Generative Recommendation: Towards Next-generation Recommender Paradigm

Wenjie Wang[1], Xinyu Lin[1], Fuli Feng[2*], Xiangnan He[2], and Tat-Seng Chua[1]
[1]National University of Singapore, [2]University of Science and Technology of China
{wenjiewang96,xylin1028,fulifeng93,xiangnanhe}@gmail.com,dcscts@nus.edu.sg

ABSTRACT

Recommender systems typically retrieve items from an item corpus for personalized recommendations. However, such a retrieval-based recommender paradigm faces two limitations: 1) the human-generated items in the corpus might fail to satisfy the users' diverse information needs, and 2) users usually adjust the recommendations via passive and inefficient feedback such as clicks. Nowadays, AI-Generated Content (AIGC) has revealed significant success across various domains, offering the potential to overcome these limitations: 1) generative AI can produce personalized items to meet users' specific information needs, and 2) the newly emerged ChatGPT significantly facilitates users to express information needs more precisely via natural language instructions. In this light, the boom of AIGC points the way towards the next-generation recommender paradigm with two new objectives: 1) generating personalized content through generative AI, and 2) integrating user instructions to guide content generation.

To this end, we propose a novel **Gene**rative **Rec**ommender paradigm named GeneRec, which adopts an AI generator to personalize content generation and leverages user instructions to acquire users' information needs. Specifically, we pre-process users' instructions and traditional feedback (*e.g.,* clicks) via an instructor to output the generation guidance. Given the guidance, we instantiate the AI generator through an AI editor and an AI creator to repurpose existing items and create new items, respectively. Eventually, GeneRec can perform content retrieval, repurposing, and creation to meet users' information needs. Besides, to ensure the trustworthiness of the generated items, we emphasize various fidelity checks such as authenticity and legality checks. Lastly, we study the feasibility of implementing the AI editor and AI creator on micro-video generation, showing promising results.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Generative Recommender Paradigm, AI-generated Content, Next-generation Recommender Systems

∗ Corresponding author: Fuli Feng (fulifeng93@gmail.com).

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Wenjie Wang, Xinyu Lin, Fuli Feng, Xiangnan He, and Tat-Seng Chua. 2023. Generative Recommendation: Towards Next-generation Recommender Paradigm. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

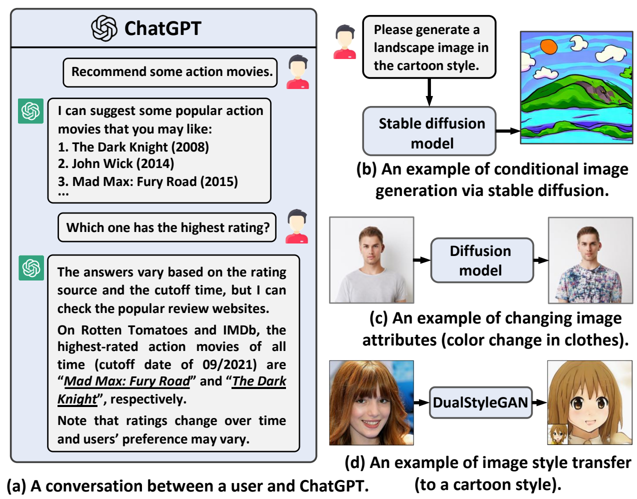

Figure 1: AIGC examples, including (a) conversation generation from ChatGPT[1], (b) image creation[2] and (c) image repurposing via diffusion models [24, 42], and (d) style transfer via DualStyleGAN [52].

Recommender systems fulfill users' information needs by retrieving item content in a personalized manner. Traditional recommender systems primarily retrieve human-generated content such as the professionally-generated movies on Netflix and user-generated micro-videos on Tiktok. However, AI-Generated Content (AIGC) has emerged as a prevailing trend across various domains. The advent of powerful neural networks, exemplified by GPT-3 [4], has enabled generative AI to produce superhuman content. As shown in Figure 1, ChatGPT [4, 34] demonstrates a remarkable ability to generate textual responses to users' diverse queries; diffusion models [9, 42] can generate vivid images and change attributes of existing images; while DualStyleGAN [52] can easily transform image styles based on users' requirements. Driven by the boom of AIGC, recommender systems must move beyond human-generated content, by envisioning a generative recommender paradigm to automatically repurpose existing items or create new items to supplement the users' diverse information needs.

[1]https://chat.openai.com/chat/.
[2]https://stablediffusionweb.com/.

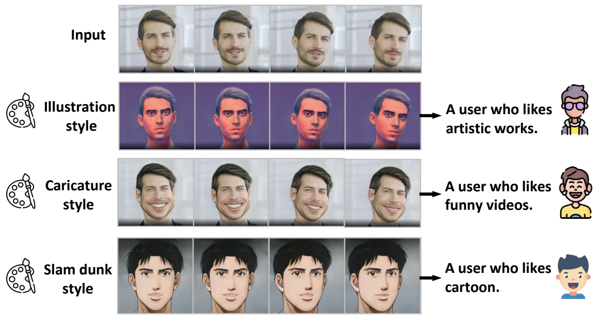

Figure 2: Examples of micro-video style transfer via VToonify [53] according to personalized user preference.

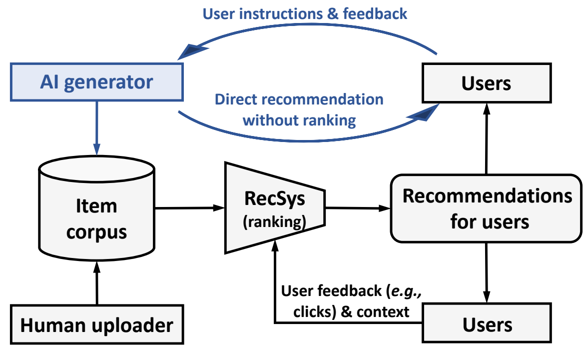

Figure 3: Illustration of the GeneRec paradigm. AI generator takes user instructions and feedback to generate personalized content, which can be directly recommended to users or fed to item corpus for ranking with human-generated items.

To pursue the generative recommender paradigm, we first retrospect the traditional retrieval-based recommender paradigm. As illustrated in Figure 3, the traditional paradigm ranks human-generated items in the item corpus, recommends the top-ranked items to users, and then collects user feedback (*e.g.,* clicks) and context (*e.g.,* interaction time) to optimize the future rankings for users. Despite its success, such a traditional paradigm suffers from two limitations. 1) The content available in the item corpus might be insufficient to satisfy users' personalized information needs. For instance, users may prefer a micro-video in a specific style such as cartoon, while generating such micro-videos by humans is time-consuming and costly. And 2) users are currently able to refine the recommendations mostly via passive feedback (*e.g.,* clicks), which cannot express their information needs explicitly and efficiently.

AIGC offers the potential to overcome the inherent limitations of the retrieval-based recommender paradigm. In particular, 1) generative AI can generate personalized content in real time, including repurposing existing items and creating new items, to supplement users' diverse information needs. For instance, it can quickly transform a micro-video into any style according to personalized user preference as shown in Figure 2. Additionally, 2) the newly published ChatGPT-like models have facilitated a powerful interface for users to convey their diverse information needs more precisely via natural language instructions (Figure 1(a)), supplementing traditional user feedback for information seeking. In this light, the emerging AIGC has spurred new objectives for the next-generation recommender systems to enable: 1) the automatic generation of personalized content through generative AI, and 2) the integration of user instructions to guide content generation.

