

Generative Recommendation: Towards Next-generation Recommender Paradigm

WENJIE WANG, National University of Singapore, Singapore

XINYU LIN, National University of Singapore, Singapore

FULI FENG*, University of Science and Technology of China, China

XIANGNAN HE, University of Science and Technology of China, China

TAT-SENG CHUA, National University of Singapore, Singapore

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 satisfy users' specific information needs, and 2) the newly emerged large language models with strong language understanding and generation abilities significantly reduce the efforts of users to precisely express information needs 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 **Generative Recommender** 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 satisfy users' information needs. Besides, to ensure the trustworthiness of the generated items, we emphasize various fidelity checks such as authenticity and legality checks. Moreover, we provide a roadmap to envision future developments of GeneRec and we present several domain-specific applications of GeneRec with some potential research tasks. 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**.

Additional Key Words and Phrases: Generative Recommender Paradigm, AI-generated Content, Next-generation Recommender Systems, Large Language Models, Generative Models

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

Authors' addresses: Wenjie Wang, wenjiewang96@gmail.com, National University of Singapore, Singapore; Xinyu Lin, xylin1028@gmail.com, National University of Singapore, Singapore; Fuli Feng, fulifeng93@gmail.com, University of Science and Technology of China, China; Xiangnan He, xiangnanhe@gmail.com, University of Science and Technology of China, China; Tat-Seng Chua, dscts@nus.edu.sg, National University of Singapore, Singapore.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://www.acm.org/permissions).

© 2024 Association for Computing Machinery.

XXXX-XXXX/2024/2-ART \$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

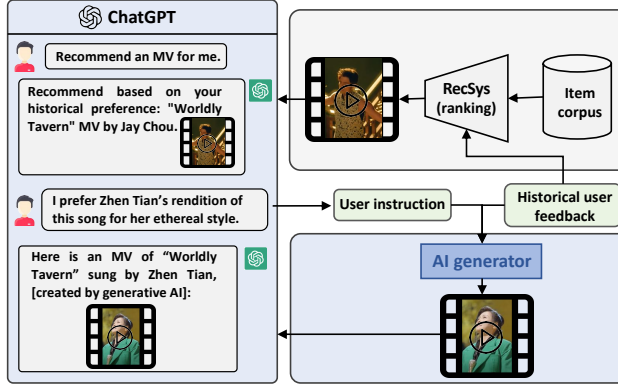


Fig. 1. An example of using generative AI to interact with users and generate new items in the micro-video domain.

ACM Reference Format:

Wenjie Wang, Xinyu Lin, Fuli Feng, Xiangnan He, and Tat-Seng Chua. 2024. Generative Recommendation: Towards Next-generation Recommender Paradigm. 1, 1 (February 2024), 25 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

Recommender systems fulfill users' information needs by retrieving item content in a personalized manner. Traditional recommender systems primarily retrieve human-generated content such as expert-generated movies on Netflix and user-generated micro-videos on Tiktok [24]. However, AI-Generated Content (AIGC) has emerged as a prevailing trend across various domains. The advent of powerful neural networks, exemplified by diffusion models [61], has enabled generative AI to produce superhuman content. As shown in Figure 1, ChatGPT [7, 53] demonstrates a remarkable ability to interact with users via natural language and generative AI is possible to create multimodal content such as video, text, and audio to form new items. Driven by the boom of AIGC, recommender systems must move beyond human-generated content, by envisioning a generative recommender paradigm to automatically repurpose existing items¹ or create new items.

To envision the generative recommender paradigm, we first retrospect the traditional retrieval-based recommender paradigm. 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 [13]. 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 music video performed by a singer in a specific style (see Figure 1), while generating such a multimodal music video by humans is impossible or costly [77]. 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 [43, 46].

AIGC offers the potential to overcome the inherent limitations of the retrieval-based recommender paradigm. In particular, 1) generative AI can generate personalized content to supplement existing items, including repurposing existing items and creating new items [6, 63]. Additionally, 2) the newly emerged Large Language Models (LLMs) show strong language understanding and generation

¹Here, repurposing means editing existing items for a different purpose, i.e., satisfying another user's personalized preference.

¹<https://chat.openai.com/chat/>.

²<https://stablediffusionweb.com/>.

abilities [75], which can effectively reduce users' efforts to convey their diverse information needs via natural language instructions (Figure 1). Compared to traditional conversational agents, users will engage more readily with advanced ChatGPT-like models, supplementing traditional user feedback. 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 **Generative Recommender** paradigm called **GeneRec**, which integrates the powerful generative AI for personalized content generation, including both repurposing and creation. Figure 2 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.

To instantiate the GeneRec paradigm, we formulate one instructor module to process the instructions, as well as two modules for item repurposing and creation. Specifically, the instructor module pre-processes user instructions and feedback to determine whether to initiate the content generation, 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 [27, 71, 77]. Moreover, in Section 2, we present a roadmap to explain the future development trends of GeneRec. Besides, we introduce several application scenarios of GeneRec across different domains in Section 3.3 and detail some potential research tasks under GeneRec in Section 3.3. 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 in various recommendation domains.
- We propose to instantiate the generative recommender paradigm by formulating three key modules: the instructor for interacting with users and processing user instructions to guide content generation, the AI editor for personalized item editing, and the AI creator for personalized item creation.
- We spotlight the essential perspectives of fidelity checks and present a roadmap with several application scenarios and potential research tasks to envision the future directions of GeneRec.
- We investigate the feasibility of utilizing existing AIGC methods to implement the AI editor and AI creator in the micro-video recommendation domain.

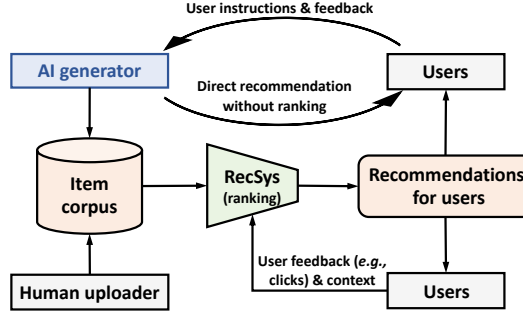


Fig. 2. Illustration of the GeneRec paradigm. The AI generator takes user instructions and feedback to generate personalized content, which can be directly recommended to users or fed to the item corpus for ranking with human-generated items.

2 GENERATIVE RECOMMENDER PARADIGM

We propose two new objectives for the next-generation recommender systems: 1) **automatically repurposing or creating items via generative AI**, and 2) **integrating rich user instructions**. To achieve these objectives, we present GeneRec to complement the traditional retrieval-based recommender paradigm.

