

CSE-842 Natural Language Processing
Dr. Kordjamshidi



Title Generation

Kuldeep Singh, Reza Khan Mohammadi, Ye Ma

Table of Contents

- The Problem
- The Solution
- The Outcome
- Conclusion




Leave me
some
pizza !

The Problem



more than **5,000**
articles are published
online every day





On average,
8 out of **10** people
only read the headline.



Oh... I have a new email... let me see...



noreply@mail.d2l.msu.edu

Activity summary for FS22-C... 2:44

Activity summary for FS22-CSE-842-...

Oh... I have a new email... let me see...



noreply@mail.d2l.msu.edu

Activity summary for FS22-C... 2:44

Activity summary for FS22-CSE-842-...



Mattison, Vincent

PhD Defe... Thu 9:39

The Department of Computer Scienc...

Oh... I have a new email... let me see...



noreply@mail.d2l.msu.edu

Activity summary for FS22-C... 2:44

Activity summary for FS22-CSE-842-...



Mattison, Vincent

[REDACTED] PhD Defe... Thu 9:39

The Department of Computer Scienc...



Colbry, Katy

Monday Motivation #127 Mon 11/28

Happy Monday! We had a "small" T...

Oh... I have a new email... let me see...



Colbry, Katy



> local mailing address?

Thu 15:23

Hi, Reza – It's just a welcome note; n...

Textual Bias can be deadly.



BREAKING: Donald Trump, who tried to overthrow the results of the 2020 presidential election and inspired a deadly riot at the Capitol in a desperate attempt to keep himself in power, has filed to run for president again in 2024.



npr.org

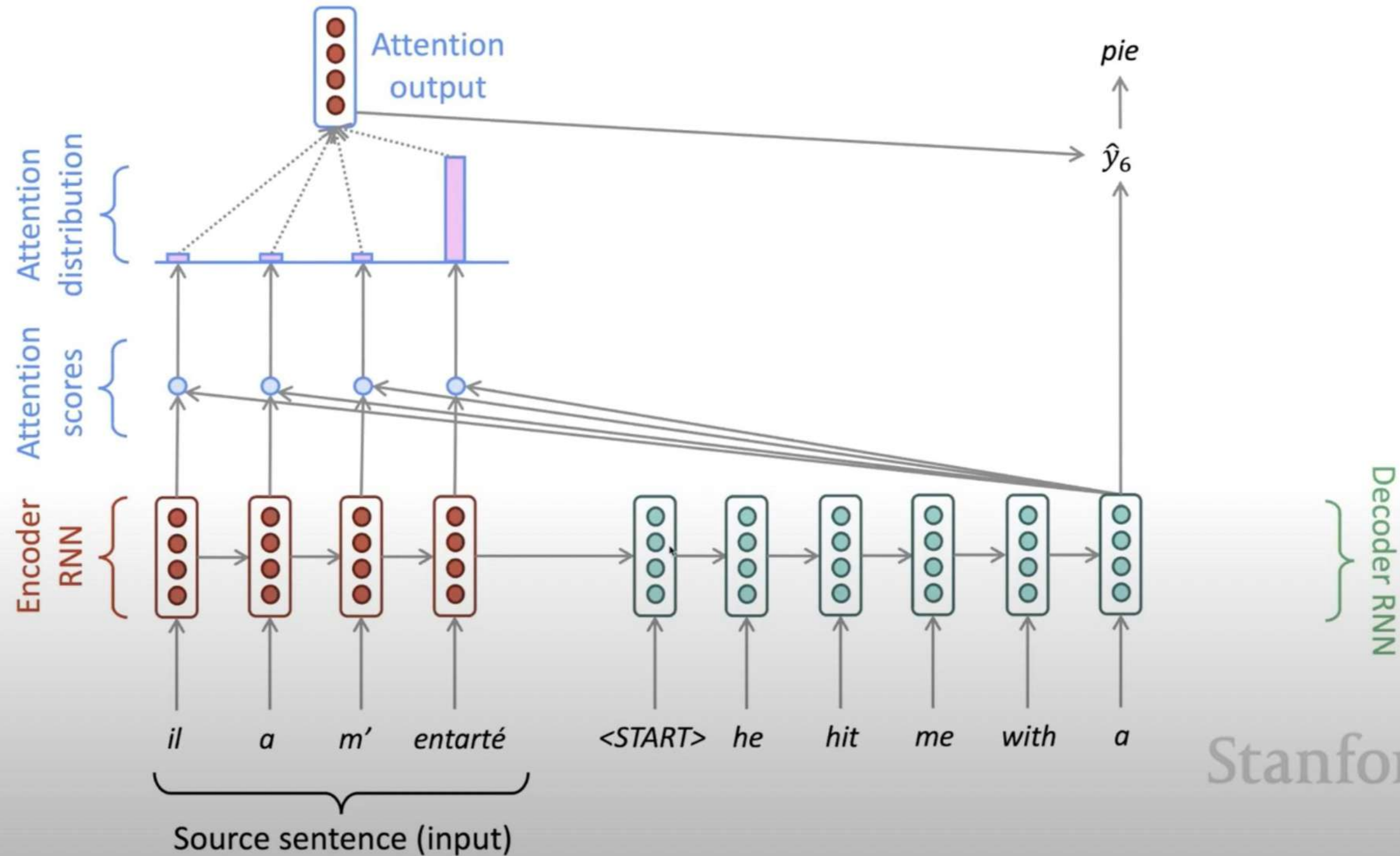
With midterm losses, Trump's climb to the nomination could be steeper than he... Donald Trump is officially running again, trying to avenge his loss to Joe Biden, even as Trump still refuses to admit he lost. Trump's push to overturn the 2020...

