Dialogue Generation for Conversational Recommendation Using User Search Interactions and Knowledge Graphs

Minsung Hwang
School of Computing, KAIST
Daejeon, Republic of Korea
hminsung@kaist.ac.kr

Donggyu Yoon
School of Computing, KAIST
Daejeon, Republic of Korea
justinyoonwk@kaist.ac.kr

Gowoon Hwang
LG Electronics
Seoul, Republic of Korea
gowoon.hwang@lge.com

Kwanghee Jang
LG Electronics
Seoul, Republic of Korea
kwanghee.jang@lge.com

Joyce Jiyoung Whang*
Department of AI Computing, KAIST
Daejeon, Republic of Korea
jjwhang@kaist.ac.kr

Abstract

Large-scale conversational recommendation datasets typically require manual dialogue construction or annotation. We propose SearCl which automatically generates conversational recommendation dialogues grounded in user search and click behaviors. SearCl segments user interactions into query sessions and derives each recommendation target from content selected after a search query. A configurable probabilistic state machine controls the dialogue flow, while two LLM agents acting as the seeker and recommender generate utterances according to turn-level constraints. To support recommendation, the recommender is provided with knowledge graphs constructed from metadata about the content items used in each dialogue, capturing their attributes and relationships. Using SearCl, we generate 3,538 dialogues from content search and click interaction data, covering 1,731 unique target items. Under an LLM-as-a-judge protocol, the resulting dialogues achieve higher quality scores than dialogues from existing benchmarks. These results demonstrate that SearCl enables controllable and knowledge-grounded dataset construction without human annotations.

CCS Concepts

• **Information systems** → **Recommender systems**; • **Computing methodologies** → *Discourse, dialogue and pragmatics.*

Keywords

Conversational Recommender System, Dialogue Generation, Large Language Model, Search-Click Interaction, Knowledge Graphs

ACM Reference Format:
Minsung Hwang, Donggyu Yoon, Gowoon Hwang, Kwanghee Jang, and Joyce Jiyoung Whang. 2026. Dialogue Generation for Conversational Recommendation Using User Search Interactions and Knowledge Graphs . In *Proceedings of The 3rd Workshop on Evaluating and Applying Recommender Systems with Large Language Models at Conference on Information and Knowledge Management (EARL Workshop@CIKM'26).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

*Corresponding author

This work is licensed under a Creative Commons Attribution 4.0 International License.
EARL Workshop@CIKM'26, Rome, Italy
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2539-5/2026/11
https://doi.org/XXXXXXX.XXXXXXX

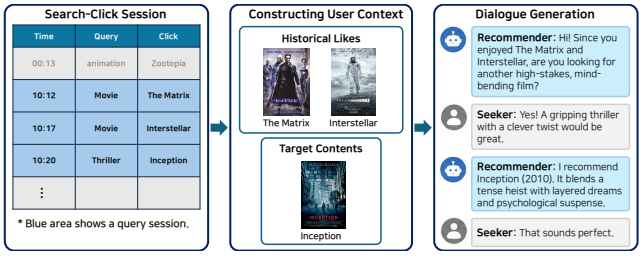


Figure 1: An example of generating a recommendation dialogue from search-click interactions. The interactions are converted into a user profile and a recommendation target, which are then used to construct the dialogue.

1 Introduction

Conversational recommender systems (CRSs) support users in discovering items through multiple conversational turns that elicit preferences, refine requirements, and explain recommendations [16, 18–21]. Developing such systems requires dialogue datasets containing realistic user intents, appropriate recommendation targets, and coherent conversational flows [3, 9, 11, 13]. Existing CRS datasets, including ReDIAL [9] and INSPIRED [3], have played an important role in advancing the field of CRS, but human dialogue collection and annotation are costly and difficult to scale. LLM-based generation can reduce this cost and improve scalability, but existing approaches often rely on manually designed dialogue templates and preference profiles derived from ratings or reviews [10]. This dependence can limit flexibility and make it difficult to account for users' search-and-click behavior.