• **Overview.** Figure 2 presents the overview of the proposed GeneRec paradigm with two loops. In the traditional retrieval-based user-system loop, human uploaders, 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 LLMs 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 1) freely enable the AI generator to generate their preferred content at any time, and 2) express their information needs more quickly and efficiently than interaction-based feedback. In particular, users cannot only express what they like, but can also indicate what they dislike via instructions. In the past, users were unwilling to make efforts to give explicit feedback such as instructions, while LLMs will be incredibly intelligent to collect user instruction by audio or multimodal inputs, reducing the burden of users via advanced interfaces in the future.

• **Content generation.** 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 [7] or instruction tuning [74]; diffusion models may need to simplify queries or extract instruction embeddings as inputs for image synthesis [61]. In addition to user instructions, user

feedback such as clicks can also guide the content generation since user instructions might ignore some implicit user preference and the AI generator can infer such preference from users' historical interactions. Learning implicit user preference from noisy and passive user feedback (e.g., clicks and dwell time) has long been a central focus in the field of recommendations. GeneRec can borrow the prior experience from the traditional recommendation domain for user preference modeling.

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 or split it into clips with different themes 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. The relevance between the generated content and users' information needs is also significant. Besides, it is also vital to ensure the trustworthiness of generated content through fidelity checks.

• **Fidelity checks.** To ensure the generated content is accurate, fair, and safe, GeneRec should pass the fidelity checks. Although different recommendation scenarios may require various fidelity checks, they generally should include but not be limited to the following perspectives.

- 1) **Bias and fairness:** the AI generator might learn from biased data [3], 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 [20, 21, 84]. For example, the news generation, especially regarding sensitive topics, should pay more attention to the checks of bias and fairness to avoid ethical and social issues.
- 2) **Privacy:** the generated content cannot disseminate any sensitive or personal information that may violate someone's privacy [62]. GeneRec may generate an item based on a user's personalized data, and then the generated item may be disseminated to other users, possibly leading to privacy leakage. Many recommendation scenarios such as news, tweets, and micro-videos are sensitive to such privacy leakage.
- 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 [11, 25].
- 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 [50]. Users need to access factual information in certain recommendation scenarios, such as daily events and historical facts, making the authenticity of the generated content extremely important.
- 5) **Legal compliance:** more importantly, AIGC must comply with all relevant laws and regulations [8]. For instance, if the generated micro-videos are about recommending healthy food, they must follow the regulations of healthcare. Besides, copyright regulation should also be considered to tackle the intricacies of authorship and ownership regarding the content edited or created by the generative AI [48]. The government should work with the recommender platform to publish new regulations to clarify the ownership regarding AI-generated content.
- 6) **Identifiability:** to assist with AIGC supervision, we suggest adding *digital watermark* [68] into AI-generated content for distinguishing human-generated and AI-generated items. For example, in the fashion domain, designers' patents are essential. Therefore, it is necessary to distinguish between AI-generated, human-generated, and AI-assisted human-generated content. We can

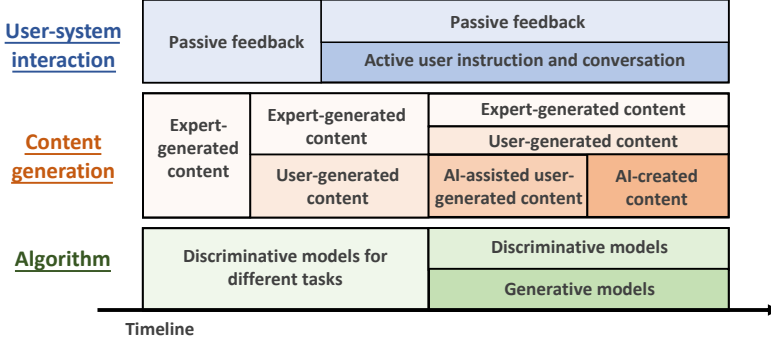


Fig. 3. A potential roadmap for the GeneRec paradigm from three aspects: user-system interaction, content generation, and recommender algorithm.

develop AI technologies to automatically identify AI-generated items [51]. Furthermore, we may consider deleting the AI-generated items after browsing by users to prevent them from being reused for inappropriate context.

• **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 [69] to measure micro-video quality), the relevance between the generated content and users' information needs, 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 with the generated content.

• **Roadmap.** In Figure 3, we outline the future development roadmap of GeneRec by examining three separate aspects: user-system interaction, content generation, and recommender algorithms.

- 1) **User-system interaction.** Users' passive feedback (e.g., clicks) actually contains many subtle preferences, which might even be difficult for users themselves to articulate. Historically, recommender systems have heavily relied on user features and such passive feedback to model user preference. Beyond passive feedback, we believe that future recommender systems will be able to engage in multimodal conversations with users through voice, visual, and textual interactions, much like the AI assistant Jarvis², to understand users' needs. By combining conversation with historical passive feedback, recommender systems can build more comprehensive user profiles, and subsequently, they can retrieve or generate content as recommendations to fulfill users' information needs.

Existing LLMs have already achieved proficient textual conversation and burgeoning capabilities in multimodal interactions [18], effectively meeting the requirements for barrier-free communication with users. In the next years, a direction worth exploring is how to enable LLMs to quickly acquire and comprehend the item content in the recommendation domain. The new item content is continually emerging and the item popularity is quickly shifting over time. By efficiently

²<https://en.wikipedia.org/wiki/J.A.R.V.I.S.>

understanding the item content, LLMs can better communicate with users and understand their information needs. Additionally, another direction to investigate in the coming years is how to combine active user conversation with passive feedback for more comprehensive user modeling.

