Paper Reading Dec 10th

2 papers about LMs
One focus on Interpretability(夹带私货→_→)
One focus on injecting knowledge into LMs

Perturbed Masking: Parameter-free Probing for Analyzing and Interpreting BERT

Zhiyong Wu, *Yun Chen, Ben Kao, Qun Liu*

HUAWEI | NOAH'S ARK LAB

Background

An emerging body of NLP work asks

“Does my neural network implicitly learns **Y** from plain text? ”

Background

An emerging body of NLP work asks

"Does my neural network implicitly learns **Y** from plain text? "

Pretrained LM
e.g., ELMo, BERT

parts-of-speech
syntax
semantics

Background

An emerging body of NLP work asks

"Does my neural network implicitly learns **Y** from plain text? "

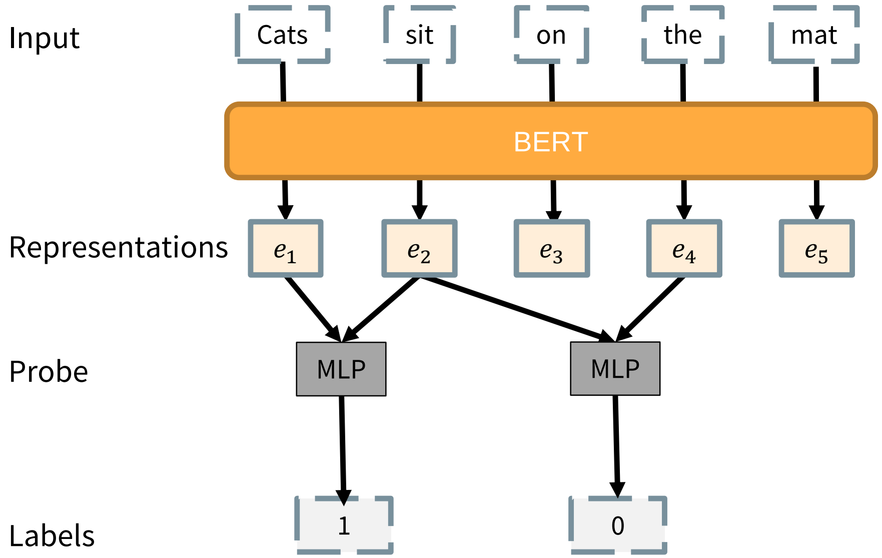
Pretrained LM
e.g., ELMo, BERT

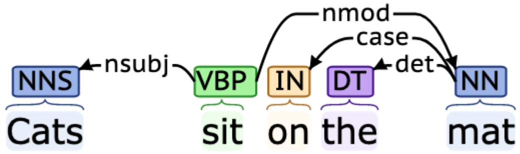
parts-of-speech
syntax
semantics

Probing:

<u>*supervised*</u> analysis of representations

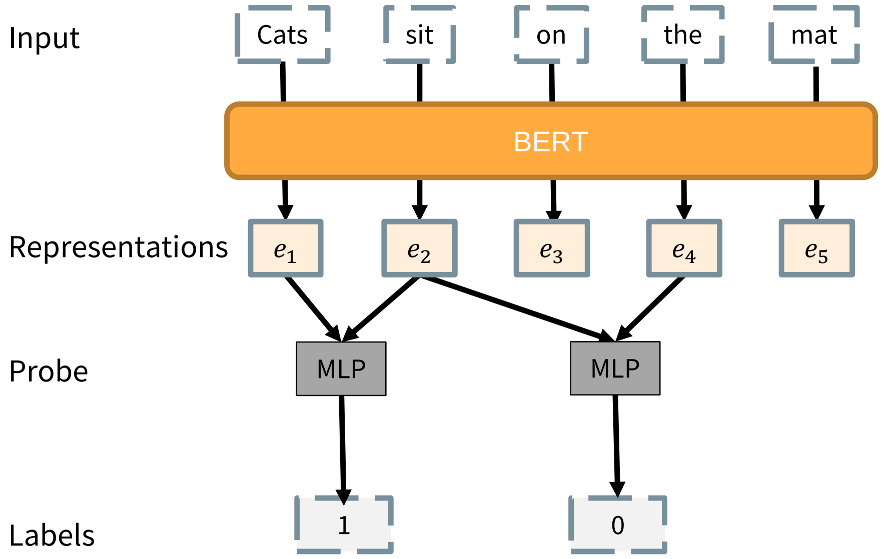
Probing: Does BERT capture syntax?

