

Language Models for Fact-Checking and Claim Assessment

Othman EL HOUFI
othman.el-houfi@etu.u-cergy.fr

Dimitris KOTZINOS
dimitrios.kotzinios@cyu.fr

Fake news & NLP

Easy to check



This tweet caused \$130 billion value drop in stock market!

This article affected COVID-19 vaccination decisions



Difficult to check

Fake news & NLP

Fake news

false, often sensational, information disseminated under the guise of news reporting.

Collins English Dictionary

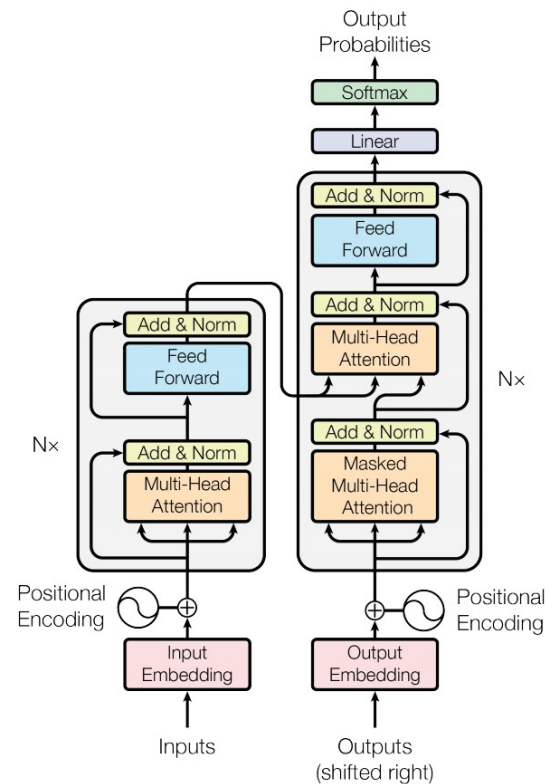
Humans

have been proven irrational and vulnerable when differentiating between real and fake news. Typical accuracy ranges between 55% and 58%.

Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. Fake news: Fundamental theories, detection strategies and challenges.

Language Models

- A new Deep Neural Network Architecture,
 - Aka Transformers
- Based on Attention mechanism,
 - Helps draw connections between any parts of the sequence input i.e Context
- Great for NLP tasks:
 - Translation, Sentiment analysis, Text summarization...
- Powerful, faster, stable than old architectures (RNNs, LSTMs)



Attention is all you need (2017).

Language Models

Pre-trained LMs comparison							
	BERT October 11, 2018	XLNet July 16, 2019	RoBERTa July 26, 2019	ALBERT September 26, 2019	DistilBERT October 2, 2019	BigBird July 28, 2020	ConvBERT August 6, 2020
Parameters	Base: 110M Large: 340M	Base: 110M Large: 340M	Base: 125M Large: 355M	Base: 12M Large: 18M	Base: 66M	Base: 110M	Base: 96M
Layers / Hidden dimensions / Self-attention heads	Base: 12 / 768 / 12 Large: 24 / 1024 / 16	Base: 12 / 768 / 12 Large: 24 / 1024 / 16	Base: 12 / 768 / 12 Large: 24 / 1024 / 16	Base: 12 / 768 / 12 Large: 24 / 1024 / 16	Base: 6 / 768 / 12	Base: 12 / 768 / 12	Base: 12 / 768 / 12
Pre-trained data	BooksCorpus + English Wikipedia ~ 16 GB	Base: 16 GB BERT data Large: 113 GB of textual data (16 GB BERT data + 97 GB additional)	BERT + CCNews + OpenWebText + Stories ~ 160 GB	BooksCorpus + English Wikipedia ~ 16 GB	BooksCorpus + English Wikipedia ~ 16 GB	BERT + CCNews + Stories ~ 97 GB	OpenWebText ~ 32 GB
Method	Bidirectional Transformer, MLM & NSP	Bidirectional Transformer with Permutation based modeling	BERT without NSP, Using Dynamic Masking	BERT with reduced parameters & SOP (not NSP)	BERT Distillation	Block Sparse Attention instead of normal attention	Improves BERT by using a span level dynamic convolution

- All Transformers
- Subtle changes in the architecture and training datasets

Datasets

FEVER

185K annotated claims
3-labels

Extracted from [Wikipedia](#) and subsequently verified without knowledge of the sentence they were derived from.

