Implementing NLU Based Bangla Chatbot Using Google Bard and Google Translator

By

Shahriar Rahman Anuvab - 21101094

Syed Aref Ahmed - 19201124

Kazi Ashif Shahrier - 21201795

Mishkat Sultana -19201028

S. M. Abdullah - 19201050

Submitted to:

Moin Mostakim

Lecturer, Department of Computer Science and Engineering, BRAC University

Abstract

Bengali, the official language of Bangladesh, stands as the sixth most widely spoken language globally, with 265 million native and non-native speakers. Addressing the growing demand for advanced language processing capabilities in the digital era, researchers have undertaken numerous experiments, building invaluable tools and techniques for the creation and manipulation of Bangla language materials. Concurrently, researchers persist to facilitate seamless integration of the Bangla language into online and technical domains.

This paper explores the integration of Google Bard's AI-powered chatbot tool, 'Bard,' into the development of a Bangla chatbot system. Grounded in Natural Language Understanding (NLU) and Neural Networks, Bard introduces a new frontier in language-based technologies. In this paper we will discuss the process of incorporating Google Bard's Neural Network API and underscore the potential it holds for the creation of an effective and user-friendly chatbot system specifically for Bengali speakers. By exploring how we can combine Google Bard's technology, we aim to contribute to the ongoing dialogue surrounding the enhancement of linguistic interactions in the digital landscape.

Introduction

A conversational agent, often referred to as a chatbot or chatterbot, is a computer program capable of engaging in natural language conversations with users. It can operate in two main ways: by drawing from a knowledge base for specific information (retrieval-based closed domain model) or by generating new sentences (generative-based open domain model). Some chatbots, known as goal-oriented, can even perform actions based on the conversation, such as a pizza ordering chatbot.

In today's digital landscape, chatbots are widely used across various domains, serving roles ranging from virtual tutors and digital assistants to customer service representatives. These conversational agents have become integral in our digitized world, playing roles as diverse as virtual therapists to enhance user experiences and streamline interactions.

This paper sets forth the ambition of constructing a conversational and generative open domain Bangla chatbot model, utilizing the capabilities of BARD. Our objective is to explore the potentials embedded within the BARD models, specifically in the field of Natural Language Understanding (NLU) and pre-training methodologies, with a vast amount of datasets in the Bangla language.

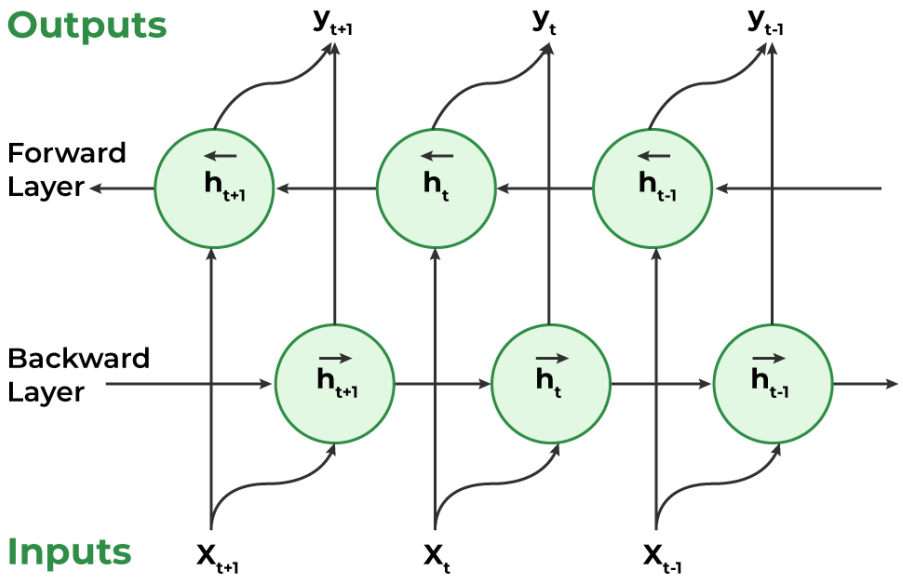
What is BARD?

When it comes to human-computer interactions, NLU plays a pivotal role. It enables machines to comprehend and interpret the meaning behind human language, making conversations between humans and AI systems more seamless and natural.

