RESEARCH-ARTICLE

Generative News Recommendation

SHEN GAO, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

JIABAO FANG, Shandong University, Jinan, Shandong, China

QUAN TU, Renmin University of China, Beijing, China

ZHITAO YAO, Shandong University, Jinan, Shandong, China

ZHUNMIN CHEN, Shandong University, Jinan, Shandong, China

PENGJIE REN, Shandong University, Jinan, Shandong, China

View all

Open Access Support provided by:

Shandong University

Leiden University

University of Electronic Science and Technology of China

Renmin University of China

PDF Download
3589334.3645448.pdf
17 January 2026
Total Citations: 19
Total Downloads: 1546

Published: 13 May 2024

Citation in BibTeX format

WWW '24: The ACM Web Conference 2024
May 13 - 17, 2024
Singapore, Singapore

Conference Sponsors:
SIGWEB

WWW '24: Proceedings of the ACM Web Conference 2024 (May 2024)
https://doi.org/10.1145/3589334.3645448
ISBN: 9798400701719

Generative News Recommendation

Shen Gao*
University of Electronic Science and Technology of China
shengao@pku.edu.cn

Jiabao Fang*
Shandong University
jiabaofang@mail.sdu.edu.cn

Quan Tu
Renmin University of China
quantu@ruc.edu.cn

Zhitao Yao
Shandong University
yaozhitao@mail.sdu.edu.cn

Zhumin Chen
Shandong University
chenzhumin@sdu.edu.cn

Pengjie Ren
Shandong University
jay.ren@outlook.com

Zhaochun Ren†
Leiden University
z.ren@liacs.leidenuniv.nl

ABSTRACT

Most existing news recommendation methods tackle this task by conducting semantic matching between candidate news and user representation produced by historical clicked news. However, they overlook the high-level connections among different news articles and also ignore the profound relationship between these news articles and users. And the definition of these methods dictates that they can only deliver news articles as-is. On the contrary, integrating several relevant news articles into a coherent narrative would assist users in gaining a quicker and more comprehensive understanding of events. In this paper, we propose a novel generative news recommendation paradigm that includes two steps: (1) Leveraging the internal knowledge and reasoning capabilities of the Large Language Model (LLM) to perform high-level matching between candidate news and user representation; (2) Generating a coherent and logically structured narrative based on the associations between related news and user interests, thus engaging users in further reading of the news. Specifically, we propose GNR to implement the generative news recommendation paradigm. First, we compose the dual-level representation of news and users by leveraging LLM to generate theme-level representations and combine them with semantic-level representations. Next, in order to generate a coherent narrative, we explore the news relation and filter the related news according to the user preference. Finally, we propose a novel training method named UIFT to train the LLM to fuse multiple news articles in a coherent narrative. Extensive experiments show that GNR can improve recommendation accuracy and eventually generate more personalized and factually consistent narratives.

CCS CONCEPTS

• **Information systems → Recommender system**.

KEYWORDS

News Recommendation, Large Language Models, Generative Recommendation

ACM Reference Format:
Shen Gao, Jiabao Fang, Quan Tu, Zhitao Yao, Zhumin Chen, Pengjie Ren, and Zhaochun Ren. 2024. Generative News Recommendation. In *Proceedings of the ACM Web Conference 2024 (WWW '24), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3589334.3645448

1 INTRODUCTION

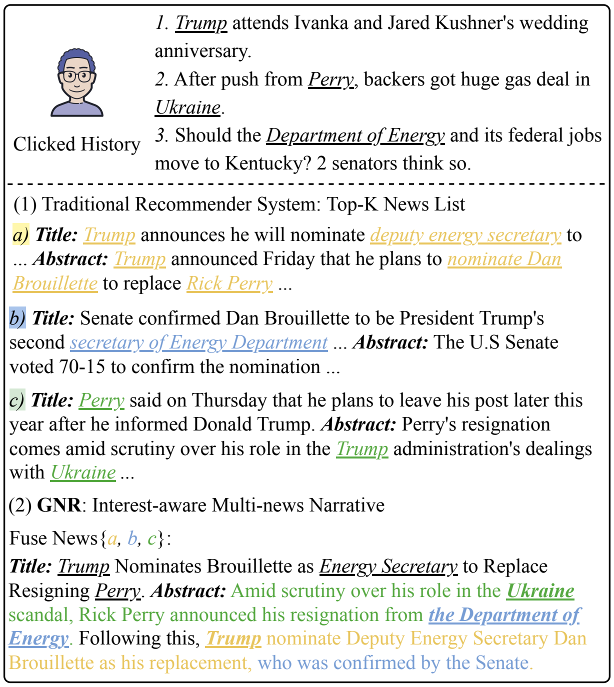
Online news platforms, such as Google News, have become crucial avenues for users to acquire daily information [5]. However, it is challenging for users to find interesting content among a large number of news articles. Hence, the news recommender system, which selects news based on user preference, is designed to improve the experience of user reading and alleviate the information overload problem [10].

Nevertheless, traditional news recommendation encounters the following limitations: (1) News recommendation is a content-based task that mainly relies on semantic matching between candidate news and user preference. These methods only capture explicit semantic relationships and overlook the equally important implicit relationships required for accurate recommendations. For example, a news article about "Argentina's win over France was the greatest World Cup final ever" may not exhibit an obvious semantic connection with another news article about "Lionel Messi cements his place among the greats after winning epic duel against Kylian Mbappé". However, a user who likes the previous news is likely to like the latter as well. This is due to their shared theme of "Messi won the World Cup final", representing an implicit relationship between them. However, finding the implicit relationship requires the knowledge of "Messi is a player of the Argentina national football team" and reasoning ability. (2) Existing news recommender systems can only recommend news in its original form. Users are required to read numerous lengthy news articles to gain an understanding of the overall context of events. Furthermore, users with different interests are presented with identical content without any personalization. Figure 1 shows an example of the recommended news list of existing methods. Although the recommended news list covers the main events of a user-preferred topic, the user will read several long news articles with redundant information. A desired outcome for a news recommender system would be to provide a

*Both authors contributed equally to this research.
†Corresponding author.

WWW '24, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0171-9/24/05.
https://doi.org/10.1145/3589334.3645448



Clicked History

1. Trump attends Ivanka and Jared Kushner's wedding anniversary.
2. After push from Perry, backers got huge gas deal in Ukraine.
3. Should the Department of Energy and its federal jobs move to Kentucky? 2 senators think so.

