Usability and Effectiveness of Bot-Assisted Group Decision Making

Esma Karahodža[1,*], Amra Delić[1,*] and Francesco Ricci[2]

[1]University of Sarajevo, Sarajevo, Bosnia and Herzegovina
[2]Free University of Bozen-Bolzano, Bolzano-Bozen, Italy

Abstract

Group recommender systems (GRSs) have been designed to support collective decision-making. Conversational GRSs are claimed to provide important advantages, related to their flexibility in replying to users' questions, and their adaptability to the dynamic evolutions of users' preferences.

This paper evaluates the usability and the effectiveness of CHARM, a domain agnostic conversational GRS leveraging a chatbot integrated into the Telegram messaging platform. CHARM is designed to support a natural group discussion where the group members suggest and discuss the alternative options. CHARM does not act as a classical RS, which is expert in a domain and therefore can push recommendations, selected for a group by enforcing a preference aggregation strategy.

We have conducted a user study with 164 participants across 44 groups, performing two decision-making tasks of different complexity, and in two experimental conditions: with and without CHARM's assistance. The experimental results show that CHARM achieves an above-average usability score (SUS = 72.02) and significantly reduces both perceived decision-making difficulty and the total number of exchanged messages, hence requiring a lower cognitive load. Correlation analysis further reveals that participants who used CHARM experienced significantly lower difficulty and higher choice satisfaction. The analysis of users' comments on the system usage highlights a generally positive outlook for CHARM's potential: many participants recognized its value for more complex decision-making scenarios, while also provided constructive suggestions for improving its reliability, interactivity, and intelligence.

We believe that the observed results are actually not specific to CHARM, but rather reflect the broader class of domain-independent, reactive bots it exemplifies.

Keywords

Group Recommender Systems, System Usability Scale, Decision-making Support, Chatbot

1. Introduction

Group Recommender Systems (GRSs) are designed to assist groups in reaching decisions, by delivering recommendations and supporting the different stages of the decision-making process [1, 2, 3]. The standard approach to generating recommendations resides on balancing the group members' individual preferences by using a preference aggregation strategy. Individual preferences are either explicitly stated by group members or derived from the user actions during the interactive group decision-making process [4].

In fact, a classical research line in this area has focused on preference aggregation, employing various strategies, either inspired by Social Choice Theory [5], their extensions [6, 7, 8, 9], or heuristic-based methods [10, 11, 12, 13], and more recently even neural network-based approaches [14, 15, 16, 17, 18, 19]. A smaller body of research has proposed tools that de-emphasize the preference aggregation task and try to support the full group decision-making process, including tasks such as eliciting preference, negotiation and discussion facilitation, and guiding the group towards a final decision [20, 3, 21, 22, 23, 24, 25, 4].

Regardless of the research focus, it is important to highlight that, differently from the widespread diffusion of single-user Recommender Systems, GRSs have failed to find traction in real-world platforms.

IntRS'25: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, September 22, 2025, Prague
*Corresponding author.
esma.karahodza@etf.unsa.ba (E. Karahodža); adelic@etf.unsa.ba (A. Delić); fricci@unibz.it (F. Ricci)
0009-0003-9999-5844 (E. Karahodža); 0000-0002-5584-6893 (A. Delić); 0000-0001-5931-5566 (F. Ricci)

To the best of our knowledge, there is no GRS in operation today that serves a large users' base, in a real-world setting, despite decades of active research in the field [26, 27]. We believe that this negative outcome could be motivated by the wrong assumption that groups will be eager to use domain-specific applications dedicated to choose either a movie, a restaurant, or an accommodation. This overlooks the fact that, differently from the individual recommendation scenario, when people make group decisions online, they need to interact not only with the RS, but also with the other group members. However, this intergroup communication functionality is offered already by general purpose communication platforms, such as, chat applications (e.g. WhatsApp). Hence, a GRS has a larger chance to enter in the common practice of a group if it does not require the group to change their common group communication and discussion tool, i.e., their chat app.

To this end, we introduce and evaluate a new kind of GRS, which is not built as a stand-alone application but is instead designed as a lightweight chatbot, embedded into an existing instant messaging chat app (Telegram), where usually decision-making processes naturally occur. Here, we assess the usability of this chatbot-based GRS and its impact on the group decision-making process. Our Telegram bot offers fundamental and necessary group decision-making functionalities, including the generation of group recommendations, but the recommended items were not proposed by the RS (as it usually happens), they were suggested by the group members as proposals that they independently identified. We analyse the GRS effect on group decision-making processes across two tasks of varying complexity, each performed with and without the assistance of the GRS chatbot, resulting in a between-subjects experimental design with four conditions.