In the world of AI, Google BARD has emerged as one of the most prominent models in the realm of NLU. Google BARD, also known as Bidirectional Encoder Representations from Transformers, is Google's answer to enhance NLU capabilities. BARD has been trained on an extensive amount of text data from diverse sources, allowing it to grasp the nuances of language and context in a more comprehensive manner. It utilizes a bidirectional approach, where it takes into account the information from both preceding and succeeding words to better understand the user's intent.

Why BARD?

- Model Architecture: The architecture of BARD is built on Google's Pathways Language Model 2 which adopts a bi-directional training approach, allowing it to capture the meaning and context of a sentence more effectively. By considering both the left and right contexts simultaneously, BARD can better understand the nuances and dependencies in natural language.



- Data size: When it comes to training language models, data is the most important factor. Google BARD benefits from Google's vast resources, allowing it to be trained on a massive amount of data, including a mixture of web text and books. The sheer scale of the training data contributes to BARD's enhanced language understanding and ability to grasp diverse user intents.
- Results and benchmarks: Numerous evaluations and benchmarks have demonstrated BARD's superiority in grasping user intent compared to other open domain models. It consistently outperforms other models in various natural language understanding tasks,

such as question-answering, sentiment analysis, and text classification. BARD's remarkable results have cemented its position as a frontrunner in the field of NLU.

Codebase and Setup

In order to demonstrate the power of Google BARD in developing a Bangla chatbot, we developed a simple chatbot incorporating Google BARD API and Google translator in colab. The codes can be found in here: https://colab.research.google.com/drive/1AaNwAqrGmFu0MUoz8E8TT6CCo8IVlk_b?usp=sharing

Installation processes as follow:

i) We installed Google Translate API with the following code- *pip install googletrans==4.0.0rc1*. 4.0.0rc1 is the version number of the API. We encountered instabilities with the latest version, so we opted for a stable version. After installing we tested a Bengali translation by importing the API and using 'bn' as destination. 'bn' is the google translators distinction of Bengali.

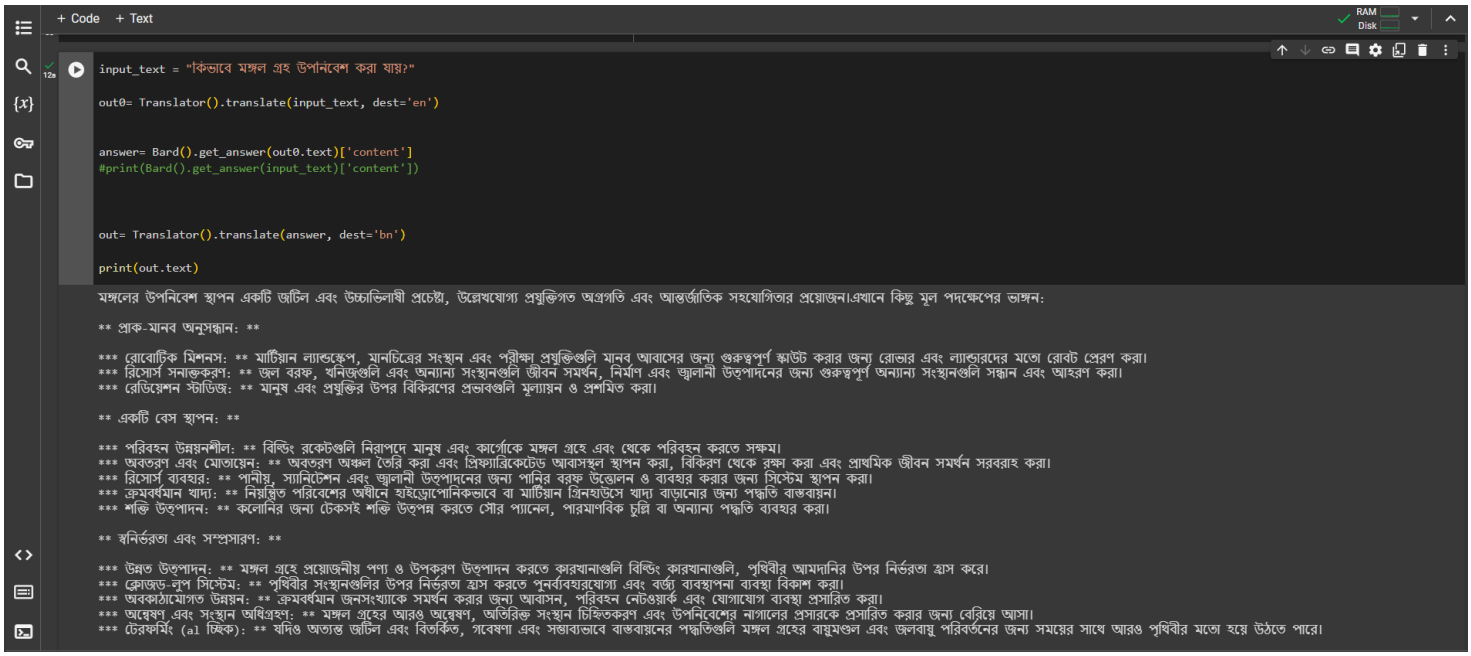
ii) Next, we installed Bard API with the following code- *!pip install bardapi.* Upon installation, we imported Bard, OS and Time functions and then inserted our google bard API key. Bard Api key will vary on each google account. So, in order to run our project on a different PC, API key related to that PC's gmail id must be provided there. By inserting the API key, we are now successfully connected to Google Bard.

