DUALRec: A Hybrid Sequential and Language Model Framework for Context-Aware Movie Recommendation

Yitong Li
yitong.li3@student.uva.nl
University of Amsterdam
Amsterdam, North Holland, Netherlands

Raoul Grasman
r.p.p.p.grasman@uva.nl
University of Amsterdam
Amsterdam, North Holland, Netherlands

ABSTRACT

Modern recommender systems face the challenge of capturing dynamic, context-rich user preferences. Traditional collaborative filtering and content-based methods struggle with temporal dynamics, while LLMs provide strong semantic reasoning but lack sequential modeling, and LSTMs capture evolving behaviors but miss semantic depth. We propose DUALRec (Dynamic User-Aware Language-based Recommender), which integrates LSTM-based temporal modeling with fine-tuned LLMs for semantically coherent recommendations. Evaluated on MovieLens-1M, DUALRec outperforms diverse baselines across Hit Rate (HR@k), NDCG@k, and genre similarity scores, demonstrating the value of bridging sequence modeling and language reasoning for context-aware recommendation. Code and full paper are available at https://github.com/SatoSakula/DUALRec.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Computing methodologies** → *Neural networks*; *Natural language processing*.

KEYWORDS

Recommender Systems, Sequential Modeling, Large Language Models, LSTM, Multi-modal Recommendations

ACM Reference Format:
Yitong Li and Raoul Grasman. 2025. DUALRec: A Hybrid Sequential and Language Model Framework for Context-Aware Movie Recommendation. In *Proceedings of the 19th ACM Conference on Recommender Systems (RecSys '25), September 22–26, 2025, Prague, Czech Republic*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Recommendation systems, according to the definition by IBM (2024), is an artificial intelligent (AI) system that suggests items to users. And it plays a crucial role in improving user engagement by predicting user preferences based on historical interactions. Traditional approaches like collaborative filtering (CB, Sarwar et al., 2001) and content-based filtering (CBF, Musto et al., 2022) have provided valuable foundations for modern recommenders. However, despite their widespread research and adoption, these approaches do fall short in two tasks: (1) capturing the temporal dynamics of user behavior as preferences evolve overtime, and (2) understanding the semantic richness behind user-item interactions, such as underlying themes, genres, or user intent shifts (Xu et al., 2025).

Traditional models often assumes user preferences are based on past behaviour and static, while recent research has started to leverage contextual information including temporal patterns like user sentiment (Yu et al., 2021), event information and emotional state (Liu et al., 2024; Liu et al., 2023) to structure dynamic recommenders. Large Language Models (LLMs) in this case have shown promising performance in understanding the complex user context and inputs (Yu et al., 2024). With LLMs capability in encoding vast amounts of common sense knowledge and understanding of complex user behaviours in complex scenarios(Wang et al., 2018). Therefore, it is worthwhile to explore how LLMs can benefit recommender systems to better represent users and items and enhance the performance of recommender accuracy (Wang et al., 2018).

The current research proposes a novel hybrid approach, named DUALRec (Dynamic User-Aware Language-based Recommender). In our approach, the LSTM model is first employed to capture the dynamic intent of user preference based on their recent activities. Then the output from LSTM and user movie watching history will then structured together as a natural language prompt template to feed into our choice of four LLMs from DeepSeek and Mistral models (DeepSeek-AI, 2024a; Mistral AI, 2023a, 2023b) as variants to predict the next movie the user might be interested in. And we will evaluate the performance of for DUALRec variants with the usage of four different LLMs using the Hit Rate at n (HR@(1,5)), Normalized Discounted Cumulative Gain at n (NDCG@(1,5)) and Jaccard Genre Similarity matrices.

