

**COMBINING STATE-OF-THE-ART MODELS FOR MULTI-DOCUMENT  
SUMMARIZATION USING MAXIMAL MARGINAL RELEVANCE**

**DAVID JOEL ADAMS**  
**Bachelor of Science, University of Lethbridge, 2019**  
**Bachelor of Arts, MacEwan University, 2013**

A thesis submitted  
in partial fulfilment of the requirements for the degree of

**MASTER OF SCIENCE**

in

**COMPUTER SCIENCE**

Department of Mathematics and Computer Science  
University of Lethbridge  
LETHBRIDGE, ALBERTA, CANADA

© David Joel Adams, 2021

COMBINING STATE-OF-THE-ART MODELS FOR MULTI-DOCUMENT  
SUMMARIZATION USING MAXIMAL MARGINAL RELEVANCE

DAVID JOEL ADAMS

Date of Defence: August 26, 2021

Dr. Y. Chali Thesis Supervisor	Professor	Ph.D.
-----------------------------------	-----------	-------

Dr. J. Anvik Thesis Examination Committee Member	Associate Professor	Ph.D.
--	---------------------	-------

Dr. J. Zhang Thesis Examination Committee Member	Associate Professor	Ph.D.
--	---------------------	-------

Dr. W. Osborn Chair, Thesis Examination Com- mittee	Associate Professor	Ph.D.
---	---------------------	-------

# Dedication

To Chris Farley, whose work helped me through those difficult moments.

# Abstract

In Natural Language Processing, multi-document summarization (MDS) poses many challenges to researchers. While advancements in deep learning approaches have led to the development of several advanced language models capable of summarization, the variety of approaches specific to the problem of multi-document summarization remains relatively limited. Current state-of-the-art models produce impressive results on multi-document datasets, but the question of whether improvements can be made via the combination of these state-of-the-art models remains. This question is particularly relevant in few-shot and zero-shot applications, in which models have little familiarity or no familiarity with the expected output, respectively. To explore one potential method, we implement a query-relevance-focused approach which combines the pretrained models' outputs using maximal marginal relevance (MMR). Our MMR-based approach shows improvement over some aspects of the current state-of-the-art results while preserving overall state-of-the-art performance, with larger improvements occurring in fewer-shot contexts.

# Acknowledgments

I would like to thank Dr. Yllias Chali for his efforts in consistently providing me support and guidance in the completion of my research. He never seemed unavailable, and always made time to be my mentor.

I would also like to thank John Anvik and John Zhang for their positions on my supervisory committee.

I would also especially like to thank Gandharv Suri, whose hard work provided welcome assistance in this research.

I would also like to thank all of my mentors from the University of Lethbridge who instructed me through courses and helped enrich my understanding of the subject.

I also would like to thank the University of Lethbridge for their financial support.

Lastly, I want to thank my mother, my brother Matt, and every other member of my family who provided support or encouragement through my pursuit of higher education.

# Contents

<b>Contents</b>	<b>vi</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Contributions . . . . .	3
1.3 Overview of the Thesis . . . . .	4
<b>2 Background</b>	<b>5</b>
2.1 Natural Language Processing (NLP) . . . . .	5
2.2 Text Summarization . . . . .	6
2.2.1 Extractive Summarization . . . . .	6
2.2.2 Abstractive Summarization . . . . .	8
2.3 Multi-Document Summarization . . . . .	9
2.4 Artificial Neural Network . . . . .	10
2.4.1 Recurrent Neural Network (RNN) . . . . .	12
2.4.2 Convolutional Neural Network (CNN) . . . . .	13
2.4.3 Long Short-Term Memory (LSTM) . . . . .	13
2.4.4 Gated Recurrent Unit (GRU) . . . . .	14
2.4.5 Sequence-to-Sequence (seq2seq) . . . . .	14
2.5 Reinforcement Learning (RL) . . . . .	16
2.6 Transformer . . . . .	17
2.6.1 Positional Encoding . . . . .	18
2.6.2 Multi-Head Attention . . . . .	18
2.7 Language Models . . . . .	19
2.7.1 BERT . . . . .	20
2.7.2 RoBERTa . . . . .	20
2.7.3 DeBERTa . . . . .	20
2.7.4 CamemBERT . . . . .	21
2.7.5 ALBERT . . . . .	21
2.7.6 BART . . . . .	21
2.7.7 PEGASUS . . . . .	22
2.7.8 GPT . . . . .	23
2.7.9 GPT-2 . . . . .	23

2.7.10	GPT-3 . . . . .	23
2.7.11	Transformer-XL . . . . .	23
2.7.12	XLNet . . . . .	24
2.7.13	XLM . . . . .	24
2.7.14	XLM-RoBERTa . . . . .	24
2.7.15	ProphetNet . . . . .	25
2.7.16	LED . . . . .	26
2.7.17	T5 . . . . .	27
2.7.18	mT5 . . . . .	27
2.7.19	UniLM . . . . .	28
2.7.20	CTRL . . . . .	28
2.7.21	Opus-MT . . . . .	28
2.7.22	BlenderBot . . . . .	28
2.7.23	TED . . . . .	28
2.8	ROUGE: Evaluation Metric . . . . .	29
2.8.1	ROUGE-N . . . . .	29
2.8.2	ROUGE-L . . . . .	30
2.8.3	ROUGE-W . . . . .	31
2.8.4	ROUGE-S . . . . .	32
2.9	Background Summary . . . . .	33
<b>3</b>	<b>MMR-Based Approach</b> . . . . .	<b>34</b>
3.1	Maximal Marginal Relevance . . . . .	34
3.2	Similarity Measures . . . . .	35
3.2.1	TF-IDF . . . . .	35
3.2.2	Doc2Vec . . . . .	36
3.2.3	Cosine Similarity . . . . .	38
3.2.4	Word Mover’s Distance (WMD) . . . . .	39
3.3	Preliminary Research Overview . . . . .	39
3.3.1	Preliminary Use of Models . . . . .	40
3.3.2	Preliminary MMR Parameters . . . . .	41
3.4	Approach Overview . . . . .	43
3.5	Model Usage . . . . .	45
3.6	MMR Usage . . . . .	45
3.6.1	Latent Dirichlet Allocation (LDA) . . . . .	46
3.6.2	MMR Reduction . . . . .	46
3.6.3	Output Combination . . . . .	47
3.7	Summary . . . . .	48
<b>4</b>	<b>Multi-News Experiment and Results</b> . . . . .	<b>49</b>
4.1	Experimental Setup . . . . .	49
4.1.1	Dataset . . . . .	49
4.1.2	Implementation Details . . . . .	49
4.1.3	Model Versions . . . . .	50
4.1.4	MMR Parameters . . . . .	52

---

4.1.5	Evaluation Metric . . . . .	54
4.1.6	Baseline Measures . . . . .	54
4.2	Results . . . . .	55
4.3	Discussion . . . . .	55
4.4	Summary . . . . .	56
<b>5</b>	<b>WCEP Experiment and Results</b>	<b>58</b>
5.1	Experimental Setup . . . . .	58
5.1.1	Dataset . . . . .	58
5.1.2	SDS Problem . . . . .	59
5.1.3	Implementation Details . . . . .	60
5.1.4	Model Versions . . . . .	60
5.1.5	MMR Parameters . . . . .	63
5.1.6	Evaluation Metric . . . . .	65
5.1.7	Baseline Measures . . . . .	65
5.2	Results . . . . .	66
5.3	Discussion . . . . .	66
5.4	Summary . . . . .	67
<b>6</b>	<b>Conclusion</b>	<b>68</b>
6.1	Summary . . . . .	68
6.2	Future Work . . . . .	68
	<b>Bibliography</b>	<b>70</b>
<b>A</b>	<b>Multi-News Sample Generated Summaries</b>	<b>82</b>
A.1	Multi-News Sample 1 . . . . .	82
A.2	Multi-News Sample 2 . . . . .	91
<b>B</b>	<b>WCEP Sample Generated Summaries</b>	<b>96</b>
B.1	WCEP Sample . . . . .	96



# List of Tables

4.1	Optimized final MMR combination parameters . . . . .	54
4.2	Results of our Multi-News experiment . . . . .	55
5.1	Optimized MMR parameters . . . . .	65
5.2	Results of our WCEP experiment . . . . .	66

# List of Figures

2.1	Diagram of MatchSum Framework (Zhong et al., 2020) . . . . .	7
2.2	Illustration of Artificial Neural Network. . . . .	11
2.3	Neuron structure including bias as input (Haykin, 2009). . . . .	12
2.4	The Transformer Architecture (Vaswani et al., 2017) . . . . .	17
2.5	BART Architecture (Lewis et al., 2020) . . . . .	21
2.6	PEGASUS Architecture (Zhang et al., 2020) . . . . .	22
2.7	ProphetNet Model (Qi et al., 2020) . . . . .	25
2.8	LED Attention Mechanisms. (Beltagy et al., 2020) . . . . .	26
3.1	Word2Vec model architectures (Mikolov et al., 2013). . . . .	37
3.2	Overview of our preliminary MMR-based approach. . . . .	42
3.3	Overview of our final MMR-based approach. . . . .	44

# Chapter 1

## Introduction

### 1.1 Motivation

Language is the structured means by which ideas are communicated between individuals (Merriam-Webster, 2021). Natural language is distinct from constructed language in that it arises without the need for conscious design, whereas constructed language is conceived with conscious intent (Lyons, 1991). Natural Language Processing (NLP) is a broad field of research focused on the representation and analysis of natural human language computationally (Khurana et al., 2017). While NLP has the potential to be applied to other representations of human language such as speech audio, text remains the most common representation used in NLP, often in the form of documents (Doszkocs, 1986). Text documents can contain large amounts of information, and in the years since the availability of the internet, the amount of available text data continues to grow. The task of extracting information from such large amounts of text data poses a significant problem for which automated assistance is warranted. The solution offered by the field of NLP to this problem lies in research aimed at text summarization.

While there exist a wide variety of methods and techniques associated with text summarization, they can broadly be divided into two types: extractive summarization and abstractive summarization. Extractive summarization involves the selection and extraction of important text elements from the source documents verbatim, while in abstractive summarization the text elements are converted into abstract representations, which are then processed and converted back into a coherent summary. Each of these types has advantages

and disadvantages. While extractive summaries tend to be more coherent at lower cost, they also tend to be limited in their inclusion of overall context. This lack of context in summary-generation tends to result in summaries which contain grammatical and factual errors, despite using the exact words from the source document. For example, a source article might describe a bike accident which occurred on June 8, 2021 and involved a man named August who was 20 and born in 2001. A flawed extractive summary of this article might claim that the bike accident occurred on August 20, 2001, all of which would contain information which is from the source text but not necessarily in factually correct format. Abstractive summaries are more likely to be reworded in a way which properly condenses the information into a smaller document, but with higher cost and the possible inclusion of words which humans would not associate with the source document. Some approaches also combine extractive and abstractive summarization.

Within the problem of text summarization are several related subproblems. Both extractive and abstractive summarization have been applied to these subproblems. These subproblems include: query-focused summarization, in which the information included in the summary is refined to prioritize relevance to a specific topic or question; and multi-document summarization (MDS), in which the source document consists of a cluster of related documents for which a comprehensive summary must be generated. MDS poses unique challenges in addition to those offered by normal text summarization, which in this context we refer to as single-document summarization (SDS). These additional challenges include increased search space and higher redundancy.

Recently, advancements in summarization have been made in the form of language models, particularly those derived from the transformer architecture proposed by Vaswani et al. (2017). Recent language models include BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019), BART (Lewis et al., 2020), UniLM (Dong et al., 2019), GPT-2 (Radford et al., 2019), PEGASUS (Zhang et al., 2020), LED (Beltagy et al., 2020), ALBERT (Lan et al., 2020), T5 (Raffel et al., 2020), ProphetNet (Qi et al., 2020), XLNet (Yang et al., 2020a),

and TED (Yang et al., 2020b). These models can produce promising summarization results, although the development of models specific to MDS remains limited.

Statistical methods have also been considered for summarization, such as Mao et al.’s (Mao et al., 2020) use of maximal marginal relevance (MMR) in a reinforcement learning MDS approach. However, to our knowledge, the use of statistical methods to improve the MDS outputs of pretrained models remains largely unexplored. The motivation of our research is to apply pretrained models and statistical methods to MDS while preserving or improving state-of-the-art performance. We hypothesize that by supplementing the best outputs of various pretrained models using MMR with a strong focus on relevance, we can produce improved MDS results over recent model outputs, particularly in zero-shot and few-shot applications. Few-shot applications are those in which the given model or system has been given limited or incomplete information about the expected output. An example of a few-shot application would be a model which is only partially pretrained on the data it will be generating output from. A zero-shot application is one in which the model or system has been given no prior information about the data to be processed or the output expected from it, such as the use of a model which has not been trained on the given data at all and is receiving entirely new information. These approaches are more commonly required in applications such as MDS in which the datasets available might be limited, increasing the likelihood that training is too expensive or not possible due to a lack of training data. The details of our approach are further explained in Chapter 3.

## **1.2 Contributions**

- We have implemented several existing pretrained models to produce summaries of document clusters. These models’ outputs were incorporated into our own approach, but also used for comparison with our approach.
- We have provided a novel approach which combines the outputs of state-of-the-art pretrained models using maximal marginal relevance (MMR). This approach involves

the determination of how the model outputs should be processed, as well as how similarity is measured within each MMR calculation. This approach is also demonstrably effective in zero-shot or few-shot applications.

- We have further explored the scalability of our MMR-based method of combination when working with larger data and fewer-shot use of the models.

### **1.3 Overview of the Thesis**

The organization of this thesis is as follows. Chapter 2 includes an overview of the literature which is relevant to the background of this thesis. Chapter 3 consists of a detailed description of our MMR-based combination approach. Chapter 4 concerns the details of our Multi-News Experiment and its results. Chapter 5 concerns our WCEP Experiment and its results. The source code for both of our experiments can be found in our public repository<sup>1</sup>. Chapter 6, the final chapter, provides an overview and summary of this thesis and offers potential direction for future research.

---

<sup>1</sup><https://github.com/ThisDavidAdams/MMR-summarization>

# Chapter 2

## Background

This chapter provides and explains background and context relevant to the tasks and challenges associated with this research.

### 2.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a branch of artificial intelligence (AI) that focuses on the interaction between human language and computers. While computers operate using intentionally-constructed languages and symbolic operations, human natural language arises organically without intentional structure or strict regulation. This lack of intentional structure and possibility for hidden contextual nuance in natural language makes translations from human to computer language difficult and likely imperfect. The difficulty of such tasks inspired their inclusion in the Turing Test proposed by Turing (1950) and later included in the Chinese room argument presented by Searle (1980). The purpose of the Turing Test was argued to be the determination of whether a machine exhibits intelligent behaviour.

Since the consideration of natural language as an integral element of determining AI, NLP has developed into a separate but related discipline concerned with the performance of various NLP tasks by computers. These tasks include voice recognition, text-to-speech, named entity recognition (NER), sentiment analysis, machine translation, question answering, and text summarization. The state-of-the-art capacity for performing these tasks tends to improve as various advancements are made in areas related to computation, such as deep

learning.

## **2.2 Text Summarization**

Text summarization involves the generation of a relatively brief statement to present the main information contained in the source document being summarized. Research has been conducted on automatic text summarization as early as Luhn's research on abstract generation in 1958 (Luhn, 1958). As noted by Borko and Bernier (1975), the benefits provided by summaries include search improvement, time-saving, and indexing improvement.

Applications for text summarization include, but are not limited to, web-scraping, news-feed generation, commercial processing of business-related documents, and product review compilation. Text summarization is increasingly important to industries which deal in large quantities of information. Research has been conducted regarding its application in industries such as the medical field (Afzal et al., 2020; Sarker et al., 2020; DeYoung et al., 2021), the legal field (Galgani et al., 2012), the oil and gas industry (Correia Marques et al., 2019), and the agricultural industry (Peng et al., 2016), for example.

### **2.2.1 Extractive Summarization**

Extractive summarization is the process by which a summary is generated via the simple extraction of input text for inclusion in the output document. While this process can involve complicated mechanisms, extractive summarization as a whole tends to be lower in cost and easier to perform due to the lack of a need for word abstraction. A common objective in extractive summarization is the selection of key information to be extracted for inclusion. Extractive summarization also tends to have less potential to create inaccurate statements due to the verbatim use of the source text. However, incorrect statements can still be generated through the use of improper grammar or the rearrangement of text in such a way to change the meaning. These possible grammatical errors are often avoided through the extraction of entire sentences as opposed to individual words.



Various methods have been implemented for the improvement of extractive summarization, such as the use of latent semantic analysis (LSA) (Ozsoy et al., 2010; Foong et al., 2015; Padmakumar and Eswaran, 2014; Steinberger and Jezek, 2004) and the use of deep learning (Verma and Nidhi, 2019; Zhang et al., 2016; Cheng and Lapata, 2016). In addition, some of these deep learning approaches are graph-based (Wang et al., 2020). A recent advancement in extractive summarization which can employ other state-of-the-art models is the state-of-the-art MatchSum framework proposed by Zhong et al. (2020), which will be utilized by our research and more thoroughly explained in the MatchSum subsection.

### MatchSum

The MatchSum framework uses text-matching for extractive summarization, and was developed by Zhong et al. (2020). A diagram of the MatchSum framework can be seen in Figure 2.1.

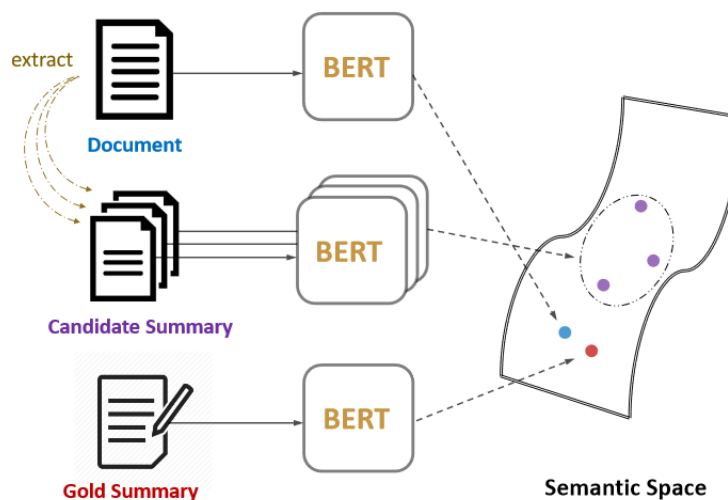


Figure 2.1: Diagram of MatchSum Framework (Zhong et al., 2020)

Figure 2.1 shows a high-level representation of the authors' objective, where the better candidate summaries are expected to be closer to the human-generated gold summaries in the semantic space. This approach is based on the observed gap between sentence-level

and summary-level extraction method, with the summary-level learning process providing less reward when the given reference summary is shorter. To address this gap, the authors propose a framework which matches the candidate summary to the source document using a Siamese-Bert architecture inspired by Bromely et al.’s siamese network structure (Bromley et al., 1993). This architecture derives semantically meaning embeddings from both the document and candidate summary rather than obtaining the sentence-level representation. They combine this approach with a pairwise loss function and a margin-based triplet loss function for document matching, which ranks the candidate summaries’ and gold summary’s ROUGE scores and allow the gold summary to have the highest matching score. The gap in ranking between the summaries determines the degree to which the candidate summary matches the document semantically (Zhong et al., 2020).

### **2.2.2 Abstractive Summarization**

Abstractive summarization produces summaries through the interpretation of source documents in order to generate a shorter text which includes the same concepts but not necessarily the same words verbatim. This summarization technique requires rephrasing and holistic document comprehension which is generally more difficult to implement than the sentence extraction methods of extractive summarization. Due to the complexity of the operations involved, research in extractive summarization was largely favored over research in abstractive summarization until the advent of deep learning techniques and advanced models decreased the difficulty of the task. Abstractive summarization approaches often include the use of an encoder-decoder architecture, in which the input text is first encoded into abstract representations to allow for reinterpretation, and then decoded as new text and generated as output (Tan et al., 2017; Lin et al., 2018; Rush et al., 2015; Nallapati et al., 2016; See et al., 2017).

Various extensions of the abstractive approach have been proposed by researchers. These extensions and improvements include the use of Abstract Meaning Representation

(AMR) with previous neural encoder-decoder approaches (Hardy and Vlachos, 2018), the use of pointer-generators (See et al., 2017), topic-awareness (Zheng et al., 2020), the use of reinforcement learning (Chen and Bansal, 2018), and the development of various state-of-the-art models (Zhang et al., 2020; Savelieva et al., 2020; Esteva et al., 2021).

Research in summarization has also included several extractive-abstractive approaches (Hsu et al., 2018; Li et al., 2021; Zhang et al., 2018; Gehrmann et al., 2018), an early iteration of which was the two-stage approach involving decomposition proposed by Jing and McKeown (1999). The aim of the extractive-abstractive concept, in principle, is to first employ the verbatim text extraction capabilities of an extractive approach to isolate the important text. The extractive output is then used as input for the abstractive step of the approach, in order to reduce the workload for the abstractive model and minimize the risk of including redundant or irrelevant information.

### **2.3 Multi-Document Summarization**

Multi-document summarization (MDS) is a subproblem within text summarization, which in this context is referred to as single-document summarization (SDS). In the more specific task of MDS, a summary, rather than being generated from a single document, is generated from a cluster of related documents. The task of MDS poses additional challenges, such as increased search space and greater potential for redundancy. In addition, summarization can be made more difficult when document clusters contain conflicting information.

A more specific MDS task can involve the presentation of the summary in a manner which progresses from general information to specific information. However, this aspect of the challenge was not considered relevant to our research, as the focus of our research was on the use of pretrained models with improved coverage of information and minimal redundancy. Therefore, the simpler task of producing a concise summary from large and diverse document clusters is the focus of this thesis.

Significant past improvements in approaches to MDS include the use of reinforcement learning (Mao et al., 2020; Rioux et al., 2014), sentence fusion (Barzilay et al., 1999; Nayeem et al., 2018), hierarchical model structures (Christensen et al., 2014; Celikyilmaz and Hakkani-Tur, 2010; Liu and Lapata, 2019), topic-modeling (Zheng et al., 2019; Li and Li, 2014; Wang et al., 2009), and determinantal point processes (Cho et al., 2019). Advancements in MDS tend to be hindered by difficulties such as the relative lack of datasets for the pretraining of MDS models, the differences in summary length and properties between datasets, and the different hierarchical properties exhibited by the source texts in different datasets. Recently, research has been conducted to generate new MDS datasets, including the Multi-News dataset (Fabbri et al., 2019), the Multi-XScience dataset (Lu et al., 2020), the WCEP dataset (Gholipour Ghalandari et al., 2020), and the GameWikiSum dataset (Antognini and Faltings, 2020). Additionally, Liu et al. (Liu et al., 2018) proposed the WikiSum architecture for generating a MDS dataset. The WikiSum method involves the generation of Wikipedia articles as summaries of references documents.

## **2.4 Artificial Neural Network**

An important development in AI was the advent of artificial neural networks (ANNs). The fundamental concept of an artificial neural network is to emulate the neural networks that occur in biological brains. These networks consist of interconnected groups of artificial neurons, also known as perceptrons or nodes in the case of ANNs, each of which processes and transfers data to the next part of the network (Graupe, 1997). These groups of nodes are known as layers, of which there are three types: input, hidden, and output. The input layer is that through which data is introduced to the ANN. A hidden layer is one in which some modification is usually performed to the data before its processing is complete. The output layer provides the final result of the alterations made by the ANN, as well as the error of the output. The network can consist of as many hidden layers as needed, with the required number being dependent on the operations the hidden layers perform and how they can

be optimized. The number of layers in an ANN is referred to as a network's depth. Each layer can consist of as many nodes as needed, which is referred to as a layer's width. The total number of nodes in an ANN is referred to as the network's size. An illustration of the concept of an ANN can be seen in Figure 2.2.

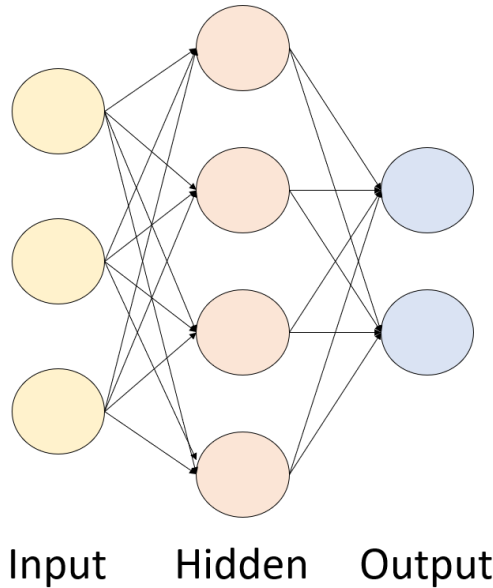


Figure 2.2: Illustration of Artificial Neural Network.

For each node in a hidden layer, each input is taken along with its weight. The inputs are multiplied by their weights and added together before the bias is added. This is known as the update function. This process can be seen in Equation 2.1.

$$z = \left( \sum_{i=0}^n w_i x_i \right) + b \quad (2.1)$$

In the equation,  $n$  is the number of inputs,  $w_i$  is the current input's weight,  $x_i$  is the current input's value,  $b$  is the bias, and  $z$  is the returned update value for the node. After the weighted inputs and the bias are added, the node passes the sum through its activation function, which determines whether the node is activated. The exact function used for activation can vary depending on the ANN's application. The sigmoid function is commonly used as an activation function, and is shown in Equation 2.2.

$$y = \sigma(z) = \left( \frac{1}{1 + e^{-z}} \right) \quad (2.2)$$

Another common and simpler activation function is the rectified linear unit function (ReLU), which can be seen in Equation 2.3.

$$y = \max(0, x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

An illustration of the overall structure of a neuron can be seen in Figure 2.3.

