

Abstractive Text Summarizer



Namratha S Kumar(016691883)

Nandini S Nair(016781830)

Pooja Manjunatha(016025659)

Priya Varahana(016780790)

Contents

❖ **Introduction**
Motivation, Originality

❖ **Related Work**
**Literature Survey of
existing solutions**

❖ **Application**
Real world uses

❖ **Data Engineering**
**Data Pre-processing,
Visualization & EDA,
Feature Engineering**

❖ **Modelling**
**ML/DL model
architecture, results &
comparison**

❖ **Conclusion**
**Summary and
future scope**



Introduction

- ❖ Data generating sources like online news, legal documents, reviews of products generates billions of documents day by day.
- ❖ Necessity to generate a significant and relevant information from the whole document.
- ❖ Four deep learning models -LSTM, biLSTM, BERT, Pointer Generator Network are implemented and evaluated to come up with an efficient abstractive text summarizer.

Motivation

- ❖ It is extremely valuable for data scientists who must gather, analyze, and interpret large amounts of data as it expedites and simplifies the process.
- ❖ The deep neural networks are designed to replicate the human brain's functions by enabling it to adapt from massive volumes of data.
- ❖ While a single-layer neural network may produce approximate predictions, additional hidden layers can help to optimize and improve for accuracy

Originality



Model Architecture

Models used in this project have original architecture. Dense, Sequential & Improved version are added to create models.



Model Comparison

We have used 4 models being compared on the basis of results obtained ,rouge and bleu scores.

Related Work



Author	Model	Description	Year	Ref.
Song et al.	LSTM	STM-CNN based ATSDL. model achieved a ROGUE-1 score of 22.9% and a ROUGE-2 score of 17.8%. Challenging to establish semantic similarity between phrases, consumes time to train the DL models	2018	[1]
Jiang, J., et al.	BiLSTM	Bi-LSTM + Attention,Seq2Seq models,DA-PN, DA-PN + Cover, DA-PN + Cover + MLO. Limitations are insufficient attention at the decoder, Limited network layers, handling of unregistered words. Evaluation results are-ROUGE 1 - 30.27 %, ROUGE 2- 8.03%, ROUGE L - 29.04%	2021	[2]
Wijayanti et al.	BERT	When compared to pre-trained english model , greater ROGUE-1 score of 44%. synonym words are considered as different, hence low rouge scores	2021	[3]
Boutkan et al.	Pointer Generation Network	Baseline Pointer-generator Network, Baseline Network + Multi-head Attention, Pointing Penalty Losses. Challenge of accurately measuring and validating abstractiveness is unresolved .Evaluation results are:ROUGE 1 - 39.53 %, ROUGE 2 - 17.28%, ROUGE L - 36.38%	2019	[4]



Applications

News Summaries

**Scientific Paper
Summaries**

**Legal Document
Summaries**

Customer Support

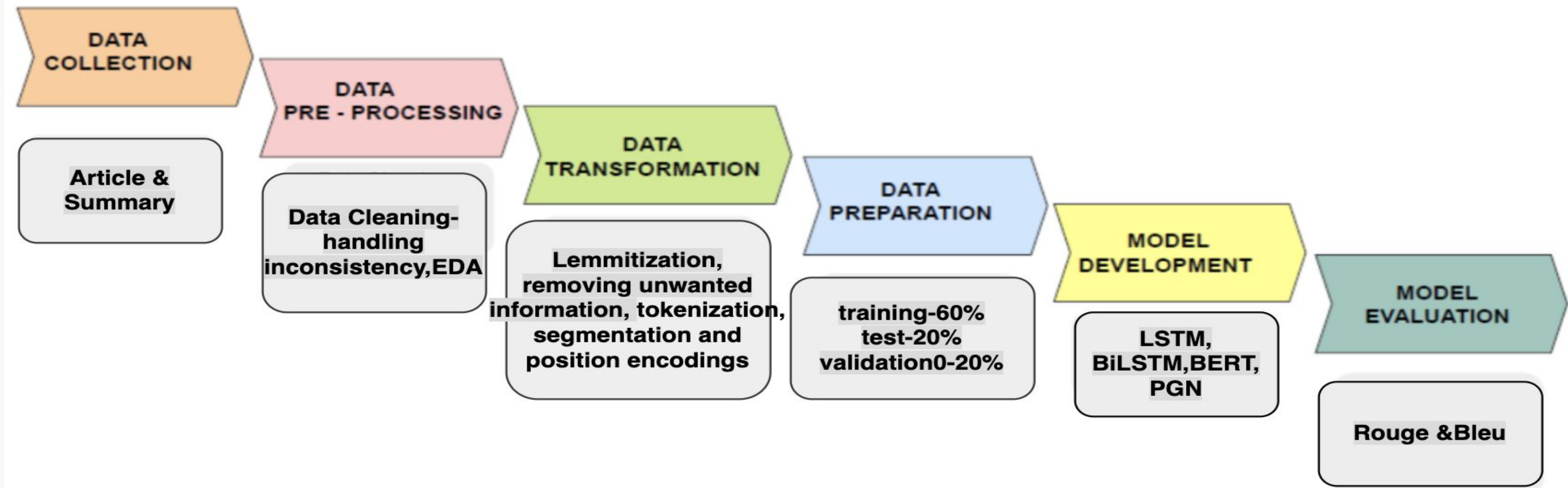
Product Reviews

Healthcare



Architecture

High Level System Design Overview



Data

Data Collection

The data is taken from kaggle - CNN Daily dataset, New York Times dataset, Wikihow dataset

Dataset Statistics & Attributes

CNN /Daily dataset(528 MB), New York Times dataset (3 GB), Wikihow dataset(590MB) - with 3 main attributes id, article, highlight



