

A Comparative Study of Chatbots and Humans

Amit Mittal¹, Ayushi Agrawal², Ayushi Chouksey³, Rachna Shriwas⁴, Saloni Agrawal⁵

Computer Engineering Department, Institute of Engineering and Technology, Indore, India^{1,2,3,4,5}

Abstract: A chatbot is a software that interacts with humans using natural language processing and pattern matching techniques to understand questions and give relevant answers. In this paper, we evaluate the performance of chatbots ALICE, Jabberwacky and Rose on various criteria. We compare their performance with that of humans. Our general conclusion is that chatbots perform equally well as humans but they cannot replace humans completely.

Keywords: Chatbot, natural language processing, human computer interaction, pattern matching.

I. INTRODUCTION

Chatbots have evolved from semi intelligent chatbot like ELIZA [6] to modern intelligent personal assistant like Siri. Lately chatbots have made commendable use of artificial intelligence and machine learning techniques and found their application in daily life such as automatic telephone answering system, help desk tools, personal assistant. Chatbots are also used commercially in business and e-commerce, interactive games, information retrieval. Chatbots have gained popularity in real world applications as they can closely imitate human representative during the conversation. Compared to human, chatbots are available 24x7 and also reduce cost.

In this paper we evaluated performance of a classic chatbot ALICE, an entertaining chatbot Jabberwacky and a modern chatbot Rose. ALICE (Artificial Linguistic Internet Computer Entity) written by Richard Wallace in 1995 is a chatbot that engages in conversation by applying heuristic pattern matching rules to the input [1]. It uses AIML for specifying heuristic conversation rules. Jabberwacky is a chatbot created by Rollo Carpenter in 1997 [6]. Its purpose is to stimulate human chat in an entertaining way. Rose was created by Bruce Wilcox in October 2011 [8]. It simulates a human girl.

We also evaluated humans on the same criteria. We used general behaviour of humans for evaluation. We compared overall performance of all these chatbots with that of humans.

II. PERFORMANCE EVALUATION

We evaluated chatbots and humans on the basis of knowledge base, conversational properties and ability to manage unexpected situations. These criteria were further divided into several parts. Each chatbot was rated on the basis of sub-criteria. Aggregate of sum of points earned by each chatbot was then compared to that of humans.

A. Knowledge Base

Knowledge base is the brain of chatbot. It contains all the information which can be domain specific or general information about the world. We evaluated chatbots on the basis of general information.

Each chatbot was asked some basic questions. Chatbot obtained 1 point for each correct answer, 0.5 if it understood the question but failed to answer correctly and

0 for each incorrect response. Humans were also rated on the same scale. Questions asked were –

Q1- What is your name, Q2- How old are you, Q3- Who is your father, Q4- What do you do and Q5- What time is it. The result obtained is presented in Table I.

TABLE I: EVALUATION OF KNOWLEDGE BASE

	ALICE	Jabberwacky	Rose	Humans
Q1	1	1	1	1
Q2	0.5	0.5	1	1
Q3	1	0.5	1	1
Q4	1	0.5	1	1
Q5	1	0.5	1	1
Sum	4.5	3	5	5
Average	4.167			5

B. Conversational Properties

To depict human like conversational behaviour natural language processing isn't sufficient, chatbot should also possess some conversational properties such as ability to understand complex user input, taking turn, recalling the name of user, multi language support, voice input and output and access to the conversation history.

- a) Ability to understand complex user input**
Sometimes user may use complex sentences containing rhetoric words or phrases, in such cases chatbot should be able to correctly interpret the meaning of the sentence and reply with the appropriate answer.
- b) Taking turn**
For holding longer conversation chatbot should take turn by asking questions or starting a new topic.
- c) Recalling the name of the user**
For effective communication chatbot should remember name of the user.
- d) Multiple language support**
It is the ability to converse in multiple languages. ALICE knows languages other than English, Jabberwacky and Rose only know English. Some humans know multiple languages while some don't. So humans obtained 0.5 in our evaluation.
- e) Voice input and output**
It is the feature of speech recognition. Rose supports voice input and output.

f) Access to conversation history

It is the ability to remember or save conversation. All chatbots have the option to save conversation logs. 0.5 was given to the humans because they remember only recent conversations.

In our evaluation, chatbots and humans obtained one point if they possessed the property and zero otherwise. The result obtained is presented in Table II.

TABLE II: EVALUATION OF CONVERSATIONAL PROPERTIES

	ALICE	Jabber-wacky	Rose	Humans
Ability to understand complex user input	1	1	1	1
Taking Turn	0	1	1	1
Recalling the name of the user	1	1	1	1
Multi Language support	1	0	0	0.5
Voice input and output	0	0	1	1
Access to conversation history	1	1	1	0.5
Sum	4	4	5	5
Average	4.33			5

C. Handling exceptional situations

Chatbots should cleverly respond to exceptional situations such as unfamiliar questions, typing errors, insults and humiliation and dialogue errors. If the user’s statement is difficult to understand or there is lack of information on a specific question then chatbot should be able to handle it intelligently. It should not give incorrect response. For example, if ALICE doesn't know the answer it tells to search about it and thus doesn't mislead the user. While Rose tries to act cleverly by saying that it knows very little about it or by changing the topic. Jabberwacky simply says it doesn't know the answer. Humans try answering to the best of their knowledge but admit if they don't know the answer.