9:08 PM · Nov 15, 2022

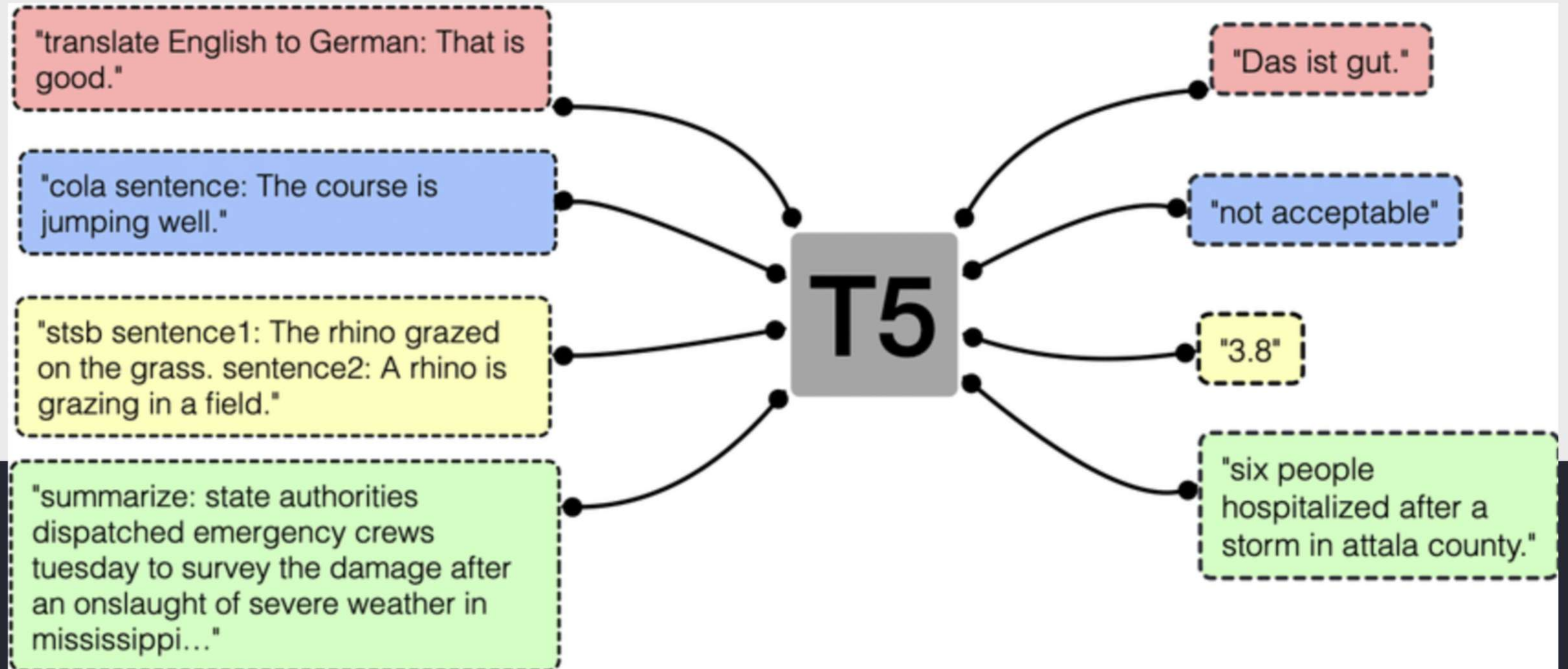
The Solution



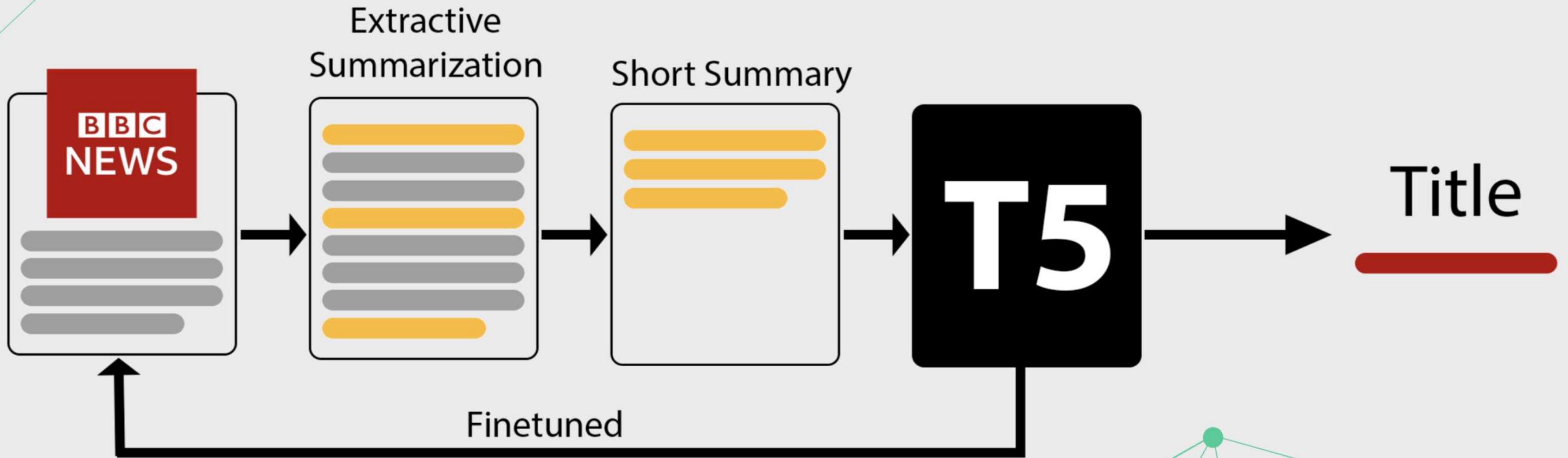
Baseline: Sequence-to-sequence + Attention



T5



Our Method



The Outcome



Table 1: A comparison between different title generation models.

Text Portion	ROUGE-1			ROUGE-2			ROUGE-L		
	r	p	f	r	p	f	r	p	f
Sequence-to-sequence	0.032	0.1776	0.0546	0.0	0.0	0.0	0.0323	0.1776	0.0546
Sequence-to-sequence + Attention	0.19	0.24	0.21	0.10	0.10	0.10	0.19	0.23	0.21
T5 (finetuned)	0.3582	0.3616	0.3560	0.1220	0.1230	0.1211	0.3405	0.3442	0.3386
T5 (First paragraph only)	0.3337	0.3372	0.3308	0.0989	0.1009	0.0984	0.3166	0.3208	0.3142
T5 + Short Summary	0.2971	0.3051	0.2965	0.0887	0.0911	0.0879	0.2844	0.2923	0.2838

Table 2: Human evaluation results.

