




# Test Everything, Carefully: Generating Questions from Text

GA DSI 26 Capstone Project



*Test everything carefully; hold  
fast to that which is good*

**1 Thessalonians 5:21**





# Table of contents

## Background

- Research Problem
- Libraries used

0  
1

0  
2

## EDA/Methodology

- Corpus/word level EDA
- Modeling


## Results/Analysis

- Scoring metrics
- Problems with metrics

0  
3

0  
4

## Conclusions

- Conclusions
  - Limitations
  - Future Work
- 

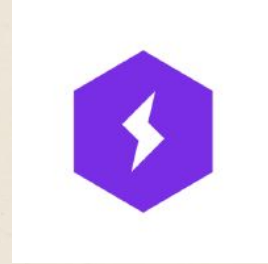


# *Background*

A brief introduction to NLP in Edtech

# Introduction

- help teachers automate tasks so that they can focus on the more human aspects of their jobs
- rise of MOOCs
- question and answer generation can populate question banks for various stakeholders





# Main problems in Question Generation

## Answer-aware

Given context and correct answer, generate question



## Answer-agnostic

Given context, generate question and answer pairs



## Distractors

Given context and question, generate correct answers and distractors (MCQ)



## Content Selection

Choosing important passages in a given text to generate questions from





# SQuAD v1.1

## Wikipedia articles

536 of the top 10,000 English articles



## Human Generated Questions

Reading comprehension questions were generated by human workers from the wikipedia articles



## Answers

Answers have to be a span in the text



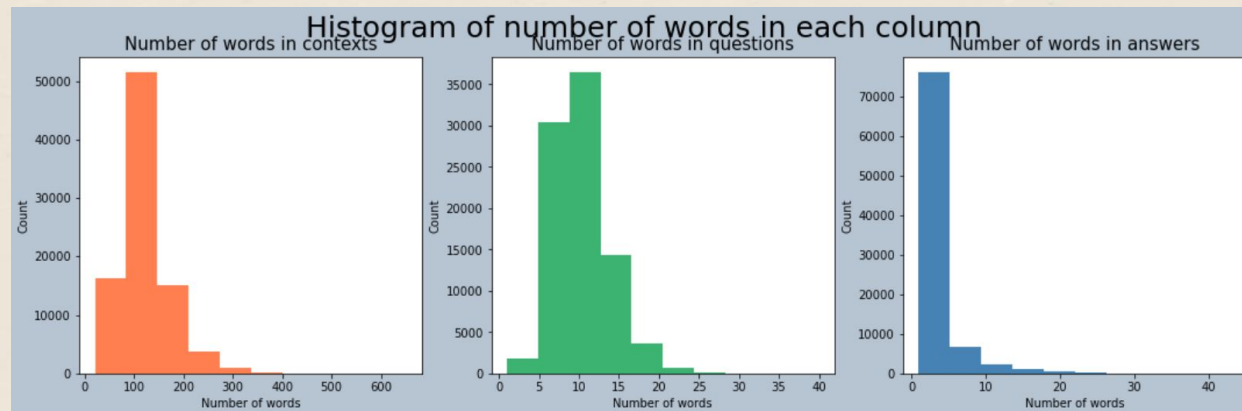
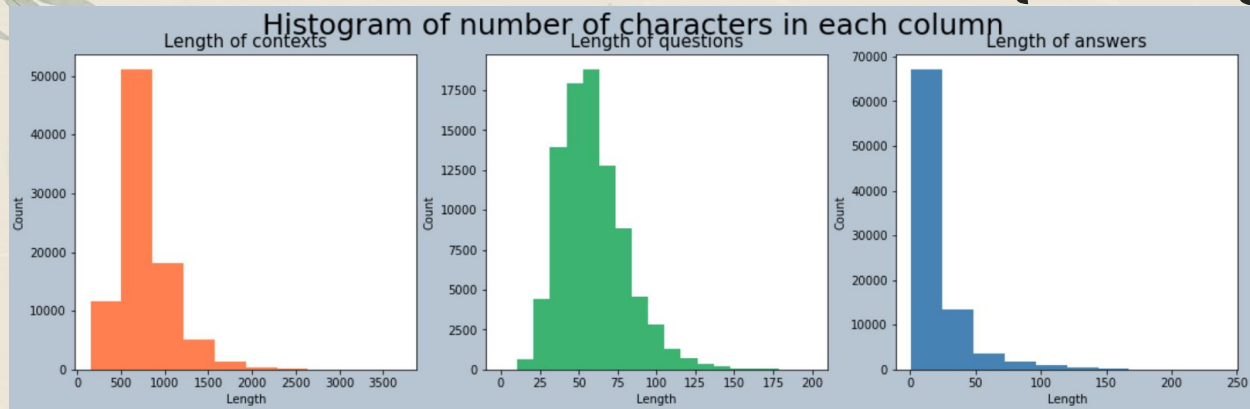


