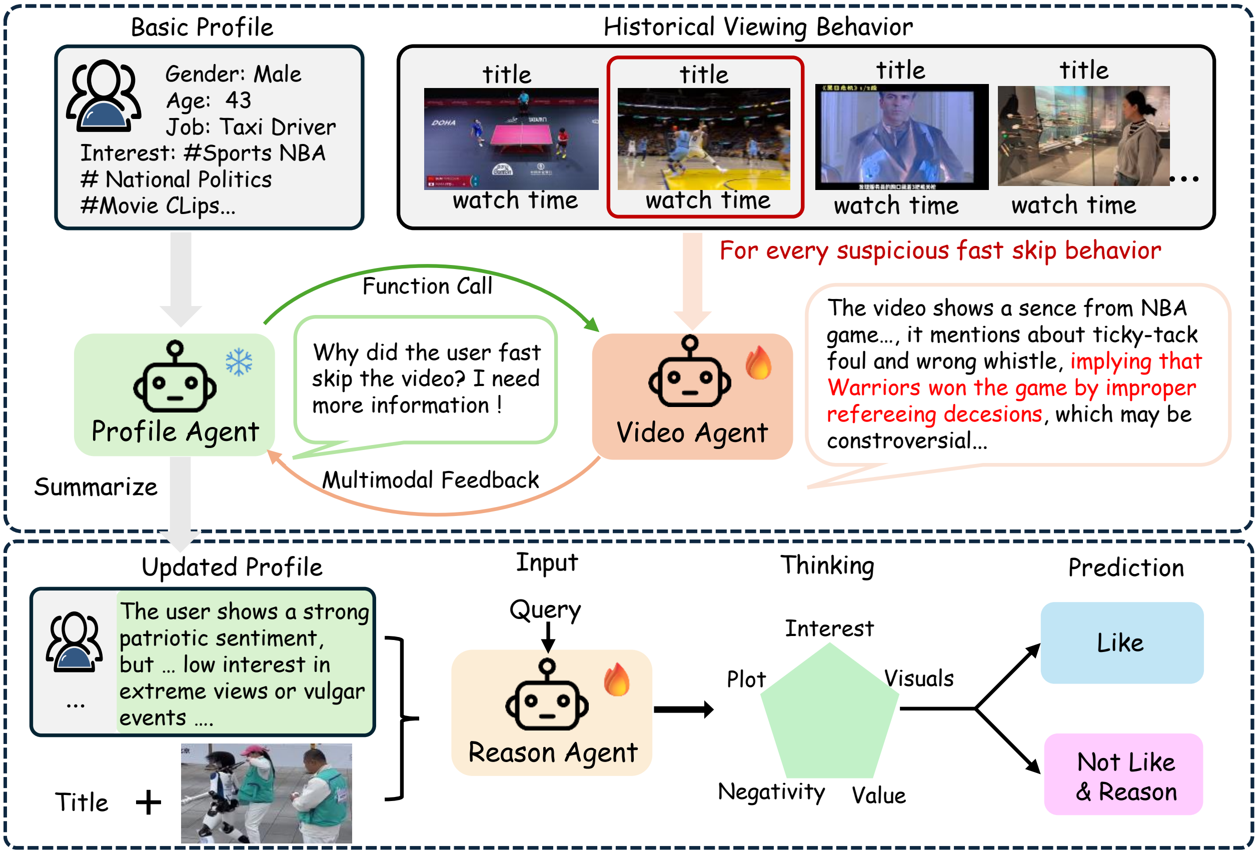

Figure 2: **Overview of our Agent-based ENF framework.** The three agents collaborate together, the Profile Agent analyze the user behaviors to get more comprehensive profile, the Video Agent aids in providing multimodal insights, and the Reason Agent uses the updated profile to predict whether a user likes the recommended video and provides explainable reasons.

in deeper psychological traits. Thus, the core goal of the Profile Agent is to uncover these latent psychological traits, enabling more nuanced and user-aligned recommendations. To achieve this, the agent leverages basic user profile information (age, gender, occupation, and recent interests) to analyze users' sequential watching behaviors, including titles and play rates. Here, the agent focuses on videos with a play rate below 0.3, which indicate user dissatisfaction with recommendations. When textual titles alone provide insufficient clues, the Profile Agent dynamically activates the Video Agent to extract multimodal clues, enhancing the depth of this analysis.