2 DATASET

The study will utilize the well-established MovieLens 1M dataset (Harper & Konstan, 2015) which is a stable benchmark dataset with 1 million ratings from 6000 users on 4000 movies. To ensure a sufficient signal for learning temporal patterns, the dataset was filtered to include only the top 1,000 most frequently watched movies, allowing the model to focus on movies with adequate user interaction data for training. Each movie is represented through three complementary modalities: (1) unique movie IDs embedded into 128-dimensional vectors, (2) tokenized movie titles using Keras Tokenizer, and (3) multi-label genre information. The dataset was then split into training (70%), validation (15%), and test (15%) sets at the user level to ensure no user history leakage between splits, and simulate real-world scenarios where models are trained on past data and evaluated on future data.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
RecSys '25, September 22–26, 2025, Prague, Czech Republic
© 2025 ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

3 METHODOLOGY

This part introduces the comprehensive methodology for DUALRec, a novel hybrid recommendation system that addresses the fundamental limitations of existing approaches by strategically combining temporal sequence modelling with semantic reasoning.

3.1 Framing the Current Recommendation Task

In a movie recommendation scenario, each user interacts with a series of movies over time. These interactions form a sequential viewing history, where each entry records the specific movie a viewer watched along with the corresponding timestamp. This sequence reflects both the user's evolving interests and the temporal patterns in their viewing behaviour.

Specifically, as shown in Figure 1, for each user, we aim to model the relationship between their past sequence of movie watching history and predict the likely next movie they will watch through our recommender model.

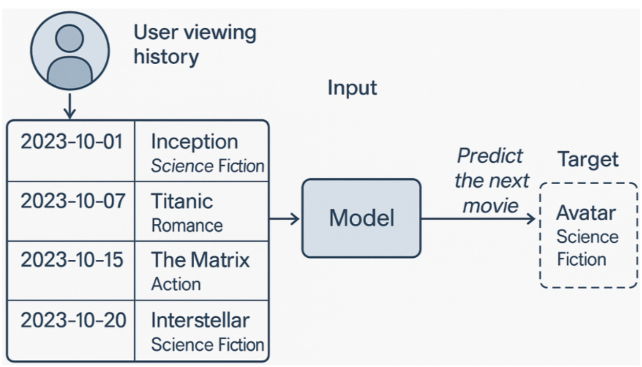


Figure 1: The current recommendation task in a workflow.

3.2 DUALRec Theoretical Framework

The Dynamic User-Aware Language-based Recommender (DUALRec) framework predicts recommendations from a three stage process. In the first stage, we trained an LSTM model on users' movie viewing sequences to perform a sequence prediction task. The LSTM predicts the next movie a user is most likely to watch. This predicted movie, along with the user's recent viewing history, is then used to construct a natural language prompt. This prompt is then fed into the DeepSeek V3, and Mistral 8B models in the second stage, which generates a response with a list of semantically relevant movie titles with release years and genre information as personalized recommendations.

And in the third stage, we further enhanced the model by adopting the LoRA-based PEFT strategy to fine-tune two other open-source language models: DeepSeek-R1-Distill-Qwen-1.5B and Mistral-7B. We evaluate the recommendation quality using several metrics: Hit Rate at k (HR@1, HR@5), Normalized Discounted Cumulative Gain (NDCG@1, NDCG@5), and genre-level Jaccard similarity, which measures the overlap in genre tags between predicted and actual next movies.

3.2.1 *Stage 1: Sequential Behaviour Modelling.*

The first stage of DUALRec models user behavior with a multimodal two-layer LSTM. User–item interactions are grouped by user, ordered chronologically, and represented by movie ID, tokenized title, and genre features. Titles are processed with a Keras tokenizer (vocabulary of 5,000, max length 10), while genres are encoded as 18-dimensional multi-hot vectors reduced via a 64-dimensional dense ReLU layer. At each timestep, the movie embedding, title, and genre representations are concatenated into a unified feature vector, which is then passed through the two-layer LSTM (Figure 2).

