User's Perception Towards Chatbot as keeping up the conversation alive

Asha Karki

Abstract: Chatbots are one of the most powerful tools that can be like a human conversation through voice, chats, or commands, which can be embedded in different platforms like Facebook Messenger, WhatsApp, or even in the website. It's not handled by human which makes more difficult to make its benchmark. People always want to communicate with humans. They might get unsatisfied when if it's off-track of the topic. The two hypotheses were introduced, which only focused on Users and their responses about the Chatbot. The first hypothesis suggests the subject and talks about the area with a reward system. The second hypothesis is focusing on the behavior of users and preventing the unambiguous situation. It is tested thoroughly using the feedback form in the survey. The survey is divided into two and tested carefully using Paired-T-test.

1. Literature Review and Hypothesis Proposed

Chatbots are changing the patterns of interactions between humans and computers [1]. Many instant messenger tools, such as Skype, Facebook Messenger, and Telegram, provide platforms to develop and deploy chatbots, which either engage with users in general conversations or help them solve domain-specific tasks [2]. In the article *Chatbots: Are they useful?* They provide real-life examples of different chatbots in different contexts. One of the models is Sophia, a developed robot to assist in mathematics at Harvard by answering students' questions [3]. This turned out to be applicable in many other contexts. As the messaging tools become platforms, traditional websites and apps provide space for this new form of human-computer interaction (HCI) [1]. The more detailed about Chatbot Conversational flow. Conversational intelligence enables the Chatbot to participate in the conversation actively and demonstrates awareness of the topic discussed, the evolving informal context, and the dialogue flow. Therefore, conversational intelligence refers to a chatbot's ability to effectively converse beyond the technical capability of achieving an everyday goal. This section

discusses social characteristics related to conversational intelligence, namely: proactivity, conscientiousness, and Communicability.[7] Most of the studies rely on data from the log of the conversations, interviews, and questionnaires. Proactivity is the capability of a system to autonomously act on the user's behalf to reduce the human effort to complete a task. In human-chatbot conversations, a proactive behavior enables a chatbot to share the initiative with the user, contributing to the discussion more naturally [7]. Chatbots may manifest proactivity when they initiate exchanges, suggests new topics, provide additional information or formulate follow-up questions .to inspire users and keep the conversation alive: proactively offering and encouraging new issues have been shown helpful to both inspire users and keep the conversation alive. Participants in the study conducted self-reported that the Chatbot's suggestions helped them get started and gave them ideas about searching topics after iteratively evaluating prototypes for a chatbot in an educational scenario. Conscientiousness is a chatbot's capacity to demonstrate attentiveness to the conversation at hand [7]. It enables a chatbot to follow the conversational flow, show understanding about the context and interpret each utterance as a meaningful part of the conversation. Communicability Interactive software is communicative because users achieve their goals by exchanging messages with the system [7]. In this context, Communicability is defined as software's capacity to convey to users its underlying design intent and interactive principles.[7]

The growth of chatbot technology changes how companies engage with their customers [1], other fields like hospitals and campuses, and many more organizations. All platform it's focused on the user. However, chatbots still fail to meet users' expectations [1] as the Chatbots playing an essential role in the customers and their satisfaction. So, it needs to be accurate. Users are the most significant assets for the business. Customers also like when things are easy and quick, and the fact of the matter is, old-school customer loyalty programs no longer fit modern lifestyles [6]. As focusing on the customers and users, the hypothesis proposed focuses on conversational intelligence, and the next one focuses on the rewards system. 77% of customers want free products, and 75% of customers expect discounts and offers.[6]. Keeping that in mind, this approach seeks to implement in chatbots. Though it needs more circumstances and people to justify the truth as per (Cummings 2004) When designing an intelligent system that provides decision support, one

must consider the human as something outside the system, but also as an integrated system component that, in the end, will ultimately determine the success or the failure of the system itself[5].

1.1 Hypothesis introduced in the chatbots

After going through the research paper and analyzing the documents, the two hypotheses are proposed as incisive for chatbot conduct and characteristics that influence the users to perceive and act towards the chatbots. The two theories are listed below:

Hypothesis one: To inspire users by suggesting new topics, keeping up the conservation alive, and adding a reward scheme for satisfaction.

