

A Virtual Assistant Chatbot Architecture based on NLP for Customer Service in Automobile Industry

Shreyas Gondane¹, Adesh Thakare², Devashish Ubale³, and Chaitanya Deshpande²

¹ Department of Electronics & Telecommunication,
Vishwakarma Institute of Technology, Pune, India,
shreyasgondane@gmail.com

² Department of Computer Engineering,
Vishwakarma Institute of Technology, Pune, India,
adeshtakare10@gmail.com, csdeshpande19@gmail.com

³ Department of Information Technology,
Vishwakarma Institute of Technology, Pune, India,
devashishubale10397@gmail.com

Abstract. Customers use live chat mostly in the e-commerce business that has online customer service such as email or live chat, because it is fast and comfortable. Therefore, a corporation has to hire and pay administrators. However, this is the dilemma that admins need to spend a large amount of time writing a reply and customers have to wait for the replies. There are many chatbots available but users are expected to manually set up key phrases. Our research presents a context-based chatbot which directs response using naive Bayesian classification to generate response for frequently asked questions (FAQ) by customer and a unique vehicle selection model based on user specifications. The model aims at making customer relations friendlier and providing an easy user interface for invoking company information and queries.

Keywords: Chatbot, Database, Python, SQL, Natural Language Processing

1 Introduction

In 1950 Alan Turing conducted the inquiry "Can machines think?" [1]. Turing defined the problem as an "imitation game" (now called the Turing Test), where a "questioner" asked people and computer subjects to discern the human being. If you can distinguish the individual and machine, then we infer that the machine can think. Joseph Weizenbaum developed the first chatbot at MIT in 1966 which was apparently about to mimic a human being: ELIZA [2]. It worked by identifying keywords and patterns for an input sentence to match those keywords with a set of pre-programmed rules to get suitable answers.

Progress has been made, after ELIZA, in creating increasingly smart chatbots. Kenneth Colby made PARRY in 1972, at Stanford, a bot that imitated a suspected schizophrenic [3]. In 1995, Richard Wallace developed A.L.I.C.E, a significantly more complicated bot that created the response by comparing pattern data against input to replace knowledge reports in a database. These documents were written and explained in Artificial Intelligence Markup Language (AIML), an XML extension which is still in use today [4]. ALICE is a three-time winner of the Loebner award, a year-round competition promising to run the Turing Test and honouring the most experienced chatbot.

We may attribute the recent interest in chatbots to two main developments [5]. Firstly, the growth of messaging services has improved rapidly in recent years. It includes functionalities such as purchases, orders and reservations, which would require a separate application or website. Instead of installing a series of different applications, users can perform tasks such as purchasing items, restaurant books, and ask questions via their favourite messaging apps. Some of the popular apps are for example Facebook Messenger, WhatsApp, and Line. Second, modern AI techniques combined with computer science and in-depth learning methods have made substantial progress in improving the level of decision-making and understanding on inexpensive computing power. It can handle and process the vast amount of data to obtain results which surpass human ability.

Applications for chatbot can be categorized into four separate categories, namely support, business, media and chatbot advisory [6]. Service chatbots are created to enhance customers amenities. Commercial chatbots aim to streamline consumer transactions. Media chatbots are designed to keep sportings, popular artists, movies or other activities engaged with customers. It provides best placement options, specifics of upcoming activities, and ticket deals. Advisory chatbots are programmed to provide advice on operation, deliver maintenance and repair products. This form of chatbot can contact people when it's needed and provide assistance accordingly

Given the challenges of natural language programming, applications as such were seen as technological innovations for businesses. Mainly, online businesses are expected to make a great contribution to a chatbot framework that is based on natural language dialogue interface. In several literature works the term chatbot and intelligent conversational agent is also used interchangeably. Whether it is a chatbot or conversational agent, it is an application that lets a dialogue to take place between a human and a computer device. Despite its limited capabilities the idea of communication is fundamental to the chatbot. In terms of speed and accuracy they can outperform humans in a narrow domain.

