

Extensive Study: Performance, Metrics and Usability of Chatbot

Shraddha Pattanshetti¹

¹Student, Department of Information Technology, D.K.T.E's Society Textile and Engineering Institute, Ichalkaranji, India

Abstract - Improvements in technology have led to digitalization. Chatbots have become significant when it comes to transformation driven by mobility. Chatbots are meant to have interaction with humans. The conversation might take place in terms of speech and text inputs. There are varieties of chatbots implemented based on the complexity of human questions. Chatbots are being implanted in a variety of areas like HealthCare, Marketing, Entertainment, Education and so on. Because of its massive usage, it becomes significant to look after the usability of chatbots. People using chatbots will look for improvised features like reduced response time, user experience, consistent reply, relevancy, accuracy and precision, Multi-lingual, personalization, increased efficiency. So, the main intention behind this paper is to provide insights regarding amelioration done in chatbots in various areas. The research paper provides in-depth information regarding limitations and heuristics followed to improve the usability of chatbots in various aspects. In addition to this, Performance metrics on which chatbots are evaluated are included.

Key Words: Usability, User Experience, Heuristics, Human perceptions, User Interface, Performance.

1.INTRODUCTION

Chatbots is a program software that allows end-users to interact with computers online via text or speech. Usually, chatbots are accessed online through mobile applications, websites or virtual assistants. Major technologies involved in the development of bots are Artificial Intelligence, Natural Language Processing and Machine learning. AI-enabled bots depend on previous references and are developed to handle complex human queries. While rule-based chatbots follow a sequential path where IVRS is used that is programmed with predefined answers and depends on current information. Some popular examples of chatbots are Adidas women, amazon's ALEXA, WHO's Whatsapp Chatbot, SEPHORA Assistant, Domino's Chatbot, IBM Watson.

Chatbots identify the context of the user and then extracts the relevant data from users' question. After this, it returns the response to the query in form of a generic text, data from the database, a Piece of information based on what the user has provided as input, the result of the interaction performed by chatbots with other backend applications, error question to re-understand the user's requirements.

The main objective to implement chatbots was to boost efficiency in terms of operations, reducing costs, offering

convenience, added customer services. The first chatbot named ELIZA was created in 1965 which could answer simple decision tree questions. Later, in 1972 chatbot named Parry was created by Kenneth Colby to simulate a person who suffered from paranoid schizophrenia. Then, ALICE was developed by Richard Wallace in 1995 that miserably failed the Turing test. Finally, A developer named Rollo Carpenter was responsible for developing a chatbot in the year 1988.

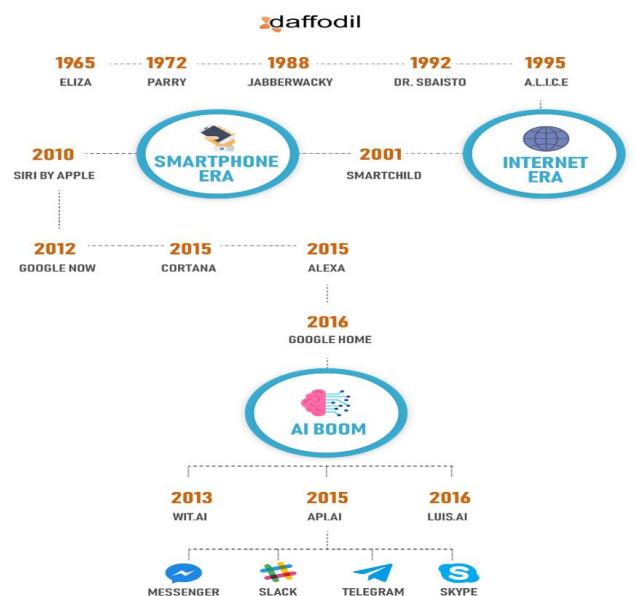


Fig -1: History of Chatbots

With benefits come limitations. There are lots of issues that people come across while using chatbots. We know, chatbots are machines. At times chatbots might not interpret human context or requirements. The reason is they are rule-based and follow predefined paths. Also, training them requires tons of the dataset and hence not every question put forth by the user is understood by a chatbot. Another limitation is that bots require maintenance and up gradation in terms of data or knowledge. The only way to solve this is to provide the bot with as much as possible old as well as new training data. Chatbots lack in taking decisions. The reason is they cannot distinguish between good and bad. For example, Microsoft build for Twitter, the bot was considered racist and rogue because of its incapability to differentiate between good and bad comments. Machines lack emotions. Hence, they are unable to show empathy to the user that might be sad, angry and annoyed. Chatbots are more obvious because they provide the same standard for multiple questions. Even

if the question is paraphrased, it provides the same apology that might frustrate users. Bots are programmed. So, if the question asked is not present in the data provided to bots, then they cannot answer.