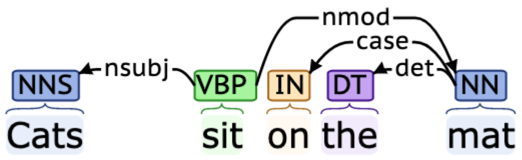




Probing: Does BERT capture syntax?

1.Select task
dependency edge prediction

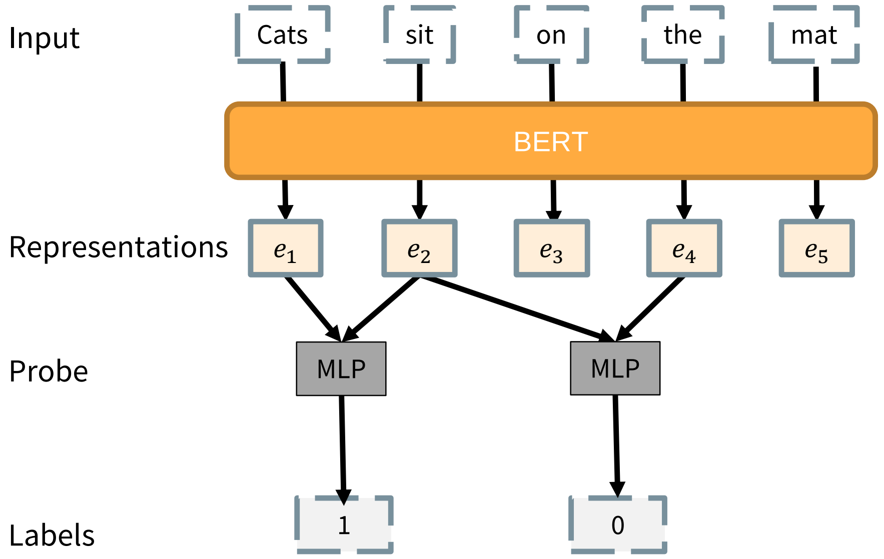


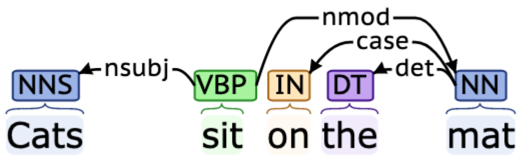


Probing: Does BERT capture syntax?

1. **Select** task
dependency edge prediction

2. **Extract** representations
from your model



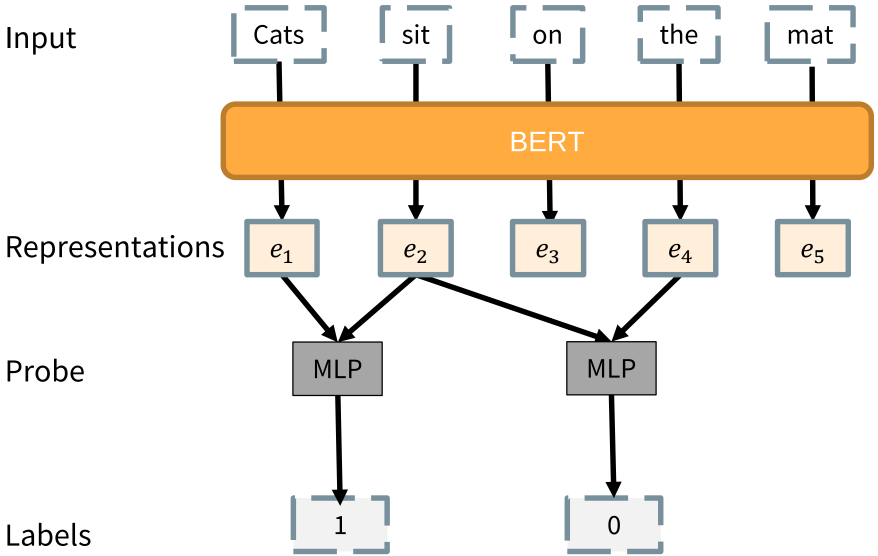


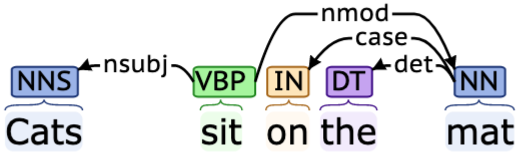
Probing: Does BERT capture syntax?