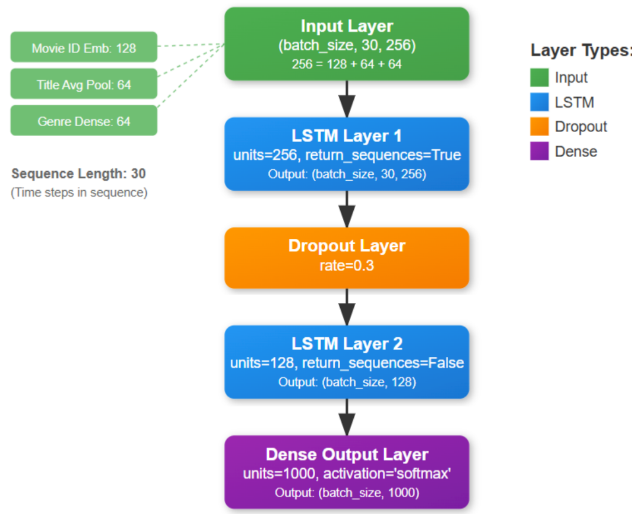


Figure 2: Two-layer LSTM architecture with input shape (30, 256), combining embeddings from movie ID, title, and genre. It outputs top-k movie predictions via a 128-unit LSTM, dropout (0.3), and dense SoftMax over 1,000 classes.

To prepare the LSTM output into Stage 2, we have the predicted movie ID mapped back to its corresponding movie title. And this title is then incorporated into a natural language prompt that is later fed into the large language model (LLM).

3.2.2 *Stage 2: Language-Based Contextual Reasoning.*

In the second stage, DUALRec enhances its recommendation performance using a Large Language Model (LLM) to perform free-form natural language generation tasks. This process builds on the output provided by the Stage 1 LSTM model, combining temporal reasoning with rich, semantic recommendation via prompt-based generation.

Example Prompt to LLM:

Below is a user's movie watching history: - Bug's Life, A (1998) (Animation, Children's, Comedy) - Antz (1998) (Animation, Children's) - Hercules (1997) (Adventure, Animation, Children's, Comedy, Musical) - Mulan (1998) (Animation, Children's) - Pocahontas (1995) (Animation, Children's, Musical, Romance) Based on this, the system (LSTM) recommends: Mission to Mars (2000). Now, as a helpful assistant, recommend 3 more full movie titles with release years and genres that this user would likely enjoy next.

The prompt is then passed through LLM using an OpenRouter API call. The model will return a generated ranked list of 3 additional movie titles. These recommendations are expected to be semantically cohesive, informed by movie themes, user viewing context and genres.

Example LLM Recommendations: *Based on the user's preference for animated, family-friendly films with adventurous and musical elements, here are three recommendations that align with their viewing history: - Tarzan (1999) Genres: Animation, Adventure, Children's, Musical - The Emperor's New Groove (2000) Genres: Animation, Adventure, Children's, Comedy - Lilo & Stitch (2002) Genres: Animation, Children's, Comedy, Science Fiction*

3.2.3 Stage 3: Language Model Fine-Tuning and Post-Generation Optimization. The third stage enhances semantic quality and personalization through LLM fine-tuning and post-generation optimization. Using sequential histories from MovieLens 1M, each user's interactions are chronologically ordered, with the earlier portion as context and the last five movies as ground truth, from which three are randomly selected as targets. Fine-tuning follows an instruction–input–output schema, where prompts combine the five most recent movies with the Stage 1 LSTM's top prediction, ensuring outputs align with user preferences rather than generic suggestions; a semantic re-ranking step is then applied for better intent alignment.

Training employs Hugging Face Transformers with LoRA fine-tuning ($r = 8$, $\alpha = 16$, dropout=0.1) applied to attention projections, using DeepSeek and Mistral tokenizers with 512-token limits and a causal language modeling objective over 3 epochs. The semantic re-ranking is then allocated by computing cosine similarity between SBERT embeddings (all-MiniLM-L6-v2, 384-dim) of LLM-generated titles and the LSTM's top prediction, reordering the three generated titles by similarity to better align with user intent.

