BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Anton Karazeev

592 group

MIPT, 2019

https://arxiv.org/abs/1810.04805

Outline

- Introduction to NLP

- BERT Model

- Conclusion

https://arxiv.org/abs/1810.04805

Introduction to NLP

http://jalammar.github.io/illustrated-bert/

https://arxiv.org/abs/1810.04805
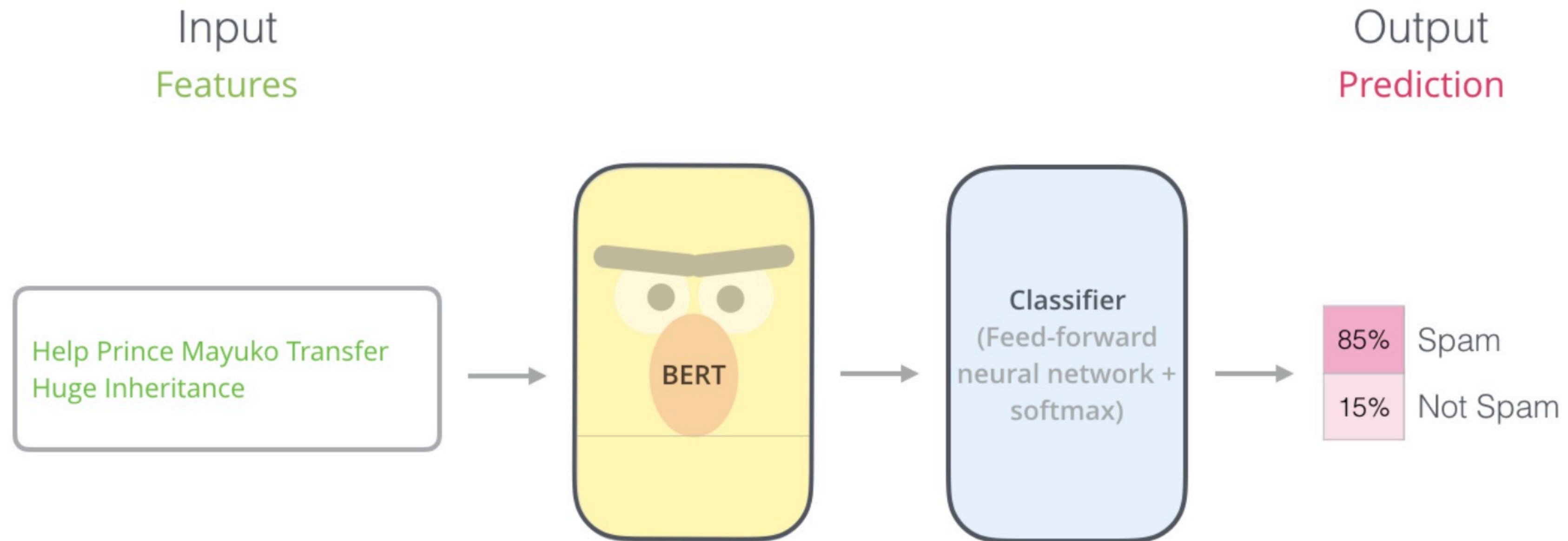
Introduction to NLP

Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached…	Not Spam

http://jalammar.github.io/illustrated-bert/

https://arxiv.org/abs/1810.04805

BERT Model

http://jalammar.github.io/illustrated-bert/

https://arxiv.org/abs/1810.04805
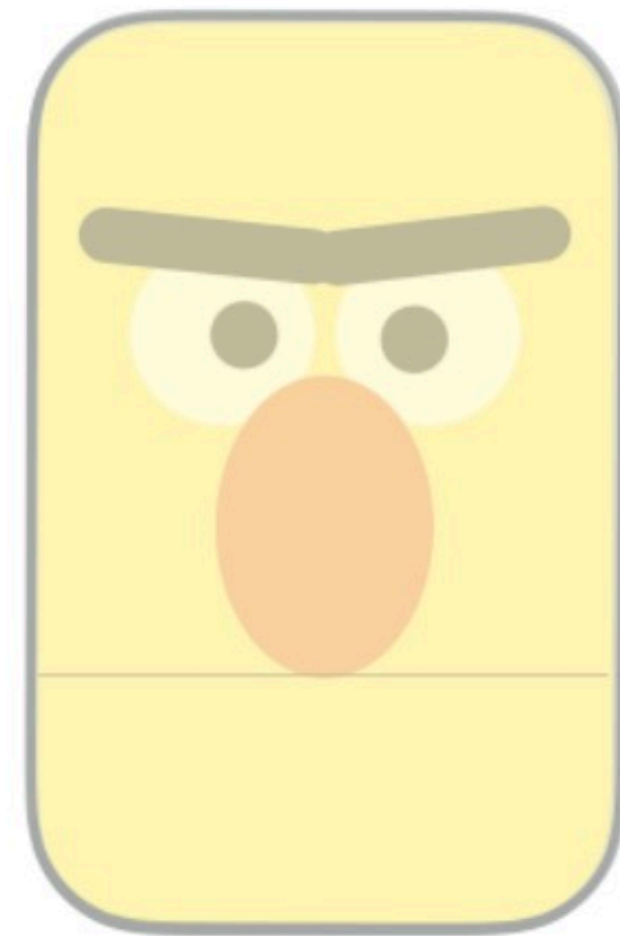
BERT Model