1.1 Literature Review

A Tool of Conversation: Chatbot, this article gives in detail insights of how chatbots are designed and factors that should be considered while designing the chatbots. In addition to this, it gives overview of design techniques, approaches and models used in the implementation (M. Dahiya, 2017).

Research paper on Personality Matters! Improving the User Experience of Chatbot Interfaces notifies us how Human perceptions like Emotions, Personality and Anthropomorphism makes usability of chatbots better. It gives in detail information about Prototype and conversation design, Personality Frameworks that includes Personality Trait Model, Design process, The chatbot Role (Smestad, 2018).

Can we improve the User Experience of Chatbots with Personalization? specifically talks about Chatbot Implementation, Chatbots in Netherland, Customer Service: Finance & Insurance Chat-bots, Conversational aspects of chatbots, Dialogue Management Strategy and Maturity Model Chatbots. This paper explains Maturity model for chatbots by Smiers. It gives insights into how chatbots can interact with humans employing conversational interface with and without personalization effect (Danielle, 2017).

Paper, The User Experience of Chatbots elaborates on the user experience. Furthermore, it mentions User Experience factors and Guidelines that should be kept in mind to while designing chatbots. Papers also discusses Good User Experience: Human-like but not Human (Zeljko, 2018).

The article, how does an Intelligence Chatbot Affect Customers Compared with Self-Service Technology for Sustainable Services? gives introduction to AI and Hospitality, Service Failure and chatbots (Taehye, 2020).

IEEE Paper, Joker Chatterbot Re-WECHAT 2016 explains the example-based system joker chatterbot that uses a database of indexed dialogue examples automatically built from a television drama subtitle corpus to manage the social open-domain dialogue (Guillaume, 2016).

2. TYPES OF CHATBOTS

Evolution in chatbots has paved the way for the growth in technology. This growth is a benefit socially and economically. Chatbots are classified into three types according to their technical complexity.

2.1 Simple Chatbots

Simple chatbots are also known as rule-based bots because they adhere to the constraints or limits that are applied to them. Generally, they possess finite capabilities. They follow IVRS (Interactive voice response system) where

the bot presents questions that already have predefined answers and the end-user needs to select an option repetitively until the user won't get the required answer to the query. This type of chatbot solely depends on its inferences on current questions or information, not on previous interactions. These are easy to build and train and well suited for straightforward conversations.

For instance, while ordering food on any online application like Zomato, flow is predefined. The bot starts by asking the customer to order the food. When asked for the menu, it displays the list and asks the customer to order. Finally, it requests for address and payment method. Customer needs to only click on options to complete the entire process.

2.2 Smart Chatbots

Smart chatbots are most commonly known as AI-enabled bots. These are designed to automate the routine tasks that humans perform. They are intelligent enough to understand the language, emotion, intent that the user wants to convey. Conversations here are free-flowing. They are programmed in a way that they can adapt to situations that they are not aware of. Training them might require tons of data. Hence, it is difficult to implement them.

Smart chatbots like Siri, Cortana, Alexa and many more are trendy because of their existing features also termed as virtual assistants. Virtual assistants don't follow commands already programmed but learns from human interaction. From playing music to having a small conversation with a person, chatbots more or less have gained more popularity.

2.3 Hybrid Chatbots

As the name suggests, hybrid chatbots are the integration of simple and smart chatbots. Looking at situations today, both the bots are equally important. They follow rules but can interpret the context the user wants to put forth. It is the perfect balanced tool that can be widely used in businesses.

An application where hybrid chatbots are used is a medical diagnosis. Visitors convey their health issues over a voice chat. This is a more efficient way to note down the symptoms rather than using the rule-based format. In this way, the patient can describe the issues in depth so that prescribing medicine for that particular disease might become easy because of the detailed information.

