

# **CREATE A CHATBOT IN PYTHON**

TEAM MEMBER

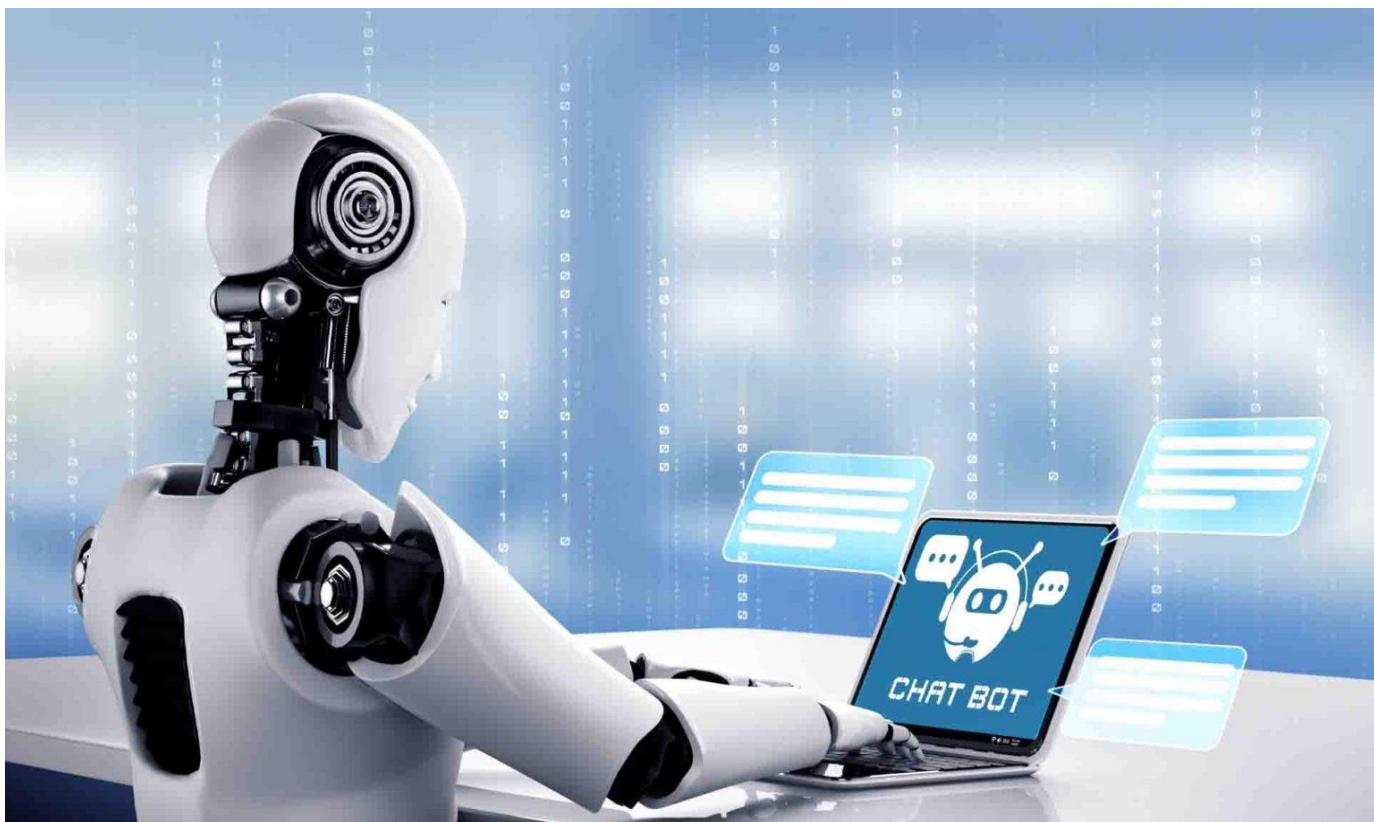
810621104012:kaviyarasan G

## **Phase-4 Submission document**

**Project Tittle: Create a chatbot in python**

**Topic:** Continue building the create a chatbot in python model by feature engineering, model training and the evaluation.