Liar

12.8K annotated claims
6-labels

Manually labeled short statements that were collected in various contexts from [PolitiFact.com](#).

MultiFC

35K annotated claims
>30-labels reduced to 5-labels

Collected from [26 fact checking websites](#) in English, paired with textual sources and rich metadata, and labeled for veracity by [human expert journalists](#).

COVID-19

6K annotated claims
2-labels

Rumors about COVID-19 extracted from various [social-media platforms](#) such as Twitter, Facebook, Instagram, etc.

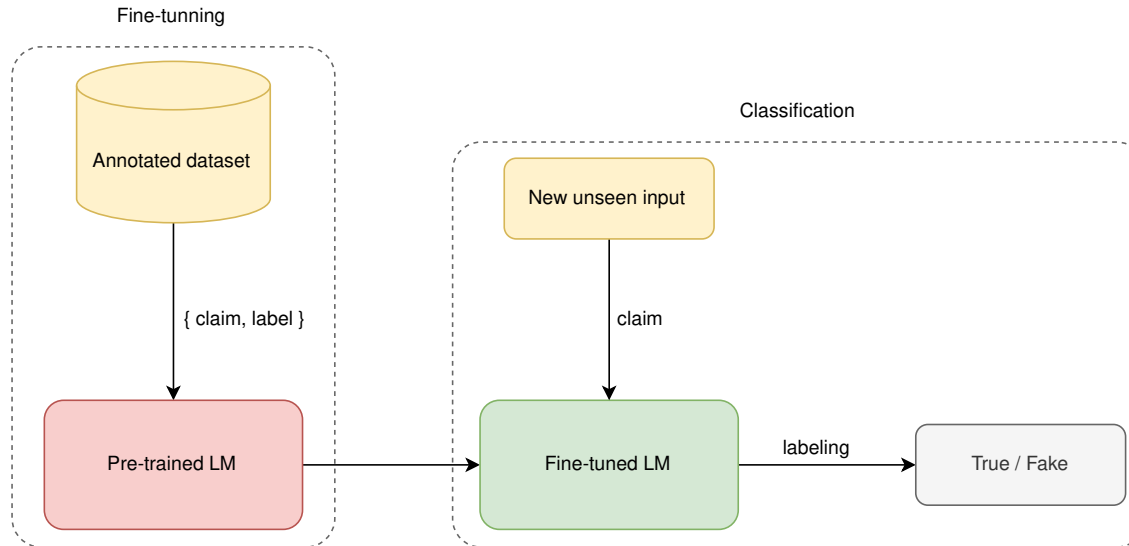
ANTi-Vax

15K annotated claims
2-labels

[Tweets](#) that were annotated as misinformation or general COVID-19 vaccine tweets using reliable sources and validated by [medical experts](#).

General Architecture

- Pre-process each dataset:
 - FEVER, Liar, MultiFC, COVID-19, ANTi-Vax
- Fine-tune a set of LMs:
 - BERT, RoBerta, ALBERT, XLNET, DistilBERT, BigBird, ConvBERT
- Deploy the best LM to assess the validity of new input claims



Results & Discussion

Dataset	Metric	2-labels	3-labels	5-labels	6-labels
FEVER	accuracy	0.81	0.64	-	
	macro f1	0.81	0.63	-	
MultiFC	accuracy	0.72	-	0.50	-
	macro f1	0.64	-	0.40	-
Liar	accuracy	0.69	-		0.31
	macro f1	0.61	-		0.30
COVID-19	accuracy	0.98	-		
	macro f1	0.98	-		
ANTi-Vax	accuracy	0.99	-		
	macro f1	0.99	-		

Conclusion & Perspective

- LMs have a great potential to solve different NLP problems,
- LMs for fact-checking are good but not the best,
 - Does not beat state-of-art traditional models.
- Different paths can be explored
 - We still have much to learn about LMs.
- In the future we plan to exploit the structural features of Complex Networks in combination with LMs.

