# Blender Bot 3.0

Amrutha S Rao
*Information Science and Engineering*
*RNS Institute of Technology*
Bengaluru, India
1RN19IS024.amrutharao@gmail.com

Anagha P Dixit
*Information Science and Engineering*
*RNS Institute of Technology*
Bengaluru, India
1RN19IS026.anagha@gmail.com

Ananya D G
*Information Science and Engineering*
*RNS Institute of Technology*
Bengaluru, India
1RN19IS027.ananyadg@gmail.com

Ashrya Verma
*Information Science and Engineering)*
*RNS Institute of Technology*
Bengaluru, India
1RN19IS038.ashryaverma@gmail.com

Dr. Satish Kumar
*professor, Guide*
*Information Science and Engineering*
*RNS Institute of Technology*
Bengaluru, India

***Abstract*—Blender Bot 3 is a large-scale dialogue model that aims to provide a more advanced conversational experience for users. It is not only capable of engaging in open-domain conversations but also has access to a long-term memory and the internet, enabling it to gather more information as needed. The model has been trained on a vast range of user-defined tasks to ensure that it can handle a variety of conversational scenarios.To deploy this model, the research team has made both the code and the model weights available for public use. The features and functions are extensible to the general public. To ensure user safety, the team has included mechanisms that can detect and flag harmful or offensive content. Considering the nature of bots and their self learning capabilities which allow them to grow, the continuous research does not stop. The data from the deployment will be used to train the bot and hence reduces the bias evoking responsible conversational agents. Overall, Blender Bot 3 is an exciting advancement in conversational AI, and the research team's focus on continuous learning through interaction will ensure that it remains relevant and responsive to user needs.**

***Index Terms*—AI, Chatbot, ML, BlenderBot**

## I. Introduction

This paper introduces BlenderBot 3 (BB3), which is an open-domain conversational software agent. The team at Meta have deployed this model as an English-speaking conversational agent on a public website that is accessible to adults in the United States. The contents and data are available to help expedate research and evolve the quality of the software agents and commitments are in line with the principles established by Sonnenburg et al. (2007) and Pineau et al. (2021). The goal is to develop models that continue to improve based on interactions with users, with a focus on becoming more responsible and useful.

### *A. Contributions of this work*

- This work introduces the BlenderBot 3 (BB3) model, which is a transformer model with 175B parameters. The model is initialized from the pre-trained model OPT-175B and fine-tuned to perform modular tasks based on recent work by the team (Shuster et al., 2022). BB3 is equipped with features such as storing information in long-term memory and searching the internet for information, inherited from its predecessors.
- Additionally, the study focuses on training BB3 with human feedback from conversations to improve its performance on tasks that are relevant to users. A full report of the study is presented in a companion paper (Xu et al., 2022b), and the findings are used to fine-tune BB3 on a large number of user-defined tasks.
- The paper provides a framework for the deployment design of the system. Additionally, the results of previous bots were studied to understand the differences.
- The recently launched system has exhibited superior performance compared to existing openly available chatbots, including its two predecessors, by a significant margin.

## II. Evolution of BlenderBot

### *A. BlenderBot 1.0*

In the year 2020, FaceBook AI Research proposed an open-domain conversationsal agent with the ability to instill skills as empathy and knowledge. Figure 1 depicts the architecture of BlenderBot 1. The model uses poly encoders to revert back to previous information and understand based on experience gained by repeated use of the response retrieval feature of the bot to increase accuracy. It is modeled around pre-determined intents from Reddit data sets that provide social media comments, which help in training natural flow of conversations. Such an asset adds to the genuineness of the responses retrieved and refined from the bots. To further fine tune the bot the Blended Skill Test dataset is used. BST is formed through combinations of various resources that provides information and evokes more empathy. Rich datasets like the above mentioned provide better chances for knowledge retrieval.

