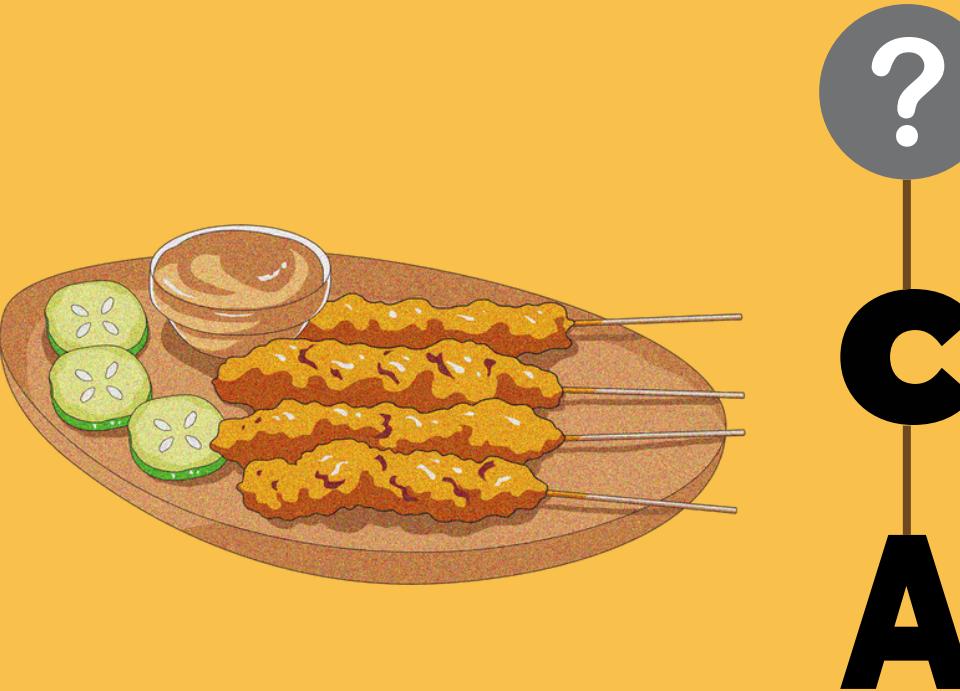


EXTRACTIVE QUESTION ANSWERING USING RECIPE DATA

Group 8 :
Dhruv Jain, Gianni Deguzman, Saumya Sinha, Tam
Huynh

BACKGROUND

- We were initially inspired by the RecipeQA paper
- Researchers scraped publicly available and non copy-written recipes
- Dataset is composed of recipes which include text and images
- Build models capable of question-answering using the different data types
- However, the original researchers focused on performing Cloze tasks (fill-in-the-blank)



Source: wgbh.org bestt recipe for beginners

Question Choose the best text for the missing blank to correctly complete the recipe
Cover. _____ . Bake. Cool, serve.

Answer A. Top, sprinkle B. Finishing touches C. Layer it up D. Ravioli bonus round

(Source: Yagcioglu et al. 2018 pg2)

MOTIVATION

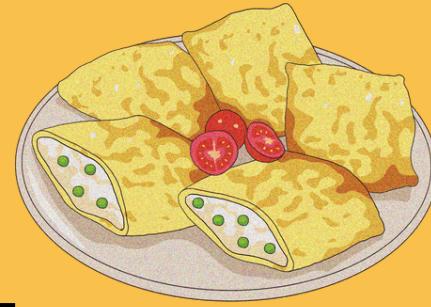
- We wanted to perform more traditional Question-Answering compared to the Cloze question answering in the paper
- Re-format the data to support models that perform extractive question answering
- Transformer-based project (BERT and BERT based models)