Bibliography

Datasets

FEVER

185K annotated claims
3-labels

The Apple Store first opened in 2001.

SUPPORTS

Adventure Time won an Oscar.

REFUTES

Yamaha Corporation produces hardware.

NEI

Liar

12.8K annotated claims
6-labels

FIFA pressured Brazil into passing a so-called Budweiser bill, allowing beer sales in soccer stadiums.

TRUE

Says Barack Obama founded ISIS. I would say the co-founder would be crooked Hillary Clinton.

FALSE

Sixty-two percent of all personal bankruptcies are caused by medical problems.

HALF-TRUE

MultiFC

35K annotated claims
>30-labels reduced to 5-labels

The government does not need a warrant to read your old emails.

TRUE

About 99% of rape allegations are fabricated.

FALSE

Husbands rarely beat up their wives. Single women get beaten up more.

IN-BETWEEN

COVID-19

6K annotated claims
2-labels

Holding your breath can let you test whether you may have COVID-19.

FAKE

WHO: We recommend systemic corticosteroids for the treatment of patients with severe and critical #COVID_19 which could be lifesaving. <https://t.co/R4HNTnEEwD>

REAL

ANTI-Vax

15K annotated claims
2-labels

Just got my appointment to be vaccinated and I'm extremely nervous and slightly excited all in one. #vaccine

NOT_MISSINFO

Disturbing and Mysterious #Death of 18yo #Camilla after #COVID #Vaccine. #AstraZeneca's Jabs stopped in #Italy for Young #People!#Experimental <https://t.co/Ssz9kKkFDu>

