Pakistan Journal of Engineering and Technology, PakJET
Multidisciplinary & Peer Reviewed
Volume: 4, Number: 2, Pages: 108- 112, Year: 2021

Artificial Intelligence Based University Chatbot using Machine Learning

Zimal Mehboob Khan[1], Hafeez-ur-Rehman[2], Maria Maqsood[3], Khalid Mehmood[4]

[1,2]Khwaja Freed university of engineering and information technology, Pakistan
[3]Tahsil Head Quarter Sadiqabad, Pakistan
[4]Govt. Associate College for Women, Rahim Yar Khan, Pakistan

Corresponding author: Zimal Mehboob Khan (email: Zemal.mehboobkhan@gmail.com)

Abstract: **Today the customer support needs 27/7 services and the people face some issues related to organizations they contact the Support Center / Help Desk and the support centers are people they are sitting to serve the customer care and they serve the customers need and demand this should increase the cost of the people setting the whole and night for services and organization need to reduce this cost and time, desire to automate this process. The proposed study is the replacement for this problem. People use this solution in university chatbot for help and support for peoples who are visiting the office this will also decrease the traffic of the admission office some universities are so far away from the city area and people travel a long distance to reach the university for a very ordinary work this will reduce this and allow the office staff to focus on the other productive activities. University chatbot build by using the data collected by user demand and interaction using the machine learning algorithms. In the proposed study the Random forest is outperform and the Decision Tree and Support vector machine also response good with normal and even spin text. This can response under the domain of the university related help and support and guide in the proper way for the use of university website and refer them the proper links and contact to reduce the traffic from physical visit office. The results show that the Random Forest is outperform for the proposed study. Our recommendations are if the universities are starts using proposed application this will strongly impact on the university's communications.**

Index Values-- Chatbot, *Machine Learning, Random forest, Decision Tree, Support vector machine.*

I. INTRODUCTION

Chatbot is Artificial Intelligence based computer software in which the software which is called chatbot interact with humans like a human in the closed domain or open domain base it depends on the need and demand of the chatbot what for that is used and what kind of situation the chatbot can face[1]. Chatbot works on the natural language processing concepts. Chatbot having the two types 1. Dialogic chatbot These chatbots are get the input from the user and used the NLP tools and features to processes the input text and find the best match for the user this kind for chat bot are mostly works on the websites and social media applications, 2. Rational Chatbot this kind of chatbot store the user input questions and quires and then based on the external knowledge (Corpora of the dataset) they response on the common sense features and responses with competence , 3. Embodied Chatbot this kind of chatbot are feels like the presence of the user which is actually feel like the chatbot talking with user as actual human so the early chatbots name like (ELIZA, CHARLIE, etc.) they feel the user that chatbot is as a human and their appearance also looks like a human this kind of chatbot are embodied in the agent and gives the impression for user[2], [3]. Today the advance chatbot like Amazon's Echo and Alexa are Apple Siri, Microsoft Cortona these bots are take the advantage of the advance machine learning technologies to processes the information and responses with effectively and efficiently[4]. There are modern using the information retrieval processes like NLP and other are using the SMT statistical machine translation this will translate the input and then processes based on the SMT and then shows output[5].

The advance chatbot use the seq2seq and SMT to processes the input and encode and decode the processes and this is current best practice. We also use the SMT but the closed domain like only for university communication and this will makes us different in the domain the best thing is that we not only use for university it can be used for open domain with some corpora changes[1], [5], [6].

Our contribution:

1. University chatbot
2. Different responses based on different algorithms.
3. Based on the run time performance can be increased based on custom training.
4. Work like human response (web links, Contact No, guidance for website, important dates information etc.)
5. Questions and answers also can be improved during run time.
6. Works like a messenger and easy to use and user friendly very effective and lite deployment.
7. Purely build on the opensource libraries