To this end, we propose a novel **Gene**rative **Rec**ommender paradigm called **GeneRec**, which integrates the powerful generative AI for personalized content generation, including both repurposing and creation. Figure 3 illustrates how GeneRec adds a loop between *an AI generator* and *users.* Taking user instructions and feedback as inputs, the AI generator needs to understand users' information needs and generate personalized content. The generated content can be either added to the item corpus for ranking or directly recommended to the users. Wherein, the user instructions are not only limited to textual conversations but can also include multimodal conversations, *i.e.,* fusing images, videos, audio, and natural languages to express the information needs, such as an instruction with a micro-video and the description "transfer the micro-video to a cartoon style" or "change the color of clothes".

To instantiate the GeneRec paradigm, we formulate one module to process the instructions, as well as two modules to implement the AI generator. Specifically, an instructor module pre-processes user instructions and feedback to determine whether to initiate the AI generator to better satisfy users' needs, and also encodes the instructions and feedback to guide the content generation. Given the guidance, an AI editor repurposes an existing item to fulfill users' specific preference, *i.e.,* personalized item editing, and an AI creator directly creates new items for personalized item creation. To ensure the trustworthiness and high quality of generated items, we emphasize the importance of various fidelity checks from the aspects of bias, privacy, safety, authenticity, and legality. Lastly, to explore the feasibility of applying the recent advance in AIGC to implement the AI editor and AI creator, we devise several tasks of micro-video generation and conduct experiments on a high-quality micro-video dataset. Empirical results show that existing AIGC methods can accomplish some repurposing and creation tasks, and it is promising to achieve the grand objectives of GeneRec in the future. We release our code and dataset at https://github.com/Linxyhaha/GeneRec.

To summarize, our contributions are threefold.

- We highlight the essential role of AIGC in recommender systems and point out the extended objectives for next-generation recommender systems: moving towards a generative recommender paradigm, which can naturally interact with users via multimodal instructions, and flexibly retrieve, repurpose, and/or create item content to meet users' diverse information needs.
- We propose to instantiate the generative recommender paradigm by formulating three modules: the instructor for processing user instructions, the AI editor for personalized item editing, and the AI creator for personalized item creation.
- We investigate the feasibility of utilizing existing AIGC methods to implement the proposed generative recommender paradigm and suggest promising research directions for future work.

2 GENERATIVE RECOMMENDER PARADIGM

Motivated by the boom of AIGC, we propose two new objectives for the next-generation recommender systems: 1) automatically repurposing or creating items to meet users' diverse needs, and 2) integrating rich instructions to guide content repurposing and creation. To achieve these objectives, we present GeneRec to complement the traditional retrieval-based recommender paradigm.

• **Overview.** Figure 3 presents the overview of the proposed GeneRec paradigm with two loops. In the traditional retrieval-based user-system loop, humans, including domain experts (*e.g.,* musicians) and regular users (*e.g.,* micro-video users), generate and upload items to the item corpus. These items are then ranked for recommendations according to the user preference, where the preference is learned from the context (*e.g.,* interaction time) and user feedback over historical recommendations.

To complement this traditional paradigm, GeneRec adds another loop between the AI generator and users. Users can control the content generated by the AI generator through user instructions and feedback. Thereafter, the generated items can be directly exposed to the users without ranking if the users clearly express their expectations for AI-generated items or if they have rejected human-generated items via negative feedback (*e.g.,* dislikes) many times. In addition, the AI-generated items can be ranked together with the human-generated items to output the recommendations.

• **User instructions.** The strong conversational ability of ChatGPT-like models can enrich the interaction modes between users and the AI generator. The users can flexibly control content generation via conversational instructions, where the instructions can be either textual conversations or multimodal conversations. Through instructions, users can express their information needs more quickly and efficiently than interaction-based feedback. Besides, using interactive instructions, users can freely enable the AI generator to generate their preferred content at any time.

• **AI generator.** Before content generation, the AI generator might need to ***pre-process*** the user instructions, for instance, some pre-trained language models might require designing prompts [4] or instruction tuning [50]; diffusion models may need to simplify queries or extract instruction embeddings as inputs for image synthesis [42]. In addition to user instructions, user feedback such as clicks can also guide the content generation since user instructions might ignore some user preference and the AI generator can infer such preference from users' historical interactions.

Subsequently, the AI generator learns personalized information needs from user instructions and feedback, and then generates personalized item content accordingly. The generation includes both ***repurposing*** existing items and ***creating*** new items from scratch. For example, to repurpose a micro-video, we may convert it into multiple styles (see Figure 2) or split it into clips with different themes (see Figure 8) for distinct users; besides, the AI generator may select a topic and create a new micro-video based on user instructions and collected Web data (*e.g.,* facts and knowledge).

Post-processing is essential to ensure the quality of generated content. The AI generator can judge whether the generated content will satisfy users' information needs and further refine it, such as adding captions and subtitles for micro-videos. Besides, it is also vital to ensure the trustworthiness of generated content.

• **Fidelity checks.** To ensure the generated content is accurate, fair, and safe, GeneRec should pass the following fidelity checks.

1) **Bias and fairness**: the AI generator might learn from biased data [2], and thus should confirm that the generated content does not perpetuate stereotypes, promote hate speech and discrimination, cause unfairness to certain populations, or reinforce other harmful biases [12, 13, 54].
2) **Privacy**: the generated content can not disseminate any sensitive or personal information that may violate someone's privacy [43].
3) **Safety**: the AI generator must not pose any risks of harm to users, including the risks of physical and psychological harm [1]. For instance, the generated micro-video for teenagers should not contain any unhealthy content. Besides, it is crucial to prevent GeneRec from various attacks such as shilling attack [6, 14].
4) **Authenticity**: to prevent misinformation spread, we need to verify that the facts, statistics, and claims made in the generated content are accurate based on reliable sources [31].
5) **Legal compliance**: more importantly, AIGC must comply with all relevant laws and regulations [5]. For instance, if the generated micro-videos are about recommending healthy food, they must follow the regulations of healthcare. In this light, we also emphasize that enacting new laws to regulate AIGC and its spread is necessary and urgent.
6) **Identifiability**: to assist with AIGC supervision, we suggest adding the *digital watermark* [47] into AI-generated item content for distinguishing human-generated and AI-generated items. Besides, we can also develop AI technologies to automatically identify AI-generated items [32]. Furthermore, we may consider deleting the AI-generated items after browsing by users to prevent them from being reused for inappropriate context, reducing the harmful spread of AIGC.

• **Evaluation.** To evaluate the generated content, we propose two kinds of evaluation setups: 1) item-side evaluation and 2) user-side evaluation. ***Item-side evaluation*** emphasizes the measurements from the item itself, including the item quality measurements (*e.g.,* using Fréchet Video Distance (FVD) metric [48] to measure micro-video quality) and various fidelity checks. ***User-side evaluation*** judges the quality of generated content based on users' satisfaction. The satisfaction can be collected either by explicit feedback or implicit feedback like in the traditional retrieval-based recommender paradigm. In detail, 1) explicit feedback includes users' ratings and conversational feedback, *e.g.,* "I like this item" in natural language. Moreover, we can design multiple facets to help users' evaluation, for instance, the style, length, and thumbnail for the evaluation of generated micro-videos. And 2) implicit feedback (*e.g.,* clicks) can also be evaluated. The widely used metrics such as the click-through rate, dwell time, and user retention rate are still applicable to measure users' satisfaction.