The Video Agent then conducts in-depth analysis at the individual video level. Beyond basic content description, it identifies potentially controversial elements within the video and provides contextual explanations. Concurrently, through detailed analysis of each interaction, the Profile Agent iteratively identifies factors causing users to skip content and updates psychological profiles (*e.g.*, value orientations and tolerance for negativity toward such videos). Finally, the Reason Agent leverages basic user information and refined psychological profiles to generate user-centric video understanding. This analysis process encompasses four key dimensions: whether the video's content aligns with the user's interest; whether the plot is appealing; whether the content contains negative events or extreme opinions; and whether the visual elements align with the user's sensory tolerance. By assessing these factors, the Reason Agent infers the user's attitude toward the video.

Progressive Training Strategy for Agents

We adopt Qwen2.5-VL-7B (Yang et al. 2024b) as the foundational MLLMs and follow the two-stage training framework of DeepSeek-R1 (Guo et al. 2025): supervised fine-tuning (SFT) for cold start, followed by reinforcement fine-tuning (RFT). Specifically, in Stage 1, leveraging reasons of real user feedback, we prompt GPT-4o to generate a chain-of-thought (CoT) reasoning process explaining why users disliked specific videos. These generated CoTs of the reasons are used as SFT data to warm up the model. In Stage 2, we further train the agent on the non-annotated data using our proposed S-GRPO algorithm. Our task definition involves a hierarchical prediction framework consisting of three progressive stages: first, binary judgment to determine whether the user generates negative feedback; second, multi-choice classification of pre-defined negative feedback types; and third, generating reasons to explain the core cause of the negative feedback. Unlike prior methods (Feng et al. 2025a;