- 2) **Content generation.** Content generation has three evolution phases. 1) Expert-generated professional content, such as carefully crafted movies and music. This type of content typically has high quality yet also comes with high production costs, making it challenging to generate large quantities quickly. 2) User-generated content, driven by the rise of micro-videos, has turned many users on the platform into content creators, greatly enriching content diversity and reach. However, user-generated content often varies widely in quality, with many items being of lower quality. 3) In the era of AIGC, AI can assist users in content creation and may even have the potential to independently generate new content. We can leverage AI to assist content creators and users themselves in content generation tasks, such as generating music for micro-videos or refining news articles. This helps improve the quality of user-generated content. During the phase of AI-assisted generation, users may still need to assume a significant portion of responsibility, including providing instructions and content design. If generative AI continues to advance, it is promising that generative AI can learn user preference directly from a vast amount of high-quality content and continuously create new content. In this way, AI-generated content will complement user-generated or expert-generated content to meet users' information needs.
- 3) **Algorithm.** Discriminative algorithms have traditionally been the mainstream approach for item rankings in recommendations, including but not limited to methods such as matrix factorization [59, 78], graph neural networks-based methods [10, 29], and transformer-based sequential algorithms [38, 80, 89]. Generative models such as variational autoencoders have some explorations [42, 73] while they receive relatively limited scrutiny. However, the recent emergence of LLMs has sparked a trend in using generative models for recommendation. Researchers are exploring the integration of various recommendation tasks, such as sequential recommendation, click-through rate prediction, recommendation explanations, and conversational recommendations, within a generative framework based on LLMs [23]. Furthermore, they are also delving into how to better leverage word knowledge within LLMs for recommendation purposes [44] and how to enable LLMs to effectively capture collaborative filtering signals in user-item interactions [88]. We believe that in the future, there will be a coexistence and mutual promotion of discriminative and generative models to accomplish various recommendation tasks. It is even possible that a unified LLM-based recommender model may emerge for various recommendation tasks, covering item retrieval, repurposing, and creation.

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 initializing the content generation and pre-processing user instructions, while the AI editor and AI creator implement the AI generator for personalized item editing and creation, respectively. Lastly, we present several application scenarios and potential research tasks.

3.1 Instructor

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

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

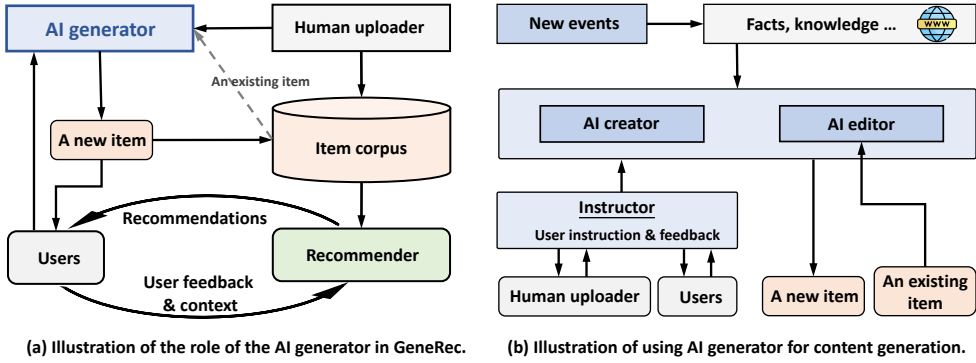


Fig. 4. A demonstration of GeneRec. Both users and human uploaders can interact with AI generators via instructions and feedback for content generation. The AI editor aims to repurpose existing items in the item corpus while the AI creator directly creates new items. The new item can be directly recommended to users or fed into the item corpus for item rankings.

- **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 Editor and AI Creator

To implement the 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(b), the AI editor intends to refine and repurpose existing items (generated by either humans or AI) in the item corpus according to the instructions and historical feedback from either human uploaders or users. Human uploaders such as video uploaders on YouTube can interact with the AI editor to generate different versions of items, which can be fed into the general item corpus for rankings. Users can also generate personalized items for themselves with the assistance of the AI editor.

- **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. Note that the "facts and knowledge" can be updated with new events and regulations continually. For instance, based on the daily events, the AI editor can help to revise and generate personalized news reports for users.

- **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 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. As illustrated in Figure 1, the AI creator may know users' preference for Jay Chou's songs from the user's historical feedback, determine the singer to Zhen Tian based on user instructions, and learn the singing skills of Zhen Tian on the Web to make a music video of "Worldly Tavern" performed by Zhen Tian.
- **Output:** A new item that fulfills users' information needs.

3.3 Domain Applications of GeneRec

High-quality AIGC via GeneRec is applicable in multiple domains. The related domains include but are not limited to the recommendations of textual content (e.g., news, articles, books, and financial reports), advertisements, videos (e.g., micro-videos, animation, and movies), images (e.g., art portraits and landscape), and audio (e.g., music). The users and human producers can leverage the AI generators to produce extensive items such as movies, music, advertisements, and micro-videos. Furthermore, it is also viable to harness AI generators for the design of various products, including fashion apparel and accessories, with the potential for a seamless transition into the manufacturing process.

In detail, we present several examples in different domains.

- **News recommendation.** Human news editors or users on the news platform can define rules by instructions for GeneRec to generate news. AI generators can also infer potential user preference by analyzing the feedback from editors' or users' historical interactions. Consequently, GeneRec can produce personalized news articles daily, taking into account newly occurring events. After news generation, fidelity checks are needed to detect misleading headlines, misinformation, satire, biased and deceptive news [90].
- **Fashion recommendation.** GeneRec can assist fashion designers in creating various versions of fashion products or directly generate personalized items for users. The passive feedback of fashion designers and users, such as clicks, plays a crucial role by providing AI generators with valuable insights into their implicit clothing preferences. AI-generated digital products with high value (e.g., popular products liked by lots of people) can be sent to factories for customized production. Currently, some fashion brands are exploring this direction, such as Tribute Brand³ and The Dematerialised⁴. In the fashion domain, fidelity checks should primarily focus on the realism of product details, ensuring image clarity, color accuracy, and high detail level while considering style consistency and compatibility.
- **Music recommendation.** AI generators have demonstrated their ability to produce high-quality music [83]. Under the GeneRec framework, to enhance the personalization of music generation, we should empower AI generators to learn users' implicit preferences for artists, lyrics, melodies, and singing styles from users' feedback. In terms of fidelity checks, music generation should address copyright and legality concerns, ensuring compliance with all relevant regulations.
- **Micro-video recommendation.** Micro-video generation within the GeneRec framework is significantly challenging, as it involves the generation of multiple modalities, including textual subtitles, cover images, videos, background music, or ambient sounds. Nevertheless, AI-assisted

³<https://www.tribute-brand.com/>.

⁴<https://thedematerialised.com/>.

user generation is a promising starting point. Extensive tools for AI-assisted content creation are emerging, encompassing features such as automatic music generation, cover image refinement, and subtitle generation. The field of AI-generated micro-videos is likely to transition progressively from AI-human collaboration toward AI-driven creation. Regarding fidelity checks, micro-video generation necessitates careful consideration of many aspects, including but not limited to bias, privacy, authenticity, safety, and realism.

3.4 Potential Research Tasks

Under the novel paradigm of GeneRec, there are various promising research tasks that deserve future exploration. We introduce some potential tasks as follows.

