SEALR: Sequential Emotion-Aware LLM-Based Personalized Recommendation System

Namjun Lee
Sungkyunkwan University
Convergence Program for Social Innovation
Department of Applied Artificial Intelligence
Seoul, South Korea
namejun12000@g.skku.edu

Jaekwang Kim*
Sungkyunkwan University
Convergence Program for Social Innovation
Department of Applied Artificial Intelligence
Seoul, South Korea
linux@skku.edu

Abstract

Large Language Models (LLMs) excel in various NLP tasks but remain underexplored in recommendation systems. This study proposes the **S**equential **E**motion-**A**ware **LLM**-Based Personalized **R**ecommendation System (**SEALR**) to leverage sentiment analysis in user-generated reviews, tracking emotional changes and extracting sentiment labels. It integrates candidate items produced by sequential models with user behavior data into an LLM, enhancing personalization. Experiments on Amazon and Yelp datasets explore the effect of varied candidate pool sizes and instruction-based fine-tuning ratios, demonstrating significant performance gains. The combination of sentiment insights and user behavior data effectively accommodates diverse user preferences and contexts.

CCS Concepts

• **Information systems** → **Recommender systems**.

Keywords

Recommendation, Personalized Recommendation, Large Language Model, Sentiment Analysis

ACM Reference Format:
Namjun Lee and Jaekwang Kim. 2025. SEALR: Sequential Emotion-Aware LLM-Based Personalized Recommendation System. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3726302.3730249

1 Introduction

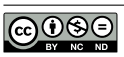
Recent advancements in Large Language Models (LLMs) have significantly improved natural language understanding and generation, achieving strong performance in tasks such as storytelling, numerical reasoning, and semantic analysis [2, 4, 32]. These capabilities have also enabled the integration of LLMs into recommendation systems, especially in tasks like search and ranking [25]. Unlike static embeddings, LLMs represent items through rich textual data and can adapt to new content in real-time [20, 27], while integrating multimodal inputs such as user reviews and metadata [13]. Furthermore, LLMs exhibit strong domain transferability, even in low-resource settings [9, 14].

Despite these benefits, several challenges hinder the use of LLMs in sequential recommendation [16, 24, 31]. High computational cost and token limitations affect real-time applicability [11], while hallucinations from ambiguous inputs can reduce user trust [22]. Additionally, dynamic and sentiment-driven user preferences are difficult to capture with static, parameterized models. Although recent work has fine-tuned LLMs on summarized user-item histories [26], such methods often overlook nuanced emotional feedback in user reviews [29], which is crucial for personalization [1].

To address these challenges, we propose SEALR, a framework designed to combine sentiment-driven user feedback with sequential recommendation models to effectively capture the emotional context of user interactions. Sentiment labels are extracted from user reviews using a RoBERTa model pre-trained on the GoEmotions dataset [17]. Additionally, candidate items are generated through a retrieval model to mitigate hallucination issues and enhance recommendation reliability. To further explore the relationship between recommendation accuracy and diversity, we investigate the effect of varying the candidate pool size. We also apply instruction-based fine-tuning of LLaMA2-7B [23] to enhance personalization. To improve computational efficiency, item IDs are encoded instead of using item names, and different fine-tuning rates are explored to assess the model's ability to generalize effectively. Our approach demonstrates that combining sentiment information with behavioral data enables personalized recommendations that adapt to diverse user needs and scenarios. In summary, our contributions are as follows:

- We propose SEALR, a novel recommendation framework that integrates sentiment labels from user reviews, improves recommendation precision, and addresses the limitations of existing systems.
- By exploring various candidate pool sizes and instruction-based fine-tuning rates, we demonstrate how recommendation performance can be optimized to suit different user scenarios.
- Through experiments using Amazon and Yelp datasets, we validate the effectiveness of SEALR and demonstrate its potential to leverage emotion-aware sequential data for personalized recommendation tasks.