# *EDA/Methodology*

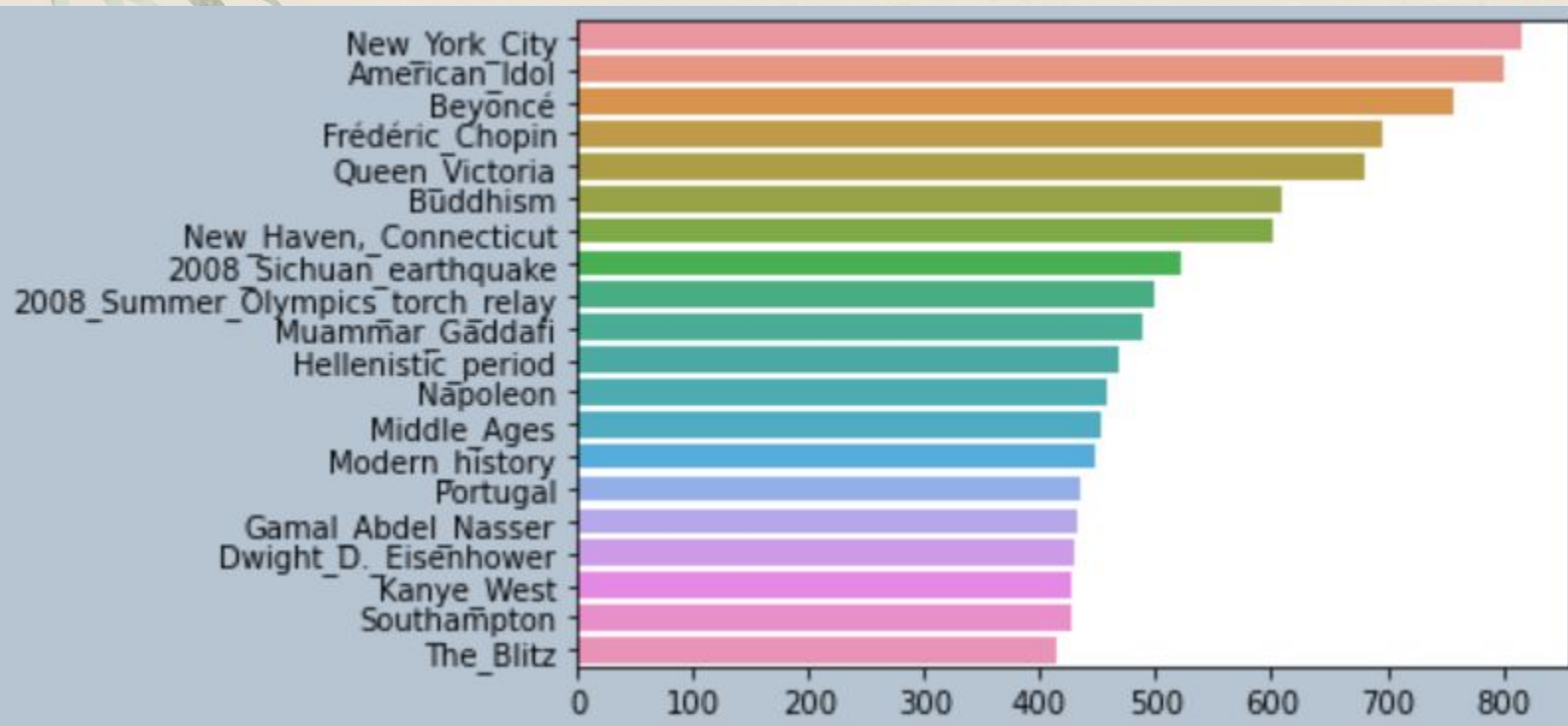
Charts and the Model



# Character and Word Frequency



# Topic Frequency



# NER Tagging

Using spaCy's pipeline, however there are many issues...