We propose SearCl, a pipeline for automatically constructing dialogues for conversational recommendation from user search and click behavior. As illustrated in Figure 1, the pipeline groups content search and click interactions into query sessions, which represent continuous periods of user interaction, and transforms each session into a conversational recommendation scenario. Within each query session, content items clicked before a target query are used to construct the user context, while the item clicked immediately after

that query is designated as the recommendation target. These inputs are then converted into a multi-turn dialogue between seeker and recommender LLM agents. A configurable probabilistic state machine determines the conversational structure, and each agent generates utterances according to constraints specified for each turn. To support recommendation explanations, we use a content knowledge graph (KG) as grounding information. KGs provide a flexible representation that can accommodate emerging entities, relations, and facts [6–8]. They can also integrate multiple modalities, including numeric attributes and visual–textual descriptions [1, 5], providing richer evidence for recommendation explanations. Accordingly, the recommender receives bounded two-hop subgraphs extracted from the KG as grounding context. SearCl also applies automatic filtering to remove target title leakage and trivial cases where the query matches the target title. Using SearCl, we generate dialogues from real-world interaction data, demonstrating that the proposed method can produce dialogues grounded in user behavior with controlled conversational structures and without manual construction. Our main contributions are as follows:

- We propose SearCl, a method for generating conversational recommendation dialogues from observed content search and click interaction data.
- We support flexible dialogue design by planning dialogue turns through a configurable probabilistic state machine, whose transition probabilities and available turn types can be adjusted to produce different conversational structures.
- We incorporate content knowledge graphs into dialogue generation, enabling recommendation explanations grounded in structured and auditable evidence.

2 Related Work

A variety of datasets have been developed to support research on conversational recommender systems. ReDIAL contains more than 10,000 human conversations centered on movie recommendations [9], while INSPIRED contains 1,001 human conversations annotated with sociable recommendation strategies such as expressing personal opinions and providing encouragement [3]. Although these datasets provide valuable examples of human recommendation dialogue, their dependence on human participation makes expansion to larger scales, additional domains, and new scenarios costly. To address this limitation, our work focuses on an automated LLM-based approach that constructs behavior-grounded recommendation dialogues without human annotations.

LLM-based dialogue generation has emerged as a scalable alternative to human data collection [10, 12, 17]. LLM-REDIAL fills manually designed dialogue templates with users' historical interactions and reviews [10]. MUSE employs a multi-agent multimodal LLM framework to synthesize dialogues from scenario-grounded user profiles [17]. DistillRecDial uses knowledge distillation to generate dialogues with smaller LLMs while simulating diverse user types [12]. These methods enable large-scale dialogue generation, but they neither incorporate observed search and click behavior into recommendation scenarios nor provide recommendation explanations grounded in KG evidence. In contrast, we propose a dialogue generation method which constructs recommendation scenarios from observed search-click interactions and provides recommendation explanations based on KG constructed by metadata.

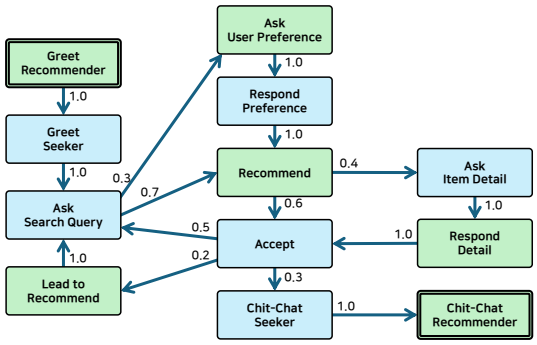


Figure 2: The probabilistic state machine configuration used to plan all dialogues in this study. Blue nodes denote seeker subgoals, whereas green nodes denote recommender subgoals. GREET RECOMMENDER and CHIT-CHAT RECOMMENDER are the initial and terminal states, respectively. Edge labels show the corresponding transition probabilities.

3 Methodology: SearCl

We propose SearCl, a method for constructing conversational recommendation dialogues from content search and click interactions. SearCl first groups these interactions into query sessions, where each observed association between a search query and a clicked item defines a recommendation scenario grounded in the user's search behavior and selected content. For each scenario, SearCl constructs the user context from content items clicked before the corresponding target query within the same query session, uses a configurable probabilistic state machine to plan each dialogue turn, and employs LLM agents assigned the roles of seeker and recommender to generate utterances according to the turn plan. To provide explanations grounded in content metadata, SearCl constructs a knowledge graph representing relations between content items and their attributes and provides it to the recommender agent. Finally, SearCl filters out dialogues containing target leakage or trivial query–title matches.