*Corresponding author



SIGIR '25, Padua, Italy

ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3730249

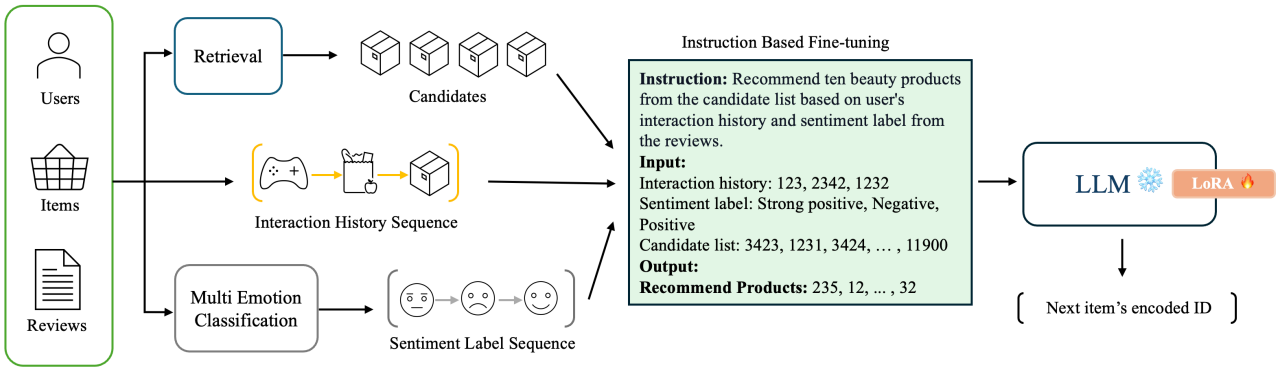


Figure 1: Overall framework of SEALR.

2 Methodology

The proposed SEALR framework aims to provide more personalized recommendations by integrating sequential user interactions, sequential sentiment labels, and candidate items retrieved through a retrieval model. The framework consists of three main stages, and the overall process is illustrated in Figure 1.

2.1 Generation of Sequential Sentiment Labels

To analyze emotions from user reviews, we utilized a RoBERTa model pre-trained on the GoEmotions dataset for multi-label emotion classification. The GoEmotions dataset [3] is a collection of 58,000 carefully curated comments extracted from Reddit, with human annotations for 28 emotion categories. These categories include 12 positive emotions (e.g., joy, admiration, gratitude), 15 negative emotions (e.g., anger, sadness, fear), and one neutral emotion (neutral), making it one of the most comprehensive fine-grained emotion datasets available. The RoBERTa model, trained on this dataset, provides probability distributions over these 28 emotion categories for each input sentence. Since a single sentence can exhibit multiple emotions simultaneously, this model is specifically designed for multi-label emotion classification, allowing it to capture the nuanced emotional context effectively.

To determine the optimal emotion label for each item, we selected the label with the highest accuracy score among the 28 emotion classes. However, using 28 distinct emotion labels can introduce excessive complexity, making it difficult for LLMs to learn key emotion patterns efficiently. To address this issue, we adopted a label consolidation strategy, grouping the emotions into five main sentiment categories: Strong Positive, Positive, Neutral, Negative, and Strong Negative (see Figure 2). This simplification allows the model to capture emotional dynamics more effectively.

This integration, while having the limitation of potentially losing some finer emotional distinctions, was a strategic choice to effectively capture overall emotional patterns and enhance personalized recommendation performance. Additionally, simplifying emotion sequences reduces computational costs and improves the efficiency of emotion-based filtering in recommendations. By integrating sequential affective labels, we refine the representation of user preferences and emotional responses, thereby enhancing the LLM-based personalized recommendation system.

2.2 Candidate Retrieval

To address issues such as hallucination and incomplete recommendations, we employ SASRec [10] as a retrieval module to ground knowledge and filter out irrelevant results. SASRec, a sequential recommendation model that captures temporal patterns in user behavior, generates a smaller, more refined candidate pool, ensuring that only the most relevant items are selected for further processing by the LLM. While SASRec is used in this framework, other sequential recommendation models could also be incorporated for similar candidate selection and filtering processes. By integrating retrieval and recommendation within a unified framework, our approach improves candidate selection efficiency, leading to more accurate and personalized recommendations.

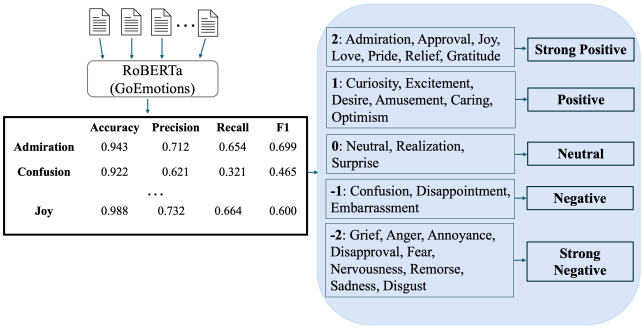


Figure 2: Sentiment label extraction from multi-emotion classification based on user reviews.