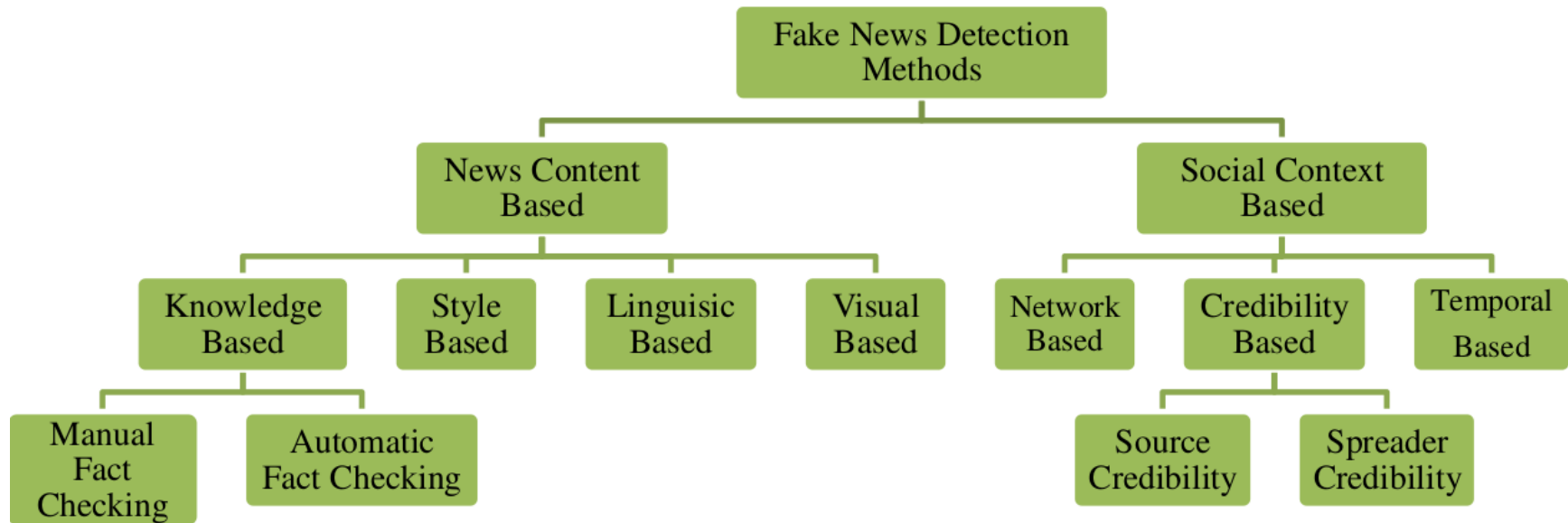
MISSINFO

Fake news & NLP

Automatic fake news detection is a practical NLP problem useful to all online content providers.

- Reduce the human time and effort to detect fake news,
 - Can sweep through huge data streams,
 - Capable of ceasing the spreading much faster.
-
- How can we differentiate fake news from real news?
 - At what level of confidence can we do so?
 - What are the existing methods that solves this problem?

Related work



Related work

Knowledge-based Fake News Detection

aims to assess news authenticity by comparing the knowledge extracted from to-be verified news content with known facts, also called fact-checking.

Anton Chernyavskiy, Dmitry Ilvovsky, and Preslav Nakov. Whatthewikifact: Fact-checking claims against wikipedia.

Style-based Fake News Detection

focuses on the style of writing, i.e. the form of a text rather than its meaning.

P. Przybyla. Capturing the style of fake news. In Proceedings of the AAAI Conference on Artificial Intelligence.

Related work

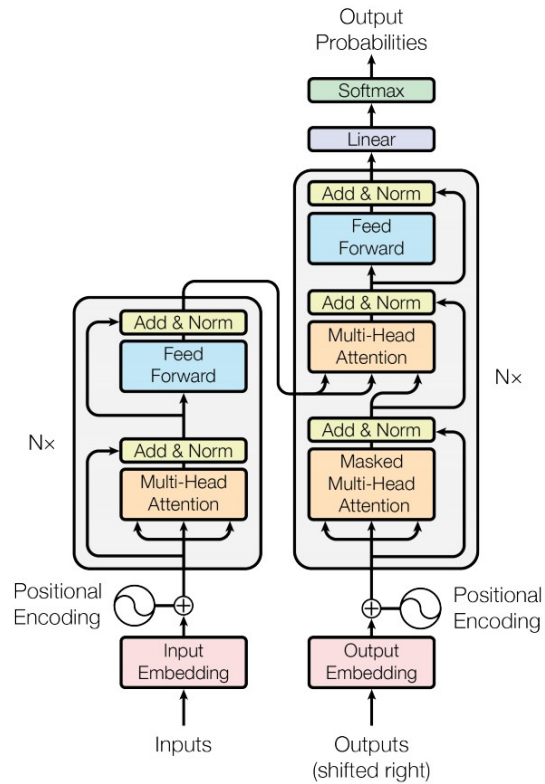
Language Model based Fact-Checking

a new approach that relies on fine-tuning state-of-art LMs like BERT that were pre-trained on Wikipedia's articles in order to solve the claim classification problem.