1.Select task
dependency edge prediction

2.Extract representations
from your model

3.Train a network
representations >> task





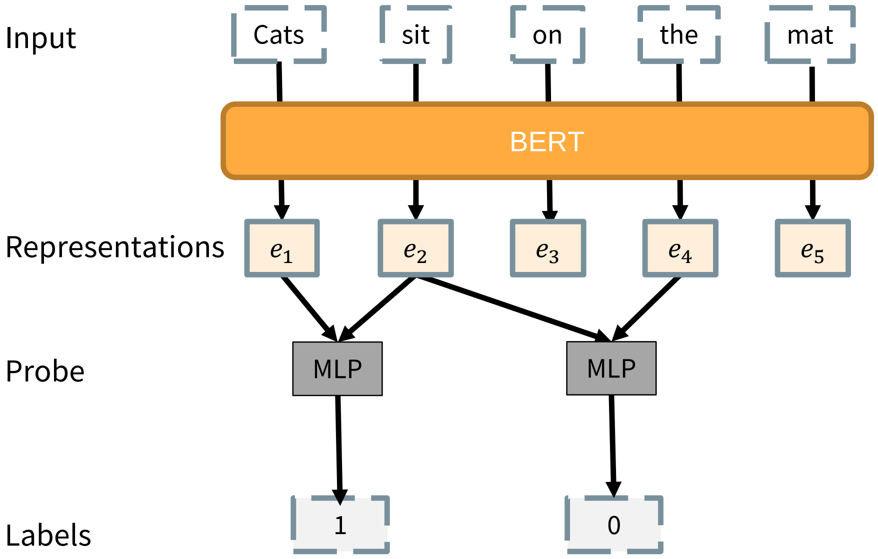
Probing: Does BERT capture syntax?

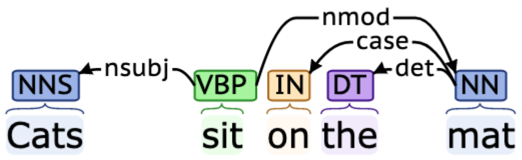
1. **Select** task
dependency edge prediction

2. **Extract** representations
from your model

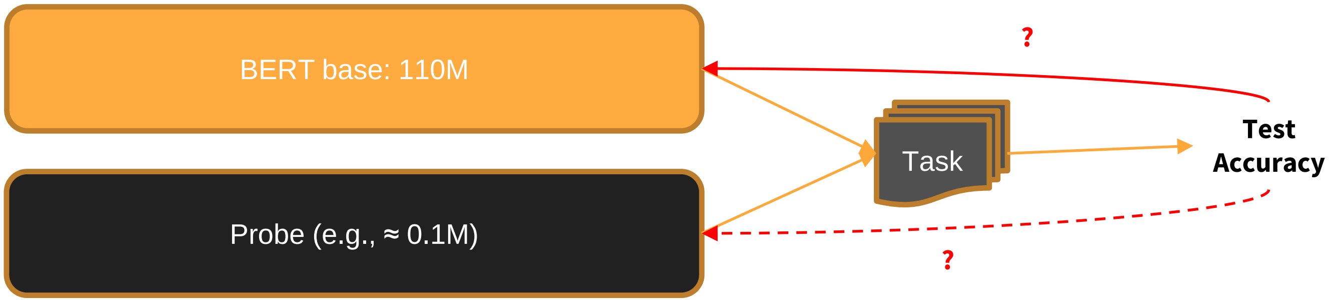
3. **Train** a network
representations >> task