3.1 Query Session Construction

We assume that the source data contain anonymized user identifiers, timestamps, search queries, content identifiers, domains, and clicked items. We retain movie and TV-show interactions that can be linked to metadata such as genres and cast members. Interactions are ordered by user and grouped into query sessions based on positive inter-event gaps (Δt). Following prior work on web sessionization [2], we fit a two-component Gaussian mixture to ($x = \log(1 + \Delta t)$). We assume that the component with the smaller mean in log-gap space captures pauses within a session, whereas the other component captures longer periods of inactivity between sessions. The inactivity threshold is defined as the first crossover of the weighted component densities between their means, converted back to seconds and clipped to 5–120 minutes. A user-specific model

requires at least five gaps, corresponding to six interactions; otherwise, we use the conventional 30-minute threshold. Under this rule, 5,632 of 65,695 users (8.6%) received user-specific thresholds, while the remaining 91.4% used the fallback. A new session begins when a consecutive-event gap exceeds the applicable threshold.

After constructing the query sessions, we distinguish historical likes from recommendation targets based on the temporal order of interactions within each session. Content items clicked before a target query are used to construct the historical preference context, while the item clicked immediately after the query is designated as the recommendation target. The query determines how the seeker expresses the current recommendation need, and the corresponding target is introduced by the recommender at the designated recommendation turn. Queries without an associated click are retained only as auxiliary context. This construction grounds each scenario in an observed transition from an information need to a selected item rather than in a manually assigned preference.

3.2 Dialogue Turn Planning

For each query session, we plan the dialogue at the turn level by sampling a sequence of subgoals using a predefined probabilistic state machine. Each node represents a seeker or recommender subgoal, and each edge represents a possible transition. Consecutive subgoals assigned to the same speaker are merged into a single dialogue turn. For each resulting turn, the plan specifies the active speaker, one or more dialogue subgoals, the content items to be used for utterance generation, and turn-specific generation constraints. The available subgoals include greeting, expressing a search need, eliciting preferences, requesting item details, recommending an item, accepting a recommendation, and closing the conversation. Both the transition probabilities and the available subgoals can be configured according to the properties of the source dataset and the objectives of dialogue generation, enabling flexible design.

Figure 2 presents the values we used as one example configuration. The main flow begins with a greeting and a search request, followed by either a direct recommendation or an optional preference-elicitation exchange. After receiving a recommendation, the seeker may accept it or request additional details. An accepted recommendation may lead to another query or a closing exchange, allowing multiple query–target pairs from the same query session to appear in one dialogue. Once all targets have been covered, the state machine prevents further transitions to additional recommendations. The resulting turn plan guides the recommendation flow while allowing the LLM agents flexibility in utterance generation.

3.3 Automatic Dialogue Generation

Following the turn plan, separate LLM agents assigned the roles of seeker and recommender generate the dialogue one turn at a time. At each turn, the active agent receives its role, the dialogue history, the current subgoal, the content items to be used in generating the utterance, and turn-specific constraints. The agent then generates a single utterance. Search-request turns prohibit disclosure of the target title, whereas recommendation turns require the designated target and relevant supporting information. This turn-by-turn process enables the agents to produce natural and contextually coherent utterances while closely following the plan.

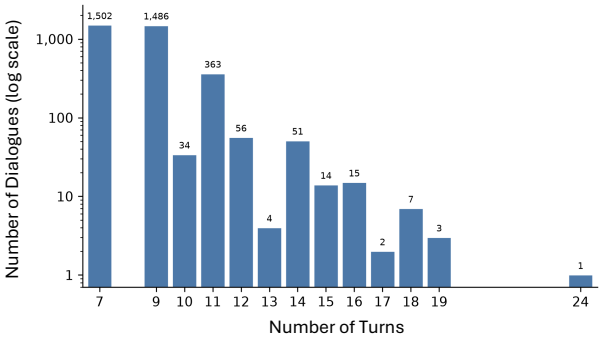


Figure 3: Distribution of dialogue counts by turn-plan length.