LITERATURE REVIEW



Name	Major Workings	Models Used	Results
Yagcioglu et al. (2018) RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipe	<ul style="list-style-type: none"> Developed the RecipeQA dataset Created baseline models 	Hasty Student approach, Impatient Reader approach	Hasty Student performed well in visual coherence and visual ordering, while Impatient Reader performed better in visual cloze and textual cloze.
Pearce et al. (2021) A Comparative Study of Transformer-Based Language Models on Extractive Question Answering	<ul style="list-style-type: none"> Performed experiments on 4 different QA datasets Compared various transformer models 	BERT, ALBERT, XLNet, RoBERTa, ConvBERT, BART (All base models)	<ul style="list-style-type: none"> RoBERTa performed best in NewsQA and SQuAD ConvBERT performed best in QuAC BERT-BiLSTM performed best in CovidQA
Liu et al. (2020) Multi-Level Multimodal Transformer Network for Multimodal Recipe Comprehension	<ul style="list-style-type: none"> Implemented Multi-Level Multi-Model Transformer 	MLMM Trans	MLMM Trans out performs PRN (Procedural Reasoning networks) in all tasks.

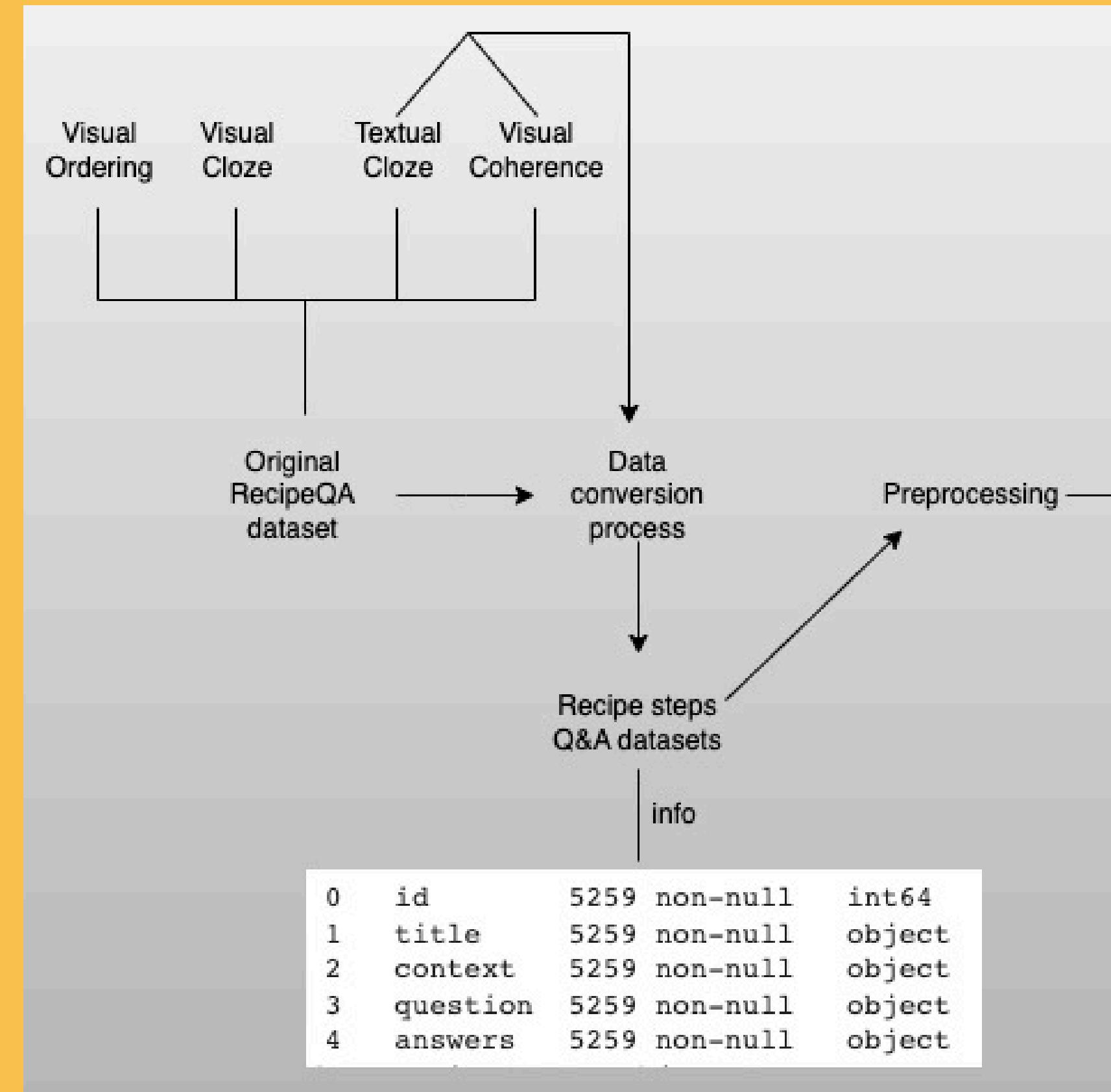
TECHNOLOGY SURVEY



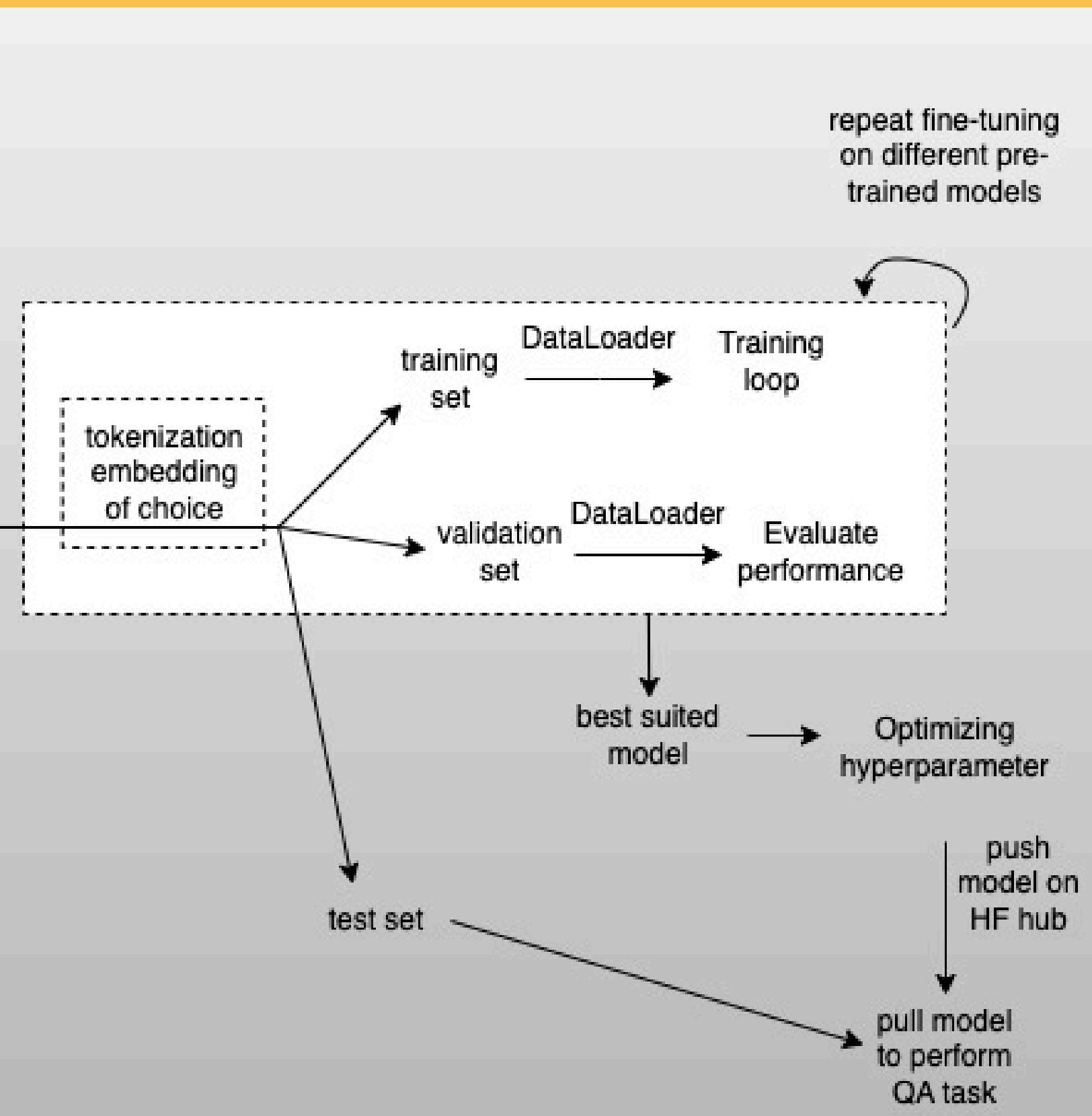
Name	Major Workings	Results
Devlin et al. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	<ul style="list-style-type: none">Introduced Bidirectional architecture	Obtained GLUE score of 80.5%, MultiNLI accuracy of 86.7%, SQuAD v1.1 QA Test F1 of 93.2%, SQuAD v2.0 Test F1 of 83.1%.
Lan et al. (2020) ALBERT: A lite BERT for self-supervised learning of language representations	<ul style="list-style-type: none">Introduced Cross-layer parameter sharingIntroduced Sentence order prediction	Performs at par (sometimes better) than BERT while reducing the number of parameters drastically.
He et al. (2021) DeBERTa: Decoding-enhanced BERT with disentangled attention	<ul style="list-style-type: none">Introduced disentangled attention mechanismImplemented enhanced mask decoder	Obtained MultiNLI score of 91.1%, SQuAD v2.0 score of 90.7%, RACE score of 86.8%