2.3 Instruction-Based Fine-Tuning & Item Recommendation

For recommendation generation, we employed LLaMA2-7B as our base model. To optimize the model for recommendation scenarios, we designed instruction-based fine-tuning using a customized prompt template. Instruction-based fine-tuning helps align the LLM's responses with domain-specific recommendation tasks by refining its ability to generate structured outputs and improving interpretability. By incorporating task-specific prompts and supervised learning, this approach enhances the model's ability to understand

```
Instruction: Recommend ten {category} from the candidate
list based on user's interaction history and sentiment label
from the reviews.

Input:
Interaction history: {interaction history sequence}
Sentiment label: {sentiment label sequence}
Candidate list: {candidates}

Output:
Recommend Products: {encoded item id}
```

Figure 3: Template used for instruction-based fine-tuning.

user preferences and provide more relevant recommendations [30]. Our prompt template is shown in Figure 3.

The input includes instructions, sequential user interactions, sentiment labels, and candidate items. For performance evaluation, sequential user interactions and sentiment labels exclude the second-to-last interaction during the evaluation phase. The output consists of the actual target items, where the first and second items correspond to the last and second-to-last items that the user interacted with. To reduce computational overhead, we replaced actual item names with encoded item IDs, as processing the actual names significantly increases token usage, thereby increasing training and inference times.

Additionally, to enable Parameter-Efficient Fine-Tuning (PEFT), we applied Low-Rank Adaptation (LoRA) [8], which optimizes memory usage and accelerates training. To demonstrate the inductive learning capability of LLM, fine-tuning was performed with 10%, 20%, and 50% of the total user data, with 5% of each subset reserved for validation to ensure stable model training. The LLM was trained to capture user preferences and emotional shifts, enhancing its ability to generate highly personalized recommendations.

3 Experiments

3.1 Experimental Setup

3.1.1 Datasets. We use three public benchmarks from the Amazon Product Reviews dataset [5, 15] for the sequential recommendation task: Beauty, Sports and Outdoors, and Toys and Games. Each dataset consists of user reviews and item metadata, covering a diverse range of products in their respective domains. In addition, we include the Yelp[1] dataset, a widely used benchmark for business recommendation tasks. Due to its large volume, we only consider transactions recorded after January 1, 2021.

Table 1: Statistics of datasets after preprocessing.

Datasets	Users	Items	Interactions	Length
Beauty	22,363	12,101	198,502	8.88
Toys	19,412	11,924	167,597	8.63
Sports	35,509	18,357	296,337	8.32
Yelp	21,306	53,511	230,186	10.80

To ensure sufficient interactions for each user and item, we followed the preprocessing procedure described in [6, 28], retaining only the 5-core dataset. Users and items with fewer than five interaction records were iteratively removed. The detailed statistics of these four datasets are presented in Table 1.

[1]https://www.yelp.com/dataset

3.1.2 Baselines. To verify the effectiveness of our proposed method, we compare it with the following representative baselines:

- GRU4Rec [7]. Uses GRU-based recurrent neural networks to model sequential user interactions.
- Caser [21]. Applies convolutional neural networks (CNN) to capture local and global user preference patterns.
- SASRec [10]. Uses self-attention combined with position embeddings for sequence modeling.
- BERT4Rec [19]. Employs a bidirectional transformer with a masked language model (MLM) objective to predict future interactions.
- BSARec [18]. Performs sequential recommendation by combining high and low-frequency information using Fourier transform and improving the self-attention mechanism.
- PALR [26]. A multi-step recommendation system that combines user behavior data and LLMs to generate personalized recommendations.

Since an official implementation of PALR is not available, we used LLaMA2-7B as the user preference summarization and recommendation LLM. To optimize memory usage, we applied LoRA with the same hyperparameter settings as our proposed model.

3.1.3 Metrics. We evaluate performance using the Leave-One-Out (LOO) strategy, where the most recent user interaction serves as the test item, the second-to-last as validation, and the remaining interactions as training data. To assess inductive learning capability, we fine-tune on 50% of users while using the other 50% for evaluation. Following prior work [12], baseline models are evaluated on the entire item set without sampling.