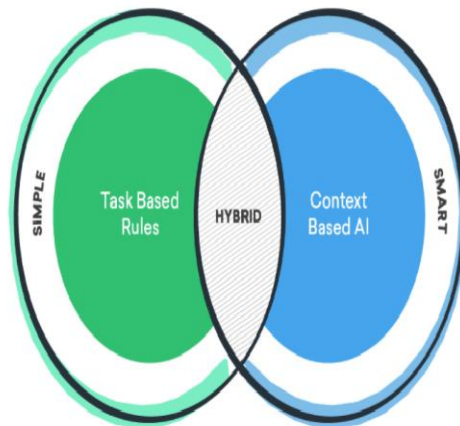


Fig -2: Types of Chatbots

3. FACTORS THAT INFLUENCE HUMAN PERCEPTION

Two major factors that dictate Human perceptions:

3.1 Emotions and Personality

Personality plays a major role that explains human behaviour and is considered to be 'the fundamental linchpin in human relationships' (Dryer 1999). It helps to measure the expression or behaviour an individual will exhibit in a particular situation or circumstances. Personality assists us to study human interactions, classifying the behaviour into simplifying traits, the response of an individual to a person's behaviour and thus is advantageous in social relationship development. Various models and theories are put forth to study personality. One of the recognized theories is the Five-factor model FFM (Costa and McCrae, 1987). These factors are identified after lexical analysis and are common traits that include people from two extreme personality ranges.

Five factors are:

3.1.1 Openness to Experience

People always ready to witness adventure, thrill, emotions are more likely to adapt to openness concerning the experience. This artlessness might build curiosity in an individual that might lead to innovation, popping unusual ideas and therefore improving the imagination skills. They dare to undertake tasks that might involve risks but work with perseverance to accomplish them. On the other hand, people who play safe are conventional thinkers. They are biased toward dogmatism and community thoughts.

3.1.2 Extraversion

Extraverts are those who actively participate in every external activity. They are acknowledged for their immense energy. Their presence might be more dominating in social activities. These people enjoy their life to the fullest since they are the most enthusiastic and action-oriented individuals. However, introverts seem to be quieter, less active, require less stimulation and more isolation.

3.1.3 Agreeableness

Agreeableness is concerned with social co-operation and harmony. People with this trait have friendly nature and always get along with other individuals. They possess qualities like kindness, generosity, humbleness, trustworthiness, helpful. They have a positive vibe that attracts people. People who are unable to adjust with other people are self-concerned and un-cooperative. They might to an extent involve in arguments and are always willing to accept challenges because of their competitive nature.

3.1.4 Conscientiousness

Being disciplined is an art that makes an individual stand out from the crowd. Being highly conscientious might end up imbibing more concentration and focus. These people plan their activities for smooth execution and are often called optimistic. On the other hand, low conscientious people lack reliability and confidence. They might be less flexible because of their less concentration.

3.1.5 Neuroticism

Neuroticism is a tendency in which people experience pessimistic thoughts such as anxiety, impatience, anger, annoyance or enmity. The thoughts can be the result of emotional instability. They have low tolerance and that is why they are more vulnerable to stress or other negative traits. These reactions might stay for a long period that adversely affects an individual himself and the people around him. Neuroticism might also negatively affect the social relationship.

3.2 Anthropomorphism

To understand how personality influences chatbots we need to first understand anthropomorphism. Anthropomorphism is a term in which nonhuman characters are assigned with human qualities. Instances that explain anthropomorphism are found in pieces of literature that include fairy tales, animated characters like Pinocchio, Donald Duck, Mickey mouse. In addition to this, the movie that specimen's anthropomorphism is a very renowned Pixar movie, 'Cars'. The vehicles in this movie align with the VW bus that is personified as Hippie. It is a human attitude that starts with childhood and continues to be throughout life and differ in complexity and not qualitatively. Anthropomorphism is perceptual. To illustrate this phenomenon, we can consider humans who can see faces in the clouds (Guthrie 1993). The humanlike mind is a cardinal constituent of Anthropomorphism. Thus, it marks as a foundation for chatbots to make the conversation more natural and follow social conducts and heuristics.

4. USABILITY HEURISTICS

Heuristic evaluation is an approach to inspect design issues correlated with a user interface concerning the efficacy,

productivity, user-friendly and complacency in terms of technology, interface, medium, product, service, latency. In 1990, Jakob Nielsen was responsible for 10 heuristics principles for assessing UI.