II. LITERATURE REVIEW

Chatbots are "natural language online human-computer conversation system". Alan Turing, who posed the question "Can machine think?" in 1950, is credited with inventing the chatbot. Since Turing, developments in natural language processing and machine learning have enhanced chatbot technology. Turing envisioned the challenge as a "imitation game" (now known as the Turing Test), in which a "interrogator" asked human and machine subjects questions with the purpose of identifying the human. We claim the machine can think if the human and machine are indistinguishable. In 1966, MIT's Joseph Weizenbaum invented ELIZA, the first chatbot that came close to simulating a person. ELIZA would recognize keywords in an input text and pattern match them against a set of pre-programmed criteria to create appropriate answers. Since SLIZA, the creation of increasingly intelligent chatbots has progressed. Kenneth Colby of Stanford invented PARRY, a bot that pretended to be a paranoid schizophrenic, in 1972. Chatbot use has risen as well, thanks to the introduction of chatbot platforms by Facebook, Kik, Slack, Skype, WeChat, Line and Telegram. Facebook Messenger has 30,000 bots and 34,000 developers on the platform by September 2016. In August 2016, the Kik Bot Shop stated that the 20,000 bots it has produced had "exchanged over 1.8 million messages"[7].

There are several different approaches of evaluating chatbot performance. Chatbot have distinct purposes in terms of information retrieval (IR): virtual assistants, question-answer bots, and domain specific bots. Evaluators should ask the chatbot questions and make requests, then measure accuracy, precision, recall, and F-score in relation to the accurate chatbot response. The bot's purpose, from the standpoint of user experience, is presumably to maximize user happiness. Users should be polled by evaluators, who will evaluate bots based on their usability and satisfaction. Bots should be tested by linguistic specialists on their capacity to construct complete, grammatical, and meaningful phrases in order to approach speech. Speech is an effective and common mode of communication that is "widely accepted as the future of interaction with phone and computer application". speech-to-text conversion start with a procedure known automatic speech recognition (ASR), which aims to achieve speaker-independent large vocabulary speech recognition (LVCR). Natural language processing (NLP) aims to convert the ASR's unstructured output into a structure representation of the text that includes spoken language understanding (SLU)[8], [9]. Dialogue act recognition is a method of extracting meaning from natural language by determining the function of the text/sentence (e.g. is this an inquiry, suggestion, offer, or command). A corpus of sentence is labelled with the function of the sentence, and a statistical machine learning model that takes in a phrase and outputs its function is developed in conversational act recognition system.

To derive information from the text, we turn unstructured text – such as ASR output or text entered into a text-only chatbot into organized grammatical data objects, which the Dialogue Manager will then handle[10]. We disregard sentence structure, order, and grammar in favor of counting the number of times each word appears. We utilize this to create a vector space model in which stop words (such as a, the, and so on) are eliminated and morphological variations (such as talk, talks, talked, and so on) are lemmatized and stored as instance of the basic lemma (e.g. talk). The bag of words strategy is straightforward it does not need syntactic knowledge, but it is not precise enough to address more difficult issues for the same reason. The most important aspect of the chatbot architecture is response generating. A structured representation of the spoken word is sent into the Response Generator (RG)[11]. This offers information about who is speaking, the context, and the discussion history.

The RG creates a response for the user, which it will give to the Dialogue Manager as an output. After the chatbot has chosen a response, the Dialogue Manager (DM) must choose from a variety of communication tactics, including linguistic techniques to make the chatbot appear human, and deliver the message. Chatbots are only as smart as the knowledge they have at their disposal. Retrieving training data for machine learning classifiers used in associative bot models or creating data corpuses for information extraction bots is essential for achieving human-like interactions. In order to improve bot accuracy, expanding the corpus of data utilized by bots is a good complement to algorithm development. This section focuses on data collecting strategies for chatbots. Initially, rule-based techniques were supplemented with manually built knowledge base for bots like ALICE, which were developed in the twentieth century[7], [11], [12]. Developers then began to employ massive amount of human-annotated conversation data. The final stage in the process is text-to-speech (TTS), which translates the produced response into voice and return it to the user. Text analysis is the initial stage of TTS, and it involves converting text into phonemes with pitch and duration. The second phase is waveform synthesis, which involves concatenating portion of recorded voice corresponding to each phoneme to produce speech. Text analysis start with a normalization procedure, in which portion of text are broken down into sentences, phrases into words and punctuation, and words into their phonemes[4].

Sentence tokenization is the process for looking for common sentence break punctuation such the exclamation mark, question mark, colon, semicolon, or period. The conversion of non-standard terms into natural English is also included in normalization. Pronunciation is the second phase in text analysis. This necessitates the use of a pronunciation lexicon, which contains mapping between words/phonemes and their pronunciations, as well as name lexicons, that relate names to their pronunciations[3].