### **Phase 4: Development part2**



# CREATE A CHATBOT IN PYTHON

## INTRODUCTION:

- ❖ A chatbot is a software program for simulating intelligent conversations with human using rules or artificial intelligence. Users interact with the chatbot via conversational interface through written or spoken text. Chatbots can live in messaging platforms like **Slack, Facebook Messenger and Telegram** and serve many purposes – **ordering products**, knowing about weather and managing your finance among other things.
- ❖ Chatbot are not a recent development. They are simulations that can understand human language, process it, and interact back with humans while performing specific tasks. For example, a chatbot can be employed as a helpdesk executive. Joseph Weizenbaum created the first Chabot in 1966, named Eliza. It all started when Alan Turing published an article named “Computer Machinery and Intelligence” and raised an intriguing question, “Can machines think?” ever since, we have seen multiple chatbots surpassing their predecessors to be more naturally conversant and technologically advanced. These advancements have led us to an era where conversations with chatbots have become as normal and natural as with another human. Before looking into the AI chatbot, learn the [foundations of artificial intelligence](#).
- ❖ Chatbot has demonstrated the power and convenience of conversational interfaces. As of now, language is emerging as the interface. According to Gartner, 30% of browsing done by users will turn out to be screen less in 2020.

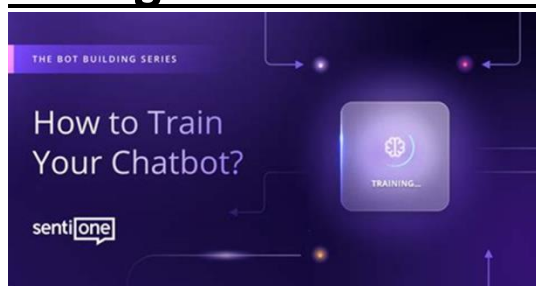
# DATASETS:

- An effective chatbot requires a massive amount of training data in order to quickly solve user inquiries without human intervention. However, the primary bottleneck in chatbot development is obtaining realistic, task-oriented dialog data to train these machine learning-based systems.
- We've put together the ultimate list of the best conversational datasets to train a chatbot, broken down into question-answer data, customer support data, dialogue data and multilingual data.

## Question-Answer Datasets for Chatbot Training:

- **Question-Answer Dataset:** This corpus includes Wikipedia articles, manually-generated factoid questions from them, and manually-generated answers to these questions, for use in academic research.
- **The WikiQA Corpus:** A publicly available set of question and sentence pairs, collected and annotated for research on open-domain question answering. In order to reflect the true information need of general users, they used Bing query logs as the question source. Each question is linked to a Wikipedia page that potentially has the answer.
- **TREC QA Collection:** TREC has had a question answering track since 1999. In each track, the task was defined such that the systems were to retrieve small snippets of text that contained an answer for open-domain, closed-class questions.

## Dialogue Datasets for Chatbot Training:



**Semantic Web Interest Group IRC Chat Logs:** This automatically generated IRC chat log is available in RDF, back to 2004, on a daily basis, including time stamps and nicknames.

**Cornell Movie-Dialogs Corpus:** This corpus contains a large metadata-rich collection of fictional conversations extracted from raw movie scripts: 220,579 conversational exchanges between 10,292 pairs of movie characters involving 9,035 characters from 617 movies.

**ConvAI2 Dataset:** The dataset contains more than 2000 dialogues for a Personal Chat competition, where human evaluators recruited via the crowdsourcing platform Yandex.Toloka chatted with bots submitted by teams.

**Santa Barbara Corpus of Spoken American English:** This dataset includes approximately 249,000 words of transcription, audio, and timestamps at the level of individual intonation units.

**The NPS Chat Corpus:** This corpus consists of 10,567 posts out of approximately 500,000 posts gathered from various online chat services in accordance with their terms of service.

**Maluba Goal-Oriented Dialogue:** Open dialogue dataset where the conversation aims at accomplishing a task or taking a decision – specifically, finding flights and a hotel. The dataset contains complex conversations and decision-making covering 250+ hotels, flights, and destinations.