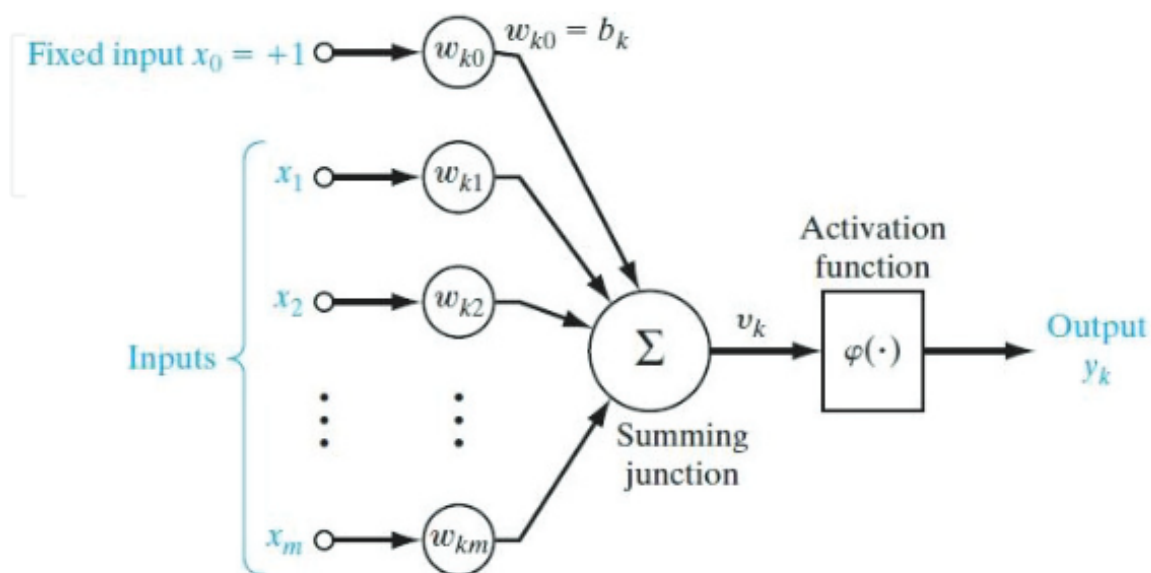


Figure 2.3: Neuron structure including bias as input (Haykin, 2009).

### 2.4.1 Recurrent Neural Network (RNN)

A recurrent neural network (RNN) is a type of ANN in which the nodes' edges connect to different time steps in the temporal sequence (Abiodun et al., 2018). The concept of the RNN is based on the work of Rumelhart et al. (1986). These connections between temporal states in the network allow for the network to retrace its steps and reattempt its calculations, which is known as backpropagation. Through backpropagation, the error signal can be sent

back through the ANN and different weights can be applied to the input values in an effort to reduce the error of the output. This signaling forms a value, the gradient, which must be minimized as its dot product with the input data is repeatedly calculated. The minimization of the gradient is known as gradient descent.

However, the RNN architecture has a potential flaw in the form of the vanishing gradient problem. As the RNN backpropagates, the gradient which is backpropagated can shrink toward zero as it is multiplied via computation. Conversely, the gradient can also increase toward infinity in what is known as the exploding gradient problem, although this problem tends to receive less direct attention in addressing as it is generally less common in neural network applications.

### **2.4.2 Convolutional Neural Network (CNN)**

A convolutional neural network (CNN) is a fully-connected RNN which contains a special layer type called the convolutional layer. The CNN was preceded to some degree by the neocognitron model (Fukushima and Miyake, 1982) before it was formally introduced by Cun et al. (1990). Each convolution layer effectively serves as a building block where increasingly complex operations are performed on the data to establish hierarchies through backpropagation. Convolutional layers are therefore particularly useful in aiding the network in the detection of patterns. This quality makes CNN's effective in the completion of NLP tasks involving patterns, such as text summarization, sentiment analysis, classification, etc., as well as a number of other data processing applications such as image processing.

### **2.4.3 Long Short-Term Memory (LSTM)**

The Long Short-Term Memory (LSTM) architecture is another proposed solution to the RNN architecture's vanishing gradient problem, and was proposed by Hochreiter and Schmidhuber (1997). A LSTM unit contains an input gate, an output gate, and a forget gate. Because this mechanism allows for gradients to pass through without being altered, it is able to avoid the vanishing gradient problem from repeated calculations.

#### 2.4.4 Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) architecture is one proposed solution to the vanishing gradient problem of RNN's, and was proposed by Cho et al. (2014). A GRU is a gating mechanism which is similar in principle to a LSTM in that it uses feedback connects and a forget gate, but also has key differences in its lack of an output gate and lower amount of parameters. This structure means the GRU allows the RNN to remember and forget certain gradients at a lower cost than when using LSTMs.

#### 2.4.5 Sequence-to-Sequence (seq2seq)

A sequence-to-sequence (seq2seq) approach is specific to language processing, and may consist of any variety of models. The approach was introduced by Google scientists Sutskever et al. in 2014 (Sutskever et al., 2014) using LSTMs for machine translation. The primary components of this approach are the *encoder* and the *decoder* mechanisms, which may be ANNs or other models. The encoder translates the input text sequence into a hidden vector containing the input and its context, and the decoder reverses the process to generate the output text sequence. Hence, the process converts an input sequence to an output sequence, inspiring the name. Each subsequent input uses the previously-generated outputs as context. The seq2seq system is usually trained using a cross-entropy loss function for the penalizing of undesirable outputs (Keneshloo et al., 2019). The seq2seq approach can also be improved via the use of attention, beam search, and bucketing.

#### Attention

Attention is a technique intended to emulate the natural human ability to pay attention to the more important information while ignoring the less important information. In neural networks, attention amplifies the inclusion of salient data while phasing out the less significant data through mathematical operations. In principle, a simple attention mechanism is achieved via gradient descent, in which the use of dot product operations gradually alters the allocation of computing power on the input data.



There are many different types of attention techniques and mechanisms used in NLP research. These attention types include scaled dot-product attention and multi-head attention, which will be further described in Section 2.6.2.

### **Beam Search**

Beam search is an algorithm which selects and stores multiple probable states to find the optimal state. The term was coined by Reddy in 1977 (Dept., 1977), although it was used in Lowerre's Harpy Speech Recognition System in 1976 (Lowerre, 1976). It is a greedy algorithm which minimizes the heuristic cost among the best partial solutions it stores. Like breadth-first search, beam search builds its search tree from the candidate states within the current level, but unlike breadth-first search, not every state is expanded. Instead, a predetermined number of best choices is selected for expansion, with the rest of the choices being ignored. This number of best choices is known as the beam width.

This algorithm is used in NLP applications such as machine translation and text summarization. In these applications, the beam search algorithm searches among the predetermined number of best generated text sequences for a more optimized search. There are also variants of the beam search that include depth-first search, such as limited discrepancy search (Harvey and Ginsberg, 1995), beam-stack search (Zhou and Hansen, 2005), and beam search using limited discrepancy (Furcy and Koenig, 2005).

### **Bucketing**

Bucketing is a simple method of padding or truncating inputs or outputs to a certain length. In this method, the bucket refers to the buffer in which the data can be loaded piece-meal or by data regions. This use of buckets allows for more efficient machine learning, if the data is appropriately grouped into the buckets.

## 2.5 Reinforcement Learning (RL)

Reinforcement learning (RL) is one of the three primary machine learning methods, the other two being supervised learning, which involves training with labeled training data, and unsupervised learning, which uses probability calculations to detect patterns in unlabeled data. Unlike these other two machine learning methods, RL uses a simple numerical score as the feedback to guide the system. A RL system uses the numerical feedback as reinforcement for both exploration of new information and exploitation of already-discovered information (Kaelbling et al., 1996).

There are multiple possible algorithms for use in RL, each designed to find the optimal policy, or action, for the system to take while ameliorating the costs. These algorithms include the criterion of optimality approach, the brute force approach, and the value function approach. In the criterion of optimality approach, the algorithm searches the policy map for the policy with the maximum expected returned reward by weighting the distant-future options less highly than the near-future options, effectively guessing which policy yields the highest reward without needing to explore the returns from it. The brute force approach, in contrast, samples the return from each policy and finds the one with the largest expected return, which is an untenable approach in scenarios with infinite numbers of policies. The value function approach maintains a list of estimated expected returns for either the current or optimal policy.

RL has received a recent surge of attention among researchers. RL has been incorporated into robotics research (Kalashnikov et al., 2018; OpenAI et al., 2019), applied to natural language tasks such as text summarization (Mao et al., 2020; Rioux et al., 2014), and been implemented in an open source applied RL platform by Facebook (Gauci et al., 2019).

## 2.6 Transformer

The transformer model architecture was proposed by Vaswani et al. (2017) as a means to distill the deep learning process to involve only attention. Because this architecture relies entirely on attention to determine which information is included, excluded, or modified, no backpropagation is necessary. A visualization of the transformer architecture can be seen in Figure 2.4.

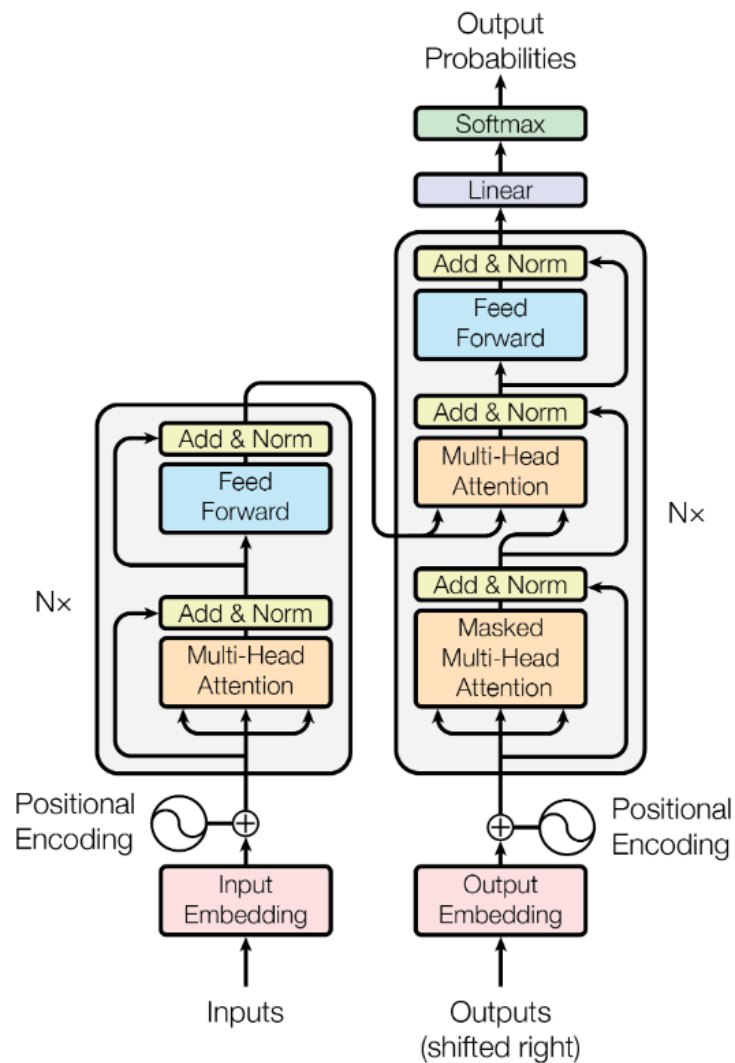


Figure 2.4: The Transformer Architecture (Vaswani et al., 2017)

The transformer consists of encoder and decoder stacks. Both the encoder and decoder consist of  $N$  identical layers, which is signified by  $N \times$  in the diagram. There were 6 identical

layers in the implementation proposed by Vaswani et al. (2017). Each encoder layer adds two sub-layers: the multi-head attention mechanism, and the positionwise fully connected feed-forward network. Each sub-layer has a residual connection (He et al., 2015) applied to it, as well as layer normalization (Ba et al., 2016).

### 2.6.1 Positional Encoding

Positional encoding is the means by which the transformer architecture is able to use the order of sequence without recurrence or convolution (Vaswani et al., 2017). The formula for these encodings can be seen in Equation 2.4 and Equation 2.5.

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}}) \quad (2.4)$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (2.5)$$

In the positional encoding equations,  $pos$  is the position, and  $i$  is the given dimension which corresponds to a sinusoid.  $d$  is the representation dimension. These encodings are added to the input embeddings at the bottoms of both the encoder and decoder stacks, and have the same dimension as the embeddings.

### 2.6.2 Multi-Head Attention

Vaswani et al. (2017) formulate attention as the mapping of a query and set of key-value pairs to an output, all of which are vectors. They employ multi-head attention, which in turn utilizes their basic attention function. They define their basic attention function as the scaled dot-product attention, the formula for which can be seen in Equation 2.6. Scaled dot-product attention divides the combined dot products of the query and all keys by the square root of the dimension of the key. The divided output then has a softmax function applied to it, the output of which is then multiplied by the value.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.6)$$

In the scaled dot-product attention formula,  $Q$  is a query,  $K$  is a key, and  $V$  is a value. Multi-head attention is a parallel application of attention of the projected versions of the queries, keys, and values. The formula for multi-head attention can be seen in Equation 2.7.

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}[\text{head}_1, \dots, \text{head}_h]W_0 \\ \text{where } \text{head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \tag{2.7}$$

In the multi-head attention formula,  $W$  is a learnable parameter matrix, while  $Q$ ,  $K$ , and  $V$  are a query, key, and value, respectively (Vaswani et al., 2017). The transformer architecture uses encoder-decoder attention layers, in which the queries are derived from the previous decoder layer and the memory keys and values are from the encoder output. The transformer also uses self-attention layers in both the encoder and decoder so they can access all positions in the previous layer. It also employs a masked multi-head attention mechanism in which the positions are prevented from attending to subsequent positions (Vaswani et al., 2017).

## 2.7 Language Models

A number of language models for NLP tasks have been developed, pretrained, and made available for use. The majority of state-of-the-art pretrained models are based on the transformer architecture. The number of hyperparameters for these pretrained models can vary depending on model size and whether the model is distilled or not. Most of these language models also include their own tokenizers and text representations to ensure proper text processing. The use of a pretrained model eliminates the need to develop a new language model from scratch, and also reduces the computation costs for NLP applications. Some of these models will be incorporated into our research.

### **2.7.1 BERT**

The Bidirectional Encoder Representations from Transformers (BERT) model architecture was developed by Google (Devlin et al., 2019). This language model was derived from the transformer architecture, and more specifically from the encoder portion of the transformer architecture. The BERT architecture is able to train using the context to both the right and the left of its target, resulting in bidirectional representations of the text. This use of left and right context is made possible via the masking of the target. The BERT model architecture has been incorporated into many subsequent approaches to NLP tasks, including the text-matching extractive summarization model developed by Zhong et al. (2020).

### **2.7.2 RoBERTa**

The Robustly optimized BERT approach (RoBERTa) model architecture was developed by Facebook to improve end-task performance of the BERT architecture (Liu et al., 2019). The modifications introduced by the RoBERTa architecture include training on more data, for more time, with larger batches. This architecture also removes the next sentence prediction objective, is trained on longer sequences, and dynamically alters the masking pattern.

### **2.7.3 DeBERTa**

The Decoding-enhanced BERT with disentangled attention (DeBERTa) model was developed by Microsoft as an improvement to the BERT and RoBERTa architectures (He et al., 2021). DeBERTa adds a novel disentangled attention mechanism through which each word is represented by a vector for its content and a vector for its position, with attention weights calculated using a disentangled matrix on the word's content and another disentangled matrix on its position. DeBERTa also adds an enhanced mask decoder for the inclusion of absolute position in the decoding layer. DeBERTa was also trained using a novel virtual adversarial method.

### 2.7.4 CamemBERT

The CamemBERT model is an adaptation of the RoBERTa model to French-language NLP tasks (Martin et al., 2020).

### 2.7.5 ALBERT

A Lite BERT (ALBERT) model architecture was developed by Google to optimize the performance of the standard BERT model architecture for greater scalability (Lan et al., 2020). This optimization was achieved via a factorization of the embedding parameters into two smaller matrices, as well as cross-layer parameter sharing.

### 2.7.6 BART

The BART model was developed by Lewis et al. (2020) for Facebook to incorporate a novel denoising autoencoder into the standard transformer architecture. This model's autoencoding approach involves mapping a transformed, corrupted version of a document to the original version of the document. The transformations used by Lewis et al. include token masking, token deletion, text infilling, sentence permutation, and document rotation. A visualization of the BART architecture can be seen in Figure 2.5.

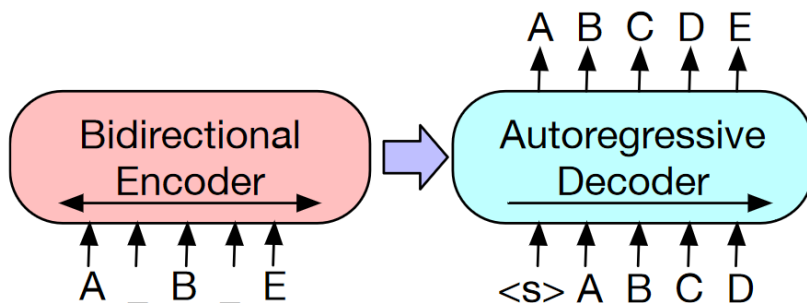


Figure 2.5: BART Architecture (Lewis et al., 2020)

In Figure 2.5, we see that the corrupted document on the left is encoded with a bidirectional model, which is then used by the autoregressive decoder to compute the probability of the original, uncorrupted document on the right. Lewis et al. note that while this model

is especially well-suited for text generation tasks, it also performs adequately for text comprehension tasks. BART is not tailored to any particular noising scheme, but rather able to incorporate any form of text corruption.

### 2.7.7 PEGASUS

The Pre-training with Extracted Gap-sentences for Abstractive SUMmarization Sequence-to-sequence (PEGASUS) models were developed by Zhang et al. (2020) for Google. This model architecture marks an attempt to reconfigure pretraining objectives to be more specific to abstractive summarization. In this model architecture, whole sentences are masked and predicted simultaneously using gap sentences generation. This masking of whole sentences is in contrast to many state-of-the-art models, which mask smaller text sequences. A visualization of the PEGASUS architecture can be seen in Figure 2.6.

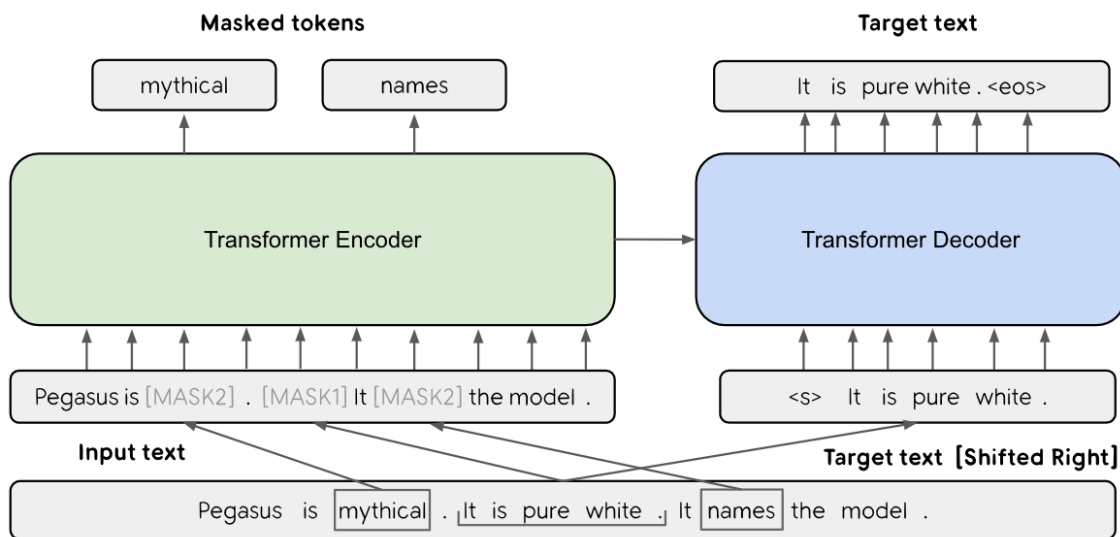


Figure 2.6: PEGASUS Architecture (Zhang et al., 2020)

In Figure 2.6, we see a candidate sequence consisting of three sentences. On the right, one of them is used as the mask sentence and extracted from the input. On the left, the remaining input has randomly-selected parts of it masked before it is passed to the encoder. Zhang et al. (2020) found that the PEGASUS model, as well as the PEGASUS<sub>Large</sub> model, produced state-of-the-art results on downstream summarization task with as little as 1000



samples (Zhang et al., 2020).

### **2.7.8 GPT**

The Generative Pre-Training (GPT) model (Radford and Narasimhan, 2018) was developed by OpenAI. This task-agnostic model uses the transformer architecture, and was pretrained on unlabeled data. After the unsupervised pretraining process, supervised fine-tuning was performed to apply the model to specific NLP tasks.

### **2.7.9 GPT-2**

The GPT-2 model is a scaled-up version of OpenAI’s previous GPT model architecture (Radford et al., 2019). Like the previous GPT model, GPT-2 uses the transformer architecture, albeit only the decoder portion. However, unlike the previous GPT model, GPT-2 was pretrained with 1.5 billion parameters, and used for zero-shot NLP tasks. Zero-shot tasks are tasks which are performed without first familiarizing the model with any of the expected output. The authors, Radford et al. (2019), note that while the GPT-2 model is capable of producing zero-shot summaries, the summaries are often confused with factual inaccuracies or a bias towards recent information.

### **2.7.10 GPT-3**

The GPT-3 model is the third released iteration of OpenAI’s GPT model architecture (Brown et al., 2020). Similar to GPT-2, this model consists of a scaled-up implementation of the previous model, with GPT-3 being trained with 175 billion parameters. Also like GPT-2, GPT-3 is applied to zero-shot NLP tasks without gradient updates or finetuning.

### **2.7.11 Transformer-XL**

The Transformer-XL model was developed by Google and Carnegie Mellon University to generate text while avoiding the limitations of a fixed-length context (Dai et al., 2019). Rather than considering only contexts of a fixed length, this model is able to learn de-

pendency for a context beyond a fixed length without rendering the context incoherent or fragmented through context selection.

### **2.7.12 XLNet**

The XLNet model was developed by Google and Carnegie Mellon University to integrate aspects from the Transformer-XL model, after which the XLNet model is named, into its autoregressive pretraining method (Yang et al., 2020a). This model overcomes the limitations of BERT by computing the maximum probability of all possible permutations of the left and right contexts, allowing the model to predict masked tokens without ignoring their interdependencies. However, these permutations are only on the factorization order of the sequence, and not on the sequence order itself. To increase the rate of convergence for their permutation objective, Yang et al. (2020a) use partial prediction, in which the last token in the factorization order is the only token being predicted. BERT also makes use of partial prediction. However, XLNet also makes use of the relative segment encoding mechanism from the Transformer-XL model, which is used to determine whether two words co-occur in the same segment of text. XLNet is shown to outperform BERT on several NLP tasks, although summarization specifically was not among the tasks used for testing.

### **2.7.13 XLM**

The cross-lingual language models (XLM) developed by Facebook researchers Lample and Conneau (2019) utilize generative pretraining for multiple languages. These models were developed using both an unsupervised pretraining approach on monolingual data and a supervised pretraining approach across parallel multilingual data.

### **2.7.14 XLM-RoBERTa**

The XLM-RoBERTa model was developed by Conneau et al. for Facebook (Conneau et al., 2020). This model leverages the previous XLM model architecture for training on larger corpora and multiple NLP tasks such as classification, question answering, and the

labeling of sequences.

### 2.7.15 ProphetNet

The ProphetNet model was developed by Microsoft and the University of Science and Technology of China (Qi et al., 2020). This model introduces a novel future n-gram prediction objective and n-stream self-attention mechanism to sequence-to-sequence NLP tasks. A diagram of the ProphetNet model can be seen in Figure 2.7.

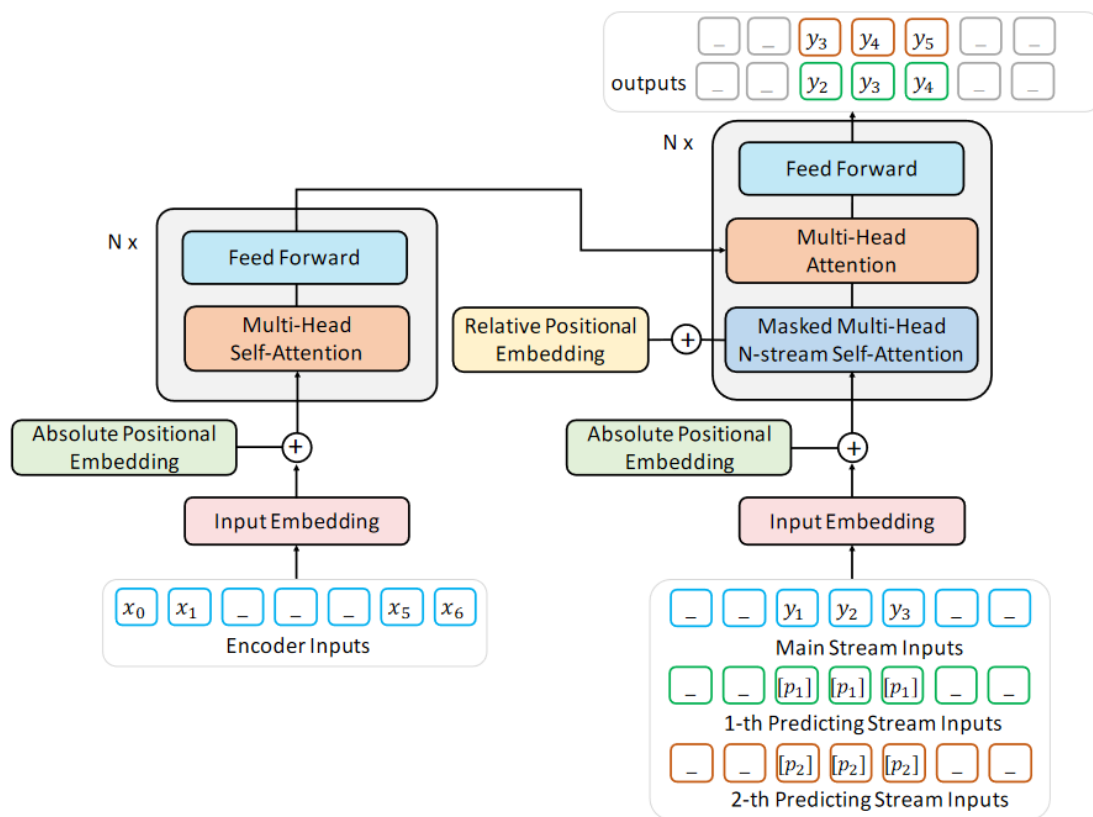


Figure 2.7: ProphetNet Model (Qi et al., 2020)

In Figure 2.7, the left portion represents the encoder, which is identical to that of the standard transformer model. On the right portion is the decoder unit which uses the n-stream self-attention mechanism. The example in the diagram is a bigram, and the underscore character represents the masking symbol. The decoder outputs are the probabilities shown in Equation 2.8.

$$p(y_t, y_{t+1} | y_{<t}, x), \dots, p(y_{t+n-1} | y_{<t}, x) = \text{Decoder}(y_{<t}, H_{enc}) \quad (2.8)$$

In the equation for ProphetNet’s probability equation,  $H_{enc}$  is the source sequence representation generated by the encoder, and the decoder outputs n probability at each time step (Qi et al., 2020). The n-stream attention mechanism is designed to allow for the simultaneous prediction of the next n continuous tokens. Each of the next n tokens are predicted by one of the extra n self-attention streams. The authors, Qi et al. (2020), tested the model with n-gram sizes of 1, 2, and 3 to determine the effect of the n-stream self-attention mechanism, and found that a greater number of parallel predictions led to improved results. They found that the ProphetNet model performed better than other state-of-the-art models in both abstractive summarization and question-answering tasks.