(1) Traditional Recommender System: Top-K News List

a) ***Title:*** Trump announces he will nominate deputy energy secretary to ... ***Abstract:*** Trump announced Friday that he plans to nominate Dan Brouillette to replace Rick Perry ...

b) ***Title:*** Senate confirmed Dan Brouillette to be President Trump's second secretary of Energy Department ... ***Abstract:*** The U.S Senate voted 70-15 to confirm the nomination ...

c) ***Title:*** Perry said on Thursday that he plans to leave his post later this year after he informed Donald Trump. ***Abstract:*** Perry's resignation comes amid scrutiny over his role in the Trump administration's dealings with Ukraine ...

(2) **GNR**: Interest-aware Multi-news Narrative

Fuse News{*a*, *b*, *c*}:

Title: Trump Nominates Brouillette as Energy Secretary to Replace Resigning Perry. ***Abstract:*** Amid scrutiny over his role in the ***Ukraine*** scandal, Rick Perry announced his resignation from ***the Department of Energy***. Following this, ***Trump*** nominate Deputy Energy Secretary Dan Brouillette as his replacement, who was confirmed by the Senate.

Figure 1: Only recommending existing news in the news corpus is one of the limitations of traditional methods.

concise paragraph that overviews the main events that the user is interested in.

In this paper, we introduce a novel generative news recommendation paradigm (GNR). Our approach incorporates the Large Language Model (LLM) as a generator to enhance news recommendations by precisely catering to user needs. As illustrated in Figure 2(a), traditional news recommendation methods perform semantic matching using candidate news and user representation, primarily composed of the user's historical clicked news list. These methods subsequently present news articles to the user in their original format. In contrast, our approach leverages the LLM to generate theme-level representations for news and users, as depicted in Figure 2(b). Then we explore a personalized related news set as the information source and generate a coherent and logically structured narrative to engage users and encourage them to read more of the news.

Specifically, the GNR consists of three modules. The first two modules aim at retrieving a news set that contains user-interested news and its related news, and the last module fuses the news in the set into a coherent narrative. The first module is **Generative Dual-level Representation**: Following the previous news recommendation methods [33, 35, 36], we first obtain semantic-level representations for both users and news. Then we leverage the LLM to map news content and user profiles to theme-level representations. Finally, we combine these representations into dual-level representations. The second module is **Personalized Related News Exploration**: To generate coherent narratives, we need to find a personalized and interconnected news set. There are three main steps in this module. We first conduct news ranking based on the dual-level news and user representations and obtain the focal news that best matches user preference. Second, we propose to explore the logical relation between news articles, aiming to discover more news articles related to the focal news. Since the second step introduces more related news articles that may not be interesting to the user, we conduct personalized filtering in the third step. Finally, we obtain a reference news set that encompasses both the main event context of the focal news and takes into account the user preference. The third module is **Interest-aware Multi-news Narrative Fusion**: The primary objective of this module is to create a coherent and logically structured narrative that encapsulates the central theme of the reference news set. To enhance the alignment between the generated narrative and user interests, and to attract users to engage more with the content, we introduce the User Interest Alignment Fine-Tuning (UIFT) method, which adjusts the probabilities of multiple multi-news narratives by optimizing for ranking loss. Extensive experiments conducted on a benchmark dataset demonstrate that GNR improves recommendation accuracy and offers users more personalized multi-news narratives.

To sum up, our contribution can be summarized as follows:

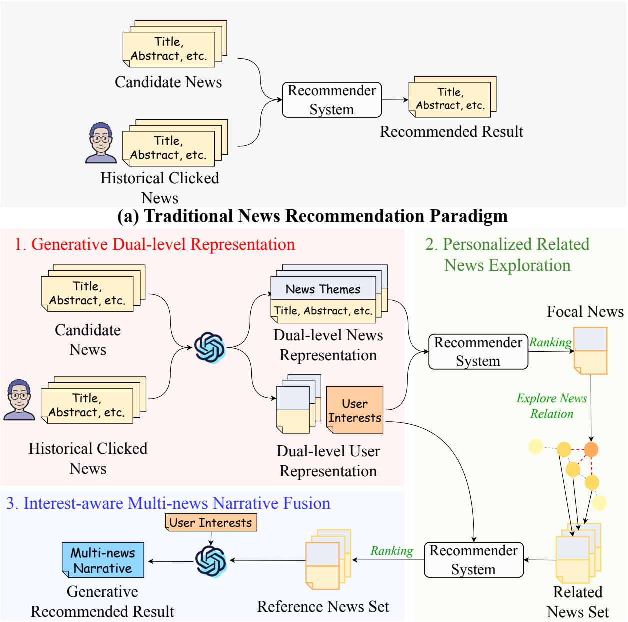
- We propose a generative news recommendation paradigm (GNR), which introduces a powerful LLM as the generator to make the recommended news meet user needs more precisely.
- In GNR, we design three modules to perform two sub-tasks: (I) Leveraging the internal knowledge and reasoning capability of LLM to retrieve a personalized related news set; (II) Generating a coherent and logically structured narrative, thus engaging users in further reading of the news.
- We propose a novel training method User Interest Alignment Fine-tuning (UIFT) which fine-tunes the LLM through ranking loss based on user interests.
- Extensive experiments on the MIND dataset demonstrate that our GNR can significantly improve the accuracy of recommender systems and the generated narratives are more personalized to fulfill the user information needs.

2 RELATED WORK

2.1 Generative LLMs for Recommendation

Recently, LLMs have achieved great success in many natural language processing tasks due to their excellent natural language understanding and natural language generation abilities [7, 23, 24, 40, 41]. And many studies have surfaced that LLMs can be used for recommendations due to their strong instruction following and common-sense reasoning abilities [4, 8, 13, 17, 18, 22].

Initially, Wang and Lim [27] proposed leveraging LLMs to conduct the sequential recommendation directly by prompting. Hou et al. [9] proposed using LLMs as a ranker, which converts the interaction history and candidate items into natural language form, and inputs them into LLMs together with the ranking instruction. However, many researchers [2, 3, 39] found that when LLMs are directly applied to recommendation tasks through prompting or in-context learning, there is a certain performance gap compared with existing recommendation models. The reason is that LLMs are primarily trained on NLP-related tasks and lack training on recommendation tasks [19, 25]. In order to solve this problem, Bao et al. [3] and Zhang et al. [39] proposed to improve the performance of LLMs for recommendation by instruction tuning, which involves



(a) Traditional News Recommendation Paradigm

(b) Generative News Recommendation Paradigm (Ours)

Figure 2: The differences between traditional news recommendation paradigm and our proposed generative news recommendation paradigm GNR.

generating natural language format instructions based on data from the recommendation tasks.