	A	B	C
1	id	article	highlights
2	0001d1afc246a7964130f43ae940af6bc6c57f01	By . Associated Press . PUBLISHED: . 14:11 E	Bishop John Folda, of North Dakota, is taking time off after being diagnosed .
3	0002095e55fcbd3a2f366d9bf92a95433dc305ef	(CNN) -- Ralph Mata was an internal affairs	Criminal complaint: Cop used his role to help cocaine traffickers .
4	00027e965c8264c35cc1bc55556db388da82b07f	A drunk driver who killed a young woman in a	Craig Eccleston-Todd, 27, had drunk at least three pints before driving car .
5	0002c17436637c4fe1837c935c04de47adb18e9a	(CNN) -- With a breezy sweep of his pen Pre	Nina dos Santos says Europe must be ready to accept sanctions will hurt
6	0003ad6ef0c37534f80b55b4235108024b407f0b	Fleetwood are the only team still to have a 1	Fleetwood top of League One after 2-0 win at Scunthorpe .
7	0004306354494f090ee2d7bc5ddbfb80b63e80de6	He's been accused of making many a fashion	Prime Minister and his family are enjoying an Easter break in Lanzarote .
8	0005d61497d21ff37a17751829bd7e3b6e4a7c5c	By . Daily Mail Reporter . PUBLISHED: . 01:1	NBA star calls for black and Hispanic communities to get tested .
9	0006021f772fad0aa78a977ce4a31b3faa6e6fe5	By . Daily Mail Reporter . This is the moment	London Midland service had been pulling into Telford Station in Shropshire .
10	00083697263e215e5e7eda753070f08aa374dd45	There are a number of job descriptions waiti	Tony Pulis believes Saido Berahino should look up to Darren Fletcher .
11	000940f2bb357ac04a236a232156d8b9b18d1667	Canberra, Australia (CNN) -- At first glance,	Black box data from Flight 370 could be analyzed at a laboratory in Australia .
12	0009ebb1967511741629926ef9f5faea2bb6be24	By . Ellie Zolfagharifard . Take a look at a ma	The distortion is the result of the Mercator map which was created in 1596 to
13	000c835555db62e319854d9f8912061cdca1893e	Two lawyers representing a woman who . cl	Alan Dershowitz has filed defamation suits against two other U.S lawyers .
14	000ca3fc9d877f8d4bb2ebd1d6858c69be571fd8	It's the moment every pet owner dreads - w	Sarah Ernhart, the owner of Sarah Beth Photography in Minneapolis, created
15	000cd1ee0098c4d510a03ddc97d11764448ebac2	Louis van Gaal said he had no option but to s	Manchester United beat Southampton 2-1 at St Mary's on Monday night .
16	000e009f6b1d954d827c9a550f3f24a5474ee82b	(CNN) -- One can hardly read the news thes	U.S. corporations merge with foreign companies, move their headquarters .
17	001097a19e2c96de11276b3cce11566ccfed0030	For most people, it has become a travel ess	Half of Brits admit to checking worke-mails while on holiday, while a third
18	0010c870d3fc53ea7f2a4a50f6496dc2df17e02f	By . Margot Peppers . Nigerian and Cameroc	Dencia's comment is hypocritical considering she recently courted controver
19	00120f91cfcab17bac165f7a4719019a628a9db3	Britain and the West must brace themselve	Ex-military chiefs suggested .
20	00128f1ba30d5e9e0f17df83285a1bc2072e2f01	A woman has been charged with reckless m	Claudia Yanira Hernandez Soriano, 25, and Juan Francisco Martinez Rojas,
21	00142065a6c92f788eacd40c9023184808a7e2d1	Beirut (CNN) -- Syria carried out an airstrike	Airstrike kills nine Syrians in refugee camp, state media reports .
22	0015194573f9b4430319683cde41e4aa17092a9d	An Australian citizen, who has awaited trial	Roman Shilov has been detained by Russian authorities without trial since
23	00156d9892fb27f1d2e100cbdd8a3997f8273781	Everton are still looking to add two new play	Everton are looking to add two new players to their squad .
24	001732b374f362d3961a510da315601e4b5e7e84	By . Daily Mail Reporter . Last updated at 1:2	Police brought in over Twitter threats believed to be made by man in Russia .
25	001789cf9b865dcac3d9fc032a6b1533e3318eda	Last updated at 3:31 PM on 19th July 2011 .	(Amish population in New York has grown by a third in the past two years .
26	00189f37b1c8bdc2b132b40270bb28ffcc622af1	Glen Johnson looks destined to leave Anfiel	Full back scored late winner as Liverpool beat Stoke 1-0 on Saturday .
27	001a6162391594e2a8607fba135bdfa154e57904	Would you believe that these breathtaking	Adventurer Tomasz Furmanek photographs Norwegian fjords from a kayak .
28	001adf6209be103cb198b8599f236b4d5760a5fe	By . Daily Mail Reporter . PUBLISHED: . 19:4	Ashleigh Davis, 32, arrested last year after a fight with another woman .

ibsection	material	headline	abstract	keywords
ics	News	12632 unique values	12604 unique values	4%
er (3283)	Op-Ed			1%
	Other (1260)			95%
	Editorial	Protect Veterans From Fraud	Congress could do much more to protect Americans who have served their country from predatory for-pr...	['Veterans', 'For-Profit Schools', 'Financial Aid (Education)', 'Frauds and Swindling', 'Colleges an...
	News	'It's Green and Slimy'	Christina Iverson and Jeff Chen ring in the New Year.	['Crossword Puzzles']
	News	Meteor Showers in 2020 That Will Light Up Night Skies	All year long, Earth passes through streams of cosmic debris. Here's a list of major meteor showers ...	['Meteors and Meteorites', 'Space and Astronomy', 'Earth', 'Solar System']
	Interactive Feature	Sync your calendar with the solar system	Never miss an eclipse, a meteor shower, a rocket launch or any other astronomical and space event th...	['Space and Astronomy', 'Moon', 'Eclipses', 'Seasons and Months', 'Solar System', 'Meteors and Meteo...

headline	title	text
214097 unique values	215365 unique values	209179 unique values
Keep related supplies in the same area., Make an effort to clean a dedicated workspace after every ...	How to Be an Organized Artist1	If you're a photographer, keep all the necessary lens, cords, and batteries in the same quadrant of...
Create a sketch in the NeoPopRealist manner of the future mural on a small piece of paper 8"x10" us...	How to Create a Neopoprealist Art Work	See the image for how this drawing develops step-by-step. However, there is an important detail: th...
Get a bachelor's degree., Enroll in a studio-based program., Train on a number of VFX computer prog...	How to Be a Visual Effects Artist1	It is possible to become a VFX artist without a college degree, but the path is often easier with o...

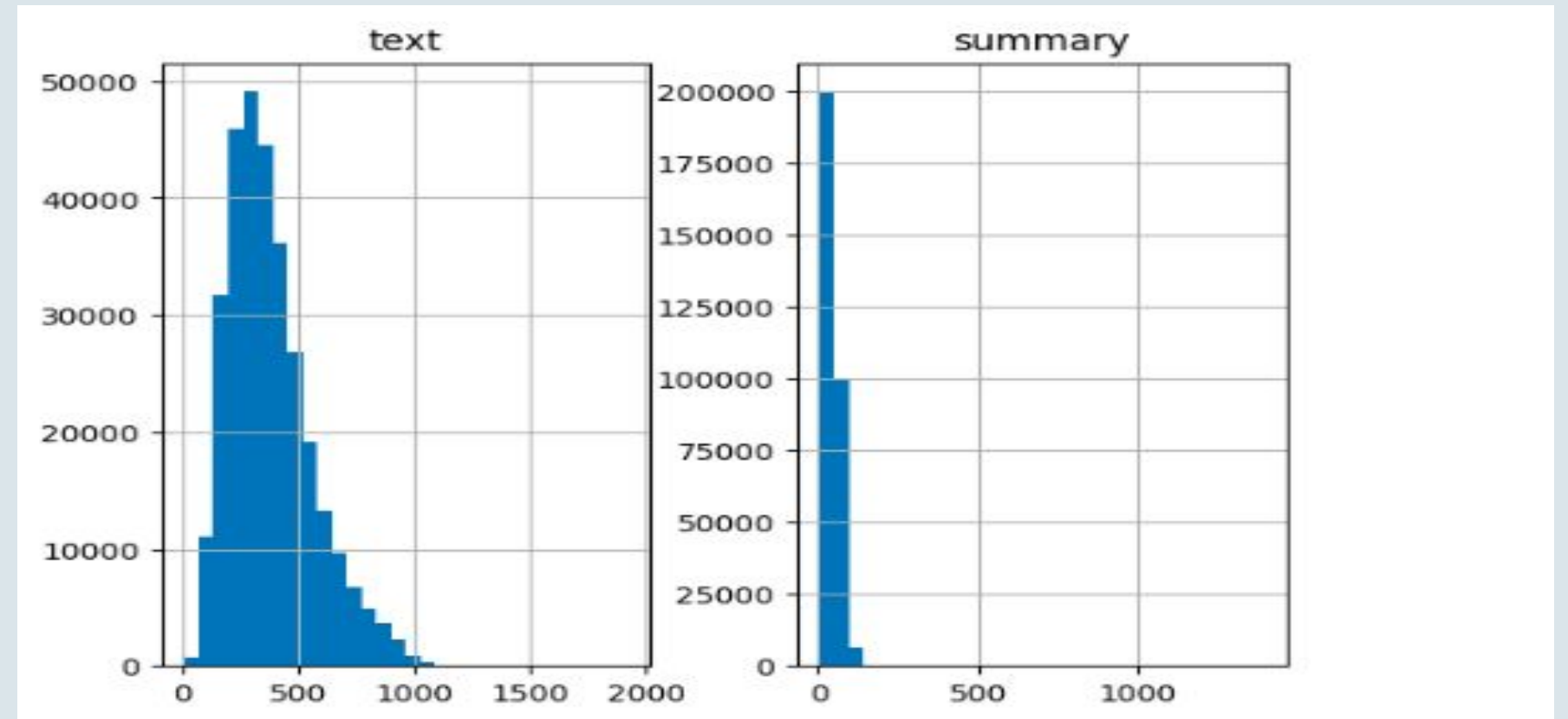
EDA

- The dataset contains 3211971 rows and 3 columns.
- There are three columns namely, id, articles and highlights and it contains no missing values

Missing Values:

```
id
article    0
highlights 0
dtype: int64
```

Combined data shape: (311971, 3)



- Most of the words in the article column fall in the range of 400-500 words.
- There are words in the highlights column that fall in the range of 70-80 words.