4. **Interpret** probing accuracy

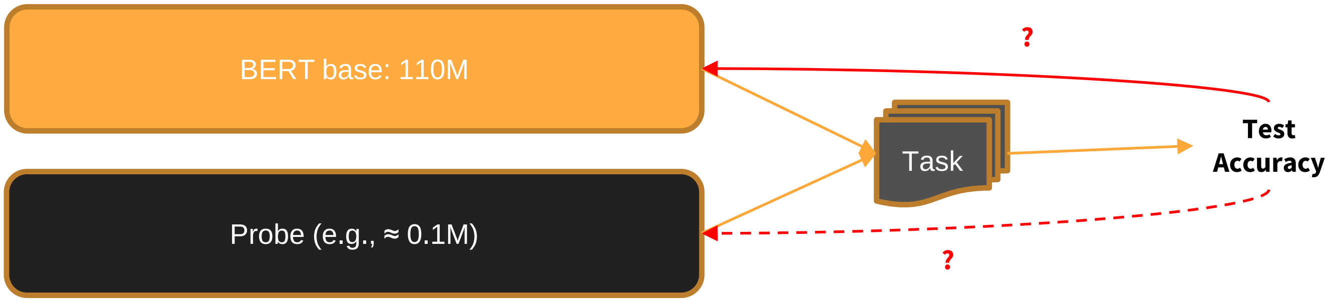




The probe confounder problem

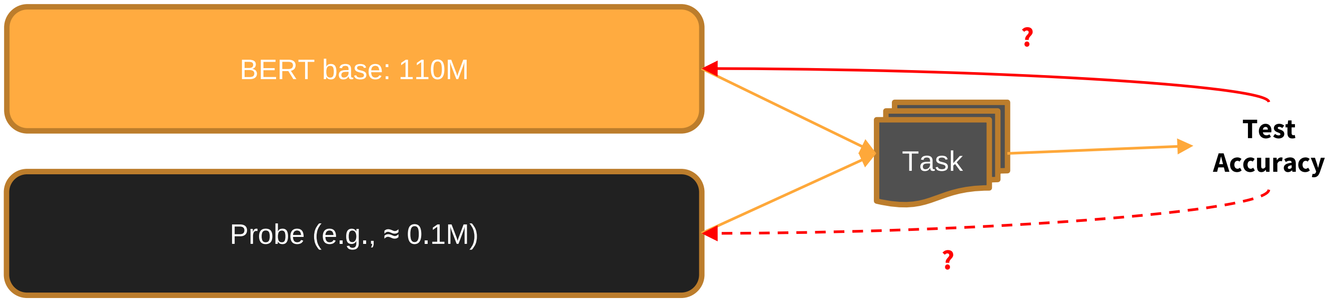


The probe confounder problem



Shall we give credit to the **representation**? and/or the **probe**?

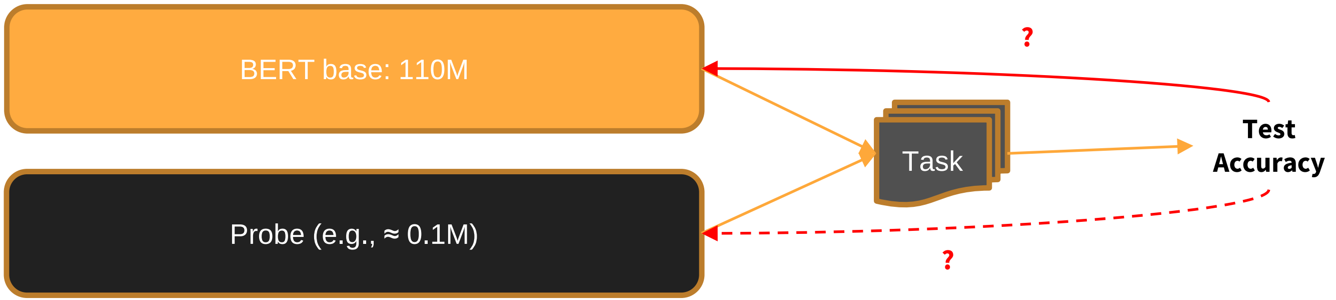
The probe confounder problem



Shall we give credit to the **representation**? and/or the **probe**?