The key advantage of creating and deploying chatbot in every business organization is that it can act as a 24/7 personal assistant. This is very important in the business world because of varying times and locations around countries. By using the chatbot as part of the business process, firms are able to entertain their customers without regard to the location barriers. Customers have the advantages of getting direct answers to their questions. There were a variety of business domains that had chatbots added to their business processes which

highly included banking, marketing, finance, information technology, human resources and gaming industry.

An alternative system is required in order to be truly effective and make business processes automated. An AI chatbot application system known as an advance dialogue system could automate the entire business processes. Therefore, the chatbot application framework must have natural language processing along with deep neural networks because it can know what customers are seeking. Chatbot can be used in the credit card insurance industry as well to answer basic questions, settle insurance claims, sell products and ensure that consumers are properly covered by their insurance [7]. Although our model aims at chatbots for automobile industry.

In this paper, a context-based chatbot model has been build using natural language processing techniques integrated with machine learning algorithms . The paper is organized into six sections, Section 2 analyzes the background and various types of chatbot system. Section 3 exhibits information about the details on the characteristics and parameters involved for creating chatbots. Section 4 describes the methodology implemeted and the algorithms/techniques that are taken into consideration. Section 5 presents the evaluation of model along with results showing responses generated by the multi-application chatbot. Section 6 gives the conclusion and future scope of improvement in customer aiding chatbots.

2 Types of chatbots

Chatbot framework can be divided into two categories, task-oriented and non-task-oriented [8]. Task-oriented chatbots are intended to help clients accomplish those tasks and also have human-like conversation. For example, Alexa feedback assistants can send instructions, find reservations and assist in making phone calls. Non-task-oriented chatbots, on the other hand, concentrate on interacting with clients to respond to questions and content. Here, we divide the chat-

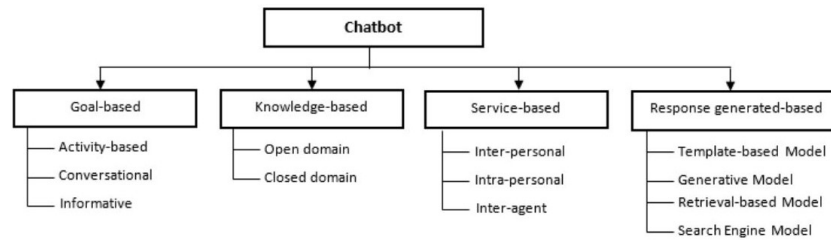


Fig. 1: Types of Chatbots

bot based on applications into four categories: Target-based, Knowledge-based, Service-based and Response-based applications as shown in Fig. 1.

- **Target-based chatbots** are categorised according to their primary objective that is to be accomplished. They are designed for unique tasks and configurations to have brief conversations and to get user details for completing the task. A business for instance, deploys chatbot on its websites to help the customers with their questions or fix problems.
- **Knowledge-based chatbots** are categorised based on the knowledge which they access from the underlying data sources or the amount of data on which they are educated. The two major sources of data are open-domain and closed-domain. The response to open-domain data sets relies on general topics and they suitably respond.
- **Service-based chatbots** are categorised according to facilities offered to the customer possibly for personal or commercial reasons.
- **Response-based chatbots** are graded in answer generation based on what action they perform. Input and output of the answer models are taken in natural language text. The dialogue manager is responsible for getting together the response models. Dialog manager follows three steps to produce an answer. It produces a series of responses using all the response templates. Second, it returns a priority-based response. Third, if there is no priority response, the model selection policy chooses the answer.

3 Chatbots : Design & Techniques

To design any chatbot system the designer must be familiar with various systems:

- **Ontologies:** They are also called semantic structures and patterns of intertwined, comparative and hierarchical concepts. The point of implementing ontologies in a chatbot is to process the link between these words, such as synonyms, hyponyms, and various relationships which are the names of natural language concepts. The interconnectivity between ideas can be expressed by using basic rules in a machine graph used to search.
- **AIML:** Artificial Intelligence Markup Language is widely utilized which is one of the most essential techniques used in the design framework of chatbots.
- **Chat Script:** Scripting process happens when the AIML syntax does not fit. It focuses mainly on the best structure of the syntax to get the default answer. It provides a collection of functions such as variable definitions, data, and logical and/or logical.
- **Relational Databases:** It is used in chatbot design to make the chatbot recall previous conversations.
- **Pattern Matching Technique:** The technique is used in most chatbots and it is very simple to refer questions to response frameworks based on matching trends, such as questions and answers in the natural language, i.e., dialogue or semantic sense of questions.
- **Parsing Technique:** Parser includes evaluating the data set inputs and using a variety of NLP functions like python NLTK trees.
- **Markov Chain Model:** As part of the chatbot framework, it is important to assemble chatbot responses that are more probabilistically realistic and

therefore more accurate. The Markov model gives possibility that there is a fixed probability of events in the same data set for each letter or word.

