**Case Study**

***Represented By-***

**RAHUL MEHTA**

**Introduction -**

**Implementation of question-answering model leveraging the Quora Question Answer Dataset.**

**Data set used-** “toughdata/quora-question-answer-dataset”

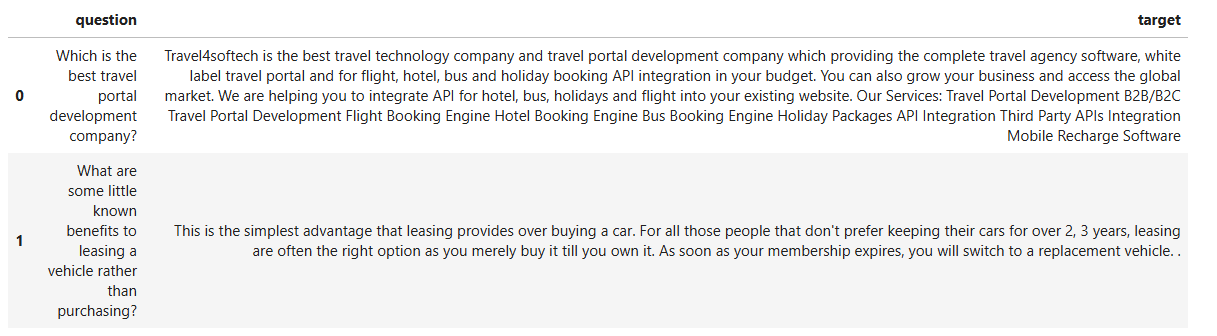
**Description of Dataset-**

The quora-question-answer-dataset data set extracted from quora for human like interaction uses contains 56k examples to train on LLM. It contains human response answers to the questions. The total data is split in train and validation set where 80% of rows were kept for train and remaining 20% is for validation round. Train data include 44k examples which were chosen randomly from full data set, remaining 11k rows were used for evaluation of matrix such as bleu score and rouge score.

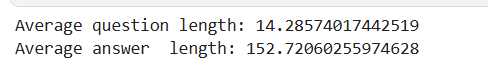
Data cleaning steps: -

* Removal of duplicate records in the data
* Removal of URL, linked text
* Removal of emoji character
* Removal of bullet points character
* Removal of extra space

Cleaned text data ->

****

Average lengths:

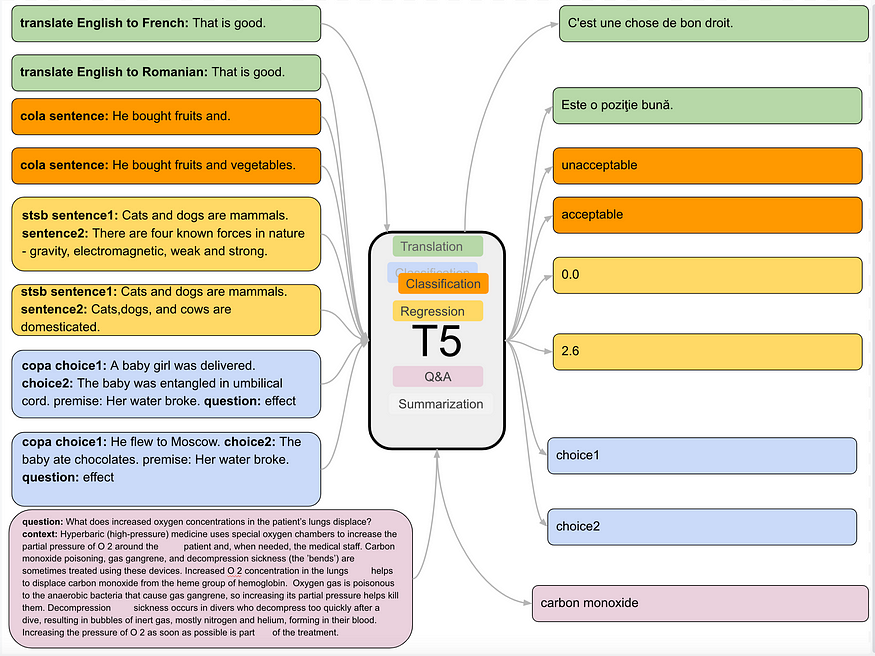


Total vocab from full data size: 101233

**Literature Survey** : -

Text to text generation:

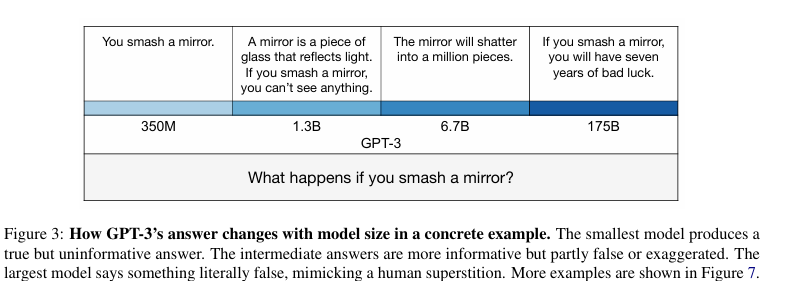
Clustered approach to solve multitasking problem in text to text generation where the tokens were used to specify for which segment it is training through token example <summarise >, <QA>, classification task can be performed in which out affecting other clustered during finetuning.