In addition to directly using LLMs to conduct recommendation, some researchers propose to use generative LLMs to assist traditional recommender systems [28, 31]. Gao et al. [6] proposed Chat-REC, which leverages an LLM as the controller of the recommender system and the interface with the users. Chat-REC allows the users to express their needs proactively and also makes the recommendation more explainable. Wang et al. [30] proposed an interactive evaluation approach based on LLMs to address the problem of past evaluation metrics that overly focus on matching with "ground-truth" items in conversational recommender systems. Wang et al. [29] proposed a generative recommender paradigm, which uses LLMs as the controller of the recommender system to determine whether to recommend an existing item from the item corpus or to generate a new item through an AI generator. Their method primarily uses LLMs as the controllers of the system and utilizes the diffusion models to transfer the micro-video style. Most of the previous works use LLMs as the recommender systems or the controllers. On the contrary, we leverage LLMs as the generator to provide dual-level representations for better recommendation and generate personalized and coherent multi-news narratives that assist users in learning news events.

2.2 News Recommendation

The main task of the news recommendation is to recommend news articles that are consistent with the user preferences [11, 26, 35, 38]. The core of the news recommender system includes the news encoder and the user encoder. Wu et al. [32] mainly focused on enhancing news representation and proposed a news recommendation approach with attentive multi-view learning (NAML), which uses word-level and view-level attention networks to select key information in news. An et al. [1] proposed a neural news recommendation approach with both long-term and short-term user representations. Wu et al. [34] proposed a neural news recommendation approach, which combines the multi-head self-attention mechanism in the news encoder and the user encoder. Following previous work, Wu et al. [36] used the more powerful pre-trained language models as the backbone for the news encoder. With the rise of LLMs, many researchers have started incorporating it into news recommendation Li et al. [12], Liu et al. [14, 15], Runfeng et al. [21]. For example, Liu et al. [14] proposed to utilize both open- and closed-source LLMs to enhance content-based recommendation, which includes the news recommendation.

However, these existing news recommendation methods are all based on human-written news corpus, which means they can only recommend raw news articles as-is. In our work, GNR can use LLMs to fuse multi-news narratives that more align with user preference.

3 TASK FORMULATION

The generative news recommendation can be formulated as retrieving a reference news set $\mathcal{N}^r$ from the news corpus $\mathcal{N}$ and generating a coherent narrative n^m to overview the main event of the reference news set $\mathcal{N}^r$. This reference news set must fulfill two key characteristics. Firstly, it should align with the user interests. Second, it should comprehensively mine related news articles by exploring implicit relationships among news articles, encompassing the full context of an event. Then the generated multi-news narrative n^m can introduce the full context of the news event. In this paper, we use $\mathcal{N} = [n_1, n_2, \ldots, n_k]$ to denote the whole news corpus which has k news articles in total. The news recommender system models the user preference based on the user's historical clicked news list $\mathcal{N}^h = [n^h_1, n^h_2, \ldots, n^h_i]$.

Then the recommender system matches the user preference and news in the candidate list $\mathcal{N}^c = [n^c_1, n^c_2, \ldots, n^c_j]$ to predict the scores and outputs a focal news n^f with the highest matching scores. Based on the focal news, we apply a filter to the whole news corpus $\mathcal{N}$ in order to find a reference news set $\mathcal{N}^r = [n^f, n^r_1, \ldots, n^r_{T-1}]$, which is both personalized and interconnected. Then we fuse the reference news set $\mathcal{N}^r$ to obtain a multi-news narrative n^m as the generative recommended result.

4 GNR METHOD

In order to help news recommendations better satisfy user needs, we propose Generative News Recommendation (GNR).

The GNR consists of three modules, as shown in Figure 2. **First**, we leverage the LLM to generate theme-level representations and combine them with the semantic-level representations to obtain the dual-level representations, as shown in § 4.1. **Second**, we design a three-step pipeline for acquiring a personalized and interconnected news set. In this process, we initially rank the candidate news, then explore the relation between the news articles, and finally filter the related news set. Details are shown in § 4.2. **Third**, we fuse the news set to generate a brief multi-news narrative, which can assist a user in quickly learning about the news event that interests him/her. Details are shown in § 4.3.

Table 1: Example prompt for theme-level news representation generation

Instruction
Based on the given news information, summarize what **topic(s)** the news is related to. Each news article is related to 1-3 topics, and each topic should not exceed five words.
Input
{"title": "Trump says the Kurds 'are no angels' and the PKK are 'probably worse' than ISIS", "abstract": "President Trump defended his decision to withdraw U.S. forces from Syria, claiming that ... and saying that the Kurds 'are no angels' ... ", "category": "politics"}
Output
This news is related to **[Trump's decision on Syria]**, **[Kurds and PKK]**.

Table 2: Example prompt for theme-level user representation generation

Instruction
You are asked to describe user interest based on his/her browsed news list. User interest includes the news **[categories]** and news **[topics]** (under each **[category]**) that users are interested in.
Input
News List: {"ID": "News 1", "title": "Lionel Messi says he wants to ...", "category": "sports", "topics": "Argentina football player Lionel Messi"} {"ID": "News 2", "title": "How the world reacted to the best World Cup final ever", "category": "sports", "topics": "World Cup final" ...
Output
According to **[News 1, News 2, News 3, News 4]**, this user is interested in news about [sports], especially **[Lionel Messi, World Cup final, Argentina's victory in the World Cup]**.

4.1 Generative Dual-level Representation

Higher-level connections can help news recommender systems better match users and news, and such connections require domain knowledge and reasoning ability to be obtained. Therefore, we propose to use the LLM to generate theme-level representations for both news and users. Finally, we combine the theme-level representations with the semantic-level representations to obtain dual-level representations and promote more accurate matching.

4.1.1 Theme-level News Representation. To obtain higher-level news representations, we leverage the common-sense knowledge in the LLM to summarize the themes for each news. For example, the original content of a news article is "In the 2022 FIFA World Cup, Lionel Messi cements his place among the greats after winning epic duel against Kylian Mbappé". Then, the theme of this news is "Messi won the World Cup", which also serves as the theme-level representation of this news. Specifically, we first manually construct a prompt template, and then put the original news content as input, including news titles, abstracts, and categories. The specific prompt construction is shown in Table 1.