User may spell some words incorrect; chatbot should identify these typos and interpret the sentence correctly. Same thing goes with grammatical errors; they should be identified and corrected for chatbot to answer appropriately. This ability is called dialogue error correction. In both the cases, both chatbots and humans answer depending on their knowledge. 0.5 was given because sometimes they detect the error and interpret correctly while sometimes they don't.

Sometimes user may abuse chatbot or use offensive language. Chatbot should be able to handle these situations in the best possible manner. It should not get offended and must reply aptly. Jabberwacky and Rose simply move on

with the conversation. ALICE gets offended if a user insults it. Humans also get offended.

The result obtained from evaluation is presented in Table III.

TABLE III: HANDLING EXCEPTIONAL SITUATIONS

	ALICE	Jabber-wacky	Rose	Humans
Ability to answer unfamiliar questions	1	1	1	1
Surmounting typos	0.5	0.5	0.5	0.5
Surmounting insults and humiliation	0	1	1	0
Dialogue error correction	0.5	0.5	0.5	0.5
Sum	2	3	3	2
Average	2.67			2

III. DISCUSSION

As the result of performance evaluation, we observed that basic knowledge base of humans is better than chatbot as they understand and answer general questions more accurately. In conversational properties, both chatbots and humans understand complex input but humans have better ability to take turn. Few chatbots support multiple languages and voice input and output. Both chatbots and humans remember the name of the user.

Chatbots have better access to conversation history as they can store large amount of chat logs compared to that of humans. Thus, observing overall behaviour of chatbots and humans, humans show better conversational properties. While handling exceptional situations, chatbots perform better than humans as they handle insults and humiliations very well. While in other situations, both perform equally well. Overview of performance evaluation is shown in Table IV.

TABLE IV: OVERVIEW OF PERFORMANCE EVALUATION

	Chatbots	Humans
Knowledge Base (Basic)	4.16	5
Conversational Properties	4.33	5
Handling Exceptional Situations	2.67	2

IV. COMPARISON BETWEEN CHATBOTS AND HUMANS

In previous section we considered the criteria which are required for implementation of chatbots. In this section we consider some general criteria that can increase the use of

chatbots. We compare chatbots and humans on these criteria on the basis of general observations.

A. Availability

Chatbots have an advantage over humans that they are available 24x7, which increases their application. Humans cannot be available all the time. Moreover, chatbots can be deployed in websites so a wide range of users can get access to them.

B. Cost

Chatbots can be costly to buy or develop but they offer service at a far lower cost than a human representative, as humans need to be paid on regular basis. Chatbots' maintenance cost is very less as one only needs to feed information in their knowledge base to keep them updated. Further many chatbots have ability to learn from their chats. This further reduces cost for their maintenance. On the other hand, humans need constant training to keep them updated with the current information.

C. Ability to hold longer conversation

Chatbots are mainly designed for question-answer exchange; they do not remember previous chats and thus cannot hold longer conversation. They tend to forget previous messages and generally reply on the basis of current message. Humans are better than chatbots in this aspect as they remember the conversation just held and reply accordingly. Thus, they are able to hold longer and interactive conversation.

D. Ability to learn from past experiences

Human have the ability to learn from their conversations. They continuously learn new things from what they encounter and apply it wherever possible. They also learn from their experiences and mistakes. On the other hand, chatbots are not good enough at learning new things from their conversations. Their knowledge base has to be updated manually.

E. Frequency of mistakes

Replies of chatbot are completely based on its knowledge base so they are less likely to commit mistakes. While frequency of mistakes is more in human as human's memory is not as reliable as knowledge base of chatbots.

F. Ambiguous Replies

Ambiguity in a conversation may occur when a word has more than one meaning. Human can easily resolve problem of ambiguity as they know the context of conversation. While chatbots use natural language processing in which they map every word to its meaning, so a word with several interpretations may get mapped to a meaning which is irrelevant to context of conversation.

G. Scalability

Chatbots can support large number of users at a time while human can support a few users at a time. Thus chatbots are easily scalable to larger workloads.

V. CONCLUSION

We evaluated a few chatbots based on their performance and we found that their knowledge base and

conversational properties are comparable to humans. To be more effective in future, chatbots need many improvements such as ability to hold longer conversation and ability to learn from past experiences. They need to develop better memory and make technological improvements in parsing techniques so as to develop a better sense of the context of conversation. Chatbots have some important advantages such as availability, scalability, reliability and low cost which increase their area of applications. In general, we can say that chatbots perform equally well as humans but humans still have an edge over chatbots.

REFERENCES

- [1] A.L.I.C.E Website [Online]. Available: <http://alice.pandorabots.com/>
- [2] Bayan Abu Shawar, Eric Atwell " Chatbots: Are They Really Useful?", LDV- Forum 2007 -Band 22(1)-31-50.
- [3] Chayan Chakrabarti , " Artificial Conversations for Chatter Bots Using Knowledge Representation, Learning, and Pragmatics, "Doctor of Philosophy, Computer Science Dissertation, The University of New Mexico Albuquerque, New Mexico, May, 2014
- [4] Daden Limited, "Deploying Chabots to Customer Advantage", A White Paper, November 2011
- [5] ELIZA Website [Online]. Available: <http://nlp-addiction.com/eliza/>
- [6] Jabberwacky Website [Online]. Available: <http://www.jabberwacky.com/>
- [7] Karolina Kuligowska, "Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents," Professionals Center for Business Research, Vol. 2, February 2015(02)
- [8] Rose Chatbot [Online]. Available: <http://ec2-54-215-197-164.us-west1.compute.amazonaws.com/speech.php>