GPT3RecBot: a universal chatbot recommender of movies, books and music in Telegram

Oleg Lashinin[1,*], Kirill Bykov[2], Marina Ananyeva[1,2] and Sergey Kolesnikov[1]

[1]Tinkoff, 2-Ya Khutorskaya Ulitsa, 38A, bld. 26, Moscow, Russia, 127287

[2]National Research University Higher School of Economics, Myasnitskaya Ulitsa, 20, Moscow, Russia, 101000

Abstract

Recent advances in large language models have extended their potential use cases to different domains. Models such as ChatGPT have an extensive internal knowledge base that enables them to provide answers to various domain-specific queries. In this paper, we explore the potential use of OpenAI's GPT3.5 model as a conversational recommender system. We designed a user-friendly chatbot capable of recommending items in three domains: books, movies, and music. Our study involved collecting explicit feedback from 517 users, and we report the results obtained. The average usefulness of our bot is 4.15 / 5. Our experimental results demonstrate the effectiveness of GPT3.5 as a personalised recommendation system. We hope that our work will inspire further research in this area. Our chatbot is available on the popular messaging platform Telegram under the name @GPT3Recbot, making it accessible to a wide range of users.

Keywords

recommender systems, large language models, user study

1. Introduction

The use of social platforms continues to grow today [1]. Platforms like Twitter [2], YouTube [3] have a lot of content on almost every topic. Since people are usually interested in a limited number of topics, recommender systems play an important role in generating personalised interfaces. They try to understand the user's interests and intentions and select suitable items to be displayed at the top.

There are many types of recommender systems. Some models work for registered users, using the user's past interactions to predict the user's next actions. Other methods, such as session-based recommenders, work even when the user is not authenticated. Such systems analyse the actions in the current session and try to predict the actions in the next session. Currently, state-of-the-art next-item recommender systems such as BERT4Rec correctly guess the next item 10%-30% of the time [4]. Although this is an incredible improvement over non-personalised methods, we are still a long way from more accurate prediction in such experimental setups.

There are many reasons for this, but one is that models deal with implicit feedback and try to understand the user's intentions implicitly. Fortunately, conversational recommenders can overcome this limitation. They make it possible to interact with the user online and better understand the user's intentions in the current moment. However, most of the proposed methods are based on some heuristics or complicated architectures [5, 6], but may fail to respond to the user in a meaningful way [7].

However, recent advances in Large Language Models (LLMs) have demonstrated a notable influence[8]. The recently released ChatGPT[1] chat model from OpenAI can solve many complex problems and learn some facts from context [9]. This is made possible by the large size of the model, extensive training techniques and utilization of high-quality data. In addition, the model has been trained on large datasets and therefore has extensive internal knowledge. Recent work has shown that such knowledge can be used for a variety of tasks [10]. However, the use of OpenAI models for recommender systems has not been well studied.

In this paper we investigate whether OpenAI models such as GPT3.5 can provide good recommendations to users. The contributions of the paper can be listed as follows:

- We describe a simple but effective scenario for using the GPT3.5 model for personalised recommendations. It allows to collect users' interests, provide them with suggestions in different domains such as movies, books and music, and collect explicit feedback on the quality of the recommendations.
- We conduct a user study with 517 real people and report the results. Our experiments demonstrate

Fifth Knowledge-aware and Conversational Recommender Systems (KaRS) Workshop @ RecSys 2023, September 18–22 2023, Singapore.

*Corresponding author.

fotol764@gmail.com (O. Lashinin); kvbykov@edu.hse.ru (K. Bykov); mananeva@hse.ru (M. Ananyeva); s.s.kolesnikov@tinkoff.ru (S. Kolesnikov)

0000-0001-8894-9592 (O. Lashinin); 1234-5678-9012 (M. Ananyeva)

[1]https://openai.com/blog/chatgpt

the effectiveness of GPT3.5 as a personalised recommendation system.

