Success of AI Writers

- Ahmet Emre Usta
- Hüseyin Yiğit Ülker

High-level Overview of the Paper

- ❖ Insufficiency of plagiarism checkers
  - ➢ Inability to detect texts paraphrased with artificial intelligence

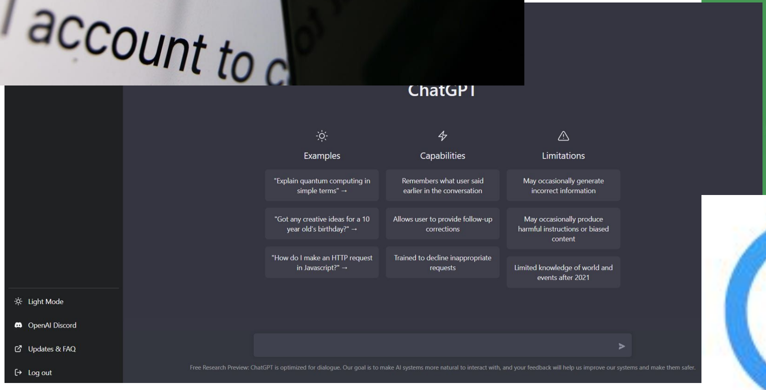
- ❖ Deep Learning Approach for plagiarism detection
  - ➢ Finding semantic similarity with Bert,RoBERTa , DeBERTa , AlBERT pre-train models

- ❖ Dataset creation

- ❖ Analysis of results

Problem Statement and Motivation

The incapability of current plagiarism detection tools to detect the similarity of writings paraphrased by AI systems.



Related Work

SpringerLink

Original Article | Open Access | Published: 24 June 2022

Reliable plagiarism detection system based on deep learning approaches

Mohamed A. El-Rashidy, Ramy G. Mohamed, Nawal A. El-Fishawy & Marwa A. Shouman

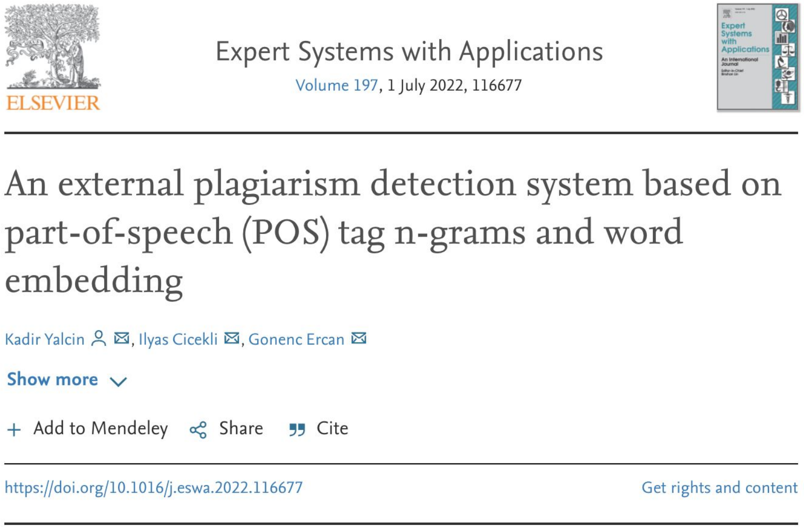
Neural Computing and Applications **34**, 18837–18858 (2022) | Cite this article

- Models : BERT ,RoBERTa, GloVe
- Similarity Measure : cosine similarity, manhattan distance, euclidean distance , dot product similarity
- Dataset : SNLI and STS benchmark
- Paper link : (PDF) NLP based Deep Learning Approach for Plagiarism Detection

Related Work

2- An external plagiarism detection system based on part-of-speech (POS) tag n-grams and word embedding

- Dataset : PAN-PC 11
- Paper link : [An external plagiarism detection system based on part-of-speech (POS) tag n-grams and word embedding - ScienceDirect]


ELSEVIER
Expert Systems with Applications
Volume 197, 1 July 2022, 116677

An external plagiarism detection system based on part-of-speech (POS) tag n-grams and word embedding