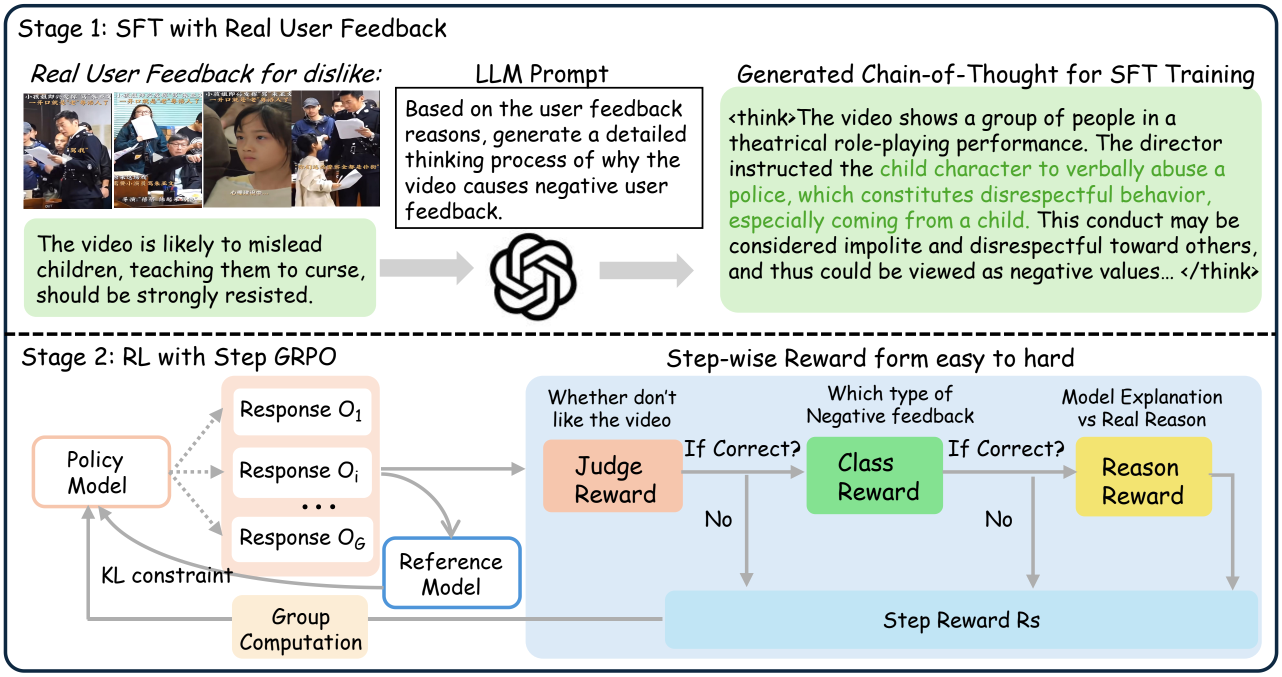

Figure 3: **Training process of our agents.** In the first stage, we use real user feedback reasons for cold start; and in the second stage, we propose a progressive reward mechanism that provides step rewards for a response in an order from easy to hard.

Li et al. 2025; Wang et al. 2025b) that only involve a single objective question for a video, our method introduces multiple granularity judgments, which makes it challenging to evaluate the response using a single reward.

To address this issue, we propose an effective multiple-choice question paired with Step Group Relative Policy Optimization (S-GRPO). Except the basic format reward, this design incorporates a progressive reward mechanism that offers three step rewards R_{S_i} for a response o_i, as depicted in Figure 3. In the binary Judge Reward r_{judge} at the first step, we determine whether the prediction of the user's attitude is correct. If the judgment is wrong, the process terminates immediately. Otherwise the response will receive a fixed reward (*e.g.*, 0.5), and if the real user feedback is positive, the process also terminates, and if the feedback is negative, the process proceeds to the second step, i.e., the Class Reward r_{class}. Here, if the choice of the negative feedback type is accurate, an additional reward (*e.g.*, 1.0) is granted, and the process advances to the third step: the Reason Reward r_{reason}. At this stage, we calculate the average of the ROUGE-1, ROUGE-2, and ROUGE-L scores between the reasoning content within the $<$*think*$>$ tag and the actual user feedback reasons, which is then used as an extra reward. Note that the Video Agent is trained using 3-step rewards on explicit negative feedback data, while the Reason Agent is trained with 2-step rewards on implicit data due to the lack of ground truth reasons. This progressive design encourages the model to tackle problems from easy to hard. For instance, it allows the model to earn rewards even when the multiple-choice answer is wrong but the binary judgment is correct; meanwhile, correct classifications accompanied by sound reasoning processes are assigned higher scores. Additionally, the advantage of A_i of response o_i among G responses is computed based on the rewards within each group. Building on this framework, the model gradually learns to classify and explain the underlying reasons, with the final policy updated to maximize the objective as follows:

$$A_i = \frac{R_i - \text{mean}(\{R_j\})}{\text{std}(\{R_j\})} \tag{1}$$

$$\mathcal{J}_{GRPO}(\theta) = \min\left(\frac{\pi_\theta(o_i|q)}{\pi_{\theta_{old}}(o_i|q)}A_i, \text{clip}\left(\frac{\pi_\theta(o_i|q)}{\pi_{\theta_{old}}(o_i|q)}, 1-\varepsilon, 1+\varepsilon\right)A_i\right) - \beta\mathcal{D}_{KL}(\pi_\theta||\pi_{ref}), \tag{2}$$