- We integrated our GPT3RecBot chatbot into Telegarm, one of the most popular messengers in the world.

2. Related Work

Conversational Recommenders. The authors of the survey [11] presented a typical conversational recommender system (CRS) architecture, which includes five components dedicated to different purposes. Due to their complexity, CRSs face many challenges [12], such as question-based user preference elicitation and dialogue understanding and generation. The authors of [6] distinguish between different types of utterances in CRSs. One of the simplest approaches is ”System is Active, User is Passive” (SAUP), where a bot asks direct questions and users respond. We choose this approach as a first step in using ChatGPT for recommender systems, and leave the other types of utterances as future work.

Approaches that follow the SAUR paradigm consist of several complicated components. Paper [13] presents a multi-memory network with query, question and search modules. The more recent CPR approach [14] models conversational recommendation as an interactive path reasoning problem on a graph. In another work, [15], researchers generate appropriate questions and recommendations taking into account online feedback from users. In general, most of the currently developed CRSs are based on complex components and are only trained for a specific domain. In addition, most of the work is evaluated using offline experiments. Some authors have conducted user studies on music [16, 17] and book [18] recommendations. As CRSs are interactive services, it seems important to conduct user studies [7].

ChatGPT. Thanks to the rapid development of large-scale language models [19], the quality of dialogue understanding has improved. ChatGPT is a good example. It has attracted the attention of researchers from physics [20], mathematics [21], computational biology [22], etc. Researchers have adopted ChatGPT for various fields of machine learning such as natural language processing [23], stock market prediction [24], binary classification [25] etc. ChatGPT was published on 30 Nov 2022, and the number of papers on Google Scholar related to ChatGPT is about 10000 as of 9 May 2023. Therefore, it is difficult to describe all the cases, but it is definitely a topic that is intensively studied by the research community.

The possible application of ChatGPT to recommender systems has not yet been well studied, due to the short time since publication. A very recent paper [26] provided a case study of ChatGPT recommendations in three typical domains: entertainment (music), high cost (smartphone) and service (travel). The authors found some limitations, such as the use of outdated data, the generation of false and misleading text, and ignorance of some facts. The researchers concluded that this model has great potential for supporting recommendation tasks. In contrast to this article, we conduct a user study that allows the calculation of qualitative metrics. In [27] the authors use ChatGPT for personalised recommendations via chat. They provide a case study for cross-domain and cold-start recommendations. In addition, this paper includes offline experiments on MovieLens-100k where the proposed ChatRec outperforms a state-of-the-art conventional recommender LightGCN. Another paper [28] is less related to our research, but still suggests using ChatGPT for bibliometricians to build recommender systems that suggest relevant articles.

To the best of our knowledge, we have not found any papers that present the results of a user study on the quality of ChatGPT's personalised recommendations. We hope that our findings will inspire new research directions in this area.

3. GPT3RecBot

In this section, we describe our GPT3RecBot in terms of dialogue scheme and implementation for end users. We have implemented a chatbot in Telegram[3] that allows users to ask for recommendations on different types of content (music, movies and books). Our bot prompts them to rate the bot based on three distinct metrics, namely Reality, Usefulness, and Recommendation Quality, using a scale from 1 to 5. The Telegram bot is available to the public[4].

3.1. Dialogue scheme

The chat flow applied to all users in GPT3RecBot is shown in Figure 1. We have implemented a SUAP paradigm where only GPT3RecBot asks questions and the user answers. During the interaction process between the user and our application, we have several steps. The first step is initialisation, where the bot randomly assigns the user to one of three test groups. These groups allow us to test three different prompts for GPT3.5 in the user study.

