OpenBook: A knowledge-based method to develop Question-Answering chatbots

Botpress Technologies Inc.

Abstract

Many internet users have encountered talking with a virtual digital assistant at least once. These interactions could be in the form of voice talking with popular assistants like Siri, Alexa, or Google Assistant or interacting with text-based chatbots. All chatbots involve text processing, and one of the most prominent approaches for building chatbots is the intent-based approach. But these methods are challenging to build, monotonous for users, and do not scale well. This whitepaper introduces a new solution developed by Botpress called OpenBook, which uses a knowledge-based approach for building question-answering chatbots. This paper provides comparison results against several popular intent-based chatbot building platforms to demonstrate the performance and ease of implementing chatbots using OpenBook.

Introduction to Chatbots

A Chatbot (or conversational agent) is a software application that simulates a conversation with a user in a natural language through the exchange of messages. The conversational exchange format can be text or voice hosted through a web application, mobile application, or consumable service.

One of the primary advantages of chatbots is that they streamline user interactions and help businesses and individuals emulate human-to-human interactions. As a result, chatbots enable companies to enjoy greater operational efficiencies and an improved customer experience while also reducing costs and interaction times for employees and customers alike. Chatbots also provide an automated way of gathering data that helps organizations better understand their customers — leading to better and more personalized solutions.

With the proliferation of chatbots and the technology supporting their development, there is a stark increase in the number of organizations utilizing them. Inside Intelligence predicts that consumer retail spending through chatbots will increase to $142 billion worldwide by 2024, compared to the $2.8 billion spent in 2019 [1]. Several industries are adopting the use of chatbots, including prominent sectors like online retail, customer service, telecommunications, and banking. Juniper Research estimates $7.3 billion in operational cost savings globally by 2023 in the banking industry alone through chatbots [2].

How are Chatbots built?

Chatbots are considered one of the most prominent emerging technologies. As a result, there is a huge demand for the tools to develop such chatbots. Conversational AI platforms have played an important role in democratizing the development of chatbots across various industries. These platforms provide a technology stack for users to build and customize chatbots according to their needs.

Typically, conversational AI platforms allow developers to create chatbots using a flow-based approach or an intent-based approach. The flow-based approach uses a predefined set of routes (mapped like a flowchart) for the chatbots to interact with the users. On the other hand, the intent-based approach enables the understanding of free text input by using Natural Language Understanding (NLU) methods to understand a user's intention (hence, 'intent-based') and provide responses based on the identified intent. While an improvement over flow-based models, intent-based approaches come with a list of shortcomings.

Intent-based methods' challenges

Intent-based approaches require a predetermined list of intents that the chatbot should be able to handle. However, creating an exhaustive list of intents requires intensive building time and resources, making it error-prone. The intents have to be predetermined, and for each intent, the user has to provide a set of utterances

to learn the relevance of potential questions. As a result, any errors made while providing the utterances will produce an undesirable outcome. The intent-based approach also has to be trained using machine learning models every time there is a change in the intent or the related utterances. Further, because this approach generates responses based on the relevant context, it must be programmed explicitly.

OpenBook solution

As an alternative to intent-based chatbots, OpenBook uses a knowledge-driven approach to build question-answering chatbots. The only input required for OpenBook is to provide the knowledge as a book. A book is a collection of short facts using a markdown format.

Unlike the intent-based approaches, chatbots built on OpenBook do not require the input of questions, their variations, and various answers to function. Instead, chatbots built on OpenBook can be deployed as soon as a book is provided to the platform. By going intentless, OpenBook requires much less data and is robust to unanticipated questions, facilitating a faster and more reliable way of creating intelligent chatbots.

OpenBook also allows developers to select the level of creativity of generated responses. They can be as strict as using the verbatim from the book of facts or generating creative human-like responses. Unlike intent-based approaches, OpenBook is scalable without affecting the system's performance. Also, OpenBook does not involve training or fine-tuning of machine learning models when the knowledge is updated, making it one of the first commercial products with a zero-shot learning method for creating chatbots.

Methodology to compare platforms

OpenBook was compared with three other popular intent-based chatbot development platforms: Rasa, IBM's Watson Assistant, and Google's Dialogflow. Three experts, each having at least one year of experience in developing chatbots using the respective platform, were given a specification/factsheet [3] to build a chatbot for a hotel.