### *B. BlenderBot 2.0*

One of the major flwas when considering a conversational software agent is that it does not remember the previous

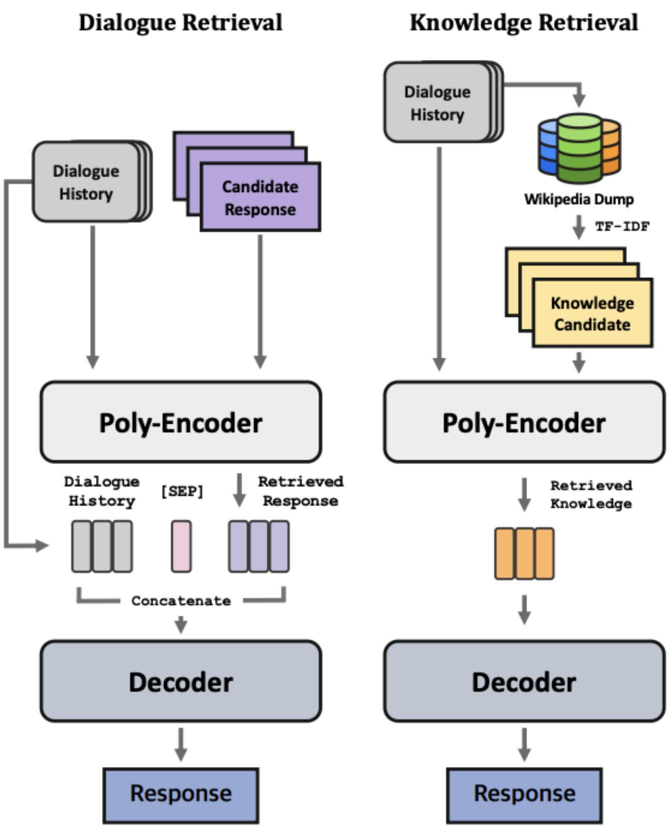

Fig. 1. Architecture of BlenderBot 1.0

conversations of the client it is servicing and treats each query as a singular conversation. For conversations to flow like the way humans converse, the chatbot needs to be trained or made intelligent enough that it remembers the context of the conversation while interacting with clients. These limitations dominantally exist in Blender Bot 1 and GPT-3. Blender Bot 2 was developed with the purpose of introducing long term memory and further enriching client experience. The architecture is shown in Figure 2. The bot creates sessions to store the data to access it if needed during a conversation. The responses also become more dynamic using the internet's knowledge.

### C. Existing Issues

- BlenderBot 1 has a simple structure to retrieve reponses for a particular intent. As the task is to search the dialogue history and access the knowledge base for providing answers, the method is crude and the responses could be incorrect. It faces limitations in providing smart reactions to client queries.
- BlenderBot 2 has limitations with response latency such as the response time it takes for a client query. Also considering it works in the way of forming sessions, there is a lack of clean data which creates bias and redundant reactions from the bot.

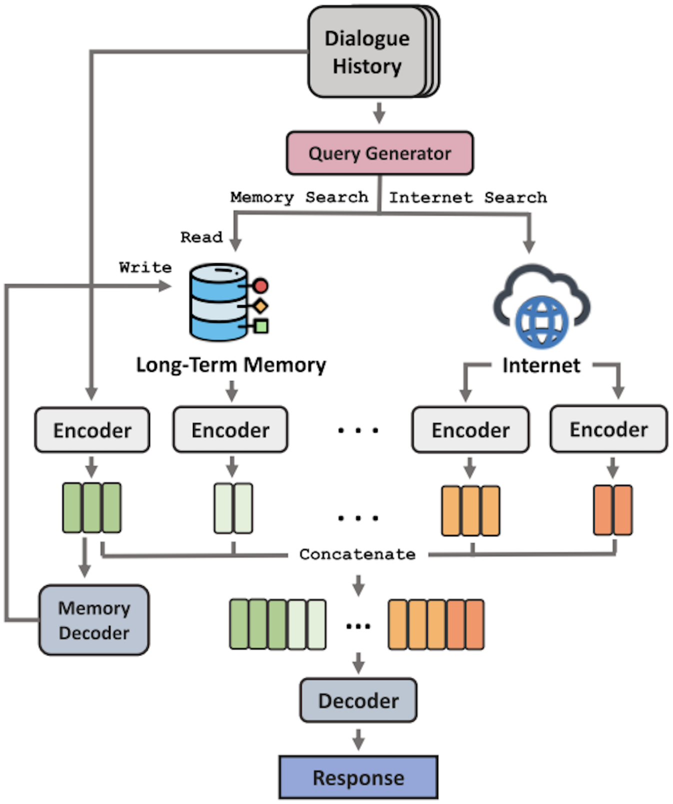

Fig. 2. Architecture of BlenderBot 2.0

## III. BlenderBot 3.0

BlenderBot 3 is a 175B- parameter chatbot that is available publically with the codes, datasets and model to construct the bot. It uses the internet to search for answers and creates a holistic approach to conversational software agents. Two recently created machine learning techniques, SeeKeR and Director are used to create conversational models that provide an environment of exchanges which are knowledgeable and also analyse the sentiment of the client. Accessing the internet provides more detailed information for intents.