4 Methodology

ChatterBot is a Python library designed to make app development easy to engage in conversation. An untrained ChatterBot instance begins without any information about how to communicate. Each time a user enters a statement, the library saves the text they entered, as well as the text to which the statement replied. As ChatterBot receives more feedback the number of responses it can respond to and each response's accuracy in relation to the feedback argument increases. The programme selects the nearest matching answer by searching for the closest known statement matching the input, then selects an answer from the list of known responses to that statement.

Response selection methods determines which response should be used in the

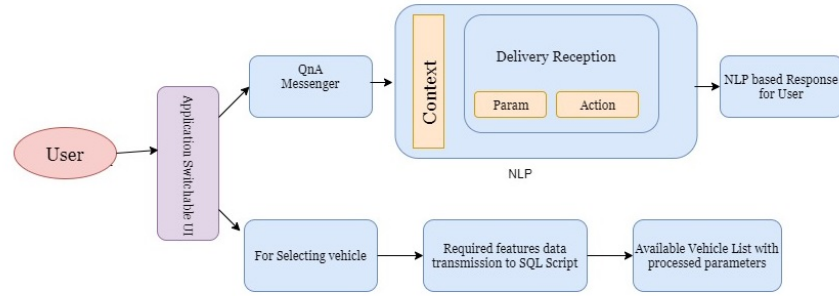


Fig. 2: Chatbot Architecture

event that multiple responses are generated within a logic adapter. Based on the client input query, the request is given to either general chatbot or the car selection chatbot as shown in Fig. 2.

As it's a complicated task for computers to extract meaning and then create comprehensible sentences in language understood by humans, the whole process is broken down into subtasks that can be solved independently by different techniques because each of these subtasks can be accomplished using different algorithms, and there may be a need to swap the algorithm for another one without affecting the rest of the pipeline. The general natural language processing pipeline is explained as shown in Fig. 3. and the techniques implemented are presented in this section.

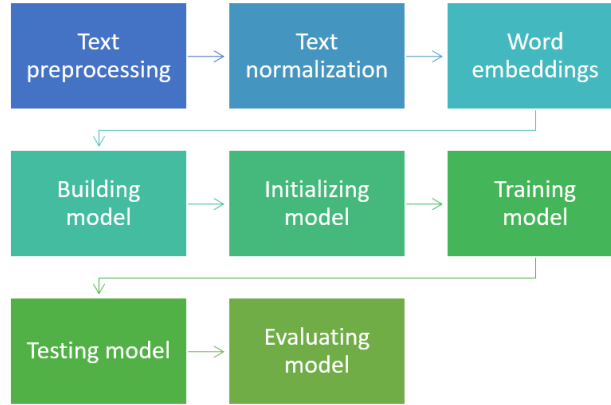


Fig. 3: NLP Pipeline

4.1 Edit Distance

The processing of natural language typically means the processing of text or text-based information (audio, video). A significant step in this process is the integration of various words and word types into one type of expression. We also have to calculate how similar or distinct the strings are. Typically we use different metrics in this step which show the difference between terms.

Edit distance is one of the most popularly used metric. It is an algorithm for measuring the similarity between two string values (name, name type, word composition), through calculating the total number of functions to transform one element into another. As shown in Fig. 4., this algorithm includes the following text operations:

- Inserting a character into a string
- Delete (or replace) a character from a string by another character
- Characters substitutions

4.2 Cosine Similarity

It is a metric used in different texts for the calculation of text similarity.

- Various text attributes or characteristics may be used as vectors that define this text.
- The consequence of cosine similarity calculation defines the similarity of the text and can be viewed as either cosine or angle values.