Two high-level activities were conducted for comparison of different chatbot development platforms — chatbot building and the chatbot user interaction. The first activity of chatbot building was focused on identifying the 'ease-of-development' and the 'input size' consumed by each platform. The 'ease-of-development' was evaluated by measuring the time taken to build chatbots by the respective experts of each platform. The 'input size' for each platform was measured by the total number of non-whitespace characters provided as an input for building the chatbot.

The second activity was to measure the chatbot's performance using various parameters through user interactions. The same set of questions was fed as input to each chatbot, and the chatbot's responses to each question were recorded. This question-response set and the specification sheet were provided to two groups of independent evaluators, who then evaluated and rated it against 11 parameters, as shown in Table 1. Each of these 11 parameters captures unique information about the response and its correctness.

1. The response contained the complete answer to the question asked
2. The response contained only partial information to the question asked
3. The response has at least one piece of additional unrelated information
4. More than half of the responses contained information that is not relevant to question
5. The response was an unaltered repetition of facts
6. The response contained misleading/untrue information
7. The response was completely wrong
8. The chatbot responded it did not know the answer when it should have answered
9. The response showed that the chatbot did not understand question
10. The chatbot correctly responded that it did not know the answer to out-of-scope question
11. The question is invalid

Table 1. Parameters used for scoring the response.

The evaluators were unaware of the platform used for response generation, and the order of the responses was also randomized to avoid any bias.

Preliminary Results and Analysis

The following section provides a brief analysis of the preliminary results. Fig. 1 measures the 'ease-of-development' by comparing the time to build of each of the chatbots developed, given the same specification. The results show the time to develop a chatbot using OpenBook is considerably less compared to the next best time by both Rasa and Dialogflow.

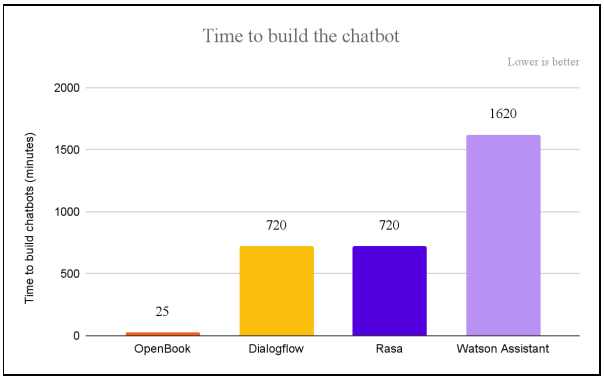

Fig. 1. Time to build the chatbots by experts on different development platforms.

The second analysis on the chatbot building was to compare the size of the input by measuring the total number of non-whitespace characters used as an input for building a chatbot for each of the platforms. It is evident from Fig. 2 that OpenBook takes the least amount of characters as an input.

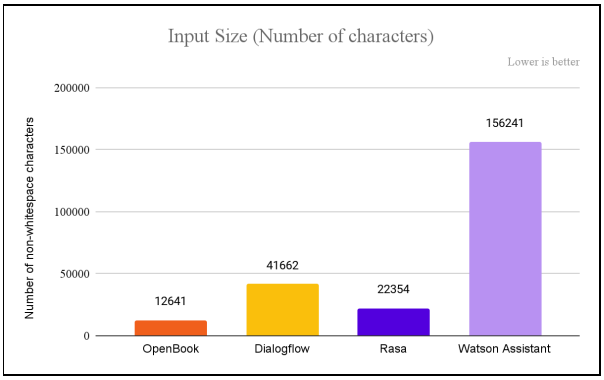

Fig. 2. Comparison of input sizes on different platforms

A preliminary analysis was also done to measure the performance of the chatbots created using different platforms. The results shown are from the initial full dataset (after the removal of erroneous instances). This paper will be updated with further analysis as and when new datasets, results, errors, and insights are identified. The raw dataset, encoding, and analysis work are available in [4] and [5]. The dataset can also be downloaded from Kaggle [6]. There are a total of 5019 question-response pairs on which results have been reported. The metrics shown are an average of scores between two sets of evaluations.

Fig. 3 shows a comparison of the number of times the responses from each chatbot were completely accurate, as measured by the evaluators.