Data Preprocessing

BEFORE PRE-PROCESSING:

By . Associated Press . PUBLISHED: . 14:11 EST, 25 October 2013 . | . UPDATED: . 15:36 EST, 25 October 2013 . The bishop of the Fargo Catholic Diocese in North Dakota has exposed potentially hundreds of church members in Fargo, Grand Forks and Jamestown to the hepatitis A virus in late September and early October. The state Health Department has issued an advisory of exposure for anyone who attended five churches and took communion. Bishop John Folda (pictured) of the Fargo Catholic Diocese in North Dakota has exposed potentially hundreds of church members in Fargo, Grand Forks and Jamestown to the hepatitis A . State Immunization Program Manager Molly Howell says the risk is low, but officials feel it's important to alert people to the possible exposure. The diocese announced on Monday that Bishop John Folda is taking time off after being diagnosed with hepatitis A. The diocese says he contracted the infection through contaminated food while attending a conference for newly ordained bishops in Italy last month. Symptoms of hepatitis A include fever, tiredness, loss of appetite, nausea and abdominal discomfort. Fargo Catholic Diocese in North Dakota (pictured) is where the bishop is located .

AFTER PRE-PROCESSING:

associated press published est october updated est october bishop fargo catholic diocese north dakota exposed potentially hundreds church members fargo grand forks jamestown hepatitis virus late september early october state health department issued advisory exposure anyone attended five churches took communion bishop john folda fargo catholic diocese north dakota exposed potentially hundreds church members fargo grand forks jamestown hepatitis state immunization program manager molly howell says risk low officials feel important alert people possible exposure diocese announced monday bishop john folda taking time diagnosed hepatitis diocese says contracted infection contaminated food attending conference newly ordained bishops italy last month symptoms hepatitis include fever tiredness loss appetite nausea abdominal discomfort fargo catholic diocese north dakota bishop located'

- Expanding contractions
- Lemmatization
- Lower casing
- Removing punctuations
- Removal of stopwords

Transformation

array(['england become overcrowded major country europe population growth rapid four times many people soon
crammed france twice many germany england overtaken netherlands become second tiny malta densely populated
nation eu squeeze england overtaken netherlands become densely populated major nation eu next years gap widen
germany france holland either decline grow slowly house commons figures based data uk eu statistical agencies
show huge impact labour open door immigration policy estimated people squeezed every square kilometre england
compared tony blair took power lax controls figures based data uk eu statistical agencies show huge impact labour
open door immigration policy revelations fuel debate immigration especially uk opening borders romanian
bulgarian workers new year day james clappison tory mp obtained figures said last labour government england
green pleasant land became england green crowded land number people expected living per square kilometre
country reasons never properly explained labour instigated policy massive expansion immigration fear must future
labour government would sir andrew green chairman migrationwatch think tank said per cent immigrants uk headed
england rapidly growing population density inevitable consequence labour mass immigration nearly four million
years added already see pressure maternity units primary schools less visible pressure housing already crisis need
build houses day next years simply new immigrants families house commons report says number people living
every square kilometre england rise astonishing people living square kilometre equivalent figure france germany
netherlands england also three times packed poland estimated one million arrivals labour originated research
raises concerns uk infrastructure cope increased pressure schools hospitals roads clash liberal democrat business
secretary vince cable left likened david cameron policies enoch powell notorious rivers blood speech large
numbers packed country also affect water power supplies increase pressure build green spaces david cameron
pressure confront electoral threat posed ukip changed law prevent eu migrants claiming benefits first three months
following arrival wake open door immigration policies deliberately pursued new labour england crowded country
europe twice many people cramped every square kilometre germany four times france three decades unless
significant tightening border controls almost people living square kilometre compared labour came power vince
cable colleagues left disgracefully suggest somehow racist worry immigration revelations show nothing race
everything schools social services nhs housing possibly expected cope unprecedented pressure officials say want
reduce pull factor uk last weekend tensions two coalition government parties boiled liberal democrat business
secretary vince cable likened tory policies enoch powell notorious rivers blood speech office national statistics
already warned britain must make room almost million people next years equivalent building city even larger london
increase mainly result immigration high migrant birthrates push numbers million home office spokesman said
immigration brought benefits uk welcome people want come contribute economy society however important control
immigration effect social cohesion public services jobs wages figures rest uk predicted scotland per square
kilometre wales northern ireland malta figure expected',

Converted to:

```
array([[ 136,  252, 14153, ...,   0,   0,   0],  
       [10258, 29405,  109, ...,   0,   0,   0],  
       [  23,  125, 1992, ...,   0,   0,   0],  
       ...,  
       [  849,  9688,  191, ...,   0,   0,   0],  
       [23115, 45137,  109, ...,   0,   0,   0],  
       [ 5560, 44202,   531, ...,   0,   0,   0]], dtype=int32)
```

- Tokenization
- Segment embeddings
- Positional encodings



Models and Performance



Base Model

Pretrained T5 (Text-To-Text Transfer Transformer)

Handle a variety of NLP tasks by converting them into a text-to-text format.

Architecture:

- SimpleT5 is a library that is built on top of the PyTorch-lightning and Transformers libraries.
- Fine-tuning involves feeding the input text to the encoder and training the decoder to generate the summary.