### A. Purpose

The main purpose of BlenderBot 3 is to have safe environment while having conversations with the bot. As its an open-domain chatbot the data training needs to be free of bias and hate speech that could be targeted against any group. It is to increase interactions while focusing on making the bot safe and create an inclusive AI system for everyone around the world. This was aimed at a time in the world when there are ongoing wars and daily bouts of discrimination around the world. Increasing safety and efficiency of AI systems would help in creating more technological advancements and allow people to accept changes more easily. It is not an easy enhancement but is a growth process by accepting large data and clearing it of any hate speech and have a clear sentiment analysis on the same.

## B. Architecture

The architecture of BlenderBot 3 is accomplished through a single transformer model through which all the modules are executed.

- Internet search is carried out when there is an absolute necessity of the action to answer a particular query.
- It has to generate a search query to be asked in the search engine.
- Knowledge response needs to be generated from the infromation base established.
- Long term memory to maintain human like conversations is attained by summarising the conversations and highlighting the contexts.
- This memory is accessed when the conversation with the user requires context understanding. This is hence called recalled memory.
- Lastly, the response is generated while analysing the best course of action available to satisfy user experience.

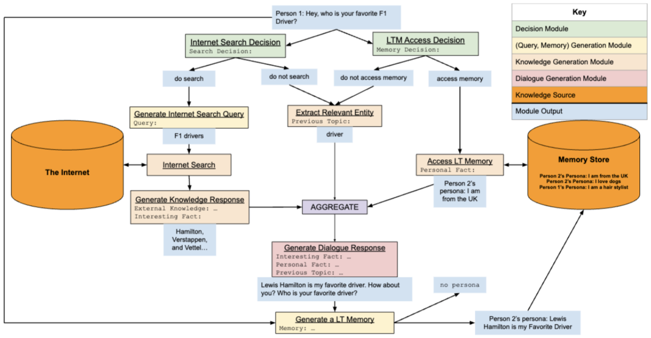
Person 1: Hey, who is your favorite F1 Driver?
Internet Search Decision
Search Decision:
LTM Access Decision
Memory Decision:
do search
do not search
do not access memory
access memory
Generate Internet Search Query
Query:
F1 drivers
Extract Relevant Entity
Previous Topic:
driver
The Internet
Internet Search
Generate Knowledge Response
External Knowledge: …
Interesting Fact:
Hamilton, Verstappen, and Vettel…
AGGREGATE
Access LT Memory
Personal Fact:
Person 2's persona: I am from the UK
Memory Store
Person 2's Persona: I am from the UK
Person 2's Persona: I love dogs
Person 1's Persona: I am a hair stylist
Generate Dialogue Response
Interesting Fact: …
Personal Fact: …
Previous Topic: …
Lewis Hamilton is my favorite driver. How about you? Who is your favorite driver?
no persona
Generate a LT Memory
Memory: …
Person 2's persona: Lewis Hamilton is my Favorite Driver
Key
Decision Module
(Query, Memory) Generation Module
Knowledge Generation Module
Dialogue Generation Module
Knowledge Source
Module Output

Fig. 3. Architecture of BlenderBot 3.0

## IV. Seeker : Search Technique

A real time search technique that calculates the probability likelihood while querying the search engine for responses. The figure below describes the the work flow of a response retrieval to query from the search engine. Utilizing the seeker, the likes and dislikes of the information can be mapped to calculate the probable output to a query. Creating sets of likes and dislikes enables to base recommendations to the clients. When the bot understands these functionalities it helps to evoke better empathy and hence enabling more human like conversations with the client. This form of searching creates an interactive system and provides the machine ability to deal with scenarios more dynamically. There are few algorithms that help with the categorisation of the terms to create a better understanding of the contexts :

- Log-Likelihood : Provides a way to measure better fit for a model. This allows to obtain cleaner data sets.
- Item recommendation : Utilizes pure exploration algorithms to clean the data and make the responses more accurate.
- Sampling : This method allows to calculate the true target value of a probability likelihood.