The recommender receives the properties of the content items designated for each turn, including plot summaries and other metadata. The recommender additionally receives subgraphs extracted from a content KG built from structured content metadata. The KG uses relations to connect content items with metadata entities such as genres, people, series, and semantic tags, as well as with related content items that share similar textual or visual features. It contains 792,395 entities, 5,618,485 triplets, and 14 relations. For every content item appearing in the user context, as a recommendation target, or in the turn plan, we extract a bounded two-hop subgraph. To prevent the input KG from expanding excessively, we limit the number of neighbors retained at each hop. The union of these subgraphs forms the dialogue-level KG context used by the recommender throughout the dialogue. This context supplies structured evidence for recommendation explanations while limiting prompt length. The KG defines an auditable evidence space but does not guarantee the factuality of every generated statement; factual support is therefore evaluated separately.

Finally, we apply automatic filters to remove dialogues that violate the intended conversation flow or make the recommendation task trivial. Specifically, we discard dialogues in which the target title is revealed before the designated recommendation turn. We also remove query–target pairs for which the original search query directly matches the target title, as these cases reflect direct item lookup rather than recommendation. These post-generation filters complement the turn-specific constraints used during generation: the constraints discourage target leakage, while the filters remove any remaining violations. Table 1 presents a retained dialogue generated from content search and click interactions. The search query is transformed into a conversational request, while the clicked item is introduced only at the designated recommendation turn.

4 Experiments

We evaluate SearCl using TV usage records that were anonymized prior to being provided to the researchers by a leading consumer electronics company in South Korea. The records contain search and viewing activities related to movie and TV show content. Our framework can be applied to other datasets containing search and viewing interactions that can be linked to metadata. Using these records, we generated dialogues using SearCl with `gpt-5.4-mini` [15].

Table 1: An example dialogue generated by SearCl from a query session containing the search query *movie*. The user context includes *Ttoltoli: The Secret of the Baby Dinosaur* as a previously liked item, and *The Super Mario Bros. Movie (2023)* is the recommendation target. The original Korean utterances are translated into English for presentation.

State(s)	Utterance
GREET_RECOMMENDER	Hello! You previously enjoyed children's animation such as *Ttoltoli: The Secret of the Baby Dinosaur*. Would you like to watch something similar today, or should I suggest something with a different atmosphere?
GREET_SEEKER → ASK_SEARCH_QUERY	Hello! Yes, please choose something for me again today. I am looking for a light and entertaining movie that is not too serious and is suitable for watching with my family. Could you recommend one?
ASK_USER_PREFERENCE	For something that you can comfortably watch with your family, animation or comedy would be a good fit. Would you prefer something completely bright and cheerful, or a story with a little more adventure?
RESPOND_PREFERENCE	I would prefer something bright and cheerful!
RECOMMEND	In that case, I recommend *The Super Mario Bros. Movie (2023)*. Its combination of comedy, animation, and adventure makes it bright and entertaining. With voice performances by Jack Black and Anya Taylor-Joy, it is also a fun choice for the whole family.
ACCEPT → CHITCHAT_SEEKER	Great, I will watch that one! Its bright and cheerful atmosphere sounds perfect for a relaxing movie night with my family. If we enjoy it, I will ask for something similar next time.
CHITCHAT_RECOMMENDER	I hope you have a great time watching it with your family! If you enjoy it, I can recommend another bright and lighthearted movie next time.

4.1 Dataset Statistics

Using SearCl, we initially generated 10,000 candidate dialogues from content search and click interaction data. After filtering out dialogues containing target leakage or trivial query–title matches, 3,538 dialogues were retained, covering 1,731 unique recommendation targets. The retained dialogues contain an average of 8.58 turns and follow 164 unique turn plans. Figure 3 presents their distribution by turn-plan length, while Figure 4 presents their distribution across binned numbers of knowledge graph triplets used for dialogue generation. Although the resulting dataset was constructed to evaluate SearCl, it can also serve as a resource for fine-tuning and evaluating conversational recommender system models.