METHODOLOGY

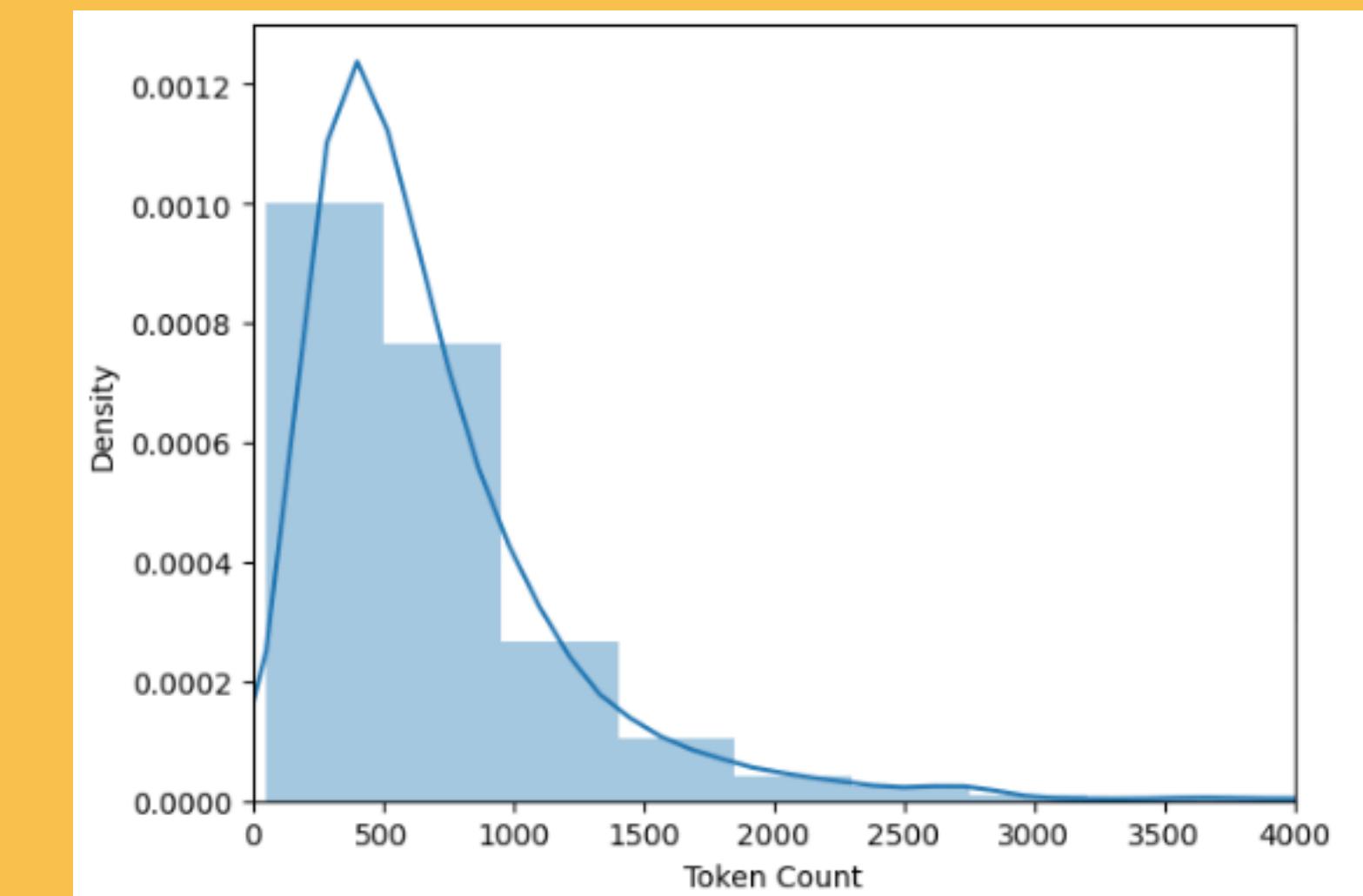
id	title	context	question	answers
1287	apple-hot-chocolate	The first step is Ingredients: You will need: Milk Nesquik/Hot Chocolate Mix Mug 2 Spoons Apple Jelly Measuring Cup Microwave. After the previous step is The Milk and Heating the Milk: In this step you will (A. Pour the milk and (B. Heat the milk in the microwave. A. Pour 1 cup of milk into a measuring cup. B. Put the measuring cup in the microwave for 1:00 Minute (or to specified temperature). Then stir.. After the previous step is Mix the Chocolate and Milk: In this step you will (A. Pour the Nesquik/Hot Chocolate mix into a mug and (B. Mix the milk with the Nesquik/Hot Chocolate Milk. A. Pour 2 tablespoons into the mug. B. Mix the hot milk and the chocolate together. Don't stir yet.. After the previous step is The Ceremony of Apple Jelly: Now is time for the moment of truth The Ceremony of the Apple Jelly... You will take 1 spoon and take a heaping spoon full of apple jelly and put into your hot chocolate. Now is the part where you stir it all together. Add apple jelly to taste....	What is the first step?	{"text": ["Ingredients"], "answer_start": [18]}
555	scallion-pancakes-1	The first step is Dough: First thing to do when making your scallion pancakes is to make the dough. I forget where I first got the scallion pancake recipe from, but it is very simple. It basically consists of 2.5 cups of flour and 1 cup of warm water. Mix the flour and water and knead until you get a slightly sticky ball of dough. I've experimented with both white flour and whole wheat flour and both work with these proportions. Once you get your ball of dough, stick it in the refrigerator for a bit to let it rest.. After the previous step is Scallions/Ginger/Other: While your dough is resting, chop the scallions for your scallion pancakes. I used an entire package of scallions as sold in my local supermarket (about six onions). I also added some chopped ginger root on this particular batch of scallion pancakes.. After the previous step is Make the Pancake: By the time you get your scallions chopped, the dough should be done resting. Take out the dough and section it out into small...	What is the step after Make the Pancake ?	{"text": ["Fry"], "answer_start": [1604]}



METHODOLOGY



Max input tokens: 22542
Min input tokens: 55



EXPERIMENTS



Model Names:

