RESEARCH-ARTICLE

Generative Recommendation: Towards Personalized Multimodal Content Generation

WENJIE WANG, National University of Singapore, Singapore City, Singapore

XINYU LIN, National University of Singapore, Singapore City, Singapore

FULI FENG, University of Science and Technology of China, Hefei, Anhui, China

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

TAT-SENG CHUA, National University of Singapore, Singapore City, Singapore

Open Access Support provided by:

University of Science and Technology of China

National University of Singapore

Published: 08 May 2025

Citation in BibTeX format

WWW '25: The ACM Web Conference 2025
April 28 - May 2, 2025
Sydney NSW, Australia

Conference Sponsors:
SIGWEB

WWW '25: Companion Proceedings of the ACM on Web Conference 2025 (May 2025)
https://doi.org/10.1145/3701716.3717529
ISBN: 9798400713316

Generative Recommendation: Towards Personalized Multimodal Content Generation

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
Hefei, China

Xiangnan He
xiangnanhe@gmail.com
University of Science and Technology of China
Hefei, China

Tat-Seng Chua
dcscts@nus.edu.sg
National University of Singapore
Singapore

Abstract

Traditional recommender systems primarily retrieve items from a pre-defined item corpus for personalized recommendations. However, this a retrieval-based paradigm faces two major constraints: 1) the existing item corpus with human-generated items may not adequately satisfy users' diverse information needs, and 2) users often rely on passive and noisy feedback (*e.g.,* clicks) to refine the recommendations. To overcome the limitations, the rapid advancement of AI-Generated Content (AIGC) presents significant potential: 1) generative AI promotes the creation of personalized multimodal content as new items to satisfy users' information needs, and 2) large language models reduce the efforts of users to express information needs proactively in natural language.

In this light, we propose a novel **Gene**rative **Rec**ommender paradigm named GeneRec with two objectives: 1) generating personalized multimodal content, and 2) integrating proactive user instructions to guide content generation. To achieve the objectives, GeneRec introduces an AI generator to personalize multimodal content generation and leverages user instructions to obtain users' information needs. Specifically, GeneRec first employs an instructor to pre-process users' instructions and traditional feedback (*e.g.,* clicks) to extract generation guidance. Following the guidance, we develop the AI generator with an AI editor to edit existing items and an AI creator to create new items, respectively. Lastly, we study the feasibility of implementing the AI creator in the fashion domain, showing promising results. Furthermore, to ensure the trustworthiness of the generated items, we emphasize various trustworthiness checks such as fairness and safety checks.

CCS Concepts

• **Information systems → Recommender systems**.

Keywords

Generative Recommender Paradigm, Personalized Multimodal Content Generation, Personalized AIGC

ACM ISBN 979-8-4007-1331-6/2025/04
https://doi.org/10.1145/3701716.3717529

ACM Reference Format:
Wenjie Wang, Xinyu Lin, Fuli Feng, Xiangnan He, and Tat-Seng Chua. 2025. Generative Recommendation: Towards Personalized Multimodal Content Generation. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3701716.3717529

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 user-generated micro-videos on Tiktok. However, advanced generative models such as diffusion models [14] empower the generation of high-quality multimodal AI-Generated Content (AIGC), exhibiting significant potential in various domains. As in Figure 1, generative AI is capable of producing multimodal content, including video, text, and audio to generate new items. With the surge of AIGC, recommender systems must move beyond human-generated content, by envisioning a generative recommender paradigm towards personalized multimodal content generation.

Before exploring the generative recommender paradigm, it is essential to revisit the traditional retrieval-based recommender framework. The traditional paradigm ranks human-generated items within a predefined corpus and recommends the highest-ranked items to users. It then collects the user feedback (*e.g.,* clicks) and context (*e.g.,* interaction duration) to continuously optimize the model for future recommendations [4]. Nevertheless, such a traditional paradigm encounters two inherent limitations. 1) The predefined corpus may be insufficient to accommodate users' diverse and personalized content preferences [15, 20]. For instance, a user may prefer a music video performed by a specific singer in a unique style (see Figure 1), yet manually creating such multimodal content is either impractical or costly [15, 18]. And 2) current recommender systems rely primarily on passive user feedback (*e.g.,* clicks), which constrains users from effectively and explicitly conveying their preferences [11, 12].