Our experimental results indicate that the proposed GRS chatbot achieves a usability score well above average, highlighting its sound design. However, explicit users' comments clearly indicate some system limitations and worth to be implemented improvements. In general, the main positive outcome is that, by using the CHARM chatbot, groups perceived the decision-making process as significantly less difficult (compared with groups that did not use CHARM). This effect was especially pronounced among the participants who rated the bot easy to use, hence revealing a positive correlation between usability and perceived decision-making complexity. Furthermore, discussions supported by CHARM produced significantly fewer messages, suggesting that CHARM provides a meaningful level of cognitive offload for participants. Moreover, even among those who did not find the bot particularly useful in the experimental setting (two rather easy decision tasks), many noted its potential value in more complex decision-making scenarios, and called for a wider intelligent and proactive system behaviour in future versions. Overall, more than 93% of participants either directly mentioned or clearly implied potential use cases, relevant domains, and openness to using the chatbot in real-world decision-making contexts.

The remainder of the paper is organized as follows. Section 2 reviews existing tools supporting the decision-making process, as well as the literature on chat-based support. In section 3, we present the functionalities of CHARM and its commands. This tool is evaluated in the user study which is described in section 4. We present the results of this study and corresponding analysis in Section 5, and in Section 6 we outline possible future directions for the chatbot.

2. Related Work

While most GRSs focus on generating single-shot recommendations by aggregating individual preferences [5], a smaller body of research has focused on the proper design of a broader set of functionalities, supporting the full group decision-making process. Early examples include TDF [3], where agents negotiate on behalf of unavailable users, and CATS [22], a GRS where users can critique recommendations or peer suggestions, prompting updated recommendations. Similarly, Where2Eat [23] supports asynchronous interaction and proposal revision during group discussion. Later, systems added more dynamic interaction. HOOTLE [24], for instance, facilitates group discussions around desirable item features, allowing members to articulate and revise their preferences. Another system, developed by Nguyen et al. [4, 28], combines long-term preference modelling with short-term feedback collected during chat-based interaction. The system supports group conversations and allows users to make suggestions, react to others' proposals, and receive aggregated recommendations within the discussion interface.

Beyond GRS-specific tools, chatbot-based support for group decision-making has also been explored. In [29, 30], the authors showed that chatbot interventions can enhance information exchange and decision quality, especially when introduced early in discussions. Similarly, Tilda [31] improves chat-based collaboration by tagging message types and generating summaries to reduce conversational overload. Chatbots have also been used as discussion facilitators. Kim et al. [32, 33] developed bots that encourage participation, manage time, and structure group deliberation leading to greater opinion diversity and higher perceived decision quality. SolutionChat [34] adds visualization and moderation support to help participants track discussion progress and highlight key arguments.

Together, these works show a growing trend of integrating intelligent and personalised decision support into natural group conversations. Our work contributes to this line of research.

In [35] a focus study was conducted to examine how groups make decisions across a variety of issues and to identify areas where the process can be supported by a tool or an agent. Than, in [36] and [37], the CHARM framework is introduced as a tool designed to support group decision-making on items proposed by group members. A more detailed explanation on CHARM is given in the next section.

3. CHARM: A Chat-bot based GRS

In this section, we describe the functionalities of the group decision-making assistive tool evaluated in this study. CHARM (CHAt-bot group RecoMmender) is a TelegramBot [36] designed to support groups in collaborative decision-making tasks. In general, a TelegramBot is simply a Telegram account operated by software, that monitors conversations and interacts with users. CHARM is domain independent, which means that the bot is not capable to autonomously suggest relevant items. In practice CHARM can only reason on the data collected while interacting with the group members and exploiting information and knowledge exchanged in the group. It is up to the group members to propose options for the group to choose.

The version of CHARM, which we used in our study, provides the following six functionalities with the goal of structuring and directing the decision-making process:

1. **Start** - initiating a decision-making task, invoked by "*\start*" command;
2. **Suggestion** - enables users to make annotated item suggestions to their fellow group members, invoked by "*\suggest*";
3. **Feedback** - enables group members' to provide explicit feedback on the suggested items, in the form of *"Love it"*, *"Like"*, *"Dislike"*, and *"Hate it"* reactions;
4. **Summary** - generates the list of the suggested items together with the feedback that participants gave to them, it is invoked by "*\summarize*";
5. **Recommendation** - invoked by "*\recommend*", shows aggregated scores of the votes, where *"Dislike"*, *"Like"* and *"Love it"* are mapped to -1, 1 and 2, respectively, and generates recommendations based on the users' feedback;
6. **Decision** - allows group members to conclude the decision-making process by using the "*\decide*", which prompts the bot to present a list of all suggestions for the group to choose from.