**Multi-Domain Wizard-of-Oz dataset (MultiWOZ):** A fully-labeled collection of written conversations spanning over multiple domains and topics. The dataset contains 10k dialogues, and is at least one order of magnitude larger than all previous annotated task-oriented corpora.

### **Features engineering:**

They can be used to enhance customer communication and boost the level of engagement. Here are some key features that make chatbots successful<sup>1</sup>:

1. **Deliver contextual responses:** Chatbots need to have the ability to understand the context so that customers feel like talking to a real person. By leveraging

the advancements in natural language processing (NLP), bots can be made to understand context without asking validating questions.

2. **Allow human handover:** Bots need to be smart to understand the sense of urgency and complexity of a conversation. Even when a chatbot template for online order fails to understand the query, it can still intelligently hand over the conversations to human support.
3. **Great UI/UX:** AI chatbots should not be complex or hard to use else they won't be able to make conversations interactive. The design has to be simple and intuitive so that users find it easy to use them for answers.
4. **Well-trained with FAQs:** Chatbot benefits are many when they are trained. Regular training can help chatbots become powerful and enable them to smoothly handle questions and interactions.
5. **Offer personalized support:** Bots that are designed using AI and machine learning can easily comprehend user conversations and respond in real-time. A customer support chatbot template can adjust the tone and language to give personalized experience.
6. **visual flow builder:** Building a chatbot should be easy and hassle-free. The visual flow builder makes it possible and helps in creating a conversational flow.
7. **Omnichannel messaging support:** Brands that leverage omnichannel strategy achieve 91% greater year-over-year customer retention rates compared to those who don't <sup>1</sup>.
8. **Live chat handover:** When a bot is unable to resolve an issue, it should be able to hand over the conversation seamlessly to a human agent.
9. **Sentiment analysis:** Chatbots can analyze customer sentiment by analyzing their text, tone, and other factors, which can help businesses improve their products or services.
10. **Chatbot Marketing:** Chatbots can be used for marketing purposes such as lead generation, customer engagement, etc.
11. **Chatbot Analytics:** Chatbot analytics provides insights into how customers interact with your bot, which can help you improve your bot's performance.
12. **Chatbot widget customization:** Customizing your chatbot widget can help you match your brand's look and feel.

### **MODEL TRAINING:**

Training a chatbot involves several steps. Here are some tips to help you get started:

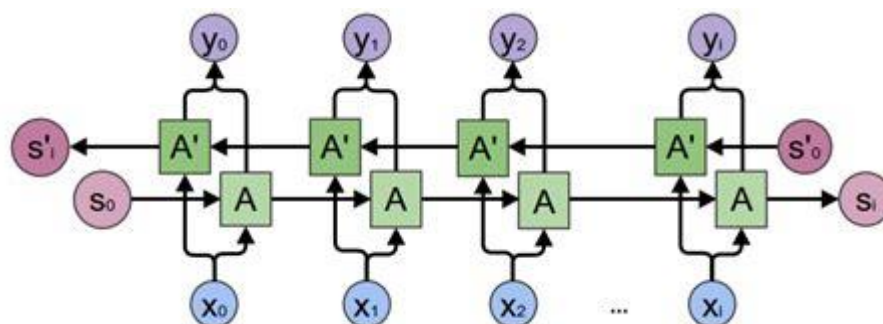
1. **Determine the chatbot use cases:** Identify the specific tasks that your chatbot will perform.

2. **Define user intent:** Understand the user's intention behind their message.
3. **Analyze conversation history:** Analyze past conversations to identify patterns and common questions.
4. **Generate variations of the user query:** Create different versions of the same question to improve the chatbot's ability to understand and respond.
5. **Ensure keywords match the intent:** Use keywords that match the user's intention to improve accuracy.
6. **Teach your team members:** Train your team members to handle complex queries that the chatbot may not be able to handle.
7. **Give your chatbot a personality:** Create a personality for your chatbot that aligns with your brand.
8. **Add media and GIFs:** Use media and GIFs to make your chatbot more engaging.
9. **Improve the chatbot frequently:** Continuously improve your chatbot by analyzing user feedback and updating its responses.

