What are Chatbots

Learn about the different types of chatbots and how they work.

Chatbots all around

From the customer support dialog boxes you find on e-commerce websites to virtual assistants like Siri and Alexa, it's likely that you've encountered chatbots frequently in your everyday life. Like its name suggests, a chatbot is a piece of software designed to conduct a conversation or dialog. They can be found in a wide range of industries to serve a variety of purposes, ranging from providing customer support to aiding in therapy to simply being a source of fun and entertainment. Let's take a closer look at how chatbots are able to do what they do!

How chatbots work

To interact with the human user, chatbots must be able to:

- parse the user input
- interpret what it means
- provide an appropriate response or output

For example, a user query could be, "Show me hotels in Los Angeles for tomorrow."

A good chatbot will be able to identify the *intent* and *entities* of the query. The intent is the purpose or category of the user query, such as to retrieve a list of hotels. Entities are extra information that describes the user's intent. In this case, the entities are "Los Angeles" and "tomorrow." With these pieces of information, chatbots should be able to respond to the user with a list of available hotels for the correct location and date.

Types of chatbots

Chatbots generally fall into a few broad categories, depending on what purpose they are designed to serve.

Response architecture models

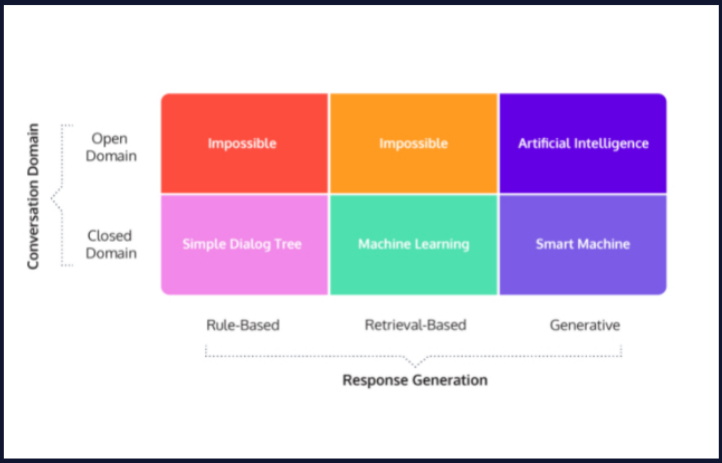
First, chatbots can be categorized according to how they generate the response that gets returned to the user. The simplest approach is the *rule-based* model, where chatbot responses are entirely predefined and returned to the user according to a series of rules. This includes decision trees that have a clear set of possible outputs defined for each step in the dialog.

Next, there is the *retrieval-based* model, where chatbot responses are pulled from an existing corpus of dialogs. Machine learning models, such as statistical NLP models and sometimes supervised neural networks, are used to interpret the user input and determine the most fitting response to retrieve. Like rule-based models, retrieval-based models rely on predefined responses, but they have the additional ability to self-learn and improve their selection of response over time.

Finally, *generative* chatbots are capable of formulating their own original responses based on user input, rather than relying on existing text. This involves the use of deep learning, such as LSTM-based seq2seq models, to train the chatbots to be able to make decisions about what is an appropriate response to return.

While generative models are very flexible and powerful in that they are not confined to a predefined set of rules or responses, they are also significantly more challenging to implement. Training these chatbots require an abundance of data, and it is often unclear what gets used for their decision-making, making them more prone to grammatical errors and nonsensical replies. By contrast, retrieval-based models can guarantee the quality of the responses since they are predefined, but these chatbots are in turn restricted to language that exists within the training data.

Thus, chatbots often use a combination of the different models in order to produce optimal results. For example, a customer support chatbot may use generative models for creating open-ended small talk with the user, but then are able to retrieve professional, predefined responses for answering the user's inquiries regarding the business or product.



Conversation domains

Chatbots can also be categorized based on the range of conversation topics they are able to cover. *Closed domain* chatbots, or *dialog agents*, are restricted to providing responses with a particular focus, such as booking a hotel room. Because they are designed with a specific goal in mind, these chatbots are often very efficient and have great success in accomplishing what they are intended to accomplish. The user-perceived quality is also high, because users don't expect the chatbots to provide responses outside of the pre-established domain.

On the other hand, *open domain* chatbots, or *conversational agents*, are capable of exploring any range of conversation topics, much like how a human-to-human interaction would be. Many of these "companion bots" have filled the roles of a friend or therapist, allowing the user to connect with them on an emotional level. While they have great potential, open domain chatbots are challenging to implement and evaluate.

It is worthwhile to note that the term "chatbot" is sometimes reserved only for open domain conversational agents, but for the purpose of this course, we include closed domain dialog agents as well.

Initiatives

Another way chatbots can be categorized is by which side – the user or bot – is able to take initiative on the conversation. Looking back at our previous example on hotel search, notice how the user is free to provide their request in their own words, and the chatbot is able to identify and piece together the relevant

words, and the chatbot is able to identify and piece together the relevant keywords to answer the query. This is an example of a *mixed-initiative* system, representative of a normal human-to-human conversation where all participants have the chance to take initiative.

By contrast, a *system-initiative* system is one where the chatbot controls the conversation and explicitly asks for each piece of information, such as the date and location for the hotel booking. While this system is more straight-forward to implement because user response can be anticipated, it lacks the flexibility and naturalness that characterize a normal human dialog.

Into the future

Chatbots have come a long way. What started out as computers that attempted to mimic human conversation has grown into elaborate systems that are able to carry out a multitude of functions and goals. But this great power and potential doesn't come without a price. As chatbots continue to evolve, there are also growing ethical concerns that arise. We will discuss this in more detail in the next article!