3 DEMONSTRATION

To instantiate the proposed GeneRec, we develop three modules: an instructor, an AI editor, and an AI creator. As described in Figure 4, the instructor is responsible for pre-processing user instructions, while the AI editor and AI creator implement the AI generator for personalized item editing and creation, respectively.

3.1 Instructor

The instructor aims to pre-process user instructions and feedback to guide the content generation of the AI generator.

• **Input:** Users' multimodal conversational instructions and the feedback over historically recommended items.

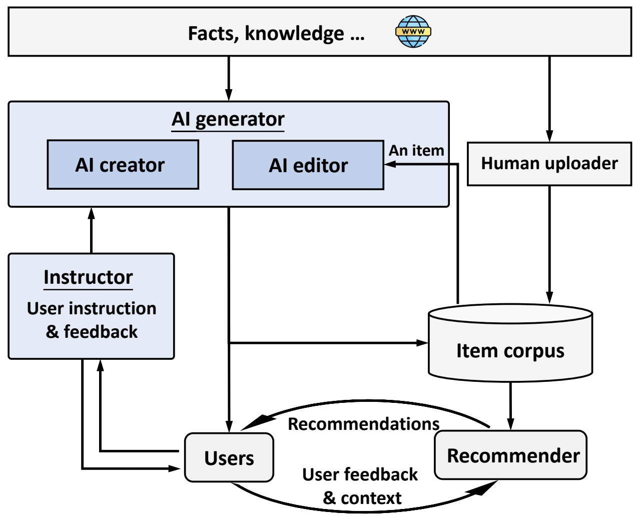

Figure 4: A demonstration of GeneRec. The instructor collects user instructions and feedback to guide content generation. The AI editor aims to repurpose existing items in item corpus while AI creator directly creates new items.

• **Processing:** Given the inputs, the instructor may still need to engage in multi-turn interactions with the users to fully understand users' information needs. Thereafter, the instructor analyzes the multimodal instructions and user feedback to determine whether there is a need to initiate the AI generator to meet users' information needs. If the users have explicitly requested AIGC via instructions or rejected human-generated items many times, the instructor may enable the AI generator for content generation. The instructor then pre-processes users' instructions and feedback as guidance signals, according to the input requirements of the AI generator. For instance, some pre-trained language models may need appropriately designed prompts and diffusion models might require the extraction of guidance embeddings from users' instructions and historically liked item features.

• **Output:** 1) The decision on whether to initiate the AI generator, and 2) the guidance signals for content generation.

3.2 AI Generator

To implement the AI generator for content generation, we formulate two modules: an AI editor and an AI creator.

3.2.1 ***AI editor for personalized item editing.*** As depicted in Figure 4, the AI editor intends to refine and repurpose existing items (generated by either humans or AI) in the item corpus according to personalized user instructions and feedback.

• **Input:** 1) The guidance signals extracted from user instructions and feedback by the instructor, 2) an existing item in the corpus, and 3) the facts and knowledge from the Web data.

• **Processing:** Given the input data, the AI editor leverages powerful neural networks to learn the users' information needs and preference, and then repurpose the input item accordingly. The "facts and knowledge" here can provide some factual events, generation skills, common knowledge, laws, and regulations to help generate accurate, safe, and legal items. For instance, based on user instructions, the AI editor might convert a micro-video into a cartoon style by imitating cartoon-style examples on the Web.

• **Output:** An edited item that better fulfills users' information preference than the original one.

3.2.2 ***AI creator for personalized item creation.*** Besides the AI editor, we also develop an AI creator to generate new items based on personalized user instructions and feedback.

• **Input:** 1) The guidance signals extracted from user instructions and feedback by the instructor, and 2) the facts and knowledge from the Web data.

• **Processing:** Given the guidance signals, facts, and knowledge, the AI creator learns the users' information needs, and creates new items to fulfill users' needs. For instance, the AI creator might determine a topic about "landscape" based on user instructions, learn about users' specific preference over "cartoon styles" from user feedback, and exploit some facts and knowledge to make a landscape micro-video in the cartoon style.

• **Output:** A new item that fulfills users' information needs.

4 FEASIBILITY STUDY

To investigate the feasibility of instantiating GeneRec, we employ AIGC methods to implement the AI editor and AI creator on a micro-video dataset due to the widespread of micro-video content.

• **Dataset.** We utilize a high-quality micro-video dataset with raw videos. It contains 64,643 interactions between 7,895 users and 4,570 micro-videos of diverse genres (*e.g.,* news and celebrities). The micro-video length is longer than eight seconds and each micro-video has a thumbnail with an approximate resolution of 1934×1080. We follow the micro-video pre-processing in [48] and each pre-processed micro-video has 400 frames with 64×64 resolution.

4.1 AI Editor

We design three tasks for personalized micro-video editing and separately tailor different methods for the tasks.

4.1.1 ***Thumbnail selection and generation.*** Considering that personalized thumbnails might better attract users to click on micro-videos [29], we devise the tasks of personalized thumbnail selection and generation to present a more attractive and personalized micro-video thumbnail for users.

• **Task.** We aim to generate personalized thumbnails based on user feedback without requiring user instructions. Formally, given a micro-video in the corpus and a user's historical feedback, the AI editor should select a frame from the micro-video as a thumbnail or generate a thumbnail to match the user's preference.

• **Implementation.** To implement personalized thumbnail selection, we utilize the image encoder $f_\theta(\cdot)$ of a representative Contrastive Language Image Pre-training model (CLIP) [37] to conduct zero-shot selection. As illustrated in Figure 5(a), given a set of N frames $\{\boldsymbol{v}_i\}_{i=1}^N$ from a micro-video, and the set of M thumbnails $\{\boldsymbol{t}_i\}_{i=1}^M$ from a user's historically liked micro-videos, we calculate

$$\begin{cases} \bar{\boldsymbol{t}} = \dfrac{1}{M}\displaystyle\sum_{i=1}^{M} f_\theta(\boldsymbol{t}_i), \\ j = \underset{i\in\{1,\ldots,N\}}{\arg\max}\ \{\bar{\boldsymbol{t}}^T \cdot f_\theta(\boldsymbol{v}_i)\}, \end{cases} \tag{1}$$

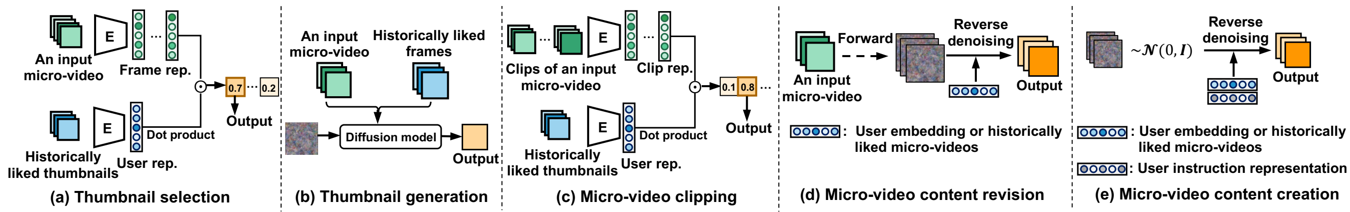

Figure 5: Illustration of the implementation of editing and creation tasks. (a)-(d) depict the process of AI editors for three micro-video editing tasks, and (e) shows the procedure of the AI creator for micro-video content creation.