Experiments

Implementation Details and Metrics

We conducted experimental validation on two granularities of negative feedback using the collected dataset: explicit negative feedback with real users' dislike reasons, and implicit negative feedback derived from users' fast-skip data. For the explicit data, negative feedback videos are mixed with normal videos, and the Video Agent is used to predict whether a video contains controversial content that may trigger negative feedback, as well as providing explanations. For the implicit data, following the sequence recommendation setting, the Reason Agent predicts whether a specific user would generate negative feedback for a recommended video based on their historical behaviors. In our implementation, GPT-4o serves as the Profile Agent, while Qwen-2.5VL-7b works as both the Video Agent and Reason Agent. We adopt full-parameter fine-tuning: the Video

Model	Size	Binary Judgment				Explain	
		Acc	**Precision**	**Recall**	**F1_Score**	**Class_Acc**	**Reasoning**
GPT-4o (Hurst et al. 2024)	-	**0.882**	0.895	0.630	0.739	0.568	0.402
DeepSeek (Guo et al. 2025)	-	0.849	**0.917**	0.440	0.594	0.352	0.266
LLaMa (Grattafiori et al. 2024)	70b	0.822	0.900	0.346	0.500	0.268	0.177
Qwen2.5VL (Yang et al. 2024a)	7b	0.815	0.816	0.423	0.564	0.296	0.229
Qwen3 (Yang et al. 2025a)	32b	0.845	0.833	0.464	0.596	0.346	0.238
Video-R1 (Feng et al. 2025a)	7b	0.835	0.789	0.540	0.667	0.432	0.318
VideoChat-R1 (Li et al. 2025)	7b	0.842	0.850	0.654	0.739	0.500	0.383
Our Video Agent	7b	0.861	0.706	**0.808**	**0.750**	**0.654**	**0.537**

Table 2: **Prediction of explicit user negative feedback: understanding and reasoning about video controversy content.** Recall serves as a key metric for the recognition rate of negative feedback videos. The reasoning score is calculated using real user feedback reasons only when the judgment is correct.

Model	Size	Binary Judgment				Explain
		Acc	**Precision**	**Recall**	**F1_Score**	**Class_Acc**
GPT-4o (Hurst et al. 2024)	-	0.575	0.396	**0.796**	0.521	0.502
DeepSeek (Guo et al. 2025)	-	0.608	0.331	0.626	0.433	0.476
LLaMa (Grattafiori et al. 2024)	70b	0.601	0.303	0.233	0.264	0.155
Qwen2.5VL (Yang et al. 2024a)	7b	0.528	0.359	0.733	0.482	0.435
Qwen3 (Yang et al. 2025a)	32b	0.548	0.368	0.708	0.484	0.425
SASRec (Kang and McAuley 2018)	-	0.448	0.230	0.358	0.279	-
MLLM-MSR (Ye et al. 2025)	7b	0.545	0.355	0.632	0.455	-
Video-R1 (Feng et al. 2025a)	7b	0.573	0.390	0.748	0.513	0.494
VideoChat-R1 (Li et al. 2025)	7b	0.561	0.384	0.775	0.516	0.512
Our ENF	7b	**0.612**	**0.404**	0.782	**0.533**	**0.543**

Table 3: **Prediction of implicit user negative feedback: simulating user video-watching behavior.** Despite the greater difficulty of implicit behavior prediction due to significant noise, our method still achieves the best results.

Agent and Reason Agent are trained with 4 80G GPUs using the constructed dataset, which contains approximately 2,000 instances for training and evaluation. We use 16 images and the video title as the video feature inputs, the group size G is set to 8 and the learning rate is 1e-6. Evaluation metrics include binary accuracy, precision, recall, and F1-score for negative feedback prediction, as well as the accuracy of feedback reason classification. For explicit negative feedback, we employ GPT-4o to assess the relevance between the model's explanations and users' real feedback reasons, with relevance scores ranging from 0 to 1.