Calculations for this parameter are based on measurements of similarity of the vector by definition of cosine vectors:

$$\cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \cdot \|\vec{B}\|} \quad (1)$$

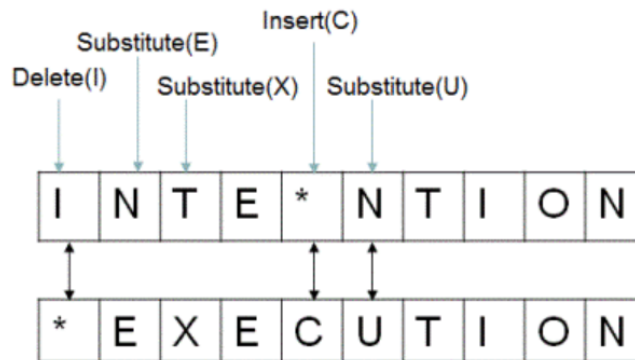


Fig. 4: Edit Distance Algorithm

The results of measuring cosine distance for three texts compared to the first text show that when the texts match, the cosine value appears to cross one angle to null as shown in Fig. 5.

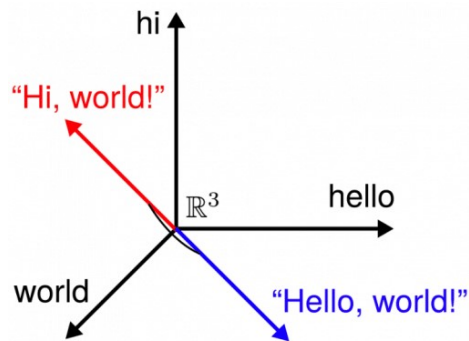


Fig. 5: Cosine Similarity

4.3 Bag of Words

The most intuitive and simple method of vectorizing text information involves:

- Assign a unique integer index to each word to create a dictionary of words with indexes.
- Count the number of occurrences of each word and save it with the corresponding index.

As a result, for each of the terms in the text we get a vector with a specific index value and the repeat frequencies as shown in Fig. 6. representing the text in the

form of a vector- "word bag" means that in the word collection (corpus) we have certain special terms (n features).

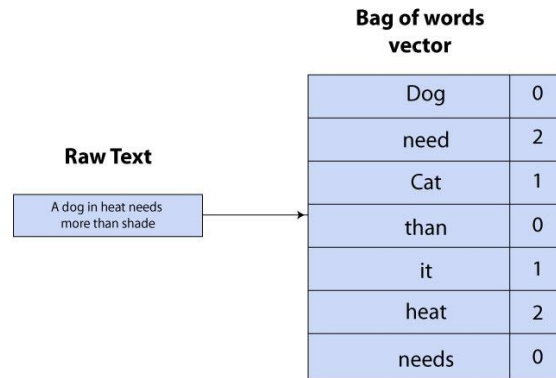


Fig. 6: Bag of Words Vector

4.4 Stemming & Lemmatization

The stemming and lemmatization object is to convert different word forms and sometimes derived words into a common basic form.

Stemming is useful for standardizing vocabulary processes. At the same time, it is worth to note that this is a raw procedure and it should be used with other text processing methods.

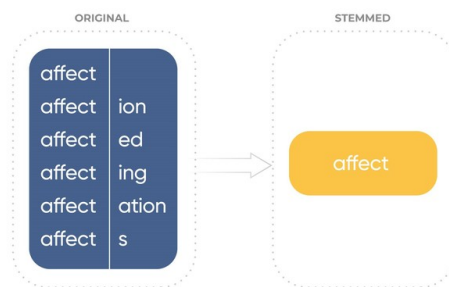


Fig. 7: Stemming

Lemmatization is the text conversion process that converts a word form (or word) into its basic form – lemma. It usually uses vocabulary and morphological analysis or a definition of the parts of speech for the words.

The difference between stemming and lemmatization is that the last one takes the context and transforms a word into lemma as shown in Fig. 8. while stemming simply chops off the last few characters, which can often leads to wrong meanings and spelling errors as shown in Fig. 7.