Fortunately, generative AI offers the potential to overcome the limitations of the traditional paradigm. To elaborate, not only can generative AI edit existing items or create new items with personalized multimodal content, the powerful generative AI such as LLMs can also significantly lessen the effort required

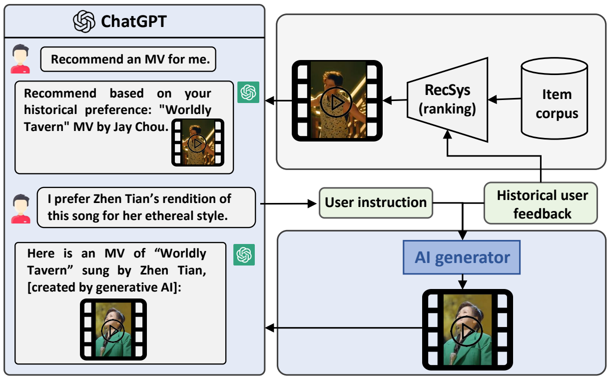


Figure 1: An example of using generative AI to interact with users and generate multimodal content as a new item.

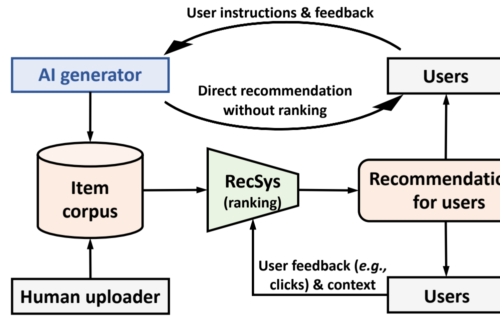


Figure 2: Illustration of GeneRec. AI generator takes user instructions and feedback to generate personalized multi-modal content, which can be directly recommended to users or fed to the item corpus for item ranking.

from users to articulate their diverse information needs using natural language instructions (Figure 1). In this light, we propose a novel **Gene**rative **Rec**ommender paradigm named **GeneRec** with two primary objectives: 1) personalized multimodal content generation, and 2) the integration of proactive user instructions and user feedback. To pursue the two objectives, GeneRec integrates the powerful generative AI for personalized multimodal content generation, including both content editing and creation. To pursue the two objectives, GeneRec adds a loop between *an AI generator* and *users* as shown in Figure 2. Precisely, the AI generator processes the user instructions in multimodal conversations and feedback to interpret the user's information needs and generate personalized multimodal content. The generated content can be added to the item corpus for ranking or recommended directly to users.

To instantiate GeneRec, we design an instructor module to process user instructions, along with two additional modules for item editing and creation. Initially, the instructor pre-processes and encodes these instructions along with user feedback to guide the generation process. Based on the guidance, an AI editor revises an existing item to align with user's specific preference (*i.e.,* personalized item editing), while an AI creator generates entirely new items (*i.e.,* personalized item creation). To explore the feasibility of GeneRec, we devise a diffusion model with user instructions and feedback as inputs to implement the AI creator in the fashion recommendation domain. Empirical results show GeneRec's promising performance in personalized content creation. Moreover, to ensure the trustworthiness and high quality of generated items, we highlight evaluation and the importance of various trustworthiness checks [10, 17, 18].

2 Generative 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 and regular users, generate and upload items to the item corpus. These items are then ranked for recommendations according to the user preference learned from the context (*e.g.,* user environments) and user feedback (*e.g.,* clicks) over historical recommendations.

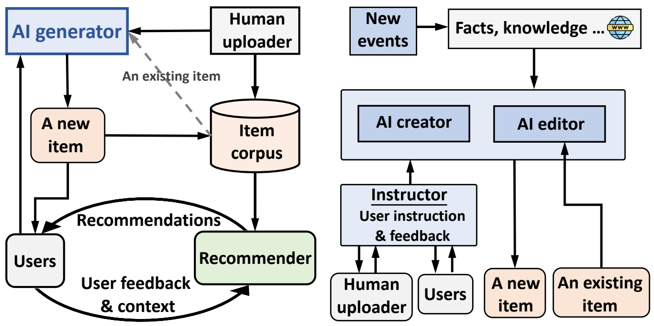
To complement this traditional paradigm, GeneRec adds another loop between the AI generator and users. Users can control the multimodal content generated by the AI generator through user instructions and feedback. Thereafter, the generated content as new items can be directly exposed to users without ranking.

• **User instructions.** The strong conversational ability of ChatGPT-like LLMs can enrich the interaction modes between users and the AI generator via conversational instructions. 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.

• **Multimodal content generation.** 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 subtle preference from users' historical interactions. The AI generator learns personalized information needs from user instructions and feedback, and then generates personalized content accordingly.

The generation includes both ***editing*** existing items and ***creating*** new items from scratch. For example, to edit a 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 video based on user instructions and Web data (*e.g.,* facts and knowledge). In addition, ***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 videos.