This hypothesis's primary purpose is to participate in the users with add-ons of reward basis actively and demonstrate understandings through this strategy. It introduces as expectation of conversation improvement and satisfaction

Hypothesis two: Focusing on user needs and motivation as a user-centric perspective with conversational intelligence.

This hypothesis's primary purpose is for more relevance and to deal with the lack of knowledge from hypothesis one chatbot. And to overcome the harassment of users.

2. Research Design

The Research is done implementing two hypotheses with the baseline chatbot. There are three versions of the Chatbot. And different questionnaire through the survey was done.

2.1 Testing two new chatbot features

The baseline chatbots were typed basis approach chatbots, but the other two chatbots are newly introduced feature approaches. The hypothesis explains above was used in baseline with two different chatbots that look similar but

different features. To know the convergence, the chatbots are split in manner H0, H1, H2, and H0, H2, H1(where H0=baseline Chatbot, H1=hypothesis one, H2=hypothesis two)

2.2 Testing Chatbots

The Chatbots were divided into two surveys and tested thoroughly with different age groups from 20-40 and with females, males' participation. As per gender-wise, there are only two users. Survey 1 was recorded for H0, H1, H2, and Survey 2 for H0, H2, and H1.

3. Building Chatbot

The Research carries focusing the user-centered. Thus, the interactive design approach was used while concentrating that in mind.

First, of baseline, chatbots simplify worked with the type general-purpose to send the link or not.

H0: Typed basis. The Chatbot was built as an initial phase to look to send the link.

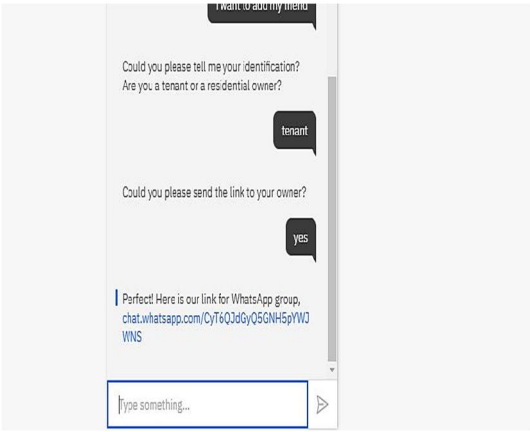

Figure1: Primary prototype

H1: To inspire users by suggesting new topics and keeping up the conservation alive, and adding a reward scheme for satisfaction. Benefits: To keep the chatbots recover from the failure and to inspire user's engagement with the chatbots

Approaches: Try to select the topic randomly and Giving a reward as a discount promo code