Fix it: control tasks (Hewitt and Liang 2019) MDL(Elena et al., 2020)

The probe confounder problem

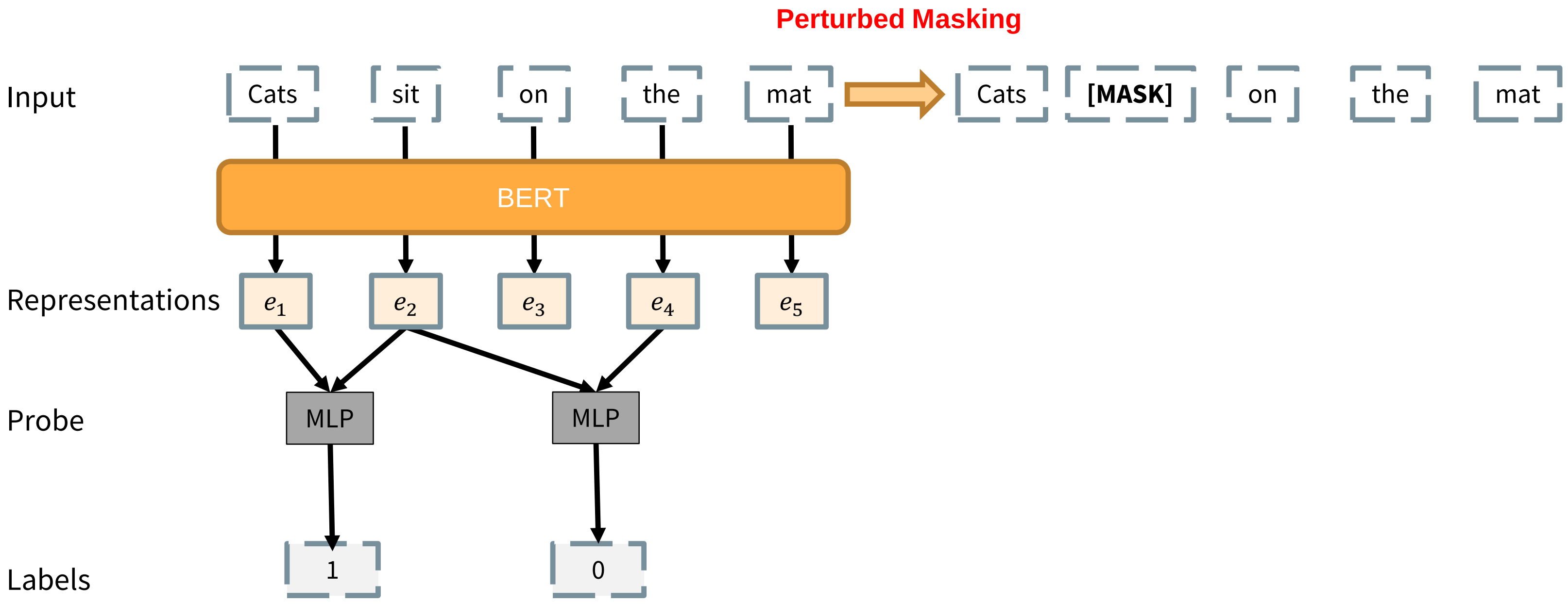


Shall we give credit to the **representation**? and/or the **probe**?