4.1 System Status Usability

System users must be instructed with the whereabouts, time required to complete the flow, progress of the user in the flow and flow to be covered. Chatbots are conversational but conversations can be transient, old messages are not current or updated and information may overwhelm and confuse the user. These issues can be resolved by sending appropriate feedback to engage the user and avoid drop-offs.

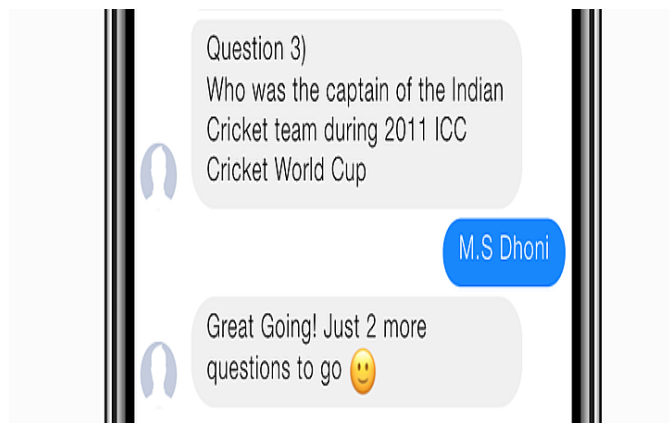


Fig -3: System Status Visibility

Secondly, the response time of bots should be reasonable and responsive. If the request is taking ample time then, responding can help reduce the user's waiting time.

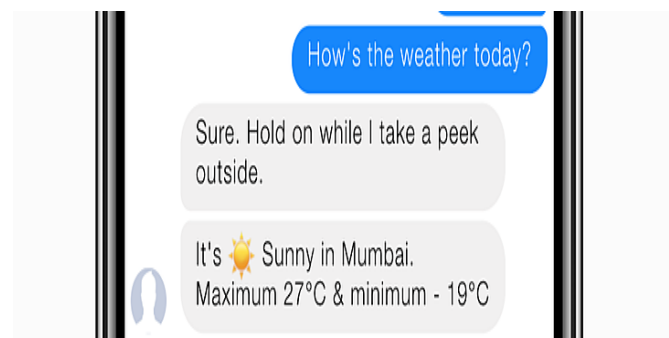


Fig -4: System Status Visibility

4.2 Match between System and Real world

Chatbots should use language that the user is familiar with, rather than using technical jargon that is uninformed. Information should be displayed in natural and chronological order. Training the bots with the command line, interactive style that users use nowadays can make the system more interpretable. In addition to this, the usage of appropriate words and phrases is most significant. Furthermore, Introduction, mid-conversation and graceful conclusions can make the user experience better.



Fig -5: Match between System and Real world

4.3 User Control and Freedom

Frequently, users can opt for functions by mistake and might require an emergency exit to quit the extended conversations. So, the user should arrange an escape hatch and inform the user of valid options at every stage in interaction. In a nutshell, the system should support undo and redo options. Users should be granted with liberty to choose the direction to continue or to step back.

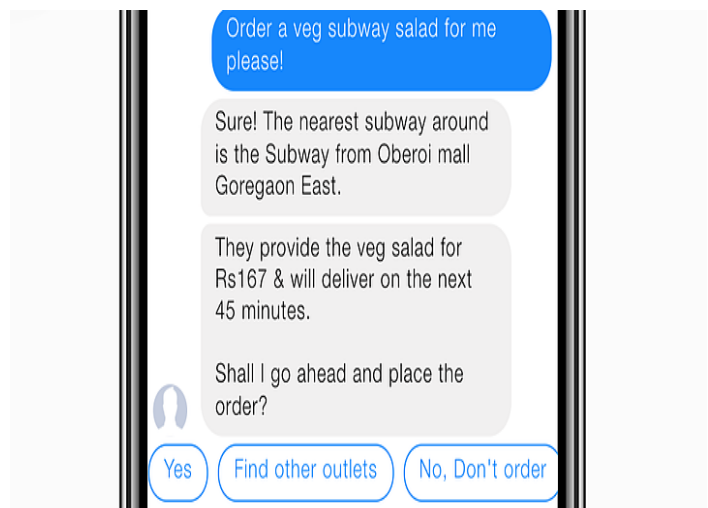


Fig -6: User Control and Freedom

4.4 Consistency and Standards

Bots need to be internally consistent. Usage of ambiguous words, situations or actions that can confound the customer should be avoided. Chatbots must align with the monotonous style of language, personality and voice rather than undertaking distinct techniques. Conventions need to be obeyed for the interaction and visual elements. Furthermore, Branding the companies to showcase their values with the

help of typography, colours on websites or applications can be done more efficiently with chatbots. Bots can prove better when it comes to brand identity, also known as voice branding.