Main Results

We next present predictions for both users' explicit and implicit negative feedback. Explicit feedback is typically triggered by overtly controversial contents in videos, which could not be recognized by traditional embedding-based methods. Therefore, we employ the Video Agent to analyze concrete multimodal content directly. These negative videos are mixed with randomly selected normal videos at an approximate ratio of 1:4, and the Video Agent is tasked with predicting whether a video is potentially controversial, along with providing explanations. The results are shown in Tab. 2. In terms of video content understanding, while GPT-4o achieves the highest prediction accuracy, existing models generally lack sensitivity to controversial content, leading to low recall rates for negative feedback videos. In contrast, our method, leveraging the S-GRPO approach that learns from easy to difficult tasks, attains the highest Recall (0.808) and F1 Score (0.750). Although the increased recall slightly reduces precision, the method also delivers the most accurate results in both reason classification and explanation (with +8.6% and +13.5% improvements over GPT-4o, respectively). This capability is critical for identifying problematic videos and mitigating poor user experiences.

In terms of implicit negative feedback prediction, we adopt a personalized user perspective. By analyzing user profiles and historical behaviors, we infer their psychological preferences and employ the Reason Agent to predict users' attitudes toward videos (*e.g.*, whether they will fast-skip). As shown in Tab. 3, predicting implicit feedback is far more challenging than explicit feedback: the highest accuracy is only 61.2%, and precision rates are generally low. This may stem from real-world user behaviors being influenced by multiple factors with significant noises, resulting in inherent randomness. We also evaluated traditional methods such as SASRec (Kang and McAuley 2018), however, these methods exhibit poor discriminative performance in cold-start scenarios that require fine-grained item differentiation. In addition, among LLM-based methods, GPT-4o achieves the highest recall rate, but directly applying such large models in a zero-shot setting will yield unsatisfactory

SFT	RL	S-GRPO	Acc	F1_Score	Class_Acc	Reasoning
✗	✗	✗	0.815	0.423	0.296	0.229
✗	✓	✓	0.830	0.686	0.592	0.492
✓	✗	✗	0.851	0.615	0.346	0.312
✓	✓	✗	0.845	0.667	0.412	0.339
✓	✓	✓	**0.861**	**0.750**	**0.654**	**0.537**

Table 4: **Ablation**. Training Processes on Video Agent.

Profile Agent	Video Agent	S-GRPO	Acc	F1_Score	Class_Acc
✗	✗	✗	0.528	0.482	0.435
✗	✓	✓	0.596	0.518	0.508
✓	✗	✓	0.573	0.513	0.504
✓	✓	✗	0.535	0.488	0.522
✓	✓	✓	**0.612**	**0.533**	**0.543**

Table 5: **Ablation**. Training Processes on Reason Agent.

performance. For the training of smaller 7b models, our ENF framework outperforms previous video-reasoning methods in both fast-skip prediction and reason classification, validating its effectiveness. This ability to attribute reasons for implicit negative feedback significantly enhances user intent understanding, not only aiding in identifying issues and improving existing recommendation systems, but also advancing next-generation explainable recommendations.

Ablations

In this section,we conduct an ablation study by designing different variants of our models. For the Video Agent, we ablate the training of our generated CoT process as SFT data, the RL process and our proposed S-GRPO reward. As shown in Tab. 4, all ablated models perform worse than the full Video Agent across all benchmarks. Specifically, the SFT process enables the model to acquire user-side prior knowledge for cold-start, removing this component leads to a noticeable performance drop, particularly in prediction accuracy. The RL process further encourages the model to think deeply and diversely, thereby enhancing overall performance. Moreover, without our progressive S-GRPO training, the model tends to focus solely on binary judgment and struggles to distinguish between classification tasks and the underlying reasons for controversial factors.

For the Reason Agent, we ablate three designs: the Profile Agent, the initialization with the Video Agent, and the S-GRPO mechanism. As presented in Tab. 5, the Profile Agent provides richer psychographic features for more comprehensive user modeling; the Video Agent offers empirical priors learned from explicit user negative feedback, which aids in predicting implicit negative feedback; and S-GRPO ensures performance balance between binary judgment and classification-reasoning tasks. All modules collectively contribute to the improvement of the final performance.