The bot's functionalities are designed to ensure that the process of making item suggestions is easily traceable, to reduce participants' cognitive load from remembering all the suggestions and others' opinions, and to assist in the preselection of potentially interesting options for the group through a simple group recommendation approach. The approach uses the average without misery [1] aggregation strategy, which averages members' feedback, but eliminates options with at least one *"Hate it"* reaction.

Figure 1 shows a simplified real-world scenario, where Bob, Alice, and Carol are deciding on a movie to watch. After Bob starts the session by using the *"\start"*, Carol and Bob make item suggestions with *"\suggestion"*. The final decision is made after CHARM displays the summary and aggregated scores of the votes (*"\summarize"* and *"\recommend"*). The decision-making process is concluded when Bob uses the *"\decided"*, to communicate the final choice to CHARM.

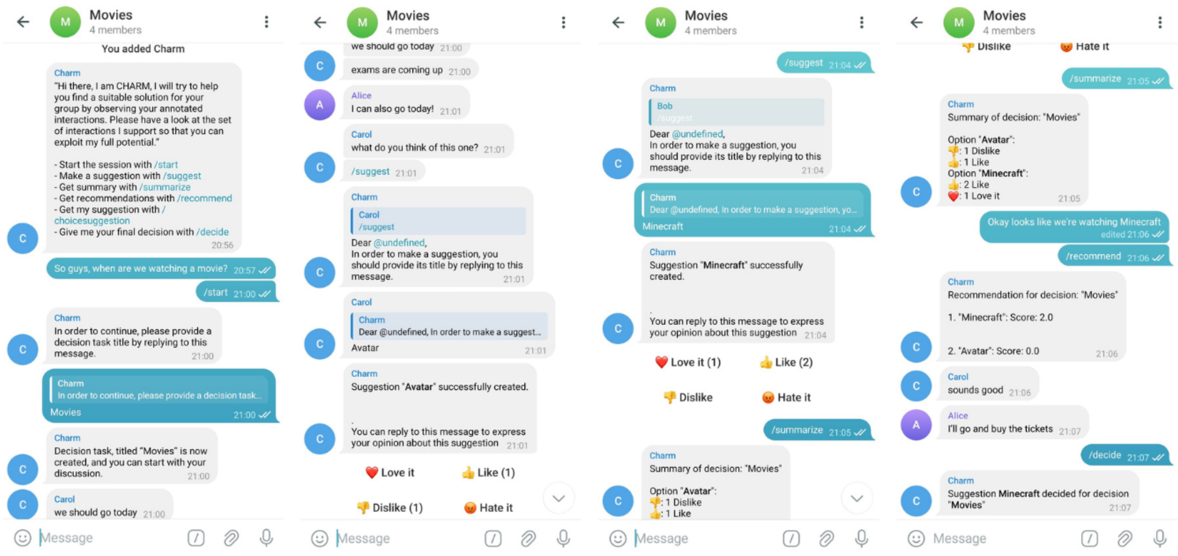

Figure 1: CHARM interaction

4. User Study Methodology

In this section, we describe the user study procedure and the analyses of the collected data, in order to understand the effects of CHARM on the group decision-making process.

Data Collection Procedure. The study was conducted at the University of Sarajevo, obtaining the ethical approval from the dean of the Faculty. The participants were first-year undergraduate students in the Computer Science and Informatics program, and the study was carried out as part of the Probability and Statistics course, as an optional activity. Prior to the study, participants were asked to create groups consisting of 2 to 6 members and to register their group with an appointed group leader. The data collection process consisted of two phases: 1) Group-chat discussion phase and 2) Post-questionnaire phase.

In the **group discussion phase**, participants were asked to invite their group members to a Telegram group chat and to discuss and make a joint decision (choice) that would satisfy all group members as much as possible. Each group was assigned to one of the four experimental conditions. The experimental conditions defined the decision-making task, as well as whether or not a group will be using CHARM assistance during the decision-making process.

The first task was to select a movie theatre (two options were available), a movie (from those played at that moment in the two theatres), and a time-slot when all group members are available. Hence, the groups were required to make three decisions.