4.1.2 Theme-level User Representation. We consider user profiles, which encompass various news themes, as higher-level representations of users. To obtain this representation, we employ the LLM to infer the connections within each news in the user's historical clicked list and generate a description of the user profile. For example, if a user's historical clicked news list contains "Why Argentina's win over France was the greatest World Cup final ever" and "Lionel Messi cements his place among the greats after winning epic duel against Kylian Mbappé", we can infer that this user is interested in the news theme "Argentina won the World Cup", which is also part of the theme-level user representation. Furthermore, we leverage the LLM through in-context learning. We manually create a prompt template and input news information, which includes news titles, categories, and theme-level news representations. The specific prompt construction is shown in Table 2.

4.1.3 Dual-level Representation Combination. As mentioned above, we leverage the LLM to generate theme-level representations for news and users. Meanwhile, we can get the semantic-level representation based on the original content of the news. Then we need to combine these two representations and provide the dual-level representations to the recommender system. Inspired by NAML [32], which is a widely known news recommendation method, we also use the multi-view attention network to fuse the representations:

$$
\begin{gathered}
\alpha_s = q_{v1}^{T} \tanh\left(\text{Linear}\left(e_s\right)\right), \alpha_t = q_{v2}^{T} \tanh\left(\text{Linear}\left(e_t\right)\right) \\
\alpha_s' = \frac{\exp\left(\alpha_s\right)}{\exp\left(\alpha_s\right) + \exp\left(\alpha_{\mathrm{t}}\right)}, \alpha_t' = \frac{\exp\left(\alpha_t\right)}{\exp\left(\alpha_s\right) + \exp\left(\alpha_t\right)} \\
e_d = \alpha_s' e_s + \alpha_t' e_t,
\end{gathered} \tag{1}
$$

where q_{v_1} and q_{v_2} both are attention query vectors, e_s is the embedding of semantic-level representation, e_t is the embedding of theme-level representation, e_d is the embedding of dual-level representation, α_s is the attention weight of the semantic-level representation, and α_t is the attention weight of the theme-level representation. In the following, we default the user embedding and the news embedding mentioned in the following are calculated based on the dual-level representations.

4.2 Personalized Related News Exploration

To ensure the coherence of the recommended narrative and cater to user interests, it is imperative to extract a personalized and interconnected news set, referred to as a "reference news set", from the news corpus $\mathcal{N}$. To achieve this goal, we have devised a three-step pipeline, which encompasses dual-level news ranking, news relation exploration, and personalized filtering. In the subsequent sections, we introduce the details of these three steps.

4.2.1 Dual-level News Ranking. To cater to user interests effectively, we adopt a candidate news ranking method similar to the conventional news recommendation paradigm. To train the ranking model, we use the negative sampling and randomly select K_{neg} non-clicked candidate news as negative samples. Then the probability of the user clicking the positive candidate news is:

$$\hat{y}_i = e_i^{user} \cdot e_i^{cand},$$
$$p_i = \frac{\exp\left(\hat{y}_i^+\right)}{\exp\left(\hat{y}_i^+\right) + \sum_{j=1}^{K_{neg}} \exp\left(\hat{y}_i^j\right)} \tag{2}$$

where e_i^{user} is the embedding of dual-level user representations, e_i^{cand} is the embedding of dual-level candidate news representations and $\hat{y}_i$ is the predicted ranking score of the candidate news.

We optimize the probability of positive sample p_i through log-likelihood loss:

$$\mathcal{L}_{ranking} = -\sum_{i \in S} \ln\left(p_i\right) \tag{3}$$

where S is the training dataset.

Then, we can obtain a ranking score for each candidate news, indicating the matching degree between the candidate news and user preference. However, the top news articles in the ranking list may not be correlated with each other, making it challenging to generate a coherent narrative based on these news articles. Thus, we only select the top-1 news as the focal news n^f in the ranking list and use it to conduct the subsequent steps.

4.2.2 News Relation Exploration. After obtaining the focal news n^f, we need to find some related news articles from the news corpus to generate a coherent narrative. Therefore, we implement a news relation classifier to judge whether two news articles are related and set a relation threshold α. In order to train the model, we collected a set of related news pairs from news websites to form a training dataset. We construct the news relation classifier based on Siamese networks [20] and train the model using contrastive learning loss. The positive pair is a pair of news articles that are related to each other, and we randomly select some unrelated news articles to construct negative data pairs. The loss function is formulated as follows:

$$\mathcal{L}_{classify} = \max\left(\left\|e_i^{news} - e_j^{news}\right\| - \left\|e_i^{news} - e_k^{news}\right\| + \epsilon, 0\right) \tag{4}$$

where e_j^{news} is the embedding of the positive news (*a.k.a.,* related news), e_k^{news} is the embedding of the negative news (*a.k.a.,* unrelated news), and ϵ is the margin between positive and negative pairs. After model training, we use this news relation classifier to explore a related news set $\mathcal{N}^{rel} = \{n_1^{rel}, n_2^{rel}, \ldots, n_j^{rel}\}$ related to the focal news n^f from the news corpus $\mathcal{N}$.

4.2.3 Personalized Filtering. Since the related news set $\mathcal{N}^{rel}$ is explored based on the news relation, which ignores the user preference. In this module, we propose to personalized filter the related news set using the traditional recommender system. We use the related news set $\mathcal{N}^{rel}$ as the candidate set and compute the matching score $\hat{y}_i$ between related news embedding e_i^{rel} and user embedding e_i^{user} as follows:

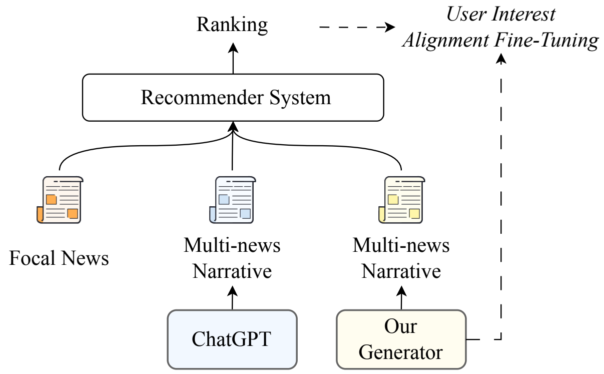


Figure 3: The framework of UIFT method.

$$\hat{y}_i = e_i^{user} \cdot e_i^{rel} \tag{5}$$

Then we select $T-1$ news articles with the highest matching score. And together with the focal news n^f obtained above, we form the reference news set $\mathcal{N}^r$.

4.3 Interest-aware Multi-news Narrative Fusion

Conventional news recommendation methods typically recommend news articles in their original form. Consequently, when users express interest in a specific news event, they need to read a multitude of related news articles to gain an understanding of the pertinent content. This process is notably time-consuming and results in a suboptimal user experience. Simultaneously, personalization diminishes as recommender systems deliver the same news content to users with varying interests. To address this challenge, we propose the fusion of multi-news narratives based on user interests. In this section, we first introduce the process of multi-news narrative fusion and subsequently introduce the UIFT method, designed to tailor the generated narrative more closely to user interests.

