Univ.AI
TA Bot

BY
ANAH VERONICA
HARSHITHA KOLUKULURU
JANHAVI KASHYAP
SHIVANI GB

AI-3 FINAL PROJECT | MAY 30, 2021

The problem

Help answer questions that are frequently asked by students in a learning setup.
We're given an FAQ dataset, need to build a chatbot to reduce workload from the team

Our Approach

Trying different datasets and as many models as we could for a better understanding

- Two Different Datasets
  - Cornell Movie-Dialogs corpus
  - The Stanford Question Answering Dataset (SQUaD)
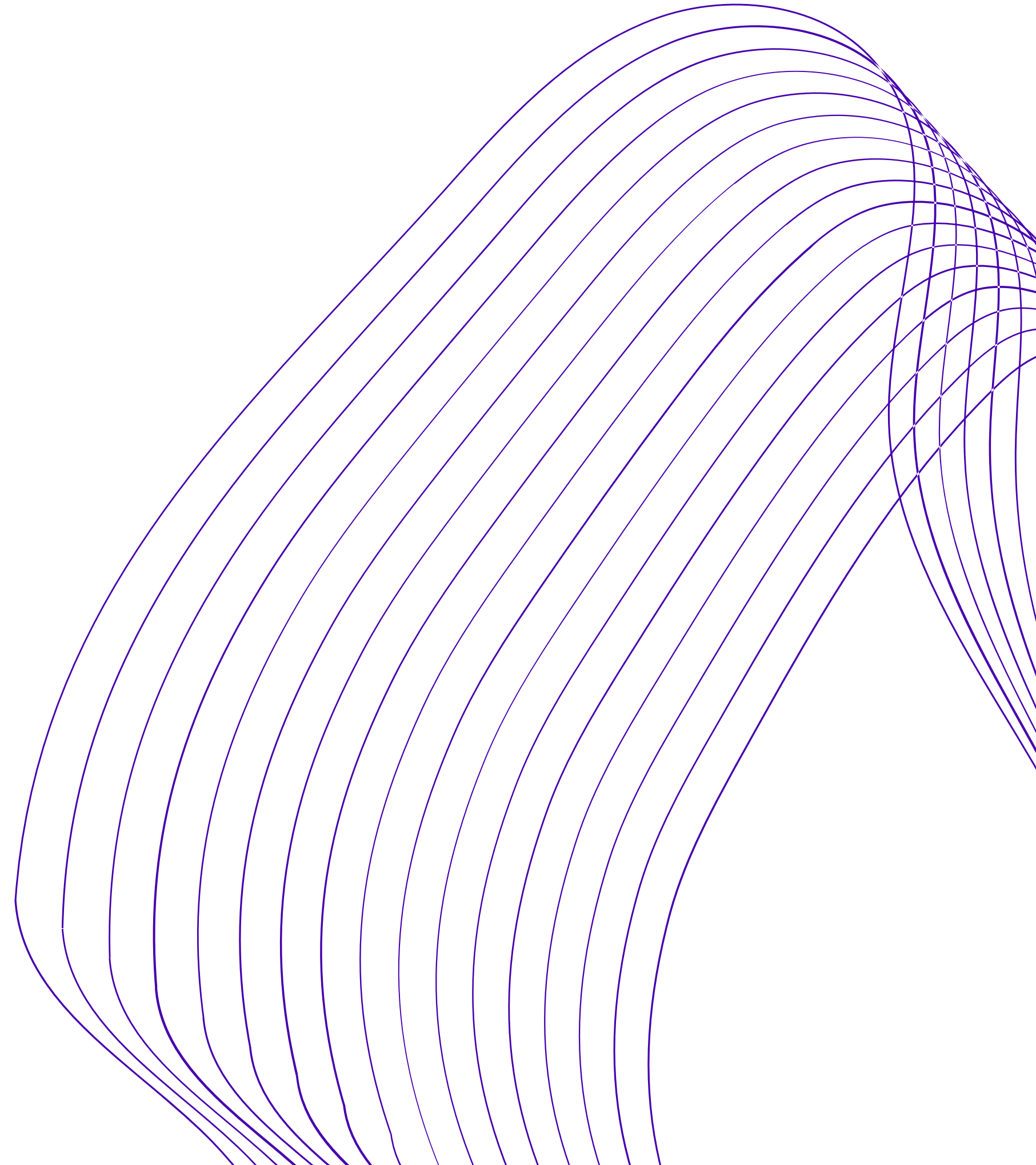
Stanford Question and answer dataset (SQuAD2.0)