### 2.7.16 LED

The Longformer-Encoder-Decoder (LED) model was developed by the Allen Institute for Artificial Intelligence as a variant of the Longformer, which decreases the scaling of self-attention cost from quadratic growth to linear growth via a new attention mechanism (Beltagy et al., 2020). The LED model applies the Longformer architecture to generative sequence-to-sequence tasks for long documents. The Longformer architecture employs different attention mechanisms from the original transformer’s self-attention mechanism, which has  $O(n^2)$  time and memory complexity. A visualization of the LED model’s attention mechanisms can be seen in Figure 2.8.

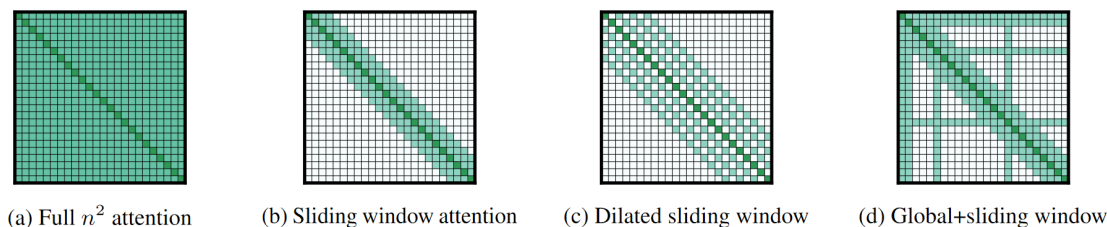


Figure 2.8: LED Attention Mechanisms. (Beltagy et al., 2020)

The sliding window attention pattern is used to surround each token with a window of a fixed size. This attention pattern incorporates all of the information pattern with a complexity of  $O(n \times w)$ , where  $n$  is the input sequence length and  $w$  is the window size. The dilated sliding window pattern is similar to the sliding window pattern, except it has gaps of size dilation  $d$ , allowing some iterations to compute longer contexts. The global attention pattern has  $O(n)$  complexity and was added to allow for task-specific pattern learning. The authors, Beltagy et al. (2020), found that the LED model outperformed other models on the summarization task.

### **2.7.17 T5**

The Text-to-Text Transfer Transformer (T5) model was developed by Google to apply the concept of transfer learning to the transformer architecture for multiple NLP tasks (Raffel et al., 2020). The use of transfer learning involves the pretraining of a model on a large corpus and then finetuning the model on a downstream task. The T5 model uses the text-to-text framework, which employs the same loss function and decoding procedure regardless of the task being performed. Raffel et al. (2020) found the text-to-text framework to perform with relatively equal effectiveness to frameworks with more task-specific loss functions and decoding procedures. The authors also find that, although their use of the encoder and decoder portions of the transformer architecture increases the number of parameters needing to be shared between the encoder and decoder, the computational cost is similar. The authors found that various versions of the model outperformed the previous best results in several NLP tasks, and could be used in abstractive summarization.

### **2.7.18 mT5**

The multilingual T5 (mT5) model was developed by Google to leverage the T5 model for NLP tasks across multiple languages (Xue et al., 2021).

### **2.7.19 UniLM**

The Unified pretrained Language Model (UniLM) was developed by Microsoft for the understanding and generation of natural language (Dong et al., 2019). This model uses self-attention masks within a shared transformer network to allow for the use of unidirectional, bidirectional, and sequence-to-sequence prediction.

### **2.7.20 CTRL**

The Conditional Transformer Language (CTRL) model was developed by Keskar et al. (2019) to allow for more human control in the generation of text. This increase in control is due to the potential for determining which parts of the training data are most influential to the model.

### **2.7.21 Opus-MT**

The OPUS-MT project was created for the development of free machine translation tools and resources, and includes over 1,000 pretrained neural machine translation models (Tiedemann and Thottingal, 2020).

### **2.7.22 BlenderBot**

The BlenderBot model was developed by Facebook to generate conversational text through the addition of skills to the standard scaled neural model approach (Roller et al., 2020). It is a chatbot model which is primarily configured to generate conversational text which is generated in relation to the input text.

### **2.7.23 TED**

The Transformer Encoder Decoder (TED) model is a summarization model which is pretrained on large-scale corpora using lead bias and theme modeling (Yang et al., 2020c).

## 2.8 ROUGE: Evaluation Metric

A standard method for measuring the quality of summaries was developed by Lin (2004) in the form of the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) collection of metrics. These metrics automatically compare the generated summaries to the human-generated gold summaries via word co-occurrence. The metrics relate to n-grams, which are word sequences of various lengths. Within each metric, three measurements are generated: Precision, recall, and F-measure or F1 score. The definitions for these measurements are defined in Equation 2.9 through Equation 2.11.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2.9)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2.10)$$

$$F\text{-measure} = \frac{(1 + \beta^2) * recall * precision}{recall + (\beta^2 * precision)} \quad (2.11)$$

True Positive refers to the relevant information which was included in the candidate sequence. False Positive refers to the irrelevant information which was included. False Negative refers to the relevant information which was not included. In the F-measure formula,  $\beta$  is the degree to which recall is considered more important than precision. In the case of ROUGE summary-level calculations, Lin (2004) proposes that  $\beta$  should be a large enough number that the final score is recall-oriented. Intuitively, precision is the amount of retrieved information which is relevant. Recall is the amount of relevant information which was retrieved. The F-measure is the combination, or harmonic mean, of precision and recall.

### 2.8.1 ROUGE-N

ROUGE-N is the set of measurements for n-grams among the candidate and reference documents, where n is the number of words in the n-gram. For example, ROUGE-1 and ROUGE-2 are common ROUGE-N metrics, and they relate to the comparisons of unigrams

and bigrams respectively between the documents. The formulas for the ROUGE-N measurements are shown in Equation 2.12 through Equation 2.14.

$$ROUGE-N(\textit{precision}) = \frac{\sum_{S \in \{\textit{ReferenceSummaries}\}} \sum_{gram_n \in S} \textit{Count}_{\textit{match}}(gram_n)}{\sum_{S \in \{\textit{GeneratedSummaries}\}} \sum_{gram_n \in S} \textit{Count}(gram_n)} \quad (2.12)$$

$$ROUGE-N(\textit{recall}) = \frac{\sum_{S \in \{\textit{ReferenceSummaries}\}} \sum_{gram_n \in S} \textit{Count}_{\textit{match}}(gram_n)}{\sum_{S \in \{\textit{ReferenceSummaries}\}} \sum_{gram_n \in S} \textit{Count}(gram_n)} \quad (2.13)$$

$$ROUGE-N(\textit{F-measure}) = \frac{(1 + \beta^2) * ROUGE-N(\textit{recall}) * ROUGE-N(\textit{precision})}{ROUGE-N(\textit{recall}) + (\beta^2 * ROUGE-N(\textit{precision}))} \quad (2.14)$$

In the formulas for these measurements,  $n$  is the length of the  $n$ -gram, which is referred to as  $gram_n$ .  $S$  is a summary.  $\textit{Count}_{\textit{match}}(gram_n)$  is the maximum co-occurring  $n$ -grams between a candidate summary and a summary from the reference summary set.  $\beta$  is set to a large number to make the calculation recall-oriented.

### 2.8.2 ROUGE-L

ROUGE-L is the metric for the longest common subsequence (LCS) between the candidate and reference summary. Naturally, we consider two documents to be more similar if they contain longer subsequences in common, such as whole sentences or paragraphs. The formulas for the ROUGE-L metrics when applied to summary-level are shown in Equation 2.15 through Equation 2.17.

$$ROUGE-L(\textit{precision}) = \frac{\sum_{i=1}^u \textit{LCS}_{\cup}(r_i, C)}{n} \quad (2.15)$$

$$ROUGE-L(\textit{recall}) = \frac{\sum_{i=1}^u \textit{LCS}_{\cup}(r_i, C)}{m} \quad (2.16)$$

$$ROUGE-L(\textit{F-measure}) = \frac{(1 + \beta^2) * ROUGE-L(\textit{recall}) * ROUGE-L(\textit{precision})}{ROUGE-L(\textit{recall}) + (\beta^2 * ROUGE-L(\textit{precision}))} \quad (2.17)$$

In the formulas for this metric,  $r_i$  is a reference summary sentence from the total set of summary sentences of size  $u$  sentences and  $m$  words.  $C$  is the candidate summary,



containing candidate summary sentences  $c_j$  for the sentence-level LCS calculation.  $LCS$  returns the length of the longest common subsequence. The candidate summary contains  $u$  sentences and  $n$  words.  $\beta$  is set to a large number to make the calculation recall-oriented.

For the sentence-level LCS calculation, the formulas are shown in Equation 2.18 through Equation 2.20.

$$LCS(\textit{precision}) = \frac{LCS(r,c)}{n} \quad (2.18)$$

$$LCS(\textit{recall}) = \frac{LCS(r,c)}{m} \quad (2.19)$$

$$LCS(\textit{F-measure}) = \frac{(1 + \beta^2) * LCS(\textit{recall}) * LCS(\textit{precision})}{LCS(\textit{recall}) + (\beta^2 * LCS(\textit{precision}))} \quad (2.20)$$

In the sentence-level LCS formulas,  $r$  is a reference summary of length  $m$ , and  $c$  is a candidate summary of length  $n$ .  $LCS(r,c)$  is the length of a longest common subsequence between summaries  $r$  and  $c$ , and  $\beta$  is  $LCS(\textit{precision})$  divided by  $LCS(\textit{recall})$ .

### 2.8.3 ROUGE-W

ROUGE-W is a metric derived from the weighted LCS method. This metric is designed to solve the problem wherein LCS cannot differentiate between consecutive LCSes and spatially divided LCSes. For example, consider the reference sequence  $R_1$  and the candidate sequence  $C_1$  and  $C_2$  as follows:

$R_1$ : [rat, bat, cat, hat, sat]

$C_1$ : [rat, bat, cat, vat, pat]

$C_2$ : [rat, mat, bat, fat, cat]

In this example,  $C_1$  should be the best choice, because it contains consecutively matching words with  $R_1$ . However, given the LCS limitation,  $C_1$  and  $C_2$  have the same ROUGE-L score. Therefore, ROUGE includes a weighted LCS (WLCS) method to compute the length of the longest consecutive matching common subsequence. The WLCS method is

performed via dynamic programming using a weighting function. The formulas for the WLCS measurements are shown in Equation 2.21 through Equation 2.23.

$$WLCS(\textit{precision}) = f^{-1} \frac{WLCS(r, c)}{f(n)} \quad (2.21)$$

$$WLCS(\textit{recall}) = f^{-1} \frac{WLCS(r, c)}{f(m)} \quad (2.22)$$

$$WLCS(\textit{F-measure}) = \frac{(1 + \beta^2) * WLCS(\textit{recall}) * WLCS(\textit{precision})}{WLCS(\textit{recall}) + (\beta^2 * WLCS(\textit{precision}))} \quad (2.23)$$

In these formulas,  $f^{-1}$  is the inverse of the weighting function  $f$ .  $f$  returns the weighted version of a given score.  $WLCS$  is the weighted LCS method previously derived from dynamic programming.  $r$  is a reference sequence of length  $m$ , and  $c$  is a candidate sequence of length  $n$ .  $\beta$  is set to a large number to make the calculation recall-oriented.  $WLCS(\textit{F-measure})$  is also known as the ROUGE-W metric.

#### 2.8.4 ROUGE-S

ROUGE-S is a metric which measures the skip-bigram co-occurrences between two sequences. A skip-bigram is any pair of words which are in the correct order in the sentence, even if arbitrary gaps separate them. For example, the sentence “I have it” contains 3 skip-bigrams: (“I have”, “I it”, “have it”) This set of metrics measures the co-occurrences between these types of bigrams. The formulas for this metric are shown in Equation 2.24 through Equation 2.26.

$$SKIP2(\textit{precision}) = \frac{SKIP2(r, c)}{Comb(n, 2)} \quad (2.24)$$

$$SKIP2(\textit{recall}) = \frac{SKIP2(r, c)}{Comb(m, 2)} \quad (2.25)$$

$$SKIP2(\textit{F-measure}) = \frac{(1 + \beta^2) * SKIP2(\textit{recall}) * SKIP2(\textit{precision})}{SKIP2(\textit{recall}) + (\beta^2 * SKIP2(\textit{precision}))} \quad (2.26)$$

In the formulas for this metric,  $SKIP2(r, c)$  is the number of skip-bigram co-occurrences

between reference sequence  $r$  of size  $m$  and candidate sequence  $c$  of size  $n$ .  $\beta$  controls the relative importance between precision and recall.  $Comb$  is the function which returns the result of a mathematical combination given the sequence length parameter.  $SKIP2(F\text{-measure})$  is also known as the ROUGE-S metric.

## 2.9 Background Summary

In this chapter, we have provided the background information for this thesis. We have explained the context for the problem of multi-document summarization and the challenges associated with it. The context we have provided includes a broader overview of the problems included within natural language processing as well. We have also described the research regarding the techniques and concepts which inform our approach, including the architectures, models, and techniques which serve as the foundation for approaches to multiple problems within computer science and natural language processing specifically.

# Chapter 3

## MMR-Based Approach

This chapter describes and explains the methods used in this MMR-based approach.

### 3.1 Maximal Marginal Relevance

Maximal Marginal Relevance (MMR) is a ranking algorithm in which documents are ranked according their combined query relevance and novelty of new information (Carbonell and Goldstein, 1998). The formula for the MMR calculation is as follows:

$$\text{Arg max}_{D_i \notin S} [\lambda(\text{Sim1}(D_i, Q)) - (1 - \lambda) (\max_{D_j \in S} \text{Sim2}(D_i, D_j))] \quad (3.1)$$

In the MMR formula, the lambda constant  $\lambda$  is a number between 0 and 1 which determines the degree to which the calculation prioritizes relevance or diversity.  $S$  is the set of documents which have already been selected, which can be whole documents or single sentences.  $D_i$  is the given candidate document which is not selected,  $D_j$  is the given previously-selected document to which the candidate document's similarity is compared, and  $Q$  is the query document to which the relevance of each candidate document is computed. Given a desired number of documents to select, the MMR calculation iterates through the unselected documents and selects the desired number of documents that are the most relevant or the most diverse, depending on the lambda constant.  $\text{Sim1}$  and  $\text{Sim2}$  are similarity measurements between documents. However, given the  $1 - \lambda$  statement preceding  $\text{Sim2}$ , the right-hand side of the calculation effectively becomes a maximization of diversity rather than similarity. A higher lambda constant increases relevance to the query, while a lower

lambda constant increases diversity among the selected documents.

MMR has been incorporated into several previous summarization approaches (Carbonell and Goldstein, 1998; Kurmi and Jain, 2014; Chaudhari and Mattukoyya, 2018; Mao et al., 2020). While various degrees of diversity and relevance were found to be optimal for summarization, all such previous approaches found at least some degree of diversity to be necessary for high-quality summarization results. However, to our knowledge, none of these approaches combine multiple state-of-the-art models’ outputs using MMR. We predict promising combinations of state-of-the-art summaries using high-relevance MMR calculations.

## **3.2 Similarity Measures**

The MMR algorithm uses two similarity calculations, which can be any similarity measurement without affecting the effectiveness of the algorithm itself. However, the outcome of the algorithm can be affected by the measures used. Different methods for determining document similarity can employ different vectorization and calculation methods, which can affect the similarity output. For the purposes of this research, we consider the cosine similarity measure in combination with the term frequency-inverse document frequency (TF-IDF) statistic, as well as the Doc2Vec model. We also consider the Word Mover’s Distance (WMD) similarity measure, but we found that it was not useful for our smaller-scale Multi-News experiment. However, we found that the WMD similarity measure was useful in our larger-scale WCEP experiment.

### **3.2.1 TF-IDF**

Term frequency-inverse document frequency (TF-IDF) is a statistic ascribed to a word or term, which denotes that word’s relative significance to the overall document. This denotation is performed by weighting the term’s frequency within the document by the term’s frequency among all the given documents. The first component of this statistic, term fre-

quency (TF), generally refers to the number of times the word occurs in the document. The second component, the inverse document frequency (IDF), refers to the logarithmically scaled fraction of documents which contain the term (Jones, 1972). The IDF acts as a weight by which the TF measure is multiplied, providing an estimate as to the significance of the term. The formulas for the versions of these statistics we use can be seen in Equation 3.2.

$$\begin{aligned}
 TF(t, d) &= \log_{10}(f_{t,d} + 1) \\
 IDF(t, d) &= \log \frac{|D|}{1 + |\{d : t \in d\}|} \\
 TF-IDF(t, d) &= TF(t, d) * IDF(t, d)
 \end{aligned}
 \tag{3.2}$$

In the formula for TF,  $f$  is the frequency of term  $t$  in document  $d$ , which is compressed with the logarithmic expression with 1 added to prevent computing the log of 0. The formula for IDF consists of the cardinality of document set  $D$  being divided by 1+ the number of documents  $d$  containing term  $t$ , and then logarithmically scaled. The TF-IDF formula simply contains a multiplication of the term's and document's TF by their IDF.

In our approach, we use a TF-IDF vectorizer, which computes the TF-IDF for every term within each document, in conjunction with the cosine similarity measure. Specifically, we use the TF-IDF vectorizer provided by the scikit-learn library<sup>2</sup>. We refer to this use of the TF-IDF statistic with cosine similarity between sentences as our TF-IDF similarity measure.

### 3.2.2 Doc2Vec

As an alternate means of vectorizing our inputs for similarity calculation, we use the Doc2Vec vectorization method, which in turn employs the same principles as the Word2Vec vectorization algorithm proposed by Mikolov et al. (2013). More specifically, we use the

<sup>2</sup>[https://github.com/scikit-learn/scikit-learn/blob/15a949460/sklearn/feature\\_extraction/text.py#L1521](https://github.com/scikit-learn/scikit-learn/blob/15a949460/sklearn/feature_extraction/text.py#L1521)

Doc2Vec model provided by Gensim<sup>3</sup>. The Word2Vec algorithm uses a shallow, one-layer neural network architecture. There are two variants of the Word2Vec architecture: the continuous bag-of-words model (CBOW), and the skip-gram model. A visualization of these two models can be seen in Figure 3.1.

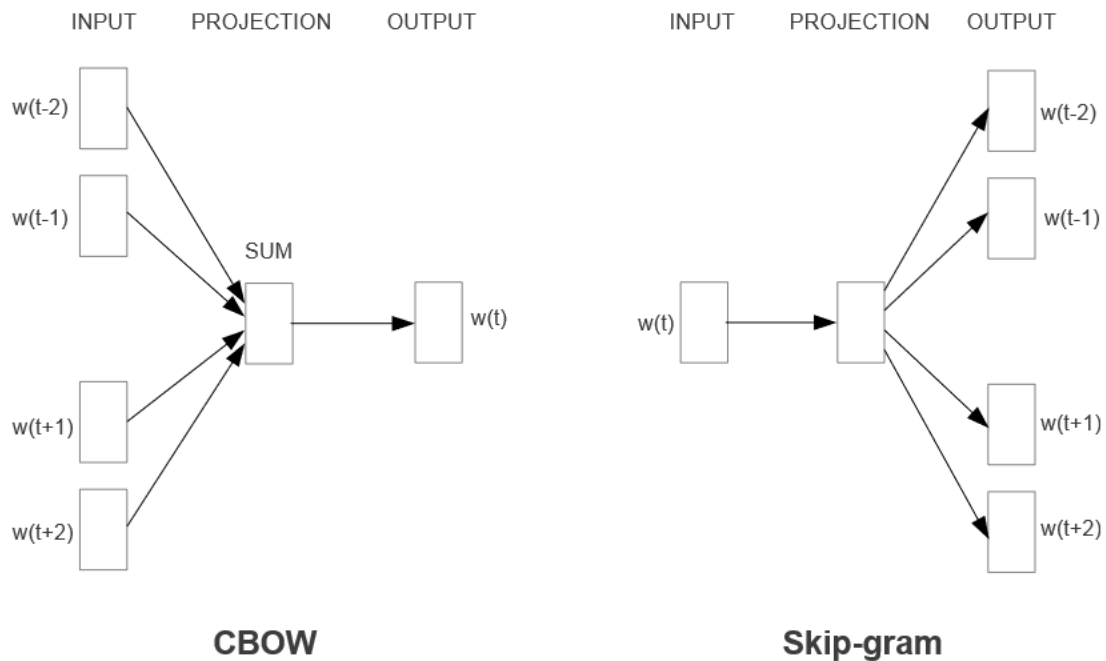


Figure 3.1: Word2Vec model architectures (Mikolov et al., 2013).

### Continuous Bag-of-Words (CBOW)

Bag-of-words (BOW) is a model for representing information which uses a simpler metric than TF-IDF. Rather than weighting the term frequencies by inverse document frequency, the BOW model considers only the frequency of each term. These frequencies are counted by considering the full document or corpus as a bag containing words, from which the words are counted one-by-one. These word counts are logged using a dictionary generated on the text. The order of the words in the dictionary, rather than the order of the words in the source document, determines the order of words in the BoW model's output vector.

<sup>3</sup><https://github.com/RaRe-Technologies/gensim/tree/develop/gensim/models>

For example, if we suppose that we have a dictionary containing the words: [‘lions’, ‘tigers’, ‘bears’, ‘ballerinas’]. Given this dictionary, if we represented the string “bears tigers lions lions lions tigers” as a BoW vector, we would have the vector representation: [3, 2, 1, 0], where the entries are in the same order as in the dictionary.

A continuous bag-of-words (CBOW) model extends this concept to use the context words surrounding a word to represent it rather than the word count itself. In this model, the inputs are the context words within a certain window size. All of the inputs are passed to the embedding layer. A dense softmax operation is then performed on the averaged context embedding to generate the predicted target word.

### **Continuous Skip-gram**

Skip-grams are continuous sequences of words that can have words between them. The skip-gram model uses pairs of words generated by moving a window of skip-gram detection. These skip-gram pairs are used by the neural network to predict the context words surrounding the given input word. Essentially, each word is the input for a classifier which can predict words within a certain distance in either direction. For our approach, we use the distributed bag-of-words (DBOW) model implementation within Gensim’s Doc2Vec architecture (Le and Mikolov, 2014), which is analogous to the skip-gram model used in the Word2Vec algorithm. The DBOW model uses predictions of randomly-sampled words within the document, the results of which can then be used for classification and the prediction of context words given an input word. This random sampling can cause the end result of an approach using Doc2Vec, such as ours, to vary somewhat.

### **3.2.3 Cosine Similarity**

Cosine similarity is a general-purpose measure of similarity between two non-zero vectors. In the context of natural language processing, cosine similarity, or the cosine kernel, measures the similarity between two vector representations of documents. This similarity is defined as the cosine of the angle between the vectors, or the Euclidean distance between



them. This distance between the vectors is calculated in the same manner as the Euclidean dot product formula for each two vectors, or the normalized dot product. The formula for cosine similarity can be seen in Equation 3.3

$$K(X, Y) = \frac{X \cdot Y}{\|X\| * \|Y\|} \quad (3.3)$$

In the formula for cosine similarity, X and Y are two inputs for the cosine kernel, and their dot product is normalized by their normed vectorspaces. This similarity measure is used with TF-IDF and BoW vectors as the inputs, which produces different similarity results.

#### **3.2.4 Word Mover’s Distance (WMD)**

Similar to possible implementations of cosine similarity, Word Mover’s Distance (WMD) employs Word2Vec vector embeddings to determine the similarity or dissimilarity between two word vectors (Kusner et al., 2015). The method effectively generates the normalized bag-of-words (nBOW) distributions for two documents before calculating the most efficient way to transport one distribution into the other. Each word in the distribution is allowed to be transformed as a whole or in parts, and the cost of the transformation is considered. The minimum weighted sum of costs for moving all words from one document to the other is known as the distance between the documents, which is calculated by the WMD similarity measure.

### **3.3 Preliminary Research Overview**

A number of alternative approaches were considered prior to the development of our current approach. This consideration with some limited-scope experimentation guided our choice of models, combination approach, and parameters. For all of our preliminary research, we used the Multi-News dataset (Fabbri et al., 2019). While we did not run each experiment on the entire test subset due to the required computing power, we ran each ex-

periment on a minimum of 100 samples to gauge the performance of the given approach and aid our decision on our final approach.

### 3.3.1 Preliminary Use of Models

We first considered the simple combination of an extractive model with a state-of-the-art abstractive model. For the extractive model, we considered the MatchSum architecture using the BERT model as proposed by Zhong et al. (2020). For the abstractive model, we considered the PEGASUS model proposed by Zhang et al. (2020). Among the methods for this combination, we considered simply generating an abstractive summary using the source document concatenated to an extractive summary as input. This approach was similar in principle to the now-common two-stage extractive-abstractive summarization approach, but employing two pretrained state-of-the-art models in place of newly-developed and newly-trained language models. We determined this approach to be ineffective by all metrics, with significantly reduced ROUGE-1, ROUGE-2, and ROUGE-L scores. While the exact reason for this decrease in summarization quality has not been determined, we suspect that it relates to the relative difference in quality between different model outputs.

This decrease in summarization quality was consistent regardless of whether the extractive output was used alone as abstractive input, used as abstractive input and then concatenated to an abstractive summary of the source text, or simply concatenated to an abstractive summary of each single document extracted from the source document cluster. Furthermore, we determined that the simple concatenation of whole model outputs effectively reduced the performance of the best model output. However, it was observed that selectively adding certain sentences from other models to the output of the best-performing model could generate improved results in at least one ROUGE metric.

We also considered the use of MMR using only one extractive model and one abstractive model. However, this approach yielded similarly unpromising results. We therefore determined that the use of a greater variety of pretrained models could increase the available

pool of summarization output for combination, which could display the potential benefits of diversity-based output combination as intended by our research.

### 3.3.2 Preliminary MMR Parameters

We also considered several combinations of parameters for our use of the MMR algorithm. The significant changeable parameters of the MMR algorithm are the query document used for relevance, the lambda constant used to determine the relative significance of the relevance, and the similarity measures used.