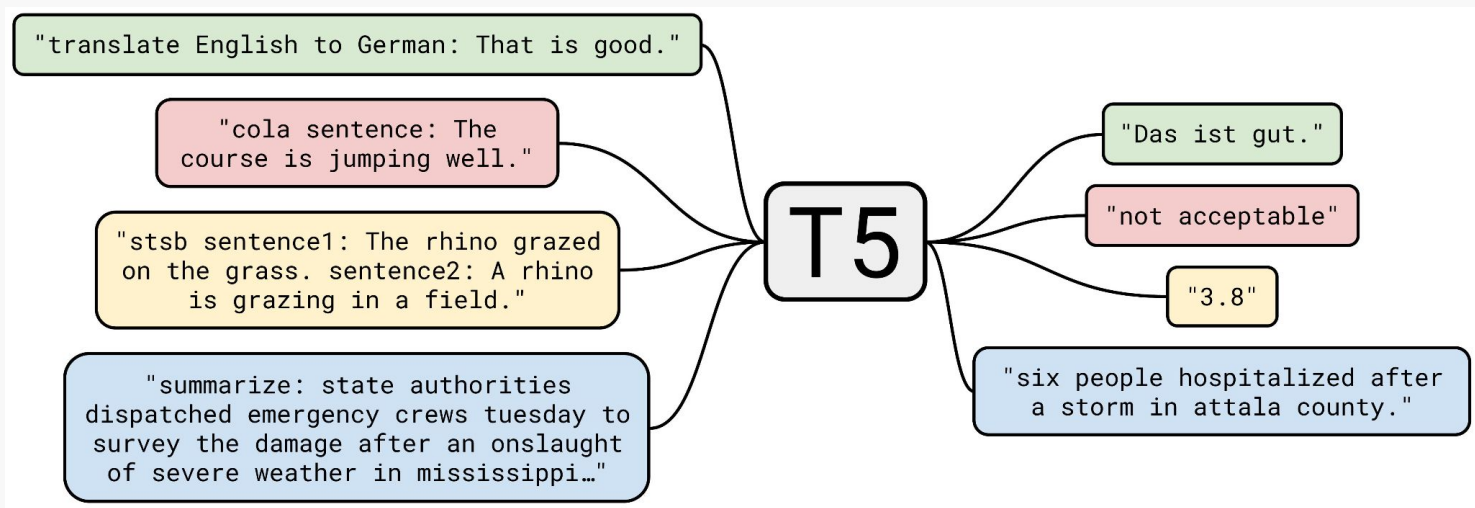


Table of hyperparameters

Hyperparameter	Value
Training Size	5000 samples
Max Sequence Length	
Input (Article)	512 tokens
Output (Summary)	150 tokens
Max Epochs	10 epochs
activation function	ReLU
optimizer	Adam
Batch Size	16
Beam Search	
Number of Beams	4
Training-Evaluation Split	60% training, 20% test 20%validation
Device	GPU

```
INFO:pytorch_lightning.callbacks.model_summary:
| Name | Type | Params
-----
0 | model | T5ForConditionalGeneration | 222 M
-----
222 M    Trainable params
0        Non-trainable params
222 M    Total params
891.614  Total estimated model params size (MB)
```



Performance Result of Base Model

Original Text :
By . Emily Crane . and Aap . A man has handed himself in almost 12 hours after allegedly running from the scene of a fatal car accident during a police chase northwest of Melbourne.....at nearby Gisborne. Police were still looking for the second man when he showed up at Gisborne Police Station .

Summary Text :
Police were pursuing a sedan with four people following reports of alleged robbery at Macedon railway station, northwest of Melbourne .
The car was found a short time later on its side early on Saturday morning .
Police found one man dead after being ejected from the vehicle .
A 27-year-old woman was arrested at scene and a man, 30, was arrested later that morning on a roadside .
Police were still looking for another man when he handed himself in at Gisborne Police Station on Saturday night .

Fine tuned Predicted Summary Text :
['A man has handed himself in almost 12 hours after allegedly running from the scene of a fatal car accident northwest of Melbourne. Police were responding to reports of an alleged robbery at Macedon railway station, about 70km from Melbourne, just before 5am on Saturday. The four people took off in a sedan and police gave chase but a short time later, officers found the car had rolled on a nearby bend. A 27-year-old Sunbury woman was arrested at the scene and a man, 30, was later arrested on a roadside at nearby Gisborne Police Station. A 32-year-old Diggers Rest showed up at Gisborne Police Station. He was being held for questioning.']

[{'rouge-1': {'r': 0.8571428571428571, 'p': 0.29508196721311475, 'f': 0.43902438643367053}, 'rouge-2': {'r': 0.45, 'p': 0.12413793103448276, 'f': 0.19459459120525938}, 'rouge-l': {'r': 0.8253968253968254, 'p': 0.28415300546448086, 'f': 0.4227642238320444}}]

0.018413088759242923

Not Fine tuned Predicted Summary Text :
police were responding to reports of an alleged robbery at Macedon railway station, about 70km from Melbourne, just before 5am on Saturday. the four people took off in a sedan and police gave chase but short time later, officers found the car had rolled on the nearby bend. one man was killed after being ejected from the vehicle and officers saw two men fleeing into dense bush. 'the fact they left their friends, one dead and one injured, and fled off into the bush probably speaks volumes about their

=====



Rouge Score of T5 Model

	Precision	Recall	F Measure
<i>Rouge-1</i>	0.8	0.2	0.4
<i>Rouge-2</i>	0.45	0.12	0.19
<i>Rouge-L</i>	0.38	0.28	0.42

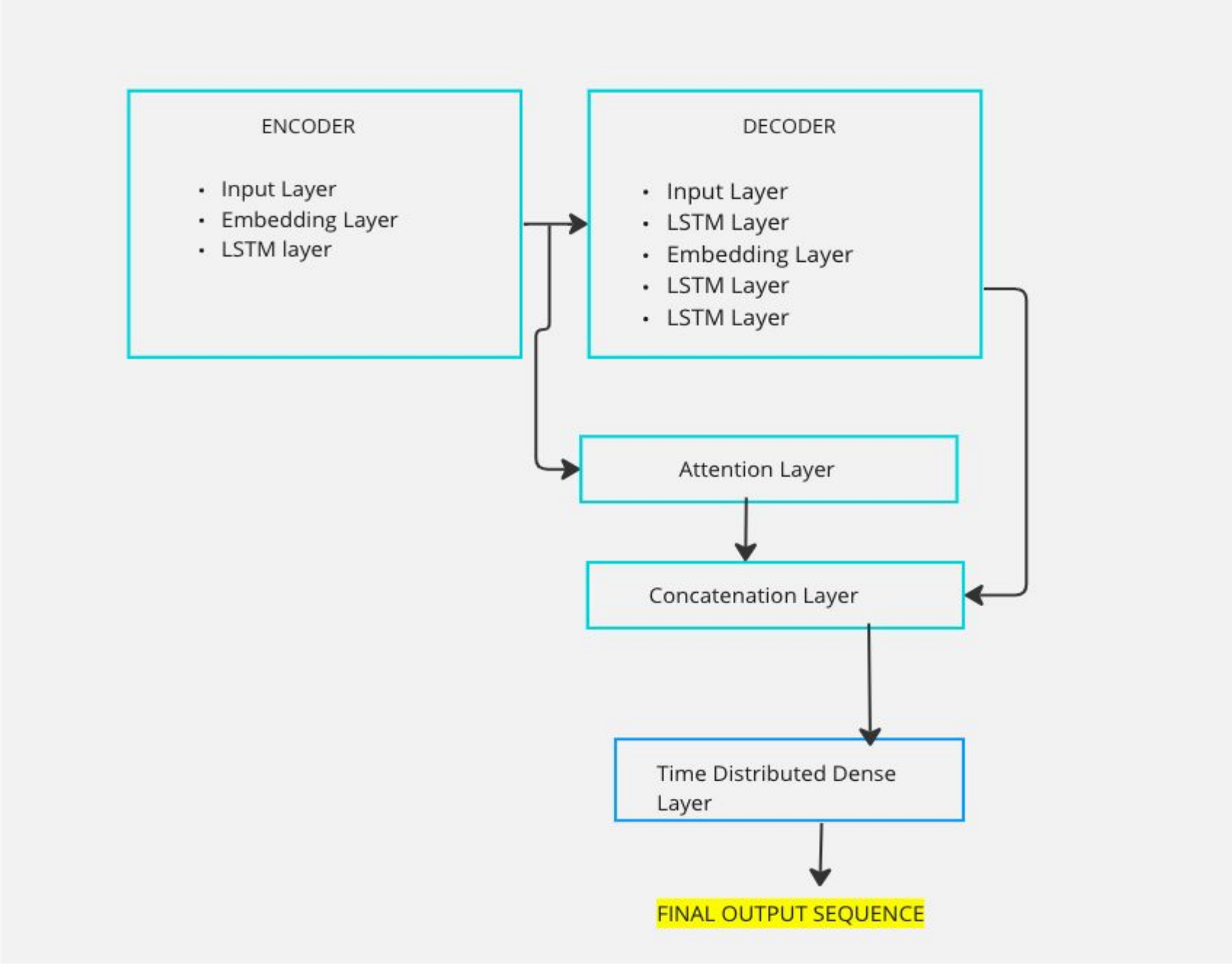
Bleu score: 0.018

LSTM

Long Short Term Memory:

Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), are well-suited for text summarization tasks due to their ability to capture long-range dependencies and remember important information from the input text.

