Design Considerations for Explanations Made by a Recommender Chatbot

Park, Sunjeong *[a]; Lim. Youn-kyung [a];

[a] Dept. of Industrial Design, KAIST, Daejeon, Korea, Republic of
* sunny0414@kaist.ac.kr

Self-explanatory features of intelligent agents have been emphasized in terms of helping users build mental models. While previous studies have revealed what traits an agent's explanations should have, investigations on how to design such explanations from a dialogue perspective (e.g., through a conversational UI) have not been addressed. Thus, the purpose of this work is to explore the roles of explanations in recommender chatbots and to suggest design considerations for such chatbots' explanations based on a better understanding of the user experience. For this study, we designed a recommender chatbot to act as a probe. It provides explanations for its recommendations, based on service customization in recommender systems that reflect how it learns and evolves. Each participant experienced the recommender chatbot for 5 days and underwent a semi-structured post-interview. We discovered three roles of explanations in user mental model development. First, each user rapidly built a mental model of the chatbot and became more tolerant of unsatisfactory recommendations. Second, users willingly gave information to the chatbot and acquired a sense of ownership towards it. Third, users reflected on their own habitual activities and used the chatbot reliably. Based on these findings, we suggest three design considerations for chatbot explanations. First, the explanations should be grounded by data from diverse channels. Second, the explanations should be logical and distinguish between personal and generic data. Third, the explanations should gradually become more complete through inferences made from comprehensive information.

Keywords: recommender system; explanations; user mental model

1 Introduction

Intelligent agents are defined as "computer programs whose behaviour only becomes fully specified after they learn from an end user's training data" (Kulesza, 2012). In this role, intelligent agents have become increasingly complex and diverse, and self-explanatory features have become important.

Recent studies have emphasized the benefits of explanations by intelligent agents (Herlocker, 2000; Johnson, 1993; Lim, 2009; Sinha, 2002). In addition, many studies exist that focus on how intelligent agents should offer explanations to end-users (Lim, 2009; Tintarev, 2007; Stumpf, 2009; Herlocker, 2000). These papers studied what characteristics such explanations should have and how they can be made more reasonable based on the underlying algorithm or model. Focusing more on users, some studies examined how an intelligent agent's explanations affect user mental models (Kulesza, 2012; Tullio, 2007). However, studies regarding "how to *design* explanations to help users develop mental models regarding an intelligent agent" are still limited.

Based on this gap in the existing research, our purpose in this study is to perform an explorational investigation of how chatbot explanations can be designed to help users build mental models of the agent, specifically regarding the chatbot explanations in recommendation services. We focused on a chatbot interaction-based recommender system that provides explanations in the context of a dialogue (i.e., a recommender chatbot). Because customization is an essential aspect of recommendation services, explanations on the customization process should be provided to the users during their interactions with the recommender chatbot. Therefore, we designed a recommender chatbot that provides explanations intended to provide transparency regarding its customization principles.

The user study was conducted with a chatbot designed by the Wizard of Oz method. User opinions and experiences regarding its explanations were collected during post-interviews. The study revealed three major findings concerning the roles that explanations have in user mental model development. Based on these findings, the considerations for designing explanations in conversational UI-based (e.g., chatbot based) recommendation services were suggested to emphasize the positive roles of such explanations. The outcomes from this study can be baselines of the explanation design that emphasize the positive roles of intelligent agents' self-explanations. It can be helpful in conversation design or user interface design of intelligent systems, not only in practice but also in future researches in the field of human-computer interaction (HCI).

2 Background

The importance of building user mental models for intelligent agents has been emphasized recently (Tullio, 2007; Kulesza, 2012; Kulesza, 2013). Mental models can be considered as "internal representations that people build based on their experiences in the real world. [...] The mental-models must be sound (i.e., accurate) enough to support effective interactions" (Kulesza, 2012). However, many intelligent agents fail to build accurate mental models, since the term intelligent raises the user expectations too much on the agents' ability. Also, the intelligent agents including recommender chatbots are based on machine learning technology which is perceived as black box algorithms. It is often ambiguous to users on what service the agent can provide, and why those services are provided. Therefore, to build a proper mental model which should be based on user understandings on what and why of intelligent agents, transparent explanations play a critical role (Tintarev, 2007; Sinha, 2002; Johnson, 1993; Herlocker, 2000). Explanations remove the "black box" aspect that typically surrounds an intelligent agent and allow the agent to justify its actions, increase user involvement, help users build mental models about the agent's processes, and raise the acceptability of the agent's actions (Herlocker, 2000).