The university first **ORDINAL** offered graduate degrees, in the form of a Master of Arts (MA **WORK\_OF\_ART**), in the 1854–1855 academic year **DATE**. The program expanded to include Master of Laws **WORK\_OF\_ART** (LL.M.) and Master of Civil Engineering in its early stages of growth, before a formal graduate school education was developed with a thesis not required to receive the degrees. This changed in 1924 **DATE** with formal requirements developed for graduate degrees, including offering Doctorate (PhD **WORK\_OF\_ART**) degrees. Today **DATE** each of the five **CARDINAL** colleges offer graduate education. Most of the departments from the College of Arts and Letters **ORG** offer **ORG** offer PhD **WORK\_OF\_ART** programs, while a professional Master of Divinity (M.Div.) program also exists. All of the departments in the College of Science **ORG** offer PhD **WORK\_OF\_ART** programs, except for the Department of Pre-Professional Studies **ORG**. The School of Architecture **ORG** offers a Master of Architecture, while each of the departments of the College of Engineering **ORG** offer PhD **WORK\_OF\_ART** programs. The College of Business **ORG** offers multiple professional programs including MBA **WORK\_OF\_ART** and Master of Science in Accountancy programs. It also operates facilities in Chicago **GPE** and Cincinnati **GPE** for its executive MBA **WORK\_OF\_ART** program. Additionally, the Alliance for Catholic Education **ORG** program offers a Master of Education **WORK\_OF\_ART** program where students study at the university during the summer and teach in Catholic **NORP** elementary schools, middle schools, and high schools across the Southern United States **GPE** for two school years **DATE**.

# Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer

Colin Raffel\*

CRAFFEL@GMAIL.COM

Noam Shazeer\*

NOAM@GOOGLE.COM

Adam Roberts\*

ADAROB@GOOGLE.COM

Katherine Lee\*

KATHERINELEE@GOOGLE.COM

Sharan Narang

SHARANNARANG@GOOGLE.COM

Michael Matena

MMATENA@GOOGLE.COM

Yanqi Zhou

YANQIZ@GOOGLE.COM

Wei Li

MWEILI@GOOGLE.COM

Peter J. Liu

PETERJLIU@GOOGLE.COM

Google, Mountain View, CA 94043, USA

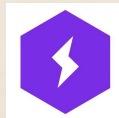
Editor: Ivan Titov

## T5 Model

- Seq2seq transformer model
- Trained on C4 corpus



# Model



## Pytorch Lightning

- A convenient wrapper for pytorch
- Allows for easy logging and monitoring of epochs

**You do the research. Lightning will do everything else.**

The ultimate PyTorch research framework. Scale your models, without the boilerplate.





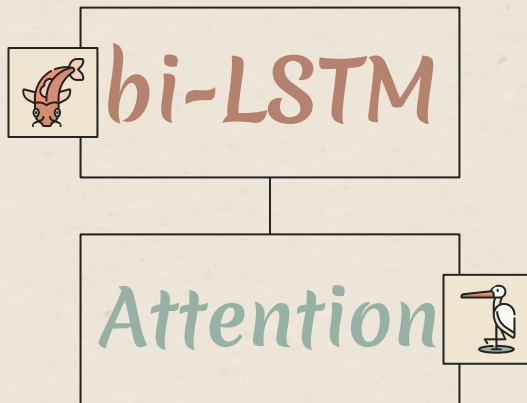
# Aside:

## What is a transformer?



### Bidirectional Long Short Term Memory

- Traditionally, RNNs suffered from vanishing gradient
- LSTM has 3 gates: input, output, and forget
- Bidirectionality = backpropagation
- Allows for memory to last longer



### Multi-headed Attention

The	The	The	The
cat	cat	cat	cat
drank	drank	drank	drank
the	the	the	the
milk	milk	milk	milk
because	because	because	because
it	it	it	it
was	was	was	was
hungry	hungry	sweet	sweet