4 EXPERIMENTS

4.1 Baselines

To evaluate the performance of the proposed DUALRec framework, we adopt a benchmark strategy similar to that used in the Xu et al. (2025) research, as both models share a fundamental architecture, we used the 11 baselines chosen by Xu et al. (2025) that allow for a more controlled performance comparison.

The baseline models include **Mostpop**, which serves as a popularity-based reference model. **SKNN** (Jannach & Ludewig, 2017), **NARM** (Li et al., 2017), **FPMC** (Rendle et al., 2010), and **STAMP** (Liu et al., 2018) represent classical sequential and session-based models. More recent graph- and intent-based methods include **GCE-GNN** (Wang et al., 2020), **MCPRN** (Wang et al., 2019), **HIDE** (Li et al., 2022), and **Atten-Mixer** (Zhang et al., 2023). We also include transfer learning and language-model baselines such as **UniSRec** (Hou et al., 2022) and **NIR** (Wang & Lim, 2023).

4.2 Evaluation Metrics

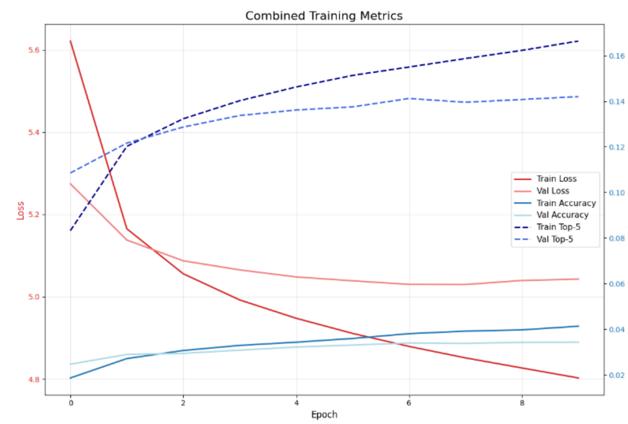
To assess the performance of the recommendation model, we employ standard metrics that are often used to assess recommender models (Xu et al., 2025; Wang & Lim, 2023), including Hit Rate at n (HR@(1,5)) and Normalized Discounted Cumulative Gain at n (NDCG@(1,5)).

In addition to ranking accuracy metrics, we also use Genre Jaccard Similarity to assess the semantic understanding of the model's recommendations. Specifically, we compute the Jaccard similarity between the top-1 recommended movie and the user's actual next movie, focusing on shared genre labels to quantify semantic relevance.

5 RESULTS AND ANALYSIS

5.1 LSTM Model Training Result

The multimodal LSTM was trained for 10 epochs using the ML-1M dataset with the preprocessing pipeline described in Section 4.2. As shown in Figure 3, LSTM demonstrated a learning progression and convergence behaviour.



Metrics	Train	Validation
Loss	4.8026	5.0427
Accuracy	4.14%	3.44%
Top-5 Accuracy	16.64%	14.20%

Figure 3: Performance of LSTM model training with ML-1M dataset.

Both training and validation losses show clear convergence, with training loss decreasing from 5.63 to 4.80 by epoch 10 and validation loss stabilizing around 5.04, indicating effective learning without overfitting. Accuracy also improves steadily, with training accuracy rising from 2.5% to 4.14% and validation accuracy reaching 3.44%, which, though modest, reflects the difficulty of next-item prediction among 1000 candidates. More importantly, top-5 accuracy achieves 16.64% on training and 14.20% on validation, demonstrating the model's ability to capture user intent in realistic recommendation settings where multiple suggestions are typically considered.

5.2 DUALRec Variant Performance Analysis and LLM Fine-tuning

As shown in Figure 4 and Figure 5(a), the DUALRec Mistral 7B fine tuned variant achieves the strongest performance among all four DUALRec configurations, and all 11 baselines with HR@1 and NDCG@ 1 of 0.1847. And its HR@5 performance of 0.2214, and NDCG@5 of 0.2078 although lower than top-performing baselines, it still demonstrated the model's ability to provide relevant recommendations within the top-5 suggestions.