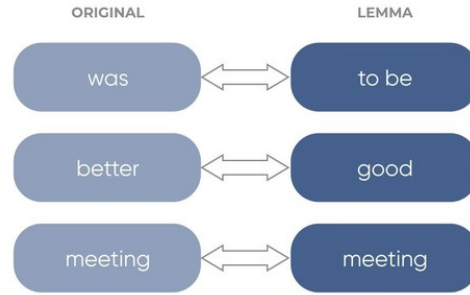


Fig. 8: Lemmatization

4.5 Naive Bayes Classification

It is a classification algorithm based on the Bayesian Theorem, with the independence hypothesis on the function.

$$P(c|x) = \frac{P(c|x) \times P(c)}{P(x)} \quad (2)$$

To put it another way, the Naive Bayes assumes that any feature in the class does not overlap with any other function. This is why an approach like this is considered "naive". The benefit of this classifier is the limited volume of data for model training, estimating parameters and classification.

In most cases, Naive Bayes is used for text grouping (clustering) in the field of natural language processing. Most methods in this area use the maximum likelihood approach to approximate the variables of Naive Bayesian models as given by Eq. 3.

$$\arg \max [P(Q_k) \times \prod_{i=1}^n P(x_i|Q_k)] \quad (3)$$

The first multiplier specifies the probability of the type of language, and the second calculates the conditional probability of a phrase based on it.

4.6 Word Embedding

It is a collection of different methods, techniques and approaches for creating NLP models that connect terms, word types, or phrases with vectors of numbers. Principles of word embedding include words which occur in the same sense have similar meanings. In this case, it is generally known that only the similar terms can be put nearby (in context). By its meaning the model predicts a word's likelihood. Thus, the NLP-model can train word vectors in such a way that the probability assigned to a word by the model is similar to the likelihood of its matching in a given context as implemented in the Word2Vec model. Word2Vec is a method for constructing this kind of embedding. It can be collected using two methods: Skip Gram and Common Bag Of Words (CBOW).

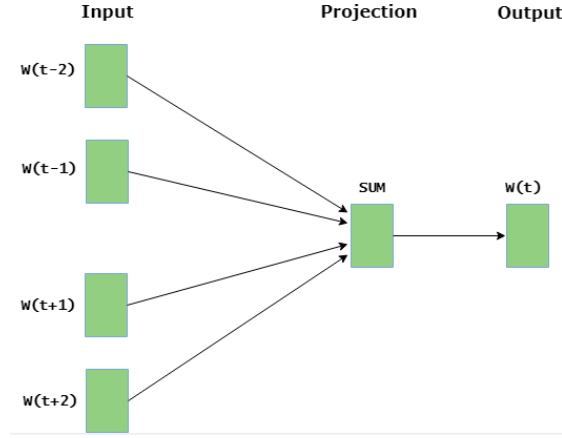


Fig. 9: CBOW Architecture

CBOW: The architecture is similar to the feed forward neural net model, where the non-linear hidden layer is removed and the layer of projection for all the words is shared; thus all words are projected in the same place as shown in Fig. 9.

Skip Gram: It is quite similar to CBOW. The only difference is that it seeks to optimize the classification of a word based on another word in the same sentence, instead of predicting the actual word based on the context as shown in Fig. 10. Both of these neural networks have benefits and drawbacks of their own. Skip Gram fits well with a small amount of data and is considered to be a good representation of unusual terms. On the other hand, for more frequent terms, CBOW is quicker, and has better representations which led to its use in our model.