For performance measurement, we use Hit Ratio (HR) to check whether the recommended list contains the target item and Normalized Discounted Cumulative Gain (NDCG) to account for ranking by considering the position of the target item in the list.

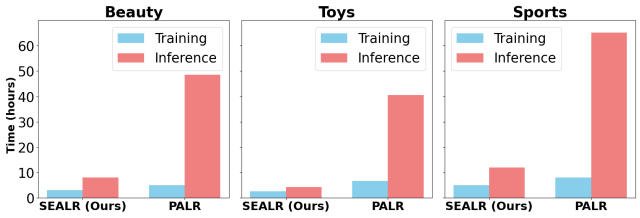


Figure 4: Efficiency improvement of SEALR over an LLM-based sequential recommendation system in training and inference time.

3.2 Result and Analysis

3.2.1 Overall Performance. Our main performance results are presented in Table 2, where the best results are highlighted in bold and the second-best results are underlined. FT denotes fine-tuning with the specified ratio, while Can refers to the number of candidates.

Compared to baselines, our approach consistently outperforms all models across the Beauty, Toys, Sports and Yelp datasets. Traditional sequential recommendation models, such as GRU4Rec, Caser, and SASRec, exhibit significantly lower performance, highlighting

Table 2: Overall recommendation performance, best results are marked in bold, second best results underlined.

Dataset		Beauty				Toys				Sports				Yelp			
Metrics		HR		NDCG		HR		NDCG		HR		NDCG		HR		NDCG	
		@5	@10	@5	@10	@5	@10	@5	@10	@5	@10	@5	@10	@5	@10	@5	@10
GRU4Rec [7]		0.0155	0.0277	0.0092	0.0132	0.0110	0.0205	0.0069	0.0100	0.0115	0.0175	0.0077	0.0096	0.0092	0.0142	0.0056	0.0072
Caser [21]		0.0202	0.0334	0.0121	0.0163	0.0152	0.0250	0.0098	0.0130	0.0092	0.0166	0.0058	0.0082	0.0036	0.0072	0.0022	0.0034
SASRec [10]		0.0276	0.0482	0.0152	0.0218	0.0363	0.0581	0.0202	0.0272	0.0160	0.0283	0.0094	0.0134	0.0143	0.0228	0.0090	0.0117
BERT4Rec [19]		0.0376	0.0631	0.0244	0.0326	0.0443	0.0679	0.0280	0.0356	0.0216	0.0379	0.0136	0.0188	0.0229	0.0365	0.0152	0.0196
BSARec [18]		0.0590	0.0927	0.0354	0.0463	0.0676	0.0987	0.0397	0.0497	0.0339	0.0510	0.0192	0.0247	0.0283	0.0466	0.0179	0.0237
PALR [26]		0.0705	0.1199	0.0426	0.0586	0.0602	0.1195	0.0364	0.0552	0.0516	0.0909	0.0309	0.0436	**0.1032**	0.1804	**0.0617**	0.0864
SEALR (Ours)	FT10 Can50	**0.1000**	**0.1922**	0.0600	0.0895	0.0990	0.1949	0.0596	0.0900	0.1046	**0.2018**	0.0627	**0.0938**	0.1003	0.2005	0.0594	0.0913
	FT20 Can50	**0.1000**	0.1906	**0.0608**	**0.0898**	0.1012	0.1951	0.0616	0.0915	0.1016	0.1960	0.0613	0.0914	0.0991	**0.2046**	0.0594	**0.0930**
	FT50 Can50	0.0936	0.1838	0.0566	0.0855	**0.1015**	**0.1952**	**0.0620**	**0.0918**	**0.1047**	0.1975	**0.0633**	0.0930	0.0933	0.1959	0.0553	0.0880

the effectiveness of LLM-based approaches in capturing user behavior. Even advanced transformer-based models like BERT4Rec and BSARec are surpassed, demonstrating the advantages of instruction-tuned LLMs for sequential recommendation. While PALR achieves slightly better results in HR@5 and NDCG@5 on the Yelp dataset, our model substantially outperforms it in HR@10 and NDCG@10, demonstrating stronger top-k ranking capabilities. This suggests that SEALR is more effective in promoting relevant items at broader recommendation positions.