First, to ensure the coherence and readability of the narrative, we use the focal news n^f (as introduced in § 4.2.1) as the central point of the multi-news narrative. Then the goal of the generated narrative is to extract the key facts of the news set $\mathcal{N}^r$ that align with user interests and fuse these key facts around the focal news n^f. An illustrative example can be found in § A.1.

Despite the in-context learning capabilities of general black-box LLM, they still exhibit deficiencies in capturing user interests, a crucial aspect of the recommendation task. So we propose User Interest Alignment Fine-Tuning (referred to as UIFT), aimed at enhancing the personalization of LLM-generated multi-news narratives. The framework of UIFT is shown in Figure 3. We first utilize ChatGPT to generate the multi-news narrative and conduct supervised fine-tuning of our multi-news narrative generator to distill the ability and knowledge from ChatGPT to our narrative generator. Then UIFT trains our multi-news narrative generator by incorporating ranking loss, thereby aligning its ranking for multiple news narratives with user interests. We collect a training set containing multiple focal news articles n^f, the corresponding multi-news narratives generated by ChatGPT $n^{ChatGPT}$, and the corresponding multi-news narratives fused by our generator n^{GNR}. Next, UIFT first allows the narrative generator π trained with supervised fine-tuning to predict the conditional probability p_i for each news:

$$p_i = \frac{\sum_t \log \mathrm{P}_\pi \left(s_{i,t} \mid n, s_{i<t}\right)}{\|s_i\|}, \tag{6}$$

where n is the news in the set $\{n_i^{ChatGPT}, n_i^{GNR}, n_i^f\}$.

In UIFT, the model learns to give higher probabilities to more personalized news, thus aligning with user interests. Specifically, we rank the news set $\{n_i^{ChatGPT}, n_i^{GNR}, n_i^f\}$ based on the user interests. Due to the high cost of manual annotation, we rely on a well-trained news recommender system to reflect user interests. Based on the predicted score computed by the recommender system, we can get the ranking between the news (i.e., $r_i^{GNR} < r_i^{ChatGPT} < r_i^f$) and the $r_i \in \{1, 2, 3\}$ denotes the ranking of the news. When r_i is smaller, it means that the corresponding news is more aligned with the user interests.

Our training goal is to give the larger probability p_i to news with better ranking r_i. We achieve this through ranking loss:

$$L_{\text{rank}} = \sum_{r_i<r_j} \max\left(0, p_i - p_j\right), \tag{7}$$

5 EXPERIMENT

In this section, we conduct extensive experiments to answer the following research questions:

RQ1: How much accuracy improvement can GNR bring to the recommendation models by combining dual-level representations?

RQ2: Can GNR generate personalized and factually consistent multi-news narratives?

RQ3: When exploring the relation between news, how does the relation threshold α affect the performance?

RQ4: When retrieving the reference news set $\mathcal{N}^r$, how does the maximum number of reference news T_{max} affect the performance?

RQ5: What are the differences between personalized multi-news narratives and non-personalized multi-news narratives?

5.1 Datasets

We conduct experiments on the MIND dataset [37], which is a large-scale dataset for news recommendation. MIND dataset contains the news dataset and the behaviors dataset. Each news item in MIND contains a news title, abstract, etc. Each behavior includes a historical clicked news list $\mathcal{N}^h$ and a candidate news list $\mathcal{N}^c$. For the news dataset, we first filtered 5145 news articles from MIND-Large under the "politics" category, which is more appropriate for the scenario of GNR. We will extend the GNR to experiment under more news categories in the future. For the behaviors dataset, we first filtered behaviors to ensure all news items in the historical clicked list and the candidate list belong to the politics news dataset. When fusing multi-news narratives, we filter behaviors to ensure that the length of the historical clicked news list to fall within the range: 5 <= $|\mathcal{H}|$ <= 15. We then structured two separate datasets for training the recommender system and the multi-news narrative generator respectively. The sizes of the training, validation, and test sets for the recommender system are 43232, 4800, and 6713. Similarly, for the multi-news narrative generator, the sizes of the training, validation, and test sets are 8926, 183, and 1956.

Furthermore, to train the classifier for news relation exploration, we employed a web crawler to extract news articles from the CNN website [1]. We also scraped the relevant news from each news webpage to construct the training dataset for the news relation classifier. While retrieving the reference news, we used them to augment the related news and enrich our news dataset.

5.2 Evaluation Metrics

The GNR proposed two sub-tasks: (1) Retrieving personalized references news sets; (2) Fusing coherent multi-news narratives. We evaluate the performance of these two sub-tasks separately.

For the first sub-task, we evaluate whether the theme-level representations generated by GNR can promote recommendation accuracy. So we leverage AUC (Area Under the Curve), MRR (Mean Reciprocal Rank), and NDCG@K (Normalized Discounted Cumulative Gain) where K=5 as the evaluation metrics.

For the second sub-task, as GNR is a novel paradigm, there are no previous benchmarks available. Therefore, we propose two automatic metrics to evaluate the personalization and consistency of multi-news narratives: (1) **Win Rate** evaluates whether the fused multi-news narratives are more personalized than the corresponding focal news. We first calculate the predicted scores of the multi-news narrative and the focal news based on a well-trained news recommendation model. The inputs of the recommender system are the dual-level news representations and user representations. We mark a situation as a "Win" when the predicted score of a multi-news narrative surpasses that of the focal news, interpreting this as an indication that the multi-news narrative aligns better with the user preference. Then we compute the Win Rate across the entire test dataset. (2) **Consistency Rate** is calculated between the reference news sets and the multi-news narratives. During the evaluation, we feed a reference news set and a multi-news narrative to the evaluator. The evaluator then determines whether the reference news set and the multi-news narrative are consistent. Subsequently, we compute the consistency rate across the entire test dataset. And inspired by Luo et al. [16], we use the ChatGPT (gpt-3.5-turbo) [2] as the evaluator to determine consistency.

5.3 Baseline Models

We evaluate GNR against the following news recommendation methods: **(1) NRMS [33]** leverages the multi-head self-attention to learn news representations and capture the relatedness between the news; **(2) PLM4NR (title) [36]** uses pre-trained language models to model news representations from news titles. We leverage the best variant PLM4NR-NRMS in our experiments; **(3) PLM4NR (title and abstract)** is similar to PLM4NR (title), but it models news representations from titles and abstracts, not just titles.