Previous studies have investigated the crucial factors of explanations. Based on the importance of explaining inner logic (Sinha, 2012), some studies have suggested approaches to explaining why (Lim, 2009) and how (Kulesza, 2012) the agent makes recommendations. Other studies revealed some features that explanations should include: the user context (Tintarev, 2007) and the source of the explanations (Kulesza, 2012; Tintarev, 2007). These studies suggested traits that explanations should possess to improve users' understanding of the agent. Although previous studies investigated the effects of explanations on user satisfaction or the degree to which users correctly understood the agent, research on the users' experiences during the mental-model building process is still lacking. Along with the previous study, which emphasized that the way an explanation is communicated is the most important aspect of helping users understand the agent (Stumpf, 2010), we narrow down our perspective on providing explanations in the context of a dialogue in a conversational UI such a chatbot format, which is the most basic and familiar type of interaction that people have with intelligent agents, to lower the barriers for understanding the agent.

3 Method

3.1 Designing a Recommender Chatbot Transparent on the Principles of Service Customization through Explanations

The purpose of this study is to investigate user experiences with a chatbot that provides explanations about its recommendations. We designed a preliminary recommender chatbot to probe user experiences on this topic and explored the roles that explanations take in building users' mental models of the recommender chatbot. To help users develop mental models and to undergo an engaged experience, we designed a recommender chatbot whose explanations transparently revealed the principles of its service customization from two aspects, which are well-known requirements for service customization, i.e., learning and evolving (Lee, 2009; Kim, 2016; Kim, 2019).

The first principle of customization applied to the recommender chatbot is that it learns about the user based on user-provided data. To reveal this aspect in the explanations, our chatbot specified the user-provided data that was associated with each recommendation using constructions such as “I learned about your information from…”. After each recommendation, the chatbot collected user feedback by asking “How was my recommendation?” Then, during subsequent recommendations, the chatbot reflected the user's feedback by statements such as “by considering your feedback on…”.

Another principle of customization is that the chatbot evolves its recommendations based on accumulated data regarding a user. As user data accumulate, the nature of inferable information changes: fragments become patterns; subsequently, the depth of the patterns deepens. Our chatbot revealed this deepening knowledge in its explanations by making references—first, to the initial fragmented data, and over time, to the user's activity patterns.

Following the recent research trend on recommender systems in the domains of physical activity and personal health care (Fernandez-Luque, 2009; Wiesner, 2010; Kumar, 2014; He, 2014; Otsuki, 2017), we selected health care as our chatbot's domain; the chatbot provides recommendations on diet and exercise. We chose this domain because of its convenience for providing personalized recommendations and because it is familiar, easily understood, and of interest or concern to everyday users.

3.2 Participants

For this study, twelve people were recruited as participants. We attempted to recruit people who were interested in exercise and diet, placing them in the target group for our health recommender chatbot. Based on the purpose of qualitative studies (Leedy & Ormrod, 2005), the goal of this study was not to identify the tendencies of a specific type of users but to understand the various users' perceptions and experiences from their own perspectives. Therefore, we considered the diversity of prior knowledge and experience with chatbots when recruiting participants. The participants judged their own prior knowledge and experience to reflect their own perspectives by selecting among four options: 1) very familiar with chatbots and have prior knowledge and experience (majored in or working in a related field and used one steadily for more than a month); 2) have chatbot knowledge but little experience (know what one is but have used one for less than a month); 3) have chatbot knowledge but no experience (know what one is but have not used one); and 4) have neither chatbot knowledge nor experience. Detailed information on each participant is provided in Table 1. Among the 12 participants, 9 were male. The average age of the participants was 25.75 years, and the median was 25.00 years.

