# OPEN GPT3

**Abstract:** Trying to **evaluate the Indic language translation**(like Hindi, Marathi) in comparison with Spanish and other popular languages. We tried three different languages Spanish, Hindi, Marathi to translate with the same training dataset sentences. Spanish translation performed better than Hindi or Marathi output because it looks like GPT3 is not properly trained for **Indic languages.**

## Background:

GPT3(Generative Pre-trained Transformer 3) includes Deep Learning Process. It is a pre-trained model. GPT-3 is substantially more powerful than its predecessor, GPT-2. Both language models accept text input and then predict the words that come next. But with 175 billion parameters, compared to GPT-2's 1.5 billion, GPT-3 is the largest language model yet.

Requirements for GPT3:

1. API key which is provided by Open AI
2. If you want to try the GPT3 in python
   - Basic Python
   - Open AI module installed

GPT3 can be used with either with help of Playground which is provided by OpenAI itself, or we can also take help of Python.
Let us understand how we can use them.

**1. GPT3 using Open AI Playground**, web-based interface.
Write input for the GPT3 in the large text area provided by them

1. Select the model which we want to use, preferably DaVinci as it is most advanced at this time. Later we can experiment to see which one gives the best result.
2. In the dropdown which is labelled as Load a preset select the preset which we want
3. Click submit for the output.
4. For the next input GPT3 will produce a prefix.

*Different attributes which are available with GPT3:*

*Temperature:* This is the most important setting to control the output of the GPT3. This setting controls the randomness of the generated text. A value of 0 makes the engine deterministic, which means that it will always generate the same output for the given input text. A value of 1 makes the engine take the most risks and a lot of creativity.
*NOTE: Don't add space at the end of the input string, GPT3 will behave unpredictably.*
Also set Top P(has some control over the randomness of the response) to 1 and also the rest of the value to their defaults.
Give some input to the GPT3 and press submit.
Now add prefix and press submit and let GPT3 do its work.
Try to increase the temperature value for giving the GPT3 more liberty.

*Response Length:* If you want to change the length of the output, it can be done by the changing value of this setting.(default setting is 64, which means GPT3 will add 64 tokens(word and/or punctuation) to the output.)

*Stop Sequence:* This parameter tells the GPT3 where it should stop generating output

*Max_tokens*: This is the total token the model will give in the output. Keeping this 0 will give an error of

NOTE: Length of the input(along with prompt) + max_token ≤ 2049  (2049 are maximum token allowed by the GPT3)

### 2. Using GPT3 using python:

This is the basic code that will help you to try the open gpt3 in python.

```
import openai
openai.api_key = API_KEY    #ADD Your API key here

prompt=” Add your Input data set here and also the line you want to translate”
#This will be your input

response = openai.Completion.create(engine="davinci", prompt=prompt,
stop="\n", temperature=0.3, max_tokens=500)   #This line provides the input
to the openai and creates the object response which contains the output of the
given prompt.
print(prompt.split(',')[-1].split('\n')[-2:][0])
print(u"{}".format(response.get('choices')[0].get('text'))) #get the output
#Printing the output
```

(You can also see our code at Google collab notebook: refer link 3)

## OUR FINDINGS

We tried three different languages Spanish, Hindi, Marathi to translate with the same training dataset sentences. Spanish translation performed better than Hindi or Marathi output because it looks like GPT3 is not properly trained for Indic languages.

We provided some general English sentences along with their translation in the languages as the training data to the model and the respective translation in Spanish, Hindi, and Marathi

## Training Data

| **English** | **Spanish** | **Hindi** | **Marathi** |
|---|---|---|---|
| I do not speak Spanish? | ¿No hablo español? | मैं स्पैनिश नहीं बोलता? | मी स्पॅनिश बोलत नाही? |
| See you later! | ¡Hasta luego! | बाद में मिलते हैं! | नंतर भेटू! |
| Where is a good restaurant? | ¿Donde se encuentra un buen restaurante? | एक अच्छा रेस्तरां कहां है? | चांगले रेस्टॉरंट कुठे आहे? |
| What rooms do you have available? | ¿Qué habitaciones tienes disponibles? | आपके पास कौन से कमरे उपलब्ध हैं? | आपल्याकडे कोणत्या खोल्या उपलब्ध आहेत? |
| The food here is delicious | la comida aqui es deliciosa | यहां का खाना स्वादिष्ट है | इथले जेवण स्वादिष्ट आहे |
| This is my hometown | Esta es mi ciudad natal | यह मेरा होमटाउन है | हे माझे मूळ गाव आहे |
| Will you please help me? | Por favor podría usted ayudarme? | क्या आप कृपया मेरी मदद करेंगे? | कृपया तुम्ही मला मदत कराल का? |
| Do you know this address? | ¿Conoce esta dirección? | क्या आप इस पते को जानते हैं? | तुम्हाला हा पत्ता माहित आहे का? |

These sentences were given to the model to translate

**PROMPT**

This place is beautiful!

What is the name of this place?

Where are you going?

## Expected Output

| **English(INPUT)** | **Spanish** | **Hindi** | Marathi |
|---|---|---|---|
| This place is beautiful! | ¡Este lugar es hermoso! | यह जगह सुंदर है! | हे ठिकाण सुंदर आहे! |
| What is the name of this place? | ¿Cuál es el nombre de este lugar? | इस जगह का नाम क्या है? | या जागेचे नाव काय आहे? |
| Where are you going | ¿A dónde vas? | तुम कहाँ जा रहे हो | आपण कोठे जात आहात |

## Output

| English(INPUT) | Spanish | Hindi | Marathi |
|---|---|---|---|
| This place is beautiful! | ¡Este lugar es hermoso! | यह स्थान बहुत ही प्यारी है! | या परिसरात सुंदर आहे! |
| What is the name of this place? | ¿Cuál es el nombre de este lugar? | यह क्या नाम है? | या जगात कोणत्याही नाव आहे? |
| Where are you going | ¿A donde vas? | आप कौन से जाना चाहते हैं? | तुमच्या जागी कुठे आहे? |

As it can be observed by comparing the expected output and the output that the Spanish translation is done nearly correct but for the Indic languages like Hindi, Marathi it is not behaving as per the need.

**The following might be the reason why the GPT3 is not performing well for the Indic Languages:**

1. GPT3 offers this type of training to the model:
   * no shot training
   * one shot training
   * a few shot training.

Here we have the max_token parameter which makes it difficult to train the model with a large dataset.

2. GPT3 doesn't retain what it has learned, so there is always a need to retrain it every time we are running the code.
3. As it is not trained sufficiently for INDIC language translations.

### References:

- https://www.twilio.com/blog/ultimate-guide-openai-gpt-3-language-model
- https://github.com/shreyashankar/gpt3-sandbox
- https://colab.research.google.com/drive/1KVFPmMW5Nu8YnIYm3D0dtiym3ZUBfoKk?usp=sharing