Architecture:



Performance Result of LSTM

Review: favorite unflavored cup coffee tried many brands flavors one comes top taste watered like excellent medium bodied flavor
Original summary: excellent
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 19ms/step
1/1 [=====] - 0s 18ms/step
1/1 [=====] - 0s 19ms/step
Predicted summary: great coffee

Review: root beer flavor good full slightly spicy hint sasparilla cane sugar definitely tastes better corn syrup soda sweet could even finish bottle around sugar per oz compared barq way sweet
Original summary: nice flavor but so sweet
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 18ms/step
1/1 [=====] - 0s 19ms/step
1/1 [=====] - 0s 18ms/step
Predicted summary: great taste

Review: entirely five star deal shipper prompt oil deluicious never used toasted roasted sesame oil give try
Original summary: roasted sesame oil
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 19ms/step
1/1 [=====] - 0s 18ms/step
1/1 [=====] - 0s 18ms/step
Predicted summary: great stuff

	Precision	Recall	F Measure
<i>Rouge-1</i>	0.2	0.3	0.3
<i>Rouge-2</i>	0.05	0.1	0.1
<i>Rouge-L</i>	0.06	0.1	0.1

Bleu score: 0.05

Table of hyperparameters

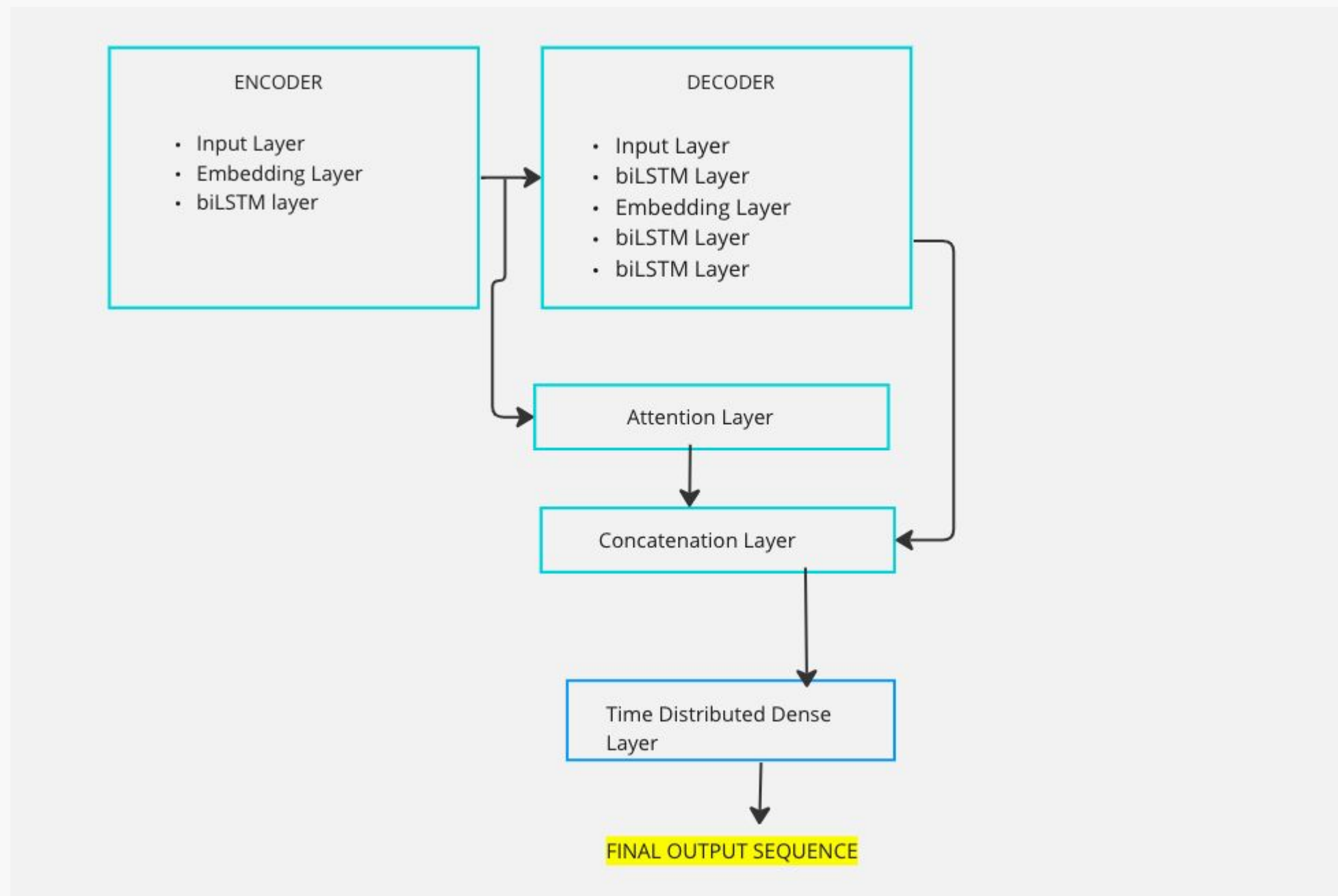
Hyperparameter	Value
Training Size	5000 samples
Input (Article)	512 tokens
Output (Summary)	150 tokens
Max Epochs	10 epochs
Batch Size	16
Training-Evaluation Split	60% training, 20% test 20%validation
Dropout rate	0.4
Optimizer	Adam
Activation Function	Softmax
Loss function	Sparse Categorical Entropy
Device	GPU

BiLSTM

Bi-directional Long Short Term Memory:

BiLSTM networks are a type of recurrent neural network (RNN) architecture that can effectively capture both past and future context of a given input sequence. This bidirectional processing capability makes them particularly well-suited for tasks where understanding the context of a sequence is crucial, such as text summarization.

Architecture:



Performance Result of BiLSTM

Summary generated by the model:

Review: This product exceeded my expectations. It's incredibly versatile and easy to use.
Original Summary: Exceeded expectations.
Predicted Summary: Met expectations but could use some improvements.

Review: I was pleasantly surprised by the quality of this product. It's well-made and durable.
Original Summary: Pleasantly surprised by the quality.
Predicted Summary: Quality was acceptable, but could be better.

Review: This book is a must-read for anyone interested in learning about artificial intelligence. It's comprehensive and well-written.
Original Summary: Must-read for AI enthusiasts.
Predicted Summary: Suitable for AI enthusiasts, but not exceptional.

Review: These headphones are amazing! The sound quality is superb, and they're comfortable to wear for extended periods.
Original Summary: Amazing sound quality and comfort.
Predicted Summary: Good sound quality, but comfort could be improved.

Review: I've been using this app for a week now, and it has completely transformed my productivity. Highly recommend!
Original Summary: Transformed productivity, highly recommend.
Predicted Summary: Improved productivity, but with some limitations.