5.4 Implementation Details

For theme-level representation generation, we select the ChatGPT (gpt-3.5-turbo) as the backbone model. For news relation exploration, we use the SBERT [20] as the backbone model. During the training, we use AdamW as the optimizer, and the learning rate is

[1] https://edition.cnn.com/

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

Table 3: Performance of recommendation accuracy. We experiment with different combinations of news representations and user representations. "Sem" denotes semantic-level representation. "Dual" denotes dual-level representation.

Input Type		NRMS			PLM4NR (title)			PLM4NR (title and abstract)		
News	User	NDCG@5	AUC	MRR	NDCG@5	AUC	MRR	NDCG@5	AUC	MRR
Sem	Sem	58.57	56.47	50.44	62.38	69.44	54.73	61.11	68.36	53.70
Sem	Dual	59.39	56.54	51.36	62.98	**70.38**	55.21	61.78	68.77	54.35
Dual	Sem	58.99	56.35	51.46	63.02	69.95	55.61	62.76	69.26	55.52
Dual	Dual	**59.81**	**57.54**	**51.80**	**63.46**	70.31	**56.03**	**62.99**	**69.42**	**55.72**

set to 1e-5. When selecting the related news, we set hyperparameter $\alpha = 0.8$. During the training of traditional news recommender systems, the learning rate is set to 1e-4. In the PLM4NR model, we employ DistilBERT as the encoders and utilize the output embedding of the "[CLS]" token. Meanwhile, we treat the next news to be clicked (i.e., ground truth) as the focal news to better compare whether the multi-news narratives fused by GNR can better match user preferences. We set the length of reference news list T to fall within the range: $2 < |T| < 5$. We exclude samples that do not have a sufficient number of related news articles, as they are not suitable for our method. For our multi-news narratives generator, we use the LLaMA 7B as the backbone model. And for RQ3, RQ4, and RQ5, we utilize ChatGPT as the generator for a fair comparison. During the supervised fine-tuning of LLaMA, the learning rate is set to 5e-5. During the UIFT, the learning rate is set to 1e-5. Besides, all LLaMA-based experiments are conducted on 8 80GB Nvidia A800 GPUs, and other experiments are conducted on 24G Nvidia 3090 GPUs.

5.5 Performance of Recommendation Accuracy (RQ1)

We first evaluate the recommendation accuracy when using the dual-level representations as input, and report the results in Table 3. From the results, we can have the following observations.

5.5.1 Result Analysis. Compared to the accuracy of using the semantic-level representation only, the performance of the backbone models increases significantly when using the dual-level news and user representations. For example, the PLM4NR (title) model improved from 62.38 to 63.46 on metric NDCG@5 and improved from 69.44 to 70.31 on metric AUC. This suggests that with the help of common-sense knowledge in the LLM, the news recommender systems can capture more levels of relationships between the news and user preference. These relationships can help systems better match the candidate news and users and improve the recommendation accuracy.

5.5.2 Ablation Study. To separately evaluate the effect of the theme-level news representation and the theme-level user representation, we conducted ablation experiments with two settings: (1) using dual-level news representations and semantic-level user representations; (2) using semantic-level user representations and dual-level news representations. As observed, the recommendation models outperform the baseline models when utilizing theme-level news (or user) representations. However, their performance is not as good as the models that employ both theme-level news representations and user representations. So these results illustrate that both theme-level news representations and theme-level user representations can promote the performance of the recommendation models.

Table 4: Performance of multi-news narrative fusion. Win Rate is evaluated by PLM4NR (title + abstract).

Generator	Win Rate	Consistency Rate
ChatGPT	72.80	96.63
Ours (SFT)	65.13	96.52
Ours (UIFT)	74.54	96.57

5.6 Performance of Multi-news Narrative Fusion (RQ2)

In this part, we evaluate whether the multi-news narratives fused by GNR can perform better than the focal news and whether the multi-news narratives are factually consistent with the corresponding reference news. The results are reported in Table 4.

When using ChatGPT as the generator, GNR has great performance in both Win Rate and Consistency Rate. It illustrates that ChatGPT is able to fuse coherent and personalized multi-news narratives. We consider this to be due to ChatGPT's excellent in-context learning capability, allowing it to perform our desired task without additional fine-tuning.

When our narrative generator is trained only through supervised fine-tuning, it can effectively fuse news and generate coherent narratives. Nevertheless, its performance in Win Rate is considerably less effective than that of ChatGPT. However, when our narrative generator is trained through UIFT, the fused multi-news narratives can achieve superior personalization while maintaining good consistency. We attribute this to the fact that we align the probabilities of our narrative generator with user interests via ranking loss. This alignment aids the model in better understanding user preference during the process of multi-news narrative fusion.

5.7 Threshold of News Relation (RQ3)

As described in § 4.2, we design to retrieve a reference news set containing T personalized and related news. Therefore, we plan to explore the news relation and find a news set related to the focal news. In this process, we define a relation threshold α to determine whether two news articles are related. We think this threshold α can influence the performance of multi-news narrative fusion. To evaluate the impact, we conduct separate experiments by setting different thresholds. To highlight the relation threshold impact, we avoid selecting the same samples for different thresholds. During the experiment with threshold α_1, certain samples are excluded if the similarity scores between their focal news and no less than $T - 1$ news articles surpass the threshold α_2, where $\alpha_2 > \alpha_1$. Then we select 100 test samples for each threshold separately. However, there are only 83 test samples that comply with $\alpha = 0.6$, so we select all of them. Finally, we evaluate the Consistency Rate of the multi-news narratives under each setting, and the results are shown in Table 5.

From the results, we can see that relation threshold α has an impact on the consistency of the fused narrative. We hypothesize that this is due to the fact that when the relevance of the reference news set is low, the generator is unable to reason correctly about the associations between the reference news, leading to the hallucination generation.

Table 5: The impact of relation threshold α on multi-news narratives.

Relation Threshold α	Consistency Rate
0.6	66
0.7	85
0.8	98

5.8 The Maximum Number of Reference News (RQ4)

In this part, we hypothesize that the maximum number of reference news T_{max} determines the quality of input information, thus affecting the final generation quality. Therefore, we experiment with how the maximum number T_{max} promotes or reduces the performance of fusion. Considering the limitation of the LLM input length, we set T_{max}=2,3,4,5,6 respectively to conduct experiments and evaluate the fused narratives in each setting.