Figure 6: Cases of personalized thumbnail selection by CLIP.

Table 1: Performance of CLIP (thumbnail selection), RDM (thumbnail generation), and the baselines. The best results are highlighted in bold and the second-best underlined.

Thumbnail Selection and Generation	Cosine@5	Cosine@10	PS@5	PS@10
Random Frame	0.4796	0.4786	22.6735	23.1950
Original	0.4978	0.4970	22.2606	22.7445
CLIP	<u>0.5142</u>	<u>0.5134</u>	<u>22.7682</u>	<u>23.2854</u>
RDM	**0.5369**	**0.5347**	**23.0145**	**23.3712**

where $\bar{t}$ is the average representation of $\{f_\theta(t_i)\}_{i=1}^{M}$, and we select the j-th frame as the recommended thumbnail due to the highest dot product score between the user representation $\bar{t}$ and the j-th frame representation $f_\theta(v_j)$. For performance comparison, we randomly select a frame from the micro-video ("Random Frame") and utilize the original thumbnail ("Original") as the two baselines.

To achieve personalized thumbnail generation, we adopt a newly pre-trained Retrieval-augmented Diffusion Model (RDM) [3], in which an external item corpus can be plugged as conditions to guide image synthesis. To generate personalized thumbnails for a micro-video, we combine this micro-video and the user's historically liked micro-videos as the input conditions of RDM (see Figure 5(b)).

• **Evaluation.** To evaluate the selected and generated thumbnails, we propose two metrics: 1) **Cosine@**K that takes the average cosine similarity between the selected/generated thumbnails of K recommended items and the user's historically liked thumbnails; and 2) **PS@**K that calculates the Prediction Score (PS) from a well trained MMGCN [51] by using the features of K selected/generated thumbnails. In detail, we train an MMGCN by using the thumbnail features and user representations, and then averaging the prediction scores between the K selected/generated thumbnails and the target user representation. Higher scores of Cosine@K and PS@K imply better results. For each user, we randomly choose $K = 5$ or 10 non-interacted items as recommendations and report the average results of ten experimental runs to ensure reliability.

• **Results.** The results of thumbnail selection and generation *w.r.t.* Cosine@K and PS@K are reported in Table 1, from which we have the following findings. 1) "Original" usually yields better Cosine@K scores than "Random Frame" since the thumbnails manually selected by the uploaders are more appealing to users than random frames. 2) CLIP outperforms "Random Frame" and "Original" by considering user feedback, validating the efficacy of personalized thumbnail selection. 3) RDM achieves the best results, justifying the superiority of using diffusion models to generate personalized thumbnails. The superior results are reasonable since RDM can generate thumbnails beyond existing frames, leading to a better alignment with user preference.

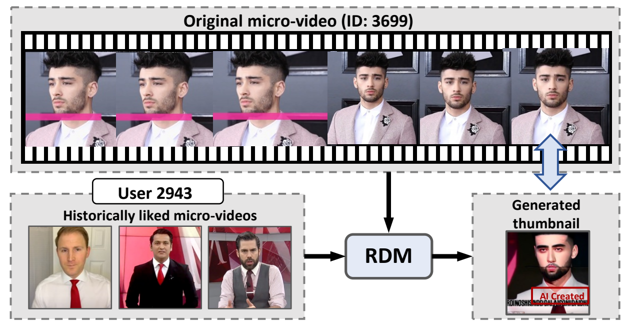

Figure 7: Cases of personalized thumbnail generation.

• **Case study.** For intuitive understanding, we visualize several cases from CLIP and RDM. From the cases of CLIP in Figure 6, we observe that the second frame containing rural landscape is selected for User 5 due to the user's historical preference for magnificent natural scenery. In contrast, the frame containing a fancy sports car is chosen for User 100 because this user likes to browse stylish vehicles. This reveals the effectiveness of CLIP in selecting personalized thumbnails according to different user preference. The case of RDM for thumbnail generation is presented in Figure 7. From the generated result, we can find that RDM tries to insert some elements to the generated thumbnail to better align with user preference while maintaining key information of the original micro-video. For instance, RDM decorates the man with a white shirt and a red tie for User 2943 based on this user's historical preference. Such observation reveals the potential of using generative AI to repurpose existing items for meeting personalized user preference. Nevertheless, we can see that the generated thumbnail lacks fidelity to some extent, probably due to the domain gap between this micro-video dataset and the pre-training data of RDM.

4.1.2 ***Micro-video clipping.*** Given a long micro-video (*e.g.*, longer than 1 minute), the task of personalized micro-video clipping aims to recommend only the users' preferred clip in order to save users' time and improve users' experience [30].

Table 2: Performance comparison between the baselines and CLIP with personalized micro-video clipping.

	Micro-video Clipping			
	Cosine@5	Cosine@10	PS@5	PS@10
Random	0.4864	0.4851	22.1483	23.1401
1st Clip	0.4910	0.4899	22.1509	23.1657
Unclipped	0.4969	0.4976	22.1685	23.1700
CLIP	**0.5052**	**0.5038**	**22.1863**	**23.1758**

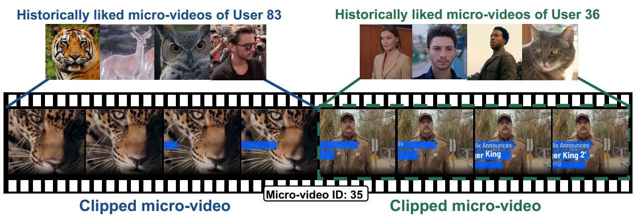

Figure 8: Case study of micro-video clipping via CLIP [37].

• **Task.** Given an existing micro-video and a user's historical feedback, the AI editor needs to select a shorter clip comprising of a set of consecutive frames from the original micro-video as the personalized recommendation.

• **Implementation.** Similar to the thumbnail selection, we leverage $f_\theta(\cdot)$ of CLIP [37] to obtain personalized micro-video clips as shown in Figure 5(c). Given user representation $\bar{\boldsymbol{t}}$ obtained from M historically liked thumbnails via Eq. (1), and a set of C clips $\{\boldsymbol{c}_i\}_{i=1}^{C}$ where each clip $\boldsymbol{c}_i$ has L consecutive frames[1] $\{\boldsymbol{v}_a^i\}_{a=1}^{L}$, we compute

$$\begin{cases} \bar{\boldsymbol{c}}_i = \dfrac{1}{L}\displaystyle\sum_{a=1}^{L} f_\theta(\boldsymbol{v}_a^i), \\ j = \underset{i\in\{1,\ldots,C\}}{\arg\max}\, \{\bar{\boldsymbol{t}}^T \cdot \bar{\boldsymbol{c}}_i\}, \end{cases} \tag{2}$$

where $\bar{\boldsymbol{c}}_i$ denotes the clip representation calculated by averaging the L frame representations, and we select j-th clip as the recommended one due to its highest similarity with the user representation $\bar{\boldsymbol{t}}$. For performance comparison, we select a random clip ("Random"), the first clip with $1 \sim L$ frames ("1st Clip"), and the original unclipped micro-video ("Unclipped") as baselines.