Model	HR@1	HR@5	NDCG@1	NDCG@5
MostPop	0.0004	0.0070	0.0004	0.0053
SLIM	0.1270	0.3600	0.1270	0.2530
FPMC	0.1132	0.3748	0.1132	0.2464
NARM	0.1692	0.5230	0.1692	0.3501
STAMP	0.1584	0.5078	0.1584	0.3367
GCE-GNN	0.1312	0.4748	0.1312	0.3044
MCPRN	0.1434	0.4758	0.1434	0.3157
HIDE	0.1498	0.4998	0.1498	0.3256
Atten-Mixer	0.1490	0.4932	0.1490	0.3216
UniSRec	0.0508	0.2508	0.0508	0.1459
NIR	0.0572	0.2326	0.0572	0.1436
DUALRec – DS V3	**0.1371**	**0.1425**	**0.1371**	**0.1406**
DUALRec – Mistral 8B	**0.0810**	**0.0907**	**0.0810**	**0.0871**
DUALRec – DS Qwen Finetuned	**0.0583**	**0.0605**	**0.0583**	**0.0597**
DUALRec – Mistral 7B Finetuned	**0.1847**	**0.2214**	**0.1847**	**0.2078**

Figure 4: Performance comparison of DUALRec variants versus baselines.

The DUALRec DeepSeek V3 variant showed moderate performance, with HR@1 and NDCG@1 scores of 0.1371, which is higher than 6 out of the 11 baselines. In contrast, the DUALRec Mistral 8B variant performed relatively poorly, with a score of only 0.0810 on the same metrics. The DUALRec DeepSeek Qwen variant exhibits the lowest performance among the four DUALRec variants, with HR@1 of 0.0583 and HR@5 of 0.0605.

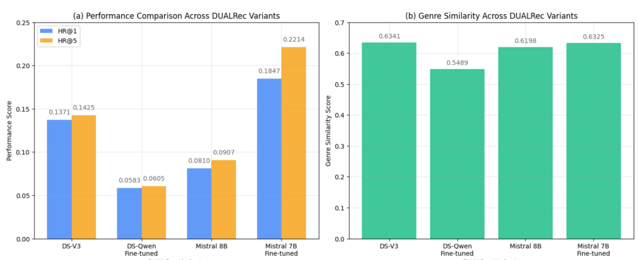


Figure 5: (a) Performance comparison across the choice of LLMs. (b) Genre similarity comparison of the LLMs predicted next movies.

As shown in Figure 5(b), the DeepSeek V3 model achieves the highest genre similarity score of 0.6341, indicating strong semantic alignment with user preference even without fine-tuning. The Mistral 7B fine tuned variant closely follows a score of 0.6325, further confirming its ability to generate both contextual and thematic relevant movie recommendations. In contrast, the DeepSeek-R1-Distill-Qwen-1.5B variant yields a lower score of 0.5489, reflecting its relatively weaker performance in capturing general level intent.

6 DISCUSSION

This study proposed DUALRec, a hybrid recommendation architecture that integrates a multimodal LSTM-based sequential predictor with a prompt-based Large Language Model (LLM) for natural language reasoning. By combining temporal user behavior with contextual inference, DUALRec aims to produce personalized and semantically coherent movie recommendations. Among the evaluated variants, a clear performance gap emerged between DeepSeek V3, DeepSeek-R1-Distill-Qwen-1.5B (fine-tuned), and Mistral 7B (fine-tuned). Notably, DeepSeek V3, despite not being fine-tuned, still outperformed the fine-tuned DeepSeek Qwen variant in both accuracy and genre similarity, suggesting possible overfitting or limited domain alignment in the latter. The fine-tuned Mistral 7B achieved the highest HR@1 and NDCG@1 among all DUALRec variants and baselines while maintaining strong genre alignment, with its architecture designed for instruction following with dense causal and sliding window attention (Mistral AI, 2024a) likely contributing to its effectiveness by better capturing structured prompts.