We briefly experimented with the use of source-extracted or summary-extracted sequences as query documents, but found this approach to lack conciseness and accuracy. We therefore instead decided upon the use of a separate generative model to extract a suitable query document from the source document cluster. We describe this process more fully in Section 3.6.1.

We also experimented with various degrees of relevance and diversity weighting in our use of the MMR algorithm. The standard use of an MMR algorithm typically begins with a lambda constant of 0.5, which evenly weights the diversity and query-relevance of the documents being selected from. However, we quickly noticed that a lambda constant either close to 1 or close 0 can vastly improve the sentence-extraction process, depending on the summarization framework and the query-generation process. These findings seem to be largely in keeping with the findings of research based on diversity-focused summarization (Nema et al., 2017; Fuad et al., 2019). We first considered a diversity-focused approach in which the entire summary output is generated by extracting the most diverse sentences from the various model outputs. A visual representation of our approach can be seen in Figure 3.2.  $l - r$  denotes each model-generated summary being reduced by at least one sentence via MMR, where  $l$  is the original number of sentences in the summary.  $n$  denotes the final number of sentences being included in the summary.

In this early approach, we first removed  $r$  least-diverse sentences from each model out-

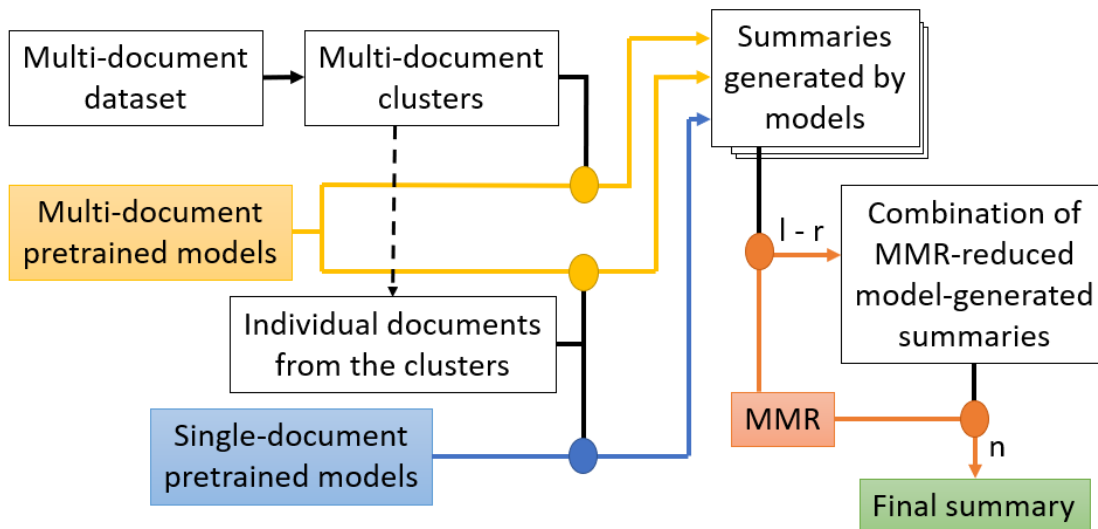


Figure 3.2: Overview of our preliminary MMR-based approach.

put using a high lambda constant, where  $r$  is  $\max(1, n * p)$ , and  $p$  is a low percentage of the summary’s number of sentences  $n$ . We then combine the remaining sentences into a summary of similar length to the MDS-pretrained model output. The use of similarity measures in the MMR algorithm allows for a significant degree of experimentation with various metrics for similarity between sentences. It is possible to use different similarity measures for the reduction of different models, or even different measures for the left and right sides of the MMR algorithm’s internal calculation. Despite this potential for the use of different similarity measures in the same MMR computation, we generally found it most effective to consistently use the same similarity measure across both sides of the algorithm, although with shorter summary lengths, the second similarity measure is rarely used. We used cosine similarity for all of the early trial runs of our approach. However, we found that the vectorization method used affected the MMR output significantly, especially when using different vectorization methods for different model outputs. The best candidates we found for similarity measures in our experiments were cosine similarity using the TF-IDF vectorizer and cosine similarity using the Doc2Vec vectorizer. We found that the Doc2Vec vectorizer worked best for MMR calculations of all models except the Longformer-Encoder-Decoder

(LED) model, for which the TF-IDF vectorizer proved most effective for MMR computations. We therefore use cosine similarity with the TF-IDF vectorizer for the LED model outputs, and cosine similarity with the Doc2Vec vectorizer for all other outputs.

However, while this early approach showed some improvement over MMR-based extraction without output reductions, the resulting ROUGE scores indicated that our complete reconstruction of the summary using MMR was inferior to that of the best-performing pre-trained model. We therefore devised our current approach, which exchanges output reduction in favor of reconstructing only a percentage of the summary using MMR.

### 3.4 Approach Overview

A visual representation of our current base approach can be seen in Figure 3.3.  $n$  denotes the output length of the best-scoring model in sentences, which in turn represents the output sentences themselves which are to be either replaced or appended with MMR-extracted sentences from other model outputs.  $m$  is the desired output length of the final summary including the MMR-extracted sentences, which is model output length  $n$  plus the desired length difference  $l$ .  $l$  can have a range of values, but we set  $l$  to  $n * p$ .  $p$  is the percentage constant, which denotes the percentage of the final output that will be constituted by MMR-extracted sentences.  $r$  is the number of sentences to be removed from the given model output using MMR, with the value of  $r$  being the maximum between 1 and the output length  $n$  times some very small percentage. The MMR reduction of sentences played no role in the WCEP experiment, because the model outputs were usually only 1 sentence in length. Additionally we use the  $k$  constant, which is the number of documents we take from the cluster for SDS generation. For datasets which have only a few documents per cluster such as Multi-News, we take all documents. For datasets which contain 100 or more documents per cluster such as WCEP, we take only  $k$  documents.

Additionally, the final output length of the dataset’s gold summaries determine whether the SDS outputs are concatenated into a single output or selected from as separate outputs.

For example, if a dataset contains 1 sentence per document and 100 documents per cluster such as some examples from the WCEP dataset, then appending our MMR-extracted output to a concatenation of multiple SDS outputs would generate a summary too large to be comparable to the dataset’s gold summary. We therefore only concatenate the SDS summaries into a single output for the model when doing so generates a summary that has fewer sentences than a single document.

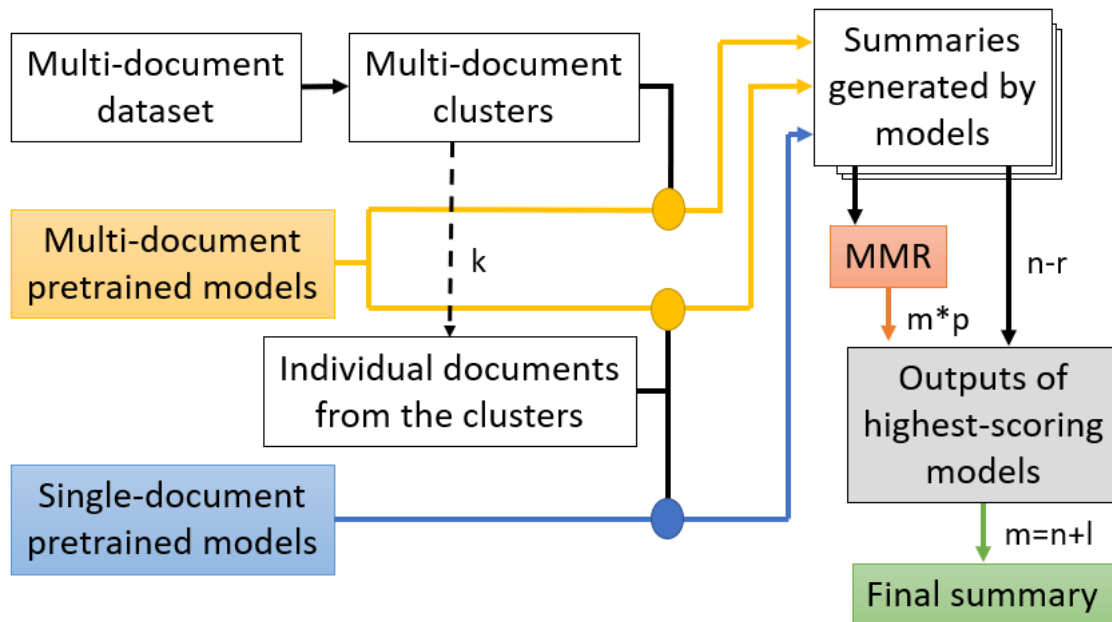


Figure 3.3: Overview of our final MMR-based approach.

For example, this approach gives the option to specify a final output of length  $n + 2$  and a  $p$  value of 0.9, which would reconstitute 90% of the final output using MMR, with the remainder of the  $n + 2$  output being composed of original output from the best-scoring model. For simplicity, we set  $l$  to  $\max(l, n*p)$ , which means the entire best model output of  $n$  length is unchanged, but these sentences are concatenated with at least 1 sentence which is extracted using MMR, and as many as  $n*p$  sentences. The number of sentences extracted with MMR therefore depends on the number of input sentences combined with the MMR percentage. Our use of a max expression ensures that the final output is never a simple repetition of the best-scoring model’s output, as the intention of our research is to explore

the improvement of model output using MMR.

### **3.5 Model Usage**

Our approach uses multiple pretrained models, including those which are pretrained on MDS datasets and those which are pretrained on SDS datasets. Training all of these models on the dataset being used would add to computation costs and be unrelated to the goals of this research, which we intend to prove effective without the need for optimal finetuning. Therefore, if a model pretrained on the MDS dataset being summarized is available, we use it. If a model is only available as either a base model or pretrained on a SDS news-related dataset, we use the available SDS pretrained dataset. Preliminary research indicates that SDS-pretrained models can perform adequately on MDS datasets, particularly if the MDS dataset clusters are split into single documents.

For the MDS-pretrained models, we run the model on the MDS cluster as a single flattened document to generate a MDS summary. We also use these MDS-pretrained models to generate SDS summaries from the single documents extracted from the MDS dataset clusters. This generation of SDS summaries from the single documents is also performed by the SDS-pretrained models. The document clusters can be split into single documents using special separation tokens, which allows each model to incorporate the entire document cluster’s context for summary generation.

There are both abstractive and extractive summarization versions of these various models available for use. We employ either the abstractive or extractive version of the models depending on preliminary evaluations of overall summary quality.

### **3.6 MMR Usage**

After the summaries of are generated by all of the pretrained models, the models’ outputs are processed using MMR. For our MMR calculations, we consider the sentences of model outputs as documents, and process the model outputs on the sentence level. The

use of unbroken sentences allows us to reconfigure the model outputs for improved summarization without compromising the readability and correctness of the model-generated summaries.

MMR calculation can favor either relevance to the query document or diversity among the selected documents. For our approach, we favor relevance by using a higher lambda constant, resulting in MMR calculations which select the k-most relevant sentence, where k is the number of sentences being selected via MMR. For our query document, we generate a small sequence of topic words from the source document cluster.

### **3.6.1 Latent Dirichlet Allocation (LDA)**

Preliminary experiments revealed that the query document selection process can strongly affect the overall quality of the produced final summary accuracy, particularly when using a relevance-focused approach. We chose to generate our query documents using Latent Dirichlet Allocation (LDA) topic modeling. LDA is a statistical model which incorporates Bayesian calculations to infer the most probable topic words from the given corpus (Blei et al., 2003). This statistical model makes use of the assumption that each document contains latent topics, each of which is distributed over the words of the document, which themselves are considered as a random blend of the various latent topics. The model functions by first choosing the topics of the document, and then selecting words to represent each topic.

### **3.6.2 MMR Reduction**

In order to reduce the chance of corrupt or irrelevant information being included in our final MMR calculation while increasing the relevance, we also make use of MMR to reduce the sentences of each model's output by a certain percentage if the model is not pretrained on the given dataset. Early experimentation revealed that this additional usage of MMR improves the results when there are a large number of sentences to select from, potentially revealing a minor shortcoming in the use of LDA topic word queries. However,



this reduction is only performed in the case of datasets that produce numerous sentences per summary. When there are sufficient sentences per summary, each model’s summary is reduced by MMR to a size of  $\max(1, n - \max(1, n \times r))$ , where  $n$  is the size of the summary in sentences and  $r$  is the percentage to reduce by, as small as 1%. Thus far in our experiments, this reduction percentage is small enough that each summary is reduced by 1, although in datasets with more sentences a more optimal percentage might be found.

Preliminary research indicates that this simple reduction of sentences improves overall summarization performance. It was also found that some model outputs may benefit from the use of certain similarity measurement methods, such as the use of TF-IDF vectorization with cosine similarity for one model and Doc2Vec vectorization with cosine similarity for another. This use of MMR for model output reduction essentially selects the  $n - 1$  most-relevant sentences from each model’s output, where  $n$  is that number of sentences originally generated by the given model. In the case of datasets that produce summaries of only one or two sentences, no MMR reduction is used.

### 3.6.3 Output Combination

The outputs are selectively combined via an extension of the best model’s output with MMR-extracted sentences from other models. After all of the model outputs are combined into a series of sentences, a number of sentences are selected from this single sequence to create the final summary using MMR. The quantity of sentences is selected to be comparable to the number of sentences generated by a model pretrained on the MDS dataset. The sentence amount produced by the best-performing model is used as the basis for our final output. Rather than composing the entire final output from various model outputs of various qualities, we extend the best model’s output length by adding a limited percentage of MMR-selected sentences. After selecting the  $j$ -most diverse sentences where  $j$  is the desired number of sentences, we add it to the current best model outputs to construct our final MMR-combined summary.

### **3.7 Summary**

In this chapter, we have described in high-level terms the means by which we produce summaries using our MMR-based combination approach. We have described our method for using state-of-the-art pretrained models together with MMR calculations to prioritize relevance among generated text. We have described how the model outputs are incorporated into the final output using MMR, as well as how the amount of MMR output is determined using the percentage parameter.

# Chapter 4

## Multi-News Experiment and Results

This chapter will describe and explain the experimentation we used in applying our MMR-based approach to the Multi-News dataset. The results of this experimentation will also be presented and discussed.

### 4.1 Experimental Setup

We ran our approach on the testing split of a dataset to generate MMR-extracted summary additions from all of the outputs of the models we used for our experiment.

#### 4.1.1 Dataset

We used the Multi-News Dataset, which consists of professionally-written summaries paired with news article document clusters which are all in the English language (Fabbri et al., 2019). The 5622 multi-document clusters from the testing subset of this dataset were split for single-document pretrained models using the dataset’s special separation token, and also used as single flattened documents for those models trained on the Multi-News dataset. We also truncated each document to a constant length of 1048 to avoid potential memory issues or token-processing errors with the various models’ tokenizers.

#### 4.1.2 Implementation Details

We ran our experiment on Google’s Colaboratory Pro platform using their GPUs<sup>4</sup>. The GPUs available on the Colaboratory platform include Nvidia K80s, T4s, P4,s and P100s,

---

<sup>4</sup><https://colab.research.google.com>

although the exact GPU being used cannot be selected by the end user (Google, 2021). For our source code implementation, we used the PyTorch numerical framework (Paszke et al., 2019).

### 4.1.3 Model Versions

The model selection process was based primarily on state-of-the-art summarization results, as well as the availability of pretrained versions of the model. If a state-of-the-art model was available pretrained on a MDS dataset, it was used. Otherwise, if it was available pretrained on a SDS dataset with similar text content such as the CNN-DM dataset (Hermann et al., 2015; Nallapati et al., 2016), it was used. If only the base pretrained model was available, the base or large version was used depending of memory concerns and overall expected increase in performance between the two versions based on preliminary experimentation. In total, we used three extractive summarization model and five abstractive models. For all models except MatchSum, XLNet, and GPT-2, we used the models available through the HuggingFace platform<sup>5</sup> created by Wolf et al. (2020) for ease of use and experimentation.

For XLNet and GPT-2, we used the model versions available through the Bert Extractive Summarizer library<sup>6</sup>. This library was developed by Derek Miller for his extractive summarization approach for lectures (Miller, 2019). The models available through this library are simply placed within the architecture with no need for finetuning (Miller, 2019). Multiple models are available through this library, although they were not included in the library’s seminal research paper.

### MatchSum

For our main extractive approach, we use the text-matching architecture used by Zhong et al. (2020) in their implementation of summarization as semantic text-matching (Match-

---

<sup>5</sup><https://huggingface.co/>

<sup>6</sup><https://pypi.org/project/bert-extractive-summarizer/>

Sum). This model was pretrained on the Multi-News dataset, and its output on the Multi-News dataset was readily available<sup>7</sup>.

## PEGASUS

We used the PEGASUS abstractive language model from Google (Zhang et al., 2020). This is the only abstractive model for which a version was available which was pretrained on a MDS dataset. As such, the PEGASUS<sub>multi-news</sub> model was used to generate a standard multi-document summary as per its pretraining method, as well as single document summaries from the single documents extracted from the document cluster. The MDS output for this model did not have the least-diverse sentence removed, as preliminary research showed that such filtering was not necessary for models which were pretrained on the MDS dataset. We use the version of this model available through the Huggingface platform<sup>8</sup>.

## BART

The BART model (Lewis et al., 2020) was used. In particular, we chose the BART<sub>large-cnn</sub> model, as this version was pretrained on a task which was most similar to that of summarizing the single Multi-News documents. This model was available through the HuggingFace platform<sup>9</sup>.

## T5

We also employed the T5 model (Raffel et al., 2020) for summarization. As no version pretrained on a similar dataset was available, we used the T5<sub>large</sub> model available through the Huggingface platform<sup>10</sup>.

---

<sup>7</sup><https://github.com/maszhongming/MatchSum>

<sup>8</sup><https://huggingface.co/>

<sup>9</sup><https://huggingface.co/>

<sup>10</sup><https://huggingface.co/>

## XLNet

We also included the results of the XLNet model (Yang et al., 2020a) in our experiment. Due to the large memory consumption potential for this model, as well as the negligible difference in performance between the base and large models, we used the XLNet<sub>base</sub> model. For ease of use, we used the extractive implementation available from the Bert Extractive Summarizer library<sup>11</sup>.

## GPT-2

The GPT-2 was also among the state-of-the-art models used (Radford et al., 2019). Due to memory concerns and negligible difference between the outputs of the large and base versions, we used the GPT-2<sub>base</sub> model. Like the XLNet model, we used the GPT-2 model available from Miller’s Bert Extractive Summarizer library<sup>12</sup>.

## ProphetNet

The ProphetNet (Qi et al., 2020) model was also included in our experiment. As a large model pretrained on the CNN-DM dataset was available, we used the ProphetNet<sub>large-cnndm</sub> model through the HuggingFace platform<sup>13</sup>.

## LED

We also used the LED model (Beltagy et al., 2020). No version of this model pretrained on a similar dataset was available, so we used the LED<sub>base</sub> model available through the Huggingface platform<sup>14</sup>.

### 4.1.4 MMR Parameters

The parameters for our MMR calculations were consistent across all usages, with the same lambda constant, query document, and similarity measure being used for each calcu-

<sup>11</sup><https://pypi.org/project/bert-extractive-summarizer/>

<sup>12</sup><https://pypi.org/project/bert-extractive-summarizer/>

<sup>13</sup><https://huggingface.co/>

<sup>14</sup><https://huggingface.co/>

lation. We optimized our MMR parameters for each similarity measure using the black box optimization library Scikit-Optimization<sup>15</sup>.

### **MMR Percentage**

Our approach involves a selective, limited inclusion of MMR-extracted sentences for our final output. Therefore, the percentage of our final output that is MMR-extracted was expected to be relatively low. Our optimization efforts confirmed our expectations by suggesting that a low MMR percentage would be optimal regardless of the similarity measure used. The resulting parameter values can be seen in Table 4.1. The optimized percentage values were used in our experiments.

### **Lambda Constant**

For all of our MMR calculations, we used a high lambda constant to prioritize relevance to the query document over diversity among the selected documents. The optimized relevance-focused lambda constant can be see in Table 4.1.

### **Query Document**

For the query documents of our MMR algorithm implementation, we extracted sequences of topic words from the source document cluster using the LDA statistic model. This sequence represents the overall topic of the document used as input. We initially experimented with the LDA model provided by the Gensim library<sup>16</sup>. However, for improved topic word selection, we use the MALLET toolkit<sup>17</sup>. Preliminary research indicated that the MALLET toolkit's implementation of the LDA model improved the relevance of the topic words, which in turn reduced the variability of the quality of our outputs. We optimized the number of topics and number of words using the Scikit-Optimization library<sup>18</sup>, the results of which can be seen in Table 4.1.

---

<sup>15</sup><https://scikit-optimize.github.io/stable/>

<sup>16</sup><https://radimrehurek.com/gensim/models/ldamodel.html>

<sup>17</sup><http://mallet.cs.umass.edu/>

<sup>18</sup><https://scikit-optimize.github.io/stable/>

### Similarity Measures

We considered the use of two similarity measures for our Multi-News experiment: our TF-IDF similarity measure, which combines cosine similarity with a TF-IDF vectorizer, and our Doc2Vec similarity measure, which combines cosine similarity with a Doc2Vec vectorizer. Because we did not need to use single documents from the clusters, and because WMD was found to be inferior for the tasks of both MMR reduction and the final MMR combination, we did not use the WMD similarity measure in this experiment. Preliminary testing revealed that the vectorization method significantly affects model output selection, particularly with LED model outputs. For the final MMR combination method, we found the Doc2Vec similarity measure to be most effective. This method was also found to be most effective for the MMR-reduction of all model outputs with the exception of LED, for which we used the TF-IDF similarity measure.

Table 4.1: Optimized final MMR combination parameters

Parameter	TF-IDF value	Doc2Vec value
lambda constant	0.9742801322275981	0.8084240487988124
MMR percentage	0.1066062697911007	0.1059268707588193
LDA topics	3	5
Words per topic	7	6

#### 4.1.5 Evaluation Metric

We used the F-measures from the ROUGE-1, ROUGE-2, and ROUGE-L metrics (Lin, 2004) to compare the term occurrences from our MMR-generated summaries with those from the human-generated summaries in the Multi-News dataset.

#### 4.1.6 Baseline Measures

We calculated the same F-measures from the ROUGE-1, ROUGE-2, and ROUGE-L metrics (Lin, 2004) for the unaltered summaries generated by the state-of-the-art models for comparison with our approach. As this experiment is to demonstrate the possibility



of few-shot improvement using MMR, we also compare to the results of the Multi-News dataset’s authors’ Hi-MAP approach (Fabbri et al., 2019).

## 4.2 Results

Our approach’s ROUGE F-measure results on the test subset of the Multi-News dataset can be seen in Table 4.2. Our MMR-based approach outperforms all models in the ROUGE-1 (R1) metric, but lags behind PEGASUS in the ROUGE-2 (R2) and ROUGE-L (RL) metrics. (SDS) denotes that the model was used to summarize the individual documents from the document cluster, and (MDS) denotes that the model was used to summarize the complete document cluster as a flattened document.

Table 4.2: Results of our Multi-News experiment

Model	R1	R2	RL
ProphetNet <sub>large-cnndm</sub>	32.01	10.43	16.46
GPT-2 <sub>base</sub>	32.56	9.87	16.34
XLNet <sub>base</sub>	32.76	9.97	16.38
T5 <sub>large</sub>	32.95	10.27	16.62
LED <sub>base</sub>	33.27	10.83	16.84
BART <sub>large-cnn</sub>	36.64	12.09	18.19
MatchSum <sub>BERT-base</sub>	43.98	15.91	20.94
PEGASUS <sub>multi-news</sub> (SDS)	40.25	16.26	19.43
Hi-MAP (Fabbri et al., 2019)	43.47	14.89	17.41
PEGASUS <sub>multi-news</sub> (MDS)	45.79	<b>18.48</b>	<b>24.27</b>
MMR-combination (ours)	<b>46.23</b>	18.30	21.25

## 4.3 Discussion

It is immediately apparent that the ROUGE-L score our approach produces for the PEGASUS model are lower than those published in the original authors’ research (Zhang et al., 2020). While the exact reason for this discrepancy remains unclear, we suspect it may be due to the fact that we use the PyTorch numerical framework while Zhang et al. (2020) use the TensorFlow framework (Abadi et al., 2016). Nevertheless, while the absolute ROUGE-

L scores may be overall lower in our experiment’s results than in some those published by other researchers, the relative differences between the ROUGE-L scores of the various summarization approaches remain adequate for effective comparison.

As shown in Table 4.2, we also observe that the ROUGE-1 score of our MMR-based combination approach is higher than all of the models used. However, the ROUGE-2 and ROUGE-L scores of our approach are lower than those of the PEGASUS model when applied to the unmodified documents clusters as flattened documents. While the ROUGE-2 and ROUGE-L scores of our MMR-based approach are lower than those of the PEGASUS model, they remain higher than those of all other models with the exception of PEGASUS when applied in the standard MDS application. This relative discrepancy in ROUGE scores seems to indicate that, while our approach does not focus on the properties of multiple-word strings within the summaries, it nevertheless preserves state-of-the-art performance related to the properties of multiple-word strings.

Due to the increased ROUGE-1 score, it seems possible that there is at least a limited direct correlation between term relevance to LDA topic words and ROUGE scores for the term length the MMR extraction is based on. Furthermore, these scores represent the use of only two models which are pretrained on the correct dataset, the combination of which was determined via preliminary experimentation to decrease all ROUGE scores relative to the highest-scoring model PEGASUS. We can therefore infer that, while the application of our method to few-shot summarization is substantial and its preservation of state-of-the-art performance is sufficient, the increase that comes from training a state-of-the-art model might be greater. Nevertheless, the question of whether this method provides further improvement in a zero-shot application remains, and will be addressed in Chapter 5.

## **4.4 Summary**

In this chapter, we have detailed the specific details of our experiments using the Multi-News dataset. We have included the manner in which we ran our experiments, the versions

and sources of the models we used, the parameters we incorporated into our MMR-based method, and the raw results of these experiments. We have also provided an explanation of the results and their implications.

# Chapter 5

## WCEP Experiment and Results

This chapter will describe and explain the experimentation we used in applying our MMR-based approach to the Wikipedia Current Events Portal (WCEP) dataset. This experiment further tests our approach using models which are not trained on the dataset in question. The results of this experimentation will also be presented and discussed.

### 5.1 Experimental Setup

We ran our approach on the testing split of a dataset to generate MMR-extracted summary additions from all of the outputs of the models we used for our experiment. The motivation for this experiment was to explore the scalability of our approach, and to test its effectiveness with 1) larger data, and 2) fewer-shot applications in which no models are pretrained on the given dataset.