Kadir Yalcin, Ilyas Cicekli, Gonenc Ercan

Show more

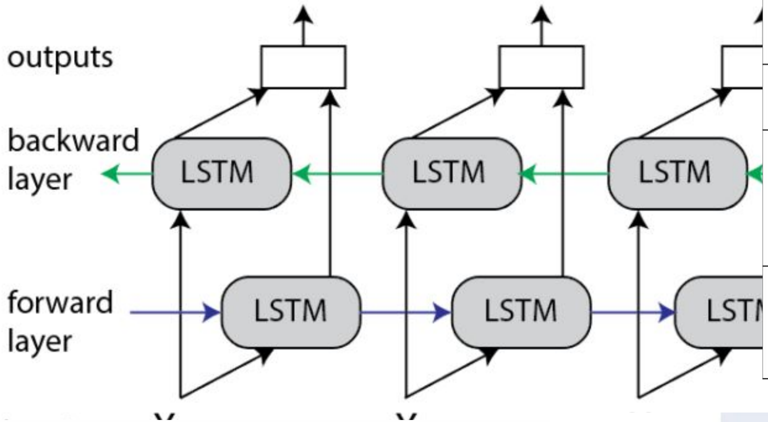
Add to Mendeley Share Cite

https://doi.org/10.1016/j.eswa.2022.116677 Get rights and content

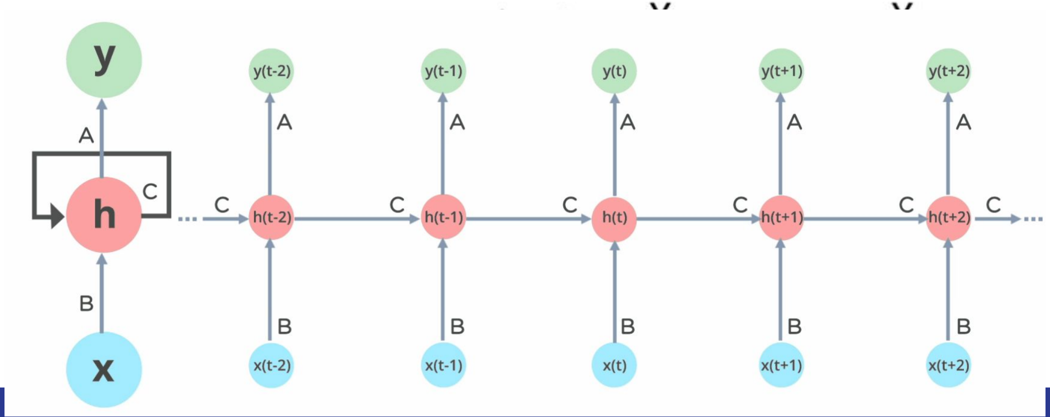
Key technical ideas

Pretrained-Models

Bert-base-uncased
Roberta-base
Alberta-base
Deberta-v3-xsmall



	BERT	RoBERTa
Size (millions)	**Base**: 110 **Large**: 340	**Base**: 110 **Large**: 340
Training Time	**Base**: 8 x V100 x 12 days* **Large:** 64 TPU Chips x 4 days (or 280 x V100 x 1 days*)	**Large:** 1024 x V100 x 1 day; 4-5 times more than BERT.
Performance	Outperforms state-of-the-art in Oct 2018	2-20% improvement over BERT
Data	16 GB BERT data (Books Corpus + Wikipedia). 3.3 Billion words.	160 GB (16 GB BERT data + 144 GB additional)
Method	BERT (Bidirectional Transformer with MLM and NSP)	BERT without NSP**



Model Structures

BERT-BASE-UNCASED

Layer (type)	Output Shape	Param #	Connected to
input_ids (InputLayer)	[(None, 128)]	0	[]
attention_masks (InputLayer)	[(None, 128)]	0	[]
token_type_ids (InputLayer)	[(None, 128)]	0	[]
bert (TFBertMainLayer)	TFBaseModelOutputWithPoolingAndCrossAttentions(last_hidden_state=(None, 128, 768), pooler_output=(Non ... one, attentions=None, cross_attentions=None)	109482240	['input_ids[0][0]', 'attention_masks[0][0]', 'token_type_ids[0][0]']
bidirectional (Bidirectional)	(None, 128, 128)	426496	['bert[0][0]']
global_average_pooling1d (GlobalAveragePooling1D)	(None, 128)	0	['bidirectional[0][0]']
global_max_pooling1d (GlobalMaxPooling1D)	(None, 128)	0	['bidirectional[0][0]']
concatenate (Concatenate)	(None, 256)	0	['global_average_pooling1d[0][0]', 'global_max_pooling1d[0][0]']
dropout_37 (Dropout)	(None, 256)	0	['concatenate[0][0]']
dense (Dense)	(None, 3)	771	['dropout_37[0][0]']