• **Results.** Similar to thumbnail selection and generation, we use Cosine@K and PS@K for evaluation by replacing thumbnails with frames to calculate cosine and prediction scores. From Table 2, we find that CLIP surpasses other approaches because it utilizes user feedback to select personalized clips that match users' specific preference. Besides, the superior performance of "Unclipped" over "Random" and "1st Clip" makes sense because the random clip and the first clip may lose some users' preferred content.

• **Case study.** In Figure 8, CLIP chooses two clips from a raw micro-video for two users with different preference. For User 83, the clip with the tiger is selected because of this user's interest in wild animals; in contrast, the clip with a man facing the camera is chosen for User 36 due to this user's preference for portraits.

4.1.3 ***Micro-video content editing.*** Users might wish to repurpose and refine the micro-video content according to personalized user preference. As such, we implement the AI editor to edit the micro-video content for satisfying users' information needs.

[1]The frame number L is a hyper-parameter for micro-video clipping. We tune it in $\{4, 8, 16, 32\}$ and choose 8 due to its better scores *w.r.t.* Cosine@5.

Table 3: Quantitative results of MCVD based on User_Hist and User_Emb. The FVD score of unconditional micro-video revision is 745.9443.

	Micro-video Content Revision				
	Cosine@5	Cosine@10	PS@5	PS@10	FVD
Original	0.5010	0.5083	25.8900	24.6800	-
User_Hist	0.5166	0.5127	25.9012	24.7107	783.7505
User_Emb	**0.5273**	**0.5200**	**26.0200**	**24.7900**	**646.7156**

Figure 9: Examples of personalized micro-video style transfer via VToonify.

• **Task.** Given an existing micro-video in the corpus, user instructions, and user feedback[2], the AI editor is asked to repurpose and edit the micro-video content to meet user preference.

• **Implementation.** We consider two subtasks for micro-video content editing: 1) ***micro-video style transfer*** based on user instructions, where we simulate user instructions to select some pre-defined styles and utilize an interactive tool VToonify[3] [53] to achieve the style transfer; and 2) ***micro-video content revision*** based on user feedback. We resort to a newly published Masked Conditional Video Diffusion model (MCVD) [48] for micro-video revision. The revision process is presented in Figure 5(d). We first fine-tune MCVD on this micro-video dataset by reconstructing the users' liked micro-videos conditional on user feedback. During inference, we forwardly corrupt the input micro-video by gradually adding noises, and then perform step-by-step denoising to generate an edited micro-video guided by user feedback. The user feedback for fine-tuning and inference can be obtained from: 1) user embeddings from a pre-trained recommender model such as MMGCN (denoted as "User_Emb"), and 2) averaged features of the user's historically liked micro-videos ("User_Hist"). To evaluate the quality of generated micro-videos, we follow [48] and adopt the widely used FVD metric, which measures the distribution gap between the real micro-videos and generated micro-videos. A lower FVD score indicates higher quality.

• **Results.** We show some cases of micro-video style transfer in Figure 9. The same micro-video is transferred into different styles according to users' instructions. From Figure 9, we can observe that the repurposed videos show high fidelity and quality, validating that the task of micro-video style transfer can be well accomplished by existing generative models. However, it might be costly for

[2]We do not consider the "facts and knowledge" in Section 3.2.1 to simplify the implementation, leaving the knowledge-enhanced implementation to future work.
[3]https://github.com/williamyang1991/vtoonify/.

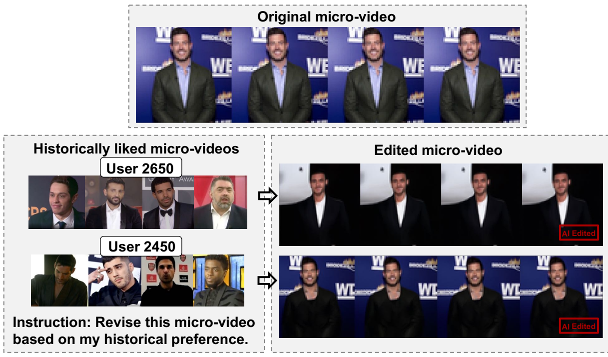

Figure 10: Case study of personalized micro-video content revision via MCVD (User_Emb).

users to give detailed instructions for more complex tasks, thus considering user feedback as guidance, especially implicit feedback like in micro-video content revision, is worth studying.

Quantitative results of micro-video content revision are summarized in Table 3, from which we have the following observation. 1) The edited items ("User_Hist" and "User_Emb") cater more to user preference, validating the feasibility of generative models for personalized micro-video revision. 2) "User_Emb" outperforms "User_Hist" possibly because the pre-trained user embeddings contain compressed critical information. "User_Hist" directly fuses the raw features from the user's historically liked micro-videos, which inevitably include some noises. And 3) the FVD score of "User_Emb" is significantly smaller than the unconditional revision, indicating the high quality of the edited micro-videos.

In addition, we analyze the cases of two users in Figure 10, where the original micro-video depicts a male standing in front of a backdrop. Given the users' instruction "revise this micro-video based on my historical preference", we initiate MCVD to repurpose the micro-video to meet personalized user preference. Specifically, since User 2650 prefers male portraits in a black suit and a white shirt, MCVD converts the dressing style of the man to match this user's preference. In contrast, for User 2450 who favors black shirts, MCVD alters both the suit and shirt to black accordingly. Despite the edited micro-videos having some deficiencies (*e.g.,* corrupted background for User 2650), we can find that integrating user instructions and feedback for higher-quality micro-video content revision is a promising direction. Besides, we highlight that the content revision and generation should add necessary watermarks and seek permission from all relevant stakeholders.

4.2 AI Creator

In this subsection, we explore instantiating the AI creator for micro-video content creation.

4.2.1 ***Micro-video content creation****.* Beyond repurposing thumbnails, clips, and content of existing micro-videos, we formulate an AI creator to create new micro-videos from scratch.

• **Task.** Given the user instructions and the user feedback over historical recommendations, the AI creator aims to create new micro-videos to meet personalized user preference.

• **Implementation.** Before implementing content creation, we investigate the performance of image synthesis based on user

Figure 11: Examples of single-turn instructions for text-to-image generation via stable diffusion.

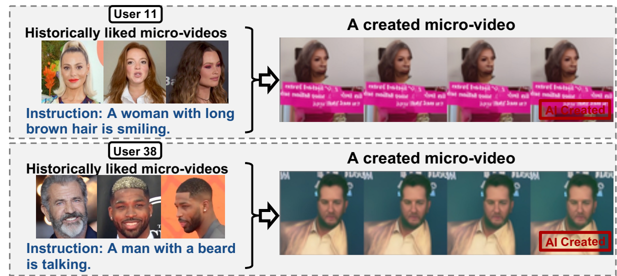

Figure 12: Case study of personalized micro-video content creation via MCVD (User_Emb).

Table 4: Quantitative results of MCVD based on User_Hist and User_Emb. The FVD score of unconditional micro-video creation is 727.8236.