BERT BASE
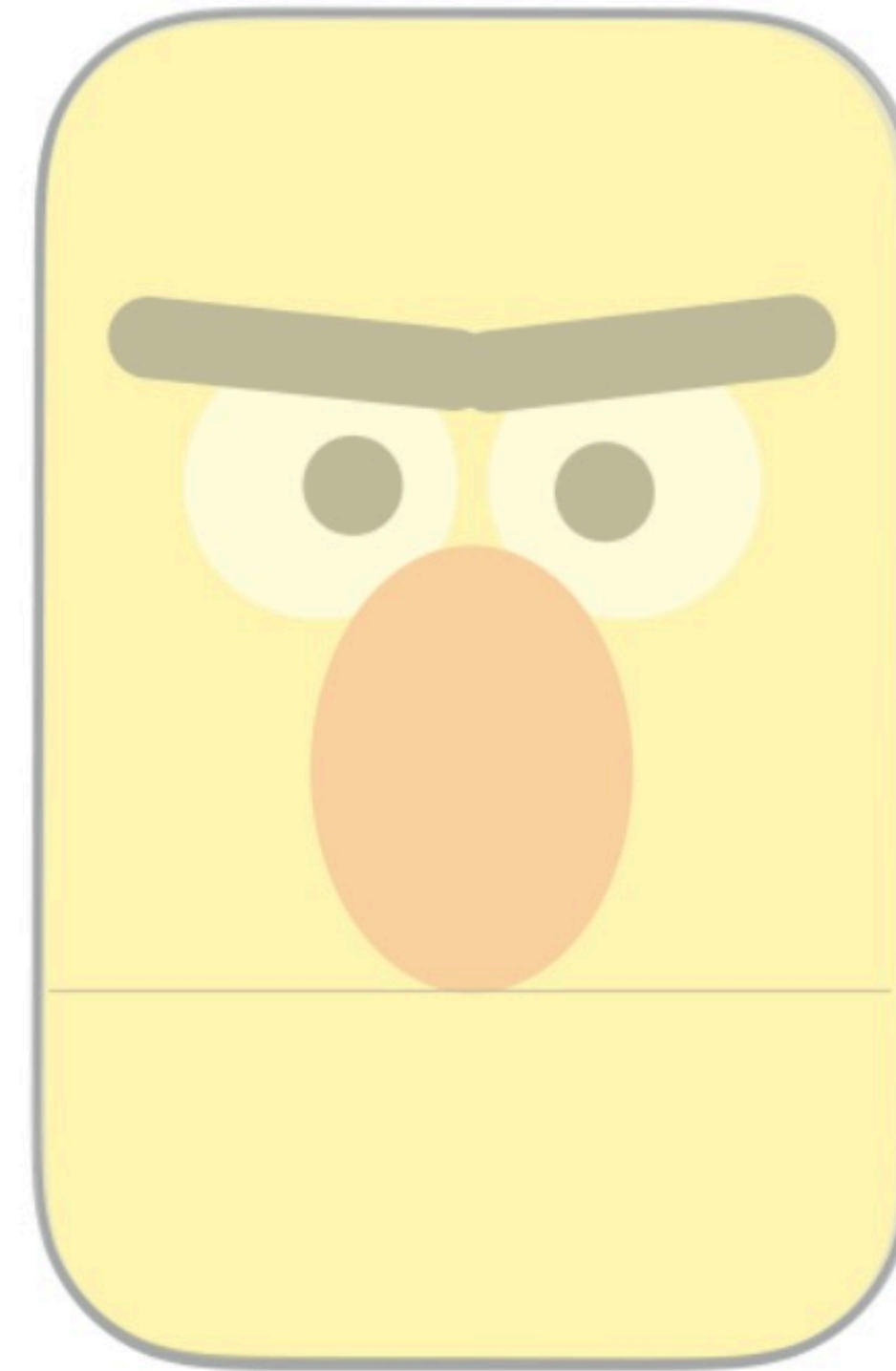
BERT LARGE