Total params: 109,909,507
Trainable params: 427,267
Non-trainable params: 109,482,240

feature extraction

Total params: 109,909,507
Trainable params: 109,909,507
Non-trainable params: 0

fine tuning

DEBERTA-V3-XSMALL

Layer (type)	Output Shape	Param #	Connected to
input_ids (InputLayer)	[(None, 128)]	0	[]
attention_masks (InputLayer)	[(None, 128)]	0	[]
token_type_ids (InputLayer)	[(None, 128)]	0	[]
deberta (TFDebertaV2MainLayer)	TFBaseModelOutput(last_hidden_state=(N ... , attentions=None)	70682112	['input_ids[0][0]', 'attention_masks[0][0]', 'token_type_ids[0][0]']
bidirectional (Bidirectional)	(None, 128, 128)	229888	['deberta[0][0]']
global_average_pooling1d (GlobalAveragePooling1D)	(None, 128)	0	['bidirectional[0][0]']
global_max_pooling1d (GlobalMaxPooling1D)	(None, 128)	0	['bidirectional[0][0]']
concatenate (Concatenate)	(None, 256)	0	['global_average_pooling1d[0][0]', 'global_max_pooling1d[0][0]']
dropout (Dropout)	(None, 256)	0	['concatenate[0][0]']
dense (Dense)	(None, 3)	771	['dropout[0][0]']

Total params: 70,912,771
Trainable params: 230,659
Non-trainable params: 70,682,112

feature extraction

Total params: 70,912,771
Trainable params: 70,912,771
Non-trainable params: 0

fine tuning

Datasets

1- SNLI (Stanford Natural Language Inference)

2- ANLI (Adversarial Natural Language Inference)

3- Paraphrased Articles using GPT-3

The Stanford Natural Language Processing Group people publications research blog software teaching join local

The Stanford Natural Language Inference (SNLI) Corpus

Natural Language Inference (NLI), also known as Recognizing Textual Entailment (RTE), is the task of determining the inference relation between two (short, ordered) texts: *entailment*, *contradiction*, or *neutral* (MacCartney and Manning 2008).

The Corpus

The Stanford Natural Language Inference (SNLI) corpus (version 1.0) is a collection of 570k human-written English sentence pairs manually labeled for balanced classification with the labels *entailment*, *contradiction*, and *neutral*. We aim for it to serve both as a benchmark for evaluating representational systems for text, especially including those induced by representation-learning methods, as well as a resource for developing NLP models of any kind.

The following paper introduces the corpus in detail. If you use the corpus in published work, please cite it:

> Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learnin
> *the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP).* [pdf] [bib]

Here are a few example pairs taken from the development portion of the corpus. Each has the judgments of five mechanical

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cat
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	entailment E E E E E	Some men are playing a sport.
A smiling costumed woman is holding an umbrella.	neutral N N E C N	A happy woman in a fairy costume holds an umbrella.

README.md

Adversarial NLI

Papers

Dataset

Adversarial NLI: A New Benchmark for Natural Language Understanding

Annotations of the Dataset for Error Analysis

ANLIzing the Adversarial Natural Language Inference Dataset

Dataset

Version 1.0 is available here: https://dl.fbaipublicfiles.com/anli/anli_v1.0.zip.

Format

The dataset files are all in JSONL format (one JSON per line). Below is one example (in JSON format) with self-explanatory fields.
Note that each example (each line) in the files contains a `uid` field represents **a unique id** across all the examples in all there rounds of ANLI.


```
{
    "uid": "8a91e1a2-9a32-4fd9-b1b6-bd2ee2287c8f",
    "premise": "Javier Torres (born May 14, 1988 in Artesia, California) is an undefeated Mexican Ar
                Torres was the second rated U.S. amateur boxer in the Super Heavyweight division anc
    "hypothesis": "Javier was born in Mexico",
    "label": "c",
    "reason": "The paragraph states that Javier was born in the California, US."
}
```