Among our fine-tuned models, FT10 Can50 achieves the highest HR@10 (0.1922) in the Beauty dataset, while FT20 Can50 attains the best NDCG@10 (0.0898), suggesting that different fine-tuning settings influence performance. In the Toys dataset, FT50 Can50 attains the best performance across all metrics, achieving HR@10 (0.1952) and NDCG@10 (0.0918). Notably, the performance gap between FT10, FT20, and FT50 remains relatively small across datasets, suggesting that LLMs can capture meaningful user-item relationships even with limited fine-tuning, emphasizing their strong inductive learning capabilities.

3.2.2 Impact of Candidate Pool Size on Performance. Table 3 presents the impact of the size of the candidate pool on the performance of SEALR using Amazon datasets. With a candidate pool size of 20, SEALR achieves the highest HR@10 and NDCG@10, suggesting that a smaller, more focused candidate set may enhance accuracy. As the candidate pool size increases, performance gradually declines, indicating that an excessively large candidate pool can introduce noise and reduce precision. These findings suggest that appropriately adjusting the candidate pool size not only optimizes performance but may also help mitigate hallucinations in LLM-based recommendation systems. In particular, an excessively large candidate set increases the likelihood of generating less reliable recommendations, whereas a well-curated pool constrains the selection process, leading to more precise and trustworthy results.

3.2.3 Efficiency of SEALR. Figure 4 compares the training and inference times of SEALR with PALR, an LLM-based sequential recommendation model, across the Beauty, Toys, and Sports datasets. For a fair comparison, we followed the same evaluation setting as PALR, with 20% of the data used for fine-tuning and 50 candidates considered for ranking. The results demonstrate that SEALR significantly improves computational efficiency while maintaining superior recommendation performance. SEALR reduces the average training time by 3 hours compared to PALR. The improvement in inference efficiency is even more pronounced, with a 43.37 hours reduction in average inference time, ensuring a more practical deployment in real-world applications. This efficiency gain is primarily due to the use of encoded item IDs instead of actual item names, which reduces token usage and lowers computational costs. These findings confirm that SEALR not only enhances recommendation accuracy but also drastically improves the computational efficiency, making it a scalable and practical solution for LLM-based sequential recommendation systems.

Table 3: Performance comparison across different candidate pool sizes with 10% of the total users used for fine-tuning.

Dataset	Beauty		Toys		Sports	
Metrics	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
Candidates 20	**0.4819**	**0.2228**	**0.4727**	**0.2191**	**0.5076**	**0.2356**
Candidates 50	0.1922	0.0895	0.1949	0.0900	0.2018	0.0938
Candidates 100	0.0970	0.0451	0.1043	0.0479	0.1012	0.0475

4 Conclusion

This study introduces SEALR, a personalized recommendation approach that enhances performance by integrating sequential user interactions and sentiment labels into an LLM. The proposed method extracts multi-label sentiments from user reviews using a pre-trained LLM and incorporates emotional dynamics into the recommendation process. Experimental results demonstrate that SEALR significantly outperforms baseline models across multiple datasets, improving recommendation accuracy and better capturing user preferences.

Acknowledgments

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) support program(No. IITP-2025-RS-2023-00259497) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation) and was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. IITP-2025-RS-2023-00254129, Graduate School of Metaverse Convergence (Sungkyunkwan University)) and was supported by the Basic Science Research Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. IITP-2025-RS-2024-00346737).

References

[1] Sumaia Mohammed Al-Ghuribi and Shahrul Azman Mohd Noah. 2021. A comprehensive overview of recommender system and sentiment analysis. *arXiv preprint arXiv:2109.08794* (2021).

[2] 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. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[3] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547* (2020).

[4] Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[5] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*. 507–517.

[6] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*. 173–182.

[7] B Hidasi. 2015. Session-based Recommendations with Recurrent Neural Networks. *arXiv preprint arXiv:1511.06939* (2015).

[8] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021).

[9] Jian Jia, Yipei Wang, Yan Li, Honggang Chen, Xuehan Bai, Zhaocheng Liu, Jian Liang, Quan Chen, Han Li, Peng Jiang, et al. 2024. Knowledge Adaptation from Large Language Model to Recommendation for Practical Industrial Application. *arXiv preprint arXiv:2405.03988* (2024).

[10] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*. IEEE, 197–206.

[11] Sein Kim, Hongseok Kang, Seungyoon Choi, Donghyun Kim, Minchul Yang, and Chanyoung Park. 2024. Large language models meet collaborative filtering: An efficient all-round llm-based recommender system. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1395–1406.