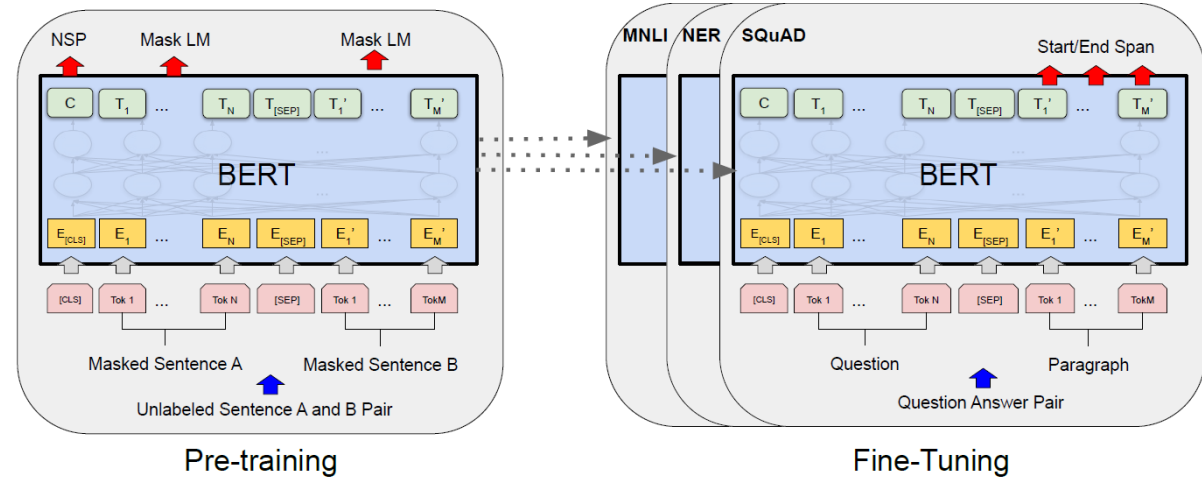
Nayeon Lee, Belinda Z Li, Sinong Wang, Wen-tau Yih, Hao Ma, and Madian Khabsa. Language models as fact checkers?

- What are Language Models?
- How can they be employed as fact-checkers?

Language Models



Attention is all you need.



Bert: Pre-training of deep bidirectional transformers for language understanding.