For the second task, participants were first presented with the following scenario: *"Your group achieved the best result on Homework 1 in the course Probability and Statistics among all groups. As a recognition of your effort and outstanding results, the Faculty has decided to reward you with a fully funded trip to an international conference on applied statistics, which will be held in Vienna from July 14 to 17 at TU Wien, address: Karlsplatz 13, 1040 Vienna, Austria."*. Then, the groups were instructed as follows: *"Your task is to find and propose suitable accommodation for the entire group for the period from July 13 to 18. The total accommodation budget is €250 per person for the five nights."*.

Hence, in the second task the group was required to make a single-decision, in contrast with the first task, a multiple-decision task, which is considered more complex. To this end, the four experimental conditions are v1: multiple-decisions task without CHARM, v2: single-decision task without CHARM, v3: multiple-decision task with CHARM, and v4: single-decision task with CHARM.

After reaching a decision, the participants were instructed to export their group chat, including multimedia content, and submit their data to the research team. To ensure anonymity and ethical research standards, an anonymisation script was provided. The script automatically hashed all usernames and mentions in the chat transcripts, preserving the participants' privacy.

The final dataset comprised 44 group chats involving a total of 164 participants, the distribution of participants and groups over the four experimental conditions is shown in table 1. All participants were aged between 19 and 20 years. The gender distribution was 61% male and 39% female.

Experimental Condition	# Participants	# Groups
v1: Multiple-decision without CHARM	54	15
v2: Single-decision without CHARM	36	9
v3: Multiple-decision with CHARM	41	11
v4: Single-decision with CHARM	33	9
Total	**164**	**44**

Table 1
Distribution of participants and groups across experimental conditions.

In the **post-questionnaire** phase, all participants were asked to state their level of agreement, on a five-point Likert scale, from 1 - *"Strongly disagree"*, to 5 - *"Strongly agree"*, with two sets of statements on choice satisfaction and decision making process difficulty. The assessment of the group choice satisfaction included three statements: *"I like the choice that we, as a group, have made"*; *"The choice we made satisfies my preferences"*; *"The choice we made was fair"*. The difficulty of the group decision-making process was evaluated with six statements: *"I Eventually I was in doubt between some alternatives"*; *"The task of making this decision was overwhelming"*; *"The decision process was frustrating"*; *"I changed my mind several times before making the final decision"*; *"I think we have chosen the best option from the available options"*; *"To make the decision was easy"*.

Finally, the participants in the bot-assisted conditions were asked to complete the System Usability Scale (SUS) questionnaire[38], and to provide feedback on their experience with CHARM. The System Usability Scale (SUS) is a standardized questionnaire used to evaluate the perceived usability of a system or product. It consists of ten items, which are as follows: *"I think that I would like to use this system frequently"*; *"I found the system unnecessarily complex"*; *"I thought the system was easy to use"*; *"I think that I would need a support of a technical person to use this system"*; *"I found the various functions in the system were well integrated"*; *"I thought there was too much inconsistency in this system"*; *"I imagine that most people would learn to use the system quickly"*; *"I found the system very cumbersome to use"*; *"I felt very confident using the system"*; *"I needed to learn a lot of things before I could get going with this system"*.

The SUS score is computed by summing the participants' responses to these items on a 5-point Likert scale, adjusting odd- and even-numbered items separately, and then multiplying the total by 2.5 to obtain a score ranging from 0 to 100. Higher scores indicate better usability. Scores above 85 indicate excellent usability, 70–84 suggest good usability, 50–69 indicate marginal usability, and scores below 50 reflect poor usability.

The experience with CHARM was also evaluated with the two additional questions: 'How helpful was the Telegram bot in making a decision?" with options: *"It was not helpful at all, it just made the decision-making process more complicated."*, *"The feedback option for suggestions was helpful."*, *"The summary functionality was helpful."*, *"The recommend functionality was helpful." "The bot was in general helpful."*, and a free-text entry; "List improvements / corrections (in bullet points), that you think would be beneficial for bot.", a free-text question.

Conducted Analyses. We investigated the usefulness and impact of the CHARM-bot in group decision-making by comparing conditions with and without its support. We first assessed the usability of CHARM by using the SUS score. Then, with Brunner–Munzel test, we investigated differences in choice satisfaction, perceived difficulty, and users-to-users interaction patterns (number of messages exchanged), between the four experimental conditions. The Brunner–Munzel test is a non-parametric statistical test that examines whether there are differences between two independent populations, that is, whether one population tends to yield larger values than the other. Spearman correlations were used to explore the relationships between usability, satisfaction, and difficulty. Lastly, we analysed how participants used the bot, the perceived usefulness, and their suggestions for improvement.