Fig -7: Consistency and Standards

4.5 Error Prevention

Humans are never born perfect and tend to make mistakes. But chatbots are software integrated with artificial intelligence. They are intelligent enough to handle mistakes carried out by users. In the worst-case, where errors can't be prevented, bots should provide users with the confirmation option before they commit the task during any critical stage in interaction.

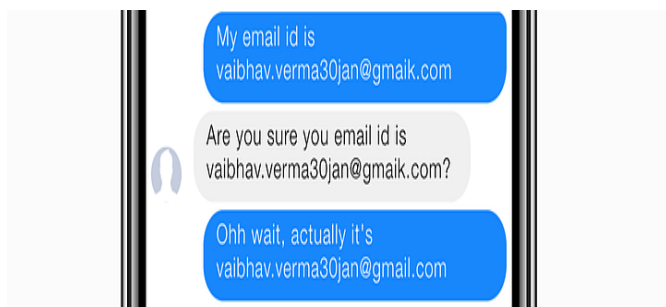


Fig -8: Error Prevention

4.6 Recognition rather than Recall

Most often, Customers read the first part of the message sent and omit the next half. This problem can be caused by several reasons. Firstly, the client is not able to comprehend which makes them puzzled. Secondly, the end-user feels the information is less interactive and tedious. Users always prefer a design that is adaptable and user-friendly. Complex designs and visuals can reduce the usability heuristics of bots. Briefly, the user's memory load must be shortened by

making options visible by reducing the reading effort or usage of emojis and graphs.

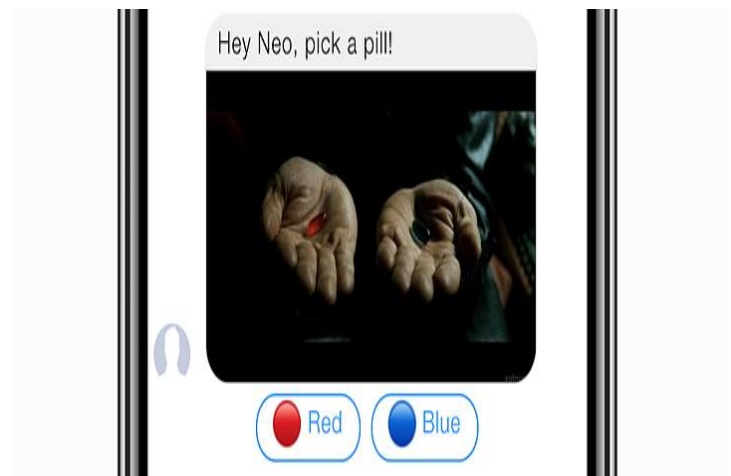


Fig -9: Recognition rather than Recall

Structured messages can solve this issue to a greater extent. They provide a distinct set of options by removing the ambiguity in conversation and allows users to select an option.

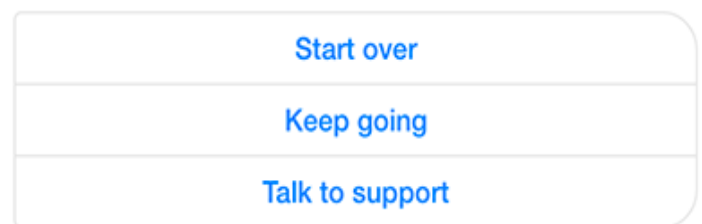


Fig -10: Structured messages

4.7 Flexibility and Efficiency of Use

Sometimes, it becomes mandatory to go with the longer flow to aid new customers through services given by bots. But this might become an impediment to expertise. Thus, Accelerators skips the redundant and unnecessary parts from the flow and jump to the main content. They avail us of the advanced level customer experience that reduces ameliorates response time and administrative burdens. Accelerators balance the customization and speed. In this way, bots adapt to both experienced and non-experienced end users.