- 1) **Instruction tuning for LLM-based instructor.** One crucial factor in implementing the instructor of GeneRec is to enhance the instruction-following capabilities, allowing the instructor to comprehend user intention and take the right actions. These actions may involve generating responses, activating AI generators for content creation, or providing guidance for content generation. Existing work still cannot effectively achieve such an instructor. On one hand, conversational recommender models previously lack strong instruction-following capabilities, struggling to understand user intention. On the other hand, LLMs exhibit limited instruction-following abilities for item recommendations [4, 86]. This limitation arises from a lack of instruction tuning within the recommendation domain. Consequently, improving LLM-based instructors for GeneRec is an essential research task in the future.
- 2) **Controlling for AI generator activation.** Under GeneRec, an essential task is to control whether to activate AI generators. Besides, after generating a new item, it needs to decide between directly recommending the generated item to users and ranking it with existing items. These two decisions depend on the recommendation context and user behaviors, including users' instructions and feedback. For instance, if a user explicitly expresses a need for a generated item through instructions or consistently provides negative feedback on existing items, the AI generator should be initiated for content generation. In the future, collecting diverse user behaviors and required actions to train the LLM-based instructor is vital for enhancing its decision-making capabilities.
- 3) **Personalized item editing.** One of the core tasks within GeneRec is the implementation of AI editors to assist human uploaders or users in personalized content editing. As mentioned in Section 3.3, item editing should be explored separately in specific domains. While Generative AI has shown the ability to generate high-quality content in some domains, it is usually guided by user instructions. However, user instructions cannot describe various nuanced aspects of an item while crafting complex instructions is challenging and time-consuming. Furthermore, many human uploaders and users may have implicit preferences that they themselves may not be aware of. As such, extracting user interests from noisy and implicit user feedback is helpful, which aligns with the central focus of recommendation models in recent years. Hence, integrating users' implicit feedback into AI editors and effectively capturing users' implicit preference for personalized item editing holds great promise.
- 4) **Personalized item creation.** Going beyond personalized item editing, personalized item creation represents a more challenging task. This endeavor can begin with relatively easier domains, such as news and music recommendations, where generative AI technologies are relatively mature.
- 5) **Domain-specific fidelity checks.** As discussed in Section 3.3, implementing GeneRec in various domains requires careful consideration of domain-specific fidelity checks. It is of paramount importance to design fidelity evaluators tailored to specific domains. For instance, in news generation, special attention should be given to aspects like news authenticity and bias.

4 DISCUSSION

In this section, we present the comparison between the GeneRec paradigm and two related tasks: conversational recommendation and traditional AI generation. Moreover, we present a possible vision for future developments of GeneRec.

4.1 Comparison

We illustrate how GeneRec differs from conversational recommendation and traditional AI generation tasks.

4.1.1 Comparison with conversational recommendation. Conversational recommender systems rely on multi-turn natural language conversations with users to acquire user preference and provide recommendations [55, 65, 87]. Although conversational recommender systems also consider multi-turn conversations, we highlight two critical differences with the GeneRec paradigm: 1) Previous conversational recommender models lack instruction-following abilities, leading to poor user experience. The dramatic development of LLMs, especially ChatGPT, has brought a revolution to traditional conversational systems by significantly improving language understanding and generative abilities. Based on LLMs, we can build a more powerful instructor for GeneRec. And 2) 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.

4.1.2 Comparison with traditional content generation. There exist various cross-modal AIGC tasks such as text, image, and audio generation conditional on users' images [81], single-turn queries [61], and multi-turn conversations [37]. Nevertheless, there are essential differences between traditional AIGC tasks and GeneRec. 1) GeneRec can utilize the user modeling techniques from traditional recommender systems to capture implicit user preference. For example, GeneRec may leverage user feedback such as clicks and dwell time to dig out the implicit user preference that is not indicated through user instructions. 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. Learning implicit user preference from user features and behaviors, including long-term, short-term, noisy, and implicit user feedback, has been the core research in the recommendation domain in the past years. Many prior techniques in the recommendation domain can be used to exploit implicit user preference for GeneRec. In this case, GeneRec can consider both users' explicit instructions and implicit preference 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, if an emergency event occurs, journalists can send the latest video reports from the scene; besides, many human content producers have unique experiences or creativity that are difficult to replicate by generative AI. Therefore, compared to previous generative AI, GeneRec considers the cooperation between AI-generated and human-generated content to meet user information needs.

4.2 A vision for future GeneRec.

In the future, generative AI might be incorporated into various information platforms to supplement human-generated content. It might start with using AI generators to assist human content producers in repurposing existing items or creating new items. With the technical advancement of AI generators, they may gradually begin to independently generate content in some simple scenarios. Moreover, from a technical view, constrained by the development of existing AIGC technologies, we need to design different modules to achieve the generation tasks across multiple modalities.

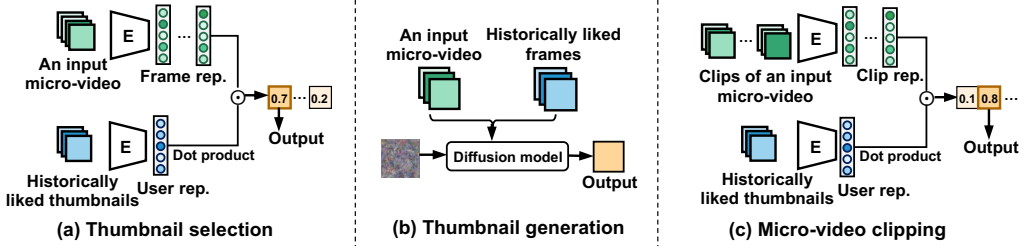


Fig. 5. Illustration of the implementation of editing tasks, including the process of AI editors for thumbnail selection, thumbnail generation, and micro-video clipping.

However, we believe that building a unified AI model for multimodal content generation is a feasible and worthwhile endeavor[40, 79]. Under the GeneRec paradigm, we expect to see the increasing maturity of 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.

5 FEASIBILITY STUDY

To investigate the feasibility of instantiating GeneRec, we employ AIGC methods to implement some simple demos of the AI editor and AI creator in micro-video application scenario due to the widespread of micro-video content. The instructor could be a ChatGPT-like interactive conversational user interface [22], an option-based interface for the user to choose “like” or “dislike” [85], or a recommender model that collects the user interactions. The obtained user instructions or user feedback will then be preprocessed to guide the AI editor or AI creator to repurpose or create the micro-video content. In our experiments, we mainly focus on implementing the AI editor and AI creator so we simply use a recommender model MMGCN [76] as the instructor and obtain the user’s historical interactions or the user embeddings from the well-trained MMGCN as the guidance.