Deberta

Electra-Small

Electra-Large

albert-base-v1

Distilbert_with _RMSProp

DistilBert

albert-base-v2

nghuyong/ernie-2.0-base-en

xlm-roberta-base

facebook/bart-base

Google/bigbird-roberta-base

Roberta-base

squeezebert/squeezebert-uncased

bert-base

Model training parameters are:

- optimizer = AdamW
- learning rate=2e-5
- epochs = 3

Metrics used :

- **F1 score:**

- Harmonic mean of **precision and recall**.
- Precision represents the **accuracy of extracted answers**, while recall measures the **ability to retrieve all the relevant answers**.
- High F1 score suggests a successful balance between relevant response retrieval and accurate extraction.

- **Exact Match:**

- Known as **Strict match** or **Binary metric**
- Measures the **exact match accuracy** of a model's response against a reference or **ground truth answer**.

- **BLEU :**

- Compare the **n - gram of generated translation to the n - grams of the references**
- Ranges from 0 to 1 and higher scores indicate a higher degree of similarity

RESULTS



Best performing models:

Model	Exact Match	F1	BLEU
<i>Deberta</i>	62.58	65.3	0.56
<i>Albert-base-v1</i>	61.43	64.5	0.60
<i>Bart-base</i>	61.76	64.94	0.54

Optimizing Hyperparameter for Deberta:

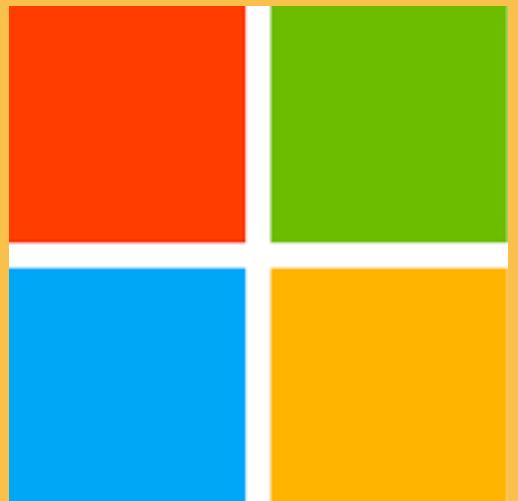
Hyperparameter tuned	Exact Match	F1	BLEU
learning rate = 2e-6, epochs=5	62.74	65.5	0.55

Example

```
question What is the step after Making the Crust ?  
answers {'text': ['The Filling'], 'answer_start': [623]}
```

```
context = """  
The first step is Making the Crust: For the crust: 10 graham cracker sheets (ie 20 individual squares) This makes about 1 1/2 cups cr  
crackers to make a fine powder. Melt the butter in the microwave and add in the butter, sugar, and water. Mix together until moistened  
minutes. Remove and let cool.. After the previous step is The Filling: Here is where you have a lot of freedom to choose any flavor f:  
keep it simple yet still very tasty. Following the instructions on the box I added 2 1/2 cups of milk and whisked until thickened. Po  
very fun step that requires a few common kitchen tools. You will need sandwich sized zip top bags (or pastry bags) and a piping tip w:  
purchased from any craft store (Michael's, Joannes, etc). Melt some of the white and dark chocolate discs and fill into a piping bag  
Using white chocolate, pipe the coordinate axis onto the pie, making sure to label both the x and y axis. For the center pipe out a c  
educational, and delicious pie you have created. Dollop each slice with whipped cream and remember to study your radian measures befo:  
"""  
question = "What is the step after Making the Crust ?"  
question_answerer(question=question, context=context)  
  
{'score': 0.45533502101898193,  
 'start': 623,  
 'end': 635,  
 'answer': ' The Filling'}
```

DISCUSSION : 3 BEST MODELS



DEBERTA

- Decoding-enhanced BERT with Disentangled Attention. - Microsoft Research 2020
- Build on BERT with their novel method of disentangled attention which separates the attention weights between the self-attention and task-specific attention - making it more capable of capturing task specific information
- "12-layer, 768-hidden, 12-heads, ~140M parameters" (Huggingface)