• **Trustworthiness checks.** To ensure the generated content is accurate, fair, and safe, GeneRec should pass the trustworthiness checks, including but not limited to the following angles. 1) **Bias and fairness**: the AI generator might learn from biased data [2], and thus should confirm that the generated multimodal content does not perpetuate stereotypes, promote hate speech and discrimination, cause unfairness to certain populations, or reinforce other harmful biases [6, 7, 21]. 2) **Privacy**: the generated content cannot disseminate any sensitive or personal information that may violate someone's privacy [16]. Many recommendation scenarios (*e.g.,* news and tweets) are sensitive to privacy leakage. 3) **Safety**: the AI generator must not pose any risks of harm to users, including physical and psychological harm [1]. For instance, the generated content for teenagers should not contain any unhealthy information.



(a) Illustration of the role of the AI generator in GeneRec.

(b) Illustration of using AI generator for content generation.

Figure 3: A demonstration of GeneRec. Both users and human uploaders can interact with AI generators via instructions and feedback for multimodal content generation. The AI editor aims to edit 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.

Besides, it is crucial to prevent GeneRec from various attacks such as shilling attack [3, 9].

• **Evaluation.** To evaluate the generated multimodal content, we propose two kinds of evaluation setups. 1) ***Item-side evaluation*** emphasizes the measurements from the item itself, including the item quality measurements, the similarity between generated items and user history, and trustworthiness checks. 2) ***User-side evaluation*** judges the quality of generated content based on users' satisfaction. The satisfaction can be collected either by explicit feedback (*e.g.,* reviews) or implicit feedback (*e.g.,* clicks) like in the traditional retrieval-based recommender paradigm.

3 Demonstration

To instantiate GeneRec, we develop three modules: an instructor, an AI editor, and an AI creator, for content generation initialization, implemented as AI generator as described in Figure 3.

3.1 Instructor

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

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

• **Processing:** Given the inputs, the instructor may need to engage in multi-turn interactions. It then analyzes the 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 extract guidance signals and enable the AI generator for personalized generation.

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

3.2 AI Generator

To implement personalized multimodal content generation, we propose two modules: an AI editor and an AI creator.

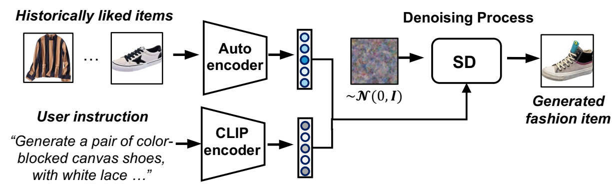


Figure 4: Illustration of implementing SD as AI creator.

3.2.1 *AI editor for personalized item editing.*

As depicted in Figure 3(b), the AI editor intends to refine and edit existing items (generated by either humans or AI) in the item corpus according to the instructions and historical feedback, assisting human uploaders or users to generate personalized content.

• **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 retrieved from the Web data.

• **Processing:** Given the input data, the AI editor leverages powerful neural networks to learn the users' information needs, and then edit the input item accordingly. It can also retrieve "facts and knowledge" from Web data to acquire factual events, generation skills, common knowledge, laws, and regulations to help generate accurate, safe, and legal items.

• **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 retrieved 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 a different singer based on user instructions, learn the singing skills of this singer from the Web data, and finally make a music video of "Worldly Tavern" performed by this singer.

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

4 Feasibility Study

To investigate the feasibility of GeneRec, we design a fashion generation task and employ diffusion models to build a simple demo of the AI creator for personalized fashion item generation. The generated designs with high value can be sent to factories for customized production like Tribute Brand[1].

• **Personalized fashion item generation.** In real-world applications, users can interact with LLMs to proactively express their needs on fashion items [8]. These multi-turn interactions can be summarized as a single instruction to guide fashion item generation together with user feedback (*e.g.,* the user's historically liked items). Given the user's historically liked fashion item images and a single

[1]https://www.tribute-brand.com/.

Table 1: Quantitative results of SD-Creator with and without historically liked items, and Random.

	Generation Quality			Personalization
	FID↓	IS↑	CS (%)↑	Cosine (%)↑
Random	-	-	14.18	49.37
SD-Creator w/o history	60.57	24.40	25.51	53.83
SD-Creator	**53.39**	**29.76**	**26.37**	**56.11**

Figure 5: Examples of SD-Creator for personalized fashion item generation.

user instruction[2], the AI creator is asked to generate personalized fashion item images that meet the user's specific needs.

• **Dataset and evaluation.** We process a high-quality fashion dataset iFashion[3], which contains 107, 396 interactions between 12, 806 users and 344, 186 fashion items from diverse genres (*e.g.,* sweater, boots, and ring). For each item, we simulate a user instruction based on item features using LLMs, *e.g.,* "Generate a pair of blue sneakers with white lace.". In each user's interaction sequence, the last liked item serves as the target item, while the preceding liked items are utilized as historically liked items. For evaluation, we assess the AI creator's 1) *generation quality* using three widely adopted metrics: FID, IS, and CLIP Score (CS) [5, 13, 19], and 2) *personalization ability* by calculating the average cosine similarity between the CLIP [13] representations of the generated item and the user's historically liked items.