- **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 [69] and each pre-processed micro-video has 400 frames with 64×64 resolution.

5.1 AI Editor

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

5.1.1 Thumbnail selection and generation. Considering that personalized thumbnails might better attract users to click on micro-videos [47], 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.

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

• **Implementation.** To implement personalized thumbnail selection, we utilize the image encoder $f_\theta(\cdot)$ of a representative Contrastive Language Image Pre-training model (CLIP) [56] to conduct zero-shot selection. As illustrated in Figure 5(a), given a set of N frames $\{v_i\}_{i=1}^N$ from a micro-video, and the set of M thumbnails $\{t_i\}_{i=1}^M$ from a user’s historically liked micro-videos, we calculate

$$\begin{cases} \mathbf{t}^* = \frac{1}{M} \sum_{i=1}^M f_\theta(\mathbf{t}_i), \\ j = \arg \max_{i \in \{1, \dots, N\}} \{\mathbf{t}^{*T} \cdot f_\theta(\mathbf{v}_i)\}, \end{cases} \quad (1)$$

where \mathbf{t}^* is the average representation of $\{f_\theta(\mathbf{t}_i)\}_{i=1}^M$, and we select the j -th frame as the recommended thumbnail due to the highest dot product score between the user representation \mathbf{t}^* and the j -th frame representation $f_\theta(\mathbf{v}_j)$. 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) [5], 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 [76] 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.

• **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

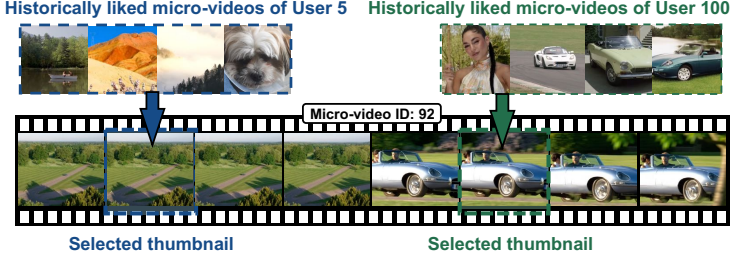


Fig. 6. Cases of personalized thumbnail selection by CLIP.

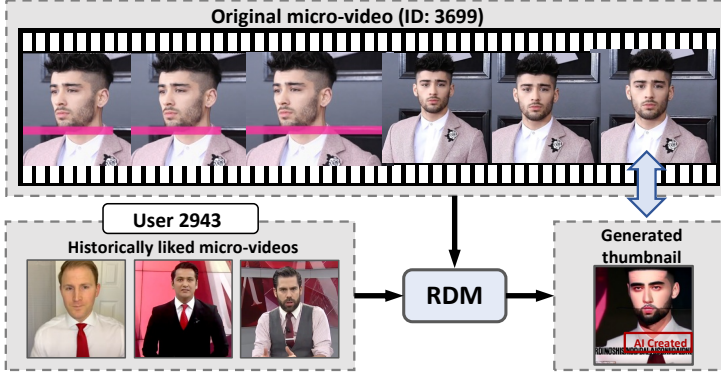


Fig. 7. Cases of personalized thumbnail generation.

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.

5.1.2 Micro-video clipping. Given a long micro-video (e.g., one 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' browsing experience [49].

• **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 [56] to obtain personalized micro-video clips as shown in Figure 5(c). Given user representation t^* obtained from M historically liked thumbnails via Eq. (1), and a set of C clips $\{c_i\}_{i=1}^C$ where each clip c_i has L consecutive frames⁵ $\{v_a^i\}_{a=1}^L$, we compute

⁵The 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 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

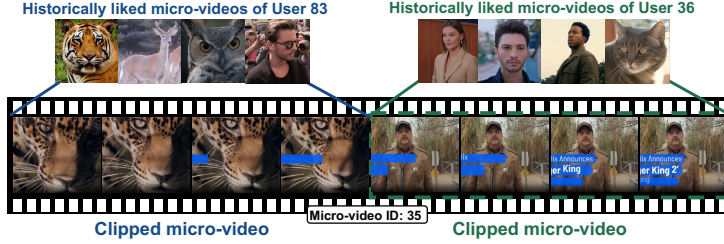


Fig. 8. Case study of micro-video clipping via CLIP [56].

$$\begin{cases} \mathbf{c}_i^* = \frac{1}{L} \sum_{a=1}^L f_{\theta}(v_a^i), \\ j = \arg \max_{i \in \{1, \dots, C\}} \{\mathbf{t}^{*T} \cdot \mathbf{c}_i^*\}, \end{cases} \quad (2)$$

where \mathbf{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 \mathbf{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.

5.1.3 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.

• **Task.** Given an existing micro-video in the corpus, user instructions, and user feedback⁶, 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 VToonify⁷ [82] to achieve the style transfer; and 2)

⁶We do not consider the “facts and knowledge” in Section 3.2.1 to simplify the implementation, leaving the knowledge-enhanced implementation to future work.

⁷<https://github.com/williamyang1991/vtoonify/>.

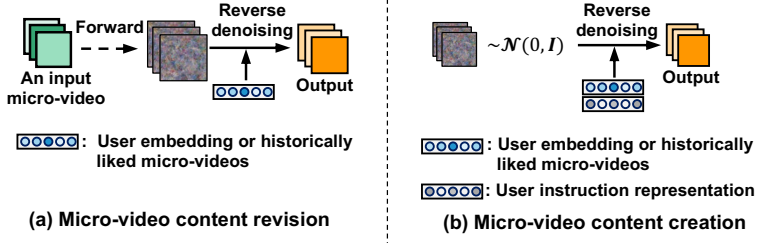


Fig. 9. Illustration of the implementation of editing and creation tasks. (a) depicts the procedure of AI editor for micro-video content revision and (b) shows the process of AI creator for micro-video content creation.