#### 5.1.1 Dataset

We used the Wikipedia Current Events Portal (WCEP) dataset, which consists of short, human-written summaries of news events, the articles for which are all extracted from the Wikipedia Current Events Portal (Gholipour Ghalandari et al., 2020). Each document cluster contains a large quantity of automatically-extracted articles. There are two primary versions of this dataset: the full version (WCEP-total) and the truncated version (WCEP-100). The full version consists of 2.39 million articles, while the truncated version consists of 650,000 articles. Also, each article cluster is limited to 100 articles in the truncated version, while the full version can contain as many as 8411 articles per cluster. The test

split of the WCEP-100 dataset consists of 1,022 article clusters, of which we perform our experiment on 500 due to cost limitations. We use the WCEP-100 dataset version available through the WCEP dataset authors' GitHub repository<sup>19</sup>.

### 5.1.2 SDS Problem

Our approach involves the use of pretrained models which are either pretrained on the given dataset or not. In this experiment, none of the pretrained models we used were trained on the given dataset. Therefore, every model is used in a single-document summarization (SDS) format, with each model generating a separate summary for each article. Due to the large number of articles per cluster, it would be prohibitively expensive to perform this SDS operation for each model on each article. Additionally, the large number of documents per cluster would far outnumber the small number of sentences per summary, thus rendering our final output incomparable with the WCEP dataset's gold summaries. Therefore, we were faced with the problem of how to reduce the number of articles being summarized without sacrificing important information from the raw dataset. Our solution was to use MMR to select the 10 articles which were most relevant to our query document. Each model then produced a SDS output for each of these 10 articles. Preliminary research indicated that some of the source articles were irrelevant, too brief, or otherwise bad data for summarization input. Therefore our MMR-based solution seemed to be useful for excluding bad data in addition to reducing cost while preserving information.

Additionally, we faced the problem of which of the SDS outputs of the best model to choose to append our MMR-selected output to. We wanted to select the best summary from the best-performing model without apriori evaluation, and with a suitable final number of sentences for comparison with the summaries of the sentences themselves. To select the best summary, we use the cosine similarity measure with the Doc2Vec vectorizer, which is equivalent to the MMR algorithm with an output size of 1 sentence. We considered appending the MMR output to each of the best model's single-document summaries prior

---

<sup>19</sup><https://github.com/complementizer/wcep-mds-dataset>

to selecting the best one. While this method did perform better than the baseline models, it was not optimal, and increased the likelihood of an unreadable final summary being generated. We therefore chose to append the MMR output only after the best summary to append to was selected using our Doc2Vec similarity measure.

### 5.1.3 Implementation Details

As with the Multi-News experiment, we ran our experiment on Google’s Colaboratory Pro platform using their GPUs<sup>20</sup>. The GPUs available on the Colaboratory platform include Nvidia K80s, T4s, P4,s and P100s, although the exact GPU being used cannot be selected by the end user (Google, 2021). For our source code implementation, we used the PyTorch numerical framework (Paszke et al., 2019).

### 5.1.4 Model Versions

As with the Multi-News experiment, the model selection process was based primarily on state-of-the-art summarization results, as well as the availability of pretrained versions of the model. If a state-of-the-art model was available pretrained on a MDS dataset such as Multi-News, it was used. Otherwise, if it was available pretrained on a SDS dataset with similar text content such as the CNN-DM dataset (Hermann et al., 2015; Nallapati et al., 2016), it was used. If only the base pretrained model was available, the base or large version was used depending of memory concerns and overall expected increase in performance between the two versions based on preliminary experimentation. None of these models were trained on the WCEP dataset, nor were their hyperparameters optimized for this application.

For this experiment, we used 7 abstractive models and 3 extractive models. The model selection was slightly different in this experiment due to the different dataset. The Match-Sum extractive model used in our Multi-News experiment was not available for this dataset without reimplementing it, so we instead included both extractive and abstractive imple-

---

<sup>20</sup><https://colab.research.google.com>

mentations of the XLNet, BART, and GPT-2 models. Additionally, we included the abstractive implementations of the ProphetNet, GPT-2, LED, T5, XLNet, PEGASUS, and BART models. As was the case with our Multi-News experiment, all of the abstractive implementations used were taken from the HuggingFace platform<sup>21</sup> created by Wolf et al. (2020). As there was no MatchSum model in this experiment, all of the extractive implementations we used were taken from the Bert Extractive Summarizer library<sup>22</sup> developed by Miller (2019). We ensured that the model versions used in the extractive implementations were the same as the model versions used in the abstractive implementations.

### **PEGASUS**

We used the PEGASUS abstractive language model from Google (Zhang et al., 2020). Unlike the Multi-News dataset, no pretrained version of the PEGASUS model was available for the WCEP dataset. We therefore used PEGASUS in an SDS application the same as all the other models. We used the Extreme Summarization (XSum) version of this model available through the Huggingface platform<sup>23</sup>. The XSum dataset (Narayan et al., 2018) consists of BBC articles with single-sentence professionally-written summaries, which is a similar summary format to that of the WCEP summaries.

### **BART**

The BART model (Lewis et al., 2020) was used. As with the previous experiment, we chose the large-cnn version of the BART model, as this version was pretrained on a task which was most similar to that of summarizing the single Multi-News documents. We acquired the abstractive implementation of this model through the HuggingFace platform<sup>24</sup>, and the extractive implementation through the Bert Extractive Summarizer library<sup>25</sup>. This is the model we chose as our best-performing model to approach our MMR-extracted output

---

<sup>21</sup><https://huggingface.co/>

<sup>22</sup><https://pypi.org/project/bert-extractive-summarizer/>

<sup>23</sup><https://huggingface.co/>

<sup>24</sup><https://huggingface.co/>

<sup>25</sup><https://pypi.org/project/bert-extractive-summarizer/>

to.

## **T5**

We also employed the T5 model (Raffel et al., 2020) for summarization. As no version pretrained on a similar dataset to WCEP was available, we used the T5<sub>large</sub> model available through the Huggingface platform<sup>26</sup>.

## **XLNet**

We also included the results of the XLNet model (Yang et al., 2020a) in our experiment. Due to the greater need for large data processing in this experiment, we used the XLNet<sub>large-cased</sub> model. As with the BART model, we acquired the abstractive implementation of this model through the HuggingFace platform<sup>27</sup>, and the extractive implementation through the Bert Extractive Summarizer library<sup>28</sup>.

## **GPT-2**

The GPT-2 was also among the state-of-the-art models used (Radford et al., 2019). Due to memory concerns with the greater number of summarizations being performed iteratively and the already-large vocabulary of the GPT-2 model, we used the GPT-2<sub>small</sub> model version. As with the BART and XLNET models, we acquired the abstractive and extractive implementations of this model through the HuggingFace platform<sup>29</sup> and Bert Extractive Summarizer library<sup>30</sup>, respectively.

## **ProphetNet**

The ProphetNet (Qi et al., 2020) model was also included in our experiment. Due to the need for additional data processing with dataset-pretrained versions of this model,

---

<sup>26</sup><https://huggingface.co/>

<sup>27</sup><https://huggingface.co/>

<sup>28</sup><https://pypi.org/project/bert-extractive-summarizer/>

<sup>29</sup><https://huggingface.co/>

<sup>30</sup><https://pypi.org/project/bert-extractive-summarizer/>



we used the ProphetNet<sub>large-uncased</sub> model in this experiment. As with all other abstractive implementations, we acquired this version from the HuggingFace platform<sup>31</sup>.

## **LED**

We also used the LED model (Beltagy et al., 2020). No version of this model pretrained on a similar dataset was available, so we used the LED<sub>base</sub> model available through the Huggingface platform<sup>32</sup>.

### **5.1.5 MMR Parameters**

The parameters for our MMR calculations were consistent across all usages, with the same lambda constant, query document, and similarity measure being used for each calculation. We optimized our MMR parameters for each similarity measure using the black box Optuna optimization library<sup>33</sup>.

### **MMR Percentage**

Our approach involves a selective, limited inclusion of MMR-extracted sentences for our final output. Therefore, the percentage of our final output that is MMR-extracted was expected to be relatively low. Our optimization efforts confirmed our expectations by suggesting that a low MMR percentage would be optimal regardless of the similarity measure used. The resulting parameter values can be seen in Table 5.1. The optimized percentage values were used in our experiments.

### **Lambda Constant**

For all of our MMR calculations, we used a high lambda constant to prioritize relevance to the query document over diversity among the selected documents. The optimized relevance-focused lambda constant can be see in Table 5.1.

---

<sup>31</sup><https://huggingface.co/>

<sup>32</sup><https://huggingface.co/>

<sup>33</sup><https://optuna.org/>

### Query Document

As in the Multi-News experiment, we used the MALLET toolkit<sup>34</sup> to extract Latent Dirichlet Allocation (LDA) topic words for our query documents. Preliminary research indicated that the MALLET toolkit’s implementation of the LDA model improved the relevance of the topic words similarly to its use in the Multi-News experiment. We optimized the number of topics and number of words using the Optuna library<sup>35</sup>, the results of which can be seen in Table 5.1.

### Similarity Measures

We considered the use of three similarity measures for our WCEP experiment: our TF-IDF similarity measure, which combines cosine similarity with a TF-IDF vectorizer; our Doc2Vec similarity measure, which combines cosine similarity with a Doc2Vec vectorizer; and the Word Mover’s Distance similarity measure. Preliminary experiments revealed that the vectorization method significantly affects model output selection similarly to in our earlier Multi-News experiment. Given the added need for similarity-based selection in this experiment with a larger dataset and smaller summaries, we consider the three uses of similarity in this implementation:  $Sim_0$ ,  $Sim_1$ , and  $Sim_2$ .  $Sim_0$  is the similarity measure used to select the best summary from the best model to append our MMR output to.  $Sim_1$  and  $Sim_2$  are the similarity measures used in the query relevance and diversity calculations of the MMR algorithm, respectively. Note that the  $Sim_2$  similarity measure is rarely used in this implementation, because of the small number of returned sentences as well as the high lambda constant and relevance prioritization. The optimized similarity measures we used are included in Table 5.1.

---

<sup>34</sup><http://mallet.cs.umass.edu/>

<sup>35</sup><https://optuna.org/>

Table 5.1: Optimized MMR parameters

Parameter	Value
lambda constant	0.9970352972330873
MMR percentage	0.2979940927185918
LDA topics	5
LDA words per topic	2
Sim <sub>0</sub>	Doc2Vec
Sim <sub>1</sub>	WMD
Sim <sub>2</sub>	TF-IDF

### 5.1.6 Evaluation Metric

We used the F-measures from the ROUGE-1, ROUGE-2, and ROUGE-L metrics (Lin, 2004) to compare the term occurrences from our MMR-generated summaries with those from the human-generated summaries in the WCEP dataset. However, as stated in Section 5.1.2, only one MMR-generated summary per document cluster is included for evaluation.

### 5.1.7 Baseline Measures

We calculated the same F-measures from the ROUGE-1, ROUGE-2, and ROUGE-L metrics (Lin, 2004) for the unaltered summaries generated by the state-of-the-art models for comparison with our approach. These measures were calculated for every summary generated by the given model, and then averaged. These average model output scores were used for baseline comparison. As one of the purposes of this experiment is to demonstrate our MMR-based method in zero-shot applications, only these baseline models are considered for comparison with our approach. Some of these models' outputs are illegible or scrambled for some of the dataset clusters, but none of the models were found to produce bad outputs consistently for all document clusters.

## 5.2 Results

Our approach’s ROUGE F-measure results on the test subset of the WCEP dataset can be seen in Table 5.2. Our MMR-based approach outperforms all models in the ROUGE-1 (R1) and ROUGE-L (RL) metrics, but lagged behind in the ROUGE-2 (R2) metric. In this experiment, all models were employed using the SDS method of summarizing each article.

Table 5.2: Results of our WCEP experiment

<b>Model</b>	<b>R1</b>	<b>R2</b>	<b>RL</b>
XLNet <sub>base</sub>	12.89	2.70	9.48
ProphetNet <sub>large-cnndm</sub>	14.19	2.36	10.56
GPT-2 <sub>small</sub>	15.89	3.08	11.18
T5 <sub>large</sub>	24.14	5.98	16.95
LED <sub>base</sub>	24.38	8.55	16.72
BART <sub>large-cnn</sub> (EXT)	27.17	9.39	19.39
PEGASUS <sub>xsum</sub>	27.33	8.38	19.19
GPT-2 <sub>small</sub> (EXT)	27.33	9.42	19.52
XLNet <sub>base</sub> (EXT)	27.41	<b>9.49</b>	19.48
BART <sub>large-cnn</sub>	28.12	8.84	19.71
MMR-combination (ours)	<b>30.74</b>	9.18	<b>21.57</b>

## 5.3 Discussion

As shown in Table 5.2, we also observe that the ROUGE-1 and ROUGE-L scores of our MMR-based combination approach are higher than all of the models used. However, the ROUGE-2 score of our approach is lower than several of the baseline models including the extractive XLNet implementation, which produced the highest ROUGE-2 score. Nevertheless, our approach’s ROUGE-2 score is still comparable to state-of-the-art performances, particularly in this fewer-shot context. Also, it can be observed that the discrepancy between XLNet(EXT)’s ROUGE-2 output and our approach’s output is less than the combined discrepancies between these approaches’ respective ROUGE-1 and ROUGE-L scores. While the reasons for this stark contrast in ROUGE scores for our approach are not entirely clear, we suspect that the reason is related to the fact that we are appending our

MMR-extracted outputs to those of our chosen best-performing model, and all of the baseline models show a relative decrease in ROUGE-2 scores compared to their ROUGE-1 and ROUGE-L scores. Since our best-performing model BART displays similar performance differences, and because our approach utilizes Doc2Vec vectorization which uses random sampling, it is not unexpected that our MMR-produced output shows similar differences despite improvement in our chosen model’s higher ROUGE scores.

These results indicate to us that this MMR-based approach is more effective in zero-shot applications than in our earlier few-shot applications, which indicates that it is still more effective to train models on the given dataset than employ a zero-shot approach. Regardless, the WCEP experiment shows great promise for our approach, particularly given that our approach effectively doubles the summary size. This increase in size might appear, on the surface, to risk reducing the overall ROUGE scores. However, the selectivity of our MMR-based extraction appears to ensure that overall state-of-the-art quality is maintained in our final summaries.

## **5.4 Summary**

In this chapter, we have detailed the specific details of our experiments using the WCEP dataset. We have included the manner in which we ran our experiments, the specific challenges related to a larger dataset with fewer-shot summarization, the versions and sources of the models we used, the parameters we incorporated into our MMR-based method, and the raw results of these experiments. We have also provided a discussion on these results on their implications.

# Chapter 6

## Conclusion

### 6.1 Summary

In this thesis, we have implemented several state-of-the-art models for the summarization of the available multi-document datasets, including a few-shot application. We have designed a method for combining pretrained models for MDS using a statistical algorithm, MMR. We have configured this MMR-based method for the models used in our experimentation. Our experimental results with the Multi-News dataset show marked improvement in ROUGE-1 scores, with good preservation of state-of-the-art performance in other ROUGE scores. Additionally, our experimental results with the WCEP dataset show improvement in both ROUGE-1 and ROUGE-L scores. Results suggest that the relevance-focused combination of pretrained model outputs using MMR can improve summary quality, despite the lack of model finetuning and the loss of performance that comes from a non-selective approach. In Appendix A, we have provided some sample summaries which were generated from our approach in our Multi-News experiment, which can be observed to contain outputs similar to the PEGASUS model but including few MMR-extracted sentences in addition. Similar examples of our WCEP experiment can be observed in Appendix B.

### 6.2 Future Work

While the ROUGE-1 improvements produced by our approach suggest that query-relevance among pretrained model outputs can improve summarization, the question of whether the relevance of n-grams improves the ROUGE scores for those n-grams could be

the basis for future research, although the computation costs may be significantly higher. The question also remains of whether the overall ROUGE scores of our approach could be improved with an implementation of biases towards certain models based on their performances. Additionally, future research could potentially be focused on how the optimal pretrained model for this or a similar approach could be automatically selected.

While ROUGE scores remain an effective evaluation tool for comparing various models' performances, human evaluation, for which our research lacked the resources, could offer a more robust evaluation of MMR-based approaches.

# Bibliography

- Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: A system for large-scale machine learning. In *arXiv*, 2016.
- Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11):e00938, 2018. ISSN 2405-8440. doi: <https://doi.org/10.1016/j.heliyon.2018.e00938>. URL <https://www.sciencedirect.com/science/article/pii/S2405844018332067>.
- Muhammad Afzal, Fakhare Alam, Khalid Mahmood Malik, and Ghaus M Malik. Clinical context-aware biomedical text summarization using deep neural network: Model development and validation. *J Med Internet Res*, 22(10):e19810, Oct 2020. ISSN 1438-8871. doi: 10.2196/19810. URL <http://www.jmir.org/2020/10/e19810/>.
- Diego Antognini and Boi Faltings. Gamewikisum: a novel large multi-document summarization dataset. In *arXiv*, 2020.
- Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. In *arXiv*, 2016.
- Regina Barzilay, Kathleen R. McKeown, and Michael Elhadad. Information fusion in the context of multi-document summarization. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 550–557, College Park, Maryland, USA, June 1999. Association for Computational Linguistics. doi: 10.3115/1034678.1034760. URL <https://www.aclweb.org/anthology/P99-1071>.
- Iz Beltagy, Matthew E. Peters, and Arman Cohan. Longformer: The long-document transformer. In *arXiv*, 2020.
- David M. Blei, Andrew Y. Ng, Michael I. Jordan, and John Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:2003, 2003.
- Harold. Borko and Charles L. Bernier. *Abstracting concepts and methods*. Academic Press New York, 1975. ISBN 0121186504.
- Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a "siamese" time delay neural network. In *Proceedings of the*



- 6th International Conference on Neural Information Processing Systems, NIPS'93*, page 737–744, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc.
- Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *arXiv*, 2020.
- Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for re-ordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '98*, page 335–336, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 1581130155. doi: 10.1145/290941.291025. URL <https://doi.org/10.1145/290941.291025>.
- Asli Celikyilmaz and Dilek Hakkani-Tur. A hybrid hierarchical model for multi-document summarization. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 815–824, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/P10-1084>.
- Mayank Chaudhari and Aakash Nelson Mattukoyya. Tone biased mmr text summarization. In *arXiv*, 2018.
- Yen-Chun Chen and Mohit Bansal. Fast abstractive summarization with reinforce-selected sentence rewriting. In *arXiv*, 2018.
- Jianpeng Cheng and Mirella Lapata. Neural summarization by extracting sentences and words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 484–494, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1046. URL <https://www.aclweb.org/anthology/P16-1046>.
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *arXiv*, 2014.
- Sangwoo Cho, Chen Li, Dong Yu, Hassan Foroosh, and Fei Liu. Multi-document summarization with determinantal point processes and contextualized representations. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pages 98–103, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-5412. URL <https://www.aclweb.org/anthology/D19-5412>.
- Janara Christensen, Stephen Soderland, Gagan Bansal, and Mausam. Hierarchical summarization: Scaling up multi-document summarization. In *Proceedings of the 52nd*

- Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 902–912, Baltimore, Maryland, June 2014. Association for Computational Linguistics. doi: 10.3115/v1/P14-1085. URL <https://www.aclweb.org/anthology/P14-1085>.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *arXiv*, 2020.
- João Marcos Correia Marques, Fabio Gagliardi Cozman, and Ismael Humberto Ferreira dos Santos. Automatic summarization of technical documents in the oil and gas industry. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 431–436, 2019. doi: 10.1109/BRACIS.2019.00082.
- Y. Le Cun, B. Boser, J. S. Denker, R. E. Howard, W. Hubbard, L. D. Jackel, and D. Henderson. *Handwritten Digit Recognition with a Back-Propagation Network*, page 396–404. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990. ISBN 1558601007.
- Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc Le, and Ruslan Salakhutdinov. Transformer-XL: Attentive language models beyond a fixed-length context. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2978–2988, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1285. URL <https://www.aclweb.org/anthology/P19-1285>.
- Carnegie-Mellon University. Computer Science Dept. Speech understanding systems: summary of results of the five-year research effort at carnegie-mellon university. Carnegie Mellon University, Aug 1977. doi: 10.1184/R1/6609821.v1. URL [https://kilthub.cmu.edu/articles/journal\\_contribution/Speech\\_understanding\\_systems\\_summary\\_of\\_results\\_of\\_the\\_five-year\\_research\\_effort\\_at\\_Carnegie-Mellon\\_University\\_/6609821/1](https://kilthub.cmu.edu/articles/journal_contribution/Speech_understanding_systems_summary_of_results_of_the_five-year_research_effort_at_Carnegie-Mellon_University_/6609821/1).
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL <https://www.aclweb.org/anthology/N19-1423>.
- Jay DeYoung, Iz Beltagy, Madeleine van Zuylen, Bailey Kuehl, and Lucy Lu Wang. Ms2: Multi-document summarization of medical studies. In *arXiv*, 2021.
- Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. Unified language model pre-training for natural language understanding and generation. In *arXiv*, 2019.
- Tamas E. Doszkoics. Natural language processing in information retrieval. *J. Am. Soc. Inf. Sci.*, 37(4):191–196, July 1986. ISSN 0002-8231.

- Andre Esteva, Anuprit Kale, Romain Paulus, Kazuma Hashimoto, Wenpeng Yin, Dragomir Radev, and Richard Socher. Covid-19 information retrieval with deep-learning based semantic search, question answering, and abstractive summarization. *npj Digital Medicine*, 4, 12 2021. doi: 10.1038/s41746-021-00437-0.
- Alexander Fabbri, Irene Li, Tianwei She, Suyi Li, and Dragomir Radev. Multi-news: A large-scale multi-document summarization dataset and abstractive hierarchical model. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1074–1084, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1102. URL <https://www.aclweb.org/anthology/P19-1102>.
- Oi-Mean Foong, Suet-Peng Yong, and Farha-Am Jaid. Text summarization using latent semantic analysis model in mobile android platform. In *2015 9th Asia Modelling Symposium (AMS)*, pages 35–39, 2015. doi: 10.1109/AMS.2015.15.
- Tanvir Ahmed Fuad, Mir Tafseer Nayeem, Asif Mahmud, and Yllias Chali. Neural sentence fusion for diversity driven abstractive multi-document summarization. *Comput. Speech Lang.*, 58:216–230, 2019.
- K. Fukushima and S. Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Pattern Recognition*, volume 15, pages 455–469, 1982.
- David Furcy and Sven Koenig. Limited discrepancy beam search. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJCAI’05*, page 125–131, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc.
- Filippo Galgani, Paul Compton, and Achim Hoffmann. Combining different summarization techniques for legal text. In *Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data*, pages 115–123, Avignon, France, April 2012. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/W12-0515>.
- Jason Gauci, Edoardo Conti, Yitao Liang, Kittipat Virochsiri, Yuchen He, Zachary Kaden, Vivek Narayanan, Xiaohui Ye, Zhengxing Chen, and Scott Fujimoto. Horizon: Facebook’s open source applied reinforcement learning platform. In *arXiv*, 2019.
- Sebastian Gehrmann, Yuntian Deng, and Alexander Rush. Bottom-up abstractive summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4098–4109, Brussels, Belgium, October–November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1443. URL <https://www.aclweb.org/anthology/D18-1443>.
- Demian Gholipour Ghalandari, Chris Hokamp, Nghia The Pham, John Glover, and Georgiana Ifrim. A large-scale multi-document summarization dataset from the Wikipedia current events portal. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1302–1308, Online, July 2020. Association

- for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.120. URL <https://www.aclweb.org/anthology/2020.acl-main.120>.
- Google. Colaboratory frequently asked questions, 2021. URL <https://research.google.com/colaboratory/faq.html>.
- Daniel Graupe. *Principles of Artificial Neural Networks*. World Scientific Publishing Co., Inc., USA, 1997. ISBN 9810225164.
- Hardy Hardy and Andreas Vlachos. Guided neural language generation for abstractive summarization using Abstract Meaning Representation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 768–773, Brussels, Belgium, October–November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1086. URL <https://www.aclweb.org/anthology/D18-1086>.
- William D. Harvey and Matthew L. Ginsberg. Limited discrepancy search. In *IJCAI*, pages 607–613, 1995.
- Simon S. Haykin. *Neural networks and learning machines*. Pearson Education, Upper Saddle River, NJ, third edition, 2009.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *arXiv*, 2015.
- Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. DeBERTa: Decoding-enhanced bert with disentangled attention. In *arXiv*, 2021.
- Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In *arXiv*, 2015.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Wan-Ting Hsu, Chieh-Kai Lin, Ming-Ying Lee, Kerui Min, Jing Tang, and Min Sun. A unified model for extractive and abstractive summarization using inconsistency loss. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 132–141, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1013. URL <https://www.aclweb.org/anthology/P18-1013>.
- Hongyan Jing and Kathleen R. McKeown. The decomposition of human-written summary sentences. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '99*, page 129–136, New York, NY, USA, 1999. Association for Computing Machinery. ISBN 1581130961. doi: 10.1145/312624.312666. URL <https://doi.org/10.1145/312624.312666>.

- Karen Spärck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28:11–21, 1972.
- Leslie Pack Kaelbling, Michael L. Littman, and Andrew W. Moore. Reinforcement learning: A survey. *J. Artif. Int. Res.*, 4(1):237–285, May 1996. ISSN 1076-9757.
- Dmitry Kalashnikov, Alex Irpan, Peter Pastor, Julian Ibarz, Alexander Herzog, Eric Jang, Deirdre Quillen, Ethan Holly, Mrinal Kalakrishnan, Vincent Vanhoucke, and Sergey Levine. Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation. In *arXiv*, 2018.
- Yaser Keneshloo, Tian Shi, Naren Ramakrishnan, and Chandan K. Reddy. Deep reinforcement learning for sequence to sequence models. In *arXiv*, 2019.
- Nitish Shirish Keskar, Bryan McCann, Lav R. Varshney, Caiming Xiong, and Richard Socher. Ctrl: A conditional transformer language model for controllable generation. In *arXiv*, 2019.
- Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh. Natural language processing: State of the art, current trends and challenges. In *arXiv*, 2017.
- Rashmi Kurmi and Pranita Jain. Text summarization using enhanced mmr technique. In *2014 International Conference on Computer Communication and Informatics*, pages 1–5, 2014. doi: 10.1109/ICCCI.2014.6921769.
- Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger. From word embeddings to document distances. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML’15*, page 957–966. JMLR.org, 2015.
- Guillaume Lample and Alexis Conneau. Cross-lingual language model pretraining. In *arXiv*, 2019.
- Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. In *arXiv*, 2020.
- Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. In *arXiv*, 2014.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.703. URL <https://www.aclweb.org/anthology/2020.acl-main.703>.