Dark colors represent higher attention (Image by Author)

Words are fed in at one shot, attention vectors get passed into next block



# *Results/Analysis*

Metrics and Visual Inspection






## Scores

**Best Cross-entropy Loss: 1.23**

**Average BLEU Score: 0.101**



**Average Cosine Similarity: 0.841**

## Validation question, answer-aware

### Context

With International Criminal Court trial dates in 2013 for both President Kenyatta and Deputy President William Ruto related to the 2007 election aftermath, US President Barack Obama chose not to visit the country during his mid-2013 African trip. Later in the summer, Kenyatta visited China at the invitation of President Xi Jinping after a stop in Russia and not having visited the United States as president. In July 2015 Obama visited Kenya, as the first American president to visit the country while in office.

### Answer

US President Barack Obama

### Generated Answer/Question

```
{'Answer': 'US President Barack Obama', 'Question': 'Who chose not to visit Kenya during his 2013 African trip?'}
```

### Original Question

Who decided not to come visit the country in 2013?

### Scores

BLEU: 0.11547005383792518

Cosine Similarity: 0.9212935566902161

## Validation question, answer-agnostic

### Context

With International Criminal Court trial dates in 2013 for both President Kenyatta and Deputy President William Ruto related to the 2007 election aftermath, US President Barack Obama chose not to visit the country during his mid-2013 African trip. Later in the summer, Kenyatta visited China at the invitation of President Xi Jinping after a stop in Russia and not having visited the United States as president. In July 2015 Obama visited Kenya, as the first American president to visit the country while in office.

### Answer

[MASK]

### Generated Answer/Question

```
{'Answer': 'President Xi Jinping', 'Question': 'Who was the first American president to visit Kenya?'}
```

### Original Question

Who decided not to come visit the country in 2013?

### Scores

BLEU: 0.1052060490523318

Cosine Similarity: 0.8921566009521484

## Test question, answer-aware

### Context

Born in Seattle, Washington, Hendrix began playing guitar at the age of 15. In 1961, he enlisted in the US Army, but was discharged the following year. Soon afterward, he moved to Clarksville then Nashville, Tennessee, and began playing gigs on the chitlin' circuit, earning a place in the Isley Brothers' backing band and later with Little Richard, with whom he continued to work through mid-1965. He then played with Curtis Knight and the Squires before moving to England in late 1966 after bassist Chas Chandler of the Animals became his manager. Within months, Hendrix had earned three UK top ten hits with the Jimi Hendrix Experience: 'Hey Joe', 'Purple Haze', and 'The Wind Cries Mary'. He achieved fame in the US after his performance at the Monterey Pop Festival in 1967, and in 1968 his third and final studio album, *Electric Ladyland*, reached number one in the US. The double LP was Hendrix's most commercially successful release and his first and only number one album. The world's highest-paid performer, he headlined the Woodstock Festival in 1969 and the Isle of Wight Festival in 1970 before his accidental death in London from barbiturate-related asphyxia on September 18, 1970.

### Answer

'Hey Joe', 'Purple Haze', and 'The Wind Cries Mary'

### Generated Answer/Question

```
{'Answer': "'Hey Joe', 'Purple Haze', and 'The Wind Cries Mary'", 'Question': 'What three hits did Hendrix have in the UK?'}
```

## Test question, answer-agnostic

### Context

Born in Seattle, Washington, Hendrix began playing guitar at the age of 15. In 1961, he enlisted in the US Army, but was discharged the following year. Soon afterward, he moved to Clarksville then Nashville, Tennessee, and began playing gigs on the chitlin' circuit, earning a place in the Isley Brothers' backing band and later with Little Richard, with whom he continued to work through mid-1965. He then played with Curtis Knight and the Squires before moving to England in late 1966 after bassist Chas Chandler of the Animals became his manager. Within months, Hendrix had earned three UK top ten hits with the Jimi Hendrix Experience: 'Hey Joe', 'Purple Haze', and 'The Wind Cries Mary'. He achieved fame in the US after his performance at the Monterey Pop Festival in 1967, and in 1968 his third and final studio album, Electric Ladyland, reached number one in the US. The double LP was Hendrix's most commercially successful release and his first and only number one album. The world's highest-paid performer, he headlined the Woodstock Festival in 1969 and the Isle of Wight Festival in 1970 before his accidental death in London from barbiturate-related asphyxia on September 18, 1970.