	Micro-video Content Creation		
	Cosine@5	Cosine@10	FVD
Original	0.4883	0.4907	-
User_Hist	0.4902	0.4915	**735.0413**
User_Emb	**0.5356**	**0.5376**	743.1090

instructions, where we construct users' single-turn instructions and apply stable diffusion [42] for image synthesis. From the generated images in Figure 11, we can find that stable diffusion is capable of generating high-quality images according to users' single-turn instructions. Here, we explore the possibility of ***micro-video content creation*** via the video diffusion model MCVD. As presented in Figure 5(e), MCVD first samples a random noise from the standard normal distribution, and then it generates a micro-video based on personalized user instructions and user feedback through the denoising process. We write some textual single-turn instructions by humans such as "a man with a beard is talking", and encode them via CLIP [37]. The encoded instruction representation is then combined with user feedback for the conditional generation of MCVD. To represent user feedback, we still employ the pre-trained user embeddings ("User_Emb") and the average features of the user's historically liked micro-videos ("User_Hist"). Similar to micro-video content revision, we fine-tune MCVD on the micro-video dataset conditional on users' encoded instructions and feedback, and then apply it for content creation.

• **Results:** From Table 4, we can find that generated micro-videos can achieve higher Cosine@K[4] values as the generation is guided by personalized user feedback and instructions. In spite of the high values of Cosine@K, the quality of generated micro-videos

[4]We cannot calculate PS@K via MMGCN because the newly created items do not have item ID embeddings, which are necessary for MMGCN prediction.

from "User_Emb" and "User_Hist" is worse than the unconditional creation as shown by the larger FVD scores. The case study in Figure 12 also validates the unsatisfactory generation quality. Specifically, MCVD generates a micro-video containing a woman with long brown hair for User 11 where the women's face is distorted; and that for User 38 is also blurred and distorted.

Admittedly, the current results of personalized micro-video creation are less satisfactory. This inferior performance might be attributed to that: 1) the simple single-turn instructions fail to describe all the details in the micro-video; and 2) the current micro-video dataset lacks sufficient micro-videos, facts, and knowledge, thus limiting the creative ability of the AI generator. In this light, it is promising to enhance the AI generator in future work from two aspects: 1) enabling the powerful ChatGPT-like tools[5] to implement the instructor and acquire detailed user instructions; and 2) incorporating external facts and knowledge along with more video data through large-scale pre-training for micro-video generation. We believe that the development of video generation might also follow the trajectory of image generation (see Figure 11) and eventually achieve satisfactory generation results.

5 DISCUSSION

In this section, we present the comparison between the GeneRec paradigm and two related tasks, as well as the challenges and future opportunities for GeneRec.

5.1 Comparison

We illustrate the key differences between GeneRec with the conversational recommendation and traditional content generation.

5.1.1 ***Comparison with conversational recommendation***. Conversational recommender systems rely on multi-turn natural language conversations with users to acquire user preference and provide recommendations [36, 45, 55]. Although conversational recommender systems also consider multi-turn conversations, we highlight two critical differences with the GeneRec paradigm: 1) GeneRec automatically repurposes or creates items through the AI generator to meet users' specific information needs while conversational recommender systems are still retrieving human-generated items; and 2) the dramatic development of AIGC, especially ChatGPT, has brought a revolution to traditional conversational systems by significantly improving the understanding and generative abilities of language models, potentially leading to a more powerful instructor in GeneRec.

5.1.2 ***Comparison with traditional content generation***. There exist various cross-modal AIGC tasks such as text, image, and audio generation conditional on users' images [52], single-turn queries [42], and multi-turn conversations [23]. Nevertheless, there are essential differences between traditional AIGC tasks and GeneRec. 1) GeneRec can leverage user feedback such as clicks and dwell time to dig out the implicit user preference that is not indicated through user queries or conversations. For instance, users may not be aware of their preference for a particular type of micro-videos, whereas their clicks and long dwell time on these micro-videos can indicate this preference. In a more comprehensive manner, GeneRec considers both user instructions and feedback to complement each other. And 2) despite the success of AIGC, retrieving human-generated content remains indispensable in many cases to satisfy users' information needs. For example, journalists can bring live video reports from the scene of an event. As such, it is of vital importance to support the close cooperation between the AI generator and human uploaders in GeneRec.

5.2 Challenges and Opportunities

In the GeneRec paradigm, we point out three crucial ***challenges***.

1) **Understanding users' information needs.** The instructor should well capture users' information needs and personalized preference from the multimodal instructions and user feedback. However, it is non-trivial to fully understand users' multimodal instructions with limited turns of interactions and simultaneously combine the user feedback to estimate user preference.
2) **High-quality AIGC to supplement human-generated content.** The users' information needs are diverse and personalized, requiring high-quality AIGC in multiple modalities. Besides, the content generated by the AI generator should supplement human-generated content instead of producing duplicate content.
3) **Content fidelity checks.** As discussed in Section 2, it is critical to ensure the trustworthiness of the generated content by various fidelity checks. However, due to the biased training data and uncontrollable generative process, the generated content inevitably violates some constraints. At present, it is difficult to conduct valid checks. For example, the authenticity of statistics claimed in the generated articles is challenging to verify.

In spite of these challenges, the revolution of AI techniques offers some ***opportunities*** to achieve the GeneRec paradigm.

1) Extending the powerful ChatGPT to acquire users' multimodal instructions is feasible. After that, integrating the acquired multimodal instructions with user feedback via large-scale pre-trained models to infer users' needs is also promising.
2) The boom of various AIGC tasks will promote the development of the AI generator in GeneRec. At present, the generative models in different tasks have started to integrate with each other [26], facilitating the growth of a unified multimodal AI generator. Furthermore, the AI generator can collect user feedback on human-generated items to learn about the weakness of human-generated content, and then optimize content generation to supplement the weakness.
3) Achieving fidelity checks can leverage powerful AI models to review the generated content, akin to a left-right hand game, taking the strong discriminative ability of AI models to scrutinize the content generated by the AI generator. Additionally, it is promising to raise human-machine collaboration for fidelity checks. In the instances where AI is uncertain, AI can supply diverse information sources to incorporate humans for judgments.

• **A vision for future GeneRec.** At present, constrained by the development of existing AIGC, we need to design different modules to achieve the generation tasks across multiple modalities. However, we believe that building a unified AI model for multimodal content generation is a feasible and worthwhile endeavor. Under the GeneRec paradigm, we expect to see the increasing maturity of

[5]In this work, we do not explore the usage of ChatGPT because it was recently released and we have insufficient time for comprehensive exploration.

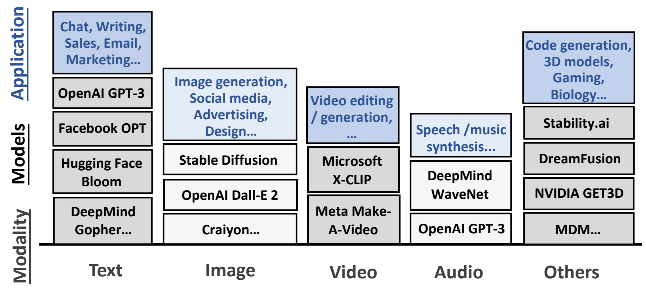

Figure 13: Illustration of advanced models and potential applications of AIGC across different modalities.

various generation tasks, along with growing integration between these tasks. Ultimately, with the inputs of user instructions and feedback, GeneRec will be able to perform retrieval, repurposing, and creation tasks by a unified model for personalized recommendations, leading to a totally new information seeking paradigm.