- Haoran Li, Arash Einolghozati, Srinivasan Iyer, Bhargavi Paranjape, Yashar Mehdad, Sonal Gupta, and Marjan Ghazvininejad. Ease: Extractive-abstractive summarization with explanations. In *arXiv*, 2021.
- Yanran Li and Sujian Li. Query-focused multi-document summarization: Combining a topic model with graph-based semi-supervised learning. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1197–1207, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/C14-1113>.
- Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/W04-1013>.
- Junyang Lin, Xu Sun, Shuming Ma, and Qi Su. Global encoding for abstractive summarization. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 163–169, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-2027. URL <https://www.aclweb.org/anthology/P18-2027>.
- Peter J. Liu, Mohammad Saleh, Etienne Pot, Ben Goodrich, Ryan Sepassi, Lukasz Kaiser, and Noam Shazeer. Generating wikipedia by summarizing long sequences. In *arXiv*, 2018.
- Yang Liu and Mirella Lapata. Hierarchical transformers for multi-document summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5070–5081, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1500. URL <https://www.aclweb.org/anthology/P19-1500>.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. In *arXiv*, 2019.
- Bruce T. Lowerre. *The Harpy Speech Recognition System*. PhD thesis, Carnegie-Mellon University, USA, 1976. AAI7619331.
- Yao Lu, Yue Dong, and Laurent Charlin. Multi-XScience: A large-scale dataset for extreme multi-document summarization of scientific articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8068–8074, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.648. URL <https://www.aclweb.org/anthology/2020.emnlp-main.648>.
- H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2):159–165, 1958. doi: 10.1147/rd.22.0159.

- John Lyons. *Natural Language and Universal Grammar: Essays in Linguistic Theory*, volume 1. Cambridge University Press, 1991. doi: 10.1017/CBO9781139165877.
- Yuning Mao, Yanru Qu, Yiqing Xie, Xiang Ren, and Jiawei Han. Multi-document summarization with maximal marginal relevance-guided reinforcement learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1737–1751, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.136. URL <https://www.aclweb.org/anthology/2020.emnlp-main.136>.
- Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. CamemBERT: a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.645. URL <https://www.aclweb.org/anthology/2020.acl-main.645>.
- Merriam-Webster. Language. In *Merriam-Webster.com dictionary*, 2021. URL <https://www.merriam-webster.com/dictionary/language>.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *arXiv*, 2013.
- Derek Miller. Leveraging bert for extractive text summarization on lectures. In *arXiv*, 2019.
- Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Çağlar Gülçehre, and Bing Xiang. Abstractive text summarization using sequence-to-sequence RNNs and beyond. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 280–290, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/K16-1028. URL <https://www.aclweb.org/anthology/K16-1028>.
- Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don’t give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *arXiv*, 2018.
- Mir Tafseer Nayeem, Tanvir Ahmed Fuad, and Yllias Chali. Abstractive unsupervised multi-document summarization using paraphrastic sentence fusion. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1191–1204, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/C18-1102>.
- Preksha Nema, Mitesh M. Khapra, Anirban Laha, and Balaraman Ravindran. Diversity driven attention model for query-based abstractive summarization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1063–1072, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1098. URL <https://www.aclweb.org/anthology/P17-1098>.

- OpenAI, Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, Jonas Schneider, Szymon Sidor, Josh Tobin, Peter Welinder, Lilian Weng, and Wojciech Zaremba. Learning dexterous in-hand manipulation. In *arXiv*, 2019.
- Makbule Ozsoy, Ilyas Cicekli, and Ferda Alpaslan. Text summarization of Turkish texts using latent semantic analysis. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 869–876, Beijing, China, August 2010. Coling 2010 Organizing Committee. URL <https://www.aclweb.org/anthology/C10-1098>.
- A. Padmakumar and D. Eswaran. Extractive text summarization using latent semantic analysis natural language processing ( fall 2014 )-project. 2014. URL <https://www.semanticscholar.org/paper/Extractive-Text-Summarization-using-Latent-Semantic-Padmakumar-Eswaran/4ab676a3dcbde030413d22da9ef825f2c6efdd33>.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- Qingxi Peng, Qisheng Lu, and Ling Shen. Automatic summarization for agricultural product review. *International Journal of u- and e- Service, Science and Technology*, 9:95–104, 11 2016. doi: 10.14257/ijunesst.2016.9.11.09.
- Weizhen Qi, Yu Yan, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang, and Ming Zhou. ProphetNet: Predicting future n-gram for sequence-to-SequencePre-training. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2401–2410, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.217. URL <https://www.aclweb.org/anthology/2020.findings-emnlp.217>.
- A. Radford, Jeffrey Wu, R. Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. In *Technical report*. OpenAI, 2019.
- Alec Radford and Karthik Narasimhan. Improving language understanding by generative pre-training. 2018. URL <https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035>.



- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. In *arXiv*, 2020.
- Cody Rioux, Sadid Hasan, and Yllias Chali. Fear the reaper: A system for automatic multi-document summarization with reinforcement learning. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pages 681–690, 01 2014. doi: 10.3115/v1/D14-1075.
- Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, and Jason Weston. Recipes for building an open-domain chatbot. In *arXiv*, 2020.
- D. Rumelhart, Geoffrey E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.
- Alexander M. Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/D15-1044. URL <https://www.aclweb.org/anthology/D15-1044>.
- Abeed Sarker, Yuan-Chi Yang, Mohammed Ali Al-Garadi, and Aamir Abbas. A light-weight text summarization system for fast access to medical evidence. *Frontiers in Digital Health*, 2:45, 2020. ISSN 2673-253X. doi: 10.3389/fdgth.2020.585559. URL <https://www.frontiersin.org/article/10.3389/fdgth.2020.585559>.
- Alexandra Savelieva, Bryan Au-Yeung, and Vasanth Ramani. Abstractive summarization of spoken and written instructions with bert. In *arXiv*, 2020.
- John R. Searle. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3:417–424, 1980.
- Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1099. URL <https://www.aclweb.org/anthology/P17-1099>.
- J. Steinberger and Karel Jezek. Using latent semantic analysis in text summarization and summary evaluation. *Proceedings of ISIM'04*, pages 93–100, 01 2004.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In *arXiv*, 2014.
- Jiwei Tan, Xiaojun Wan, and Jianguo Xiao. Abstractive document summarization with a graph-based attentional neural model. In *Proceedings of the 55th Annual Meeting of*

- the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1171–1181, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1108. URL <https://www.aclweb.org/anthology/P17-1108>.
- Jörg Tiedemann and Santhosh Thottingal. OPUS-MT – building open translation services for the world. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480, Lisboa, Portugal, November 2020. European Association for Machine Translation. URL <https://www.aclweb.org/anthology/2020.eamt-1.61>.
- A. M. Turing. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236): 433–460, 10 1950. ISSN 0026-4423. doi: 10.1093/mind/LIX.236.433. URL <https://doi.org/10.1093/mind/LIX.236.433>.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *arXiv*, 2017.
- Sukriti Verma and Vagisha Nidhi. Extractive summarization using deep learning. In *arXiv*, 2019.
- Danqing Wang, Pengfei Liu, Yining Zheng, Xipeng Qiu, and Xuanjing Huang. Heterogeneous graph neural networks for extractive document summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6209–6219, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.553. URL <https://www.aclweb.org/anthology/2020.acl-main.553>.
- Dingding Wang, Shenghuo Zhu, Tao Li, and Yihong Gong. Multi-document summarization using sentence-based topic models. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 297–300, Suntec, Singapore, August 2009. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/P09-2075>.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Huggingface’s transformers: State-of-the-art natural language processing. In *arXiv*, 2020.
- Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. mt5: A massively multilingual pre-trained text-to-text transformer. In *arXiv*, 2021.
- Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. Xlnet: Generalized autoregressive pretraining for language understanding. In *arXiv*, 2020a.
- Ziyi Yang, Chenguang Zhu, Robert Gmyr, Michael Zeng, Xuedong Huang, and Eric Darve. TED: A pretrained unsupervised summarization model with theme modeling and denoising. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages

- 1865–1874, Online, November 2020b. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.168. URL <https://www.aclweb.org/anthology/2020.findings-emnlp.168>.
- Ziyi Yang, Chenguang Zhu, Robert Gmyr, Michael Zeng, Xuedong Huang, and Eric Darve. Ted: A pretrained unsupervised summarization model with theme modeling and denoising. In *arXiv*, 2020c.
- Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *arXiv*, 2020.
- Yong Zhang, Joo Er Meng, and Mahardhika Pratama. Extractive document summarization based on convolutional neural networks. In *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, pages 918–922, 2016. doi: 10.1109/IECON.2016.7793761.
- Yong Zhang, Erdan Chen, and Weidong Xiao. Extractive-abstractive summarization with pointer and coverage mechanism. In *Proceedings of 2018 International Conference on Big Data Technologies, ICBDT '18*, page 69–74, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450364270. doi: 10.1145/3226116.3226126. URL <https://doi.org/10.1145/3226116.3226126>.
- Chujie Zheng, Kunpeng Zhang, Harry Jiannan Wang, and Ling Fan. Topic-aware abstractive text summarization. In *arXiv*, 2020.
- Xin Zheng, Aixin Sun, Jing Li, and Karthik Muthuswamy. Subtopic-driven multi-document summarization. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3153–3162, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1311. URL <https://www.aclweb.org/anthology/D19-1311>.
- Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. Extractive summarization as text matching. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6197–6208, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.552. URL <https://www.aclweb.org/anthology/2020.acl-main.552>.
- Rong Zhou and Eric Hansen. Beam-stack search: Integrating backtracking with beam search. In *ICAPS*, pages 90–98, 01 2005.

# Appendix A

## Multi-News Sample Generated Summaries

### A.1 Multi-News Sample 1

**Source Text:** After a year in which liberals scored impressive, high-profile Supreme Court victories, conservatives could be in line for wins on some of this term's most contentious issues, as the justices consider cases that could gut public sector labor unions and roll back affirmative action at state universities. \n \n However, as the court's new term kicks off Monday, uncertainty surrounds several other politically potent cases that could wind up on the court's agenda. \n \n Story Continued Below \n \n Litigation over state efforts to limit abortion by regulating clinics and doctors is making its way to the high court. And the justices are already facing a batch of petitions involving the rights of religious institutions to opt out of providing contraception under Obamacare. \n \n Both issues seem likely to land on this term's docket, although the justices haven't formally taken up either. \n \n Many in the Obama administration would also like to see the court weigh in on immigration in coming months, upholding the president's right to grant quasi-legal status and work permits to millions of immigrants who entered the U.S. illegally as children. \n \n But it's unclear whether that fight will get to the justices in time for a decision this term or whether Obama's effort to expand his executive actions on immigration will remain blocked by a lower court order until the president leaves office. \n \n Here's POLITICO's look at five of the most important cases the justices could grapple with soon: \n \n A potential body blow to labor \n \n Public-employee unions and politicians of both parties are keenly focused on a California dispute about whether states can compel government employees to pay union dues. A loss for the unions could sharply diminish the clout of a movement already struggling with its political relevance. \n \n The case, *Friedrichs v. California Teachers Association*, was brought by Orange County, Calif. schoolteacher Rebecca Friedrichs and other teachers, who are arguing that forcing them to pay union dues violates their First Amendment rights. They also contend that unions should have to get permission before collecting dues used for political purposes, as opposed to the current system that requires objecting employees to opt out. \n \n "The significance is substantial, either way it comes out," said University of California at Irvine Law Professor Catherine Fisk. "The reason why conservative lawyers are bringing these case is the hope that a significant number of government employees choose not to join the union and certain government employee unions will be weaker." \n \n The Roberts court has not been friendly to unions, issuing a 5-4 ruling last year that prohibited mandatory union fees for home health work-

ers but stopped short of banning so-called "agency shops" in government. The new case directly asks the justices to overturn a 38-year-old precedent that allows all workers covered by union negotiations to be charged for representation. How the case will be resolved is unclear, partly because the conservative justices often see limits on government employee's First Amendment rights when their speech is at issue. Fisk said the unions are "rationally fearful" about what the court will do, but she thinks the justices might end up dumping the case after it's heard. "I think the case raises so many doctrinal problems for them," she said.

Higher ed affirmative action back in the crosshairs

Two years after punting the case back to an appeals court, the justices will take a second crack at resolving a dispute about the constitutionality of the University of Texas at Austin's affirmative action program. The case was brought by rejected applicant Abigail Fisher, who contends she was rejected because of her race. The last time Fisher's case went before the high court, affirmative action opponents hoped it would serve as a vehicle to pare back preferences for racial and ethnic groups at government-run schools. However, the justices instead told the 5th Circuit it had been too deferential to the University of Texas's claims that the programs were narrowly tailored to promote diversity. Justice Anthony Kennedy won the support of six other justices for a decision that said such programs must be handled with "strict scrutiny," but the decision did not suggest they were automatically unconstitutional. Justice Elena Kagan has recused herself because she was involved in the case during her previous service as solicitor general. The real question is whether Kennedy will join the four other Republican appointees in setting such a high bar for affirmative action that many public colleges will abandon the preferences and admissions practices they use to achieve racial and ethnic diversity.

The meaning of 'one person, one vote'

A Texas case has the potential to deal a blow to Latino political clout, tilting the balance of power away from urban areas and towards suburban and rural areas with more white voters. *Evenwel v. Abbott* presents the question of whether state legislative districts can be apportioned using a count of eligible voters rather than a count of all people. If immigrants (both illegal and legal) as well as children can be left out of the count, "the rural areas where voters tend to have fewer non-citizens or where there are fewer young people concentrated would necessarily gain," said New York University Law Professor Rick Pildes. "It's a reasonable inference if the urban areas are more Democratic leaning that they would lose power to more Republican rural areas."

Congressional redistricting shouldn't be directly affected by the case, Pildes said, because the Constitution says the U.S. census used for that reapportionment should be based on each state's population. But others say the ruling could spill over into Congressional redistricting down the line.

Religious nonprofits and Obamacare

In the Hobby Lobby case last year, the justices allowed for-profit companies to seek religious exemptions from Obamacare's coverage requirements. But now, several religious-affiliated schools and institutions—including the Little Sisters of the Poor nursing home in Denver—have filed lawsuits, too. They argue that the administration's process for allowing religious nonprofits to opt out of the contraception requirement requires them to violate their religious beliefs. The Supreme Court has eight petitions—including one that just arrived from the federal government—on this issue.

Some prominent judges are effectively begging the Supreme Court to jump in by lamenting the refusal of some courts to protect the largely-Catholic religious entities from sanctions for failing to fill out paperwork that triggers the contraception exemption but

also sets in motion coverage from others. "How ironic that this most consequential claim of religious free exercise, with literally millions of dollars in fines and immortal souls on the line, should be denied when nearly every other individual religious freedom claim has been upheld by this court," 5th Circuit Judge Edith Jones complained in a dissent last week. "How tragic to see the humiliation of sincere religious practitioners, which, coming from the federal government and its courts, implicitly denigrates the orthodoxy to which their lives bear testament. And both ironic and tragic is the harm to the Judeo-Christian heritage whose practitioners brought religious toleration to full fruition in this nation. Undermine this heritage, as our founders knew, and the props of morality and civic virtue will be destroyed."

The justices haven't signaled which of the challenges, if any, they'll consider but are expected to do so in the coming weeks. Testing when abortion clinic regulations go too far

Two of the latest tactics in the abortion wars could wind up before the justices this term: requirements that doctors performing abortions have admitting privileges at nearby hospitals and that abortion clinics meet standards for hospitals or surgical centers.

An appeals court has upheld most such limits in Texas, but in June the Supreme Court voted, 5-4, to block key parts of the law until the justices decide whether to weigh in. Petitions to take up that case and a similar law in Mississippi are already pending at the Supreme Court.

Supporters of the laws say they are designed to protect women's health, but abortion providers and abortion rights advocates say the laws would force many clinics to close and penalize poor women who could not afford travel to distant clinics.

A similar Wisconsin law led to combative oral arguments in front of the 7th Circuit last week. Judge Richard Posner suggested the law was a transparent effort to prevent abortions, not aid women.

"Governor Walker, before he withdrew from the presidential competition, said he thought abortion should be forbidden even if the mother dies as a result," Judge Richard Posner said to a lawyer for the state, in remarks first reported by the Milwaukee Journal-Sentinel. "Is that kind of official Wisconsin policy?"

"That perhaps is Governor Walker's personal view, but it's not a state policy," Assistant Attorney General Brian Keenan replied. Keenan insisted requiring doctors to have admitting privileges was a reasonable precaution to aid women.

"The admitting privileges would benefit the continuity of care for the woman when she goes to that hospital," he said.

Posner said the fact that the law was intended to kick in one business day after it was passed made clear the authors' intentions. "That statute can't be justified in terms of women's health," the judge said.

Jennifer Haberkorn contributed to this report.

————— The new term's biggest rulings will land in June, as the 2016 presidential campaign enters its final stretch, and they will help shape the political debate.

"Constitutional law and politics are certainly not the same thing, but they are interrelated, never more so than in a presidential election year that will likely determine who gets to appoint the next justice or two or three," said Vikram D. Amar, dean of the University of Illinois College of Law.

By the time the next president is inaugurated, Justice Stephen G. Breyer will be 78, Justices Scalia and Kennedy will be 80, and Justice Ruth Bader Ginsburg will be 83.

"This coming term will again put into focus that the court is divided along partisan lines and that the 2016 presidential elections will be hugely consequential in shaping constitutional and other law for perhaps a generation or more," said Neal E. Devins, a law professor at William & Mary.

The current court is the first in history split along partisan lines, where the party of the president who appointed each justice is a reli-

able predictor of judicial ideology. Put another way, all five Republican appointees are to the right of all four Democratic appointees. It was not long ago that Republican appointees like Justices John Paul Stevens and David H. Souter routinely voted with the court's liberal wing.

As a consequence of the current alignment, Professor Devins said, "the Roberts court has generated more marquee decisions divided by party alignment than all other courts combined."

The last term's big cases did not for the most part follow that pattern because Justice Kennedy, who was appointed by President Ronald Reagan and sits at the court's ideological fulcrum, voted with the court's liberal wing at an unusually high rate.

Advertisement Continue reading the main story

"The story of the last term is that the left side of the court did a lot of winning," said Irving L. Gornstein, the executive director of Georgetown's Supreme Court Institute.

"This term," he added, "I would expect a return to the norm, with the right side of the court winning a majority but by no means all of the big cases, with Justice Kennedy again the key vote."

The cases on unions and affirmative action, for instance, were almost certainly added to the docket by the more conservative justices in the confidence that they would be able to move the law to the right. Both cases were created by legal entrepreneurs and brought on behalf of plaintiffs recruited by conservative groups.

Photo

The case on unions, *Friedrichs v. California Teachers Association*, No. 14-915, may deal a blow to organized labor. "It could set the stage for a Citizens United-style reconsideration in the area of union dues," said John P. Elwood, a lawyer at Vinson & Elkins, referring to the 2010 decision that transformed campaign finance law.

The new case takes aim at a compromise fashioned by the court in 1977 in *Abood v. Detroit Board of Education*.

In *Abood*, the court said public workers who decline to join a union can nevertheless be required to pay for the union's collective bargaining efforts to prevent freeloading and ensure "labor peace." But nonmembers, the court went on, cannot be forced to pay for the union's purely political activities, as that would amount to forbidden compelled speech under the First Amendment.

Newsletter Sign Up Continue reading the main story Please verify you're not a robot by clicking the box. Invalid email address. Please re-enter. You must select a newsletter to subscribe to. Sign Up You agree to receive occasional updates and special offers for The New York Times's products and services. Thank you for subscribing. An error has occurred. Please try again later. View all New York Times newsletters.

The California teachers who brought the new case say t collective bargaining is itself political, as it concerns public policy on spending, seniority, class size and the like. Unions respond that the case is a First Amendment Trojan horse designed to further weaken the power of organized labor.

The unions have reason to be nervous. The court has twice signaled that it may be ready to overrule *Abood* notwithstanding the doctrine of *stare decisis*, Latin for "to stand by things decided." Justice Alito, the court's leading critic of *Abood*, offered a joking alternative definition in public remarks last month.

"It is a Latin phrase," he said. "It means 'to leave things decided when it suits our purposes.'"

Advertisement Continue reading the main story

The case on unions is not the only sequel on the docket.

In *Fisher v. University of Texas at Austin*, No. 14-981, the court will return to the subject of whether the Constitution permits public colleges and universities to take account of race in admissions decisions.

In 2013, in a short, vague compromise ruling in the case, the court refused to decide whether the admissions plan at the University of Texas at Austin—which combines race-neutral and race-conscious tools to achieve diversity—is

constitutional. The court's return to the subject after an appeals court sustained the hybrid plan has struck many supporters of affirmative action as an ominous sign. \n \n The case was brought by the Project on Fair Representation, a small conservative advocacy group that successfully mounted a challenge to the Voting Rights Act in 2013. The group is also behind this term's most important case on voting, *Evenwel v. Abbott*, No. 14-940, which asks the court to address the meaning of "one person, one vote." \n \n The court has never resolved whether state voting districts should have the same number of people, including unauthorized immigrants, children and others not eligible to vote, or the same number of voters. Allowing states to count only voters would in many parts of the country shift political power from cities to rural areas, a move that would generally benefit Republicans. \n \n On the last day of the term in June, Justices Breyer and Ginsburg announced that they had grave doubts about the constitutionality of the death penalty and seemed to invite a broad challenge. It has not yet arrived, and it is hardly clear that a majority would be receptive to such a challenge. \n \n The new term does have an unusually high number of capital cases presenting more focused issues, including a challenge to Florida's sentencing scheme, *Hurst v. Florida*, No. 14-7505, and a case on race discrimination in jury selection, *Foster v. Chatman*, No. 14-8349. \n \n The court has not heard an abortion case since 2007, when it upheld the federal Partial Birth Abortion Ban Act. That seems about to change. \n \n The most likely candidate is a challenge to a Texas law that threatens to reduce the number of abortion clinics in the state to about 10, down from more than 40. Should the court agree to hear the case, *Whole Woman's Health Center v. Cole*, No. 15-274, it is likely to produce the most important abortion ruling since 1992, when *Planned Parenthood v. Casey* reaffirmed the constitutional right to abortion identified in *Roe v. Wade* in 1973. \n \n Advertisement Continue reading the main story \n \n The question in the Texas case is whether two parts of a 2013 state law imposed an "undue burden" on the constitutional right to abortion. One part of the law requires all clinics in the state to meet the standards for "ambulatory surgical centers," including regulations concerning buildings, equipment and staffing. The other requires doctors performing abortions to have admitting privileges at a nearby hospital. \n \n An appeals court largely upheld the contested provisions, but the Supreme Court in June, by a 5-to-4 vote, stepped in to block the ruling while it considered whether to hear the case. \n \n That suggests three things: that the court is likely to hear the case, that its decision will be closely divided and that the ruling will land in June, thrusting a volatile and divisive issue into the middle of the presidential race. ————— WASHINGTON—The death penalty is shaping up to be a big issue for the Supreme Court as it begins a new term Monday, with at least six capital-punishment cases on the docket and a recent wave of executions keeping the justices up late to field last-minute appeals. \n \n In the weeks ahead, the court is set to hear arguments over the constitutionality of capital sentences in Florida, Georgia, Kansas and Pennsylvania. The focus on execution issues follows a 5-4 ruling last term involving a sedative used for lethal injections. The split... —————

**Gold Summary:** –The Supreme Court is facing a docket of high-profile political cases that will test whether recent liberal victories were more fluke or firm conviction, the *New York Times* reports. The court—which is divided 5-4 for conservatives, but saw Justice Roberts vote liberal on Obamacare and same-sex marriage—will look at cases including unions, affirmative action, and possibly abortion. A primer: Unions: Since 1977, unions



have been allowed to charge non-union workers for dues that go to collective bargaining efforts, but not political ones. Now California teachers have brought a case saying collective bargaining is itself political. It could set the stage for a Citizens United-style reconsideration in the area of union dues, a lawyer says. Affirmative Action: Abigail Fisher says that being white played a role in the University of Texas denying her admission back in 2008. The Supreme Court punted on her case in 2013, and now it's back on the docket. Like the unions case, this was brought by a conservative group that recruited the plaintiffs. The death penalty: Justices will decide on capital-punishment cases in Pennsylvania, Kansas, Georgia, and Florida, the Wall Street Journal reports. Justices Ruth Bader Ginsburg and Stephen Breyer have already expressed doubts about whether capital punishment is constitutional. One person, one vote: Should state legislative districts be drawn based on their number of people or eligible voters? If justices choose the latter—leaving out immigrants and children—Latinos could lose political clout and rural areas will gain, Politico reports. Abortion: Justices may opt to revisit a Texas law that could reduce the state's abortion clinics from more than 40 to roughly 10. At issue is whether new clinic requirements are an undue burden on women's right to an abortion. One commentator believes this Supreme Court session will be ugly for liberals.