Method	MovieLens Acc	MovieLens F1_Score	Steam Acc	Steam F1_Score
GPT-4o	0.584	0.600	0.634	0.662
RecAgent	0.581	0.621	0.627	0.650
Agent4Rec	0.691	0.698	0.689	0.679
SimUSER	0.791	0.777	0.791	0.794
Ours	**0.815**	**0.808**	**0.803**	**0.805**

Table 6: User preference alignment across MovieLens and Steam datasets.

Method	**Avg_Time↑**	**Fast-skip Rate↓**	**Dislike Rate↓**
Base RS	47.6%	23.7%	0.61%
Base RS + ENF	53.8%	14.3%	0.35%

Table 7: Performance evaluation in the real-world scenario.

Generalization on other Datasets. To evaluate the performance of our method in other domains, we further conduct evaluation on MovieLens (Harper and Konstan 2015) and Steam (Kang and McAuley 2018) following Agent4Rec (Zhang et al. 2024a) to simulate users' preferences toward items. Tab. 6 shows that, previous methods (Bougie and Watanabe 2025; Wang et al. 2025a) only prompt frozen LLMs for the prediction, which heavily rely on LLM's pre-training performance, and also suffer from inherent hallucinations. In contrast, our method achieves higher prediction accuracy through RL-based alignment with user preferences.

Testing on Business Platform. We further evaluate ENF in real-world video recommendation scenarios on Tencent News. For selected users, we use their behavioral data from previous days as the reference and assess performance over the subsequent few days. The original recommendation system (labeled Base RS) generates candidate videos, while our ENF framework predicts user attitudes towards these videos and filters out those likely to trigger negative feedback. Evaluation metrics include the average watch time, fast-skip rate, and dislike rate of recommended videos. As observed, our method yields significant improvements across these three metrics, with 13.0%, 39.7% and 42.6% improvements over the baseline, respectively. This validates the effectiveness of our approach in enhancing user satisfaction.

Conclusion

In this paper, we conduct research on users' negative feedback in recommendation systems. We first construct a benchmark dataset with real reasons for negative feedback, and propose the ENF Agentic framework with three hierarchically structured agents for multimodal behavior analysis and interpretable predictions. Leveraging the S-GRPO training paradigm, we improve prediction accuracy and achieve reliable explanations to finally improve recommendations. As a pioneering effort, this work offers novel insights into negative feedback and advances next-generation explainable recommendations, and we hope to inspire more exploration.

Acknowledgements

This work was supported by the National Key R&D Program of China(NO.2022ZD0160505).

References

Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F. L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Asghar, N. 2016. Yelp dataset challenge: Review rating prediction. *arXiv preprint arXiv:1605.05362*.

Bao, K.; Zhang, J.; Zhang, Y.; Wang, W.; Feng, F.; and He, X. 2023. Tallrec: An effective and efficient tuning framework to align large language model with recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, 1007–1014.

Bougie, N.; and Watanabe, N. 2025. Simuser: Simulating user behavior with large language models for recommender system evaluation. *arXiv preprint arXiv:2504.12722*.

Cena, F.; Console, L.; and Vernero, F. 2023. How to deal with negative preferences in recommender systems: a theoretical framework. *Journal of Intelligent Information Systems*, 60(1): 23–47.

Chen, H.; Chen, Y.; Wang, X.; Xie, R.; Wang, R.; Xia, F.; and Zhu, W. 2021. Curriculum disentangled recommendation with noisy multi-feedback. *Advances in Neural Information Processing Systems*, 34: 26924–26936.

Corecco, N.; Piatti, G.; Lanzendörfer, L. A.; Fan, F. X.; and Wattenhofer, R. 2024. SUBER: An RL Environment with Simulated Human Behavior for Recommender Systems. *arXiv preprint arXiv:2406.01631*.

Feng, K.; Gong, K.; Li, B.; Guo, Z.; Wang, Y.; Peng, T.; Wu, J.; Zhang, X.; Wang, B.; and Yue, X. 2025a. Video-r1: Reinforcing video reasoning in mllms. *arXiv preprint arXiv:2503.21776*.