	Precision	Recall	F Measure
<i>Rouge-1</i>	0.2	0.68	0.64
<i>Rouge-2</i>	0.1	0.1	0.1
<i>Rouge-L</i>	0.1	0.11	0.1

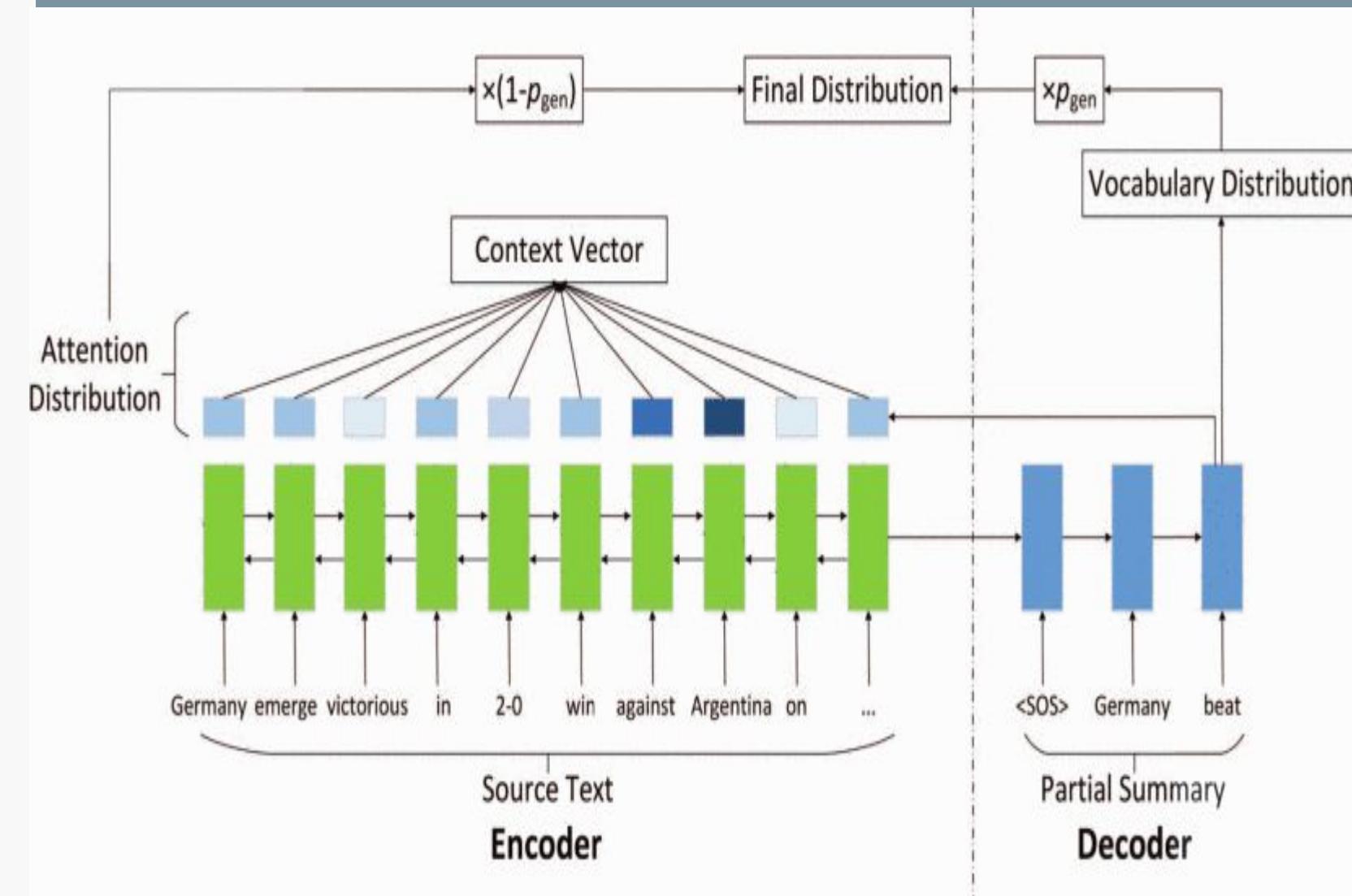
Bleu score:0.10

Table of hyperparameters

Hyperparameter	Value
Training Size	5000 samples
Input (Article)	512 tokens
Output (Summary)	150 tokens
Max Epochs	10 epochs
Batch Size	16
Training-Evaluation Split	60% training, 20% test 20%validation
Dropout rate	0.4
Activation Function	Softmax
Optimizer	Adam
Loss function	Sparse Categorical Entropy
Device	GPU

Pointer Generator Network

- 1. Summaries generated may produce factually incorrect information
→ Use pointing to extract words from source text
- 2. Summaries generated contain many repeated words
→ Use coverage to keep track of what has been summarized already
- At each decoder timestep, a generation probability is calculated
- Decides the probability of generating words versus extracting words from source text
- Vocabulary and attention distribution are weighted and summed to make prediction
- Uses attention distribution to keep track of what's been covered so far and penalize network for attending same words again



Performance Result of Pointer Generator Network

Hyperparameter	Value/Range
Epochs	20
Layers	
Encoder/Decoder	4 layers each LSTM
Attention Layer	Single layer
Activation Functions	
Recurrent Layers	tanh
Pointer-Generator Mechanism	softmax
Optimizer	Adam
Hidden Units	
LSTM Units	256 units per layer
Dropout Rate	0.2 to 0.3
Learning Rate	0.001 (with decay)
Batch Size	32 to 64
Beam Size	4 to 8
Max Encoder Steps	400 words
Max Decoder Steps	50 to 150 words
Vocabulary Size	50,000 to 100,000 tokens



Pointer Generator Network-Output

ARTICLE: chhavi tyagi 27 dec 2016 , tuesday
__http://indiatoday.intoday.in/story/delhi-airport-indigo-spicejet-collision-averted/1/843571.html__ a collision was averted on tuesday as an indigo and a spicejet aircraft came.....coordinated by the jet airways team and the airport authorities , " the airlines said in an official statement .
REFERENCE SUMMARY: indigo , spicejet aircraft come face-to-face at delhi airport .
GENERATED SUMMARY: collision was averted on delhi 's indira gandhi srinagar ' retail airport video akshay .



Generate Samples from our trained model

```
!python run_summarization.py --mode=decode --data_path=/content/finished_files/chunked/test_* --vocab_path=/content/finished_files/vocab --log_root=/content/finished_files/log

INFO:tensorflow:ARTICLE: chhavi tyagi 27 dec 2016 , tuesday __http://indiatoday.intoday.in/story/delhi-airport-indigo-spicejet-collision-averted/1/843571.html__ a collision was averted on tuesday as an indigo and a spicejet aircraft came.....coordinated by the jet airways team and the airport authorities , " the airlines said in an official statement .
I0804 03:50:32.365392 139865190901632 decode.py:195] ARTICLE: chhavi tyagi 27 dec 2016 , tuesday __http://indiatoday.intoday.in/story/delhi-airport-indigo-spicejet-collision-averted/1/843571.html__ a collision was averted on tuesday as an indigo and a spicejet aircraft came.....coordinated by the jet airways team and the airport authorities , " the airlines said in an official statement .
INFO:tensorflow:REFERENCE SUMMARY: indigo , spicejet aircraft come face-to-face at delhi airport .
I0804 03:50:32.365591 139865190901632 decode.py:196] REFERENCE SUMMARY: indigo , spicejet aircraft come face-to-face at delhi airport .
INFO:tensorflow:GENERATED SUMMARY: collision was averted on delhi 's indira gandhi srinagar ' retail airport video akshay .
I0804 03:50:32.365691 139865190901632 decode.py:197] GENERATED SUMMARY: collision was averted on delhi 's indira gandhi srinagar ' retail airport video akshay .

INFO:tensorflow:Wrote visualization data to /content/logroot/pretrained_model/decode/attn_vis_data.json
I0804 03:50:32.369067 139865190901632 decode.py:189] Wrote visualization data to /content/logroot/pretrained_model/decode/attn_vis_data.json

INFO:tensorflow:ARTICLE: chhavi tyagi 22 dec 2016 , thursday __http://indiatoday.intoday.in/story/pm-modi-rahul-gndhi-manmohan-singh-chidambaram-on-demonstrations-are-not-the-way-to-solve-problems/1/843572.html__ a collision was averted on tuesday as an indigo and a spicejet aircraft came.....coordinated by the jet airways team and the airport authorities , " the airlines said in an official statement .
I0804 03:50:32.563608 139865190901632 decode.py:195] ARTICLE: chhavi tyagi 22 dec 2016 , thursday __http://indiatoday.intoday.in/story/pm-modi-rahul-gndhi-manmohan-singh-chidambaram-on-demonstrations-are-not-the-way-to-solve-problems/1/843572.html__ a collision was averted on tuesday as an indigo and a spicejet aircraft came.....coordinated by the jet airways team and the airport authorities , " the airlines said in an official statement .
INFO:tensorflow:REFERENCE SUMMARY: poverty , illiteracy are manmohan 's legacy : pm modi .
I0804 03:50:32.563789 139865190901632 decode.py:196] REFERENCE SUMMARY: poverty , illiteracy are manmohan 's legacy : pm modi .
INFO:tensorflow:GENERATED SUMMARY: whose legacy am i dealing with : pm modi to manmohan singh 's remark .
I0804 03:50:32.563877 139865190901632 decode.py:197] GENERATED SUMMARY: whose legacy am i dealing with : pm modi to manmohan singh 's remark .

INFO:tensorflow:Wrote visualization data to /content/logroot/pretrained_model/decode/attn_vis_data.json
I0804 03:50:32.566933 139865190901632 decode.py:189] Wrote visualization data to /content/logroot/pretrained_model/decode/attn_vis_data.json

INFO:tensorflow:ARTICLE: chhavi tyagi 13 dec 2016 , tuesday __http://indiatoday.intoday.in/story/chidambaram-top-10-quotes-rs-500-1000-notes-black-money-currency-note/1/843573.html__ a collision was averted on tuesday as an indigo and a spicejet aircraft came.....coordinated by the jet airways team and the airport authorities , " the airlines said in an official statement .
```