AHMET EMRE AND 1 COLLABORATOR · UPDATED 4 DAYS AGO

6 New Notebook Download (162 kB)

Paraphrased Articles using GPT-3

Paraphrased Academic Article Dataset Generated using GPT-3

Model Train Details

DATASET INFORMATIONs				
Dataset Name	Original Dataset Name	Data Length	Size(MB)	Details
df_train	snli_data	549361	65	SNLI Train Data
df_train_longer_sentences	snli_data	266582	38	Half of SNLI test data (only longer sentences)
df_validation	snli_data	9842	1.2	SNLI Validation Data
df_test	snli_data	9824	1.2	SNLI Test Data
ANLI	anli_data	169265	66.4	All ANLI data (train+validation+test)

Model Train Details

MODELS						
Model Name	**Train Datasets**	**Val Dataset**	**Max String Length**	**Batch Size**	**Epoch**	**Total Parameters**
albert-base	*df_train*	df_validation	128	32	*8*	12,110,851
bert-base-uncased	*df_train*	df_validation	128	32	*2*	109,909,507
roberta-base	*df_train_longer_sentences*	df_validation	128	32	*4*	125,072,899
deberta-v3-xsmall	***df_train+anli***	df_validation	128	32	*4*	70,912,771

Technical Metrics

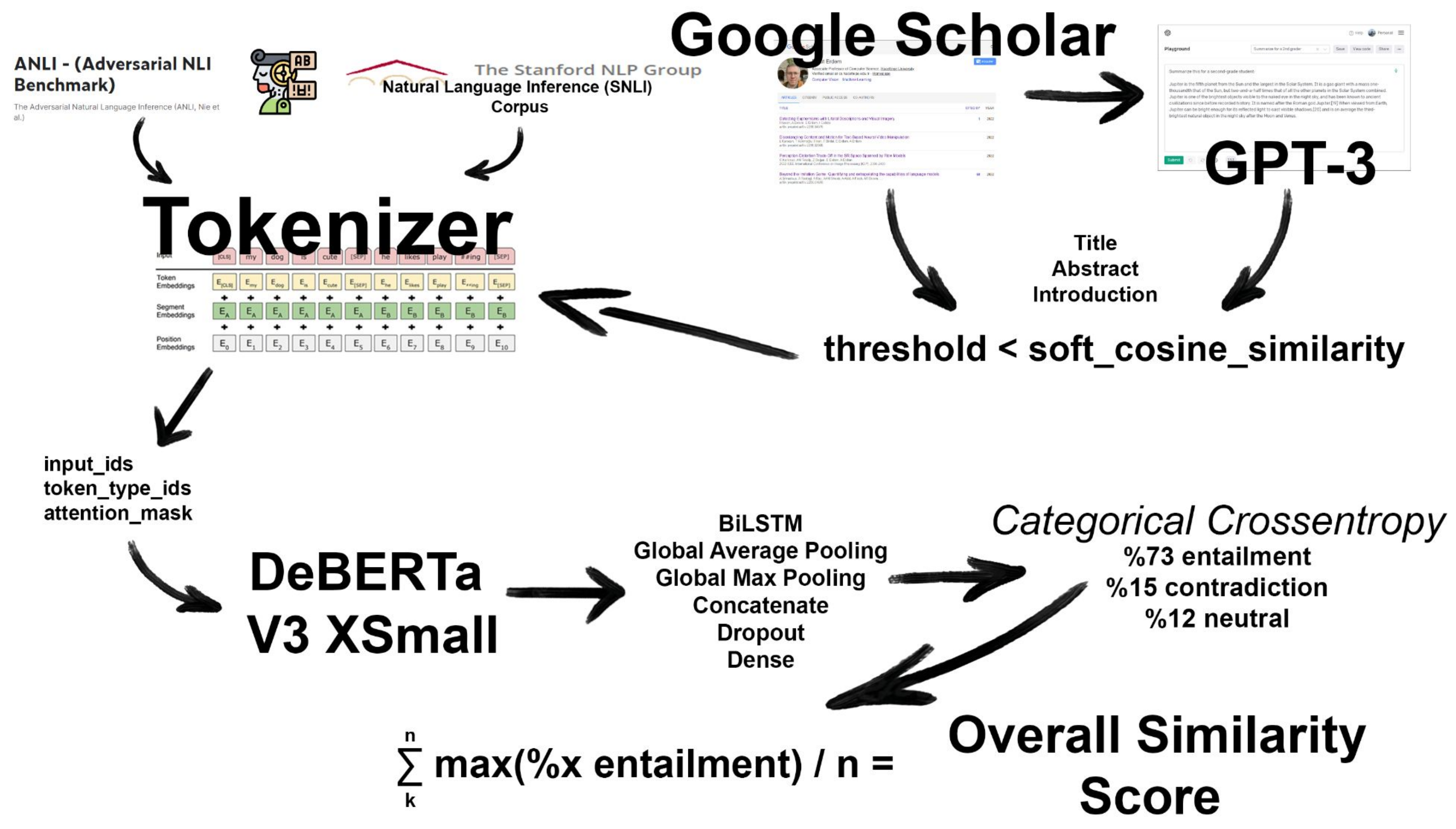
SNLI DATA TEST RESULTS				
Model Name	**df_test_time(second)**	**df_test_loss(x10^-4)**	**df_test_acc(x10^-4)**	**Test Environment**
albert-base	17.7	4179	8391	Google Colab Pro
bert-base-uncased	22	2778	9022	Google Colab Pro
roberta-base	17.1	2749	9053	Google Colab Pro
deberta-v3-xsmall	23	2535	***9108***	Google Colab Pro