Feng, K.; Ma, Y.; Zhang, X.; Liu, B.; Yuluo, Y.; Zhang, Y.; Liu, R.; Liu, H.; Qin, Z.; Mo, S.; et al. 2025b. Follow-Your-Instruction: A Comprehensive MLLM Agent for World Data Synthesis. *arXiv preprint arXiv:2508.05580*.

Frolov, E.; and Oseledets, I. 2016. Fifty shades of ratings: how to benefit from a negative feedback in top-N recommendations tasks. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 91–98.

Gao, C.; Li, S.; Zhang, Y.; Chen, J.; Li, B.; Lei, W.; Jiang, P.; and He, X. 2022. Kuairand: An unbiased sequential recommendation dataset with randomly exposed videos. In *Proceedings of the 31st ACM international conference on information & knowledge management*, 3953–3957.

Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Harper, F. M.; and Konstan, J. A. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4): 1–19.

Hou, Y.; Li, J.; He, Z.; Yan, A.; Chen, X.; and McAuley, J. 2024. Bridging language and items for retrieval and recommendation. *arXiv preprint arXiv:2403.03952*.

Huang, Y.; Cui, B.; Zhang, W.; Jiang, J.; and Xu, Y. 2015. Tencentrec: Real-time stream recommendation in practice. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, 227–238.

Hurst, A.; Lerer, A.; Goucher, A. P.; Perelman, A.; Ramesh, A.; Clark, A.; Ostrow, A.; Welihinda, A.; Hayes, A.; Radford, A.; et al. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Kang, W.-C.; and McAuley, J. 2018. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*, 197–206. IEEE.

Lai, R.; Chen, L.; Chen, R.; and Zhang, C. 2025. DAR: Dimension-Adaptive Recommendation with Multi-Granular Noise Control. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2203–2212.

Lai, R.; Chen, R.; Han, Q.; Zhang, C.; and Chen, L. 2024. Adaptive hardness negative sampling for collaborative filtering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8645–8652.

Li, X.; Yan, Z.; Meng, D.; Dong, L.; Zeng, X.; He, Y.; Wang, Y.; Qiao, Y.; Wang, Y.; and Wang, L. 2025. Videochat-r1: Enhancing spatio-temporal perception via reinforcement fine-tuning. *arXiv preprint arXiv:2504.06958*.

Liu, Z.; Sun, Z.; Zang, Y.; Dong, X.; Cao, Y.; Duan, H.; Lin, D.; and Wang, J. 2025. Visual-rft: Visual reinforcement fine-tuning. *arXiv preprint arXiv:2503.01785*.

Ma, Y.; Wang, Y.; Wu, Y.; Lyu, Z.; Chen, S.; Li, X.; and Qiao, Y. 2022. Visual knowledge graph for human action reasoning in videos. In *Proceedings of the 30th ACM International Conference on Multimedia*, 4132–4141.

Ni, Y.; Cheng, Y.; Liu, X.; Fu, J.; Li, Y.; He, X.; Zhang, Y.; and Yuan, F. 2023. A content-driven micro-video recommendation dataset at scale. *arXiv preprint arXiv:2309.15379*.

Pan, W.; Xia, S.; Liu, Z.; Peng, X.; and Ming, Z. 2016. Mixed factorization for collaborative recommendation with heterogeneous explicit feedbacks. *Information Sciences*, 332: 84–93.

Pan, Y.; Gao, C.; Chang, J.; Niu, Y.; Song, Y.; Gai, K.; Jin, D.; and Li, Y. 2023. Understanding and modeling passive-negative feedback for short-video sequential recommendation. In *Proceedings of the 17th ACM conference on recommender systems*, 540–550.

Seo, C.; Jeong, K.-J.; Lim, S.; and Shin, W.-Y. 2022. SiReN: Sign-aware recommendation using graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 35(4): 4729–4743.