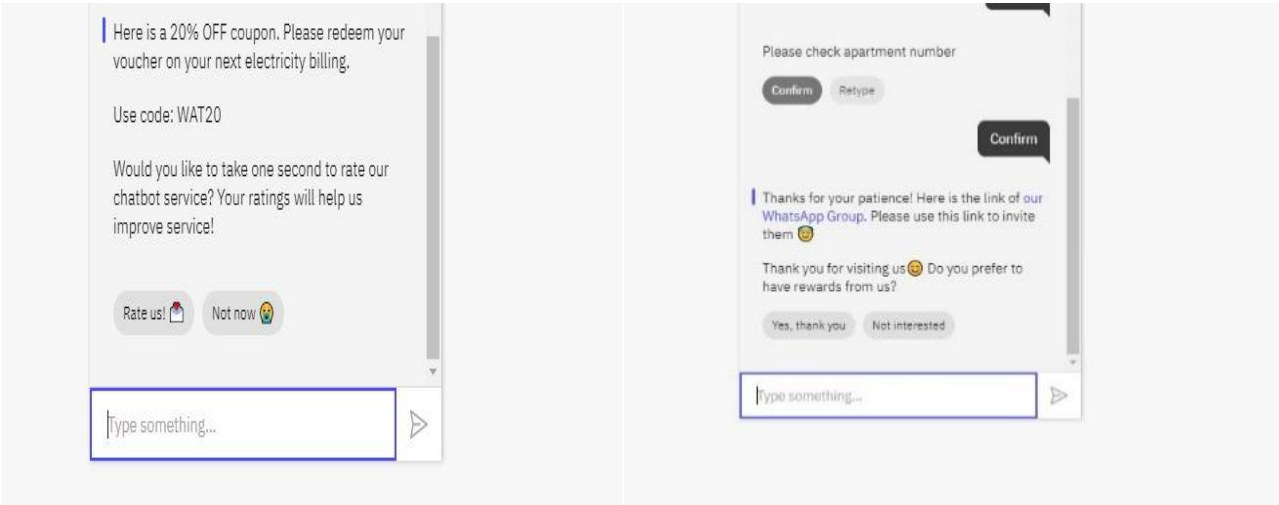

Figure 2: Reward scheme implementation in Chatbot

H2: Focusing on user needs and motivation as a user-centric perspective with conversational intelligence. Benefits: To keep the conversation on track, to manage the user's expectations, to identify the unfriendly users, and predict user satisfaction through ranking.

Approaches: To clarify the purpose of the Chatbot, functionality for the next step, and to ignore the abusive exclamation and constructive emotional reaction, and to predict user's satisfaction.

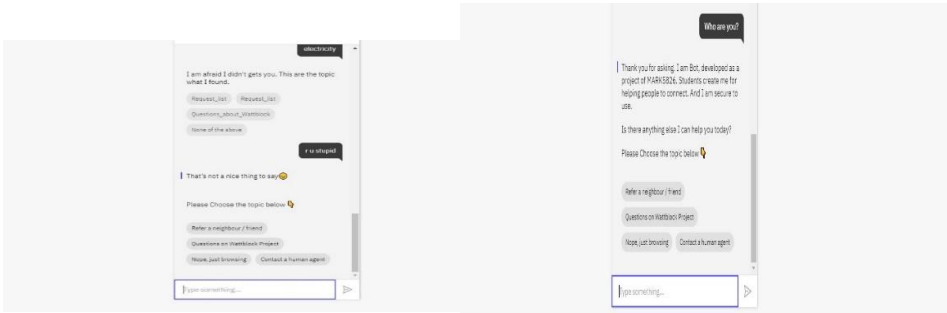

Figure 3: Controlling harassment by users and try to solve queries of users

4. Experiment Design

4.1 Online survey form

The two-survey form from the google form been made and shared with the 8-8 users, and their response s were recorded and measured. The interface for the user was given through the web link from IBM.

4.2 Outcome Measures

Chatbots' outcome measurement is done through the performance rate, satisfaction rate, smartness level, and experience with chatbots. The data obtained by the survey is measured using a paired t-test. The paired sample t-test, sometimes called the dependent sample t-test, is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. In a paired sample t-test, each subject or entity is measured twice, resulting in pairs of statements [9].

5.Analysis

5.1 Demographics

The participation in the survey was from different ages range 20-40 and with the gender male-female. Most of them are the student, with the intuition that they are quite aware of the Chatbot.

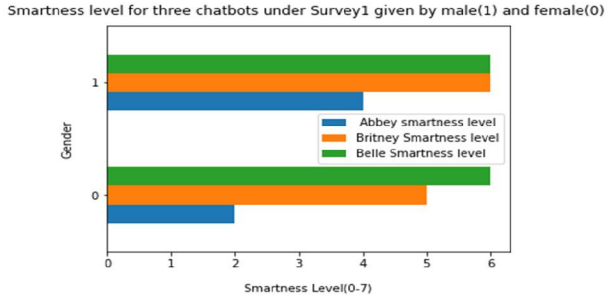

Figure 4: Smartness level given by the participation as per gender (female, male) for Survey1

Figure 4 above shows the smartness level given by males and females in survey one, where Abbey is baseline chatbots, Britney is for hypothesis 1, and Belle is for hypothesis 2. As from the figure, we can see the smartness level for baseline is around 3-4, but the smartness level of the Chatbot with hypotheses 1 and 2
comparatively high around 5-7. The female seems they like hypothesis 2 chatbot than hypothesis 1 chatbot but for male its same range.