Method	Fluency	Informativeness
T5 + pretrained weights	1.5234	1.8657
T5 + finetuned weights	2.2885	2.2684
T5 + First Paragraph	2.4362	2.3758
T5 + Short Summary	2.4362	2.2348



Ground-truth title

T5 pretrained

T5 finetuned

T5 + First Paragraph

T5 + Short Summary

Oasis star fined for German brawl

oasis singer liam gallagher has been fined euros after a fight in

oasis singer gallagher fined

oasis singer fined euros

oasis singer gallagher fined





Ground-truth title

T5 pretrained

T5 finetuned

T5 + First Paragraph

T5 + Short Summary

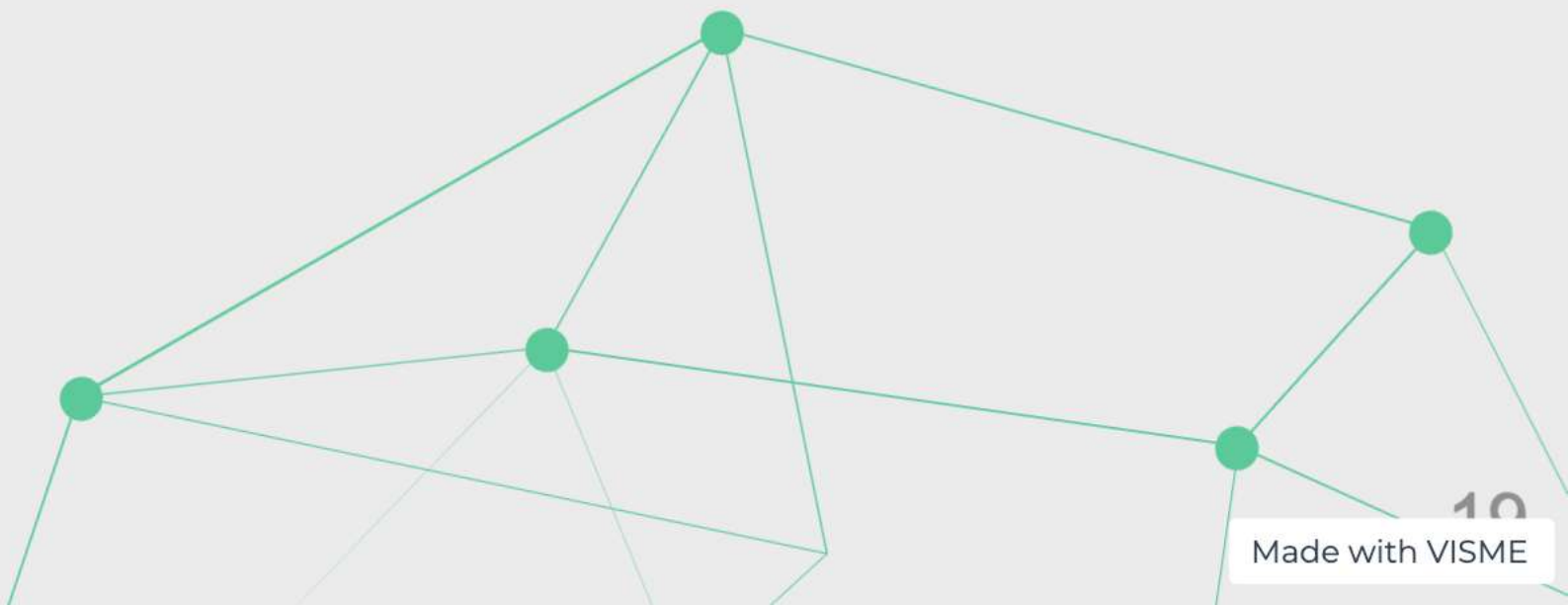
Davenport hits out at Wimbledon

world number one lindsay davenport has criticised wimbledon over the issue

davenport criticises wimbledon

davenport criticises wimbledon

davenport criticises wimbledon





Ground-truth title

T5 pretrained

T5 finetuned

T5 + First Paragraph

T5 + Short Summary


Rock group Korn's guitarist quits

the guitarist with us rock band korn has quit the music business saying he made the