(more accurate better)

Technical Metrics

ANLI DATA TEST RESULTS				
Model Name	**ANLI_test_time(second)**	**ANLI_test_loss(x10^-4)**	**ANLI_test_acc(x10^-4)**	**Test Environment**
albert-base	285	11476	4985	Google Colab Pro
bert-base-uncased	364	11211	6025	Google Colab Pro
roberta-base	277	15230	5424	Google Colab Pro
deberta-v3-xsmall	***232***	2199	*9226*	Google Colab Pro

(faster better)



ANLI - (Adversarial NLI Benchmark)
The Adversarial Natural Language Inference (ANLI, Nie et al.)
The Stanford NLP Group
Natural Language Inference (SNLI) Corpus
Google Scholar
GPT-3
Tokenizer
Input
[CLS] my dog is cute [SEP] he likes play ##ing [SEP]
Token Embeddings
Segment Embeddings
Position Embeddings
Title
Abstract
Introduction
threshold < soft_cosine_similarity
input_ids
token_type_ids
attention_mask
DeBERTa V3 XSmall
BiLSTM
Global Average Pooling
Global Max Pooling
Concatenate
Dropout
Dense
Categorical Crossentropy
%73 entailment
%15 contradiction
%12 neutral
$\sum_{k}^{n}$ max(%x entailment) / n =
Overall Similarity Score