However, the current study has limitations, including evaluation solely on the MovieLens 1M dataset representing a narrow movie domain and training on offline, static data that limits real-world responsiveness. Future research should extend DUALRec to diverse multimodal datasets incorporating features such as images, video clips, music, or user-generated reviews, while optimizing the architecture for large-scale, real-time inference with incremental learning algorithms to update user representations dynamically. Overall, this study demonstrates the feasibility and effectiveness of combining temporal sequence modeling with large-scale language understanding for recommendation, with future iterations benefiting from broader datasets, multimodal features, and real-time adaptation techniques.

7 ETHICS AND REPRODUCIBILITY

This research uses the publicly available MovieLens-1M dataset following established ethical guidelines for recommendation research. All experiments are conducted with proper data handling procedures, and no personally identifiable information is utilized. Code, experimental configurations, and detailed implementation notes are made available at the provided GitHub repository to ensure full reproducibility of results.

8 ACKNOWLEDGMENTS

We thank the University of Amsterdam for supporting this research. We also acknowledge the Group Lens for providing the open source datasets that made this work possible.

REFERENCES

Bakhshizadeh, M. (2024). Supporting knowledge workers through personal information assistance with context aware recommender systems. In *Proceedings of the 18th ACM Conference on Recommender Systems* (pp. 1296-1301).

Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., & Shah, H. (2016). Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems* (pp. 7-10).

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 191-198).

DeepSeek-AI. (2024a). DeepSeek: Scaling large language models with compute-optimal training and serving. *arXiv preprint arXiv:2401.02954.*

DeepSeek-AI. (2024b). DeepSeek-V3 [Pretrained language model]. Hugging Face. https://huggingface.co/deepseek-ai/DeepSeek-V3

DeepSeek-AI. (2024c). DeepSeek-R1-Distill-Qwen-1.5B [Pretrained language model]. Hugging Face. https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Qwen-1.5B

Donkers, T., Loepp, B., & Ziegler, J. (2017). Sequential user-based recurrent neural network recommendations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (pp. 152–160).

Gao, C., Lei, W., He, X., de Rijke, M., & Chua, T.-S. (2021). Advances and challenges in conversational recommender systems: A survey. *AI Open*, 2, 100-126.

Harper, F. M., & Konstan, J. A. (2015). The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 5(4), 1–19.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 173-182).

Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2015). Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*.

Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2020). The curious case of neural text degeneration. In *Advances in Neural Information Processing Systems 33*.

Hou, Y., Mu, S., Zhao, W. X., Li, Y., Ding, B., & Wen, J.-R. (2022). Towards universal sequence representation learning for recommender systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 585-593).

Hu, E., Li, Y., Shen, Y., Wang, S., Wallis, P., Wang, L., Allen-Zhu, Z., & Chen, W. (2021). LoRA: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Hugging Face. (2023). Parameter-efficient fine-tuning (PEFT). Hugging Face Blog. https://huggingface.co/blog/peft

IBM. (2024). Recommendation engine. Retrieved July 16, 2025, from https://www.ibm.com/think/topics/recommendation-engine

Jagerman, R., Markov, I., & de Rijke, M. (2019). When people change their mind: Off-policy evaluation in non-stationary recommendation environments. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining* (pp. 447-455).

Jannach, D., & Ludewig, M. (2017). When recurrent neural networks meet the neighborhood for session-based recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (pp. 306-310).

Kang, W.-C., & McAuley, J. (2018). Self-attentive sequential recommendation. In *2018 IEEE International Conference on Data Mining* (pp. 197–206). IEEE.

Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T., & Ma, J. (2017). Neural attentive session-based recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 1419-1428).

Li, Y., Gao, C., Luo, H., Jin, D., & Li, Y. (2022). Enhancing hypergraph neural networks with intent disentanglement for session-based recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1997-2002).

Liu, Q., Zeng, Y., Mokhosi, R., & Zhang, H. (2018). STAMP: Short-term attention/memory priority model for session-based recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1831-1839).