Chatbots have been used in a variety of consumer-facing applications, including online travel companies (OTAs), thanks to artificial intelligence. The goal of this research is to define five quality aspects of chatbot services and evaluate their impact on user confirmation, which leads to continued usage. The effect of technology fear as a moderator in the links between chatbot quality factors and post-use confirmation is also investigated. 295 Chinese OTA user participated in the survey. To examine

measurement and structural models, partial least square (PLS) was applied. Post- use confirmation is positively correlated with understandability, dependability, assurance, and interaction, and technological fear moderates the associations between four chatbot quality aspects and confirmation. Confirmation has a positive relationship with contentment, which promotes the inclination to continue using something. By studying the moderating influence of technology anxiety, this study investigates how customers perceive chatbot services in OTAs (human-like agents vs. technology-enabled services)[2], [3].
Obesity and overweight are severe health problem that have diverse and interconnected causes across the world. Chatbots, on the other hand, are becoming more popular as a tool to connect with consumers in mobile health applications[9]. The user-centered concept and feasible study of a chatbot to gather connected data to enable the 32 research of individual and societal overweight and obesity causes in communities are described in this paper. Through an open poll on 52 wireframes developed by 150 design students, we first analyzed the customers' demands and gathered their graphical preferences it included inquiries on sociodemographic characteristics, food and exercise patterns, the need for overweight and obesity applications, and desired functionality. We also spoke with a panel of experts.
After that, we planned and built a chatbot. Finally, we completed a feasibility assessment through a pilot study [13]. We received 452 to the poll and spoke with four experts. A user's status score was established as a normalized sum (0-100) of diet, physical activity, BMI, and social network ratings. We tested the chatbot deployment on 85 healthy volunteers in a pilot study. We discovered 8 underweight persons (11%), 5 overweight individuals (7%), and no obese cases among the 74 individuals who completed all parts. The average body mass index (BMI) was 21.4 kg/m2, normal weight. Wakamola, a Telegram chatbot, is a useful tool for gathering information from a community regarding sociodemographic, dietary habits, physical activity, BMI, and particular ailments. Furthermore, chatbot allows users in social network to connect and investigate the reason of overweight and obesity from both individual and a societal perspective[14], [15].

III. DATA COLLECTION

The Dataset if collected from the university questions and answer help desk and manually collected on the question and answer. We actually build a form with different questions for different departments and help meetings with front desk people to track the most frequent question and what will be the answer for that question according to that department the hard paper forms and meetings and discussions and website chat room data are collected and combined and then we find the final look for the dataset as shown in Fig. 1. The dataset also included the some open domain dataset points that helps us to make the training good for chatbot.

	hi, how are you doing?	i'm fine. how about yourself?
0	i'm fine. how about yourself?	i'm pretty good. thanks for asking.
1	i'm pretty good. thanks for asking.	no problem. so how have you been?
2	no problem. so how have you been?	i've been great. what about you?
3	i've been great. what about you?	i've been good. i'm in school right now.
4	i've been good. i'm in school right now.	what school do you go to?

FIGURE 1.0: The Dataset view used for training of the UCB

IV. EXPERIMENTS AND RESULTS

We have adopted the processes mode shown in the figure 2.0. The processes model shows that the collection of the dataset is the first step and then we perform the necessary data cleaning processes and then the feature extraction like TFIDF, with N-gram unigram, bi-gram, uni-bi-gram and the after features extraction we build the dataset for training and the features are also improved by add some critical datapoints. The training and test are divided into two parts with 80% for the training and the 20% for test we use the machine leaning algorithm are Decision Tree, Random Forest and support vector machine with linear svc.
Our features for Decision Tree, Random Forest, and Support Vector Machine are TFIDF and Bag-of-words in the Bag-of-words we use the count vectorization that is also used for Decision Tree and Random Forest and Support Vector Machine. The Decision tree, Random Forest and Support vector machine are tested on the real time environment and check the responses of the chatbot we have shows some experiments we perform same input on the different algorithms and find the best one for proposed study.
In Fig. 2, we presented the processes model of the proposed study the model shows the three main parts data collection or gathering , cleaning second the using the feature engineering for feature extraction this feature extraction only the part to extract feature we have to find the which feature extraction method is optimal solution for our proposed study and meet the desired results., third one is the final part which is model building and testing the chatbot which is called prediction on the input text and response of the bot how much the accurate the bot understand our input and the processing of natural language processing and pattern recognition of the input text it is the whole processes model of the proposed study.

a. How the Chatbot works

The university chatbot the work flow is shown in Fig. 3. Step wise working of the chatbot is the when the user come on the website of the university and visit the university pages the click and open the like is stored in the dataset during the user interact with the website this will helps us for the training the chatbot mad based on the interaction we generate the questions and responses.

