

IndiGo Hack to Hire Data Science (4010)

By: Aditya Pandey
adpandey2102003@gmail.com



Case Study Objectives

- To develop a robust NLP model capable of generating coherent and contextually accurate text.
- Fine - tuned on quora question answer dataset
- The bot is able to mimic human like interaction for answering wide range of queries.



Data exploration and Preprocessing

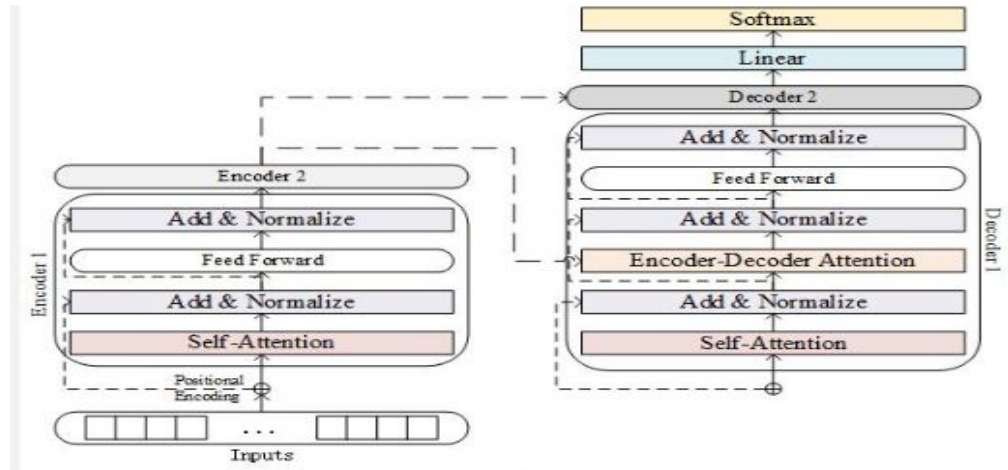
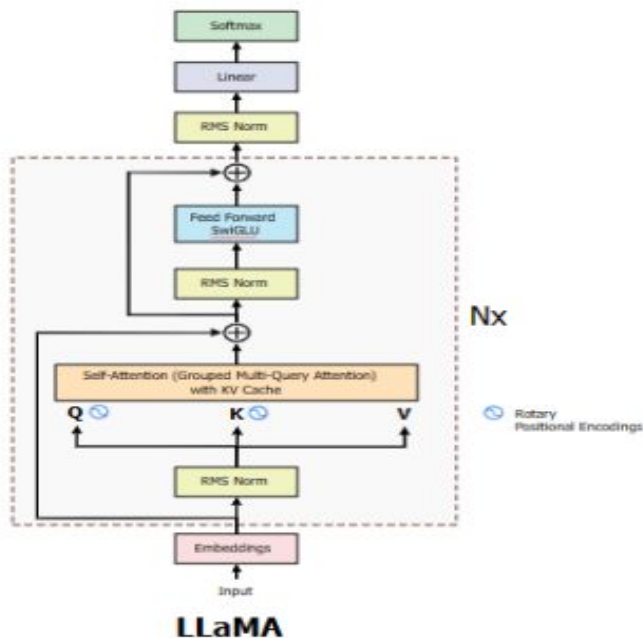
- Fine tuned using our dataset containing 56.4k question/answer pairs scraped from the question board Quora.
- For every question, there is one answer. Here are some sample rows from the dataset -

question (string)	answer (string)
"How do people who eat only once or twice a day manage to not feel hungry between meals?"	"The body adjusts after few days and even if they feel bit hungry it is not an issue. Humans have evolved in...
"Can you take control of your blood sugar?"	"Hey there.What if I told you that you could lower your blood sugar, lose weight, boost your energy and...
"Having at 62, I have no pension. I haven't payed enough in for old age pension and I am alone. What...	"Don't worry, you can claim Pension Credit which is about £180 a week plus a heating payment in the...
"What is the best opening sentence you have read in a book?"	"Some of the most enthralling opening lines from novels I have read and enjoyed are: Anna Karenina -...