There are several ways to train a chatbot, including using sentiment analysis, Python coding language, and Named Entity Recognition (NER) <sup>1</sup>. However, if you're not a professional developer or tech-savvy, you might want to consider using a platform that provides artificial intelligence (AI) and natural language processing (NLP) bots <sup>1</sup>.



1



2

## **CHATBOT TABLE EXAMPLE:**

Here are some resources that might help you with chatbot tables:

1. **Chatbot Scripts: A Step By Step Guide (With Examples & Templates):** This article provides a comprehensive guide to chatbot scripting, including how to get started with scripting your first bot. It also shows you how to create your first bot using Hubspot's Chatflows, a free, intuitive tool that integrates seamlessly with Hubspot CRM. The article also includes a list of programs to consider for writing, editing and sharing your bot scripts before implementation<sup>1</sup>.
2. **25 Chatbot Platforms: A Comparative Table:** This table compares 25 of the best-known platforms for building chatbots, such as IBM Watson, Microsoft Bot Framework, LUIS, Wit.ai, Api.ai, Chatfuel, and others. The table provides an overview of the most popular bot platforms and can help you select a chatbot platform that suits your needs<sup>2</sup>.
3. **How to Use ChatGPT to Present Data in Table:** This article explains how to create a ChatGPT table. You can start by providing ChatGPT with the data you want to include in your table. You can do this by entering the data into the chatbot's input field or asking the chatbot to provide data. Next, provide ChatGPT with the format you want your table to take<sup>3</sup>.
4. **Chatbot (bot) table/entity reference (Microsoft Dataverse):** This resource provides information about the systemuser\_bot\_publishedby one-to-many relationship for the systemuser table/entity. It also provides relationship details where the bot table is the first table in the relationship<sup>4</sup>.
5. **Create chatbots from Power Apps:** This resource explains how app makers can create and manage their chatbots directly from Power Apps. You can create and manage your chatbots using the Chatbots option in the left navigation pane. To create a chatbot, select Chatbots, then Create, and then select New chatbot<sup>5</sup>.

Are you trying to get your degree from a foreign university?

Just now

Yes

Just now



That's a great decision

Is this your first time applying?

☒ Yes

☐ No

Just now

1

Create a table showing the top 5 countries by population, including their population and land area in square kilometers.



Sure, here's a table showing the top 5 countries by population, along with their population and land area in square kilometers:

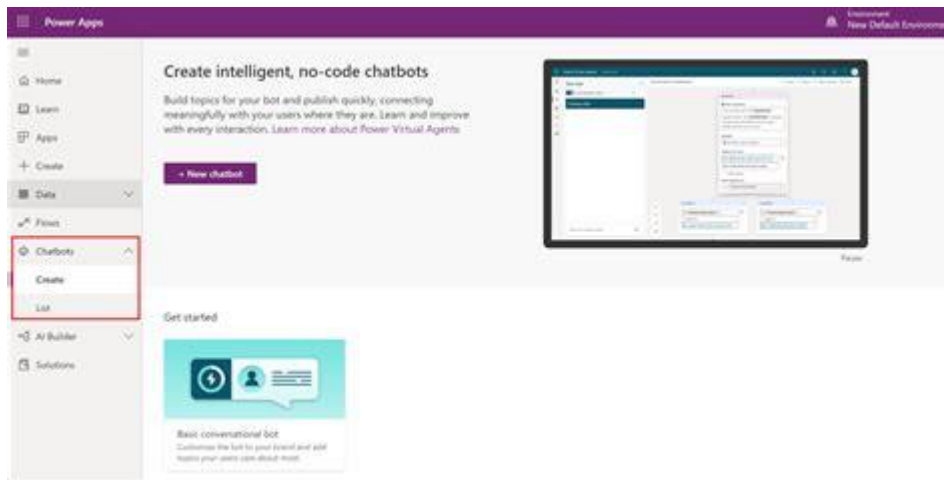


Country	Population	Land Area (sq km)
China	1,397,715,000	9,596,961
India	1,366,417,754	2,973,190
United States	332,915,073	9,147,593
Indonesia	276,361,783	1,811,570
Pakistan	225,199,937	770,880

Note: The population and land area data used in this table is based on the latest estimates available as of my knowledge cutoff of September 2021.