- SQuAD2.0 is an extractive question-answering dataset. It's the good stuff!
- SQUAD has a context column - removed for the baseline and Seq2Seq models
- Took only 10,000 QA Sets
- Had to add a custom context tailored to the FAQ data in order to fine-tune our models after training on SQuAD
- Models trained:
  - Baseline Model
  - Seq2Seq
  - BERT
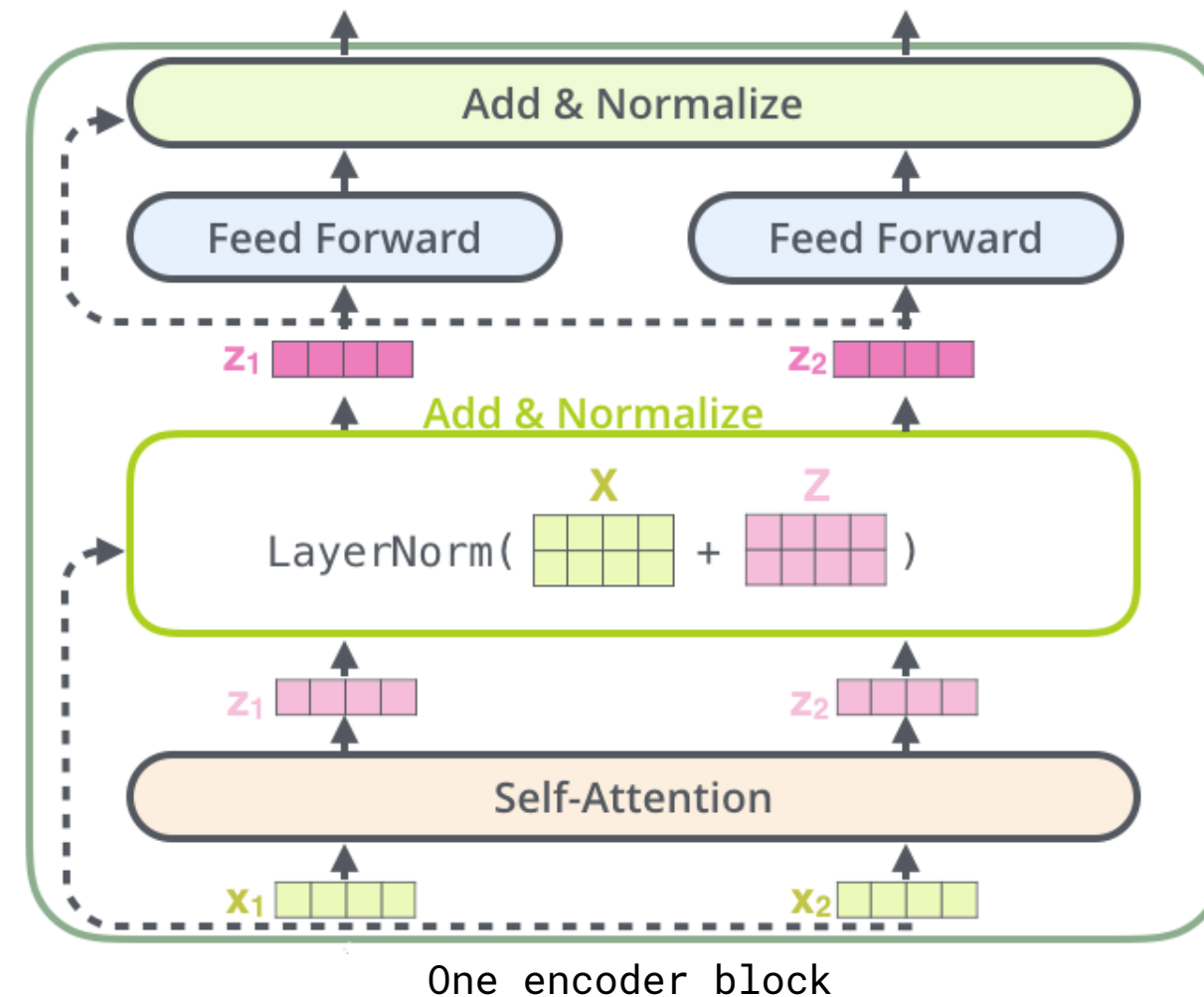
Article: Endangered Species Act
Paragraph: " ... *Other legislation followed, including the Migratory Bird Conservation Act of 1929, a 1937 treaty prohibiting the hunting of right and gray whales, and the Bald Eagle Protection Act of 1940. These later laws had a low cost to society—the species were relatively rare—and little opposition was raised.*"