Fig. 10. Examples of personalized micro-video style transfer via VToonify.

micro-video content revision based on user feedback. We resort to a newly published Masked Conditional Video Diffusion model (MCVD) [69] for micro-video revision. The revision process is presented in Figure 9(a). 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 [69] and adopt the widely used FVD metric [67], which measures the distribution gap between the real micro-videos and generated micro-videos. Specifically, FVD builds on the principles underlying Frechet Inception Distance (FID [31]) and additionally considers the temporal coherence for videos. The FVD metric is formally written as:

$$\text{FVD} = |\mu_R - \mu_G|^2 + \text{Tr} \left(\Sigma_R + \Sigma_G - 2(\Sigma_R \Sigma_G)^{\frac{1}{2}} \right), \quad (3)$$

where μ_R and μ_G are the means and Σ_R and Σ_G are the co-variance matrices of the distribution of feature representation of real-world videos and the generated videos. The feature representations are obtained by the pre-trained Inflated 3D Convnet (I3D [9]), which considers the temporal coherence of the visual content across a sequence of frames. A lower FVD score indicates higher quality.

• **Results.** We show some cases of micro-video style transfer in Figure 10. The same micro-video is transferred into different styles according to users' instructions. From Figure 10, we can observe that the repurposed videos show high fidelity and quality, validating that the task of micro-video

Table 3. Quantitative results of MCVD based on User_Hist and User_Emb. As a reference, 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



Fig. 11. Case study of personalized micro-video content revision via MCVD (User_Emb).

style transfer can be well accomplished by existing generative models. However, it might be costly for 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 11, 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.

5.2 AI Creator

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



Fig. 12. Examples of single-turn instructions for text-to-image generation via stable diffusion.

Table 4. Quantitative results of MCVD based on User_Hist and User_Emb. As a reference, 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

5.2.1 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 instructions, where we construct users' single-turn instructions and apply stable diffusion [61] for image synthesis. From the generated images in Figure 12, 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 9(b), 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 [56]. 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@K⁸ 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 from "User_Emb" and "User_Hist" is worse than the unconditional creation as shown by the larger FVD scores. The case study in Figure 13(a) 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, evident in the blurred rendering of the mouth and the generated hair extending to the chin region (see Figure 13(b)). And the generated portrait for User 38 is also blurred and distorted as shown in Figure 13(a) and (b). We can find that the collar and coat are conjoined, and the left arm remains absent in the generated content.

⁸We cannot calculate PS@K via MMGCN because the newly created items do not have item ID embeddings, which are necessary for MMGCN prediction.

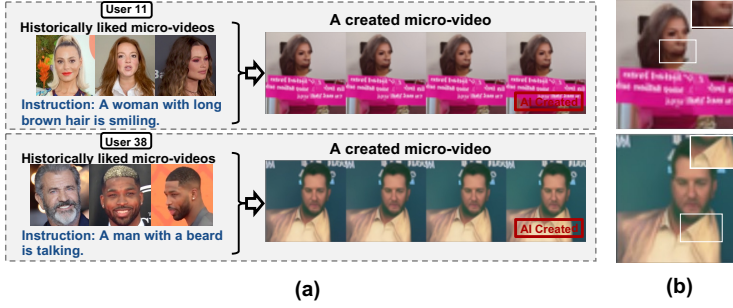


Fig. 13. Case study of personalized micro-video content creation via MCVD (User_Emb).

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 three aspects: 1) enabling the powerful ChatGPT-like tools⁹ to implement the instructor and acquire detailed user instructions; 2) pursuing more comprehensive user modeling and better encoding users' implicit preference to AI generators; and 3) using more advanced algorithms of generative AI with strong prior world knowledge through pre-training on more data in different modalities, such as image (e.g., Midjourney), video (e.g., PikaLab), and audio (e.g., AudioCraft). We believe that the development of video generation might also follow the trajectory of image generation (see Figure 12) and eventually achieve satisfactory generation results.

Moreover, we emphasize the importance of addressing copyright issues regarding AI-generated content. For GeneRec in the micro-video domain, we offer potential solutions from two perspectives: 1) authorship belongs to the platform when the micro-video is fully generated by the AI creator; 2) the uploaders hold the authorship of the generated micro-video if the generated content is edited from the uploader-created content. Nevertheless, the policy for tackling copyright violations needs formal regulations, especially from the government's view.

• **Running example of GeneRec on micro-video recommendation.** To run the whole GeneRec paradigm for micro-video recommendation, it involves the steps as follows: 1) A user watches the recommended micro-videos, or directly searches or expresses what they prefer to watch (i.e., information needs) at the moment; 2) Instructor with conversational interaction interface acquires the user's information needs and pre-processes instructions. 3) Based on the pre-processed user's instructions and historical feedback, the instructor will decide whether to use the AI editor or the AI creator for generating personalized micro-videos. For example, if the user requests the instructor to provide creative AI-generated content, the AI creator will then be called to generate a new micro-video as the response. 4) Post-processing of the micro-video such as the checks of quality, relevance, and fidelity will then be conducted. 5) Lastly, the instructor will decide to directly recommend the AI-generated micro-video to the user, or rank the AI-generated micro-video with all existing micro-videos in the item corpus. 6) Provide the final recommendation and go back to step 1) or 2) based on the user's information needs.

⁹In this work, we do not explore the usage of ChatGPT because it was recently released and we have insufficient time for comprehensive exploration.

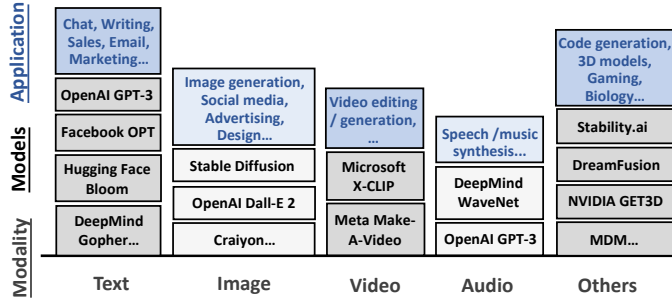


Fig. 14. Illustration of advanced models and potential applications of AIGC across different modalities.

6 RELATED WORK

• **Recommender Systems.** The traditional retrieval-based recommender paradigm constructs a loop between the recommender models and users [58, 91]. Recommender models rank the items in a corpus and recommend the top-ranked items to the users [36, 84]. 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 [19, 30, 39]. Early approaches directly utilize collaborative behaviors of similar users (*i.e.*, user-based CF) or items (*i.e.*, item-based CF). Later on, MF [59] decomposes the interaction matrix into user and item matrices separately, laying the foundation for subsequent neural CF methods [30] and graph-based CF methods [29]. Beyond purely using user-item interactions, prior work considers incorporating context [60] and the content features of users and items [28, 76] for session-based, sequential, and multimedia recommendations [15, 32, 38]. In recent years, various new recommender frameworks have been proposed, such as conversational recommender systems [65, 66] acquiring user preference via conversations and user-controllable recommender systems [72] for controlling the attributes of recommended items. Recently, some researchers have considered using large language models for recommendations [4, 12, 35, 41, 45] while they mainly enhance recommender algorithms instead of producing content through generative AI or revolutionizing the user-system interaction interfaces.