5. Experimental Results and Analysis

System Usability Score. The system usability score of the CHARM-bot was 72.02 which exceeds the benchmark of 69 and is considered as *acceptable usability with minor improvements needed.*

Moreover, CHARM received a slightly higher score (73.70) when used for the single-decision task, compared to the multiple-decision task (71.03) - although this difference is not statistically significant.

Comparisons of decision-making process with and without CHARM assistance. To compare the differences between two independent populations, specifically, the group decision-making processes conducted with and without CHARM, while considering choice satisfaction, perceived decision-making difficulty, and the number of messages exchanged, as we previously noted, we used the non-parametric Brunner–Munzel test. The results are shown in Table 2, and in sake of conciseness, we present only statistically significant differences at a threshold of $p < 0.05$. The results indicate that perceived difficulty is significantly lower for groups using CHARM, particularly in the more complex task. However, CHARM does not lead to increased choice satisfaction; in fact, satisfaction was slightly higher in groups that did not use CHARM for the simpler task. Conversely, the total number of messages exchanged is significantly lower in groups that used CHARM. We justify these results by guessing that CHARM is reducing the cognitive load of group members by helping them to keep track of the various item suggestions and the others' opinions.

Table 2
Brunner–Munzel test results comparing perceptual and behavioural variables across experimental conditions

Comparison	Measure Description	B-M test *p*-value	Mean NoCHARM	Mean CHARM
Overall	Eventually I was in doubt between some alternatives.	**0.0093**	3.31	2.90
Multiple decisions	Eventually I was in doubt between some alternatives.	**0.0040**	3.35	2.76
Single decision	I like the choice that we, as a group, have made.	**0.0090**	4.97	4.61
Overall	Total #messages	**0.0007**	271.28	184.24
Multiple decisions	Total #messages	**0.0423**	304.02	211.36

Correlation Analysis. In order to better understand these results, we performed a Spearman correlation analysis, measuring the relationship between the system usability on the one side and perceived difficulty and choice satisfaction on the other side. Significant correlations are shown in Table 3. These results strongly indicate that the decision-making process is significantly less difficult and choice satisfaction significantly higher for participants who scored CHARM higher in usability.

CHARM Functionalities Usage Analysis. Next, to illustrate how participants used CHARM, in Figure 2 we show the frequency, and box-plots of CHARM and mistyped commands usage. The most frequently used functionality is *"\suggest"* (a group member suggests an item), and the frequency of the other functionalities is quite uniform. The figure shows a relatively high number of mistyped commands, but the majority of these were actually made by 3 groups.

Participants Feedback Analysis. When asked *how helpful the bot was in the decision-making process*, 30 subjects indicated that the bot was in general helpful, 31 pointed out the summary functionality as helpful, 29 indicated recommendations functionality, 20 the feedback option, while 11 out of 74 (14.86%) indicated that the bot was not helpful at all. Participants who did not find the CHARM helpful

Table 3
Significant correlations ($r \geq 0.30$) between SUS items and outcome variables (corrected for multiple comparisons)

SUS Item	Outcome variable	corr (p-value)
I would like to use this system frequently	The decision process was frustrating.	-0.42^{**} (0.0080)
The system was unnecessarily complex	The decision process was frustrating.	0.50^{**} (0.0010)
System functions were well integrated	The decision process was frustrating.	-0.47^{**} (0.0016)
System functions were well integrated	I like the choice that we, as a group, have made.	0.35^{*} (0.0357)
There was too much inconsistency	The decision process was frustrating.	0.41^{**} (0.0080)
People would learn to use it quickly	The decision process was frustrating.	-0.37^{*} (0.0270)
I felt very confident using the system	The decision process was frustrating.	-0.36^{*} (0.0270)