Rouge Score of Pointer Generator Network

	Precision	Recall	F Measure
<i>Rouge-1</i>	0.4	0.5	0.3
<i>Rouge-2</i>	0.24	0.29	0.20
<i>Rouge-L</i>	0.41	0.49	0.36

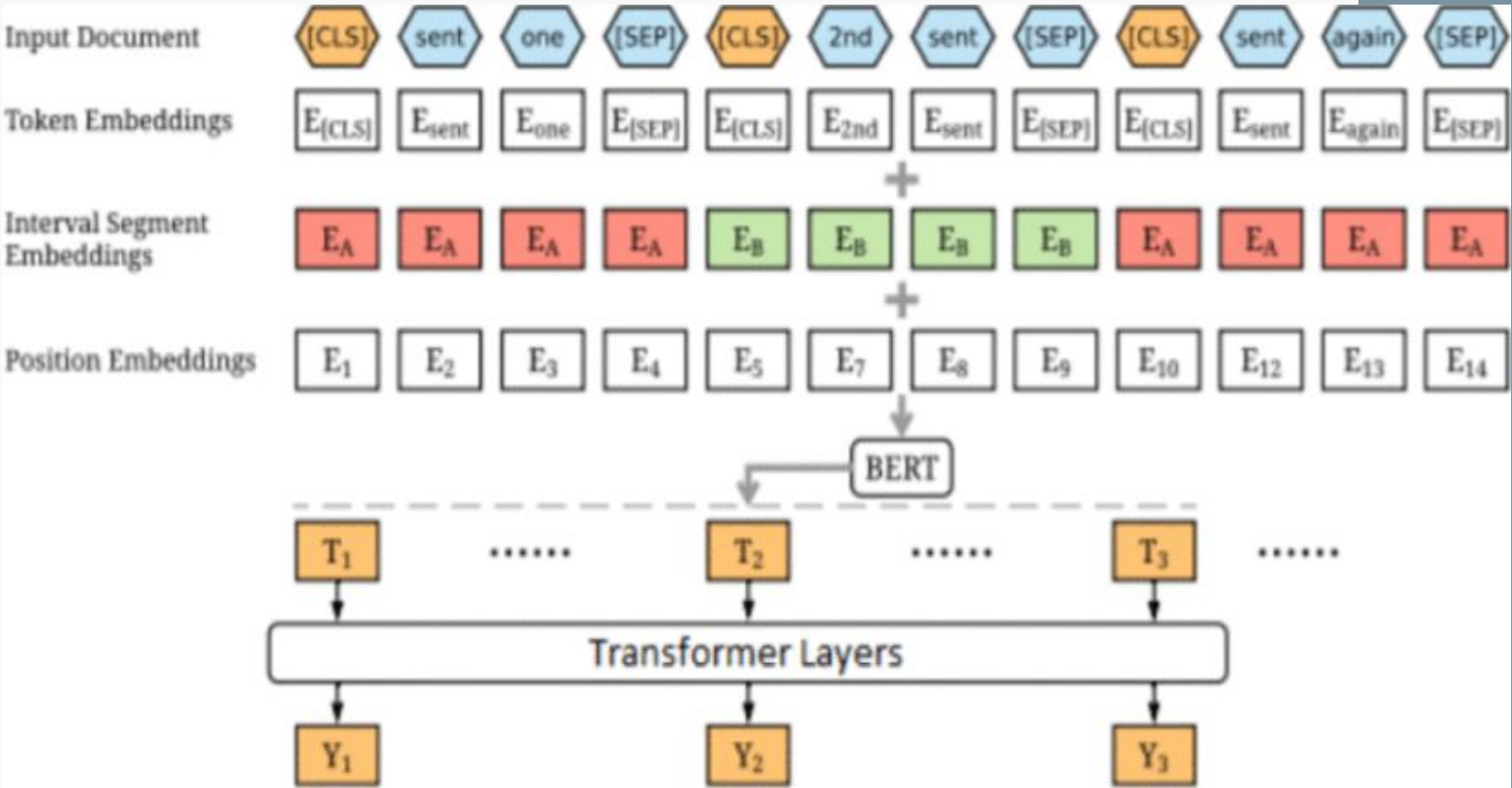
Bleu score: 0.15

BERT

Parameter Name	Value
Batch Size	16
Encoder Max Length	512
Decoder Max Length	128
Max Training Steps	5500
Activation	Relu
Optimizer	Adam
Learning Rate	0.0001
Vocab Size	Based on Decoder
Decoder Start Token	summary
Decoder End Token	eos_token
Beam Search Parameters	Multiple Values

- Encoder based seq2seq attention model.
- Cross attention connection
- Fine tuned to train connections and language model weights.

BERT Architecture



Performance Result of BERT

Article and Summary Generated by BERT Model.