- Created a function to prefix "answer the question:" to all of the questions, then tokenize them. Then it will tokenize the answers. The "inputs" for training the model will be the tokenized and prefixed 'questions', and the "labels" will be the answers.

Modal Selection

Google Flan T5 base - FLAN-T5 is an open-source, sequence-to-sequence, large language model that can be also used commercially. The model was published by Google researchers in late 2022, and has been fine-tuned on multiple tasks.



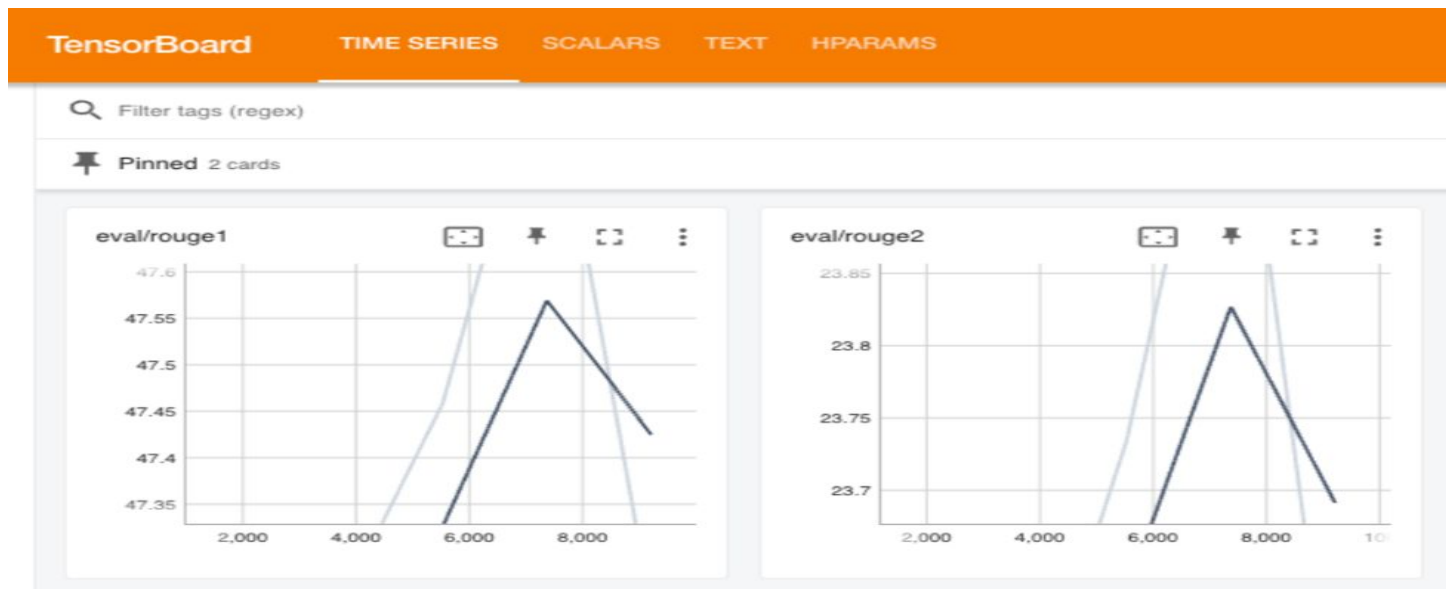
LLaMA (Large Language Model Meta AI) is a collection of state-of-the-art foundation language models ranging from 7B to 65B parameters. The foundation models were trained on large unlabeled datasets, making them ideal for fine-tuning on a variety of tasks. The model was trained on the following source: 67.0% CommonCrawl, 15% C4, 4.5% GitHub, 4.5% Wikipedia, 4.5% Books, 4.5% misc