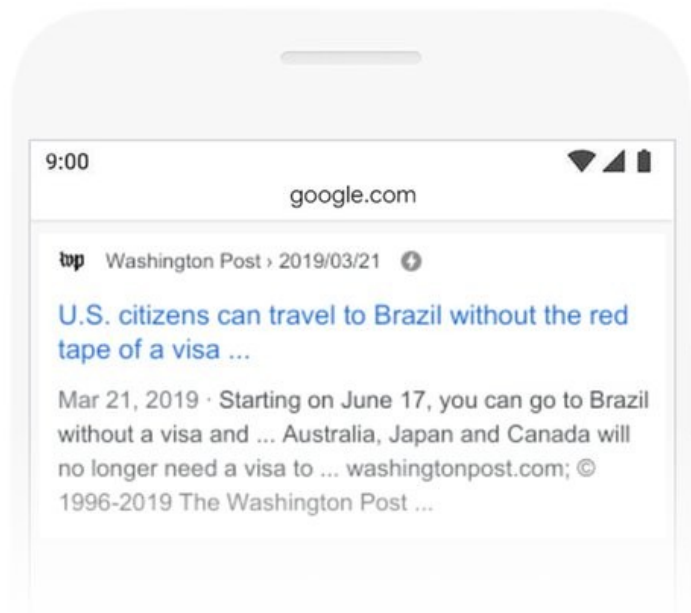
Language Models

BERT

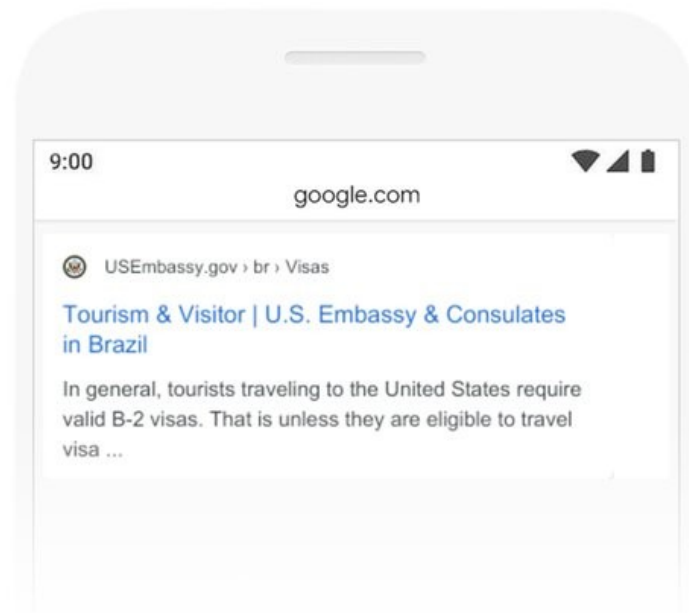


2019 brazil traveler to usa need a visa

BEFORE



AFTER

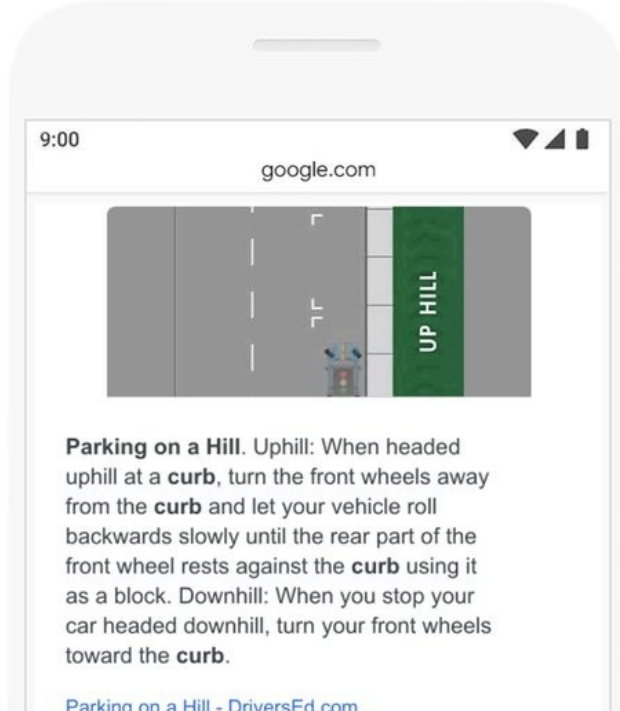


Language Models

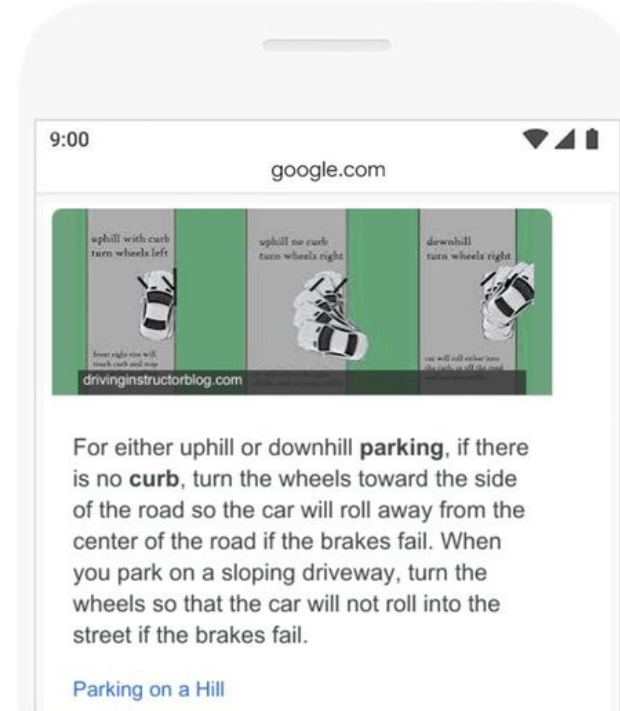
BERT

🔍 parking on a hill with no curb

BEFORE



AFTER



Results & Discussion

Fine-tuned model	Label	prec	recall	f1	accuracy	macro prec	macro recall	macro f1
<i>BERT-base-uncased</i>	SUPPORTS	0.55	0.78	0.64	0.62	0.63	0.62	0.61
	REFUTES	0.75	0.59	0.66				
	NEI	0.61	0.47	0.53				
<i>ALBERT-base-v2</i>	SUPPORTS	0.46	0.81	0.59	0.53	0.58	0.53	0.52
	REFUTES	0.77	0.46	0.58				
	NEI	0.50	0.33	0.40				
<i>DistilBERT-base-uncased</i>	SUPPORTS	0.54	0.78	0.64	0.61	0.63	0.61	0.61
	REFUTES	0.75	0.58	0.65				
	NEI	0.60	0.47	0.53				
<i>RoBERTa-base</i>	SUPPORTS	0.54	0.81	0.65	0.62	0.64	0.62	0.61
	REFUTES	0.75	0.59	0.66				
	NEI	0.63	0.45	0.53				
<i>BigBird-RoBERTa-base</i>	SUPPORTS	0.53	0.81	0.64	0.61	0.64	0.61	0.60
	REFUTES	0.75	0.58	0.66				
	NEI	0.63	0.44	0.52				
<i>XLNET-base-cased</i>	SUPPORTS	0.53	0.81	0.64	0.61	0.63	0.61	0.60
	REFUTES	0.74	0.59	0.65				
	NEI	0.63	0.43	0.51				
Related work	Label	prec	recall	f1	accuracy	macro prec	macro recall	macro f1
<i>BERT-large</i> [7]	SUPPORTS	0.54	0.67	0.59	0.57	0.57	0.57	0.57
	REFUTES	0.62	0.55	0.58				
	NEI	0.57	0.49	0.53				
<i>FEVER Baseline</i> [19]	-	-	-	-	0.49	-	-	-