As we can see in Figure 4, the Win Rate increases first and then decreases with an increase in the maximum number of reference news T_{max}. The optimal performance is achieved when $T_{max} = 4$. This observation demonstrates when T_{max} is less than 4, the reference news set contains insufficient information to adequately cover the main events and cater to user preference. When T_{max} exceeds 4, there is an abundance of information within the reference news set that may lack relevance to the user interests, thereby introducing noise during the fusion process.

5.9 Personalized Evaluation of Multi-news Narratives (RQ5)

In GNR, personalized multi-news narratives should prioritize content that aligns with user preference while minimizing irrelevant material. In this section, we provide a quantitative comparison between personalized and non-personalized multi-news narratives. We utilize the same LLM to summarize the reference news sets and obtain non-personalized multi-news narratives. In the experiment, we sample 100 test cases and conduct both GPT-4 evaluation and human evaluation. The evaluator should select narratives that highlight content engaging with the user interests while excluding content that doesn't align with those interests. As shown in Table 6, the results demonstrate that our personalized narratives are more aligned with the user preference.

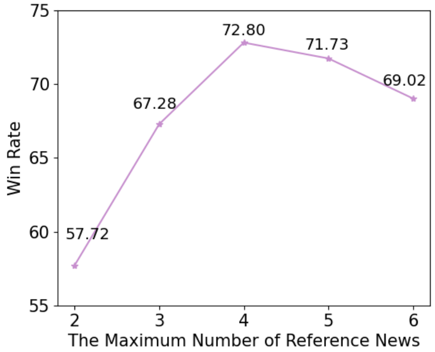


Figure 4: The impact of the maximum number of reference news T_{max} on multi-news narratives.

Table 6: The comparison between non-personalized multi-news narratives and personalized multi-news narratives. The Cohen's kappa between two results is more than 0.4.

Evaluator	Non-Personalized Narrative Wins	Tie	Personalized Narrative Wins
Human	4%	53%	43%
GPT-4	7%	56%	37%

6 CONCLUSION

In this paper, we introduce a novel generative news recommendation paradigm (GNR), which aims at enhancing news recommendation and fulfilling user needs more precisely using LLM. By harnessing the internal knowledge and reasoning capabilities of LLM, we generate theme-level representations for news and users. These representations are then combined with semantic-level representations to create dual-level representations. Subsequently, we explore the news relation to find personalized related news sets based on dual-level representations. Afterward, we fuse the personalized related news sets to create coherent and logically structured multi-news narratives, engaging users further in further reading of the news. Extensive experiments demonstrate that our GNR enhances recommendation performance and generates personalized and factually consistent multi-news narratives.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (T2293773, 62372275, 62272274, 62202271, 62102234, 62072279), the Natural Science Foundation of Shandong Province (ZR2023QF159), the National Key R&D Program of China with grant No.2022YFC3303004.

REFERENCES

[1] Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019. Neural News Recommendation with Long- and Short-term User Representations. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, Anna Korhonen, David R. Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, 336–345. https://doi.org/10.18653/v1/p19-1033

[2] Keqin Bao, Jizhi Zhang, Wenjie Wang, Yang Zhang, Zhengyi Yang, Yancheng Luo, Fuli Feng, Xiangnaan He, and Qi Tian. 2023. A bi-step grounding paradigm for large language models in recommendation systems. *arXiv preprint arXiv:2308.08434* (2023).

[3] 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. *arXiv preprint arXiv:2305.00447* (2023).

[4] 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 preprint arXiv:2305.02182* (2023).

[5] Abhinandan Das, Mayur Datar, Ashutosh Garg, and Shyamsundar Rajaram. 2007. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, Carey L. Williamson, Mary Ellen Zurko, Peter F. Patel-Schneider, and Prashant J. Shenoy (Eds.). ACM, 271–280. https://doi.org/10.1145/1242572.1242610

[6] 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 preprint arXiv:2303.14524* (2023).

[7] Chunxi Guo, Zhiliang Tian, Jintao Tang, Shasha Li, Zhihua Wen, Kaixuan Wang, and Ting Wang. 2023. Retrieval-augmented gpt-3.5-based text-to-sql framework with sample-aware prompting and dynamic revision chain. In *International Conference on Neural Information Processing*. Springer, 341–356.

[8] Jesse Harte, Wouter Zorgdrager, Panos Louridas, Asterios Katsifodimos, Dietmar Jannach, and Marios Fragkoulis. 2023. Leveraging large language models for sequential recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 1096–1102.

[9] 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 preprint arXiv:2305.08845* (2023).

[10] Miaomiao Li and Licheng Wang. 2019. A Survey on Personalized News Recommendation Technology. *IEEE Access* 7 (2019), 145861–145879. https://doi.org/10.1109/ACCESS.2019.2944927

[11] Xinyi Li, Yongfeng Zhang, and Edward C Malthouse. 2023. Exploring Fine-tuning ChatGPT for News Recommendation. *arXiv preprint arXiv:2311.05850* (2023).

[12] Xinyi Li, Yongfeng Zhang, and Edward C Malthouse. 2023. A Preliminary Study of ChatGPT on News Recommendation: Personalization, Provider Fairness, Fake News. *arXiv preprint arXiv:2306.10702* (2023).

[13] Jiayi Liao, Sihang Li, Zhengyi Yang, Jiancan Wu, Yancheng Yuan, Xiang Wang, and Xiangnan He. 2023. LLaRA: Aligning Large Language Models with Sequential Recommenders. *arXiv preprint arXiv:2312.02445* (2023).

[14] Qijiong Liu, Nuo Chen, Tetsuya Sakai, and Xiao-Ming Wu. 2023. ONCE: Boosting Content-based Recommendation with Both Open- and Closed-source Large Language Models. arXiv:2305.06566 (Aug. 2023). http://arxiv.org/abs/2305.06566 arXiv:2305.06566 [cs].

[15] Qijiong Liu, Nuo Chen, Tetsuya Sakai, and Xiao-Ming Wu. 2023. ONCE: Boosting Content-based Recommendation with Both Open-and Closed-source Large Language Models. *arXiv preprint arXiv:2305.06566* (2023).

[16] Zheheng Luo, Qianqian Xie, and Sophia Ananiadou. 2023. Chatgpt as a factual inconsistency evaluator for text summarization.

[17] Hanjia Lyu, Song Jiang, Hanqing Zeng, Yinglong Xia, and Jiebo Luo. 2023. LLM-Rec: Personalized Recommendation via Prompting Large Language Models. *arXiv preprint arXiv:2307.15780* (2023).

[18] Sheshera Mysore, Andrew McCallum, and Hamed Zamani. 2023. Large Language Model Augmented Narrative Driven Recommendations. *arXiv preprint arXiv:2306.02250* (2023).