After initialisation, we are ready to communicate with the user and ask them to select the type of content. We do not ask for specific types of content and the users choose for themselves. The next step is to ask explicitly about the user's interests. This helps ChatGPT to understand the user's preferences. The next question is asked to users in the second and third test groups. This is the question

[2]BPMN - Business Process Model and Notation diagram
[3]https://telegram.org/
[4]https://t.me/GPT3Recbot

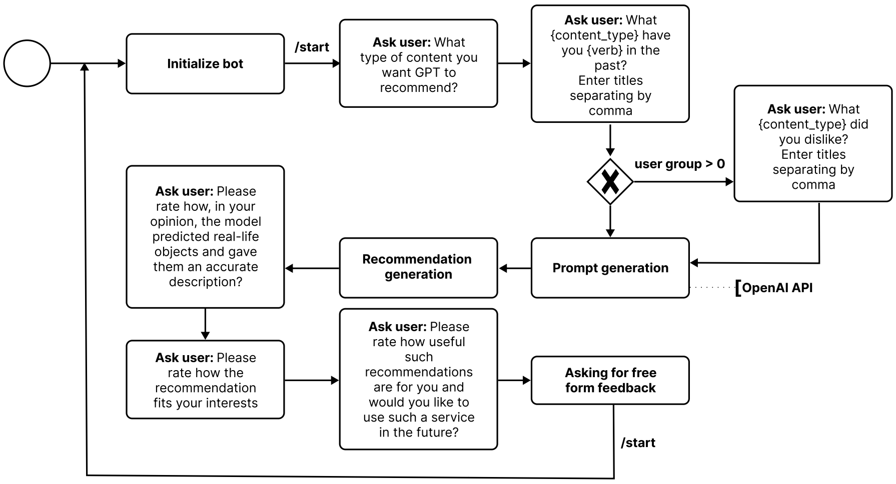

Figure 1: The full BPMN[2] diagram of the flow that the user can go through when using the experimental recommendation bot. The starting point of the flow is the circle, which signals the start of the bot instance. A diamond with a cross inside it indicates an exclusive path, which is identical to if statements in programming.

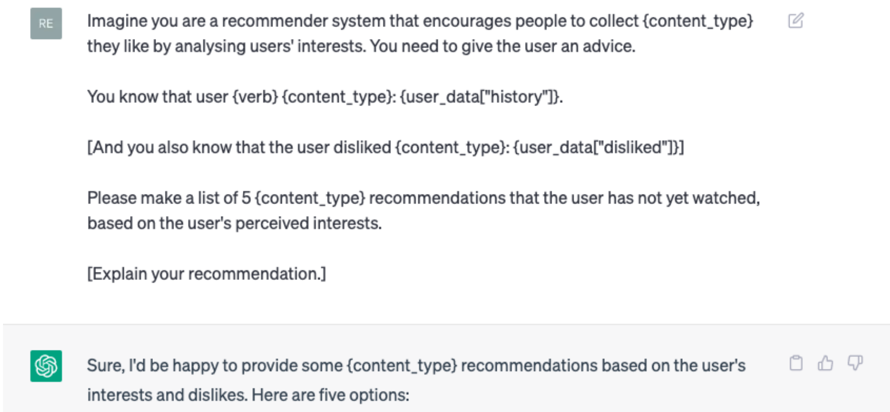

Figure 2: Our prompt example is in the interactive mode of ChatGPT. The prompt with inserted user responses depends on the content type and user preferences. Our user study has three different test groups. The prompts for the second and third groups contain information about the user's dislikes. The third group contains a prompt to explain recommendations. A full ChatGPT response to the prompt pattern is hidden due to space limitations.

about the content the user does not like. We expect the model to use this information to further restrict the types of content genres that are recommended.

After the questions have been asked, the prompt generation phase begins, whereby ChatGPT is supplied with generated prompts. During this phase, a unique prompt is generated for the bot, making it specific for each test group G, by filling in pre-existing templates with user responses. We keep three different prompt templates, and each subsequent group's prompt extends the previous one. The example is shown in Figure 2. After generating the prompt, we send it to gpt-3.5-turbo (training data until September 2021) using the OpenAI API, which provides us with a convenient interface to use the power of ChatGPT in our application. After a waiting period, the answer is sent back to the user.