<i>Ohio State University</i> [19]	-	-	-	-	0.50	-	-	-
<i>Columbia NLP</i> [19]	-	-	-	-	0.58	-	-	-
<i>Papelo</i> [19]	-	-	-	-	0.61	-	-	-
<i>UNC-NLP</i> [19]	-	-	-	-	0.68	-	-	-
<i>DREAM</i> [20]	-	-	-	-	0.77	-	-	-

Proposed Method

Language Models for classification

Comparison	BERT October 11, 2018	RoBERTa July 26, 2019	DistilBERT October 2, 2019	ALBERT September 26, 2019
Parameters	Base: 110M Large: 340M	Base: 125 Large: 355	Base: 66	Base: 12M Large: 18M
Layers / Hidden Dimensions / Self-Attention Heads	Base: 12 / 768 / 12 Large: 24 / 1024 / 16	Base: 12 / 768 / 12 Large: 24 / 1024 / 16	Base: 6 / 768 / 12	Base: 12 / 768 / 12 Large: 24 / 1024 / 16
Training Time	Base: 8 x V100 x 12d Large: 280 x V100 x 1d	1024 x V100 x 1 day (4-5x more than BERT)	Base: 8 x V100 x 3.5d (4 times less than BERT)	[not given] Large: 1.7x faster
Performance	Outperforming SOTA in Oct 2018	88.5 on GLUE	97% of BERT-base's performance on GLUE	89.4 on GLUE
Pre-Training Data	BooksCorpus + English Wikipedia = 16 GB	BERT + CCNews + OpenWebText + Stories = 160 GB	BooksCorpus + English Wikipedia = 16 GB	BooksCorpus + English Wikipedia = 16 GB
Method	Bidirectional Transformer, MLM & NSP	BERT without NSP, Using Dynamic Masking	BERT Distillation	BERT with reduced parameters & SOP (not NSP)

https://humboldt-wi.github.io/blog/research/information_systems_1920/uncertainty_identification_transformers/