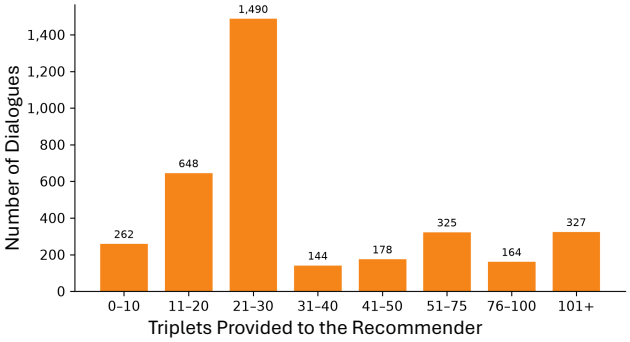


Figure 4: Distribution of the number of triplets provided to the recommender for each dialogue.

4.2 Dialogue Quality Evaluation

We compare our dataset with 100-dialogue samples from ReDIAL [9] and INSPIRED [3]. All corpora are evaluated using the same LLM-as-a-judge protocol [22] with `gpt-5.6-luna` [14]. Following LLM-REDIAL [10], we evaluate utterances in terms of *Fluency*, *Informativeness*, *Logicality*, and *Coherence* using a 0–2 scale. We additionally introduce *Constraint adherence* and *Grounding*, also scored from 0 to 2, to evaluate properties of our method. For conversation-level evaluation, we introduce *Completeness*, *History utilization*, and *Overall quality*, which are scored on a standard five-point Likert scale.

At the utterance level, *Fluency* measures grammatical correctness and naturalness, while *Informativeness* assesses whether an utterance provides useful and specific information. *Logicality* evaluates internal consistency, and *Coherence* measures how well the utterance follows the preceding dialogue context. *Constraint adherence* determines whether the utterance satisfies the conditions of the turn plan, including the designated speaker and subgoal, whether the target content item is included, and whether the utterance does not include test leakage. *Grounding* evaluates whether recommendations are supported by the user context, content metadata, or KG evidence. At the conversation level, *Completeness* measures whether the dialogue contains the expected recommendation stages in an appropriate order. *History utilization* assesses how effectively the interaction history is incorporated, while *Overall quality* captures the dialogue's overall naturalness and consistency as a dialogue.

As shown in Table 2, our dataset receives higher average scores than the sampled ReDIAL [9] and INSPIRED [3] dialogues across all comparable dimensions. At the utterance level, the generated dialogues achieve high fluency, logicality, coherence, and constraint adherence, while also outperforming both comparison corpora in informativeness and grounding. At the conversation level, the largest improvement is observed in overall quality. These results indicate that the proposed method produces coherent dialogues that follow the intended recommendation structure and effectively use the grounding information provided to the generator.

4.3 Ablation Study

To isolate the effect of KG context, we generate a pair of dialogues for every query session in the dataset. Each pair uses the same user context, recommendation target, turn plan, generation model,

Table 2: LLM-based quality evaluation results of dialogues from SearCl and sampled dialogues from ReDIAL [9] and INSPIRED [3]. Utterance-level scores are averaged over all individual utterances, whereas conversation-level scores are averaged over all conversations.

Level	Metric	SearCl	ReDIAL	INSPIRED
Utterance (0–2)	Fluency	**1.98**	1.67	1.71
	Informativeness	**1.35**	0.99	1.03
	Logicality	**1.95**	1.86	1.86
	Coherence	**1.96**	1.89	1.92
	Constraint adherence	**1.99**	–	–
	Grounding	**1.81**	1.52	1.33
Conversation (1–5)	Completeness	**4.67**	3.35	3.99
	History utilization	**4.54**	4.03	3.48
	Overall quality	**4.62**	3.18	3.53

and interaction context, with KG context as the only experimental variable. In the condition without the KG, the recommender still receives all properties of the recommended content, including its plot summary and other metadata; only the relational context derived from the KG is removed. We compare the overall quality scores between the two conditions using a two-sided paired t test. Providing KG context produces a statistically significant increase in mean overall quality, from 4.556 without KG context to 4.620 with it ($p = 6.3 \times 10^{-8}$). This result indicates that relational information about content entities helps the recommender produce more appropriate and well-supported responses beyond what can be achieved using content properties alone.