**Question 1:** "*Which laws faced significant opposition?*"
Plausible Answer: *later laws*

Question 2: "*What was the name of the 1937 treaty?*"
Plausible Answer: *Bald Eagle Protection Act*

One encoder block

```
structure:
  article (dict)
  ├── title (str)
  └── paragraphs (dict)
      ├── context (str)
      └── qas (dict)
          ├── question (str)
          ├── id (str)
          └── answers (dict)
              ├── answer_start (integer)
              └── text (str)
  ...
  └── paragraphs n
      ├── context
      └── qas
```

Cornell Movie-Dialogs Corpus
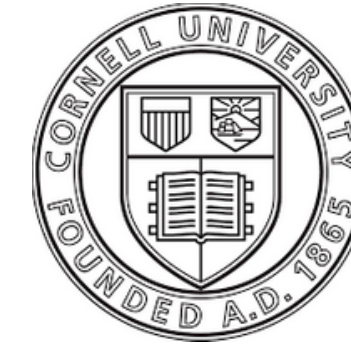
- It is a very common dataset to use for building chatbots
- Preprocessing: Filtering the data to have only statements which have a response, size restrictions, etc
- Models trained with this data (including the preprocessing steps, explaining the models, outputs
- Additional points, model architecture, theory, etc.
  - Baseline
  - Seq2Seq
  - Transformers

GPT-2

In addition, we fine-tuned a GPT-2 model on the FAQ data.