We then ask the user to rate the reality of the recommendation, which measures how realistic the recommended items and descriptions seemed to the user. The next question is about relevance to the user, which is a measure of how well the model matches the user's preferences. The last is usefulness, where we measure whether GPT3RecBot can help users in their daily lives. In addition, users can send us an open-ended feedback

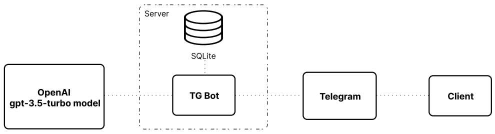

Figure 3: Technical architecture of the implemented applications, including the main components and how they are connected. A connection is a direct communication between two components, e.g. between the OpenAI API and the bot instance of a server application.

form.

3.2. Implementation

The architecture described in this section is typical of messaging bots. It's shown in Figure 3. We have a simple client-server application, where the client is Telegram and the server is our own application. Let us describe the communication flow between the main components we have in the whole system.

A user starts the bot and sends a start command using the Telegram client when he decides to try our application and participate in the experiment. The Telegram client sends calls to our server application via the Telegram API. This process continues until there are no more requests for messages. The main application, the core of the server, stores all conversations and talks to OpenAI to get recommendations via ChatGPT based on the user's specific requests when it's time. All responses, messages and technical information (logs) are stored in the database, which is based on SQLite [29], a relational database popular for small projects. The reason for choosing this database is that it can handle the full load of the system. It is convenient for research purposes to download the database from the server and monitor the experiment online, as this database stores all the information in a single file. At the end of the full run, we ask the user if they want to try again. This creates a loop and allows the user to do a few iterations.

Implementing a bot in Telegram for messaging with users is free. The messenger has reached 550 million monthly active users in 2022 [30] and it is quite easy to register new users. Use of the OpenAI API is paid. However, OpenAI grants $18 for new accounts. The price is $0.002 / 1K tokens actual on 9 May 2023. The average length of our requests is about 100 tokens, according to the user study. This means that we can get about 5000 responses from GPT3.5 for free and continue to get them for $0.0002 / response.

We will maintain this bot until it reaches the limit for free responses. At the time of submission, we have about 4500 responses available to the OpenAI API, which seems sufficient to demonstrate the bot to conference attendees. If we reach this limit before the conference, we will provide an opportunity to try it via another OpenAI account belonging to another author of this paper. In case of force majeure, anyone can run our GPT3RecBot as soon as we release an implementation code under the MIT licence[5].

4. User study design

Statement on ethics. Our respondents shared their interests without providing any personal information. As this was done through a crowdsourcing platform, participants knew that their ratings would be used for anonymised aggregation. We informed them that this information would only be used for research purposes. The Research Ethics Committee (REC) of the HSE university approved this study.

Description. We conducted a user study to evaluate the potential of GPT3RecBot. As GPT3RecBot is a chatbot that recommends items on request, real users were involved in the experiments. We used a crowdsourcing platform to get high quality feedback on the bot. All crowdworkers were asked to read the questions carefully and share their thoughts on the quality of the recommendations and the formatting of the recommendations. We asked students from an HSE university to test our chatbot. It is important to note that respondents chose the type of content they wanted to be recommended. For example, some people don't like to read books, but they are music fans, so they are likely to be interested in music recommendations. We believe that a set of music, book and movie recommendations can cover the interests of the majority of people. However, in future research, we aim to expand the applicability of our methodology to encompass a wider range of domains.

To gain some insight into prompt engineering for GPT-based recommendation models, we randomly divided the users into three groups, with each user being assigned to a group for the duration of the study. The first group was only asked about their preferences in one selected domain. In the second group, the users had to indicate the items they did not like. This information could theoretically help the model to filter recommendations and avoid suggesting obviously irrelevant content. According to recent research [26], ChatGPT are said to be good explainers. In preliminary experiments, we found that GPT3.5 could explain its recommendations to users, even linking items to their interests. So we added a third group in which GPT3.5 was instructed to explain its recommendations. We assumed that explaining recommendations could improve the user experience, based on the results of other studies [31, 32].