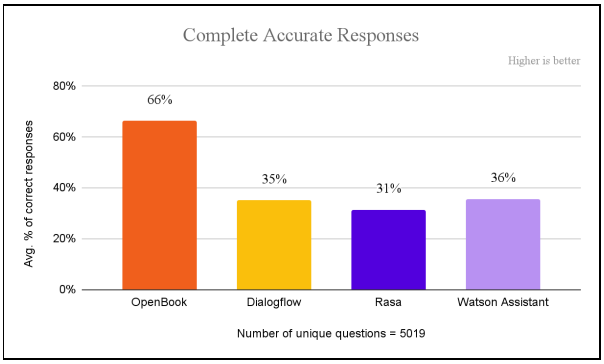

Fig. 3. Comparison of performance of chatbots by measuring the number of accurate responses (average of 2 rounds of evaluation).

A comparison was made to measure the number of times the response generated was not correct, as shown in Fig. 4. Inaccurate responses are responses that are irrelevant to the question asked but relevant to the information provided to the chatbot while building it. From Fig. 4, it can be seen that OpenBook performs better than the intent-based methods.

Another parameter that determines the creativity of a chatbot to provide more human-like responses was to identify the number of times that the chatbot repeated the information verbatim from the information provided to build the chatbot. This shows that OpenBook can provide responses that do not sound monotonous.

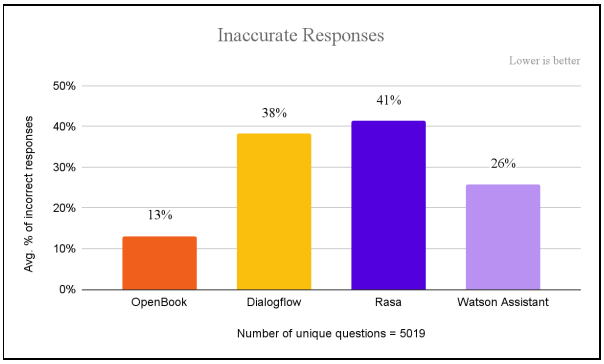

Fig. 4. Comparison of performance of chatbots by measuring the number of incorrect responses (average of 2 rounds of evaluation).

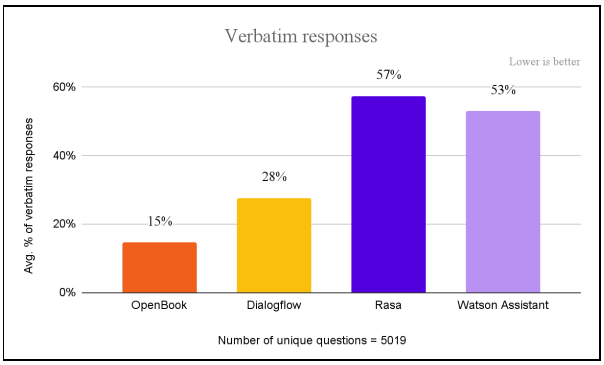

Fig. 5. Comparison of verbatim responses by different chatbots (average of 2 rounds of evaluation).

Another analysis was done to see if the chatbots were able to provide information that was completely false, as opposed to providing inaccurate responses, as shown in Fig. 6. Providing false responses is an area where OpenBook is below par against the intent-based chatbots and suggests clear opportunities to improve the intentless chatbot platform further.

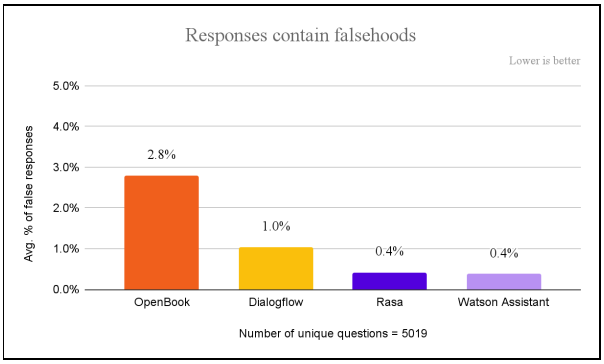

Fig. 6. The number of instances where the chatbots created fake responses irrelevant to the information used for the bots (average of 2 rounds of evaluation).

There were also two rounds of evaluation to determine the best and worst responses for each question given by the four chatbots. The chatbots' best and worst responses are shown in Fig. 7 and Fig. 8, respectively. The evaluator can rate more than one response as the best or worst for each question, given that the responses by different chatbots might be similar. Both numbers show that OpenBook fared well compared to the other chatbots in having the most number of best responses and the least number of worst answers.