3





5

## **EVALUATION :**

CHATBOT EVALUATION RESEARCH From the FIVE papers focusing on evaluation methods, two of them [47,61] consisted of meta-evaluations of other papers.

### **First meta-evaluation:**

In this paper published in 2017, Radzi will and Benton [61] worked on reviewing evaluations methods from 42 papers and articles published between 1995 and 2017 about chatbots. One of their main discoveries is that the methods are generally aligned with the ISO9214 concept of usability.

### **Second meta-evaluation:**

Maroengsit's et al. [47] published their paper in 2019. They went through 30 contributions representing chatbots from distinct do-mains (such as e-commerce, health or open-domain), 17 of which had evaluation methods. They focused on evaluation methods and also ended up with three main categories.

### **Perspectives for conversational:**

Al Jadeja and Varia [37] propose in their 2017 paper four perspective on evaluation which should be used individually or in hybrid models depending on the objectives of the chatbot. The User Perspective, which measures user satisfaction, usability, etc. The Information Retrieval Perspective, which measures, for example, the accuracy and speed of the answer to the user's query.

### **Perspectives for Hcl:**

In their 2018 paper, Jain et al. [38] focus on an HCI (human-computer interaction) perspective to evaluate design implications for chatbot designers. The participants were asked to interact briefly each day for three days with chatbots without any specific task, and the team measured quantitative data (such as total interaction time, message count and number of interactive elements) during the discussions. At the end of the three days, they asked users to answer some questions about the chatbots in semi-structured interviews, from which they spotlighted four elements as follows: functionality, conversational intelligence, chatbot personality and chatbot interface.

### **Perspectives for conversational experience:**

The Alexa Prize, where users initiated over a million conversations with chatbots participating in the competition, then rated them. In this 2018 paper, they strongly reject the Turing Test as an appropriate way of evaluating chatbots, notably because the goal of evaluating a chatbot is evaluating conversational experience, which a chatbot may be able to provide without passing itself as a human. Instead, they propose the six following metrics: conversational user experience, engagement, coherence, domain coverage, conversational depth and topical diversity/conversational breadth.

### **EXAMPLE PROGRAM:**

```
# Import "chatbot" from
# chatterbot package.
from chatterbot import ChatBot

# In order to train our bot, we have
# to import a trainer package
# "ChatterBotCorpusTrainer"
from chatterbot.trainers import ChatterBotCorpusTrainer

# Give a name to the chatbot "corona bot"
# and assign a trainer component.
chatbot=ChatBot('corona bot')

# Create a new trainer for the chatbot
trainer = ChatterBotCorpusTrainer(chatbot)

# Now let us train our bot with multiple corpus
```

```
trainer.train("chatterbot.corpus.english.greetings",  
             "chatterbot.corpus.english.conversations" )
```

```
response = chatbot.get_response('What is your Number')  
print(response)
```

```
response = chatbot.get_response('Who are you?')  
print(response)
```

### OUTPUT:

```
Training greetings.yml: [#####] 100%  
Training conversations.yml: [#####] 33%
```

```
[nltk_data] Downloading package stopwords to /home/nikhil/nltk_data...  
[nltk_data] Package stopwords is already up-to-date!  
[nltk_data] Downloading package averaged_perceptron_tagger to  
[nltk_data] /home/nikhil/nltk_data...  
[nltk_data] Package averaged_perceptron_tagger is already up-to-  
[nltk_data] date!
```

```
Training conversations.yml: [#####] 100%  
I don't have any number  
I am just an artificial intelligence.
```

## CHATBOT LIBRARIES:

### spaCy

spaCy is an open-source library for Natural Language Processing (NLP) in Python language. The library is designed specifically for developers to build interactive NLP applications, which can process and 'understand' large volumes of text. spaCy can also be utilized to create information extraction or NLU systems, as well as pre-process text for deep learning.