[5]https://github.com/AnonymousRecsys/tgbot

Table 1
Overall results. Users chose between 3 different domains and there were 3 different prompts for GPT-3.5. Ratings are averaged. The best score within a pair (domain, metric) is in bold. The best score between domains is underlined. Significant improvements are marked with † (one-way ANOVA + Turkey's HSD test [33] for pairwise comparison).

	Books			Movies			Music		
	1st group	2nd group	3rd group	1st group	2nd group	3rd group	1st group	2nd group	3rd group
User's Interests	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠
User's Dislikes		⊠	⊠		⊠	⊠		⊠	⊠
Explanations			⊠			⊠			⊠
# users per domain		109			269			139	
Reality		4.23†			4.05			3.86	
Recommendation quality		3.93			3.89			3.85	
Usefulness		4.39†			4.15			3.98	
Diversity		0.68			0.65			0.9†	
# users per group	39	33	37	98	90	81	62	29	48
Reality	4.18	**4.37**†	4.18	3.95	**4.14**	4.09	3.73	**4.26**†	3.8
Recommendation quality	3.79	**4.23**	3.82	3.77	3.9	**4.04**	3.78	3.85	**3.93**
Usefulness	4.29	**4.73**†	4.21	4.1	4.19	**4.2**	3.82	4.07	**4.13**
Diversity	0.61	0.64	**0.86**	0.77	0.6	**0.94**	0.68	0.65	**0.9**

5. Results

517 people completed our task and left feedback. They gave explicit feedback on three different metrics, such as the reality of the suggestions, the relevance of the recommended items and the usefulness of GPT3RecBot. Note that we do not count users who did not give explicit feedback. The summarised results are listed in Table 1. There are a few things to note.

Firstly, people chose films more often than books and music combined. We shuffled the domain selection buttons during the experiments, so this bias is not introduced by the GPT3RecBot interface. It is presumed that among the users who participated in our system's testing, films watching is a more prevalent hobby compared to books or music. The other suggestion is that people are most likely to want to try out movie recommendations, as they have experienced with movie recommendation engines in the past. We leave the investigation of this effect to future work.

Secondly, it is clear that there is a difference in user ratings depending on the type of content. GPT3.5 may have a different knowledge capacity for different types of content. The best results were obtained when recommending books, which received the highest ratings in all questions asked. The reason for this may be that ChatGPT has more knowledge about text-based content and is able to process this type of content. This is followed by films, where it also performed well. The poorer performance may also be explained by the information available for GPT. Films have descriptions that are much shorter than book texts. The worst performance is for music recommendations. When analysing free-form questions, we were able to catch fake music titles generated by ChatGPT. It may try to synthesise additional information if it is not available in sufficient quantity in the training data.

If we analyse the results within groups, we can see that GPT3.5 significantly improves its recommendations when it is aware of negative user experiences, which may help to filter recommendations. Asking GPT3.5 to explain its recommendations improved the results for music and film recommendations, which may have helped to convince users of the appropriateness of the recommendations generated.

To test the hypothesis that GPT only recommends a limited number of content titles, we measured content diversity, which is the proportion of unique recommended content to the total number of content titles [34]. We can clearly see in Table 1 that ChatGPT is able to provide users with content that is different and new to them. In Figure 4 we show 5 of the most popular recommended movies and books. The model recommended Nineteen Eighty-Four[6] 16% of the time. Other items were recommended less than 4% of the time. The top recommended items appear to be popular and this may be due to popularity bias [35].

For the open-ended question, we had to come up with

[6]Nineteen Eighty-Four is a dystopian social science fiction novel and cautionary tale by George Orwell.