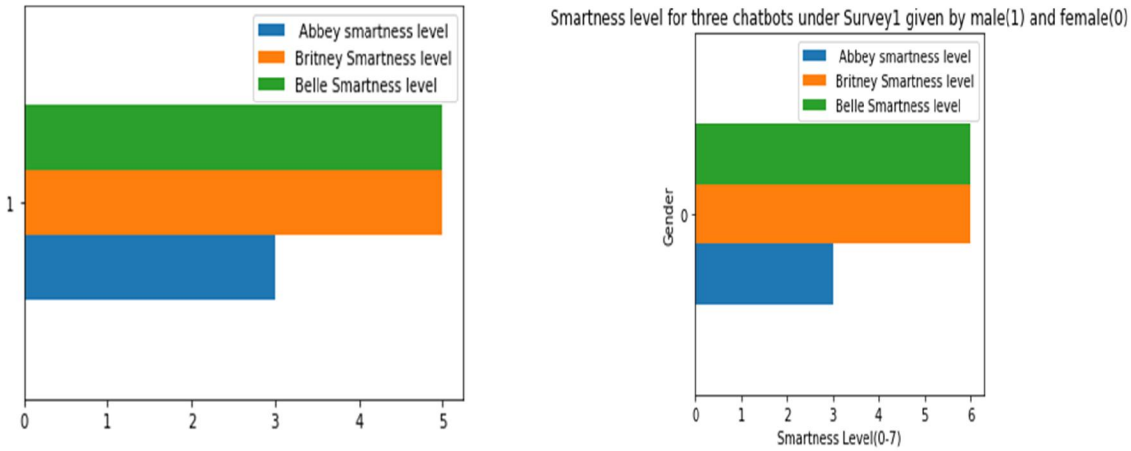

Figure 5: Smartness level is given by the participation as per gender (female, male) for Survey 2

Figure 5 above shows the smartness level given by males and females in survey two, where Abbey is baseline chatbots, Britney is for hypothesis 1, and Belle is for hypothesis 2. From the figure, we can see the smartness level for baseline is around 3-4, but the Chatbot's smartness level with hypotheses 1 and 2 comparatively high, approximately 5-7. As for the male, It seems they had given less level rate, i.e., around five, but females had given up to 6.

5.2 Quantitative questions

For all three chatbots, the question about the good and bad experience was asked. The three chatbots, the better experience was with the two hypothesis bots where people comment the familiar, ending was smoother. After that, the data was obtained it been tested using paired t-test. The paired sample *t*-test has two competing hypotheses, the null

hypothesis, and the alternative hypothesis. The null hypothesis assumes that the actual mean difference between the paired samples is zero. Under this model, all apparent discrepancies are explained by random variation. Conversely, the alternative hypothesis assumes that the paired samples' actual mean difference is not equal to zero. The alternative hypothesis can take one of several forms depending on the expected outcome.[8]

The result for hypothesis 1 with the baseline chatbot in Survey 1, the p-value is around 0.0011, $p < 0.05$, and in survey two, it's 0.0, which is $p < 0.05$. And for hypothesis 2, Chatbot with the baseline in Survey 1, the p-value is around 0.0052, and for survey two, it's 0.0, which is also less than 0.05. Thus, we can reject the null hypothesis. This means there are significant changes in the baseline chatbot and hypothesis chatbot.

As the mean difference for hypothesis 1 chatbot with baseline is 2.125 in Survey1 and 3.5 in survey2. For the mean difference for hypothesis 2, the mean difference is 2 in Survey1 and 3.5 in Survey2. Thus, there can be positive differences; hence Chatbot with a hypothesis is satisfied by users than baseline.

5.3 Qualitative questions

For the qualitative questions, in the survey, the chatbot smartness level and the impression about the performance were asked. The comments were not pleasant for the baseline, not intelligent, the ending, but for the other two, the words were satisfying, the absolute one, engaging. The appendix table shows that even the maximum smartness level for the H1 and H2 Chatbot is seven, but for the H0 baseline chatbot, it's around 3-4.