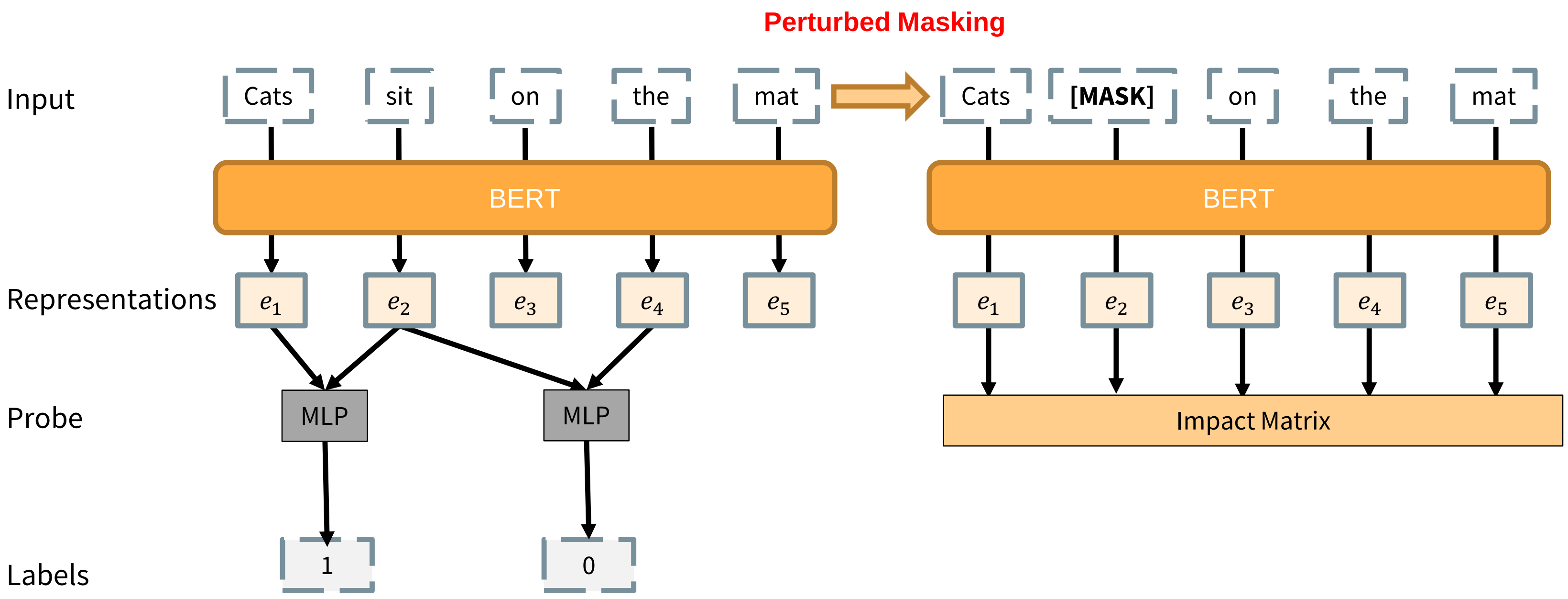
Fix it: control tasks (Hewitt and Liang 2019) MDL(Elena et al., 2020)

This work: **parameter-free(unsupervised) probing**

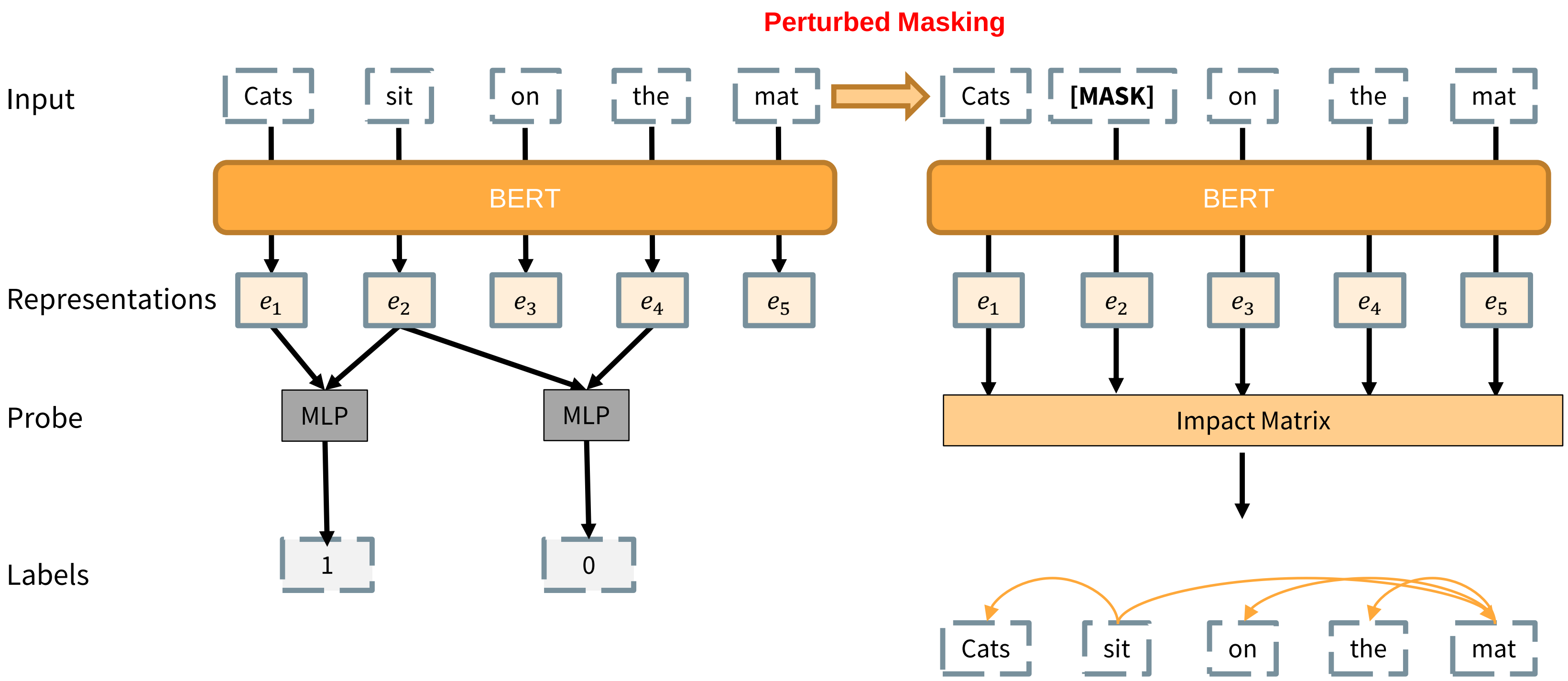
Unsupervised Probing with Perturbed Masking