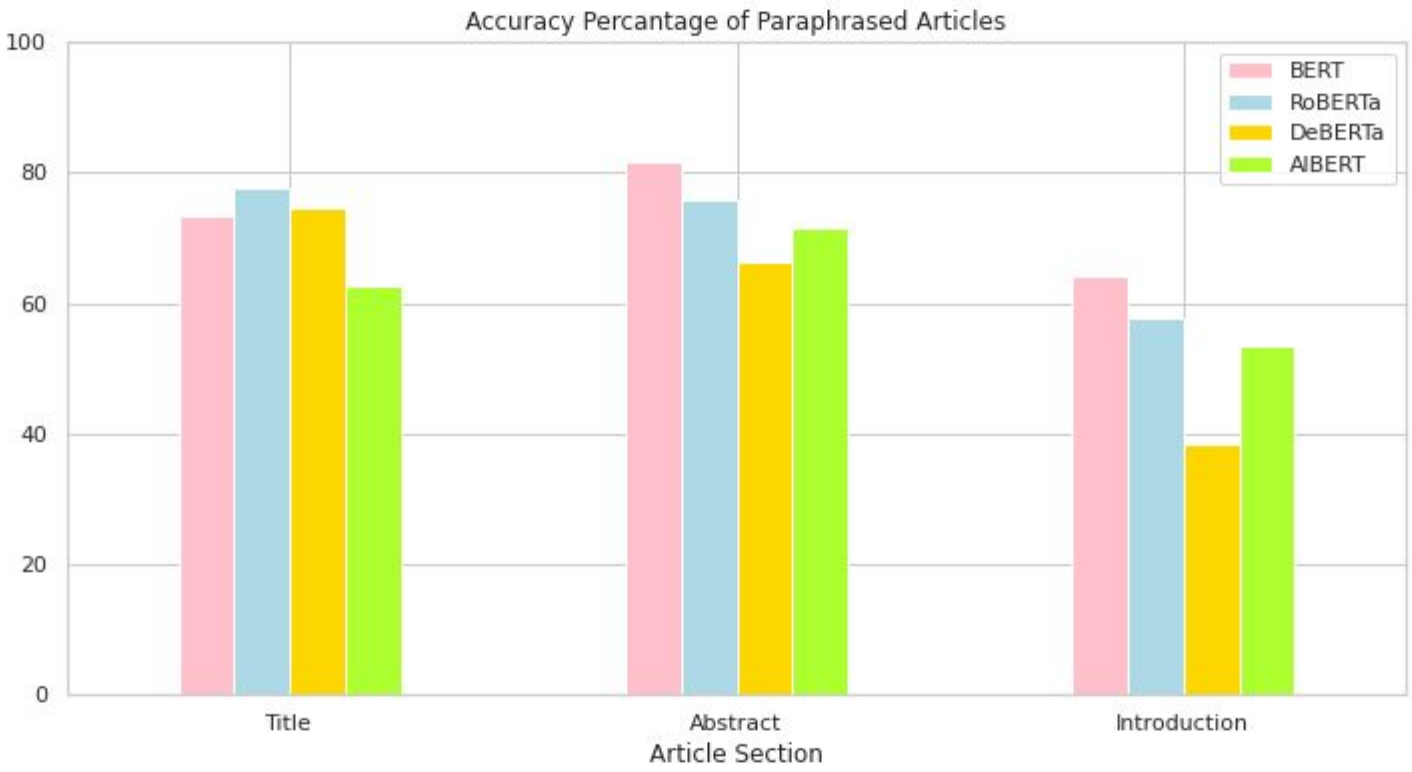
Experiment Results

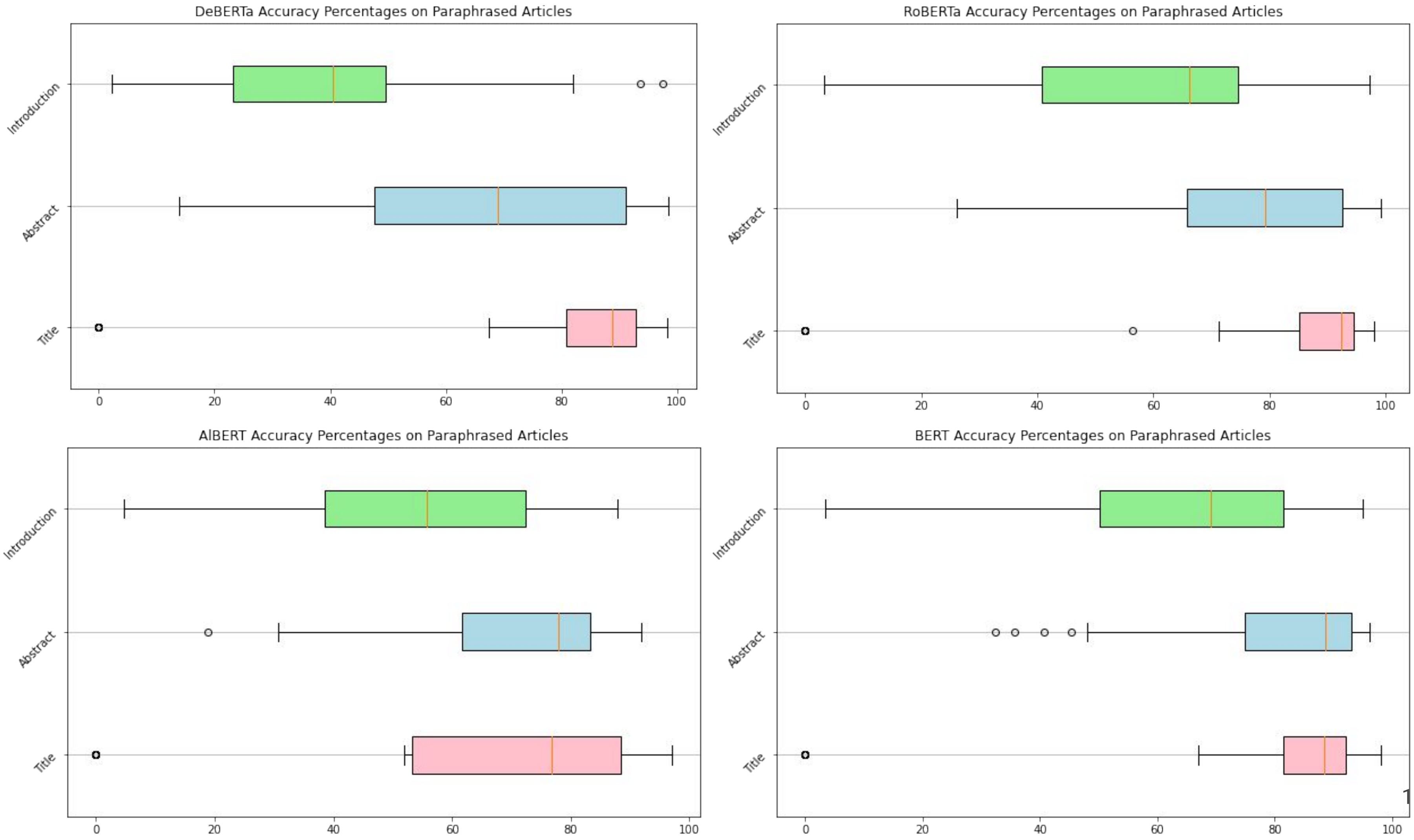
Paraphrased Articles using GPT-3			
Model	**Title Similarity Percentage**	**Abstract Similarity Percentage**	**Introduction Similarity Percentage**
BERT	73,3	***81,8***	***64,1***
RoBERTa	***77,8***	75,9	57,7
DeBERTa	74,6	66,4	38,5
AlBERT	62,6	71,5	53,3