• **Implementation of AI creator.** We employ the latest Stable Diffusion-v2 (SD) [14] to implement the AI creator (SD-Creator). To adapt SD to our designed task, we use the user's historically interacted fashion item images and the constructed user instruction as visual and textual conditions, respectively. As shown in Figure 4, we encode the visual item images via a pre-trained autoencoder [14] and encode the textual user instruction via CLIP [13]. The encoded representations then guide the generation of the personalized fashion item image in the denoising process of SD.

• **Results.** We evaluate SD-Creator with (w/) and without (w/o) historically liked items and compare it with a Random method, which randomly selects an item from the item corpus as the recommended fashion item. The quantitative results are presented in Table 1, from which we can observe that: 1) The CS scores of both SD-based creators surpass Random by a large margin, showing their better alignment with the user's textual instructions. Notably, SD-Creator with users' historical interactions achieves superior generation quality, potentially underscoring the usefulness of integrating visual conditions as a reference for content generation. 2) SD-Creator can generate items that are more similar to the user's historically liked items, possibly indicating its strong personalization ability in capturing implicit user preference. These findings highlight the potential of harnessing AIGC methods for personalized content creation.

We also provide some case studies as shown in Figure 5. We can find that 1) SD-Creator can generate high-quality fashion item images that closely align with user instructions, *e.g.,* a black skirt that has white stripes for user 9436, and a V-neck patterned dress for user 9570; and 2) SD-Creator can capture user preference from the user's historical interactions and generate personalized fashion items. For instance, the generated skirt for user 9436 has the A-line style in black and white, showing similar patterns with the user's historically liked items. Besides, for user 9570, the color of the generated dress is red, a color likely preferred by the user.

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

5 Conclusion and Future Work

In this work, we empowered recommender systems with the abilities of personalized multimodal 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, editing, and creation to meet users' information needs. To implement GeneRec, we formulated three modules: an instructor, an AI editor, and an AI creator. We explored the feasibility of implementing the key AI creator of GeneRec in the fashion design domain. The experiments reveal promising results of existing generative models on personalized multimodal content generation.

This work formulates a new generative recommender paradigm, leaving many valuable research directions for future exploration. In particular, 1) it is critical to learn users' information needs from users' multimodal instructions and feedback in multiple domains. 2) Developing more powerful generative modules for various tasks is essential. Besides, we might implement some generation tasks through a unified model, where multiple tasks may promote each other. And 3) it is promising to devise new metrics, standards, and technologies to enrich the evaluation and trustworthiness checks of AIGC. It is promising to introduce human-AI collaboration and multi-agent collaboration for interactive evaluation and various trustworthiness 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] Paul-Alexandru Chirita, Wolfgang Nejdl, and Cristian Zamfir. 2005. Preventing Shilling Attacks in Online Recommender Systems. In *WIDM.* ACM, 67–74.
[4] 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.
[5] Ming Ding, Zhuoyi Yang, Wenyi Hong, Wendi Zheng, Chang Zhou, Da Yin, Junyang Lin, Xu Zou, Zhou Shao, Hongxia Yang, et al. 2021. Cogview: Mastering text-to-image generation via transformers. *NeurIPS* 34, 19822–19835.
[6] 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.

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

[8] 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*.

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

[10] 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.

[11] 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.

[12] Nathan N Liu, Evan W Xiang, Min Zhao, and Qiang Yang. 2010. Unifying explicit and implicit feedback for collaborative filtering. In *CIKM*. 1445–1448.

[13] 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.

[14] 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.

[15] Xiaoteng Shen, Rui Zhang, Xiaoyan Zhao, Jieming Zhu, and Xi Xiao. 2024. PMG: Personalized Multimodal Generation with Large Language Models. In *WWW*. ACM, 3833–3843.

[16] 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.

[17] 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).

[18] Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, and Hong Lin. 2023. Ai-generated content (aigc): A survey. *arXiv:2304.06632* (2023).

[19] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. 2023. Imagereward: Learning and evaluating human preferences for text-to-image generation. *NeurIPS* 36.

[20] Yiyan Xu, Wenjie Wang, Fuli Feng, Yunshan Ma, Jizhi Zhang, and Xiangnan He. 2024. Diffusion Models for Generative Outfit Recommendation. In *SIGIR*. ACM, 1350–1359.

[21] 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.