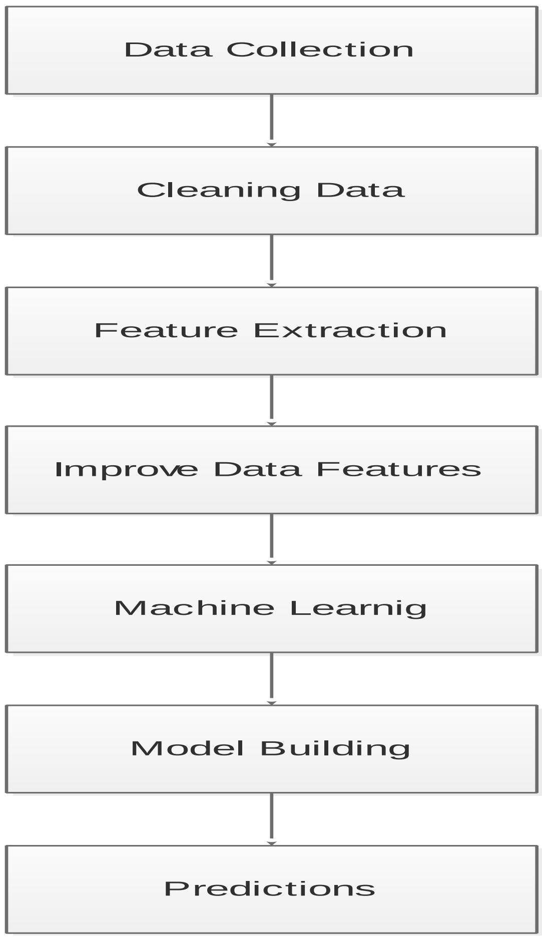

FIGURE 2: Processes Model for UCB (University Chat Bot)