Some features of spaCy are mentioned below:

**Tokenization:** Tokenization helps in segmenting text into words, punctuations, etc.

**Part-of-speech (POS) Tagging:** This feature helps assign word types to tokens, like a verb.

**Sentence Boundary Detection (SBD):** This feature helps find and segment individual sentences in a text.

**Similarity:**

This feature helps compare words, text spans, etc. and matches the similarities between them.

### **Text Classification:**

It helps assign categories or labels to a document. Rule-based Matching: This feature helps search and find sequences of tokens based on their texts and linguistic annotations.

## **2| ChatterBot**

ChatterBot is a Python library designed to make it easy to create software that can engage in conversation. It uses a selection of machine learning algorithms to produce different types of responses. This helps create chatbots and automate conversations with users.

### **Installing ChatterBot**

To install this library, a user is recommended to use pip. From PyPi using pip run the following command in your terminal: `pip install chatterbot`

## **3| NLTK**

Natural Language Tool Kit – or NLTK – is an open-source suite of libraries and programs for building programs in Python language. The toolkit provides easy-to-use interfaces with several corpora and lexical resources, such as Word Net, along with a suite of text processing libraries for tokenization, stemming, tagging, parsing, classification and semantic reasoning.

### **Installing NLTK**

NLTK requires Python 3.5, 3.6, 3.7, or 3.8 versions. After installing the NLTK package, you need to install the necessary datasets/models for specific functions to work.

## **4| TextBlob**

TextBlob is a library for processing textual data which is written in Python language. The library provides a simple API for working into common NLP tasks, such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and more. This library runs on Python versions 2 and 3, and it focuses on providing access to common text-processing operations through a familiar interface.

## **Installing TextBlob**

```
pip install -U textblob
```

```
python -m textblob.download_corpora
```

## **5| DeepPavlov**

DeepPavlov is an open-source conversational AI library built on Tensor Flow and Keras. It has comprehensive and flexible tools that let developers and NLP researchers create production-ready conversational skills and complex multi-skill conversational assistants.

### **Installing DeepPavlov**

The library supports Python 3.6 and 3.7. To install the package, type: `pip install deep pavlov`