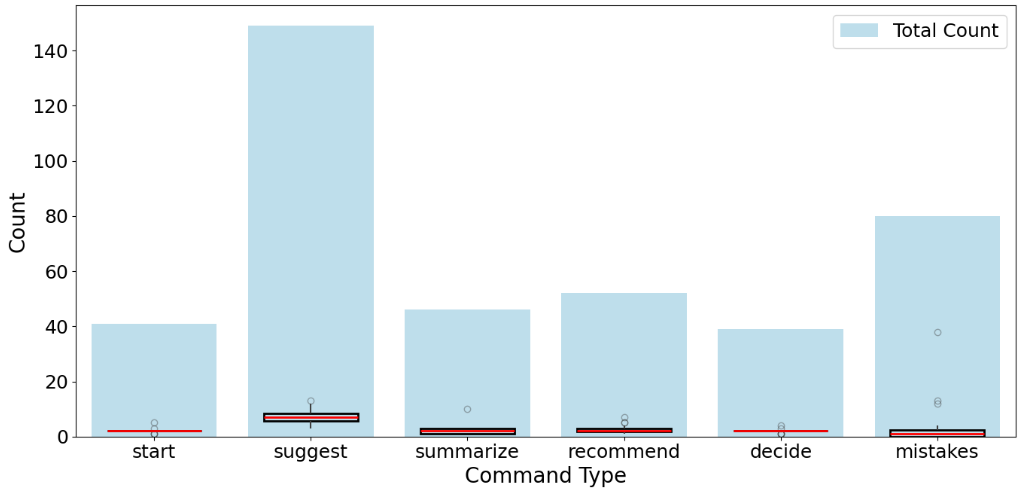

Figure 2: Total counts and distributions of commands and mistyped commands made in v3 and v4

emphasized that this was either because the task was too simple to require such support, or because of broken and unreliable core commands, like summary and recommendations. Notably, 14 out of 74 participants explicitly reported these command failures, which directly contributed to their frustration and reduced perceived usefulness.

Despite these issues, the overall sentiment leans toward optimism about the bot's potential: participants consistently acknowledge that in more complex decision-making scenarios, such a tool could be highly valuable. Suggestions for future improvements were abundant and constructive. Participants called for smarter, context-aware behaviour, for instance, generating suggestions and recommendations based on the ongoing conversation; more intuitive interaction, such as, buttons or menus instead of typed commands, improved naming conventions, and an easier way to modify or remove suggestions; further support for visualising preferences and feedback; and finally clearer onboarding and better feedback when something goes wrong. The absence of helpful error messages or step-by-step guidance made troubleshooting difficult and added to confusion. Many envisioned the bot evolving into an intelligent assistant that simplifies group coordination, increases fairness, and enables more efficient decision-making in real-world chat environments.

6. Conclusion

In this paper, we have presented the system usability analysis of the CHARM-bot and an initial comparison of group decision-making processes with and without its assistance. Our findings suggest that even in its current, simple form, a domain independent and purely reactive bot like CHARM can offer some support to group decision-making by reducing perceived difficulty and by lowering the total number of exchanged messages. We attribute this result to CHARM's design goal of easing cognitive load, especially by tracking suggestions and opinions. However, no differences in choice satisfaction were found, which we believe may be due to the bot ensuring complete transparency of the decision-making process: when all suggestions and feedback remain visible and accessible at any time, there is less room for manipulation or persuasion, and participants are also less likely to forget their initially preferred options. Although in both cases (with and without CHARM assistance), the average choice satisfaction score was well above 4 (out of 5).

The usage of bot functionalities was generally balanced, except for the higher use of the suggestion function. The summarization command was seen as most helpful, while 11 out of 74 participants did not find CHARM useful. Most of these users felt the task was too simple to require assistance and encountered issues with core features. Importantly, participants offered valuable suggestions for improving the bot, many of which align with current directions in GRS research [26, 27, 39], such as generating summaries of the discussion, automatically extracting suggestions from chat, and providing additional recommendations through web search.

We acknowledge the limitations of our study, including the relatively small sample size, as well as the convenient sample, i.e., all participants were students, and the preliminary nature of our analysis.

In our upcoming work, our plan is to first improve the robustness of the bot by eliminating failures in its core functionalities. With the new CHARM version, we then intend to conduct a more in-depth, content-focused comparison of decision-making processes with and without CHARM. Specifically, we aim to investigate the following hypotheses:

H1: *CHARM fosters a more task-focused and structured decision-making process* — as suggested by the observed reduction in exchanged messages, we assume that the bot helps groups stay oriented on their goal by providing simple mechanisms to track options and preferences.

H2: *CHARM increases transparency in group decisions* — by keeping all suggestions and feedback constantly visible, the bot makes it harder to disregard opinions or dominate the discussion, ensuring more transparent and balanced participation.

H3: *Decisions reached with CHARM are perceived as fairer* — we hypothesize that improved transparency and equal access to information lead participants to feel more represented and treated equitably.