ALBERT-BASE-V1

- A Lite BERT - Google Research an Toyota Technological Institute at Chicago (2020)
- Researchers set out to address BERT's inclination for memory issues and training time
- Reduced total parameters through parameter-sharing between layers
- "12 repeating layers, 128 embedding, 768-hidden, 12-heads, 11M parameters" (Huggingface)

BART-BASE



- "Denoising autencoder for pre-training sequence-to-sequence models" - Facebook AI 2019
- Has key differences compared to BERT:
 - Designed more for summarization and machine translation (text generation)
 - Bidirectional encoder-decoder to allow output generation from both directions of the input sequence
 - Pretrained using denoising autencoder objective vs masked language- aim to reconstruct original input from noised versions
 - Left to right sequence-tosequence denoising autoencoder
- "12-layer, 768-hidden, 16-heads, 139M parameters" (Huggingface)

DISCUSSION: CHALLENGES AND LIMITATIONS

- Cloze tasks vs Question Answering
 - Recipe data is vastly different than other text data
 - Initial dataset and paper are focused on cloze tasks
 - Data required conversion
 - Not many similar papers to draw inspiration from
- Implementing models
 - Out-of-memory errors
 - High training time w/ GPU
 - Exhausting free-tier GPU in Colab
- Limited Dataset

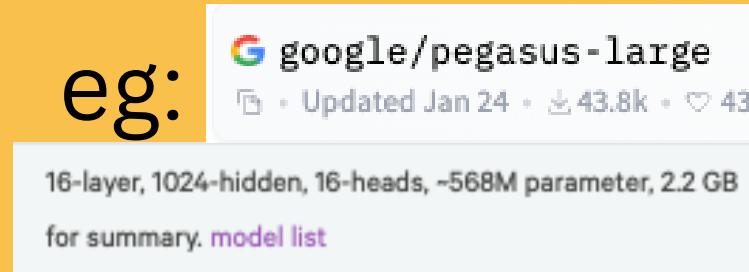
The first step is Ingredients Halal Vanilla Extract: 3 until 5 whole vanilla beans250 gram of vegetable glycerin food gradeEvery 100 gram of vanilla beans have 35 until 40 of whole vanilla beans. After the previous step is Scrape Vanilla Beans: Scrape Vanilla Beans and get the seeds into vegetable glycerin. After the previous step is Vegetable Glycerin and Vanilla Beans: Vanilla Beans Seed and Vegetable Glycerin. The last step is Vanilla Beans Can Use With Vegetable Glycerin: Whole Vanilla Beans put in a bottle with seeds and vegetable glycerin



FUTURE WORK



eg:



A

Can you tell me the
first step in making
adobo?

Sure, the first step
is to ...

EXPAND DATASET

- More recipes
- Different languages
- Expand question variety

MORE RESOURCES

- Faster computing
- Facilitate more training
- Larger models

FURTHER OPTIMIZATION

- Facilitate more training

GENERATIVE MODEL

- More robust question-answering
- Recipe chatbot

Team Member Work Distribution

Dhruv

- Modeling - DeBERTa-base, DeBERTa-large, ELECTRA-base, ELECTRA-large, DeBERTa hyper parameter tuning
- Slides: Literature Review, Technology Survey

Gianni

- Slides:Background/motivation, Discussion, Future Work
- Modeling: Bert-base, squeezebert-uncased, Albert-xlarge-v2, Albert-xxlarge-v2, pegasus-qa, xlnet-base-cased, Deberta tuning

Saumya

- Modeling Albert-base-v1, distilbert-base-uncased, distilbert-base-uncased hyper tuning , Deberta hyperparameter tunning
- Slides Experiments and results, demo

Tam

- Convert original dataset to QA dataset, preprocessing
- Modeling: roberta-base, albert-base-v2, nghuyong/ernie-2.0-base-en, xlm-roberta-base, facebook/bart-base, google/bigbird-roberta-base
- Slides: methodology, demo/example