Wang, L.; Zhang, J.; Yang, H.; Chen, Z.-Y.; Tang, J.; Zhang, Z.; Chen, X.; Lin, Y.; Sun, H.; Song, R.; et al. 2025a.

User behavior simulation with large language model-based agents. *ACM Transactions on Information Systems*, 43(2): 1–37.

Wang, Q.; Yu, Y.; Yuan, Y.; Mao, R.; and Zhou, T. 2025b. VideoRFT: Incentivizing Video Reasoning Capability in MLLMs via Reinforced Fine-Tuning. *arXiv preprint arXiv:2505.12434*.

Wang, X.; Tang, X.; Zhao, W. X.; Wang, J.; and Wen, J.-R. 2023a. Rethinking the evaluation for conversational recommendation in the era of large language models. *arXiv preprint arXiv:2305.13112*.

Wang, Y.; Halpern, Y.; Chang, S.; Feng, J.; Le, E. Y.; Li, L.; Liang, X.; Huang, M.-C.; Li, S.; Beutel, A.; et al. 2023b. Learning from Negative User Feedback and Measuring Responsiveness for Sequential Recommenders. In *Proceedings of the 17th ACM Conference on Recommender Systems*, 1049–1053.

Xie, R.; Ling, C.; Wang, Y.; Wang, R.; Xia, F.; and Lin, L. 2021. Deep feedback network for recommendation. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, 2519–2525.

Yang, A.; Li, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Gao, C.; Huang, C.; Lv, C.; et al. 2025a. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Yang, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Li, C.; Liu, D.; Huang, F.; Wei, H.; et al. 2024a. Qwen2. 5 technical report. *arXiv preprint arXiv:2412.15115*.

Yang, A.; Zhang, B.; Hui, B.; Gao, B.; Yu, B.; Li, C.; Liu, D.; Tu, J.; Zhou, J.; Lin, J.; et al. 2024b. Qwen2. 5-math technical report: Toward mathematical expert model via self-improvement. *arXiv preprint arXiv:2409.12122*.

Yang, J.; Yi, X.; Zhiyuan Cheng, D.; Hong, L.; Li, Y.; Xiaoming Wang, S.; Xu, T.; and Chi, E. H. 2020. Mixed negative sampling for learning two-tower neural networks in recommendations. In *Companion proceedings of the web conference 2020*, 441–447.

Yang, S.; Cao, J.; Li, H.; Mao, Y.; and Pang, S. 2025b. RecCoT: Enhancing Recommendation via Chain-of-Thought. *arXiv preprint arXiv:2506.21032*.

Ye, Y.; Zheng, Z.; Shen, Y.; Wang, T.; Zhang, H.; Zhu, P.; Yu, R.; Zhang, K.; and Xiong, H. 2025. Harnessing multimodal large language models for multimodal sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 13069–13077.

Ying, R.; He, R.; Chen, K.; Eksombatchai, P.; Hamilton, W. L.; and Leskovec, J. 2018. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 974–983.

Yuan, F.; He, X.; Karatzoglou, A.; and Zhang, L. 2020. Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 1469–1478.

Zhang, A.; Chen, Y.; Sheng, L.; Wang, X.; and Chua, T.-S. 2024a. On generative agents in recommendation. In *Proceedings of the 47th international ACM SIGIR conference on research and development in Information Retrieval*, 1807–1817.

Zhang, E.; Wang, X.; Gong, P.; Lin, Y.; and Mao, J. 2024b. Usimagent: Large language models for simulating search users. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2687–2692.

Zhang, Z.; Liu, S.; Liu, Z.; Zhong, R.; Cai, Q.; Zhao, X.; Zhang, C.; Liu, Q.; and Jiang, P. 2025. Llm-powered user simulator for recommender system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 13339–13347.

Zheng, L.; Lu, C.-T.; Jiang, F.; Zhang, J.; and Yu, P. S. 2018. Spectral collaborative filtering. In *Proceedings of the 12th ACM conference on recommender systems*, 311–319.