Results
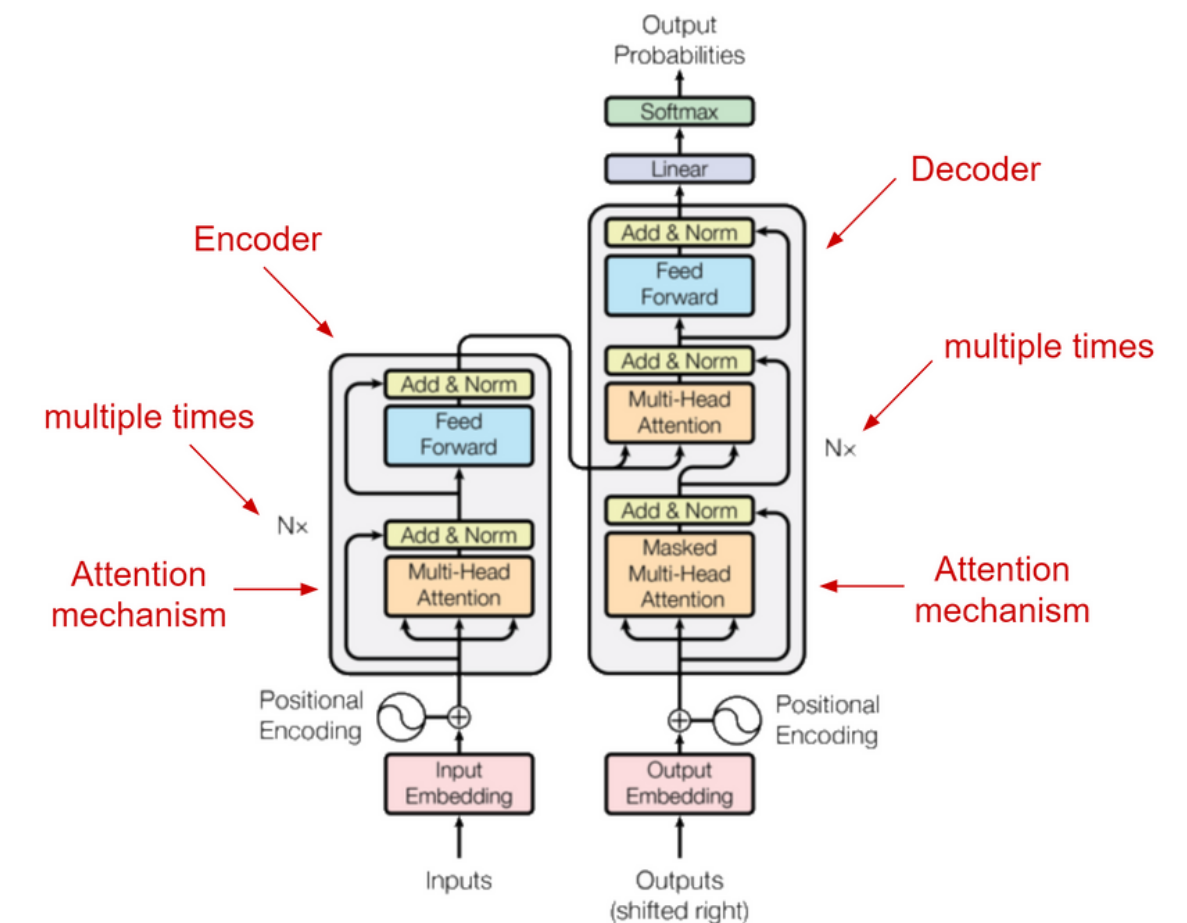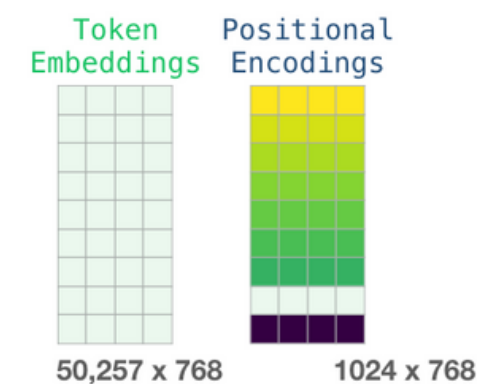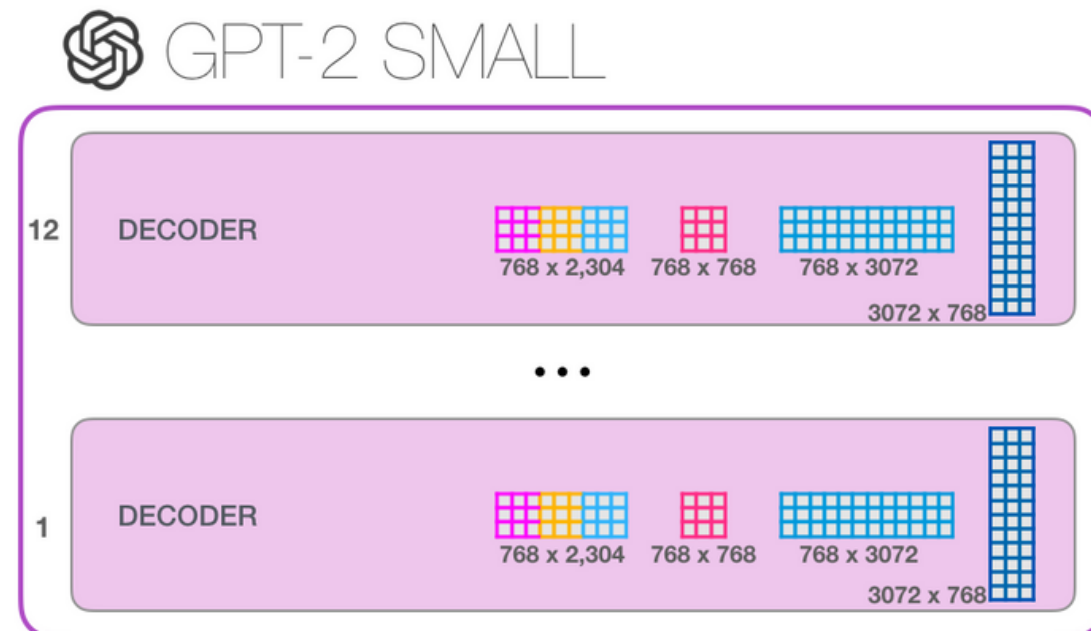
SQUaD Dataset   Cornell Dataset

RNN - Baseline Model

Basically had no predictive power but good accuracies on both datasets

Seq2Seq Model

Predictions didn't make a lot of sense, influenced greatly by the training data

BERT

Transformer Model

GPT-2

The runner-up

And the winner is BERT!