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 to meet users' information needs and present a novel generative paradigm for next-generation recommender systems. Satisfying personalized information needs has been studied in the Human-Computer Interaction (HCI) domain [2]. While studies from HCI mainly focus on user modeling based on device- and context-specific information [70], we emphasize the harness of ChatGPT-like models for advanced user-system interactions, supplementing traditional user feedback with better user engagement.

• **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 14. 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 [16, 52, 64] as well as high-coherence videos [33, 34, 69]. Besides, previous work has explored audio synthesis [57]. For instance, [17] proposes a generative model that aligns text with waves, endowing high-fidelity text-to-speech generation. Furthermore, extensive generative models are crafted for other fields, such as 3D generation [26], gaming [14], and chemistry [54].

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 roadmap, application scenarios, and future research tasks of GeneRec. We explored the feasibility of implementing GeneRec on micro-video generation and the experiments reveal some weaknesses and promising results of existing AIGC methods 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 integrate 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 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] Ofer Arazy, Oded Nov, and Nanda Kumar. 2015. Personalization: UI personalization, theoretical grounding in HCI and design research. *THCI* 7, 2 (2015), 43–69.
- [3] Ricardo Baeza-Yates. 2020. Bias in Search and Recommender Systems. In *RecSys*. ACM, 2–2.
- [4] Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. 2023. Tallrec: An effective and efficient tuning framework to align large language model with recommendation. In *RecSys*. ACM.
- [5] Andreas Blattmann, Robin Rombach, Kaan Oktay, Jonas Müller, and Björn Ommer. 2022. Semi-Parametric Neural Image Synthesis. In *NeurIPS*. Curran Associates, Inc.
- [6] Tim Brooks, Aleksander Holynski, and Alexei A Efros. 2023. Instructpix2pix: Learning to follow image editing instructions. In *CVPR*. 18392–18402.

- [7] 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.
- [8] Miriam C Buiten. 2019. Towards Intelligent Regulation of Artificial Intelligence. *Eur J Risk Regul* 10, 1 (2019), 41–59.
- [9] Joao Carreira and Andrew Zisserman. 2017. Quo vadis, action recognition? a new model and the kinetics dataset. In *CVPR*. 6299–6308.
- [10] Jianxin Chang, Chen Gao, Yu Zheng, Yiqun Hui, Yanan Niu, Yang Song, Depeng Jin, and Yong Li. 2021. Sequential recommendation with graph neural networks. In *SIGIR*. 378–387.
- [11] Paul-Alexandru Chirita, Wolfgang Nejdl, and Cristian Zamfir. 2005. Preventing Shilling Attacks in Online Recommender Systems. In *WIDM*. ACM, 67–74.
- [12] Sunhao Dai, Ninglu Shao, Haiyuan Zhao, Weijie Yu, Zihua Si, Chen Xu, Zhongxiang Sun, Xiao Zhang, and Jun Xu. 2023. Uncovering ChatGPT’s Capabilities in Recommender Systems. *arXiv:2305.02182*.
- [13] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. 2010. The YouTube video recommendation system. In *RecSys*. 293–296.
- [14] Rafael Guerra de Pontes and Herman Martins Gomes. 2020. Evolutionary Procedural Content Generation for An Endless Platform Game. In *SBGames*. IEEE, 80–89.
- [15] 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.
- [16] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion Models Beat Gans on Image Synthesis. In *NeurIPS*. Curran Associates, Inc., 8780–8794.
- [17] Jeff Donahue, Sander Dieleman, Mikołaj Bińkowski, Erich Elsen, and Karen Simonyan. 2021. End-to-end Adversarial Text-to-speech. In *ICLR*.
- [18] Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. 2023. Palm-e: An embodied multimodal language model. *arXiv:2303.03378*.
- [19] Travis Ebesu, Bin Shen, and Yi Fang. 2018. Collaborative Memory Network for Recommendation Systems. In *SIGIR*. ACM, 515–524.
- [20] 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.
- [21] Ruoyuan Gao and Chirag Shah. 2020. Counteracting Bias and Increasing Fairness in Search and Recommender Systems. In *RecSys*. ACM, 745–747.
- [22] Yunfan Gao, Tao Sheng, Youlin Xiang, Yun Xiong, Haofen Wang, and Jiawei Zhang. 2023. Chat-rec: Towards interactive and explainable llms-augmented recommender system. *arXiv:2303.14524*.
- [23] Shijie Geng, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2022. Recommendation as language processing (rlp): A unified pretrain, personalized prompt & predict paradigm (p5). In *RecSys*. ACM, 299–315.
- [24] Carlos A Gomez-Urbe and Neil Hunt. 2015. The netflix recommender system: Algorithms, business value, and innovation. *TMS* 6, 4 (2015), 1–19.
- [25] Ihsan Gunes, Cihan Kaleli, Alper Bilge, and Huseyin Polat. 2014. Shilling Attacks Against Recommender Systems: A Comprehensive Survey. *Artificial Intelligence Review* 42, 4 (2014).
- [26] 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.
- [27] Danhuai Guo, Huixuan Chen, Ruoling Wu, and Yangang Wang. 2023. AIGC challenges and opportunities related to public safety: a case study of ChatGPT. *Journal of Safety Science and Resilience* 4, 4 (2023), 329–339.
- [28] 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.
- [29] 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.
- [30] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *WWW*. ACM, 173–182.
- [31] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *NeurIPS* 30 (2017).
- [32] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *ICLR*.
- [33] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. 2022. Video Diffusion Models. *arXiv:2204.03458*.
- [34] 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*.