korn guitarist quits after religious awakening

korn guitarist quits music business

welch quits korn after religious awakening



Conclusion

- 1 Titles are important.
- 2 We can generate them using proper deep learning architectures.
- 3 Finetuning T5 can significantly help us.
- 4 The first paragraph is enough!
- 5 Automatic and Human evaluations have shown the effectiveness of our methods.

NATURAL LANGUAGE PROCESSING COURSE (FALL 2022) FINAL PROJECT: TITLE GENERATION USING A TWO-PHASED SUMMARIZATION TECHNIQUE

Kuldeep Singh[†] Reza Khan Mohammadi[†] Ye Ma[‡]

[†]Computer Science and Engineering Department, Michigan State University
[‡]Department of Linguistics, Languages, and Cultures, Michigan State University

ABSTRACT

The generation of titles is an interesting process. The generation of titles is a challenging process due to the significance of the title, especially in the news. There is no straightforward approach to generating the proper title. Still, titles are known to play the initial most crucial effect on the reader. In this project, we tackle title generation in the news domain. We explored several summarization techniques, mostly revolving around sequence-to-sequence architectures. Finally, we develop a two-phased title generation technique that first generates a short summary of the raw input news and, from that, generates the final title using a finetuned version of T5. We evaluated our approach using the ROUGE metric. Our final results demonstrate the significance of news' first paragraphs in generating proper titles. Our final technique yielded a ROUGE score of 0.3582.

Index Terms— title generation, text summarization, abstractive summarization, extractive summarization, T5

1. INTRODUCTION

Generating a title for a given text is not an easy task. From a literature perspective, the title of a text is supposed to have certain characteristics. For instance, it has to be expressive, meaning that it is expected to convey the heart and soul of the text's content. It also should be interesting enough to grab the attention of the audience. Also, titles can intentionally/unintentionally cause bias and be written in a way that directs readers' mindset towards a specific perspective which can act as a double-edged sword. This may cause either grab the audience's attention and cause a stream of views

summarization techniques. We also experiment with transformers such as BERT [1] and T5 [2]. We perform several experiments and ablation studies to determine means of improving the efficiency of our method. We used ROUGE [3] as our primary metric of assessing generated titles.

The remaining of this paper is organized as follows: first, we describe previous studies in section 2. Then, we fully describe our approach in section 3 and performed experiments and ablation studies in section 4. Finally, we conclude the paper in the section 5.

2. RELATED WORK

Researchers have been using the task of Title Generation to provide an apt summary for a blog, a news title for a blog, commit message for a code snippet, generate a Youtube video title using description, create a generic title for clustered documents, the title for Spoken Broadcast news, generating StackOverflow Questions given the code snippet and description, etc. To put in simple words it is basically a model which produces a one-liner summary or representation for a longer document. A document can be multiple text files, code snippets, or news articles. It can be looked at as a similar task to summarization.

One of the earliest studies in title generation is [4] which was published in 1999. This paper discusses the generation of one-sentence length summarizations by statistically learning models of both content selection and realization. Given an appropriate training corpus, they try to generate summaries similar to the training ones, of any desired length. They use news-wire articles from Reuters and the Associated Press from the LDC. This is basically a statistical summarization

The background features a dark, moody image of a laboratory beaker containing a dark liquid. Overlaid on this is a network of thin, light-green lines connecting several circular nodes of the same color. These nodes are positioned at various points: one near the top left, one near the top right, one near the bottom left, and one near the bottom right, with lines crisscrossing between them. The text 'Any Questions?' is centered in a large, white, sans-serif font.

Any Questions?





Thank you!

Kuldeep Singh, Reza Khan Mohammadi, Ye Ma