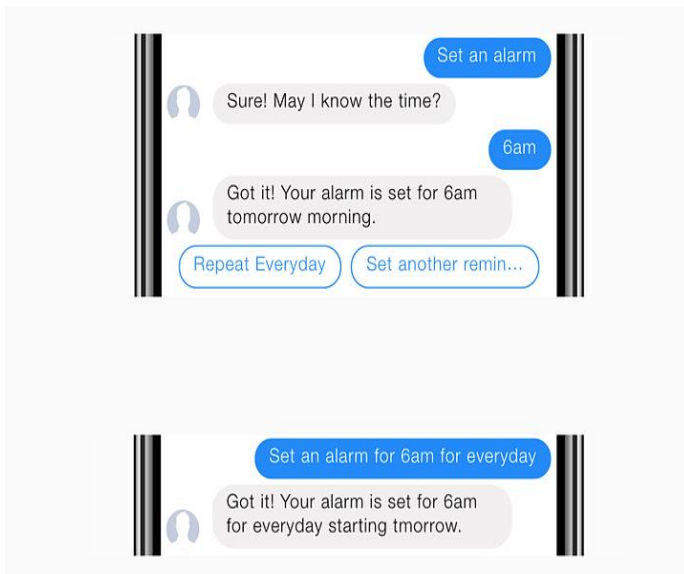


Fig -11: Flexibility and Efficiency of Use




Medium	Advantages	When to use	Examples
Voice 	Hands Free	Control and command	"Switch off the lights"
Video 	Informative, Engaging	Comprehensive content that needs full attention	"Show the highlights of the today's Wimbledon match"
Text/ Image 	Fast & Instant feedback	Glanceable Information	"Show my tasks for today"

Fig -13: Aesthetic and Minimal Design

4.8 Aesthetic and Minimal Design

Aesthetics plays a significant role when it comes to design qualities like balance movement, pattern, colour. It impacts the inner qualities and memory of the design. Minimalism includes essentials by maximizing utilization and usability. Aesthetics and minimalism can be obtained by the coherence and accuracy. Content should be to the point and crisp or else irrelevant information can diminish relative visibility of pertinent units of information.

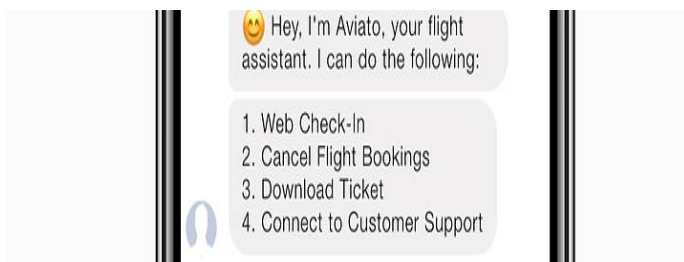


Fig -12: Aesthetic and Minimal Design

4.9 Help users Recognize, Diagnose and Recover from errors

At times, chatbots might find it hard to interpret what the user is trying to say. The reason can be unclear or incomplete information provided by the user. In such cases, bots should harness AI-enabled solutions where the frequency of errors and uncertainty of the outcomes are managed professionally.

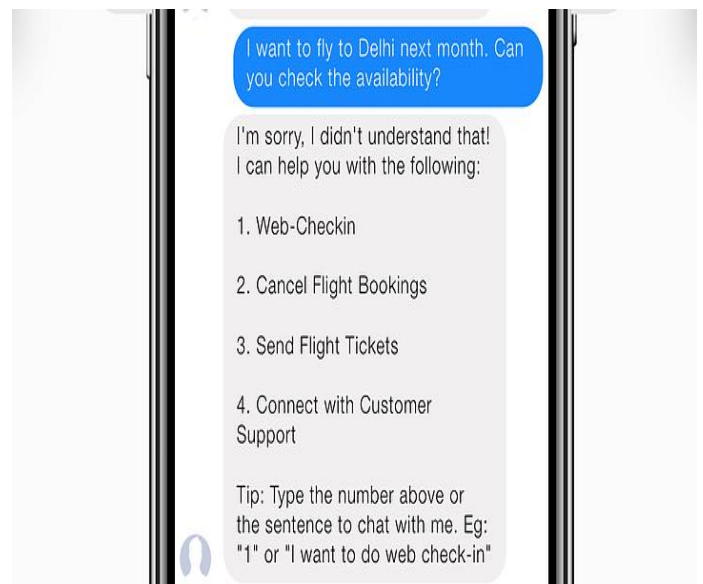


Fig -14: Help users Recognize, Diagnose and Recover from errors

4.10 Help and Documentation

Documentations can be required to acquire thorough official information regarding the system, company, object, process. Users might demand data to authenticate and

increase their knowledge. So, the human assistant should be far-sighted to answer huge and long queries. The answer provided to the question should be relevant and to the point.