Table 2
Summary of the feedback collected in an open form. Reviews are separated by Distillbert into positive and negative classes. Then a summarisation is performed for each domain and class.

	Negative	Positive
Books	This form is convenient, interesting to use and provides a good selection of books for the future. It would be great if it also described the genre. The first three recommendations were to the author's liking, but they chose the wrong genre. The bot makes good recommendations, but most of the choices are those that have already been read. It is useful as a handy assistant. The recommendation of a collection of poems is based on preference, but not on the theme of war	The narrator is looking for more thrillers and detective stories. They have a good selection of recommendations, 4/5 of which are known and read. They like the format of the recommendations, which is simple and clear. They hope that the next recommendation will be more interesting.
Movies	The bot correctly identified genres and recommended films, but it would be nice to be able to select films released by year and add additional criteria to get a more satisfactory recommendation. It would be ideal to link a conditional account from IMDB to the bot. The AI should be more selective and understand general preferences to make recommendations for films, not just popular world films, but also lesser known good films. The bot should be able to recognise genres and make a recommendation based on the person's interests	The selection, the prompt and quick response, the clear way of providing information, the clarity and format to meet the needs were all appreciated. The service was good and the bot made good recommendations for each genre, making it convenient and easy to register.
Music	The idea of the bot is interesting, but the description of the recommended compositions is imprecise. The format and quality are good, but the recommendations need to be refined with interest filters. This service is useful for discovering new music and new artists, but it does not match the user's interests.	The most important details are that a bot can be used to select music to listen to, and that the recommendations are of high quality and match preferred genres. The service is convenient and efficient, and the recommendations are of high quality and in preferred genres.

a special method to process such a large amount of feedback. First, we split all the feedback into domains and then used the Distilbert model[7] for sentiment analysis [36] and selected only those reviews where the model was at least 70% sure about the sentiment of the feedback. We then combined the feedback into a single text and used the Quillbot[8] to summarise it. The results are shown in Table 2. The responses in the negative and positive reviews seem a little contradictory. Users who left negative feedback point out that GPT3RecBot does not correctly guess the user's intersets. However, users who left positive feedback wrote the opposite opinion. It is important to note that our bot received 82% positive feedback and 18% negative feedback according to Distillbert's classification.

[7] distilbert-base-uncased-finetuned-sst-2-english available at huggingface.co, revision af0f99b
[8] https://quillbot.com/summarize

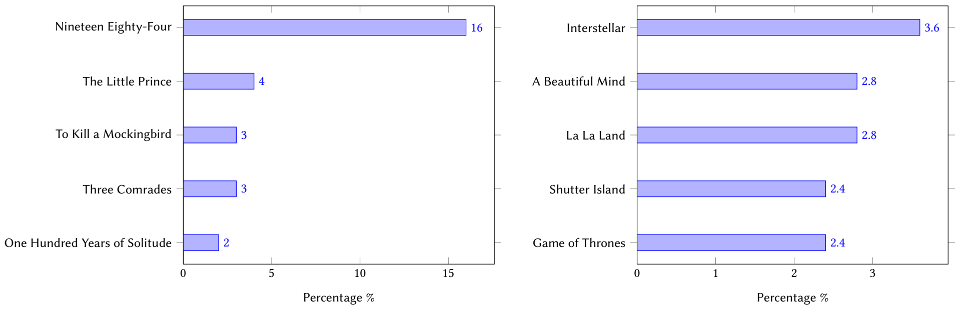

Figure 4: Top 5 recommended movies and books in our user study. With the exception of Nineteen Eighty-Four, the other items are recommended less than 4% of the time.

6. Conclusion

In this paper, we present GPT3RecBot, which can personally recommend movies, books or music. This bot is available to Telegram users for research purposes, and each user has 10 free requests. We share a chat flow, implementation code, and a prompt used to facilitate future research. Our paper is the first to present the results of a user study on the recommendation quality of ChatGPT. 517 users participated in the study and gave an average rating of 4.01 / 5 for reality of recommended content, 3.86 / 5 for relevance and 4.15 / 5 for usefulness of GPT3RecBot. We hope that such high ratings will increase the interest of the research community in future work.