- [35] Yupeng Hou, Junjie Zhang, Zihan Lin, Hongyu Lu, Ruobing Xie, Julian McAuley, and Wayne Xin Zhao. 2023. Large language models are zero-shot rankers for recommender systems. *arXiv:2305.08845*.
- [36] Chenhao Hu, Shuhua Huang, Yansen Zhang, and Yubao Liu. 2022. Learning to Infer User Implicit Preference in Conversational Recommendation. In *SIGIR*. ACM, 256–266.
- [37] 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.
- [38] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *ICDM*. IEEE, 197–206.
- [39] Ioannis Konstas, Vassilios Stathopoulos, and Joemon M Jose. 2009. On Social Networks and Collaborative Recommendation. In *SIGIR*. ACM, 195–202.
- [40] Chunyuan Li, Zhe Gan, Zhengyuan Yang, Jianwei Yang, Linjie Li, Lijuan Wang, and Jianfeng Gao. 2023. Multimodal foundation models: From specialists to general-purpose assistants. *arXiv:2309.10020* 1, 2 (2023), 2.
- [41] Jinming Li, Wentao Zhang, Tian Wang, Guanglei Xiong, Alan Lu, and Gerard Medioni. 2023. GPT4Rec: A Generative Framework for Personalized Recommendation and User Interests Interpretation. *arXiv preprint arXiv:2304.03879* (2023).
- [42] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. In *WWW*. ACM, 689–698.
- [43] Yu Liang and Martijn C Willemsen. 2023. Promoting music exploration through personalized nudging in a genre exploration recommender. *International Journal of Human–Computer Interaction* 39, 7 (2023), 1495–1518.
- [44] Xinyu Lin, Wenjie Wang, Yongqi Li, Fuli Feng, See-Kiong Ng, and Tat-Seng Chua. 2023. A Multi-facet Paradigm to Bridge Large Language Model and Recommendation. *arXiv:2310.06491* (2023).
- [45] Junling Liu, Chao Liu, Renjie Lv, Kang Zhou, and Yan Zhang. 2023. Is ChatGPT a Good Recommender? A Preliminary Study. *arXiv preprint arXiv:2304.10149* (2023).
- [46] Nathan N Liu, Evan W Xiang, Min Zhao, and Qiang Yang. 2010. Unifying explicit and implicit feedback for collaborative filtering. In *CIKM*. 1445–1448.
- [47] 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.
- [48] Nicola Lucchi. 2023. ChatGPT: a case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation* (2023), 1–23.
- [49] 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.
- [50] 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.
- [51] 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*.
- [52] Alexander Quinn Nichol and Prafulla Dhariwal. 2021. Improved Denoising Diffusion Probabilistic Models. In *ICML*. PMLR, 8162–8171.
- [53] 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*.
- [54] 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.
- [55] 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.
- [56] 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.
- [57] 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.
- [58] 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.
- [59] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI*. AUAI Press, 452–461.
- [60] Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2011. Fast Context-aware Recommendations with Factorization Machines. In *SIGIR*. ACM, 635–644.

- [61] 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.
- [62] 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.
- [63] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiuyan Hu, Harry Yang, Oron Ashual, Oran Gafni, et al. 2022. Make-a-video: Text-to-video generation without text-video data. *arXiv:2209.14792*.
- [64] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2021. Denoising Diffusion Implicit Models. In *ICLR*.
- [65] Yueming Sun and Yi Zhang. 2018. Conversational Recommender System. In *SIGIR*. ACM, 235–244.
- [66] 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.
- [67] Thomas Unterthiner, Sjoerd van Steenkiste, Karol Kurach, Raphaël Marinier, Marcin Michalski, and Sylvain Gelly. 2019. FVD: A new metric for video generation. (2019).
- [68] Ron G Van Schyndel, Andrew Z Tirkel, and Charles F Osborne. 1994. A Digital Watermark. In *ICIP*. IEEE, 86–90.
- [69] Vikram Voleti, Alexia Jolicoeur-Martineau, and Christopher Pal. 2022. MCVD: Masked Conditional Video Diffusion for Prediction, Generation, and Interpolation. In *NeurIPS*. Curran Associates, Inc.
- [70] Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Quay Au, Bernd Bischl, Markus Bühner, and Heinrich Hussmann. 2019. Opportunities and challenges of utilizing personality traits for personalization in HCI. *Personalized Human-Computer Interaction* 31 (2019).
- [71] Tao Wang, Yushu Zhang, Shuren Qi, Ruoyu Zhao, Zhihua Xia, and Jian Weng. 2023. Security and privacy on generative data in aigc: A survey. *arXiv:2309.09435* (2023).
- [72] Wenjie Wang, Fuli Feng, Liqiang Nie, and Tat-Seng Chua. 2022. User-controllable Recommendation Against Filter Bubbles. In *SIGIR*. ACM, 1251–1261.
- [73] Wenjie Wang, Xinyu Lin, Fuli Feng, Xiangnan He, Min Lin, and Tat-Seng Chua. 2022. Causal Representation Learning for Out-of-Distribution Recommendation. In *WWW*. ACM, 3562–3571.
- [74] 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*.
- [75] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022. Emergent abilities of large language models. *arXiv:2206.07682* (2022).
- [76] 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.
- [77] Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, and Hong Lin. 2023. Ai-generated content (aigc): A survey. *arXiv:2304.06632* (2023).
- [78] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. 2021. Self-supervised graph learning for recommendation. In *SIGIR*. 726–735.
- [79] Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. 2023. Next-gpt: Any-to-any multimodal llm. *arXiv:2309.05519* (2023).
- [80] Zhe Xie, Chengxuan Liu, Yichi Zhang, Hongtao Lu, Dong Wang, and Yue Ding. 2021. Adversarial and contrastive variational autoencoder for sequential recommendation. In *WWW*. ACM, 449–459.
- [81] 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.
- [82] 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.
- [83] Dingyao Yu, Kaitao Song, Peiling Lu, Tianyu He, Xu Tan, Wei Ye, Shikun Zhang, and Jiang Bian. 2023. MusicAgent: An AI Agent for Music Understanding and Generation with Large Language Models. *arXiv:2310.11954*.
- [84] 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.
- [85] Bo Zhang and S Shyam Sundar. 2019. Proactive vs. reactive personalization: Can customization of privacy enhance user experience? *Int. J. Hum. Comput.* 128 (2019), 86–99.
- [86] Junjie Zhang, Ruobing Xie, Yupeng Hou, Wayne Xin Zhao, Leyu Lin, and Ji-Rong Wen. 2023. Recommendation as instruction following: A large language model empowered recommendation approach. *arXiv:2305.07001*.
- [87] 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.
- [88] Yang Zhang, Fuli Feng, Jizhi Zhang, Keqin Bao, Qifan Wang, and Xiangnan He. 2023. CoLLM: Integrating Collaborative Embeddings into Large Language Models for Recommendation. *arXiv:2310.19488*.
- [89] Kun Zhou, Hui Wang, Wayne Xin Zhao, Yutao Zhu, Sirui Wang, Fuzheng Zhang, Zhongyuan Wang, and Ji-Rong Wen. 2020. S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization. In *CIKM*. 1893–1902.

- [90] Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *CSUR* 53, 5 (2020), 1–40.
- [91] 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.