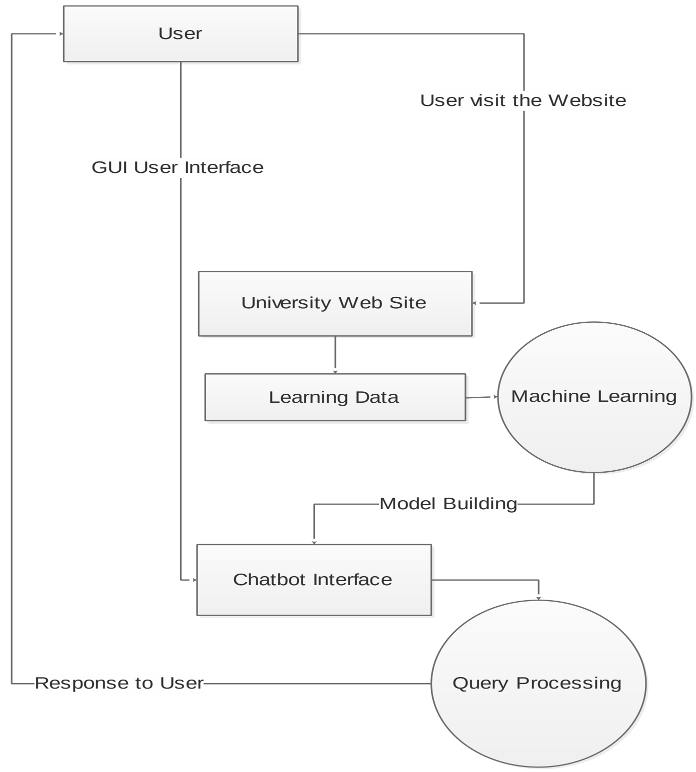

FIGURE 3.0: Working of the UCB from end-to-end processes

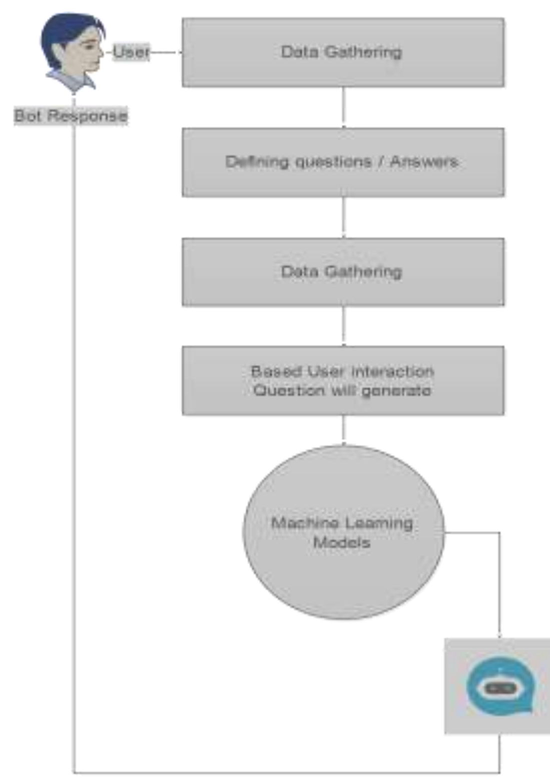

FIGURE 4: Real time work flow and interface of the UCB with human.

Then the dataset is boiled with the time to time and we use this kind of information for training for the chatbot and processes goes on and the time to time the training dataset and corpora are increased that will also increase the performance of the chatbot. We also can manually add the question and answers with append command and also improve the learning dataset for chatbot. We make it automated by tracking the visit and click of the user and get the URL and title and make the pipeline with requirements of the user and then this process goes so on and to collect the dataset on the run time and used them for training with updated and dynamic change of the website this will update the training and corpora for batter response. The processes and interface used in the development of UCB we present in Fig. 3.

In the proposed study check the different question with one algorithm and same questions with different algorithms and the same questions with different text spin than we notice that the UCB gives us the robust response this makes us confident for the deployment in the university website and portal. The proposed work is shown in Fig. 4. All the questions and the follow up responses are given in Tab. I and Tab. II. The proposed technique shows the achievement in terms of the responses over the asked question and its truthfulness.

Table I: Testing the processes of the RF, DT, SVM with same questions and response of the models.

Pipe.predict(['hi'])[0] *# chat With Decision Tree model*
'hello'
Piper.predict(['hi'])[0] *# chat With Random Forest model*
'hello'
Pipe.predict(['how are you'])[0] *# Chat with DT Model*
"i'm fine. how about yourself?"
Piper.predict(['how are you'])[0] *# Chat with RF Model*
"i'm fine. how about yourself?"
Piper.predict(['can you help me ?'])[0] *# Chat with RF Model*
'why me?'
Piper.predict(['how can i get the admission in university?'])[0] *# With RF*
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Pipe.predict(['how can i get the admission in university?'])[0] *# With DT*
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Piper.predict(['i want to get admission with you'])[0] *# With RF*
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Pipe.predict(['i want to get admission with you'])[0] *# With DT*
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'

Table II: Spin text and check the responses of the all models developed.

Piper.predict(['how can i get the admission in university?'])[0] *# With RF*
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Pipe.predict(['how can i get the admission in university?'])[0] *# With DT*
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Piper.predict(['i want to get admission with you'])[0] *# With RF*
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Pipe.predict(['i want to get admission with you'])[0] *# With DT*
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Pipe.predict(['great'])[0]
'i appreciate that.'
Piper.predict(['how can i get the admission in university?'])[0]
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Piper.predict(['how can i get the admission'])[0]
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Piper.predict(['what is admission processes'])[0]
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Piper.predict(['programs in university?'])[0]
'Contact Us Admission contact: 068-5882433 Admission Mobile: 0331-2869464 or Email Address: admissions@kfueit.edu.pk'
Piper.predict(['what is admission office numner ?'])[0]
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Piper.predict(['admission office contact number ?'])[0]
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Piper.predict(['how to connect with admission office ?'])[0]
'Please visit the http://www.kfueit.edu.pk and apply for Admission test as per your interested program'
Piper.predict(['what is university contact ?'])[0]
'Contact Us Admission contact: 068-5882433 Admission Mobile: 0331-2869464 or Email Address: admissions@kfueit.edu.pk'
Piper.predict(['what is phone number of university?'])[0]
'Contact Us Admission contact: 068-5882433 Admission Mobile: 0331-2869464 or Email Address: admissions@kfueit.edu.pk'
Piper.predict(['university contact ?'])[0]
'Contact Us Admission contact: 068-5882433 Admission Mobile: 0331-2869464 or Email Address: admissions@kfueit.edu.pk'

From the results, we notice that the random forest is out perform than decision tree and support vector machine. The recognition of the random forest in *pipr* in the code.