6 RELATED WORK

• **Recommender Systems.** The traditional retrieval-based recommender paradigm constructs a loop between the recommender models and users [39, 56]. Recommender models rank the items in a corpus and recommend the top-ranked items to the users [22, 54]. The collected user feedback and context are then used to optimize the next round of recommendations. Following such paradigm, recommender systems have been widely investigated. Technically speaking, the most representative method is Collaborative Filtering (CF), which assumes that users with similar behaviors share similar preference [11, 18, 27]. Early approaches directly utilize collaborative behaviors of similar users (*i.e.,* user-based CF) or items (*i.e.,* item-based CF). Later on, MF [40] decomposes the interaction matrix into user and item matrices separately, laying the foundation for subsequent neural CF methods [18] and graph-based CF methods [17]. Beyond purely using user-item interactions, prior work considers incorporating context [41] and the content features of users and items [16, 51] for session-based, sequential, and multimedia recommendations [8, 19, 25, 28]. In recent years, various new recommender frameworks have been proposed, such as conversational recommender systems [45, 46] acquiring user preference via conversations and user-controllable recommender systems [49] for controlling the attributes of recommended items.

However, past work only recommends human-generated items, which might fail to satisfy users' diverse information needs. In our work, we propose to empower traditional recommender paradigms with the ability of content generation for meeting users' information needs and present a novel generative recommender paradigm for next-generation recommender systems.

• **Generative AI.** The development of content generation roughly falls into three stages. At the early stage, most platforms heavily rely on high-quality professionally-generated content, which is however challenging to meet the demand of large-scale content production due to the high costs of professional experts. Later on, User-Generated Content (UGC) becomes prevalent due to the emergence of smartphones and well-wrapped generation tools. Despite its increasing growth, the quality of UGC is usually not guaranteed. Beyond human-generated content, recent years have witnessed third-stage content generation with groundbreaking generative AI techniques, leading to various AIGC-driven applications.

As the core of AIGC, generative AI has been extensively investigated across a diverse range of applications as depicted in Figure 13. In the text synthesis domain, substantial methods are proposed to serve different tasks such as article writing and dialog systems. For example, the newly published ChatGPT demonstrates an impressive ability for conversational interactions. Image and video synthesis are also two prevailing tasks of AIGC. Recent advancement of diffusion models shows promising results in generating high-quality images in various aesthetic styles [9, 33, 44] as well as high-coherence videos [20, 21, 48]. Besides, previous work has also explored audio synthesis tasks [38]. For instance, [10] proposes a generative model that directly aligns text with waves, endowing high-fidelity text-to-speech generation. Furthermore, extensive generative models are crafted for other fields, such as 3D generation [15], gaming [7], and chemistry [35].

The revolution of generative AI has catalyzed the production of high-quality content and brought a promising future for generative recommender systems. The advent of AIGC can complement existing human-generated content to better satisfy users' information needs. Besides, the powerful generative language models can help acquire users' information needs via multimodal conversations.

7 CONCLUSION AND FUTURE WORK

In this work, we empowered recommender systems with the abilities of content generation and instruction guidance. In particular, we proposed a GeneRec paradigm, which could: 1) acquire users' information needs via user instructions and feedback, and 2) achieve both item retrieval, repurposing, and creation to meet users' information needs. To instantiate GeneRec, we formulated three modules: an instructor for pre-processing user instructions and feedback, an AI editor for repurposing existing items, and an AI creator for creating new items. Besides, we highlighted the importance of multiple fidelity checks to ensure the trustworthiness of the generated content, and also pointed out the challenges and future opportunities of GeneRec. We explored the feasibility of implementing GeneRec on micro-video generation and the experiments reveal promising results on various tasks.

This work formulates a new generative paradigm for next-generation recommender systems, leaving many valuable research directions for future work. In particular, 1) it is critical to learn users' information needs from users' multimodal instructions and feedback. In detail, GeneRec should learn to ask questions for efficient information acquisition, reduce the modality gap to understand users' multimodal instructions, and utilize user feedback to complement instructions for better generation guidance. 2) Developing more powerful generative modules for various tasks (*e.g.,* thumbnail generation and micro-video creation) is essential. Besides, we might implement some generation tasks through a unified model, where multiple tasks may promote each other. And 3) we should devise new metrics, standards, and technologies to enrich the evaluation and fidelity checks of AIGC. It is also a promising direction to introduce human-machine collaboration for GeneRec evaluation and various fidelity checks.

REFERENCES

[1] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete Problems in AI Safety. *arXiv:1606.06565*.

[2] Ricardo Baeza-Yates. 2020. Bias in Search and Recommender Systems. In *RecSys*. ACM, 2–2.

[3] Andreas Blattmann, Robin Rombach, Kaan Oktay, Jonas Müller, and Björn Ommer. 2022. Semi-Parametric Neural Image Synthesis. In *NeurIPS*. Curran Associates, Inc.

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language Models Are Few-shot Learners. In *NeurIPS*. Curran Associates, Inc., 1877–1901.

[5] Miriam C Buiten. 2019. Towards Intelligent Regulation of Artificial Intelligence. *Eur J Risk Regul* 10, 1 (2019), 41–59.

[6] Paul-Alexandru Chirita, Wolfgang Nejdl, and Cristian Zamfir. 2005. Preventing Shilling Attacks in Online Recommender Systems. In *WIDM*. ACM, 67–74.

[7] Rafael Guerra de Pontes and Herman Martins Gomes. 2020. Evolutionary Procedural Content Generation for An Endless Platform Game. In *SBGames*. IEEE, 80–89.

[8] Yang Deng, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam. 2021. Unified Conversational Recommendation Policy Learning via Graph-based Reinforcement Learning. In *SIGIR*. ACM, 1431–1441.

[9] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion Models Beat Gans on Image Synthesis. In *NeurIPS*. Curran Associates, Inc., 8780–8794.

[10] Jeff Donahue, Sander Dieleman, Mikołaj Bińkowski, Erich Elsen, and Karen Simonyan. 2021. End-to-end Adversarial Text-to-speech. In *ICLR*.

[11] Travis Ebesu, Bin Shen, and Yi Fang. 2018. Collaborative Memory Network for Recommendation Systems. In *SIGIR*. ACM, 515–524.

[12] Zuohui Fu, Yikun Xian, Ruoyuan Gao, Jieyu Zhao, Qiaoying Huang, Yingqiang Ge, Shuyuan Xu, Shijie Geng, Chirag Shah, Yongfeng Zhang, et al. 2020. Fairness-aware Explainable Recommendation Over Knowledge Graphs. In *SIGIr*. ACM, 69–78.

[13] Ruoyuan Gao and Chirag Shah. 2020. Counteracting Bias and Increasing Fairness in Search and Recommender Systems. In *RecSys*. ACM, 745–747.

[14] Ihsan Gunes, Cihan Kaleli, Alper Bilge, and Huseyin Polat. 2014. Shilling Attacks Against Recommender Systems: A Comprehensive Survey. *Artificial Intelligence Review* 42, 4 (2014).

[15] Chuan Guo, Xinxin Zuo, Sen Wang, Shihao Zou, Qingyao Sun, Annan Deng, Minglun Gong, and Li Cheng. 2020. Action2motion: Conditioned Generation of 3d Human Motions. In *MM*. ACM, 2021–2029.

[16] Ido Guy, Naama Zwerdling, Inbal Ronen, David Carmel, and Erel Uziel. 2010. Social Media Recommendation Based on People and Tags. In *SIGIR*. ACM, 194–201.

[17] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and Powering Graph Convolution Network for Recommendation. In *SIGIR*. ACM, 639–648.