6.Conclusion

The survey recommends that H1 and H2 chatbots are smart and recommend one. Though the peripheral measurement was small, as from the bar diagram in Figures 4 and 5, we can see the participants didn't find a difference in the interface between h1 and h2. So, more surveys need to be done for better results and better understanding, and more

modifications need to do to act like humans. This Chatbot is the target for the business customer. It can be used by other organizations too.

7. Future Research

The chatbots need more modification and more training to be in a benchmark. So, there are lots of features that need to add to it. As in the future, I would like to add more speech approach chatbot, which can understand human beings' emotions and Artificial Intelligence for better relevance and stopping crashes. Natural language processing will also be a plus point for bots, helping users communicate in a human-like approach.

References:

1) https://www.researchgate.net/publication/317920872_Chatbots_and_the_new_world_of_HCI
2) DALE, R. (2016). The return of the chatbots. *Natural Language Engineering, 22*(5), 811-817. doi:10.1017/S1351324916000243
3) Shawar, Bayan & Atwell, Eric. (2007). Chatbots: Are they Really Useful? LDV Forum. 22. 29-49.
4) https://www.researchgate.net/publication/323255891_Single_or_Multiple_Conversational_Agents_An_Interactional_Coherence_Comparison
5) Cummings, M., 2004. Automation bias in intelligent time critical decision support systems, in:AIAA 1st Intelligent Systems Technical Conference.
6) https://medium.com/marketing-and-entrepreneurship/how-to-run-a-customer-loyalty-program-with-a-chatbot-53bcfc619df9
7) https://arxiv.org/abs/1904.02743
8) https://www.statisticssolutions.com/manova-analysis-paired-sample-t-test/#:~:text=The%20paired%20sample%20t-test,resulting%20in%20pairs%20of%20o

Appendix

Out[11]:

	Abbey conversation satisfaction rate	Abbey smartness level	Belle conversation satisfaction rate	Belle Smartness level	Britney conversation satisfaction rate	Britney Smartness level	difference
count	8.000000	8.000000	8.000000	8.000000	8.00000	8.000000	8.000000
mean	2.375000	2.250000	5.875000	5.625000	5.87500	5.625000	3.500000
std	0.916125	0.886405	0.834523	0.744024	0.64087	0.744024	1.069045
min	1.000000	1.000000	5.000000	5.000000	5.00000	5.000000	2.000000
25%	1.750000	1.750000	5.000000	5.000000	5.75000	5.000000	2.750000
50%	3.000000	2.500000	6.000000	5.500000	6.00000	5.500000	4.000000
75%	3.000000	3.000000	6.250000	6.000000	6.00000	6.000000	4.000000
max	3.000000	3.000000	7.000000	7.000000	7.00000	7.000000	5.000000

Table1: Description of response for Survey 1

[68]:

	Please put your gender	Abbey conversation satisfaction rate	Abbey smartness level	Britney conversation satisfaction rate	Britney Smartness level	Belleconversation satisfaction rate	Belle Smartness level	difference
count	8.000000	8.000000	8.000000	8.000000	8.00000	8.000000	8.000000	8.000000
mean	0.500000	4.250000	3.125000	6.375000	6.00000	6.250000	6.375000	2.000000
std	0.534522	1.035098	0.834523	0.517549	0.92582	0.707107	0.517549	1.414214
min	0.000000	3.000000	2.000000	6.000000	5.00000	5.000000	6.000000	-1.000000
25%	0.000000	3.750000	2.750000	6.000000	5.00000	6.000000	6.000000	1.750000
50%	0.500000	4.000000	3.000000	6.000000	6.00000	6.000000	6.000000	2.500000
75%	1.000000	5.000000	4.000000	7.000000	7.00000	7.000000	7.000000	3.000000
max	1.000000	6.000000	4.000000	7.000000	7.00000	7.000000	7.000000	3.000000

Table 2: Description of rate for Survey 2.