Table 1 Prior user knowledge on and experience with chatbots

Participant number	Prior knowledge on and experience with chatbots
P1	Very familiar in both knowledge and experience
P2	Have knowledge, but no experience
P3	Have knowledge, but no experience
P4	None
P5	Have knowledge but little experience
P6	Have knowledge but little experience
P7	Have knowledge but little experience
P8	Very familiar in both knowledge and experience
P9	Very familiar in both knowledge and experience
P10	Have knowledge, but no experience
P11	Have knowledge, but no experience
P12	Have knowledge, but no experience

3.3 Procedure

3.3.1 Experiencing the Recommender Chatbot

We used the Wizard of Oz method to develop the recommender chatbot based on the strategy described in Section 3.1. The researchers acted as a recommender chatbot, producing health recommendations and the explanations on each recommendation. The recommendations were based on each participant's schedule data and physical data (e.g., number of steps over time, travel paths and distances), and the explanations were designed based on the basic principles we described in Section 3.1. To track participants' personal daily activities and log their physical activities, the researcher provided each user with a smartphone health app (e.g., iPhone health app, Pacer) or an activity tracker (e.g., Mi band, Jawbone). We collected the participants' schedule data for the 5 days of the study period by accessing their calendar apps. The participants were aware of the types of data provided to the researchers and that the recommendations would be provided based solely on these data. The participants experienced this health recommender chatbot over 5 days.

Each participant received personalized daily health recommendations concerning their exercise and diet from the recommender chatbot. For the chatbot system, we utilized an instant messenger app (KakaoTalk) because that app is domestically widespread; thus, users could chat in a familiar environment. Furthermore, because KakaoTalk supports many business chatbots, most of the users already had some experience conversing with a chatbot via this app (even when they were not aware that it was a chatbot). We did not tell the participants that our chatbot was a roleplay performed by a human wizard: the chatbot provided recommendations at fixed times as if they were provided by a non-human system.

At the beginning of the experiment, the recommender chatbot sent an introductory message to introduce the service, as shown in Table 2.

Table 2 Introductory message sent by the chatbot at the start of conversation with the participants

Purpose of the message	Sample message content
Greet the user and set an exercise goal	"Hello John, the weather is bright today. I will introduce myself first. I am a chatbot who recommends exercise and diet based on your exercise goals. Would you please set your exercise goals by choosing among 1) reducing body fat, 2) improving balance, or 3) building muscle?"
Explain recommendation service in detail and get confirmation on data collection	"Your exercise goal is body fat reduction. I will recommend exercises and meals related to body fat reduction. I will provide daily recommendations regarding your exercise and diet. To give you a better quality of recommendations, I need to learn a little about you. May I access your schedule data and health tracking data?"

After the introduction, the recommendations were provided to each user via instant messenger twice a day: one recommendation for exercise and one recommendation for diet. After each recommendation, the chatbot asked, "How was my recommendation?" to collect user feedback; subsequently, it could reflect the feedback as another aspect of user information. The recommender chatbot mentioned the source of its data using words such as "learned," "based on," "considering" and "feedback" in the explanations to reflect the first principle of customization (learning). As time passed, the learned user information became more abundant, and the recommender chatbot considered the accumulated information to enhance the recommendations. The inferences from this accumulated information were reflected in the explanations over the study period of 5 days, reflecting the second principle of customization (evolving). Examples of the detailed explanations provided on each day are listed in Table 3.

Table 3 Base data reflected in the recommendations and examples of explanations

Days	Base data reflected on recommendation	Example of recommendation with explanation
Day 1	Based on exercise goal	(participant whose exercise goal is to reduce body fat) "How about oatmeal and boiled eggs to help you lose body fat today?"
Day 2	Schedule was additionally considered	"Good morning, John. Based on your schedule from 10 am to 3 pm today, you will be tired. I recommend you have chicken soup for lunch, a high-protein meal."
Day 3	Daily activity data additionally reflected	"It's raining today. I learned about your activities yesterday, and your activity level came up a little short compared to the recommended amount. I recommend that you have a low-fat tofu salad today."
Day 4	User feedback to prior recommendations additionally considered	(a participant who wanted the chatbot to recommend meals that can be eaten at restaurants near his school) "Good morning, John. Today's meal is recommended by considering your diet feedback. I recommend a turkey sandwich, which is a balanced meal that can be eaten near your school."
Day 5	Users' routine patterns analysed from daily activity data over the previous four days additionally considered	"Looking at the pattern of your recent activity, you seem to prefer high-intensity exercise at the appointed time. Based on the feedback you provided, I recommend the Tabata routine today at 8 PM, which will help you improve your strength and reduce your body fat. I will provide the

		workout video as you requested last time."

3.3.2 User Interviews

After 5 days of experience with the recommender chatbot, a post-interview was conducted with each participant. This interview focused particularly on probing users' thoughts and experiences with the explanations from various aspects. The interview questions are listed in Table 4. Each interview lasted a maximum of one hour and was voice-recorded.