**Matchsum Summary:** however , as the court ' s new term kicks off monday , uncertainty surrounds several other politically potent cases that could wind up on the court ' s agenda .\n story continued below litigation over state efforts to limit abortion by regulating clinics and doctors is making its way to the high court .\n and the justices are already facing a batch of petitions involving the rights of religious institutions to opt out of providing contraception under obamacare .\n both issues seem likely to land on this term ' s docket , although the justices haven ' t formally taken up either .\n many in the obama administration would also like to see the court weigh in on immigration in coming months , upholding the president ' s right to grant quasi-legal status and work permits to millions of immigrants who entered the u.s. illegally as\n ” constitutional law and politics are certainly not the same thing , but they are interrelated , never more so than in a presidential election year that will likely determine who gets to appoint the next justice or two or three , ” said vikram d. amar , dean of the university of illinois college of law .\n ” this coming term will again put into focus that the court is divided along partisan lines and that the 2016 presidential elections will be hugely consequential in shaping constitutional and other law for perhaps a generation or more , ” said neal e. devins , a law professor at william & amp ; mary .\n the current court is the first in history split along partisan lines , where the party of the president who appointed each justice is a reliable\n washington —the death penalty is shaping up to be a big issue for the supreme court as it begins a new term monday , with at least six capital-punishment cases on the docket and a recent wave of executions keeping the justices up late to field last-minute appeals .\n

**Pegasus (MDS) Summary:** –The Supreme Court kicks off its new term today, and there's no shortage of high-profile cases on its docket. Here's a look at some of the bigger ones: Abortion: The court could rule as soon as next month on whether states can regulate abortion clinics, the New York Times reports. A lower court had already struck down a key part of California's law, but a ruling in favor of the state could open the door to similar

laws in other states. Immigration: The court could rule on President Obama's right to grant work permits to millions of immigrants who entered the US illegally as children, Politico reports. But it's unclear whether that fight will get to the justices in time for a decision this term or whether Obama's effort to expand his executive actions on immigration will remain blocked by a lower court order until the president leaves office. Affirmative action: The court could rule on the constitutionality of Texas' affirmative action program, which uses race as a factor in college admissions. The last time the case came before the court, Anthony Kennedy voted with the other six justices to declare the program unconstitutional, but he's not expected to do the same this time around. Public sector unions: The court could

**Pegasus (SDS) Summary:** –The Supreme Court kicks off its new term Monday, with at least six capital-punishment cases on the docket and a recent wave of executions keeping the justices up late to field last-minute appeals. In the weeks ahead, the court is set to hear arguments over the constitutionality of capital sentences in Florida, Georgia, Kansas, and Pennsylvania. The focus on execution issues follows a 5-4 ruling last term involving a sedative used for lethal injections. The split: Chief Justice John Roberts wrote the majority opinion in that case. –The Supreme Court kicks off its new term today, and the cases on the docket range from affirmative action to campaign finance. Here's a look at what's on the docket, via the New York Times, the Wall Street Journal, SCOTUSblog, and SCOTUS-Blog: Friedrichs v. California Teachers Association: A group of California teachers who say they're being forced to pay dues to a union they didn't join are challenging a 1977 ruling that allows public workers who don't want to belong to a union to be required to pay dues even if they don't want to participate in the union. The plaintiffs say that violates their First Amendment right to free speech. affirmative action: A group of University of Texas students who say they're being denied admission because of their race are challenging a lower court ruling allowing the school to take into account race in its admissions process. Abortion: The court will decide whether a state can ban abortions after 20 weeks, in a case that could have major implications for states that ban abortions after that point. Gay marriage: The court will decide whether a state can ban same-sex marriage, in a case that could have major implications for states that ban same-sex marriage. –The Supreme Court kicks off its new term today, and there's no shortage of high-profile cases on its docket. Here's a look at some of the bigger ones: Abortion: The court could rule as soon as next month on whether states can regulate abortion clinics, the New York Times reports. A lower court had already struck down a key part of California's law, but a ruling in favor of the state could open the door to similar laws in other states. Immigration: The court could rule on President Obama's right to grant work permits to millions of immigrants who entered the US illegally as children, Politico reports. But it's unclear whether that fight will get to the justices in time for a decision this term or whether Obama's effort to expand his executive actions on immigration will remain blocked by a lower court order until the president leaves office. Affirmative action: The court could rule on the constitutionality of Texas' affirmative action program, which uses race as a factor in college admissions. The last time the case came before the court, Anthony Kennedy voted with the other six justices to declare the program unconstitutional, but he's not expected to do the same this time around. Public sector unions: The court could

**XLNet Summary:** WASHINGTON—The death penalty is shaping up to be a big issue for the Supreme Court as it begins a new term Monday, with at least six capital-punishment cases on the docket and a recent wave of executions keeping the justices up late to field last-minute appeals. In the weeks ahead, the court is set to hear arguments over the constitutionality of capital sentences in Florida, Georgia, Kansas and Pennsylvania. The new term’s biggest rulings will land in June, as the 2016 presidential campaign enters its final stretch, and they will help shape the political debate. By the time the next president is inaugurated, Justice Stephen G. Breyer will be 78, Justices Scalia and Kennedy will be 80, and Justice Ruth Bader Ginsburg will be 83. However, as the court’s new term kicks off Monday, uncertainty surrounds several other politically potent cases that could wind up on the court’s agenda. Story Continued Below Litigation over state efforts to limit abortion by regulating clinics and doctors is making its way to the high court.

**T5 Summary:** the death penalty is shaping up to be a big issue for the supreme court as it begins a new term . at least six capital-punishment cases are on the docket . a recent wave of executions has kept the justices up late to field last-minute appeals . the new term’s biggest rulings will land in June, as the 2016 presidential campaign enters the final stretch . constitutional law and politics are certainly not the same thing, dean says . by the time the next president is inaugurated, the next justices will be gone . conservatives could be in line for wins on some of this term’s most contentious issues . justices consider cases that could gut public sector labor unions and roll back affirmative action . uncertainty surrounds several other politically potent cases on court’s agenda .

**BART Summary:** The Supreme Court begins a new term Monday. At least six capital-punishment cases are on the docket. In the weeks ahead, the court is set to hear arguments over the constitutionality of capital sentences. The focus on execution issues follows a 5-4 ruling last term. The new term’s biggest rulings will land in June, as the 2016 presidential campaign enters its final stretch. By the time the next president is inaugurated, Justice Stephen G. Breyer will be 78, Justices Scalia and Kennedy will be 80, and Justice Ruth Bader Ginsburg will be 83. The current court is the first in history split along partisan lines. The Supreme Court’s new term kicks off Monday. Conservatives could be in line for wins on some of this term’s most contentious issues. The justices are already facing a batch of petitions involving the rights of religious institutions to opt out of providing contraception under Obamacare. Many in the Obama administration would also like to see the court weigh in on immigration in coming months.

**GPT-2 Summary:** WASHINGTON—The death penalty is shaping up to be a big issue for the Supreme Court as it begins a new term Monday, with at least six capital-punishment cases on the docket and a recent wave of executions keeping the justices up late to field last-minute appeals. In the weeks ahead, the court is set to hear arguments over the constitutionality of capital sentences in Florida, Georgia, Kansas and Pennsylvania. The new term’s biggest rulings will land in June, as the 2016 presidential campaign enters its final stretch, and they will help shape the political debate. However, as the court’s new term kicks off Monday, uncertainty surrounds several other politically potent cases that could wind up on the court’s agenda.

**ProphetNet Summary:** at least six capital - punishment cases are on the supreme court docket . the focus on execution issues follows a 5 - 4 ruling last term . by the time the next president is inaugurated , justice stephen g . breyer will be 78 . justices scalia and kennedy will be 80 , and ruth bader ginsburg will be 83 . the current court is the first in history split along partisan lines . conservatives could be in line for wins on some of this term ' s most contentious issues . cases could gut public sector labor unions and roll back affirmative action . uncertainty surrounds several other politically potent cases that could wind up on the court ' s agenda .

**LED Summary:** WASHINGTON—The death penalty is shaping up to be a big issue for the Supreme Court as it begins a new term Monday, with at least six capital-punishment cases on the docket and a recent wave of executions keeping the justices up late to field last-minute appeals. In the weeks ahead, the court is set to hear arguments over the constitutionality of capital sentences in Florida, Georgia, Kansas and Pennsylvania. The focus on execution issues follows a 5-4 ruling last term involving a sedative used for lethal injections. The split... The new term's biggest rulings will land in June, as the 2016 presidential campaign enters its final stretch, and they will help shape the political debate. After a year in which liberals scored impressive, high-profile Supreme Court victories, conservatives could be in line for wins on some of this term's most contentious issues, as the justices consider cases that could gut public sector labor unions and roll back affirmative action at state universities. Many in the Obama administration would also like to see the court weigh in on immigration in coming months, upholding the preside

**MMR-Combination Summary (Ours):** at least six capital - punishment cases are on the supreme court docket . -The Supreme Court kicks off its new term today, and there's no shortage of high-profile cases on its docket. Here's a look at some of the bigger ones: Abortion: The court could rule as soon as next month on whether states can regulate abortion clinics, the New York Times reports. A lower court had already struck down a key part of California's law, but a ruling in favor of the state could open the door to similar laws in other states. Immigration: The court could rule on President Obama's right to grant work permits to millions of immigrants who entered the US illegally as children, Politico reports. But it's unclear whether that fight will get to the justices in time for a decision this term or whether Obama's effort to expand his executive actions on immigration will remain blocked by a lower court order until the president leaves office. Affirmative action: The court could rule on the constitutionality of Texas' affirmative action program, which uses race as a factor in college admissions. The last time the case came before the court, Anthony Kennedy voted with the other six justices to declare the program unconstitutional, but he's not expected to do the same this time around. Public sector unions: The court could In the weeks ahead, the court is set to hear arguments over the constitutionality of capital sentences.

## A.2 Multi-News Sample 2

**Source Text:** A still image taken from Israeli Defence Forces (IDF) video footage shows what they say is a small unidentified aircraft shot down in a mid-air interception after it crossed into southern Israel October 6, 2012. \n \n DUBAI (Reuters) - The incursion by an unmanned aircraft into Israeli airspace at the weekend exposed the weakness of Israeli air defenses, an Iranian military official was quoted as saying on Monday. \n \n The Israeli air force shot down a drone on Saturday after it crossed into southern Israel, the military said, but it remained unclear where the aircraft had come from. \n \n Jamaluddin Aberoumand, deputy coordinator for Iran's Islamic Revolutionary Guard Corps, said the incident indicated that Israel's Iron Dome anti-missile defense system does not work and lacks the necessary capacity, Fars news agency reported. \n \n The Iron Dome system, jointly funded with the United States, is designed to shoot down short-range guerrilla rockets, not slow-flying aircraft. It intercepted more than 80 percent of the targets it engaged in March when nearly 300 rockets and mortars were fired at southern Israel, the Pentagon said at the time. \n \n The drone was first spotted above the Mediterranean near the Hamas-ruled Gaza Strip to the west of Israel, said military spokeswoman Avital Leibovich. An Israeli warplane shot it down above a forest near the occupied West Bank. \n \n Israeli parliament member Miri Regev, a former chief spokesman of the military, wrote on Twitter it was an Iranian drone launched by Hezbollah, referring to the Lebanese Shi'ite group that fought a war with Israel in 2006. \n \n Israeli defense officials have not confirmed this. \n \n Aberoumand attributed claims the drone was made by Iran to a psychological operation by Israel, but did not confirm or deny them. The Zionist regime (Israel) has many enemies, he added. \n \n On at least one occasion, Iranian-backed Hezbollah has sent a drone into Israeli airspace. And in 2010, an Israeli warplane shot down an apparently unmanned balloon in the Negev near the country's Dimona nuclear reactor. \n \n The Israeli military released a 10-second video clip of what it said was Saturday's mid-air interception, showing a small aircraft just before a missile from a fighter jet destroys it. \n \n Israel has threatened to bomb Iran's nuclear sites if diplomatic efforts fail to stop the nuclear work it believes is aimed at getting weapons capability, a charge Tehran denies. \n \n Iran has responded with threats to attack U.S. military bases in the region and retaliate against Israel if attacked. \n \n (Reporting By Yeganeh Torbati; Editing by Alistair Lyon) —————

GAZA Israel said it struck targets in the Gaza Strip on Monday after Palestinian militants fired rockets at southern Israel, in what they said was a response to an Israeli air strike that killed one militant and wounded a second a day earlier. \n \n Israel said its air raid targeted 25-year-old Mohammed Makawi whom it linked to a radical group involved in a recent Sinai border attack in which an Israeli was killed. Hospital sources in Gaza said Makawi died of his wounds. \n \n The armed wing of Hamas, the Islamists who control the Gaza Strip, said it had joined in Monday's rocket attack along with the smaller Islamic Jihad group. \n \n Gaza has been under the control of Hamas since 2007. The Islamist group rejects permanent peace with Israel and the two sides fought a three-week war in December-January 2008-2009. The border is tense, with frequent clashes. \n \n The Israeli army says over 470 rockets have been fired from Gaza this year, but it was the first time since June that Hamas had acknowledged launching rockets at Israel. A Hamas spokesman said the movement would not remain passive in the face of what it called one-sided Israeli violence. \n \n The Israeli army said it had targeted Hamas terror activity sites and terrorist squads

responsible for the rocket fire; but gave no details. Israeli forces say they will not tolerate such attacks and will hold Hamas responsible for them. \n \n Gaza hospital officials said one Islamic Jihad militant thought to have been involved in the rocket attack had been wounded by Israeli tank fire east of the town of Rafah. \n \n Residents of Khan Younis in southern Gaza said an Israeli tank fired at the suspected launch area, slightly wounding four children and damaging a minaret and a water tower. \n \n Abu Ubaida, spokesman of the Hamas armed wing Izz El-Deen Al-Qassam Brigades, said the rocket firing was a message to Israel that it would not accept a formula of a one-sided aggression by the occupation on flimsy pretexts. \n \n With Egyptian mediation, Hamas has made efforts in the past to clamp down on smaller militant groups that persist in mounting attacks on Israel, partly in order to avoid another devastating war. The Israeli offensive in the winter of 2009 killed more than 1,000 Palestinians. \n \n But on Monday Abu Ubaida said Hamas had displayed a high level of coordination with Islamic Jihad in unleashing the latest rocket barrage. \n \n Should the enemy continue its aggression against the Gaza Strip the reaction by the resistance will be stronger and broader, he said. \n \n The homemade weapons fired from Gaza are inaccurate, but potentially lethal. An Israeli spokeswoman said some exploded harmlessly on Monday near the border with the Gaza Strip. But one kibbutz resident said children could have been killed. \n \n It was a very powerful barrage, Ilan Yosef of Kibbutz Nir Yitzhak near the Gaza border told Israel radio. The goat pen in the (children's) petting corner was severely hit and a vet is treating the animals still alive and dealing with those that aren't. \n \n (Reporting by Nidal al-Mughrabi Writing by Ori Lewis; Editing by Alistair Lyon)

**Gold Summary:**— Israel launched a round of airstrikes on Gaza today, wounding two militants and eight bystanders, including some children. Israel says the attack was retaliation for a round of Palestinian rockets fired yesterday, which Palestinians say were in response to another Israeli airstrike earlier in the day, Reuters reports. More than 470 rockets have been fired from Gaza since June, but this was the first time Hamas acknowledged responsibility. As usual the rockets failed to harm anyone, though one did kill some goats at a petting zoo. The incident comes in the wake of a drone flight into Israeli airspace this weekend, which some have speculated came from Iran. An Iranian military official today boasted that the incident proved that Israel is vulnerable, and that its anti-missile defense system does not work and lacks the necessary capacity. But he denied that the drone was made by Iran, saying such accusations represented a psychological operation on Israel's part. The Zionist regime has many enemies, he said. Iran, meanwhile, is blaming Israel for a hack attack on its oil rigs.

**Matchsum Summary:** the israeli air force shot down a drone on saturday after it crossed into southern israel , the military said , but it remained unclear where the aircraft had come from . \n \n jamaluddin aberoumand , deputy coordinator for iran ' s islamic revolutionary guard corps , said the incident indicated that israel ' s iron dome anti-missile defense system “ does not work and lacks the necessary capacity “ , fars news agency reported . \n \n it intercepted more than 80 percent of the targets it engaged in march when nearly 300 rockets and mortars were fired at southern israel , the pentagon said at the time . \n \n the drone was first spotted above the mediterranean near the hamas-ruled gaza strip to the west of israel

, said military spokeswoman avital leibovich .\n an israeli warplane shot it down above a forest near the occupied west bank .\n gaza israel said it struck targets in the gaza strip on monday after palestinian militants fired rockets at southern israel , in what they said was a response to an israeli air strike that killed one militant and wounded a second a day earlier .\n israel said its air raid targeted 25-year-old mohammed makawi whom it linked to a radical group involved in a recent sinai border attack in which an israeli was killed .\n hospital sources in gaza said makawi died of his wounds .\n a hamas spokesman said the movement would not remain passive in the face of what it called “ one-sided “ israeli violence .\n

**Pegasus (MDS) Summary:**– The Israeli air force shot down an unmanned drone that infiltrated its airspace over the weekend, and an Iranian military official says the incident shows that Israel’s air defenses does not work and lacks the necessary capacity, Reuters reports. Jamaluddin Aberoumand, deputy coordinator for Iran’s Islamic Revolutionary Guard Corps, said the incident indicated that Israel’s Iron Dome anti-missile defense system does not work and lacks the necessary capacity. He also blamed the drone on a psychological operation by Israel, but didn’t confirm or deny it. The Zionist regime (Israel) has many enemies, he added. On the other hand, a Hamas spokesman said the militant group would not accept a formula of a one-sided aggression by the occupation on flimsy pretexts. Meanwhile, the Israeli army says it struck Hamas terror activity sites and terrorist squads responsible for the rocket fire in response to a rocket attack from the Gaza Strip that wounded four children and damaged a water tower, Haaretz reports.

**Pegasus (SDS) Summary:** – Israel shot down an unmanned drone over the weekend, but the incident may just be the tip of the iceberg. An Iranian military official today said Israel’s Iron Dome missile defense system does not work and lacks the necessary capacity, Reuters reports. The Israeli military released a 10-second video clip of what it said was Saturday’s mid-air interception, showing a small aircraft just before a missile from a fighter jet destroys it. The Israeli military released a 10-second video clip of what it said was Saturday’s mid-air interception, showing a small aircraft just before a missile from a fighter jet destroys it. Jamaluddin Aberoumand, the deputy coordinator for Iran’s Islamic Revolutionary Guard Corps, said the incident indicated that Israel’s Iron Dome anti-missile defense system does not work and lacks the necessary capacity, Reuters reports. The Israeli military says over 470 rockets have been fired from Gaza this year, but it was the first time since June that Hamas had acknowledged launching rockets at Israel. A Hamas spokesman said the movement would not remain passive in the face of what it called one-sided Israeli violence. Israeli forces say they will not tolerate such attacks and will hold Hamas responsible for them. Gaza hospital officials said one Islamic Jihad militant – Israel’s downing of an unmanned drone over the weekend may have been a one-off, but it’s just the latest sign that the country’s Iron Dome missile defense system does not work and lacks the necessary capacity, an Iranian military official was quoted as saying today. Jamaluddin Aberoumand, the deputy coordinator for Iran’s Islamic Revolutionary Guard Corps, said the incident indicated that Israel’s Iron Dome anti-missile defense system does not work and lacks the necessary capacity, Reuters reports. The Israeli military released a 10-second video clip of what it said was Saturday’s mid-air interception, showing a small aircraft just before a missile from a fighter jet destroys it. Aberoumand attributed claims the drone was made by

Iran to a psychological operation by Israel, but did not confirm or deny them. The Zionist regime (Israel) has many enemies, he added. On at least one occasion, Iranian-backed Hezbollah has sent a drone into Israeli airspace. And in 2010, an Israeli warplane shot down an apparently unmanned balloon in the Negev near the country's Dimona nuclear reactor. The Israeli military released a 10-second video clip of what it said was Saturday's mid-air interception, showing

**XLNet Summary:** GAZA Israel said it struck targets in the Gaza Strip on Monday after Palestinian militants fired rockets at southern Israel, in what they said was a response to an Israeli air strike that killed one militant and wounded a second a day earlier. The Israeli army says over 470 rockets have been fired from Gaza this year, but it was the first time since June that Hamas had acknowledged launching r A still image taken from Israeli Defence Forces (IDF) video footage shows what they say is a small unidentified aircraft shot down in a mid-air interception after it crossed into southern Israel October 6, 2012. Jamaluddin Aberoumand, deputy coordinator for Iran's Islamic Revolutionary Guard Corps, said the incident indicated that Israel's Iron Dome anti-missile defense system does not work and lacks the necessary capacity; Fars news agency reported.

**T5 Summary:** Israel said it struck targets in the Gaza strip on monday after militants fired rockets at southern Israel . militants said it was a response to an air strike that killed one militant and wounded a second a day earlier . it said its air raid targeted 25-year-old Mohammed Makawi whom it linked to a radical group . the incursion by an unmanned aircraft into Israeli airspace at the weekend exposed the weakness of Israeli air defenses, an Iranian military official was quoted as saying . the Israeli air force shot down a drone on Saturday after it crossed into southern Israel .

**BART Summary:** Israel says it targeted 25-year-old Mohammed Makawi. Hospital sources in Gaza said Makawi died of his wounds. The armed wing of Hamas, the Islamists who control the Gaza Strip, said it had joined in the rocket attack along with the smaller Islamic Jihad group. Incursion by unmanned aircraft into Israeli airspace at the weekend exposed the weakness of Israeli air defenses, an Iranian military official was quoted as saying. Incident indicated that Israel's Iron Dome anti-missile defense system does not work and lacks the necessary capacity; Fars news agency reported. Israeli air force shot down a drone on Saturday after it crossed into southern Israel.

**GPT-2 Summary:** GAZA Israel said it struck targets in the Gaza Strip on Monday after Palestinian militants fired rockets at southern Israel, in what they said was a response to an Israeli air strike that killed one militant and wounded a second a day earlier. The Islamist group rejects permanent peace with Israel and the two sides fought a three-week war in December-January 2008-2009. A still image taken from Israeli Defence Forces (IDF) video footage shows what they say is a small unidentified aircraft shot down in a mid-air interception after it crossed into southern Israel October 6, 2012. DUBAI (Reuters) - The incursion by an unmanned aircraft into Israeli airspace at the weekend exposed the weakness of Israeli air defenses, an Iranian military official was quoted as saying on Monday.



**ProphetNet Summary:** gaza israel says it strikes targets after palestinian militants fire rockets at southern israel . israel says its air strike killed one militant and wounded a second a day earlier . hospital sources in gaza say mohammed makawi died of his wounds . iranian military official says incursion exposed weakness of israeli air defenses . israeli air force shot down a drone on saturday after it crossed into southern israel . iranian official says incident indicated israel ' s anti - missile defense system does not work "

**LED Summary:** GAZA Israel said it struck targets in the Gaza Strip on Monday after Palestinian militants fired rockets at southern Israel, in what they said was a response to an Israeli air strike that killed one militant and wounded a second a day earlier. The Israeli army says over 470 rockets have been fired from Gaza this year, but it was the first time since June that Hamas had acknowledged launching r A still image taken from Israeli Defence Forces (IDF) video footage shows what they say is a small unidentified aircraft shot down in a mid-air interception after it crossed into southern Israel October 6, 2012. The Iron Dome system, jointly funded with the United States, is designed to shoot down short-range guerrilla rockets, not slow-flying aircraft. It intercepted more than 80 percent

**MMR-Combination Summary (Ours):** The Israeli army says over 470 rockets have been fired from Gaza this year, but it was the first time since June that Hamas had acknowledged launching rockets at Israel. –The Israeli air force shot down an unmanned drone that infiltrated its airspace over the weekend, and an Iranian military official says the incident shows that Israel's air defenses does not work and lacks the necessary capacity, Reuters reports. Jamaluddin Aberoumand, deputy coordinator for Iran's Islamic Revolutionary Guard Corps, said the incident indicated that Israel's Iron Dome anti-missile defense system does not work and lacks the necessary capacity. He also blamed the drone on a psychological operation by Israel, but didn't confirm or deny it. The Zionist regime (Israel) has many enemies, he added. On the other hand, a Hamas spokesman said the militant group would not accept a formula of a one-sided aggression by the occupation on flimsy pretexts. Meanwhile, the Israeli army says it struck Hamas terror activity sites and terrorist squads responsible for the rocket fire in response to a rocket attack from the Gaza Strip that wounded four children and damaged a water tower, Haaretz reports.

# Appendix B

## WCEP Sample Generated Summaries

### B.1 WCEP Sample

**Source Text (one document per paragraph):** KABUL, Afghanistan —Taliban militants overran a cluster of government outposts in northwestern Afghanistan on Friday, leaving more than a dozen Afghan soldiers dead, including several troops from an elite police unit, officials said.\n\n The brutal fighting came a day after the United States and the Taliban ended their sixth round of peace negotiations and the militant group said that it would continue to mount attacks, including ones like the car bomb directed at two Western aid organizations in Kabul on Wednesday that killed nine and wounded at least 20.\n\n Col. Qais Mangal, a spokesman for Afghanistan’s Defense Ministry, said the Taliban had attacked two outposts in the Murghab district of Badghis Province that are shared by the army, the police and the National Directorate of Security. He said the Afghan military had sent reinforcements as the fighting continued.\n\n ‘ ‘ Unfortunately, the Afghan forces suffered casualties, but there isn’t any information about the exact number,” Colonel Mangal said.

10Taliban militants have been killed in an airstrike in Afghanistan. The airstrike was carried out by NATO-led coalition forces’ pilot-less aircraft in a Taliban hideout.\n\n The strike, which was conducted in Barmal district, killed 10 militants and destroyed a militants’ vehicle.\n\n In a separate incident in Wardak province, a Taliban militant was killed while three were arrested. The security forces seized five motorcycles during the raid.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with “some progress” made on a draft agreement for when foreign troops might withdraw, a Taliban official said.\n\n The talks, in which the United States has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks, began on April 30 in Doha, the capital of Qatar.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country’s beleaguered security forces.\n\n Another 11 soldiers were hurt in the onslaught early Friday.\n\n Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district.\n\n Defense Ministry spokesman Col. Qais Mangal confirmed the attack but refused to discuss casualties.\n\n The Taliban have inflicted staggering casualties

on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States. The latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress,' Taliban officials said.

The incident took place on Thursday, the TOLONews broadcaster specified. The spokesman of Faryab police, Abdul Karim Yurish, told the broadcaster that militants had placed the bomb near a road in order to target Afghan security forces.  
 \n \n READ MORE: At Least 30 Civilians Killed in US Airstrikes in Afghanistan —Official  
 \n \n Afghan government forces and the Taliban movement have long been engaged in an armed confrontation. The Taliban, who have previously seized vast territories in Afghan rural areas, are carrying out attacks across the country, and the Afghan security forces conduct regular counterterrorist operations against the insurgents.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces.  
 \n \n Another 11 soldiers were hurt in the onslaught early Friday.  
 \n \n Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district.  
 \n \n Defense Ministry spokesman Col. Qais Mangal confirmed the attack but refused to discuss casualties.  
 \n \n The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States. The latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress,' Taliban officials said.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces.  
 \n \n Another 11 soldiers were hurt in the onslaught early Friday.  
 \n \n Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district.  
 \n \n Defense Ministry spokesman Col. Qais Mangal confirmed the attack but refused to discuss casualties.  
 \n \n The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States. The latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress,' Taliban officials said.

KABUL —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces.  
 \n \n Another 11 soldiers were hurt in the onslaught early Friday.  
 \n \n Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district.  
 \n \n Defence Ministry spokesman Col. Qais Mangal confirmed the attack but refused to discuss casualties.  
 \n \n The Taliban have inflicted staggering casualties on Afghanistan's security

forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States. The latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress,' Taliban officials said.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said. \n \n The talks, in which the United States has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks, began on April 30 in Doha, the capital of Qatar.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. \n \n Another 11 soldiers were hurt in the onslaught early Friday. \n \n Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district. \n \n Defense Ministry spokesman Col. Qais Mangal confirmed the attack but refused to discuss casualties. \n \n The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States. The latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress,' Taliban officials said.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said. \n \n The talks, in which the United States has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks, began on April 30 in Doha, the capital of Qatar.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. \n \n Another 11 soldiers were hurt in the onslaught early Friday. \n \n Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district. \n \n Defense Ministry spokesman Col. Qais Mangal confirmed the attack but refused to discuss casualties. \n \n The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States. The latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress,' Taliban officials said.