H4: *CHARM promotes exploration of a broader set of alternatives* — since suggesting and recalling options is simplified, we expect participants to contribute more diverse proposals, ultimately enriching the decision space.

Furthermore, by adding simple functionalities to CHARM-bot, such as calling less active (or even passive) users to take part in providing their feedback, we aim to investigate how such interventions would affect various group decision-making aspects, including evenness of participation, fairness, satisfaction, and perceived complexity. Finally, in situations where conflicts are detected (e.g., opposing views or polarization), we envision extending the bot's functionality to find alternatives outside the user-made suggestions that can better satisfy multiple perspectives.

In the longer term, we also plan to analyse the influence of social relationships, participant roles, and emotional dynamics during group discussions, tailoring the bot's proactiveness accordingly. This includes enabling the bot to engage in one-on-one conversations with individual users when appropriate, with the goal of easing conflict resolution and guiding groups toward reaching more balanced and satisfactory decisions overall.

Declaration on Generative AI

During the preparation of this work, the author(s) used Chat-GPT-4 for sentence polishing. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] J. Masthoff, A. Delic, Group recommender systems: Beyond preference aggregation, in: Recommender Systems Handbook, Springer US, 2022, pp. 381–420.

[2] A. Felfernig, L. Boratto, M. Stettinger, M. Tkalčič (Eds.), Group recommender systems, Signals and communication technology, Springer Nature Switzerland, 2024.

[3] A. Jameson, More than the sum of its members: challenges for group recommender systems, in: Proceedings of the working conference on Advanced visual interfaces, 2004, pp. 48–54.

[4] T. N. Nguyen, F. Ricci, A. Delic, D. G. Bridge, Conflict resolution in group decision making: insights from a simulation study, User Model. User Adapt. Interact. 29 (2019) 895–941.

[5] J. Masthoff, Group recommender systems: aggregation, satisfaction and group attributes, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer, 2015, pp. 743–776.

[6] I. Ali, S.-W. Kim, Group recommendations: approaches and evaluation, in: Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication, ACM, 2015, p. 105.

[7] S. Berkovsky, J. Freyne, Group-based recipe recommendations: analysis of data aggregation strategies, in: Proceedings of the fourth ACM conference on Recommender systems, ACM, 2010, pp. 111–118.

[8] L. Quijano-Sanchez, J. A. Recio-Garcia, B. Diaz-Agudo, G. Jimenez-Diaz, Social factors in group recommender systems, ACM Transactions on Intelligent Systems and Technology (TIST) 4 (2013) 8.

[9] E. Quintarelli, E. Rabosio, L. Tanca, Recommending new items to ephemeral groups using contextual user influence, in: Proceedings of the 10th ACM Conference on Recommender Systems, ACM, 2016, pp. 285–292.

[10] M. Salamó, K. McCarthy, B. Smyth, Generating recommendations for consensus negotiation in group personalization services, Personal and Ubiquitous Computing 16 (2012) 597–610.

[11] L. A. M. C. Carvalho, H. T. Macedo, Users' satisfaction in recommendation systems for groups: an approach based on noncooperative games, in: Proceedings of the 22nd International Conference on World Wide Web, 2013, pp. 951–958.

[12] L. Baltrunas, T. Makcinskas, F. Ricci, Group recommendations with rank aggregation and collaborative filtering, in: Proceedings of the fourth ACM conference on Recommender systems, 2010, pp. 119–126.

[13] H.-N. Kim, M. Bloess, A. El Saddik, Folkommender: a group recommender system based on a graph-based ranking algorithm, Multimedia systems 19 (2013) 509–525.

[14] D. Cao, X. He, L. Miao, Y. An, C. Yang, R. Hong, Attentive group recommendation, in: The 41st International ACM SIGIR conference on research & development in information retrieval, 2018, pp. 645–654.

[15] J. Zhang, M. Gao, J. Yu, L. Guo, J. Li, H. Yin, Double-scale self-supervised hypergraph learning for group recommendation, in: Proceedings of the 30th ACM international conference on information & knowledge management, 2021, pp. 2557–2567.

[16] Z. Huang, Y. Liu, C. Zhan, C. Lin, W. Cai, Y. Chen, A novel group recommendation model with two-stage deep learning, IEEE Transactions on Systems, Man, and Cybernetics: Systems 52 (2021) 5853–5864.

[17] H. Yin, Q. Wang, K. Zheng, Z. Li, J. Yang, X. Zhou, Social influence-based group representation learning for group recommendation, in: 2019 IEEE 35th International Conference on Data Engineering (ICDE), IEEE, 2019, pp. 566–577.