5 Conclusion

We present SearCl, which constructs conversational recommendation datasets from search and click interactions. SearCl grounds recommendation targets in observed user actions, plans dialogue turns through a configurable probabilistic state machine, and uses a content knowledge graph to support recommendation explanations. Our experiments demonstrate that SearCl generates diverse and coherent dialogues at scale without manual dialogue writing, while achieving higher average quality scores than representative conversational recommendation benchmarks. The transition probabilities and the set of turn types considered during planning are configurable, allowing SearCl to accommodate different dialogue patterns and interaction settings.

The flexibility of SearCl allows it to incorporate richer behavioral signals that distinguish sustained preferences from temporary interest or curiosity and provide evidence of negative preferences. These signals can support dialogue plans involving rejection, dissatisfaction, and unsuccessful recommendations. This configurability also provides a natural extension toward replacing globally specified transition probabilities with context-dependent ones. Recent work has used the current query and dialogue history to determine whether a system should continue the interaction or provide a recommendation [4]. Applying this idea to the finer-grained subgoals in SearCl could produce dialogue plans that dynamically adapt to the current query, dialogue state, and available user context.

GenAI Usage Disclosure

In this work, we used generative AI tools to generate synthetic conversational recommendation data and conduct LLM-based quality evaluation. As described in Sections 3.3 and 4.2, OpenAI's gpt-5.4-mini was used to generate utterances for the seeker and recommender agents under author-defined roles, dialogue histories, turn-level subgoals, content constraints, and a knowledge graph. OpenAI's gpt-5.6-luna was used as a judge to evaluate SearCl and the comparison datasets using author-defined evaluation criteria and prompts. The underlying search-and-click logs, query-session segmentation, user contexts, recommendation targets, dialogue plans, and KG subgraphs were obtained or constructed programmatically. The generated dialogues were processed using predefined constraints and post-generation filters.

Additionally, generative AI tools were used solely for proofreading, grammar correction, and minor wording and stylistic edits to improve the readability of the manuscript. They were not used to develop the research ideas, methods, or scientific claims. The authors reviewed all AI-assisted manuscript changes and take responsibility for the final content of this work, including its text, claims, generated data, evaluation procedure, and conclusions.

We did not use generative AI tools to formulate or refine research hypotheses, design or provide feedback on the research methodology or experiments, implement the proposed methods, preprocess or clean the source interaction logs, construct query sessions, dialogue plans, or knowledge graphs, perform statistical tests, or interpret the experimental results.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2025-00559066).

References

[1] Chanyoung Chung, Jaejun Lee, and Joyce Jiyoung Whang. 2023. Representation Learning on Hyper-Relational and Numeric Knowledge Graphs with Transformers. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 310–322. doi:10.1145/3580305.3599490

[2] Aaron Halfaker, Os Keyes, Daniel Kluver, Jacob Thebault-Spieker, Tien Nguyen, Kate Grandprey-Shores, Anuradha Uduwage, and Morten Warncke-Wang. 2015. User Session Identification Based on Strong Regularities in Inter-activity Time. In *Proceedings of the 24th International Conference on World Wide Web*.

[3] Shirley Anugrah Hayati, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, and Zhou Yu. 2020. INSPIRED: Toward Sociable Recommendation Dialog Systems. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 8142–8152. doi:10.18653/v1/2020.emnlp-main.654

[4] Ming He, Jiwen Wang, Tianyu Ding, and Tong Shen. 2023. Conversation and recommendation: knowledge-enhanced personalized dialog system. *Knowledge and Information Systems* 65, 1 (2023), 261–279.

[5] Jaejun Lee, Chanyoung Chung, Hochang Lee, Sungho Jo, and Joyce Jiyoung Whang. 2023. VISTA: Visual-Textual Knowledge Graph Representation Learning. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. 7314–7328. doi:10.18653/v1/2023.findings-emnlp.488

[6] Jaejun Lee, Chanyoung Chung, and Joyce Jiyoung Whang. 2023. InGram: Inductive Knowledge Graph Embedding via Relation Graphs. In *Proceedings of the 40th International Conference on Machine Learning*. 18796–18809.