Experiment Results

BERT gave us *better results* even though it has *less parameters* and is trained with *data consisting of fewer but longer sentences*

Experiment Results




DeBERTa Accuracy Percentages on Paraphrased Articles
RoBERTa Accuracy Percentages on Paraphrased Articles
AlBERT Accuracy Percentages on Paraphrased Articles
BERT Accuracy Percentages on Paraphrased Articles
Introduction
Abstract
Title
0
20
40
60
80
100

Strength and Weakness

W - 128 string per sentence limit

W - longer run times

S - semantic similarity check

S - similarity check on every sentences of two paragraph

Future Direction

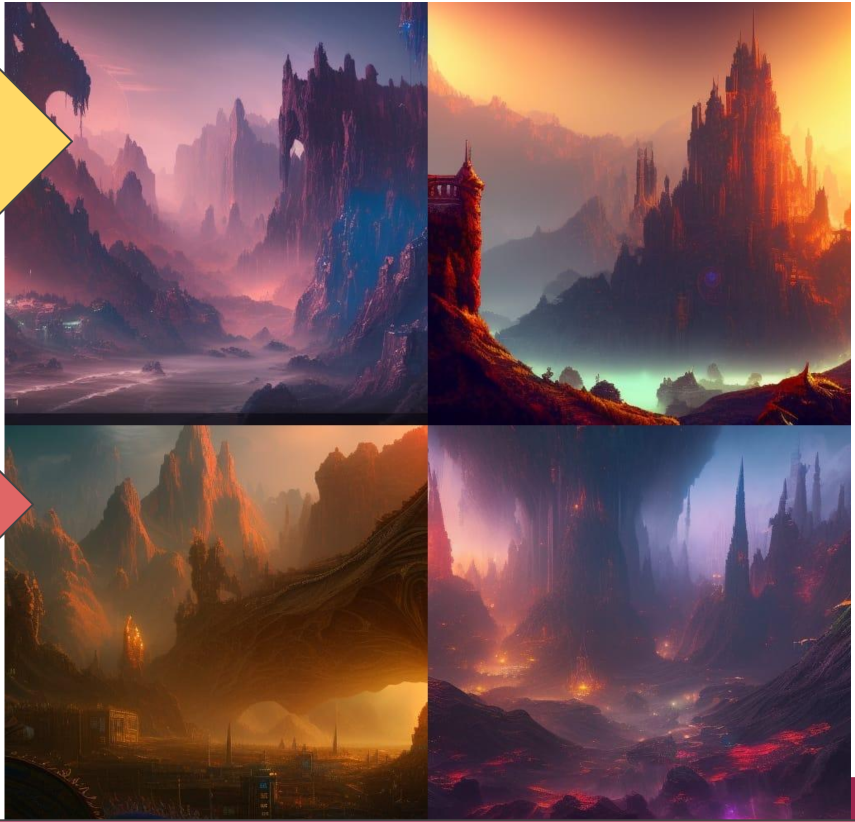
Training with Higher **Epochs** Numbers

Working on **PAN-PC 11** Dataset

Finding a **Faster** Solutions

Trying **Longer** String Sequences

Fully Functional **Hugging Face** Spaces



This pictures created the prompt *"future of ai"* using Night Cafe's stable diffusion model

DEMO

Semantic Similarity Checker

Original Text

Moussaka is one of the best known Greek dishes – a baked casserole consisting of ground lamb meat and layers of sliced eggplant, covered with a thick layer of bechamel sauce that gets golden and crusty as it bakes. The lamb is sometimes replaced with beef, while the eggplants might be replaced with zucchini or potatoes. It is likely that moussaka has Middle-Eastern origins, and it was introduced when the Arabs brought the eggplant to Greece. Its Greek name mousakás is derived from the Turkish musakka, which came from the Arabic word musaqqa'ah, meaning chilled.Moussaka is not an everyday dish–it is baked as a special treat for guests and family on festive days. An exotic version of lasagna, without the pasta, moussaka is exceptionally healthy due to the abundance of vegetables used in the dish. It is commonly cut into squares and served warm, not hot, as the dish needs some resting time in order to firm up.

Suspected Text

Moussaka is a traditional Greek dish made up of ground lamb (or sometimes beef) and layers of eggplant, all covered in a thick and golden bechamel sauce. It is thought to have originated from the Middle East, as the name is derived from the Turkish musakka, which comes from the Arabic word musaqqa'ah, meaning chilled. It is not a meal eaten every day, but is a special treat that is served on festive occasions. It is similar to lasagna, but without the pasta, and is very healthy due to the vegetables used. It is usually cut into squares and served warm after it has had some time to cool and firm up.

Clear | Submit

Overall Similarity

46.058001120885216

Flag

https://c26df07d-df3d-479f.gradio.live/

(link expires in a few hours)

Examples

Original Text	Suspected Text
Moussaka is one of the best known Greek dishes – a baked casserole consisting of ground lamb meat and layers of sliced eggplant, covered with a thick layer of bechamel sauce that gets golden and crusty as it bakes. The lamb is sometimes replaced with beef, while the eggplants might be replaced with zucchini or potatoes. It is likely that moussaka has Middle-Eastern origins, and it was introduced when the Arabs brought the eggplant to Greece. Its Greek name mousakás is derived from the Turkish musakka, which came from the Arabic word musaqqa'ah, meaning chilled.Moussaka is not an everyday dish–it is baked as a special treat for guests and family on festive days. An exotic version of lasagna, without the pasta, moussaka is exceptionally healthy due to the abundance of vegetables used in the dish. It is commonly cut into squares and served warm, not hot, as the dish needs some resting time in order to firm up.	Moussaka is a traditional Greek dish made up of ground lamb (or sometimes beef) and layers of eggplant, all covered in a thick and golden bechamel sauce. It is thought to have originated from the Middle East, as the name is derived from the Turkish musakka, which comes from the Arabic word musaqqa'ah, meaning chilled. It is not a meal eaten every day, but is a special treat that is served on festive occasions. It is similar to lasagna, but without the pasta, and is very healthy due to the vegetables used. It is usually cut into squares and served warm after it has had some time to cool and firm up.

For Future Works

https://www.linkedin.com/in/a-emreusta/

https://www.linkedin.com/in/huseyin-yigit-ulker/

https://www.kaggle.com/datasets/aemreusta/paraphrased-articles-using-gpt3

https://github.com/a-emreusta/success-of-ai-writers