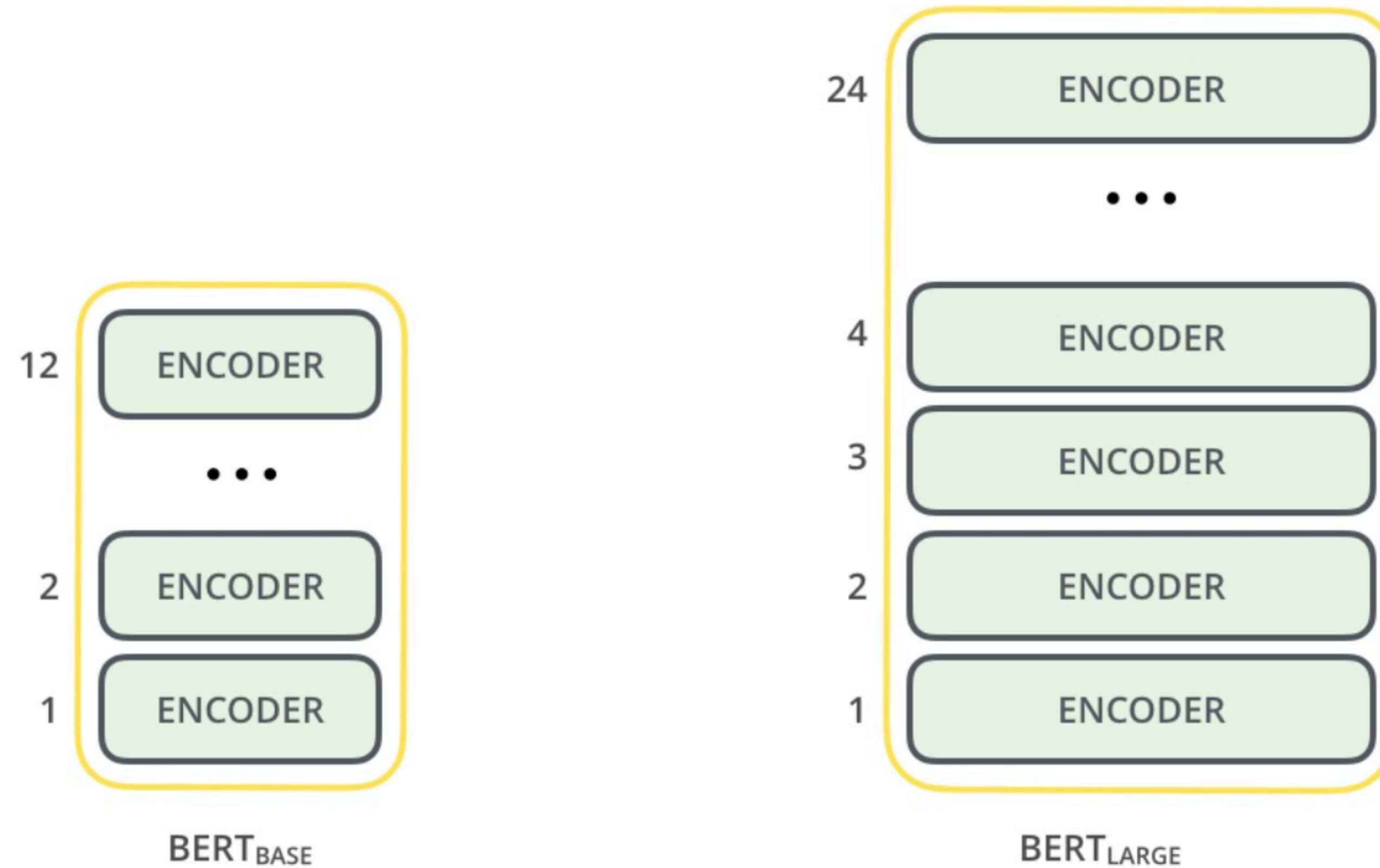
BERT Model