[7] Jaejun Lee, Seheon Kim, and Joyce Jiyoung Whang. 2026. Generative Representation Learning on Hyper-relational Knowledge Graphs via Masked Discrete Diffusion. *arXiv preprint arXiv:2605.24064*. doi:10.48550/arXiv.2605.24064

[8] Jaejun Lee and Joyce Jiyoung Whang. 2025. Structure Is All You Need: Structural Representation Learning on Hyper-Relational Knowledge Graphs. In *Proceedings of the 42nd International Conference on Machine Learning*. 33765–33783.

[9] Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards Deep Conversational Recommendations. In

Proceedings of the 32nd International Conference on Neural Information Processing Systems.

[10] Tingting Liang, Chenxin Jin, Lingzhi Wang, Wenqi Fan, Congying Xia, Kai Chen, and Yuyu Yin. 2024. LLM-REDIAL: A Large-Scale Dataset for Conversational Recommender Systems Created from User Behaviors with LLMs. In *Findings of the Association for Computational Linguistics: ACL 2024*. 8926–8939. doi:10.18653/v1/2024.findings-acl.529

[11] Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. Towards Conversational Recommendation over Multi-Type Dialogs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 1036–1049. doi:10.18653/v1/2020.acl-main.98

[12] Alessandro Francesco Maria Martina, Alessandro Petruzzelli, Cataldo Musto, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2025. DistillRecDial: A Knowledge-Distilled Dataset Capturing User Diversity in Conversational Recommendation. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems*. 726–735. doi:10.1145/3705328.3748161

[13] Seungwhan Moon, Pararth Shah, Anuj Kumar, and Rajen Subba. 2019. OpenDialKG: Explainable Conversational Reasoning with Attention-Based Walks over Knowledge Graphs. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 845–854. doi:10.18653/v1/P19-1081

[14] OpenAI. 2026. GPT-5.6 Model Guidance. https://developers.openai.com/api/docs/guides/latest-model. Accessed: 2026-08-17.

[15] OpenAI. 2026. Introducing GPT-5.4 Mini and Nano. https://openai.com/index/introducing-gpt-5-4-mini-and-nano/. Accessed: 2026-08-20.

[16] Subham Raj, Aman Vaibhav Jha, Mayank Anand, and Sriparna Saha. 2026. HARPO: Hierarchical Agentic Reasoning for User-Aligned Conversational Recommendation. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*. 35580–35599. doi:10.18653/v1/2026.acl-long.1646

[17] Zihan Wang, Xiaocui Yang, YongKang Liu, Shi Feng, Daling Wang, and Yifei Zhang. 2025. MUSE: A Multimodal Conversational Recommendation Dataset with Scenario-Grounded User Profiles. In *Findings of the Association for Computational Linguistics: ACL 2025*. 1027–1053. doi:10.18653/v1/2025.findings-acl.58

[18] Yunjia Xi, Weiwen Liu, Jianghao Lin, Bo Chen, Ruiming Tang, Weinan Zhang, and Yong Yu. 2024. MemoCRS: Memory-Enhanced Sequential Conversational Recommender Systems with Large Language Models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 2585–2595. doi:10.1145/3627673.3679599

[19] Feng Xia, Shuo Zhang, and Xi Wang. 2026. When and How to Ask: Dynamic Preference Elicitation Strategies for Conversational Recommendation. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2039–2048. doi:10.1145/3805712.3809624

[20] Zhenye Yang, Jinpeng Chen, Huan Li, Xiongnan Jin, Xuanyang Li, Junwei Zhang, Hongbo Gao, Kaimin Wei, and Senzhang Wang. 2025. STEP: Stepwise Curriculum Learning for Context-Knowledge Fusion in Conversational Recommendation. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. 3824–3833. doi:10.1145/3746252.3761186

[21] Haocheng Yu, Yaxiong Wu, Hao Wang, Wei Guo, Yong Liu, Yawen Li, Yuyang Ye, Junping Du, and Enhong Chen. 2026. Thought-Augmented Planning for LLM-Powered Interactive Recommender Agent. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1821–1832. doi:10.1145/3770854.3780286

[22] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, Vol. 36. 46595–46623.