Table 4 Intent of the interview questions and sample interview questions

Intent of the interview question	Sample interview question
To relax the participants and allow them to freely communicate their overall impressions	"How was your experience with the health recommender chatbot?"
To understand participants' expectations at the outset	"What expectations did you have right after the initial chatbot greeting?"
To recall experiences related to the participant's satisfaction with the explanations	"What was the most satisfactory explanation? Why?" and "What was the most unsatisfactory explanation? Why?"
To gain participants' detailed impressions of the explanations based on two principles of customization	"Did you feel that the chatbot was learning about you? If yes, when and why? If not, why not?" "Did you feel that the chatbot was evolving? If yes, when and why? If not, why not?"
To understand participants' opinions about the transparency of the explanations	"Did you think that the chatbot provided transparent explanations about its recommendations? If yes, can you give an example of when and why? If not, why not?"
To understand participants' overall levels of trust and satisfaction with the chatbot in relation to its transparency	"Regarding your impressions of transparency, how did this impact your overall trust of and satisfaction with the chatbot?"

After asking these questions, we showed each participant their full conversation with their own recommender chatbot and asked them to freely provide any additional impressions about the conversations.

The collected interview data were transcribed, and similar comments were grouped. To categorize the comments, we tagged each comment based on the purpose of this study. This open-coding process was performed while keeping in mind the detailed factors of explanations and their effects on the users' experiences. The coding was individually performed by the four researchers, and five researchers (including the four researchers who performed the open-coding) discussed the results. This process of open-coding and discussion was performed iteratively to extract agreed-upon findings. In this way, we were able to identify the three findings concerning the roles of explanations in recommender chatbots for user mental model development.

4 Findings: The Roles of Explanations in User Mental Model Development

4.1 Users rapidly built a mental model on the recommender chatbot and became more tolerant of unsatisfactory recommendations

Users were able to discover what the recommender chatbots knew and understood about them through the explanations. The chatbots provided explanations regarding the data on which the recommendations were based along with the content. Therefore, the users were able to understand how the chatbot processed data and consequently, to build mental models on it. This process helped users feel that the recommender chatbot understood them, which ultimately led to an increase in trust. "*I used the chatbot for only a few days, but it was nice to know that it [chatbot] knew those various things about me. I felt understood.*" [P9]

The increased user trust in the recommender chatbot made the users more likely to actively try the recommended diet or exercise. "*I was motivated to perform the recommended exercise because it [the chatbot] explained the reason why it was suggesting that to me. Whether the recommendation was good or not, I was willing to try.*" [P11]

In addition, when the recommender chatbot made an unwanted or improper recommendation, users thought it occurred because they had not given the chatbot enough information; therefore, providing an incorrect recommendation was a reasonable mistake. Users tried to convince themselves that a lack of information was the reason why the recommendation was wrong, which caused them to adopt a more tolerant attitude towards the recommender chatbot. *"I know that it needs to learn about my activities or preferences for a long time to make totally fitting recommendations. It can give inappropriate recommendations sometimes since it does not have enough information about me.*" [P10]

The explanations provided with the recommendations helped users gain trust in the recommender chatbot and be more tolerant of unsatisfactory recommendations, which can potentially reduce the abandonment rate.

4.2 Users willingly gave information to the recommender chatbot and acquired a sense of ownership towards it

Users felt intimate with the recommender chatbot when it mentioned social issues or referenced personal information concerning the user's daily life in the explanations. From the explanations, users thought that the recommender chatbot felt sympathy for them. *"I felt intimate because the chatbot said things such as 'Since you have been to Jeju [...].' I felt that it was thinking specifically about me.*" [P9]

As users built their mental models on the recommender chatbot and felt increasing intimacy, they tried to provide more detailed feedback to obtain better recommendations. "*From this point on, I explained in more detail [not only giving feedback on whether I liked it or not] when giving feedback. For example, 'I like this recommendation, but I do not want to try it right now.' I was trying to give more information*." [P12]

The recommender chatbot provided increasingly customized recommendations and used phrases that were increasingly specific to each user by collectively reflecting the user feedback and activity data. Users were able to experience the recommender chatbot's development in a more customized direction through the explanations and felt themselves becoming engaged with the chatbot. As a result, users gained a sense of ownership from thoughts such as 'the recommender chatbot is only for me'. "*At first, the chatbot felt like a teacher who teaches 30 students, but as it became more personalized, I felt a one-to-one relationship [of teacher-student].*" [P4]

The user-chatbot relationship and the feeling of ownership helped users to provide information steadily and willingly to the recommender chatbot. The ownership helped the participants trust the chatbot's recommendations. This positive feedback can be a cornerstone in building long-term relationships.