[18] A. Sankar, Y. Wu, Y. Wu, W. Zhang, H. Yang, H. Sundaram, Groupim: A mutual information maximization framework for neural group recommendation, in: Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, 2020, pp. 1279–1288.

[19] L. Guo, H. Yin, T. Chen, X. Zhang, K. Zheng, Hierarchical hyperedge embedding-based representation learning for group recommendation, ACM Transactions on Information Systems (TOIS) 40

(2021) 1–27.
[20] L. Ardissono, A. Goy, G. Petrone, M. Segnan, P. Torasso, Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices, Applied artificial intelligence 17 (2003) 687–714.
[21] A. Venturini, F. Ricci, Applying trip@ dvice recommendation technology to www. visiteurope. com, Frontiers in Artificial Intelligence and Applications 141 (2006) 607.
[22] K. McCarthy, M. Salamó, L. Coyle, L. McGinty, B. Smyth, P. Nixon, The needs of the many: a case-based group recommender system, Advances in Case-Based Reasoning (2006) 196–210.
[23] F. Guzzi, F. Ricci, R. D. Burke, Interactive multi-party critiquing for group recommendation, in: RecSys, ACM, 2011, pp. 265–268.
[24] J. O. Á. Márquez, J. Ziegler, Preference elicitation and negotiation in a group recommender system, in: Human-Computer Interaction, Springer, 2015, pp. 20–37.
[25] D. Contreras, M. Salamó, L. Boratto, Integrating collaboration and leadership in conversational group recommender systems, ACM Transactions on Information Systems (TOIS) 39 (2021) 1–32.
[26] F. Ricci, A. Delić, Widening the role of group recommender systems with cajo, 2025. URL: https://arxiv.org/abs/2504.05934. arXiv:2504.05934.
[27] D. Jannach, A. Delić, F. Ricci, M. Zanker, Rethinking group recommender systems in the era of generative ai: From one-shot recommendations to agentic group decision support, arXiv preprint arXiv:2507.00535 (2025).
[28] T. N. Nguyen, F. Ricci, A chat-based group recommender system for tourism, Information Technology & Tourism 18 (2018) 5–28.
[29] N. Gurkan, B. Yan, Chatbot catalysts: Improving team decision-making through cognitive diversity and information elaboration, in: ICIS 2023 Proceedings, ???? URL: https://par.nsf.gov/biblio/10545067.
[30] B. Yan, N. Gurkan, It depends on the timing: The ripple effect of ai on team decision-making, in: Proceedings of the Hawaii International Conference on System Sciences, 2023.
[31] A. X. Zhang, J. Cranshaw, Making sense of group chat through collaborative tagging and summarization, Proc. ACM Hum.-Comput. Interact. 2 (2018). doi:10.1145/3274465.
[32] S. Kim, J. Eun, C. Oh, B. Suh, J. Lee, Bot in the bunch: Facilitating group chat discussion by improving efficiency and participation with a chatbot, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–13.
[33] S. Kim, J. Eun, J. Seering, J. Lee, Moderator chatbot for deliberative discussion: Effects of discussion structure and discussant facilitation, Proc. ACM Hum. Comput. Interact. 5 (2021) 87:1–87:26. doi:10.1145/3449161.
[34] S.-C. Lee, J. Song, E.-Y. Ko, S. Park, J. Kim, J. Kim, Solutionchat: Real-time moderator support for chat-based structured discussion, in: CHI '20, 2020, pp. 1–12. doi:10.1145/3313831.3376609.
[35] A. Delic, H. Emamgholizadeh, F. Ricci, J. Masthoff, Supporting group decision-making: Insights from a focus group study, in: UMAP, ACM, 2024, pp. 301–306.
[36] A. Delic, H. Emamgholizadeh, T. N. Nguyen, F. Ricci, Charm: a group decision making support chatbot, in: Companion Proceedings of the 29th International Conference on Intelligent User Interfaces, 2024, pp. 7–10.
[37] H. Emamgholizadeh, A. Delic, F. Ricci, Supporting a group member to make a group choice, in: Adjunct Proceedings of the 31st ACM conference on user modeling, adaptation and personalization, 2023, pp. 96–99.
[38] J. Brooke, et al., Sus-a quick and dirty usability scale, Usability evaluation in industry 189 (1996) 4–7.
[39] E. Karahodža, A. Delić, F. Ricci, Conceptual framework for group dynamics modeling from group chat interactions, in: Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization, 2025, pp. 23–27.