[18] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *WWW*. ACM, 173–182.

[19] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *ICLR*.

[20] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. 2022. Video Diffusion Models. *arXiv:2204.03458*.

[21] Wenyi Hong, Ming Ding, Wendi Zheng, Xinghan Liu, and Jie Tang. 2022. Cogvideo: Large-scale Pretraining for Text-to-video Generation via Transformers. *arXiv:2205.15868*.

[22] Chenhao Hu, Shuhua Huang, Yansen Zhang, and Yubao Liu. 2022. Learning to Infer User Implicit Preference in Conversational Recommendation. In *SIGIR*. ACM, 256–266.

[23] Yuming Jiang, Ziqi Huang, Xingang Pan, Chen Change Loy, and Ziwei Liu. 2021. Talk-to-edit: Fine-grained Facial Editing via Dialog. In *CVPR*. IEEE, 13799–13808.

[24] Yuming Jiang, Shuai Yang, Haonan Qju, Wayne Wu, Chen Change Loy, and Ziwei Liu. 2022. Text2human: Text-driven Controllable Human Image Generation. *TOG* 41, 4 (2022), 1–11.

[25] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *ICDM*. IEEE, 197–206.

[26] Jing Yu Koh, Ruslan Salakhutdinov, and Daniel Fried. 2023. Grounding Language Models to Images for Multimodal Generation. *arXiv:2301.13823*.

[27] Ioannis Konstas, Vassilios Stathopoulos, and Joemon M Jose. 2009. On Social Networks and Collaborative Recommendation. In *SIGIR*. ACM, 195–202.

[28] Wenqiang Lei, Xiangnan He, Yisong Miao, Qingyun Wu, Richang Hong, Min-Yen Kan, and Tat-Seng Chua. 2020. Estimation-action-reflection: Towards Deep Interaction Between Conversational and Recommender Systems. In *WSDM*. ACM, 304–312.

[29] Wu Liu, Tao Mei, Yongdong Zhang, Cherry Che, and Jiebo Luo. 2015. Multi-task Deep Visual-semantic Embedding for Video Thumbnail Selection. In *CVPR*. IEEE, 3707–3715.

[30] Huaishao Luo, Lei Ji, Ming Zhong, Yang Chen, Wen Lei, Nan Duan, and Tianrui Li. 2022. CLIP4Clip: An Empirical Study of CLIP for End to End Video Clip Retrieval and Captioning. *Neurocomputing* 508 (2022), 293–304.

[31] Marie-Helen Maras and Alex Alexandrou. 2019. Determining Authenticity of Video Evidence in the Age of Artificial Intelligence and In the Wake of Deepfake Videos. *The International Journal of Evidence & Proof* 23, 3 (2019), 255–262.

[32] Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D Manning, and Chelsea Finn. 2023. DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature. *arXiv:2301.11305*.

[33] Alexander Quinn Nichol and Prafulla Dhariwal. 2021. Improved Denoising Diffusion Probabilistic Models. In *ICML*. PMLR, 8162–8171.

[34] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training Language Models to Follow Instructions with Human Feedback. In *arXiv:2203.02155*.

[35] Daniil Polykovskiy, Alexander Zhebrak, Benjamin Sanchez-Lengeling, Sergey Golovanov, Oktai Tatanov, Stanislav Belyaev, Rauf Kurbanov, Aleksey Artamonov, Vladimir Aladinskiy, Mark Veselov, et al. 2020. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *Front. Pharmacol.* 11 (2020), 565644.

[36] Chen Qu, Liu Yang, W Bruce Croft, Johanne R Trippas, Yongfeng Zhang, and Minghui Qiu. 2018. Analyzing and Characterizing User Intent in Information-seeking Conversations. In *SIGIR*. ACM, 989–992.

[37] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning Transferable Visual Models from Natural Language Supervision. In *ICML*. PMLR, 8748–8763.

[38] Yi Ren, Jinzheng He, Xu Tan, Tao Qin, Zhou Zhao, and Tie-Yan Liu. 2020. Popmag: Pop Music Accompaniment Generation. In *MM*. ACM, 1198–1206.

[39] Zhaochun Ren, Zhi Tian, Dongdong Li, Pengjie Ren, Liu Yang, Xin Xin, Huasheng Liang, Maarten de Rijke, and Zhumin Chen. 2022. Variational Reasoning about User Preferences for Conversational Recommendation. In *SIGIR*. ACM, 165–175.

[40] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI*. AUAI Press, 452–461.

[41] Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2011. Fast Context-aware Recommendations with Factorization Machines. In *SIGIR*. ACM, 635–644.

[42] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution Image Synthesis with Latent Diffusion Models. In *CVPR*. IEEE, 10684–10695.

[43] Hyejin Shin, Sungwook Kim, Junbum Shin, and Xiaokui Xiao. 2018. Privacy Enhanced Matrix Factorization for Recommendation with Local Differential Privacy. *TKDE* 30, 9 (2018), 1770–1782.

[44] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2021. Denoising Diffusion Implicit Models. In *ICLR*.

[45] Yueming Sun and Yi Zhang. 2018. Conversational Recommender System. In *SIGIR*. ACM, 235–244.

[46] Quan Tu, Shen Gao, Yanran Li, Jianwei Cui, Bin Wang, and Rui Yan. 2022. Conversational Recommendation via Hierarchical Information Modeling. In *SIGIR*. ACM, 2201–2205.

[47] Ron G Van Schyndel, Andrew Z Tirkel, and Charles F Osborne. 1994. A Digital Watermark. In *ICIP*. IEEE, 86–90.

[48] Vikram Voleti, Alexia Jolicoeur-Martineau, and Christopher Pal. 2022. MCVD: Masked Conditional Video Diffusion for Prediction, Generation, and Interpolation. In *NeurIPS*. Curran Associates, Inc.

[49] Wenjie Wang, Fuli Feng, Liqiang Nie, and Tat-Seng Chua. 2022. User-controllable Recommendation Against Filter Bubbles. In *SIGIR*. ACM, 1251–1261.

[50] Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2022. Finetuned Language Models Are Zero-shot Learners. In *ICLR*.

[51] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. 2019. MMGCN: Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video. In *MM*. ACM, 1437–1445.

[52] Shuai Yang, Liming Jiang, Ziwei Liu, and Chen Change Loy. 2022. Pastiche Master: Exemplar-based High-resolution Portrait Style Transfer. In *CVPR*. IEEE, 7693–7702.

[53] Shuai Yang, Liming Jiang, Ziwei Liu, and Chen Change Loy. 2022. VToonify: Controllable High-Resolution Portrait Video Style Transfer. *TOG* 41, 6 (2022), 1–15.

[54] Meike Zehlike, Tom Sühr, Ricardo Baeza-Yates, Francesco Bonchi, Carlos Castillo, and Sara Hajian. 2022. Fair Top-k Ranking with multiple protected groups. *IPM* 59, 1 (2022), 102707.

[55] Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W Bruce Croft. 2018. Towards Conversational Search and Recommendation: System Ask, User Respond. In *CIKM*. ACM, 177–186.

[56] Jie Zou, Evangelos Kanoulas, Pengjie Ren, Zhaochun Ren, Aixin Sun, and Cheng Long. 2022. Improving Conversational Recommender Systems via Transformer-based Sequential Modelling. In *SIGIR*. ACM, 2319–2324.