Model Performance

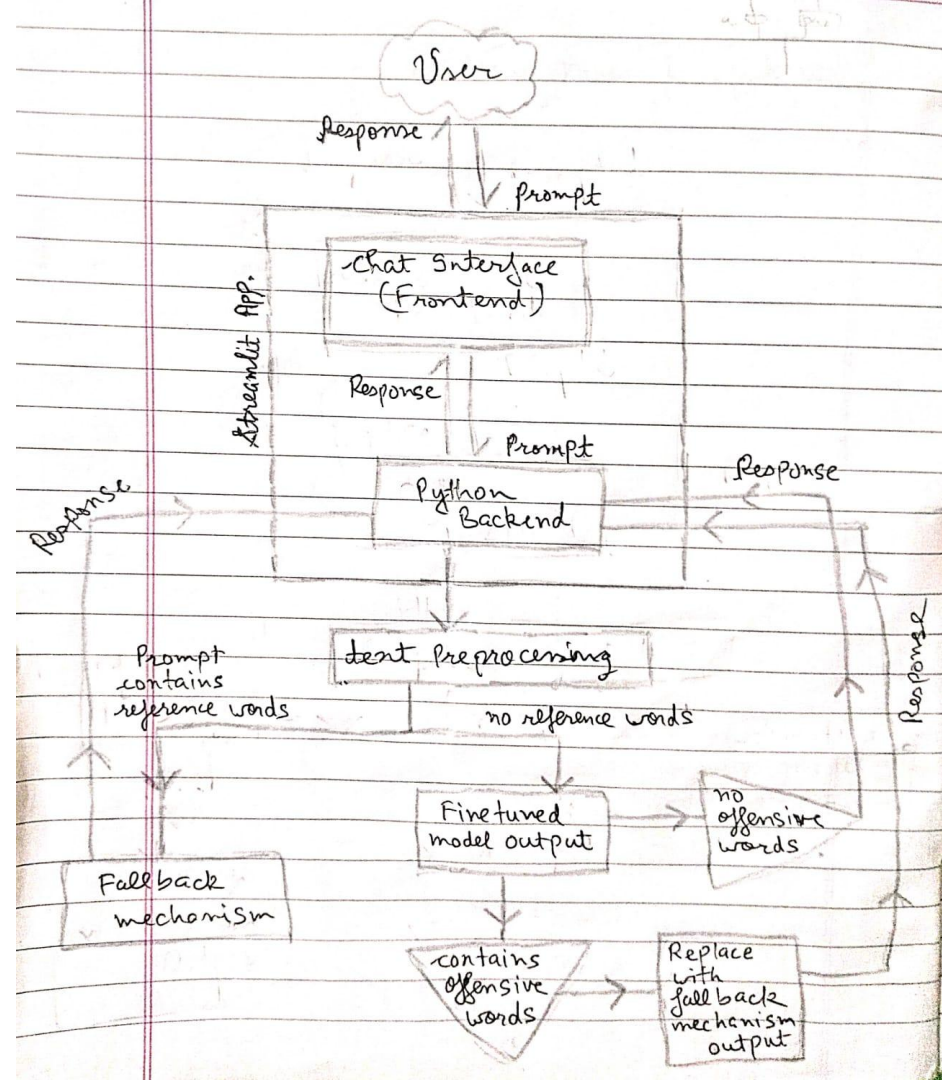
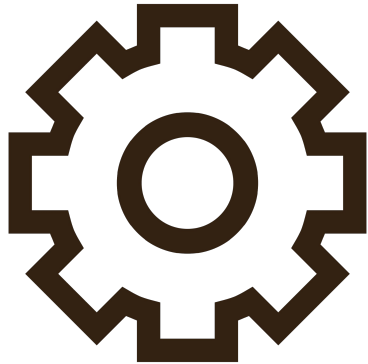
First Training

Epoch	Training Loss	Validation Loss	Rouge1	Rouge2	RougeL	RougeSum
1	3.153300	2.982270	0.111572	0.023446	0.090942	0.101183