[12] Walid Krichene and Steffen Rendle. 2020. On sampled metrics for item recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 1748–1757.

[13] Sichun Luo, Yuxuan Yao, Bowei He, Yinya Huang, Aojun Zhou, Xinyi Zhang, Yuanzhang Xiao, Mingjie Zhan, and Linqi Song. 2024. Integrating large language models into recommendation via mutual augmentation and adaptive aggregation. *arXiv preprint arXiv:2401.13870* (2024).

[14] Weiqing Luo, Chonggang Song, Lingling Yi, and Gong Cheng. 2024. Trawl: External knowledge-enhanced recommendation with llm assistance. *arXiv preprint arXiv:2403.06642* (2024).

[15] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. 43–52.

[16] Bo Peng, Ben Burns, Ziqi Chen, Srinivasan Parthasarathy, and Xia Ning. 2023. Towards Efficient and Effective Adaptation of Large Language Models for Sequential Recommendation. *arXiv preprint arXiv:2310.01612* (2023).

[17] Sam Lowe. 2024. roberta-base-go_emotions (Revision 58b6c5b). doi:10.57967/hf/3548

[18] Yehjin Shin, Jeongwhan Choi, Hyowon Wi, and Noseong Park. 2024. An attentive inductive bias for sequential recommendation beyond the self-attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 8984–8992.

[19] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1441–1450.

[20] Juntao Tan, Shuyuan Xu, Wenyue Hua, Yingqiang Ge, Zelong Li, and Yongfeng Zhang. 2024. Idgenrec: Llm-recsys alignment with textual id learning. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 355–364.

[21] Jiaxi Tang and Ke Wang. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In *Proceedings of the eleventh ACM international conference on web search and data mining*. 565–573.

[22] SM Tonmoy, SM Zaman, Vinija Jain, Anku Rani, Vipula Rawte, Aman Chadha, and Amitava Das. 2024. A comprehensive survey of hallucination mitigation techniques in large language models. *arXiv preprint arXiv:2401.01313* (2024).

[23] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).

[24] Yu Wang, Zhiwei Liu, Jianguo Zhang, Weiran Yao, Shelby Heinecke, and Philip S Yu. 2023. Drdt: Dynamic reflection with divergent thinking for llm-based sequential recommendation. *arXiv preprint arXiv:2312.11336* (2023).

[25] Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, et al. 2024. A survey on large language models for recommendation. *World Wide Web* 27, 5 (2024), 60.

[26] Fan Yang, Zheng Chen, Ziyan Jiang, Eunah Cho, Xiaojiang Huang, and Yanbin Lu. 2023. Palr: Personalization aware llms for recommendation, 2023. *URL: https://arxiv. org/abs/2305.07622* (2023).

[27] Bin Yin, Junjie Xie, Yu Qin, Zixiang Ding, Zhichao Feng, Xiang Li, and Wei Lin. 2023. Heterogeneous knowledge fusion: A novel approach for personalized recommendation via llm. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 599–601.

[28] Zhenrui Yue, Huimin Zeng, Ziyi Kou, Lanyu Shang, and Dong Wang. 2022. Defending substitution-based profile pollution attacks on sequential recommenders. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 59–70.

[29] Kai Zhang, Hao Qian, Qi Liu, Zhiqiang Zhang, Jun Zhou, Jianhui Ma, and Enhong Chen. 2021. Sifn: A sentiment-aware interactive fusion network for review-based item recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 3627–3631.

[30] Shengyu Zhang, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li, Runyi Hu, Tianwei Zhang, Fei Wu, et al. 2023. Instruction tuning for large language models: A survey. *arXiv preprint arXiv:2308.10792* (2023).

[31] Zihuai Zhao, Wenqi Fan, Jiatong Li, Yunqing Liu, Xiaowei Mei, Yiqi Wang, Zhen Wen, Fei Wang, Xiangyu Zhao, Jiliang Tang, et al. 2024. Recommender systems in the era of large language models (llms). *IEEE Transactions on Knowledge and Data Engineering* (2024).

[32] Zhi Zheng, Wenshuo Chao, Zhaopeng Qiu, Hengshu Zhu, and Hui Xiong. 2024. Harnessing large language models for text-rich sequential recommendation. In *Proceedings of the ACM on Web Conference 2024*. 3207–3216.