Use the output of the masked word's position to predict the masked word

Possible classes: All English words

0.1%	Aardvark
...	...
10%	Improvisation
...	...
0%	Zyzzyva

FFNN + Softmax

1 2 3 4 5 6 7 8 ... 512
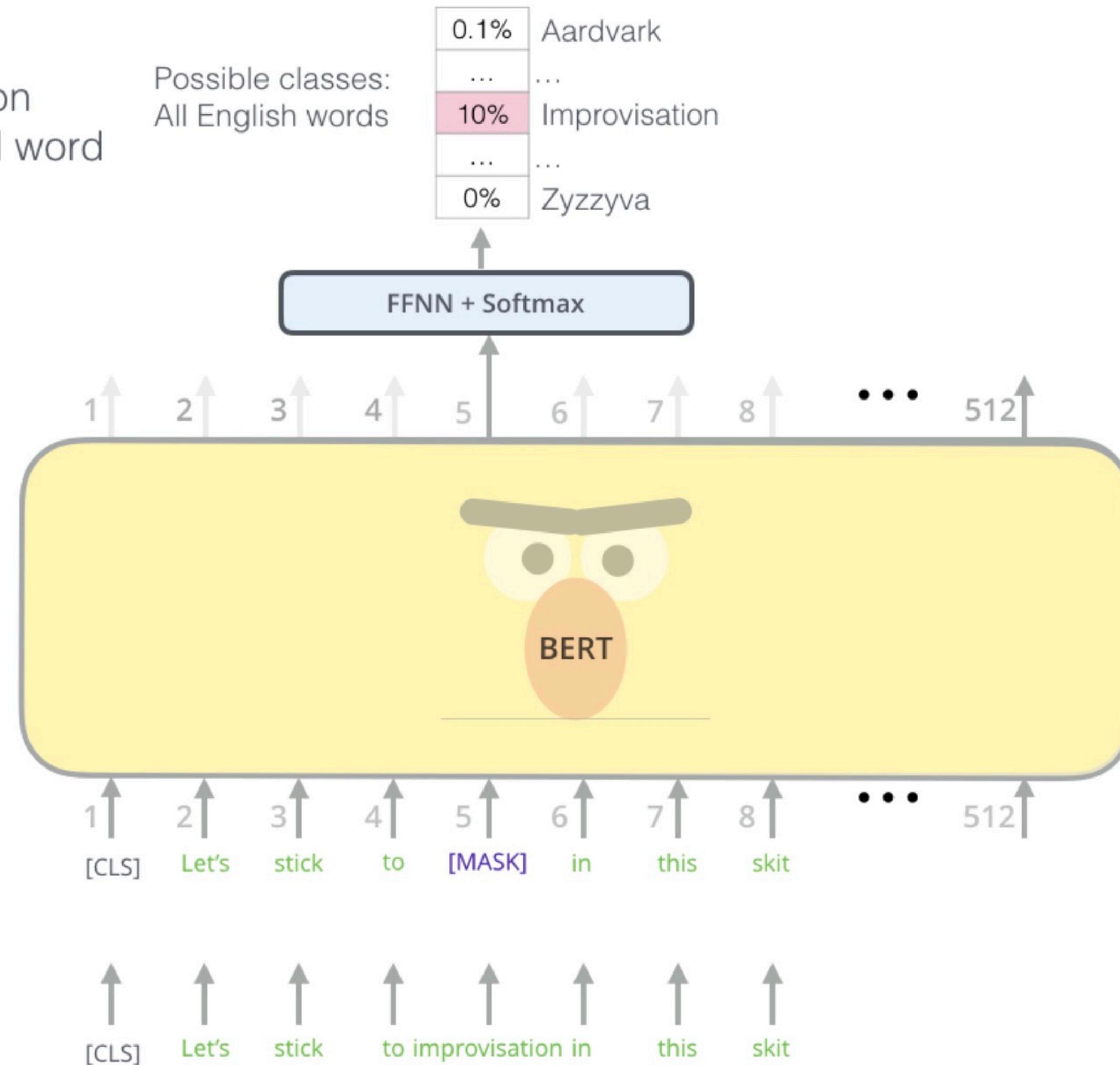
BERT

Randomly mask 15% of tokens

1 2 3 4 5 6 7 8 ... 512

[CLS] Let's stick to [MASK] in this skit

Input

[CLS] Let's stick to improvisation in this skit

http://jalammar.github.io/illustrated-bert/

7

https://arxiv.org/abs/1810.04805

BERT Model

System	MNLI-(m/mm)	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE	**Average**
	392k	363k	108k	67k	8.5k	5.7k	3.5k	2.5k	-
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.9	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
BERT$_{BASE}$	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
BERT$_{LARGE}$	**86.7/85.9**	**72.1**	**91.1**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**81.9**

Table 1: GLUE Test results, scored by the GLUE evaluation server. The number below each task denotes the number of training examples. The "Average" column is slightly different than the official GLUE score, since we exclude the problematic WNLI set. OpenAI GPT = (L=12, H=768, A=12); BERT$_{BASE}$ = (L=12, H=768, A=12); BERT$_{LARGE}$ = (L=24, H=1024, A=16). BERT and OpenAI GPT are single-model, single task. All results obtained from https://gluebenchmark.com/leaderboard and https://blog.openai.com/language-unsupervised/.

https://arxiv.org/abs/1810.04805

Summary

- The Illustrated BERT, ELMo, and co.: http://jalammar.github.io/illustrated-bert/

- About BERT in Google AI Blog: https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

- Bert image's source: https://twitter.com/bertsesame

https://arxiv.org/abs/1810.04805

Conclusion

- Proposed and trained model is a **huge improvement** in the field of Natural Language Processing

- It makes the future in which **people and robots** can freely interact with each other much closer to nowadays

https://arxiv.org/abs/1810.04805