Through these methods the seeker generates metrics for streamling data usage and the bot functionalities.

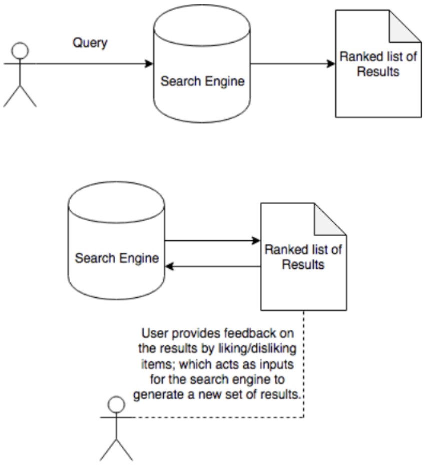
Query
Search Engine
Ranked list of Results
Search Engine
Ranked list of Results
User provides feedback on the results by liking/disliking items; which acts as inputs for the search engine to generate a new set of results.

Fig. 4. Search Workflow

## V. Director : Classifier Technique

Director is a classifier technique utilizing language models to predict the next token during a conversation understanding the context. The process takes place as shown in figure 5. There two heads : 1) Language Head and 2) Classifier Head. The two heads are used to categorize the terms into their sentiments of positive or negative based on the context and grammar understanding. The classifier head maps to each token and is processed through the shared transformer. This technique is a unified structure which has the following main features :

- Classifiers are not bidirectional but rather involuntarily combative which allows newer tokens to use previous information.
- Classifiers allows parallelism which enhances quality of candidate separation
- The language model of all classifiers is the same at the core reducing computational work.

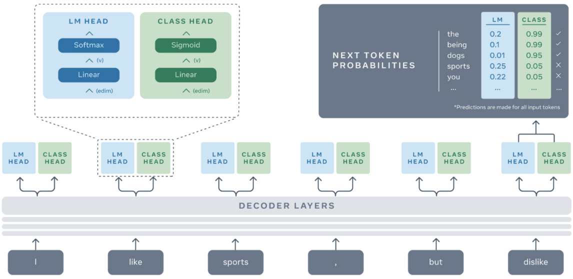

Fig. 5. Director Model For Positive And Negative Terms

## Conclusion

Chabots were invented with the sole purpose of providing a response to a query, introducing the internet and expanding the knowledge base increases the quality of such responses. Understanding that developing sentiments and language models we bridge the gap between the complexity of a subject. Also incorporating newer ML techniques such as Seeker and Director, new methods are being studied and established to improve efficiency of a software that is inherently made for the convenience of humans. The above paper discusses the new generation of BlenderBot which provides an array of possible applications to enhance human experience and help with further technology advancements. Creating a safe environment to use the machines which will help the bots in reaching a wider user space.

## References

[1] Kurt Shuster, Jing Xu, Mojtaba Komeili, ”BlenderBot 3: a deployed conversational agent that continually learns to responsibly engage”

[2] Ari Biswas Thai T. Pham, ”Seeker: Real-Time Interactive Search”

[3] META AI, ”BlenderBot 3: A 175B parameter, publicly available chatbot that improves its skills and safety over time”

[4] Kushal Arora, Kurt Shuster, Sainbayar Sukhbaatar, Jason Weston, ”DIRECTOR: Generator-Classifiers For Supervised Language Modeling”

[5] Jungseob Lee, Midan Shim , ”There is no rose without a thorn: Finding weaknesses on BlenderBot 2.0’s in terms of Model, Data and User-Centric Approach”

[6] Vanshika Arya, Rukhsar Khan, Mukul Aggarwal, ”A Chatbot Application by using Natural Language Processing and Artificial Intelligence Markup Language”

[7] Anagha P Dixit, Amrutha S Rao, Padmachandana, Smriti S, ”Conversational AI and Artificial Neural Networks”