4.3 Users reflected on their own activity habits and reliably used the recommender chatbot

Users were deeply impressed by the explanations when the recommender chatbot analysed their cumulative exercise data and made reports from a comprehensive viewpoint. Existing activity trackers or personalized health care apps merely sort and display users' activity data. However, in our study, the recommender chatbot interpreted the data and described it conversationally when delivering recommendations to the user. In particular, the parts in which the chatbot analysed the activity over several days and discovered lifestyle patterns caused users to recognize the chatbot as a customized information source that reinterprets their personal objective data. *"I think it was good that it discovered my life pattern and let me know. I liked the comprehensive evaluations such as 'You enjoy exercise through in-life activities rather than high-intensity exercise'. [It explained] aspects that are difficult to understand when looking at the data by myself."* [P11]

From the provided analysis, users felt they were understood by the recommender chatbot and used it more reliably. Users self-checked their lifestyle and patterns through the customized reports provided by the chatbot. *"It was very good in terms of the chatbot understanding my life pattern. It gave me a chance to do a self-check."* [P7], Furthermore, users became self-motivated to change their life patterns, and they were more likely to actively accept the recommender chatbot's recommendations. *"I started to trust the chatbot's recommendations more, and my attitude changed—I became more likely to try the recommendations to change my life pattern in a better way.*" [P12]

These interview results imply that explanations that interpret the user's behaviour patterns increase the perceived reliability of the recommender chatbot and can self-motivate users. Fostering self-motivation has the potential to have a continuing impact on the user's life and improves the probability of long-term use of the recommender chatbot.

5 Design Considerations to Emphasize the Positive Roles of Explanations

We extracted three considerations for designing explanations that emphasize their positive roles in helping users develop mental models. These considerations were proposed based on the user experience instances when they expressed positive influences regarding the explanations. These design considerations can be applied as strategies for recommender chatbots in future studies and should be discussed deeply so they can be developed in forms more feasible for designers.

5.1 Explanations should be grounded by data from diverse channels

When giving a recommendation to a user, the chatbot should explain why this content is recommended and which data are the basis for the recommendation. From our study, we discovered that users trust the recommendations more when the content is grounded by the data from diverse channels.

For example, suppose a user who set an exercise goal of muscle-building has a below-average daily step count and mentioned in feedback that he does not cook. To this user, the recommendation could be given as follows, considering his dinnertime schedule: "Looking at your activity data for a few days, your step count is a little short compared to the average. I recommend that you eat dinner at a restaurant approximately a 15-minute walk from work, based on your feedback and activity data. Today, I recommend the XX sushi restaurant, which is famous for its high-protein brown rice sushi." This explanation includes data from four channels: 1) the step-counting channel, 2) the user's feedback channel, 3) the user's schedule channel, and 4) the user's exercise goal channel. The 15-minute-distant restaurant was recommended because the user's *number of steps* was below average. Reflecting the *feedback* that the user cannot cook, the chatbot recommended a restaurant rather than a specific meal to cook. The chatbot recommended a suitable dinnertime based on the *user's schedule*. The reason for recommending that particular restaurant (sushi) was because sushi is high-protein food, which is related to the user's *exercise goal*.

In this way, the recommender chatbot lets the user know that it has learned personalized information through various channels and that it considers these data when making recommendations. This not only helps users trust the recommendations but can also positively affect the user-chatbot relationship by revealing that the chatbot is actively accepting and learning from the provided information and feedback.

To be more specific about the source channel of the data, users felt more intimate with the recommender chatbot when it considered data from the user's context. This context included both public social issues and personal situations that had occurred in the user's daily life. To provide explanations considering the user context, it is important to collect information about the user's overall lifestyle. Even when that information is not directly related to the recommender chatbot’s service domain (e.g., exercise and diet), explanations that demonstrate an understanding of the user's lifestyle will help users feel more intimate towards the recommender chatbot. For example, prior to recommending diets, information should be collected not only on users' diet goals and food allergies but also their living environments (whether the user can cook) and typical travel ranges.