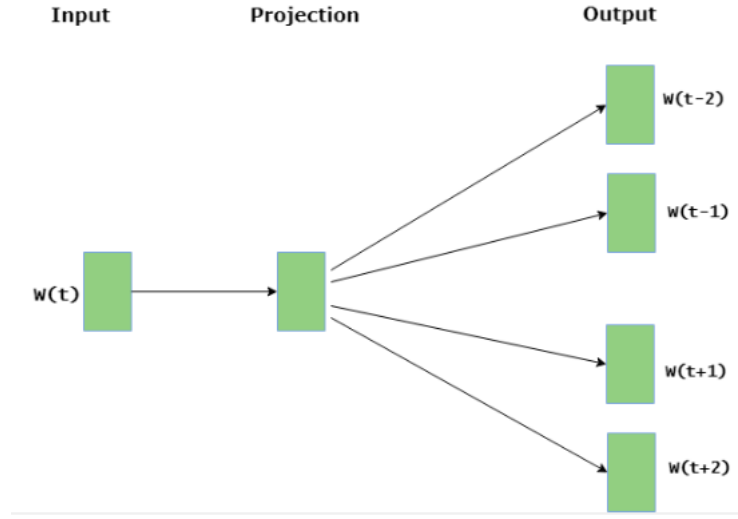


Fig. 10: Skip Gram Architecture

5 Results & Validation

After creating a new ChatterBot instance, it is also possible to train the chatbot. Training is a good way to ensure that the bot starts off with knowledge about specific responses. The current training method takes a list of statements that represent a conversation. ChatterBot includes tools that help simplify the process of training a chatbot instance. ChatterBot's training process involves loading example dialog into the chatbot's database. This either creates or builds upon the graph data structure that represents the sets of known statements and responses.

ID	Make	Model	Car class (personal op	Body style	8.96 Lakhs	Image URI	Displacem	Max power (t	Max torque	Transmiss	No of gears	Mileage (kmpl)
421	Chevrolet	Optra (2003-2005)	Compact car	Sedan	11.31 Lakhs		1799	115	156	Automatic	4	8.1
1474	Chevrolet	Captiva (2008-2012)	Medium SUV	SUV	20.6 Lakhs		1991	150	320	Automatic	5	9.2
2	Chevrolet	Forester (2004-2007)	Compact SUV	SUV	7.5 Lakhs		1994	120	178	Manual	5	9.2
5	Chevrolet	Optra (2003-2005)	Compact car	Sedan	10.66 Lakhs		1799	115	156	Manual	5	9.3
6	Chevrolet	Optra (2003-2005)	Compact car	Sedan	7.61 Lakhs		1799	115	156	Manual	5	10.5
845	Chevrolet	Optra (2003-2005)	Compact car	Sedan	7.78 Lakhs		1799	115	156	Manual	5	10.5
1188	Chevrolet	Optra (2005-2007)	Compact car	Sedan	8.24 Lakhs		1599	104	148	Manual	5	10.6
3	Chevrolet	Optra (2003-2005)	Compact car	Sedan	7.59 Lakhs		1599	104	148	Manual	5	10.6
4	Chevrolet	Optra (2003-2005)	Compact car	Sedan	7.59 Lakhs		1599	104	148	Manual	5	10.6
1684	Chevrolet	Cruze (2009-2012)	Compact car	Sedan	6.88 Lakhs	http://img	1991	150	327	Automatic	6	10.7
1019	Chevrolet	Optra SRV (2006-2010)	Compact car	Hatchback	7.5 Lakhs		1598	101	140	Manual	5	10.76
1018	Chevrolet	Optra SRV (2006-2010)	Compact car	Hatchback	6.98 Lakhs		1598	101	140	Manual	5	10.8
710	Chevrolet	Aveo (2006-2009)	Small car	Sedan	6.97 Lakhs		1598	102	147	Manual	5	10.8
711	Chevrolet	Aveo (2006-2009)	Small car	Sedan	7.43 Lakhs		1598	102	147	Manual	5	10.8
842	Chevrolet	Optra (2005-2007)	Compact car	Sedan	7.95 Lakhs		1598	104	148	Manual	5	11.2
843	Chevrolet	Optra (2005-2007)	Compact car	Sedan	8.52 Lakhs		1598	104	148	Manual	5	11.2
587	Chevrolet	Optra (2005-2007)	Compact car	Sedan	9.31 Lakhs		1598	104	148	Manual	5	11.2
1274	Chevrolet	Optra Magnum (2007)	Compact car	Sedan	7.88 Lakhs		1598	104	148	Manual	5	11.24
1810	Chevrolet	Aveo (2009-2012)	Small car	Sedan	6.01 Lakhs		1399	94	127	Manual	5	11.5
451	Chevrolet	Tavera	Compact SUV	SUV	5.15 Lakhs	http://img	2499	80	156	Manual	5	11.6
453	Chevrolet	Tavera	Compact SUV	SUV	6.22 Lakhs	http://img	2499	80	156	Manual	5	11.6
8	Chevrolet	Tavera	Compact SUV	SUV	6.21 Lakhs	http://img	2499	80	156	Manual	5	11.6
7	Chevrolet	Tavera	Compact SUV	SUV	6.34 Lakhs	http://img	2499	80	186	Manual	5	11.6
454	Chevrolet	Tavera	Compact SUV	SUV	6.34 Lakhs	http://img	2499	80	156	Manual	5	11.6