### Answer

[MASK]

### Generated Answer/Question

```
{'Answer': '1969', 'Question': 'In what year did Hendrix headline the Isle of Wight Festival?'}
```





# *Conclusions*

What could be improved upon?



# Conclusions

## *Answer-agnostic needs some work*

The answer-aware generation does relatively well but the answer-agnostic needs more work



## *Need for better scoring metrics*

The current scoring metrics are problematic, as we will see



## *Identifying question types*

Work could be done on identifying what kind of question should be generated (5W1H)

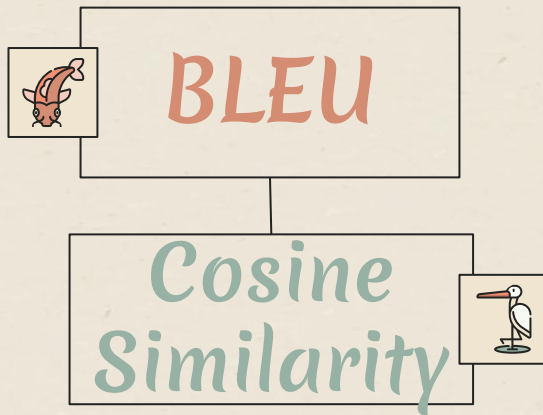


# Aside: Scoring Metrics



## BiLingual Evaluation Understudy

- Traditionally, used for machine translation
- Uses n-grams
- Doesn't capture context
- Very common in the literature



## Cosine Similarity

- Similarity between phrases in an vector space for word embeddings
- Considers context for word embeddings



# Scoring Literature Review

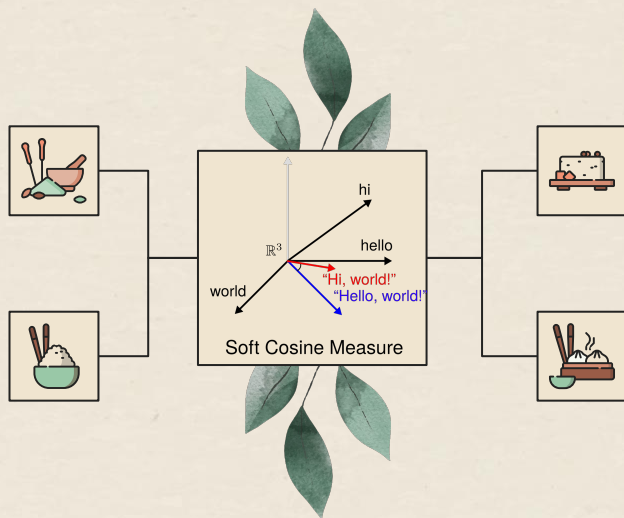
Scialom, 2019

5 elements of a good question:

- Answerability
- Relevance
- (grammatical) Correctness
- Soundness
- Fluency

Nema, 2018

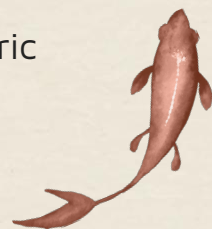
Add weights to words based on question type



Callison-Burch, 2006

"BLEU cannot guarantee correlation with human judgments"

Liu, 2016

- Embedding-based metrics
  - Consider context
  - Use human-centric metrics
- 

# Limitations

## Types of questions

Only for short comprehension answers directly from text,  
not for inference questions



## Scoring metric

A better scoring metric is needed



# Future work

## *Additional linguistic features*

Dependency parsing, better NER, spellchecker might help improve model performance



## *Better scoring metrics*


To come up with a better scoring metric that can capture the 5 elements



## *Multilingual*

Training on datasets from other mother tongues





知是行之始，行是知之成。  
*Knowledge is the beginning of practice;  
doing is the completion of knowing*

Wang Yangming







# Thanks!

**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

**Detailed bibliography in notebooks**