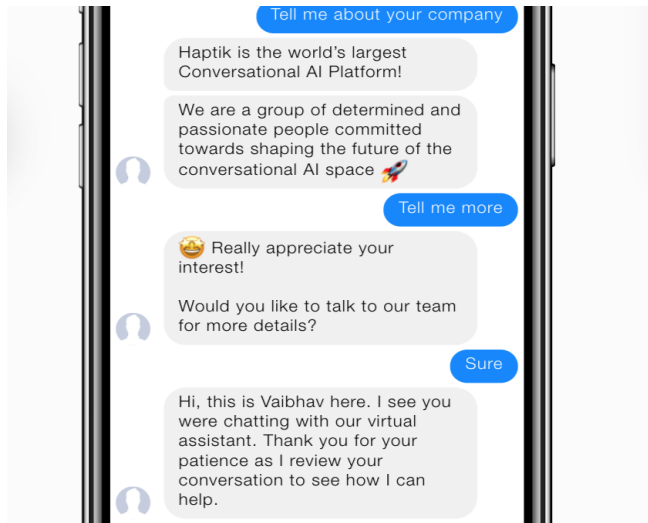


Fig -15: Help and Documentation

5. CHATBOT METRICS

After developing software, it is important to keep track of the software progress. To track the progress performance metrics are used. Metrics measure the quality of the product thereby conveying whether client requirements can be met or not. If the software developed has promised quality and standard then the organization as a whole is at advantage in terms of profits, opportunities and most important customer satisfaction. In addition to this, choosing correct metrics is very important to bring out the effectiveness of the process. Metrics vary according to industry standards like safety, legal or contractual purposes used to incur high profits, reduce complaints, increase effectiveness, improve savings.

Chatbots need to be monitored by companies to meet the expectations of the end-user. When it comes to the usability of chatbots, user experience is always a top priority. Recovery from errors, aesthetic design, flexibility, consistency freedom in using chatbots and many more are checked. This inspection can be done if correct metrics are used. At times, analytics are underestimated and unnoticed. Metrics and Analytics do not solely define the chatbot success but they provide valuable insights that improve heuristics and user experience.

KPI is also known as Key performance indicators are used to track the success of a specific target market for chatbots. It is widely used by organizations to meet their business objectives. High-level KPI's mainly focus on the overall performance of the company. On the other hand, low-level KPI's focus on subordinate departments like finance, sales, marketing, accounts, services, support, HR and others.

Below are the aspects used as chatbot metrics to test the performance of bots.

Metric	Type	Data Collection Method
Total elapsed time	Efficiency	Quantitative Analysis
Total number of user/system turns	Efficiency	Quantitative Analysis
Total number of system turns	Efficiency	Quantitative Analysis
Total number of turns per task	Efficiency	Quantitative Analysis
Total elapsed time per turn	Efficiency	Quantitative Analysis
Number of re-prompts	Qualitative	Quantitative Analysis
Number of user barge-ins	Qualitative	Quantitative Analysis
Number of inappropriate system responses	Qualitative	Quantitative Analysis
Concept accuracy	Qualitative	Quantitative Analysis
Turn correction ratio	Qualitative	Quantitative Analysis
Ease of usage	Qualitative	Questionnaire
Clarity	Qualitative	Questionnaire
Naturalness	Qualitative	Questionnaire
Friendliness	Qualitative	Questionnaire
Robustness regarding misunderstandings	Qualitative	Questionnaire
Willingness to use system again	Qualitative	Questionnaire

Fig -16: Chatbot Metrics

6. CONCLUSION

To conclude, every aspect has advantages and disadvantages. Also, these issues cannot be avoided because of their connection with business and might be an obstacle in the growth of the company and revenue generation. Regardless of this fact, organizations are still investing in chatbots. A survey asserts that 55% and more clients are attracted to chatbot abilities because of their availability 24*7. We can say those virtual assistants and smart bots if ameliorated they can change the direction of the world. Apart from social media, they are being introduced in critical sectors like banking, healthcare, e-commerce. Moreover, chatbots are a promising tool for collecting necessary data that can be helpful in future.