V. CONCLUSION

The proposed study gives us the data is actual the base for the performance of the UCB so the proposed study switches the user input as for dataset appending value this value stored and the dataset increased and increased the performance of the UCB. The Machine learning algorithms DT, RF, SVM are all good perform but in the some tests the DT and SVM not response like RF so we finalize the RF is the good machine learning algorithm for training and test. The results and test show the UCB is productive and efficient approach for universities that is helpful for university departments and as well as for E-Business websites.

REFERENCES

[1] Y. W. Chandra and S. Suyanto, “Indonesian chatbot of university admission using a question answering system based on sequence-to-sequence model,” *Procedia Comput. Sci.*, vol. 157, pp. 367–374, 2019.

[2] M. Dahiya, “A tool of conversation: Chatbot,” *Int. J. Comput. Sci. Eng.*, vol. 5, no. 5, pp. 158–161, 2017.

[3] B. R. Ranoliya, N. Raghuwanshi, and S. Singh, “Chatbot for university related FAQs,” in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 1525–1530.

[4] S. A. Abdul-Kader and J. C. Woods, “Survey on chatbot design techniques in speech conversation systems,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 7, 2015.

[5] A. Følstad *et al.*, *Chatbot research and design*. Springer, 2020.

[6] A. Przegalinska, L. Ciechanowski, A. Stroz, P. Gloor, and G. Mazurek, “In bot we trust: A new methodology of chatbot performance measures,” *Bus. Horiz.*, vol. 62, no. 6, pp. 785–797, 2019.

[7] M. S. Satu and M. H. Parvez, “Review of integrated applications with aiml based chatbot,” in *2015 International Conference on Computer and Information Engineering (ICCIE)*, 2015, pp. 87–90.

[8] N. Bhartiya, N. Jangid, S. Jannu, P. Shukla, and R. Chapaneri, “Artificial Neural Network Based University Chatbot System,” in *2019 IEEE Bombay Section Signature Conference (IBSSC)*, 2019, pp. 1–6.

[9] S. Asensio-Cuesta *et al.*, “A User-Centered Chatbot (Wakamola) to Collect Linked Data in Population Networks to Support Studies of Overweight and Obesity Causes: Design and Pilot Study,” *JMIR Med. Inform.*, vol. 9, no. 4, p. e17503, Apr. 2021, doi: 10.2196/17503.

[10] J. Huang, M. Zhou, and D. Yang, “Extracting Chatbot Knowledge from Online Discussion Forums.,” in *IJCAI*, 2007, vol. 7, pp. 423–428.

[11] J. Hill, W. R. Ford, and I. G. Farreras, “Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations,” *Comput. Hum. Behav.*, vol. 49, pp. 245–250, 2015.

[12] L. Cui, S. Huang, F. Wei, C. Tan, C. Duan, and M. Zhou, “Superagent: A customer service chatbot for e-commerce websites,” in *Proceedings of ACL 2017, System Demonstrations*, 2017, pp. 97–102.

[13] B. AbuShawar and E. Atwell, “ALICE chatbot: Trials and outputs,” *Comput. Sist.*, vol. 19, no. 4, pp. 625–632, 2015.

[14] A. Barnett *et al.*, “Enacting ‘more-than-human’ care: Clients’ and counsellors’ views on the multiple affordances of chatbots in alcohol and other drug counselling,” *Int. J. Drug Policy*, p. 102910, Oct. 2020, doi: 10.1016/j.drugpo.2020.102910.

[15] L. Li, K. Y. Lee, E. Emokpae, and S.-B. Yang, “What makes you continuously use chatbot services? Evidence from chinese online travel agencies,” *Electron. Mark.*, Jan. 2021, doi: 10.1007/s12525-020-00454-z.