iii) Now. we opted to take a text input in Bengali and then converted the input into English by setting the dest= ‘en’ on google translator. Then we have inserted the translated input into Google Bard with the following command- *answer= Bard().get_answer(out0.text)['content']*.
Now we have fetched the Bard’s output and translated the output into Bangla in the background using Google translate. We have printed the translated output which gives a thoughtful and comprehensive reply in Bangla against any question. Because of using the translator tool, even if we ask a question in English or any other language in the world in our program, our program will always reply meaningfully in Bangla which is exclusive to our chatbot.

We found that by incorporating Google BARD we were able to build a very convenient and user satisfactory chatbot tool. This further directs us to our assumption that the models used in BARD are trained on large amounts of dataset, and they have enormous possibilities in the field of developing chatbot tools.

Results

1:



```
input_text = "কিভাবে মঙ্গল গ্রহ উপনিবেশ করা যায়?"

out0= Translator().translate(input_text, dest='en')


answer= Bard().get_answer(out0.text)['content']
#print(Bard().get_answer(input_text)['content'])



out= Translator().translate(answer, dest='bn')

print(out.text)
```

মঙ্গলের উপনিবেশ স্থাপন একটি জটিল এবং উচ্চাভিলাষী প্রচেষ্টা, উল্লেখযোগ্য প্রযুক্তিগত অগ্রগতি এবং আন্তর্জাতিক সহযোগিতার প্রয়োজন।এখানে কিছু মূল পদক্ষেপের ভাঙ্গন:

** প্রাক-মানব অনুসন্ধান: **

*** রোবোটিক মিশনস: ** ম্যাটিয়ান ল্যান্ডস্কেপ, মানচিত্রের সংস্থান এবং পরীক্ষা প্রযুক্তিগুলি মানব আবাসের জন্য গুরুত্বপূর্ণ স্কাউট করার জন্য রোভার এবং ল্যান্ডারদের মতো রোবট প্রেরণ করা।
*** রিসোর্স সনাক্তকরণ: ** জল বরফ, খনিজগুলি এবং অন্যান্য সংস্থানগুলি জীবন সমর্থন, নির্মাণ এবং জ্বালানী উত্পাদনের জন্য গুরুত্বপূর্ণ অন্যান্য সংস্থানগুলি সন্ধান এবং আহরণ করা।
*** রেডিয়েশন স্টাডিজ: ** মানুষ এবং প্রযুক্তির উপর বিকিরণের প্রভাবগুলি মূল্যায়ন ও প্রশমিত করা।