References

[1] E. Ortiz-Ospina, M. Roser, The rise of social media, Our world in data (2023).

[2] V. W. Anelli, S. Kalloori, B. Ferwerda, L. Belli, A. Tejani, F. Portman, A. Lung-Yut-Fong, B. Chamberlain, Y. Xie, J. Hunt, et al., Recsys 2021 challenge workshop: Fairness-aware engagement prediction at scale on twitter's home timeline, in: Proceedings of the 15th ACM Conference on Recommender Systems, 2021, pp. 819–824.

[3] P. Covington, J. Adams, E. Sargin, Deep neural networks for youtube recommendations, in: Proceedings of the 10th ACM conference on recommender systems, 2016, pp. 191–198.

[4] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, P. Jiang, Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer, in: Proceedings of the 28th ACM international conference on information and knowledge management, 2019, pp. 1441–1450.

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

[6] Z. Fu, Y. Xian, Y. Zhang, Y. Zhang, Tutorial on conversational recommendation systems, in: Proceedings of the 14th ACM Conference on Recommender Systems, 2020, pp. 751–753.

[7] A. Manzoor, D. Jannach, Conversational recommendation based on end-to-end learning: How far are we?, Computers in Human Behavior Reports 4 (2021) 100139.

[8] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al., A survey of large language models, arXiv preprint arXiv:2303.18223 (2023).

[9] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.

[10] X. Hu, Y. Tian, K. Nagato, M. Nakao, A. Liu, Opportunities and challenges of chatgpt for design knowledge management, arXiv preprint arXiv:2304.02796 (2023).

[11] D. Jannach, A. Manzoor, W. Cai, L. Chen, A survey on conversational recommender systems, ACM Computing Surveys (CSUR) 54 (2021) 1–36.

[12] C. Gao, W. Lei, X. He, M. de Rijke, T.-S. Chua, Advances and challenges in conversational recommender systems: A survey, AI Open 2 (2021) 100–126. URL: https://www.sciencedirect.com/science/article/pii/S2666651021000164. doi:https://doi.org/10.1016/j.aiopen.2021.06.002.

[13] Y. Zhang, X. Chen, Q. Ai, L. Yang, W. B. Croft, Towards conversational search and recommendation: System ask, user respond, in: Proceedings of the 27th acm international conference on information and knowledge management, 2018, pp. 177–186.

[14] W. Lei, G. Zhang, X. He, Y. Miao, X. Wang, L. Chen, T.-S. Chua, Interactive path reason-

ing on graph for conversational recommendation, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 2073–2083. URL: https://doi.org/10.1145/3394486.3403258. doi:10.1145/3394486.3403258.

[15] W. Lei, X. He, Y. Miao, Q. Wu, R. Hong, M.-Y. Kan, T.-S. Chua, Estimation-action-reflection: Towards deep interaction between conversational and recommender systems, in: Proceedings of the 13th International Conference on Web Search and Data Mining, WSDM '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 304–312. URL: https://doi.org/10.1145/3336191.3371769. doi:10.1145/3336191.3371769.

[16] Y. Jin, W. Cai, L. Chen, N. N. Htun, K. Verbert, Musicbot: Evaluating critiquing-based music recommenders with conversational interaction, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 951–960. URL: https://doi.org/10.1145/3357384.3357923. doi:10.1145/3357384.3357923.

[17] W. Cai, Y. Jin, L. Chen, Impacts of personal characteristics on user trust in conversational recommender systems, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22, Association for Computing Machinery, New York, NY, USA, 2022. URL: https://doi.org/10.1145/3491102.3517471. doi:10.1145/3491102.3517471.