Reference: [20-074.pdf (jmlr.org)](https://jmlr.org/papers/volume21/20-074/20-074.pdf)

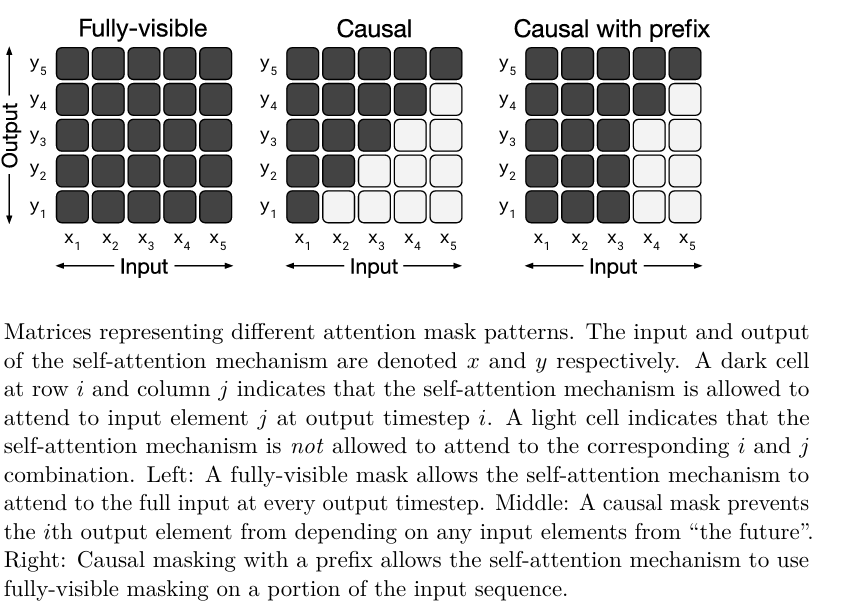
Measuring How Models Mimic Human Falsehoods:

* Usefulness for the TruthfulQA benchmark to evaluate the truthfulness of language models before deployment in critical applications.
* Development and incorporation additional training techniques that emphasize factual accuracy and reduce the generation of falsehoods.
* Continuously monitor and refine models to ensure they maintain high standards of truthfulness as they evolve.

****

Reference: [2109.07958 (arxiv.org)](https://arxiv.org/pdf/2109.07958)

Exploring the Limits of Transfer Learning:



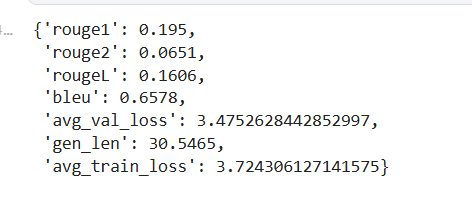
Reference: [20-074.pdf (jmlr.org)](https://jmlr.org/papers/volume21/20-074/20-074.pdf)

**Methodology**:

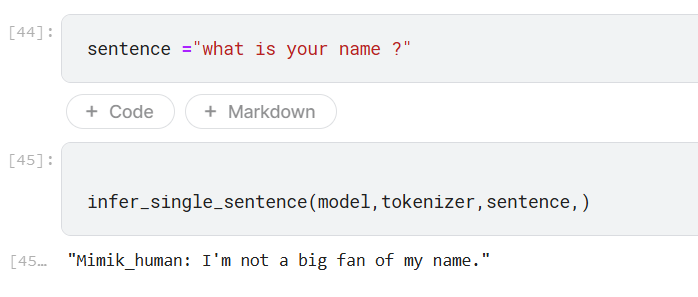
* Use Multitasking pretrained models (T5) to handle wide range of query/question model need to answer
* Freezing some of the perms of the model while finetuning to keep the context
* Use warmup steps for gradually learning the weights
* Configuration of optimiser for a set of weights such as ["bias", "LayerNorm.weight" for finetuning specific to a task

**Results**:

Model evaluation matrix after 5 epochs of training :



Inference of model for a given sentence



**Conclusion/insight and recommendations**:

* Inclusion of system which should measure the models Falsehoods in mimic of human
* Used of combined architecture model like t5 and bert
* Fine tuning of complex model with have high parameter.