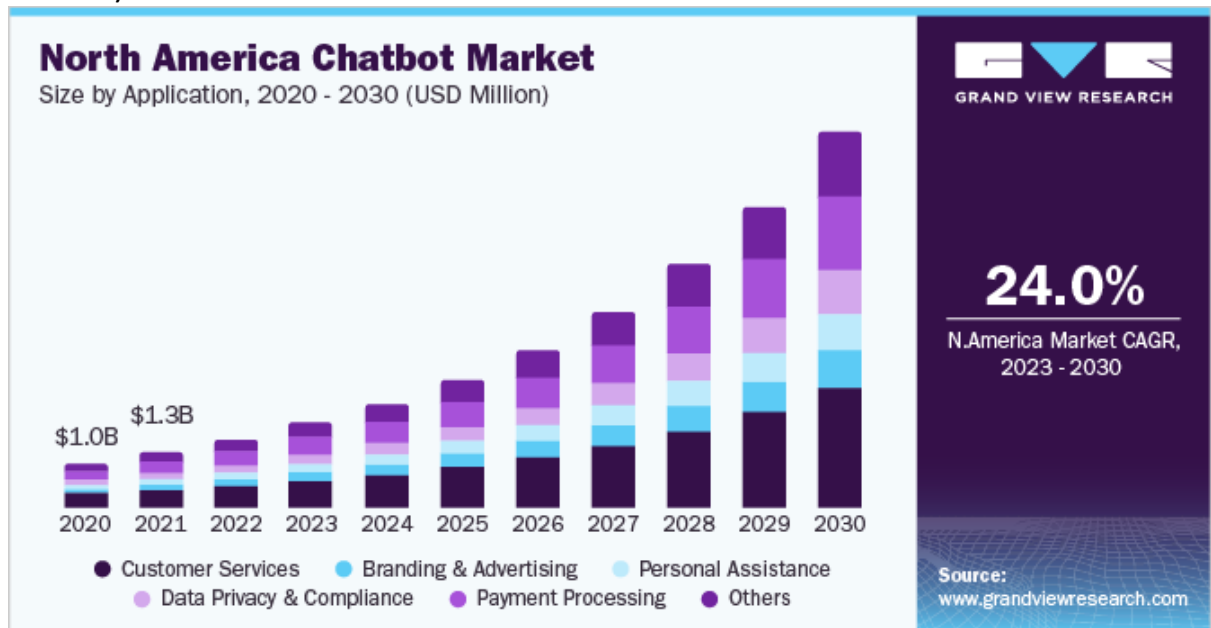
## **6| PyNLPI**

PyNLPI, pronounced as 'pineapple', is a Python library for NLP. It can be used for basic tasks, such as the extraction of n-grams and frequency lists, and to build a simple language model. The library is divided into several packages and modules. It works on Python 2.7, as well as Python 3.

## **GLOBAL MARKET IN CHATBOT:**

- The global chatbot market size was estimated at USD 5,132.8 million in 2022 and is expected to expand at a compound annual growth rate (CAGR) of 23.3% from 2023 to 2030. The rising acceptance of customer service activities among businesses for the reduction in operating costs is anticipated to ascend the market growth. A chatbot is a collaborative application programmed using artificial intelligence technology and specific set of rules. It is intended for the interaction with humans over textual conversation process and is incorporated with various messaging facilities, thus supporting users in various sectors. Various innovations carried out in artificial intelligence, and machine learning technologies are expected to enhance the features of chatbot, which, in turn, would drive demand for the market.

- North America Chatbot market size by application, 2020 - 2030 (USD Million)



- To learn more about this report, request a free sample copy
- The chatbot market has shown a significant growth in the last few years. The need to provide 24x7 customer services and reducing the operational cost by signing the jobs to the chatbot, the rise in customer's demands for self-service operations has led to the increase of the demand. This can be attributed to the growing preference of individuals toward messaging applications over social networking sites.
- Chatbots that are developed for integration with messaging applications are also expected to witness significant market demand in the future. Moreover, chatbot are gaining popularity as the businesses continue to automate their sales and customer services. The use of chatbot enables organizations to deliver timely services at reduced costs.
- Today there is a significant rise in the usage of messaging services, which is expected to surpass the social networking sites in terms of the number of users. There is also a significant development in the

Natural language processing field, which made the interactions between computer and human languages more streamlined. The entire ecosystem of the chatbot comprises various cross-linking features that help in optimum usage of the application.

- The users interacting with the customer service interrelate with the chatbot through the chatbot interface. The chatbot interact with users via webpage or mobile applications or through SMS etc. The chatbots are deployed on various channels such as standalone, website, and third-party messenger platforms. In standalone, bots are integrated with the applications and other software of the enterprises. The users interact with the chatbots deployed to raise the requests or the complaints to customer service.
- The chatbot is a messaging service that was developed by using a set of rules and artificial intelligence that can be interacted via a chat interface. There are two types of chatbots, one which is developed by a set of rules and the other which uses artificial intelligence and machine learning algorithms. Some chatbots also include voice recognition in order to offer an enhanced customer service experience. Currently, chatbot developers are including analytics into software application in order to get better insights into customer behavior and buying patterns.
- Open.AI, an American artificial intelligence company, released Chat GPT, an AI chatbot that answers every question. It is the latest series of AI which is referred to as GPT. The chat structure enables the AI to respond to "follow-up questions, admit its errors, contest false assumptions, and decline unsuitable requests.
- With the spread of the coronavirus, various marketing agencies, financial organizations, and grocery outlets have initialized the full fledge implementation of chatbots and conversational AI tools to enhance customer service in the crisis. Amid the pandemic, businesses face challenges in sustaining and are willing to try new technological avenues such as voice-based assistants and chatbots.

- Also, chatbots have assisted the medical and healthcare verticals during the difficult times. Several medical insurance companies and public health institutions worldwide have deployed chatbots to assist users/patients in providing proper COVID-19 advice and suggestions.

consists of artificial intelligence, natural language processing, and chatbot frameworks. **CHATBOT EVALUATION**