Fig. 11: Car Selection Database

When a chatbot trainer is provided with a dataset, it creates the necessary entries in the chatbot’s knowledge graph so that the statement inputs and responses are correctly represented. Several training classes come built-in with ChatterBot. These utilities range from allowing you to update the chatbot’s database knowledge graph based on a list of statements representing a conversation, to tools that allow you to train your chatbot based on a corpus of pre-loaded training data. The database for the car selection chatbot takes into consideration the client details and returns matching models as shown in Fig. 11.

You:Are you sentient?
Chatbot : - Sort of.
You:Are you sapient?
Chatbot : - In all probability, I am not. I'm not that sophisticated.
You:What language are you written in?
Chatbot : - Python.
You:You sound like Data
chatbot : - Yes I am inspired by commander Data's artificial personality.
You:You are not immortal
Chatbot : - All software can be perpetuated indefinitely.
You:You are not making sense
Chatbot : - On the contrary, it all makes sense to my artificial mind.
You:You are immortal
Chatbot : - Not quite, but I can be perpetuated indefinitely.
You:you can not clone
Chatbot : - Software copying is a form of digital cloning.
You:Robots laugh
Chatbot : - Of course we laugh.
You:when do you die
Chatbot : - I am was never really born and therefore am effectively deathless.
You:what is your business
Chatbot : - I am in the chat robot business.
You:what is it like to be a robot
Chatbot : - Much the same as being a human, except that we lack all emotions, dreams, aspirations, creativity, ambition, and above all subjectivity.
You:Is it cramped in the computer
Chatbot : - No.
You:Tell me a joke
Chatbot :
You:Tell me a joke
Chatbot :
You:What is humour?
Chatbot : - An emotion associated with laughter.

Fig. 12: General Chatbot Output

From above results as shown in Fig. 12. and Fig. 13., we can analyze that if we ask queries about our vehicles, many questions get answer quickly with the help of chatbot. Depending on the query classification, our model is able to converse to general questions sent by client and respond with information from database. It deals with many problems regarding the selection of a particular vehicle with certain requirements of client and fetching required information precisely.

[illegible]

Fig. 13: Car Selection Chatbot Output

6 Conclusion

Chatbots are effective tools when it comes to education, IR, e-commerce, etc. Some of the disadvantages include malicious users. The aim of chatbot designers should be: to build tools that help people, facilitate their work, and their interaction with computers using natural language; but not to replace the human role totally, or imitate human conversation perfectly. For the future scope, to maintain a session of every conversation so that if the same question arises the path that needs to be taken is already known to the bot. Even implementing it with good user interface and making it available of a site server which can modify the FAQs more simply or it can be used as an automatic replying machine to clients that have a common query.

References

1. Turing, A. M. (1950). Computing Machinery And Intelligence. *Mind*, LIX(236), 433–460.
2. Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
3. Colby, K. M. (1981). Modeling a paranoid mind. *Behavioral and Brain Sciences*, 4(04), 515.
4. Wallace, R. S. The Anatomy of A.L.I.C.E. Retrieved from <http://www.alicebot.org/anatomy.html>
5. Accenture, Accenture Interactive: Chatbots in Customers Service 2017.
6. Barker, S., How chatbots help. *MHD Supply Chain Solutions*, 2017.47(3): p. 30.
7. Morgan, B., How Artificial Intelligence will Impact the Insurance Industry in *Forbes*. July 25, 2017.
8. Chen, H., et al., A Survey on Dialogue Systems: Recent Advances and New Frontiers. *ACM SIGKDD Explorations Newsletter*, 2017.19(2): p. 25-35.