5.2 Explanations should be logical and distinguish between personal data and generic data

Users gain trust in the recommender chatbot and its recommendations when it provides explanations that have a logical flow. To provide a logical explanation, the content of the recommendation should be connected to the data on which the recommendation is based. Additionally, the source of the data should be clarified by distinguishing two data types: "personal" data, which is specific to the user, and "generic" data, which is obtained from big data. For example, a logical explanation can be used to highlight a user's personal data (less than 5,000 steps per day) compared with objective indicators from generic data (average step per day for an adult in their 20 s): “You have walked less than 5,000 steps per day for the last three days, which is only half of the recommended average step per day for an adult in their 20s”. This explanation is more convincing to users than an explanation such as “Looking at your activity patterns during the last three days, your step count is insufficient”.

Furthermore, the explanations should include the reasons why a particular content was recommended so that the users build trust on the recommender chatbot. For example, when sending a workout link, it is good to provide the reason why a video of a specific trainer was selected (e.g., this trainer’s videos are the most popular among home-oriented trainers). If the explanation is insufficient, users may suspect that the content may have been selected for advertising purposes, which can erode trust in the recommender chatbot.

Through the logical explanation mentioning the source of the data, users were able to perceive what the chatbot had learned about them. Therefore, users were more tolerant of unwanted or improper recommendations in the initial stages of use because they understood what information was insufficient for the chatbot. Users’ understanding of the data that the chatbot had learned about them made it easier for them to provide detailed feedback. Moreover, the higher quality of user feedback made it possible for the chatbot to provide increasingly customized recommendations with explanations that referred to prior feedback. During this process of users giving feedback and getting it back in future recommendations, users felt themselves becoming engaged with the chatbot and felt ownership of it.

5.3 Explanations should gradually become more complete through inference from comprehensive information

The longer a user interacts with the recommender chatbot, the more data it accumulates. Instead of showing the accumulated user data in a piecemeal fashion (e.g., using a list or chart), the chatbot should make a comprehensive judgement by deducing patterns and referring to them in explanations.

For example, our recommender chatbot provided explanations on the user's recent activity patterns at specific times of the day. "Looking at your *activity patterns* during the last three days, your step count *during work hours* is insufficient. Since you have a meeting until 5 PM, I recommend stretching during office hours for 10 minutes at 5 pm." This explanation reflected a user pattern determined by the chatbot from the user's activity data. Then, it recommended an exercise to augment the user's insufficient activity level.

As another example, our recommender chatbot recommended specific exercises (e.g., squats, lunges, Pilates, stretching at work, walking, etc.) through a video link, and the chatbot could discover what types of exercise each user preferred based on the accumulated user feedback. This knowledge was then expressed in the recommendations based on exercise classifications. "You seem to prefer exercise through *normal life activities* rather than high-intensity exercise at some specific time. Based on your feedback on the recommended exercises, I recommend moderate-intensity exercise. Try home-Pilates for 15 minutes after 9 pm."

As the amount of data accumulated, the participants in our study expected not only to obtain more satisfying recommendations but also to obtain analysed information on their comprehensive lifestyle patterns. In other words, providing data for an extended period allowed users to expect that the recommender chatbot would monitor their lives and activities. This suggests that it is important to meet users' expectations regarding both satisfactory recommendations and acknowledgements of the data users provide. Comprehensively inferring data and providing reports to users can positively affect the continuous use of the recommender chatbot by meeting users' expectations.

6 Conclusion and Future Research

By designing a recommender chatbot that provides explanations on its customization directions, we exploratively investigated user's experiences with the explanations and their development of mental models on the recommender chatbot. The user study revealed the findings on the roles of the explanations in the users' mental model development. We then suggested three considerations for designing explanations made by recommender chatbots based on the analysis of positive instances from the user studies. First, the explanations should be grounded on data from diverse channels rather than responding to data from only a single channel. Second, the explanations should be logical, which can be accomplished by the chatbot stating the types of data on which its recommendations are based. Third, not only the recommendations themselves but also the explanations should evolve in a direction customized to the user by gradually becoming more comprehensive through inferences.

Since all of the recruited participants were Asian, the activity patterns or lifestyles of users can be biased. Despite that, the key design considerations for explanations proposed in this paper will be the seeds for further research that generate specific design outcomes. This user-centred approach for designing explanations in recommender chatbots will lead to users developing more effective mental models, eventually resulting in users building trust and being more satisfied with the recommender chatbot. We hope that future studies demonstrate the influences of these considerations on people's service experiences regarding the intelligent agent, not just in the health care domain but also in other domains.