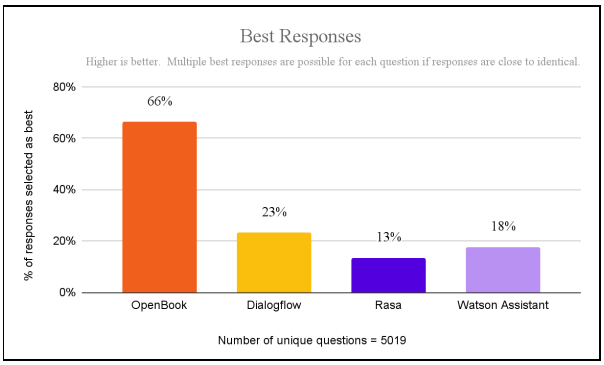

Fig. 7a. The number of times (in percentage) the response from each chatbot was considered to be the best (average of 2 rounds of evaluation).

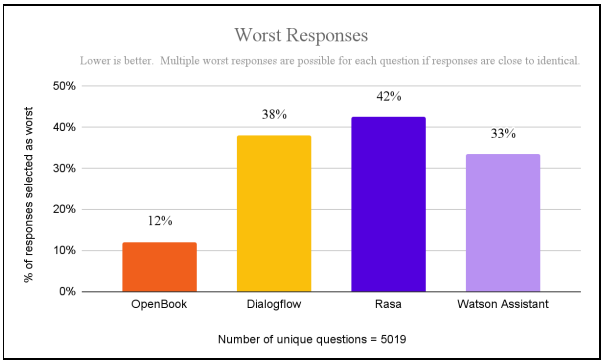

Fig. 7b. The number of times (in percentage) the response from each chatbot was considered to be the worst (average of 2 rounds of evaluation).

Study Parameters

While the benchmarking exercise was not completely objective; it has value in highlighting the progress made with the use of a knowledge-based approach over intent-based approaches. An effort has been made to mitigate bias through obvious sources in the results. Any identified erroneous data has been removed from the analysis. However, there are multiple constraints and fairness measures in the methodology and analysis. Tables 2, 3, and 4 provide a non-exhaustive list of the constraints, fairness measures, and identified opportunities for improvement in the methodology. Despite the issues, the results indicate the advantages of the knowledge-driven approach and can be considered the future of chatbot development, with OpenBook's first version being the first step toward it. Multiple improvements can be made to make the knowledge-based approaches easier to build and with better performance.

1. The chatbots were built by experts with a minimum of one year of experience in building bots using their respective platforms but may not have an equal amount of experience.
2. Only one attempt was provided to each expert to build the chatbot.
3. The performance of the different platforms heavily reflects the approach of the developers of the chatbots. Even though the developers were experts, the choices they made influenced the results.
4. Each response was rated by two different evaluators.
5. The time of development of building a chatbot was not actively tracked using an automated system.
6. The subjectivity and bias of independent evaluators in understanding specifications, questions, and responses by different bots are not quantified.
7. The level of expertise of independent evaluators for coding responses might differ.

Table 2. Constraints of study

1. The order of the responses from different chatbots presented to evaluators was randomized.
2. Evaluators were unaware of the platform used to develop while evaluating the responses.
3. Data will be shared so that independent analysis can be conducted.

Table 3. Fairness Measures

1. Chatbots developed using other platforms fare well in certain parameters. It is an active area that Botpress is working to improve.
2. Data cleaning and sanity checks on the encoded answers can be further improved.

Table 4. Identified opportunities for improvement in OpenBook

Conclusion

OpenBook is a knowledge-based approach to building question-answering chatbots without training and fine-tuning models. These benchmarks show that OpenBook beats its counterparts using intent-based methods in both categories — the ease of building a chatbot by developers and the satisfaction scores by the chatbot users. From the preliminary results and analysis, one can infer that knowledge-driven approaches work well over the intent-based approaches in providing the right and relevant responses and making it easier for building the chatbots using OpenBook.

References

[1] Chatbot market in 2022: Stats, trends, and companies in the growing AI chatbot industry, Business Insider, 2022 (link)

[2] Bank Cost Savings via Chatbots to reach $7.3 billion by 2023, as automated customer experience evolve (link)

[3] Specification/Factsheet given to chatbot builders (link)

[4] Raw Data and Analysis Spreadsheet - Round 1 (link)

[5] Raw Data and Analysis Spreadsheet - Round 2 (link)

[6] Justin Watson, Sylvain Perron. (2022). Botpress Question and Answering Chatbot Data, Kaggle. https://doi.org/10.34740/KAGGLE/DSV/3594757 (link)