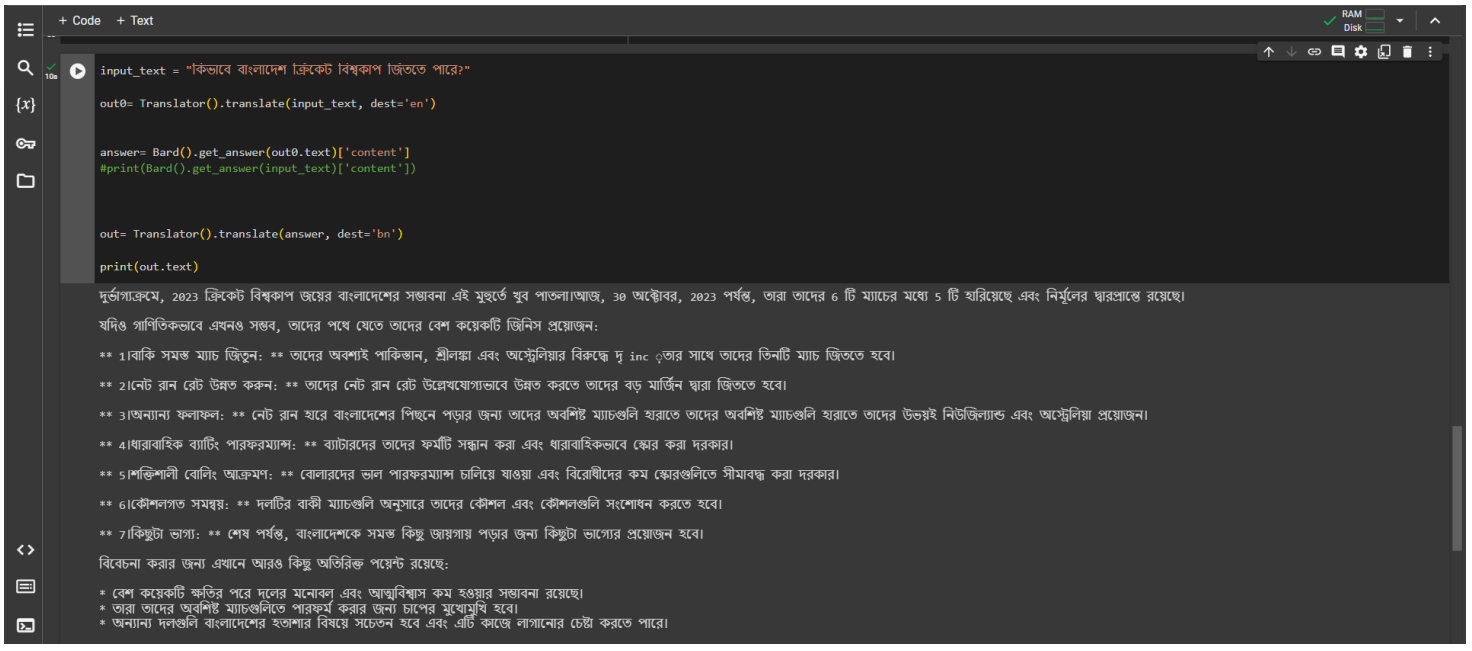
** একটি বেস স্থাপন: **

*** পরিবহন উন্নয়নশীল: ** বিল্ডিং রকেটগুলি নিরাপদে মানুষ এবং কার্গোকে মঙ্গল গ্রহে এবং থেকে পরিবহন করতে সক্ষম।
*** অবতরণ এবং মোতায়েন: ** অবতরণ অঞ্চল তৈরি করা এবং প্রিফ্যাব্রিকেটেড আবাসস্থল স্থাপন করা, বিকিরণ থেকে রক্ষা করা এবং প্রাথমিক জীবন সমর্থন সরবরাহ করা।
*** রিসোর্স ব্যবহার: ** পানীয়, স্যানিটেশন এবং জ্বালানী উত্পাদনের জন্য পানির বরফ উত্তোলন ও ব্যবহার করার জন্য সিস্টেম স্থাপন করা।
*** ক্রমবর্ধমান খাদ্য: ** নিয়ন্ত্রিত পরিবেশের অধীনে হাইড্রোপোনিকভাবে বা ম্যাটিয়ান গ্রিনহাউসে খাদ্য বাড়ানোর জন্য পদ্ধতি বাস্তবায়ন।
*** শক্তি উত্পাদন: ** কলোনির জন্য টেকসই শক্তি উত্পন্ন করতে সৌর প্যানেল, পারমাণবিক চুল্লি বা অন্যান্য পদ্ধতি ব্যবহার করা।