[18] A. Ghazimatin, S. Pramanik, R. Saha Roy, G. Weikum, Elixir: learning from user feedback on explanations to improve recommender models, in: Proceedings of the Web Conference 2021, 2021, pp. 3850–3860.

[19] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., Training language models to follow instructions with human feedback, Advances in Neural Information Processing Systems 35 (2022) 27730–27744.

[20] J. Holmes, Z. Liu, L. Zhang, Y. Ding, T. T. Sio, L. A. McGee, J. B. Ashman, X. Li, T. Liu, J. Shen, et al., Evaluating large language models on a highly-specialized topic, radiation oncology physics, arXiv preprint arXiv:2304.01938 (2023).

[21] S. Frieder, L. Pinchetti, R.-R. Griffiths, T. Salvatori, T. Lukasiewicz, P. C. Petersen, A. Chevalier, J. Berner, Mathematical capabilities of chatgpt, arXiv preprint arXiv:2301.13867 (2023).

[22] T. Lubiana, R. Lopes, P. Medeiros, J. C. Silva, A. N. A. Goncalves, V. Maracaja-Coutinho, H. I. Nakaya, Ten quick tips for harnessing the power of chatgpt/gpt-4 in computational biology, arXiv preprint arXiv:2303.16429 (2023).

[23] C. Qin, A. Zhang, Z. Zhang, J. Chen, M. Yasunaga, D. Yang, Is chatgpt a general-purpose natural language processing task solver?, arXiv preprint arXiv:2302.06476 (2023).

[24] A. Lopez-Lira, Y. Tang, Can chatgpt forecast stock price movements? return predictability and large language models, arXiv preprint arXiv:2304.07619 (2023).

[25] Z. Luo, Q. Xie, S. Ananiadou, Chatgpt as a factual inconsistency evaluator for abstractive text summarization, arXiv preprint arXiv:2303.15621 (2023).

[26] X. ZHAO, Using chatgpt as a recommender system: A case study of multiple product domains xianglin zhao li chen yucheng jin about 7 min chatgpt recommender systems (????).

[27] Y. Gao, T. Sheng, Y. Xiang, Y. Xiong, H. Wang, J. Zhang, Chat-rec: Towards interactive and explainable llms-augmented recommender system, arXiv preprint arXiv:2303.14524 (2023).

[28] D. K. Kirtania, Chatgpt as a tool for bibliometrics analysis: Interview with chatgpt, Available at SSRN 4391794 (2023).

[29] J. Kreibich, Using SQLite, " O'Reilly Media, Inc.", 2010.

[30] T. von Arx, K. G. Paterson, On the cryptographic fragility of the telegram ecosystem, Cryptology ePrint Archive (2022).

[31] K. Damak, O. Nasraoui, W. S. Sanders, Sequence-based explainable hybrid song recommendation, Frontiers in big Data 4 (2021) 693494.

[32] Y. Xian, T. Zhao, J. Li, J. Chan, A. Kan, J. Ma, X. L. Dong, C. Faloutsos, G. Karypis, S. Muthukrishnan, Y. Zhang, Ex3: Explainable attribute-aware item-set recommendations, in: Proceedings of the 15th ACM Conference on Recommender Systems, RecSys '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 484–494. URL: https://doi.org/10.1145/3460231.3474240. doi:10.1145/3460231.3474240.

[33] H. Abdi, L. J. Williams, Tukey's honestly significant difference (hsd) test, Encyclopedia of research design 3 (2010) 1–5.

[34] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, Y.-C. Zhang, Solving the apparent diversity-accuracy dilemma of recommender systems, Proceedings of the National Academy of Sciences 107 (2010) 4511–4515.

[35] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, X. He, Bias and debias in recommender system: A survey and future directions, ACM Transactions on Information Systems 41 (2023) 1–39.

[36] L. Zhang, S. Wang, B. Liu, Deep learning for sen-

timent analysis: A survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8 (2018) e1253.