Second Training



Application UML Diagram/Architecture



Insights and Recommendations


6.1 Key Insights


- **Model Performance:** T5 showed versatile performance even though it was only trained 1 epoch due to resource constraint.
- **LLama 2 -7b chat hf-** model is too large to finetune on local machine or google colab free tier resulting in halting of training mid-way due to GPU vram overflow.
- **Data Sensitivity:** The models were sensitive to data quality, highlighting the importance of thorough data cleaning and preprocessing, also different models have different requirements for training and inference data structures and format.
- **Model responses** - The bot can answer questions related to common phenomena and universal truths as they formed the majority in the quora dataset. But it stumbles while answering logical and subjective(directed towards it) questions due to the very same reason.


6.2 Recommendations and future scope

- **Data Augmentation:** To improve model robustness, consider using data augmentation techniques, such as paraphrasing and synonym replacement, to diversify training data.
- **Fine-tuning for Specific Tasks:** Further fine-tuning on domain-specific data can significantly enhance model performance in specialized areas. I could only run the fine-tuning for 1 epoch (cycle) due to resource constraint, while the recommended epochs for proper fine-tuning is around 5. If in future more computing power is available then the model performance can be greatly increased.
- **Hybrid Models:** Exploring hybrid models that combine the strengths of BERT's contextual understanding with GPT's generative capabilities could provide superior results. This can be done by allowing different models to compare/vote for the best output amongst themselves using some comparison mechanism or pass one model's answer through another so it can be further improved with both model's knowledge.
- **Retrain with varying hyperparameters** - One most common method is to use a variety of hyperparameters and observe the performance of the model, then choosing the best one. But again this is only possible with abundant computing power so that training the same model 5-6 times with different parameters can be done in parallel.

How can i help you...

 hello, how are you

 I am doing well, thank you for asking! I am excited to be able to help people with their tasks and to learn more about the world. How are you doing today?

 how can businesses increase sales

 Boost sales by leveraging your sales funnel. Here are some tips to help you increase sales: 1. Create a sales pipeline: Create an online sales channel that connects your customers with your products or services. This can be e-commerce, social media, or email marketing. 2. Use social marketing: Use reputable social networks to reach your target audience. These platforms can help increase your brand's visibility and attract new customers. 3. Use email campaigns: Email marketing is incredibly effective at increasing sales. Use emails to send out promotional messages and engage with potential customers and build recurring sales streams. 4. Use lead generation: Invest in lead nurturing programs like LeadGenerator, which can generate leads for your business. Lead generation is an important part of sales marketing, as it helps you build

|What is AI?



can you guide me through the steps to book a flight




Sure, here are the steps to book a flight:

1. Choose a website or app to book your flight. There are many different websites and apps that allow you to book flights, so you can choose one that is convenient for you.
2. Enter your travel information. This includes your departure and arrival cities, the dates of your trip, and the number of passengers.
3. Select your flight. Once you have entered your travel information, you will be able to select a flight that meets your needs. You can choose from different airlines, flight times, and prices.
4. Enter your personal information. This includes your name, address, and credit card information.
5. Review your booking. Once you have entered your personal information, you will be able to review your booking. Make sure that all of the information is correct before you proceed to payment.
6. Pay for your flight. You can pay for your flight with a credit card or debit card.
7. Receive your confirmation email. Once you have paid for your flight, you will receive a confirmation email. This email will include your flight information and a link to check in for your flight

|What is AI?



Conclusion



This project successfully explored the capabilities of various NLP models for text generation. Developing a state-of-the-art question-answering model leveraging the Quora Question Answer Dataset. I created an AI system capable of understanding and generating accurate responses to a variety of user queries, mimicking a human-like interaction. The findings suggest a path forward for improving model performance and practical applications, including deploying these models in real-world applications.

Now to Application Demonstration