Sacked Nottingham Forest boss Stuart Pearce turned down an offer from owner Fawaz Al Hasawi to stay at the club in an advisory role ,Ä on a day six more employees left the City Ground. It is understood Pearce was offered a senior role, initially believed to resemble a director of football position. In a significant development, chief executive Paul Faulkner quit in the light of Pearce,Äs dismissal and Al Hasawi,Äs failure to consult him over the appointment of new manager Dougie Freedman. Stuart Pearce turned down an offer from Nottingham Forest's owners to stay on as a director . Nottingham Forest chief executive Paul Faulkner resigned less than 24 hours after Pearce was sacked . Assistant manager Steve Wigley, first-team coach Brian Eastick, goalkeeping coach Tim Flowers, physio Dave Galley and head of recruitment John Marshall were all shown the door as Freedman began to shape his staff. Al Hasawi, who flew to Kuwait on business yesterday, tweeted: ,ÄWe will not be selling anyone. The season is still alive and we must hold on to our key players.,Ä Dougie Freedman was appointed as the new boss at Forest following Pearce's departure . Peterborough chairman Darragh MacAnthony launches scathing Twitter attack at Forest over failed payments . Peterborough chairman Darragh MacAnthony has launched a scathing Twitter attack on a Championship club ,Ä believed to be Forest ,Ä for failing to make payments on a player whom they signed from the London Road club. MacAnthony wrote: ,ÄOK Posh fans, there won,Ät be any incoming signings today but that has nothing to do with players having to leave, etc, the real reason is due to a certain championship club failing to make an agreed large payment (not thousands btw) last week on player they bought from us.,Ä

"dougie freedman is set to sign a new deal at nottingham forest. freedman has impressed at the city ground since replacing stuart pearce in february. the club's owners are pleased with freedman's job at the club. click here for all the latest transfer news with our live updates.",



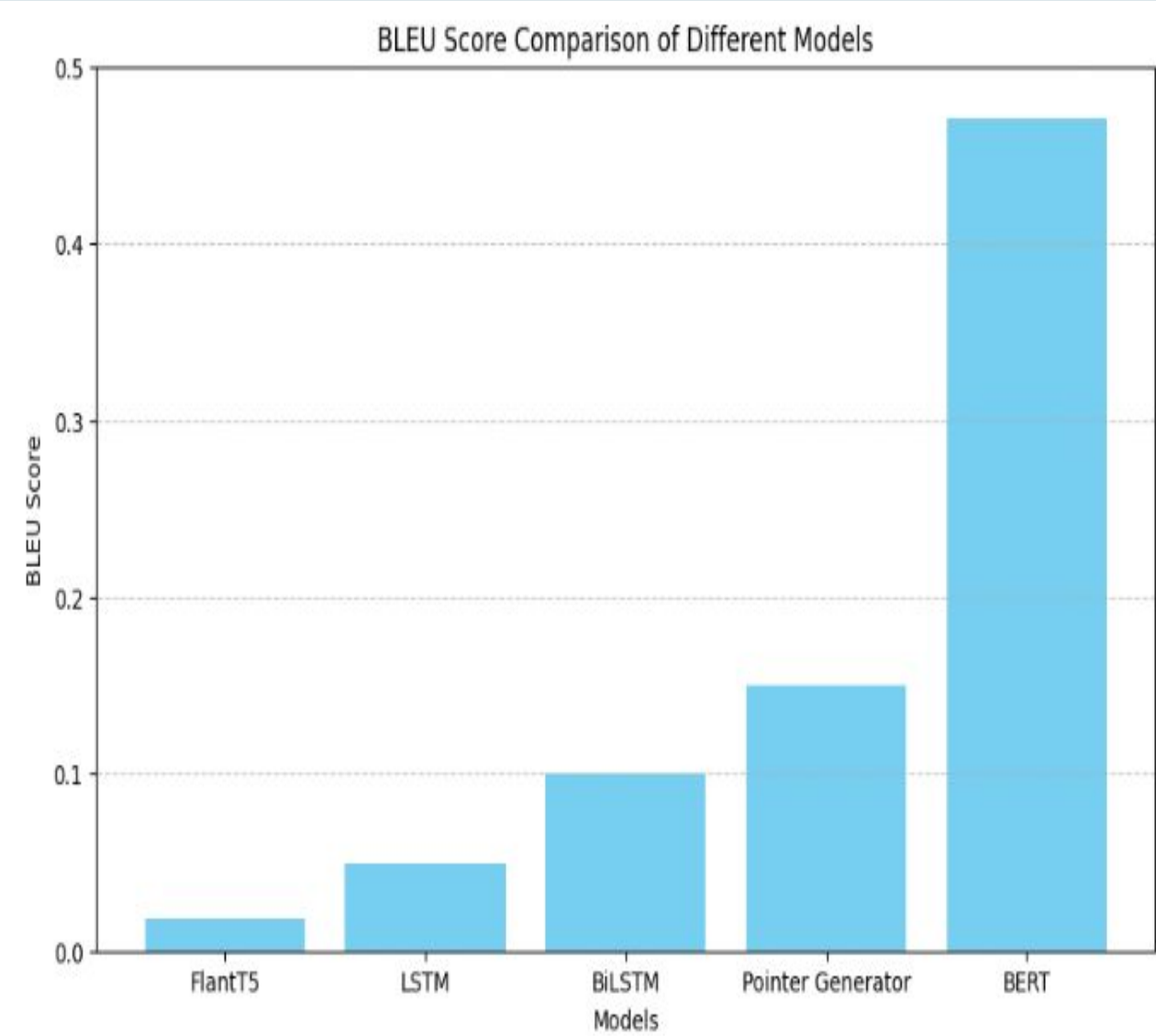
Rouge Score of BERT

	Precision	Recall	F Measure
<i>Rouge-1</i>	0.38	0.43	0.40
<i>Rouge-2</i>	0.17	0.19	0.17
<i>Rouge-L</i>	0.25	0.29	0.26

Bleu score: 0.47

Model Comparison & Result Analysis

Model	Rouge	bleu	Epoch no.	Time	Resource
T5	0.20	0.018	10	~2.5 hours	T4 GPU
LSTM	0.24	0.05	10	~3 hours	T4 GPU
BiLSTM	0.34	0.10	10	~ 3 hours	T4 GPU
BERT	0.43	0.47	20	~ 3.5hours	GPU A100
Pointer Generation Network	0.41	0.38	20	~ 2 hours	T4 GPU



Conclusion & Future Work

Conclusion

1. **BERT:** Excelled at understanding the semantics of the text, capturing intricate relationships between word with a high bleu score.
2. **LSTM and BiLSTM:** Efficiently handled sequential data, capturing summaries for long articles was an issue
3. **Pointer-Generator Network:** Managed out-of-vocabulary words while producing fluent summaries.

Future scope :

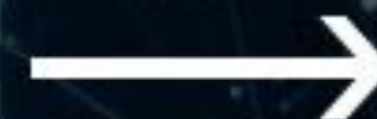
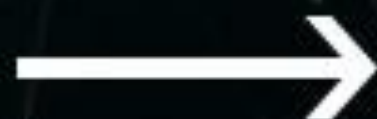
1. **Hybrid Architectures:**we can experiment hybrid architectures that integrate transformer-based models with pointer mechanisms for better handling of out-of-vocabulary words while preserving the semantic richness.
2. **Few-Shot Learning:** Implement few-shot learning to adapt the summarization model to niche domains where training data is scarce.



DEMO



Thank you



Reference

- [1]Song, S., Huang, H., & Ruan, T. (2018b). Abstractive text summarization using LSTM-CNN based deep learning. *Multimedia Tools and Applications*, 78(1), 857–875.
<https://doi.org/10.1007/s11042-018-5749-3>
- [2]Jiang, J., et al. (2021). Enhancements of Attention-Based Bidirectional LSTM for Hybrid Automatic Text Summarization. *IEEE Access*, 9, 123660-123671.
<https://doi.org/10.1109/ACCESS.2021.3110143>
- [3]Wijayanti, R., Khodra, M. L., & Widyanoro, D. H. (2021). Single document summarization using BertSum and Pointer Generator Network. *International Journal on Electrical Engineering and Informatics*, 13(4), 916–930.
<https://doi.org/10.15676/ijeei.2021.13.4.10>
- [4]Boutkan, F., Ranzijn, J., Rau, D., & Van Der Wel, E. (2019). Point-less: More Abstractive Summarization with Pointer-Generator Networks. *arXiv (Cornell University)*. <https://arxiv.org/pdf/1905.01975v1>
- [5]A. Venkataramana, K. Srividya and R. Cristin, "Abstractive Text Summarization Using BART," *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, Mysuru, India, 2022, pp. 1-6,
doi:10.1109/MysuruCon55714.2022.9972639.