** স্বনির্ভরতা এবং সম্প্রসারণ: **

*** উন্নত উত্পাদন: ** মঙ্গল গ্রহে প্রয়োজনীয় পণ্য ও উপকরণ উত্পাদন করতে কারখানাগুলি বিল্ডিং কারখানাগুলি, পৃথিবীর আমদানির উপর নির্ভরতা হ্রাস করে।
*** ক্লোজড-লুপ সিস্টেম: ** পৃথিবীর সংস্থানগুলির উপর নির্ভরতা হ্রাস করতে পুনর্ব্যবহারযোগ্য এবং বর্জ্য ব্যবস্থাপনা ব্যবস্থা বিকাশ করা।
*** অবকাঠামোগত উন্নয়ন: ** ক্রমবর্ধমান জনসংখ্যাকে সমর্থন করার জন্য আবাসন, পরিবহন নেটওয়ার্ক এবং যোগাযোগ ব্যবস্থা প্রসারিত করা।
*** অন্বেষণ এবং সংস্থান অধিগ্রহণ: ** মঙ্গল গ্রহের আরও অন্বেষণ, অতিরিক্ত সংস্থান চিহ্নিতকরণ এবং উপনিবেশের নাগালের প্রসারকে প্রসারিত করার জন্য বেরিয়ে আসা।
*** টেরফর্মিং (al ছিক): ** যদিও অত্যন্ত জটিল এবং বিতর্কিত, গবেষণা এবং সম্ভাব্যভাবে বাস্তবায়নের পদ্ধতিগুলি মঙ্গল গ্রহের বায়ুমণ্ডল এবং জলবায়ু পরিবর্তনের জন্য সময়ের সাথে আরও পৃথিবীর মতো হয়ে উঠতে পারে।

2:



```
input_text = "কিভাবে বাংলাদেশ ক্রিকেট বিশ্বকাপ জিততে পারে?"

out0= Translator().translate(input_text, dest='en')


answer= Bard().get_answer(out0.text)['content']
#print(Bard().get_answer(input_text)['content'])



out= Translator().translate(answer, dest='bn')

print(out.text)
```

দুর্ভাগ্যক্রমে, 2023 ক্রিকেট বিশ্বকাপ জয়ের বাংলাদেশের সম্ভাবনা এই মুহূর্তে খুব পাতলা।আজ, 30 অক্টোবর, 2023 পর্যন্ত, তারা তাদের 6 টি ম্যাচের মধ্যে 5 টি হারিয়েছে এবং নির্মূলের দ্বারপ্রান্তে রয়েছে।

যদিও গাণিতিকভাবে এখনও সম্ভব, তাদের পথে যেতে তাদের বেশ কয়েকটি জিনিস প্রয়োজন:

** 1।বাকি সমস্ত ম্যাচ জিতুন: ** তাদের অবশ্যই পাকিস্তান, শ্রীলঙ্কা এবং অস্ট্রেলিয়ার বিরুদ্ধে দু inc ুতার সাথে তাদের তিনটি ম্যাচ জিততে হবে।

** 2।নেট রান রেট উন্নত করুন: ** তাদের নেট রান রেট উল্লেখযোগ্যভাবে উন্নত করতে তাদের বড় মার্জিন দ্বারা জিততে হবে।

** 3।অন্যান্য ফলাফল: ** নেট রান হারে বাংলাদেশের পিছনে পড়ার জন্য তাদের অবশিষ্ট ম্যাচগুলি হারাতে তাদের অবশিষ্ট ম্যাচগুলি হারাতে তাদের উভয়ই নিউজিল্যান্ড এবং অস্ট্রেলিয়া প্রয়োজন।

** 4।ধারাবাহিক ব্যাটিং পারফরম্যান্স: ** ব্যাটারদের তাদের ফর্মটি সন্ধান করা এবং ধারাবাহিকভাবে স্কোর করা দরকার।

** 5।শক্তিশালী বোলিং আক্রমণ: ** বোলারদের ভাল পারফরম্যান্স চালিয়ে যাওয়া এবং বিরোধীদের কম স্কোরগুলিতে সীমাবদ্ধ করা দরকার।