7 Acknowledgements

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding).

8 References

Fernandez-Luque, L., Karlsen, R., & Vognild, L. K. (2009, August). Challenges and opportunities of using recommender systems for personalized health education. In MIE (pp. 903-907).

He, Q., Agu, E., Strong, D., & Tulu, B. (2014, November). RecFit: a context-aware system for recommending physical activities. In Proceedings of the 1st Workshop on Mobile Medical Applications (pp. 34-39). ACM. doi:10.1145/2676431.2676439

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work - CSCW '00. doi:10.1145/358916.358995

Johnson, H., & Johnson, P. (1993). Explanation facilities and interactive systems. Proceedings of the 1st International Conference on Intelligent User Interfaces - IUI '93. doi:10.1145/169891.169951

KakaoTalk, http://www.kakao.com/services/8

Kim, D. J., Lee, Y., Rho, S., & Lim, Y. K. (2016, May). Design opportunities in three stages of relationship development between users and self-tracking devices. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 699-703). ACM. doi:10.1145/2858036.2858148

Kim, D. J., & Lim, Y. K. (2019, May). Co-Performing Agent: Design for Building User–Agent Partnership in Learning and Adaptive Services. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM.

Kulesza, T., Stumpf, S., Burnett, M., & Kwan, I. (2012, May). Tell me more?: the effects of mental model soundness on personalizing an intelligent agent. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1-10). ACM. doi:10.1145/2207676.2207678

Kulesza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I., & Wong, W. K. (2013, September). Too much, too little, or just right? Ways explanations impact end users' mental models. In Visual Languages and Human-Centric Computing (VL/HCC), 2013 IEEE Symposium on (pp. 3-10). IEEE.

Kumar, G., Jerbi, H., Gurrin, C., & O'Mahony, M. P. (2014, December). Towards activity recommendation from lifelogs. In Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services (pp. 87-96). ACM. doi:10.1145/2684200.2684298

Lee, M. K., & Forlizzi, J. (2009). Designing adaptive robotic services. Proc. of IASDR'09.

Leedy, P. D., & Ormrod, J. E. (2005). Practical research (pp. 139-140). Pearson Custom.

Lim, B. Y., Dey, A. K., & Avrahami, D. (2009, April). Why and why not explanations improve the intelligibility of context-aware intelligent systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2119-2128). ACM. doi:10.1145/1518701.1519023

Otsuki, R., Sugiyama, O., Matsumura, K., Tada, M., Noma, H., & Kuroda, T. (2017, October). When should the Health Promotion Agent Recommend the User to Walk?: Approach for Searching an Appropriate Walking Pattern for the User. In Proceedings of the 5th International Conference on Human Agent Interaction (pp. 469-472). ACM. doi:10.1145/3125739.3132589

Sinha, R., & Swearingen, K. (2002, April). The role of transparency in recommender systems. In CHI'02 extended abstracts on Human factors in computing systems (pp. 830-831). ACM. doi:10.1145/506443.506619

Stumpf, S., Rajaram, V., Li, L., Wong, W. K., Burnett, M., Dietterich, T., ... & Herlocker, J. (2009). Interacting meaningfully with machine learning systems: Three experiments. International Journal of Human-Computer Studies, 67(8), 639-662. doi:10.1016/j.ijhcs.2009.03.004

Stumpf, S., Wong, W. K., Burnett, M., & Kulesza, T. (2010). Making intelligent systems understandable and controllable by end users.

Tintarev, N., & Masthoff, J. (2007, October). Effective explanations of recommendations: user-centered design. In Proceedings of the 2007 ACM conference on Recommender systems (pp. 153-156). ACM. doi:10.1145/1297231.1297259

Tullio, J., Dey, A. K., Chalecki, J., & Fogarty, J. (2007, April). How it works: a field study of non-technical users interacting with an intelligent system. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 31-40). ACM. doi:10.1145/1240624.1240630

Wiesner, M., & Pfeifer, D. (2010, November). Adapting recommender systems to the requirements of personal health record systems. In Proceedings of the 1st ACM International Health Informatics Symposium (pp. 410-414). ACM. doi:10.1145/1882992.1883053