[19] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems* 35 (2022), 27730–27744.

[20] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[21] Xie Runfeng, Cui Xiangyang, Yan Zhou, Wang Xin, Xuan Zhanwei, Zhang Kai, et al. 2023. Lkpnr: Llm and kg for personalized news recommendation framework. *arXiv preprint arXiv:2308.12028* (2023).

[22] Scott Sanner, Krisztian Balog, Filip Radlinski, Ben Wedin, and Lucas Dixon. 2023. Large Language Models are Competitive Near Cold-start Recommenders for Language-and Item-based Preferences. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 890–896.

[23] Hongda Sun, Quan Tu, Jinpeng Li, and Rui Yan. 2023. Convntm: conversational neural topic model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 13609–13617.

[24] Hongda Sun, Weikai Xu, Wei Liu, Jian Luan, Bin Wang, Shuo Shang, Ji-Rong Wen, and Rui Yan. 2023. From Indeterminacy to Determinacy: Augmenting Logical Reasoning Capabilities with Large Language Models. *arXiv preprint arXiv:2310.18659* (2023).

[25] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[26] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 world wide web conference*. 1835–1844.

[27] Lei Wang and Ee-Peng Lim. 2023. Zero-Shot Next-Item Recommendation using Large Pretrained Language Models. *arXiv preprint arXiv:2304.03153* (2023).

[28] Lei Wang, Jingsen Zhang, Hao Yang, Zhiyuan Chen, Jiakai Tang, Zeyu Zhang, Xu Chen, Yankai Lin, Ruihua Song, Wayne Xin Zhao, et al. 2023. When large language model based agent meets user behavior analysis: A novel user simulation paradigm. *arXiv preprint ArXiv:2306.02552* (2023).

[29] Wenjie Wang, Xinyu Lin, Fuli Feng, Xiangnan He, and Tat-Seng Chua. 2023. Generative recommendation: Towards next-generation recommender paradigm. *arXiv preprint arXiv:2304.03516* (2023).

[30] Xiaolei Wang, Xinyu Tang, Wayne Xin Zhao, Jingyuan Wang, and Ji-Rong Wen. 2023. Rethinking the Evaluation for Conversational Recommendation in the Era of Large Language Models. *arXiv preprint arXiv:2305.13112* (2023).

[31] Yan Wang, Zhixuan Chu, Xin Ouyang, Simeng Wang, Hongyan Hao, Yue Shen, Jinjie Gu, Siqiao Xue, James Y Zhang, Qing Cui, et al. 2023. Enhancing recommender systems with large language model reasoning graphs. *arXiv preprint arXiv:2308.10835* (2023).

[32] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. Neural News Recommendation with Attentive Multi-View Learning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, Sarit Kraus (Ed.). ijcai.org, 3863–3869. https://doi.org/10.24963/ijcai.2019/536

[33] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. Neural News Recommendation with Multi-Head Self-Attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 6389–6394. https://doi.org/10.18653/v1/D19-1671

[34] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. Neural News Recommendation with Multi-Head Self-Attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, 6388–6393. https://doi.org/10.18653/v1/D19-1671

[35] Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. 2023. Personalized News Recommendation: Methods and Challenges. *ACM Trans. Inf. Syst.* 41, 1 (2023), 24:1–24:50. https://doi.org/10.1145/3530257

[36] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2021. Empowering News Recommendation with Pre-trained Language Models. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai (Eds.). ACM, 1652–1656. https://doi.org/10.1145/3404835.3463069

[37] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. 2020. MIND: A Large-scale Dataset for News Recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault (Eds.). Association for Computational Linguistics, 3597–3606. https://doi.org/10.18653/v1/2020.acl-main.331

[38] Boming Yang, Dairui Liu, Toyotaro Suzumura, Ruihai Dong, and Irene Li. 2023. Going Beyond Local: Global Graph-Enhanced Personalized News Recommendations. *arXiv preprint arXiv:2307.06576* (2023).

[39] 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 preprint arXiv:2305.07001* (2023).

[40] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023).

[41] Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, et al. 2023. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. *arXiv preprint arXiv:2302.09419* (2023).

A APPENDIX

A.1 Example prompts for multi-news narrative generation

Instruction

You are a personalized text generator. First, I will provide you with a news list that includes both the **[main news]** and **[topic-related news]**. Second, I will provide you with user interests, including the **[categories]** and **[topics]** of news that the user is interested in. Based on the input news list and user interests, you are required to generate a **{personalized news summary}** centered around the **[main news]**.

Input

News List:
{"ID": "Main News", "title": "Brett Kavanaugh calls Ruth Bader Ginsburg 'inspiration,' heaps gratitude on allies", "abstract": 'In his first speech as a Supreme Court justice, Brett Kavanaugh heaped "gratitude" on his supporters and hailed Ruth Bader Ginsburg as an "inspiration."", "topic": "Brett Kavanaugh, ..."}
{"ID": "N****", "title": "Ruth Bader Ginsburg misses court due to illness", "abstract": "Supreme Court Justice Ruth Bader Ginsburg was not on the bench for oral arguments Wednesday due to illness ...", "topic": "Supreme Court, Ruth Bader Ginsburg's illness ..."}
{"ID": "N****", "title": "Ruth Bader Ginsburg defends Kavanaugh, Gorsuch as very decent and very smart", "abstract": "Supreme Court Justice Ruth Bader Ginsburg came to the defense of her more conservative colleagues on the bench, Justices Brett ...", "topics": "Ruth Bader Ginsburg, Supreme Court Justices ..."}
{"ID": "N****", "title": "Ruth Bader Ginsburg back at work after stomach bug", "abstract": "Supreme Court Justice Ruth Bader Ginsburg returned to work Friday morning after missing a day of oral arguments ...", "topics": "Ruth Bader Ginsburg ..."}
User Interest:
This user is interested in news about[politics], especially [..., Washington Post, impeachment inquiry, Supreme Court, ...].

Output

{"title": "Brett Kavanaugh Gives First Speech as Justice, Praises Ruth Bader Ginsburg Being World Champion", "category": "politics", "topics": "Brett Kavanaugh, Ruth Bader Ginsburg, Supreme Court", "abstract": "In his first speech as a Supreme Court justice, Brett Kavanaugh expressed gratitude to his supporters and hailed Ruth Bader Ginsburg as an inspiration. Supreme Court Justice Ginsburg, who recently missed court due to illness, defended Kavanaugh and Gorsuch as 'very decent and very smart.' Despite being absent due to a stomach bug earlier this week, Ginsburg returned to work on Friday."}