Liu, W., Cheng, S., Zeng, D., & Qu, H. (2023). Enhancing document-level event argument extraction with contextual clues and role relevance. *arXiv preprint arXiv:2310.05991*.

Liu, W., Zhou, L., Zeng, D., Xiao, Y., Cheng, S., Zhang, C., Lee, G., Zhang, M., & Chen, W. (2024). Beyond single-event extraction: Towards efficient document-level multi-event argument extraction. *arXiv preprint arXiv:2405.01884*.

Mangrulkar, S., Gugger, S., Debut, L., Belkada, Y., Paul, S., & Bossan, B. (2022). PEFT: State-of-the-art parameter-efficient fine-tuning methods [Computer software]. Hugging Face. https://github.com/huggingface/peft

Mistral AI. (2024a). Mistral-7B-Instruct-v0.3 [Pretrained language model]. Hugging Face. https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3

Mistral AI. (2024b). Mistral-8B-Instruct-2410 [Pretrained language model]. Hugging Face. https://huggingface.co/mistralai/Ministral-8B-Instruct-2410

Musto, C., de Gemmis, M., Lops, P., Narducci, F., & Semeraro, G. (2022). Semantics and content-based recommendations. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 251-295). Springer.

Quadrana, M., Karatzoglou, A., Hidasi, B., & Cremonesi, P. (2017). Personalizing session-based recommendations with hierarchical recurrent neural networks. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (pp. 130–137).

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 3982–3992). Association for Computational Linguistics.

Ren, X., Wei, W., Xia, L., Su, L., Cheng, S., Wang, J., Yin, D., & Huang, C. (2024). Representation learning with large language models for recommendation. In *Proceedings of the ACM Web Conference 2024* (pp. 3464–3475).

Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2009). BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*.

Rendle, S., Freudenthaler, C., & Schmidt-Thieme, L. (2010). Factorizing personalized Markov chains for next-basket recommendation. In *Proceedings of the 19th International Conference on the World Wide Web* (pp. 811–820).

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on the World Wide Web* (pp. 285–295).

Sun, F., Liu, Q., Wu, S., Pei, Y., Lin, Z., & Wang, L. (2019). BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 1441–1450). ACM.

Wang, Q., Li, J., Wang, S., Xing, Q., Niu, R., Kong, H., Li, R., Long, G., Chang, Y., & Zhang, C. (2018). Towards next-generation LLM-based recommender systems: A survey and beyond. *ACM Computing Surveys*.

Wang, S., Hu, L., Wang, Y., Sheng, Q. Z., Orgun, M., & Cao, L. (2019). Modeling multi-purpose sessions for next-item recommendations via mixture-channel purpose routing networks. In *International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence.

Wang, Z., Wei, W., Cong, G., Li, X. L., Mao, X. L., & Qiu, M. (2020). Global context enhanced graph neural networks for session-based recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 169-178).

Wang, L., & Lim, E.-P. (2023). Zero-shot next-item recommendation using large pretrained language models. *arXiv preprint arXiv:2305.00962*.

Wang, S., Zhang, X., Wang, Y., & Ricci, F. (2024). Trustworthy recommender systems. *ACM Transactions on Intelligent Systems and Technology*, 15(4), 1–20.

Xu, X., Zhang, Y., Wang, L., Chen, J., & Li, H. (2025). Enhancing user intent for recommendation systems via large language models. *Preprints*.

Yu, P., Cui, V. Y., & Guan, J. (2021). Text classification by using natural language processing. *Journal of Physics: Conference Series*, 1802(4), 042010. IOP Publishing.

Yu, P., Xu, X., & Wang, J. (2024). Applications of large language models in multimodal learning. *Journal of Computer Technology and Applied Mathematics*, 1(4), 108-116.

Zhang, P., Guo, J., Li, C., Xie, Y., Kim, J. B., Zhang, Y., Xie, X., Wang, H., & Kim, S. (2023). Efficiently leveraging multi-level user intent for session-based recommendation via atten-mixer network. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining* (pp. 168–176). Association for Computing Machinery.