Could answer a few questions but not accurately enough. Attention *is* all you need.

```
Question:  What is the deadline for quiz submission?
Answer:  on the day of the following lecture . Extensions
---------------------------------------------

Question:  What is the deadline for exercise submission?
Answer:  on the day of the following lecture . Extensions
---------------------------------------------

Question:  How many hours do I need to complete this course?
Answer:  fifteen hours a week
---------------------------------------------

Question:  Who will grade the exercise?
Answer:  the edstem platform . You need to be
---------------------------------------------

Question:  Why is the auto-grader failing me?
Answer:  alternative solutions please follow the instructions . Ten percent of your grade
---------------------------------------------

Question:  Do I do the exercises individually?
Answer:  have to be done individually but can be discussed in breakout rooms
---------------------------------------------

Question:  Is the lab compulsory?
Answer:  are not graded
---------------------------------------------

Question:  Will the sessions be recorded?
Answer:  sessions will be recorded
---------------------------------------------
```

```
Input: What is the deadline for quiz submission?
Output: 5 PM on the day of the following lecture.
Input: What is the deadline for exercise submission?
Output: All material is available on edStem.
Input: How many hours do I need to complete this course?
Output: this is he ?
Input: Who will grade the exercise?
Output: The exercises are auto-graded once you click the submit button.
Input: Why is the auto-grader failing me?
Output: All material is available on edStem.
Input: Do I do the exercises individually?
Output: sir , everything is the night there before the first time someone hours.
Input: Is the lab compulsory?
Output: Please post a private message on edStem.
Input: Will the sessions be recorded?
Output: Yes. Each session has exercise and there is a homework after each week.
Input: Can I have access to the recorded videos?
Output: Please post a private message on edStem.
Input: Where are the recordings?
Output: Recording links will be posted on edStem.
Input: Where can I ask questions regarding reading material
Output: Please post a private message on edStem.
Input: Where to find course material?
Output: Please post a private message on edStem.
Input: Do we have homework?
Output: idea . . . find itknow .
Input: Where can I find the homework?
Output: All material is available on edStem.
```

```
<|question|>:Will the pre-class session be recorded?
<|answer|>:no preclass quiz is
<|question|>:What is the deadline for quiz submission?
<|answer|>:5 pm on the day
<|question|>:What is the deadline for exercise submission?
<|answer|>:5 pm on the day
<|question|>:How many hours do I need to complete this course?
<|answer|>:about 15 hours per week
<|question|>:Who will grade the exercise?
<|answer|>:the exercises are autog
<|question|>:Why is the auto-grader failing me?
<|answer|>:Because the autogr
<|question|>:Do I do the exercises individually?
<|answer|>:unfortunately the exercises are
<|question|>:Is the lab compulsory?
<|answer|>:labs are not compulsory
<|question|>:Will the sessions be recorded?
<|answer|>:no preclass quizzes
<|question|>:Can I have access to the recorded videos?
<|answer|>:no preclass quiz is
<|question|>:Where are the recordings?
<|answer|>:recording links will be
<|question|>:Where can I ask questions regarding reading material
<|answer|>:all written and analyzed solutions
<|question|>:Where to find course material?
<|answer|>:<|endoftext
<|question|>:Do we have homework?
<|answer|>:no preclass quiz is
<|question|>:Where can I find the homework?
<|answer|>:exercise submission time is
<|question|>:Where to submit the homework?
<|answer|>:exercise submission time is
```

Conclusion and Future work

I

We'd like to formally evaluate our models with established metrics.

II

We fine-tuned the GPT-2 model with the FAQ data which was just too small to make any conclusions about how it worked on QA tasks. Despite that, some of the outputs we got actually made a little sense. We'd like to train it on a bigger dataset with augmented data for longer.

III

Text generation and question-answering are really interesting areas and there are a lot of ways we could've made our models better.

Thank you!

We had a lot of fun working on this project!

Even the person who had to annotate the FAQ data

```
' to jail are continued with the holy offense with the holy offense <EOS>',
' to jail <EOS>',
' in the edstem lessons tab <EOS>',
' in the edstem lessons tab <EOS>',
' in the edstem lessons tab <EOS>',
' the exercises continued is careful to talk him the the submit button <EOS>',
' in the edstem lessons tab <EOS>',
' in the edstem lessons tab <EOS>',
' in the edstem lessons tab <EOS>',
' in the edstem lessons tab <EOS>',
' to jail are continued with the holy offense with the holy offense <EOS>',
' a jupyter notebook with the code <EOS>',
' a jupyter notebook with the code <EOS>',
' a jupyter notebook with the code <EOS>',
' the exercises continued is careful to talk him the the submit button <EOS>',
' the exercises for course least happens <EOS>']
```

```
[5]
<question>:where are the recordings
<|answer|>:www.youtube.com
====================
<question>:where are the recordings
<|answer|>:www.thetheft
====================
<question>:where are the recordings
<|answer|>:www.youtube.com
====================
```