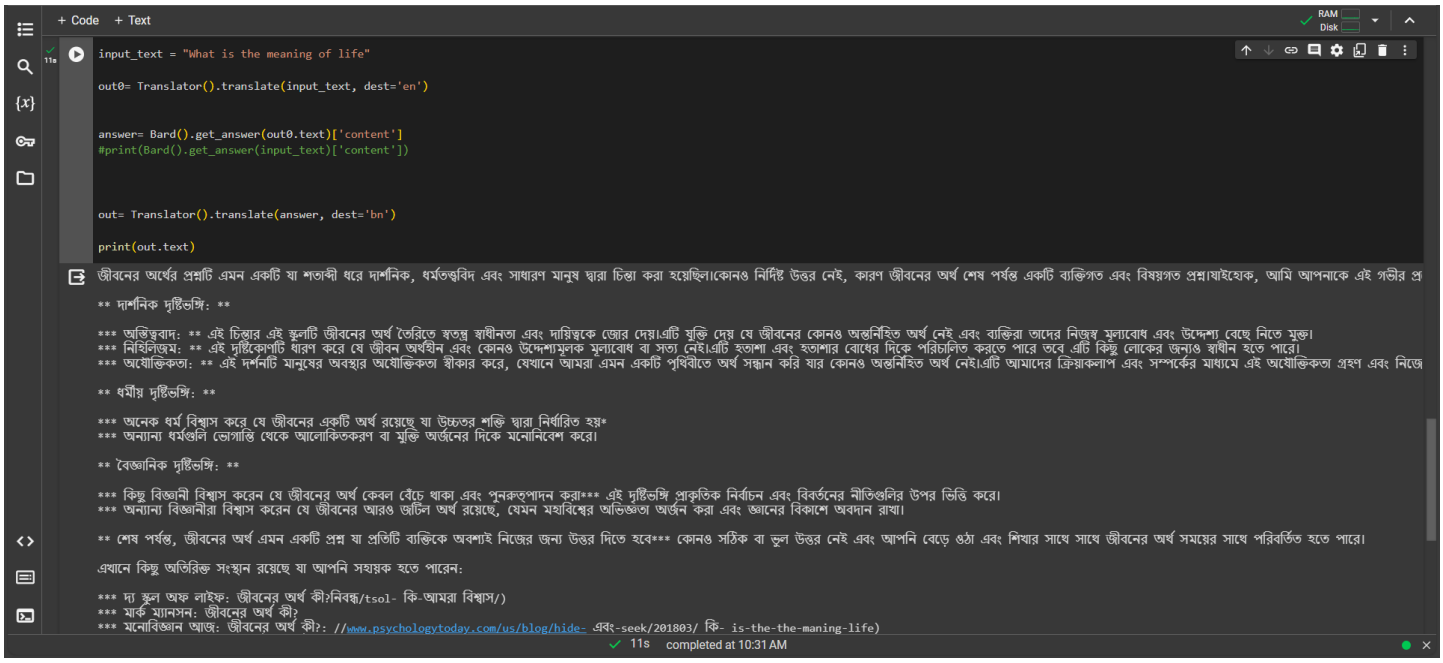
** 6।কৌশলগত সমন্বয়: ** দলটির বাকী ম্যাচগুলি অনুসারে তাদের কৌশল এবং কৌশলগুলি সংশোধন করতে হবে।

** 7।কিছুটা ভাগ্য: ** শেষ পর্যন্ত, বাংলাদেশকে সমস্ত কিছু জায়গায় পড়ার জন্য কিছুটা ভাগ্যের প্রয়োজন হবে।

বিবেচনা করার জন্য এখানে আরও কিছু অতিরিক্ত পয়েন্ট রয়েছে:

* বেশ কয়েকটি ক্ষতির পরে দলের মনোবল এবং আত্মবিশ্বাস কম হওয়ার সম্ভাবনা রয়েছে।
* তারা তাদের অবশিষ্ট ম্যাচগুলিতে পারফর্ম করার জন্য চাপের মুখোমুখি হবে।
* অন্যান্য দলগুলি বাংলাদেশের হতাশার বিষয়ে সচেতন হবে এবং এটি কাজে লাগানোর চেষ্টা করতে পারে।

3.



Further Possibilities in Developing Bangla Chatbot Tool

In conclusion, we already mentioned and demonstrated the accuracy of the models used in Google BARD API. We further explained that one of the main reasons that the BARD works so well is because of the fact that it was trained on an enormous number of datasets.

So, one of the most significant future possibilities in developing Bangla chatbot tool is that we can incorporate the models to train a large size of dataset in Bangla, as this: https://raw.githubusercontent.com/csebuetnlp/banglabert/master/question_answering/sample_inputs/train.json

By doing so we will be able to eliminate the need of using a translator in building a chat bot, rather we will be able to generate responses based on training and testing in Bangla. We also hope that this will be further convenient and satisfactory for the Bangla speaking users. Our plan of future improvement includes the feature of inserting images as input and derive out meaningful replies based on questions asked regarding the input image.