**Gold Summary:** One day after peace talks between the Taliban and the United States ended with no resolution, Taliban militants attack government outposts in northwestern Afghanistan, killing at least twelve Afghan soldiers.

**Pegasus Summaries (one summary per paragraph):** At least 13 Afghan soldiers

have been killed in fierce fighting with the Taliban in Badghis Province, the Afghan Defense Ministry has said, in one of the deadliest single incidents for the country's security forces since the 2001 American-led invasion.

At least 10 militants have been killed in an airstrike in Afghanistan, while three others have been arrested in a separate incident in Wardak province, officials say. In a separate incident on Sunday, a Taliban militant was killed while three were arrested.

The United States and the Taliban say they have made progress in talks aimed at ending more than 14 years of war in Afghanistan, but remain far apart on key issues such as troop withdrawal and the fate of the country's intelligence service, the National Directorate of Security.

The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States. The latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress.'"

At least six Afghan policemen have been killed by a roadside bomb in the northern Faryab province, local media report, citing local officials. The Taliban, who have previously seized vast territories in rural areas, are carrying out attacks across the country, and the Afghan security forces conduct regular counterterrorist operations against the insurgents.

The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States. The latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress.'"

The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States. The latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress.'"

The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States and the US is holding direct peace talks with the group's political office in Qatar.

The United States and the Taliban say they have made progress in talks aimed at ending more than 14 years of war in Afghanistan, but remain far apart on key issues such as troop withdrawal and the fate of the country's intelligence service, the National Directorate of Security.

The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United

States. the latest round of talks ended Thursday in Doha, where the Taliban maintain a political office, with 'some progress'

**XLNet Summaries (one summary per paragraph):** ("The Taliban") —The Taliban ("The Taliban,"") – The Taliban "The ";") ("the ", "the,"), "The,",, """,...),.. (" The ", the ", and. "., ("THE ", a """), the,, the. . ),, and, (" the "" ", in. ; ; "" the

(1) (2) (3) (4) (5) (6) (7) (9) (10) (11) (12) (13) (14) (15) (16) (17) (18) (19) (20) (21) (22) (23) (24) (25) (26) (27) (28) (30) (29) (31) (32) (33) (34) (35) (36) (37) (38) (

(Reuters, Reuters) (- - - - and - (- and (-) - and -) (-and (- — and — (- —and —) The (-, (- or (-? (-.... ( ) ( (? (-) And (, ) And the (: (! (...),

(AP) —Afghan officials say the Taliban have attacked two security outpost in western Badgis province, killed 15 security personnel and injured 11 soldiers in the assault early Friday. Defense Ministry Spokesman Col. Qai Mangal confirmed both the attack and the casualties. The insurgents have inflicted devastating casualties on the security forces over the past decade, even as they talk peace with the U.S. The Taliban are waging a relentless attack on the security outposts across the country, killing dozens of soldiers and wounding hundreds of others over the past two years. The insurgents are talking a peace with the American government but have not made any progress. The Taliban claim both the attack, and the casualties, were their own. The insurgents claim both the

of the Taliban movement, Abdul Karim Yush, told a TOLOnews broadcast that the bomb had been placed near a road to target Afghan government forces. The spokesman for Faryab Police, Abdul Karim Rish, told that the incident took place in Thursday, and the police conducted regular counterterrorist operation against the insurgents... " "The Taliban movement has previously seized vast regions in rural areas, and the government forces conduct regular anti-terrorist operations against their insurgents..., and the Taliban are conducting regular counter-terrorist operations. ;, "" (. ) ", ; ;" (,

(AP) —Afghan officials say the Taliban have attacked two security outpost in western Badgis province, killed 15 security personnel and injured 11 soldiers in the assault early Friday. Defense Ministry Spokesman Col. Qai Mangal confirmed both the attack and the casualties. The insurgents have inflicted devastating casualties on the security forces over the past decade, even as they talk peace with the U.S. The Taliban are waging a relentless attack on the security outposts across the country, killing dozens of soldiers and wounding hundreds of others over the past two years. The insurgents are talking a peace with the American government but have not made any progress. The Taliban claim both the attack, and the casualties, were their own. The insurgents claim both the

(AP) —Afghan officials say the Taliban have attacked two security outpost in western Badgis province, killed 15 security personnel and injured 11 soldiers in the assault early Friday. Defense Ministry Spokesman Col. Qai Mangal confirmed both the attack and the casualties. The insurgents have inflicted devastating casualties on the security forces over

the past decade, even as they talk peace with the U.S. The Taliban are waging a relentless attack on the security outposts across the country, killing dozens of soldiers and wounding hundreds of others over the past two years. The insurgents are talking a peace with the American government but have not made any progress. The Taliban claim both the attack, and the casualties, were their own. The insurgents claim both the

—The Taliban have launched a massive assault on two military outposts in west Badghis provincial, killing 15 soldiers, Taliban officials said Friday. The attack occurred early Friday, a day after the Taliban held a peace conference with the United Nations. The Taliban are negotiating with the United Nation over the future of Afghanistan, the United States and the Taliban, the United Nations said. The Taliban also hold a political office in Doha, Qatar, where the insurgents maintain an office, the United Nation said. The United Nations also hold a peace conference in Doha, the Taliban said. The U.S. and the Taliban also hold an office in Doha. The U,S. and The Taliban also have an office in Qatar, the U.S,

(Reuters, Reuters) (- - - - and - (- and (-) - and -) (-and (- — and — (- —and —) The (-, (- or (-? (-... ( ) ( (? (-) And (, ) And the (: (! (...),

(AP) —A Taliban official says the insurgents have launched an attack on two security posts in western Badghis province, killed 15 security personnel and injured 11 soldiers. Defense Ministry spokeswoman Col. Qai Mangal confirmed that attack. The Taliban are talking peace, even as they are talking peace. The Taliban claim to have overran the outposts in B Bala Murgab district, killed 15 soldiers and injured 11 security personnel. The Taliban deny the attack. The insurgents claim to have killed 15 security staff, injured 11 soldiers and killed 15 insurgents. The Taliban refuse the attack. TL;D: (APs) —The Taliban claim they have launched an offensive on two security post in western Badghis

**XLNet EXT Summaries (one summary per paragraph):** KABUL, Afghanistan —Taliban militants overran a cluster of government outposts in northwestern Afghanistan on Friday, leaving more than a dozen Afghan soldiers dead, including several troops from an elite police unit, officials said. He said the Afghan military had sent reinforcements as the fighting continued.

10Taliban militants have been killed in an airstrike in Afghanistan. The strike, which was conducted in Barmal district, killed 10 militants and destroyed a militants' vehicle.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Qais Mangal confirmed the attack but refused to discuss casualties.

The incident took place on Thursday, the TOLONews broadcaster specified. READ MORE: At Least 30 Civilians Killed in US Airstrikes in Afghanistan —Official  
Afghan government forces and the Taliban movement have long been engaged in an armed confrontation.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Qais Mangal confirmed the attack but refused to discuss casualties.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Qais Mangal confirmed the attack but refused to discuss casualties.

KABUL —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district.

**T5 Summaries (one summary per paragraph):** the fighting comes a day after the united states and the Taliban ended their sixth round of peace negotiations. the militant group said that it would continue to mount attacks, including ones like the car bomb directed at two western aid organizations in Kabul on Wednesday.

NATO-led coalition forces' pilot-less aircraft carried out the airstrike. the strike was conducted in barmal district, killed 10 militants. in a separate incident, a Taliban militant was killed while three were arrested.

peace talks wrap up with "some progress" made on draft agreement for when foreign troops might withdraw. talks began on April 30 in doha, the capital of Qatar. u.s. has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks.



the attack is the latest in a series of deadly attacks targeting the country's security forces. the Taliban have inflicted staggering casualties on the country's security forces in attacks. the latest round of peace talks ended in doha, where the Taliban maintain a political office.

the incident took place on Thursday, the broadcaster specified. police say militants placed the bomb near a road in order to target security forces. official Afghan government forces and the Taliban movement have long been engaged in an armed confrontation.

the attack is the latest in a series of deadly attacks targeting the country's security forces. the Taliban have inflicted staggering casualties on the country's security forces in attacks. the latest round of peace talks ended in doha, where the Taliban maintain a political office.

the attack is the latest in a series of deadly attacks targeting the country's security forces. the Taliban have inflicted staggering casualties on the country's security forces in attacks. the latest round of peace talks ended in doha, where the Taliban maintain a political office.

the attack is the latest in a series of deadly attacks targeting the country's security forces. the Taliban have inflicted staggering casualties on the country's security forces in attacks. the latest round of peace talks ended in doha, where the Taliban maintain a political office.

peace talks wrap up with "some progress" made on draft agreement for when foreign troops might withdraw. talks began on April 30 in doha, the capital of Qatar. u.s. has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks.

the attack is the latest in a series of deadly attacks targeting the country's security forces. another 11 soldiers were hurt in the onslaught early friday. the Taliban have inflicted staggering casualties on the country's security forces in relentless attacks over the past year.

**BART Summaries (one summary per paragraph):** Taliban militants overran a cluster of government outposts in northwestern Afghanistan. More than a dozen Afghan soldiers dead, including several troops from an elite police unit. The brutal fighting came a day after the United States and the Taliban ended their sixth round of peace negotiations.

10 Taliban militants have been killed in an airstrike in Afghanistan. The airstrike was carried out by NATO-led coalition forces' pilot-less aircraft. In a separate incident in Wardak province, a Taliban militant was killed while three were arrested. The security forces seized five motorcycles during the raid.

The talks began on April 30 in Doha, the capital of Qatar. The U.S. has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks. "Some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said.

Taliban attack two security outposts in Badghis province, killing 15 security personnel.

Another 11 soldiers were hurt in the onslaught early Friday. Defense Ministry spokesman Col. Qais Mangal confirmed the attack but refused to discuss casualties. Taliban have inflicted staggering casualties on Afghanistan's security forces.

The incident took place on Thursday, the TOLONews broadcaster specified. The spokesman of Faryab police, Abdul Karim Yurish, told the broadcaster that militants had placed the bomb near a road. The Taliban have previously seized vast territories in Afghan rural areas.

Taliban attack two security outposts in Badghis province, killing 15 security personnel. Another 11 soldiers were hurt in the onslaught early Friday. Defense Ministry spokesman Col. Qais Mangal confirmed the attack but refused to discuss casualties. Taliban have inflicted staggering casualties on Afghanistan's security forces.

Taliban attack two security outposts in Badghis province, killing 15 security personnel. Another 11 soldiers were hurt in the onslaught early Friday. Defense Ministry spokesman Col. Qais Mangal confirmed the attack but refused to discuss casualties. Taliban have inflicted staggering casualties on Afghanistan's security forces.

Taliban attack two outposts in Badghis province. The attacks are the latest in a series of attacks on Afghan security forces. The Taliban say the attacks are in response to U.S. drone strikes in the region. The U.S.-led coalition has been conducting drone strikes in Afghanistan since 2001.

The talks began on April 30 in Doha, the capital of Qatar. The U.S. has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks. "Some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said.

Taliban attack two security outposts in Badghis province, killing 15 security personnel. Another 11 soldiers were hurt in the onslaught early Friday. Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year. The insurgents are talking peace with the United States.

**BART EXT Summaries (one summary per paragraph):** KABUL, Afghanistan —Taliban militants overran a cluster of government outposts in northwestern Afghanistan on Friday, leaving more than a dozen Afghan soldiers dead, including several troops from an elite police unit, officials said. He said the Afghan military had sent reinforcements as the fighting continued.

10Taliban militants have been killed in an airstrike in Afghanistan. The strike, which was conducted in Barmal district, killed 10 militants and destroyed a militants' vehicle.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said. The talks, in which the United States has also sought

assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks, began on April 30 in Doha, the capital of Qatar.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday.

The incident took place on Thursday, the TOLONews broadcaster specified. The spokesman of Faryab police, Abdul Karim Yurish, told the broadcaster that militants had placed the bomb near a road in order to target Afghan security forces.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday.

KABUL —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said. The talks, in which the United States has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks, began on April 30 in Doha, the capital of Qatar.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday.

**GPT-2 Summaries (one summary per paragraph):** The Taliban are killing civilians.\n\n "The Taliban are killing our soldiers, and we are not going to let them go," he said.\n .@Mangal: "We are not going anywhere. We are not going with the Taliban." —Reuters (@Reuters) August 19, 2016\n\n The Taliban, which has been fighting the government since the fall of the Taliban in 2001, has been fighting for control of the area since the fall, and has been fighting to gain control of the city of Kunduz, which is controlled by the Taliban.\n\n In a statement, the Taliban said it had killed at least 10 civilians in the

city, including three children. A Taliban spokesman said the group had killed at most three

KABUL, Afghanistan —Taliban militants overran a cluster of government outposts in northwestern Afghanistan on Friday, leaving more than a dozen Afghan soldiers dead, including several troops from an elite police unit, officials said. Qais Mangal, a spokesman for Afghanistan's Defense Ministry, said the Taliban had attacked two outposts in the Murghab district of Badghis Province that are shared by the army, the police and the National Directorate of Security. He said the Afghan military had sent reinforcements as the fighting continued.

Taliban militants killed in airstrike in Wardak. A Taliban militant was arrested in Wardak on Thursday. The security force seized five motorcycles and a vehicle. "The Taliban has been responsible for the killing of over 100 people in the last two weeks," said the Taliban's spokesman, Mullah Omar. "The Taliban has also been responsible for killing over 100 people. The Taliban has been involved in the killing of more than 100 people in Afghanistan." The Taliban has claimed responsibility for the attack. On Thursday, the Taliban claimed responsibility for a suicide attack in the Afghan capital Kabul. Taliban spokesman Zabihullah Mujahid

10Taliban militants have been killed in an airstrike in Afghanistan. The strike, which was conducted in Barmal district, killed 10 militants and destroyed a militants' vehicle. The security forces seized five motorcycles during the raid.

U.S.-led coalition forces have been fighting the Taliban since the end of the war in 2001. "We are now ready to begin the process of withdrawing the troops," the Taliban official said, speaking on condition of anonymity. U.S. Secretary of State John Kerry said the Taliban had agreed to withdraw troops from Afghanistan by the end of March. But the Taliban said it would not accept any withdrawal from Afghanistan unless it was accompanied by a new peace agreement. A Taliban spokesman said the Taliban would not accept a withdrawal from Afghanistan if it was accompanied with a new peace deal. In a statement, the Taliban said the United States had agreed to a new peace plan for Afghanistan, which would include a

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said.

Taliban attack on security outpost in Bala. The attack comes as the Taliban continue to push for a political solution to the country's civil war. "The Taliban have been attacking the security forces in the province for a long time, and now they are attacking the security outposts," said a Taliban official. Afghanistan's security forces have been fighting the Taliban for more than a year, and the Taliban have been fighting for more than two years. In the past, the Taliban have attacked security forces in Bala, but the Tal-

iban have not attacked the security outpost in Bala since the start of the year.\n \n Last Update: Friday, 15 May 2014 KSA 15:15 - GMT 15:

The US has been bombing the Taliban in Afghanistan for years.\n \n The US has been conducting airstrikes against the Taliban in the country since September 2014, when the US-led coalition began bombing the Taliban. The US has also conducted airstrikes against the Afghan Taliban in the past year, and has conducted airstrikes against Taliban positions in the country.\n \n In the past, the US has conducted airstrikes on the Taliban in Pakistan, and has also conducted strikes against the Taliban.\n \n On Thursday, the US military said it had conducted a "targeted strike" against a Taliban position in the country, but did not specify what type of strike it was.\n \n "The US military has conducted a targeted strike against a Taliban location in the country," the

Taliban attack on security outpost in Bala\n \n The attack comes as the Taliban continue to push for a political solution to the country's civil war.\n \n "The Taliban have been attacking the security forces in the province for a long time, and now they are attacking the security outposts," said a Taliban official.\n \n Afghanistan's security forces have been fighting the Taliban for more than a year, and the Taliban have been fighting for more than two years.\n \n In the past, the Taliban have attacked security forces in Bala, but the Taliban have not attacked the security outpost in Bala since the start of the year.\n \n Last Update: Friday, 15 May 2014 KSA 15:15 - GMT 15:

Taliban attack on security outpost in Bala\n \n The attack comes as the Taliban continue to push for a political solution to the country's civil war.\n \n "The Taliban have been attacking the security forces in the province for a long time, and now they are attacking the security outposts," said a Taliban official.\n \n Afghanistan's security forces have been fighting the Taliban for more than a year, and the Taliban have been fighting for more than two years.\n \n In the past, the Taliban have attacked security forces in Bala, but the Taliban have not attacked the security outpost in Bala since the start of the year.\n \n Last Update: Friday, 15 May 2014 KSA 15:15 - GMT 15:

Taliban are killing Afghan security forces. —Reuters\n \n The attack comes as the Taliban continue to push for a political solution to the country's political crisis.\n \n In a statement, the Taliban said it had killed 15 security personnel and wounded 12 others in the latest attack.\n \n "The Taliban have killed 15 security forces and wounded 12 other security personnel in a brazen attack on the security forces of the Afghan government," the statement said.\n \n Afghanistan's security forces have been under heavy fire since the Taliban took over the country in 2014.\n \n Last month, the Taliban claimed responsibility for a series of attacks on security forces in the country, including the killing of a security guard in the capital Kabul.\n \n A Taliban spokesman said

U.S.-led coalition forces have been fighting the Taliban since the end of the war in 2001.\n \n "We are now ready to begin the process of withdrawing the troops," the Taliban official said, speaking on condition of anonymity.\n \n U.S. Secretary of State John Kerry said the Taliban had agreed to withdraw troops from Afghanistan by the end of March.\n \n

\n But the Taliban said it would not accept any withdrawal from Afghanistan unless it was accompanied by a new peace agreement.\n \n A Taliban spokesman said the Taliban would not accept a withdrawal from Afghanistan if it was accompanied with a new peace deal.\n \n In a statement, the Taliban said the United States had agreed to a new peace plan for Afghanistan, which would include a

Taliban attack on security outpost in Bala Makhmour\n \n The attack comes as the Taliban continue to push for a political solution to the country's civil war.\n \n "The Taliban have been attacking security outposts for years, and they are now attacking the security outposts of the country," said a Taliban official.\n \n Afghanistan's security forces have been fighting the Taliban for more than a year, and the Taliban have been fighting for more than two years.\n \n In the past, the Taliban have attacked security outposts, but the Taliban have not attacked them.\n \n Last Update: Friday, 11 May 2014 KSA 11:30 - GMT 11:30

**GPT-2 EXT Summaries (one summary per paragraph):** KABUL, Afghanistan —Taliban militants overran a cluster of government outposts in northwestern Afghanistan on Friday, leaving more than a dozen Afghan soldiers dead, including several troops from an elite police unit, officials said. Qais Mangal, a spokesman for Afghanistan's Defense Ministry, said the Taliban had attacked two outposts in the Murghab district of Badghis Province that are shared by the army, the police and the National Directorate of Security. He said the Afghan military had sent reinforcements as the fighting continued.

10Taliban militants have been killed in an airstrike in Afghanistan. The strike, which was conducted in Barmal district, killed 10 militants and destroyed a militants' vehicle. The security forces seized five motorcycles during the raid.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday. Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district.

The incident took place on Thursday, the TOLONews broadcaster specified. The spokesman of Faryab police, Abdul Karim Yurish, told the broadcaster that militants had placed the bomb near a road in order to target Afghan security forces.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday. Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday. Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district.

KABUL —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday. The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. Another 11 soldiers were hurt in the onslaught early Friday. The Taliban have inflicted staggering casualties on Afghanistan's security forces in relentless attacks over the past year, even as the insurgents are talking peace with the United States.

**ProphetNet Summaries (one summary per paragraph):** to summarise the brutal fighting began the day before, officials said. the brutal fighting continued as the fighting continued. the fighting continued on friday, officials said, adding that the fighting continued and that the fighting would continue. the fighting, which began on friday, the day after the fighting began. the fighting continues.

to summarise the strike took place in afghanistan. the strike, which was conducted by the us - led coalition forces, targeted the militants ' hideout. the strike was conducted in response to a tip - off from the taliban. the strike also targeted a taliban hideout, which was located in a taliban hideout. in a separate incident, a taliban militant was killed.

to summarise, a taliban official said. the talks, which began on april 30 in doha, qatar, which began april 30, began on april 29, 2015, in which the u. s., and the taliban, and the united states, and the u. k., and a taliban official in doha, where the talks are being held, began on apr. 30, 2015.

to summarise —the taliban have launched an assault on two security outposts in western badghis province, afghanistan's second - most - populous province. the assault began early friday morning, when the taliban launched an attack on two outposts in the province. the attack came early friday morning. the attack began early friday.

to summarise the incident took place on thursday, the broadcaster specified. according to the broadcaster, militants had placed the bomb near a road in order to target afghan security forces. the broadcaster specified that militants had placed a bomb near the road on thursday. according to, the police specified. the bomb was placed near a road near a road, the bbc specified. the taliban, who have previously engaged in armed confrontations with the afghan security forces, and the taliban, and the afghan government and the taliban.

to summarise —the taliban have launched an assault on two security outposts in western badghis province, afghanistan's second - most - populous province. the assault began early friday morning, when the taliban launched an attack on two outposts in the province. the attack came early friday morning. the attack began early friday.

to summarise —the taliban have launched an assault on two security outposts in western badghis province, afghanistan's second - most - populous province. the assault began early friday morning, when the taliban launched an attack on two outposts in the province. the attack came early friday morning. the attack began early friday.

to summarise —the taliban have launched an assault on two security outposts in western badghis province, afghanistan ' s second - most populous, and the latest in a string of deadly attacks in which the taliban have claimed at least 15 lives, officials said.

to summarise, a taliban official said. the talks, which began on april 30 in doha, qatar, which began april 30, began on april 29, 2015, in which the u. s., and the taliban, and the united states, and the u. k., and a taliban official in doha, where the talks are being held, began on apr. 30, 2015.

to summarise —the taliban have launched an assault on two security outposts in western badghis province, afghanistan ' s second - most - populous province. the onslaught began early friday morning, when the taliban launched an attack on two outposts in the province. the assault began early friday, with the attack in bala murghab district.

**LED Summaries (one summary per paragraph):** KABUL, Afghanistan —Taliban militants overran a cluster of government outposts in northwestern Afghanistan on Friday, leaving more than a dozen Afghan soldiers dead, including several troops from an elite police unit, officials said. The Taliban also killed at least two police officers and a soldier from an elite security unit. The brutal fighting came a day after the United States and the Taliban ended their sixth round of peace negotiations and the militant group said that it would continue to mount attacks, including ones like the car bomb directed at two Western aid organizations in Kabul on Wednesday that killed nine and wounded at least 20. The Taliban said it was planning to attack the outposts again. Col. Qais Mangal, a spokesman for Afghanistan's Defense Ministry

10Taliban militants have been killed in an airstrike in Afghanistan. The airstrike was carried out by NATO-led coalition forces' pilot-less aircraft in a Taliban hideout. The mil-



itants were killed in the attack. The strike, which was conducted in Barmal district, killed 10 militants and destroyed a militants' vehicle. The militants had fled the area. In a separate incident in Wardak province, a Taliban militant was killed while three were arrested. The security forces seized five motorcycles during the raid.

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said. The Taliban said it had reached a deal with the United States on the withdrawal of foreign troops from Afghanistan. The talks, in which the United States has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks, began on April 30 in Doha, the capital of Qatar.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. The Taliban said in a statement that the attack was carried out by the Afghan security forces. Another 11 soldiers were hurt in the onslaught early Friday. The Taliban said the attack was part of a wider offensive. Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district. He said the Taliban had also attacked the two outpost in the same district. Defense Ministry spokesman Col. Q

The incident took place on Thursday, the TOLONews broadcaster specified. The spokesman of Faryab police, Abdul Karim Yurish, told the broadcaster that militants had placed the bomb near a road in order to target Afghan security forces. **READ MORE: At Least 30 Civilians Killed in US Airstrikes in Afghanistan —Official —USA** Afghan government forces and the Taliban movement have long been engaged in an armed confrontation. The Taliban, who have previously seized vast territories in Afghan rural areas, are carrying out attacks across the country, and the Afghan security forces conduct regular counterterrorist operations against the insurgents.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. The Taliban said in a statement that the attack was carried out by the Afghan security forces. Another 11 soldiers were hurt in the onslaught early Friday. The Taliban said the attack was part of a wider offensive. Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district. He said the Taliban had also attacked the two outpost in the same district. Defense Ministry spokesman Col. Q

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. The Taliban said in a statement that the attack was carried out by the Afghan security forces. Another 11 soldiers were hurt in the onslaught early Friday. The Taliban said the attack was part of a wider offensive. Ziauddin Akazai, a lawmaker in the province, said

the Taliban overran the two outposts in Bala Murghab district. He said the Taliban had also attacked the two outpost in the same district. Defense Ministry spokesman Col. Q

KABUL —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. The Taliban said in a statement that the attack was carried out by the Afghan security forces. Another 11 soldiers were hurt in the onslaught early Friday. The Taliban said the attack was part of a wider offensive. Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district. He said the Taliban had also attacked the two outpost in the province's eastern Ghazni district. Defence

KABUL (Reuters) - U.S. and Taliban negotiators wrapped up their sixth round of peace talks on Thursday with "some progress" made on a draft agreement for when foreign troops might withdraw, a Taliban official said. The Taliban said it had reached a deal with the United States on the withdrawal of foreign troops from Afghanistan. The talks, in which the United States has also sought assurances that the Taliban will not allow militant groups to use Afghanistan to stage attacks, began on April 30 in Doha, the capital of Qatar.

KABUL, Afghanistan (AP) —An Afghan official says the Taliban have launched an assault on two security outposts in western Badghis province, killing 15 security personnel in the latest in a series of deadly attacks targeting the country's beleaguered security forces. The Taliban said in a statement that the attack was carried out by the Afghan security forces. Another 11 soldiers were hurt in the onslaught early Friday. The Taliban said the attack was part of a wider offensive. Ziauddin Akazai, a lawmaker in the province, said the Taliban overran the two outposts in Bala Murghab district. He said the Taliban had also attacked the two outpost in the same district. Defense Ministry spokesman Col

**MMR-Combination Summary (Ours):** Taliban attack two outposts in Badghis province. The attacks are the latest in a series of attacks on Afghan security forces. The Taliban say the attacks are in response to U. S. drone strikes in the region. The U. S.-led coalition has been conducting drone strikes in Afghanistan since 2001. Taliban attack two security outposts in Badghis province, killing 15 security personnel.