7. FUTURE WORK

AI and ML are surrounded by us helping us in every aspect making our life much easier. Statistics by HubSpot suggest that 21% of people admit that chatbots are very useful if integrated with business because of the benefits they bring. After thorough research, we know that chatbots are born with some disadvantages that might drastically affect the industry and areas associated with it. So clearly, some improvements are required to embrace the future of chatbots.

Improvements can be done in programming techniques. Chatbots can be implemented using advanced frameworks like TensorFlow, Keras or PyTorch which will surely deliver better results and will pave a better future for chatbots. The research and development team of organizations all over the world are working to implant chatbots with human features. This will increase customer satisfaction because chatbots might work more like humans than the machine. Evaluation of chatbot depends on a key factor that is user experience. User Interface not being up to mark is creating issues. Hence, utilizing top-notch UI-UX will help chatbot rise.

REFERENCES

- [1] Bergart, R., & McLaughlin, J. (2020). Can User Experience Research Be Trusted? A Study of the UX Practitioner Experience in Academic Libraries. *Weave: Journal of Library User Experience*, 3(2). <https://doi.org/10.3998/weave.12535642.0003.201>
- [2] Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, 100006. <https://doi.org/10.1016/j.mlwa.2020.100006>
- [3] Dilmegani, W. B. C., & Dilmegani, C. (2021a, July 5). Top Chatbot Development Best Practices in 2021. *AIMultiple*. <https://research.aimultiple.com/chatbot-best-practices/>
- [4] Dilmegani, W. B. C., & Dilmegani, C. (2021b, September 8). The Ultimate Guide to The Top 14 Benefits of Chatbots. *AIMultiple*. <https://research.aimultiple.com/chatbot-benefits/>
- [5] Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience*, 5(1). <https://doi.org/10.1007/s41233-020-00033-2>
- [6] Kuligowska, K. (2015). Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents. *Professionals Center for Business Research*, 2(02), 1-16. <https://doi.org/10.18483/pcbr.22>
- [7] Malhotra, A. (2021). Character Chatbot - A Conversational AI Chatbot which can alter its Character according to User Specifications. *International Journal for Research in Applied Science and Engineering Technology*, 9(VI), 971-977. <https://doi.org/10.22214/ijraset.2021.35090>
- [8] McTear, M. (2020). Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots. *Synthesis Lectures on Human Language Technologies*, 13(3), 1-251. <https://doi.org/10.2200/s01060ed1v01y202010hlt048>
- [9] Rauf, I., Troubitsyna, E., & Porres, I. (2019). A systematic mapping study of API usability evaluation methods. *Computer Science Review*, 33, 49-68. <https://doi.org/10.1016/j.cosrev.2019.05.001>
- [10] Ruikar, V. (2016). Interactive Voice/Web Response System in clinical research. *Perspectives in Clinical Research*, 7(1), 15. <https://doi.org/10.4103/2229-3485.173781>
- [11] Skjuve, M., Haugstveit, I. M., Følstad, A., & Brandtzaeg, P. B. (2019). Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human Technology*, 30-54. <https://doi.org/10.17011/ht/urn.201902201607>
- [12] Um, T., Kim, T., & Chung, N. (2020). How does an Intelligence Chatbot Affect Customers Compared with Self-Service Technology for Sustainable Services? *Sustainability*, 12(12), 5119. <https://doi.org/10.3390/su12125119>
- [13] Tucker, K. (2018, May 26). A Chatbot Abstract - Chatbots Life. *Medium*. <https://chatbotlife.com/a-chatbot-abstract-1cd002e7a480>
- [14] Verma, V. (2019, November 19). 10 Usability Heuristics every Chatbot company should follow. *Medium*. <https://uxdesign.cc/10-usability-heuristics-to-design-better-chatbots-654223552533>
- [15] Wikipedia contributors. (2021, September 8). Big Five personality traits. *Wikipedia*. https://en.wikipedia.org/wiki/Big_Five_personality_traits
- [16] Skjuve, M., Haugstveit, I. M., Følstad, A., & Brandtzaeg, P. B. (2019b). Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human Technology*, 30-54. <https://doi.org/10.17011/ht/urn.201902201607>
- [17] Malhotra, A. (2021b). Character Chatbot - A Conversational AI Chatbot which can alter its Character according to User Specifications. *International Journal for Research in Applied Science and Engineering Technology*, 9(VI), 971-977. <https://doi.org/10.22214/ijraset.2021.35090>