Unsupervised Probing with Perturbed Masking



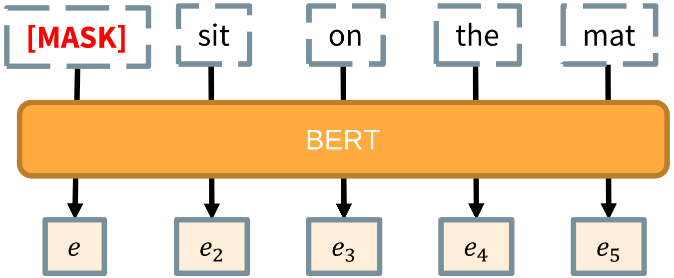
Unsupervised Probing with Perturbed Masking



Perturbed Masking

Example: Calculate impact *sit* has on *Cats:* $f(Cats, sit) = d(e, e')$

$e = E(Cats|S\backslash\{Cats\})$

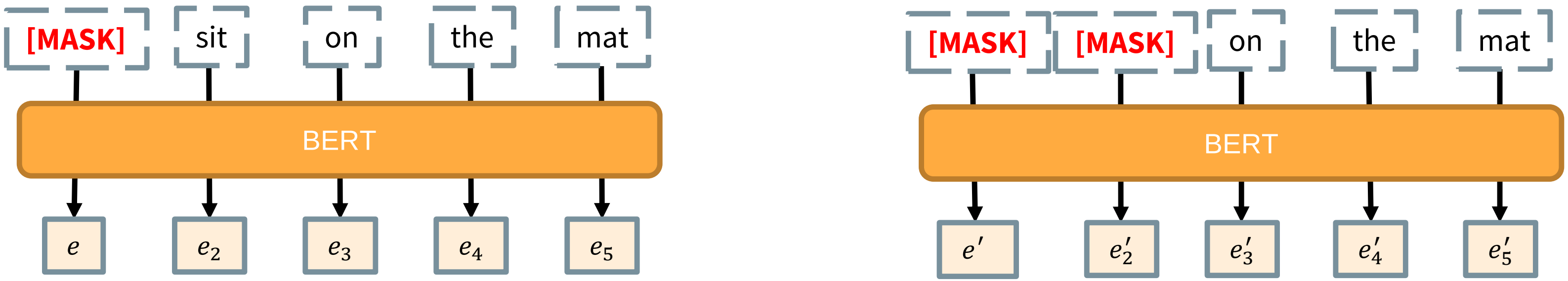


Perturbed Masking

Example: Calculate impact *sit* has on *Cats:* $f(Cats, sit) = d(e, e')$

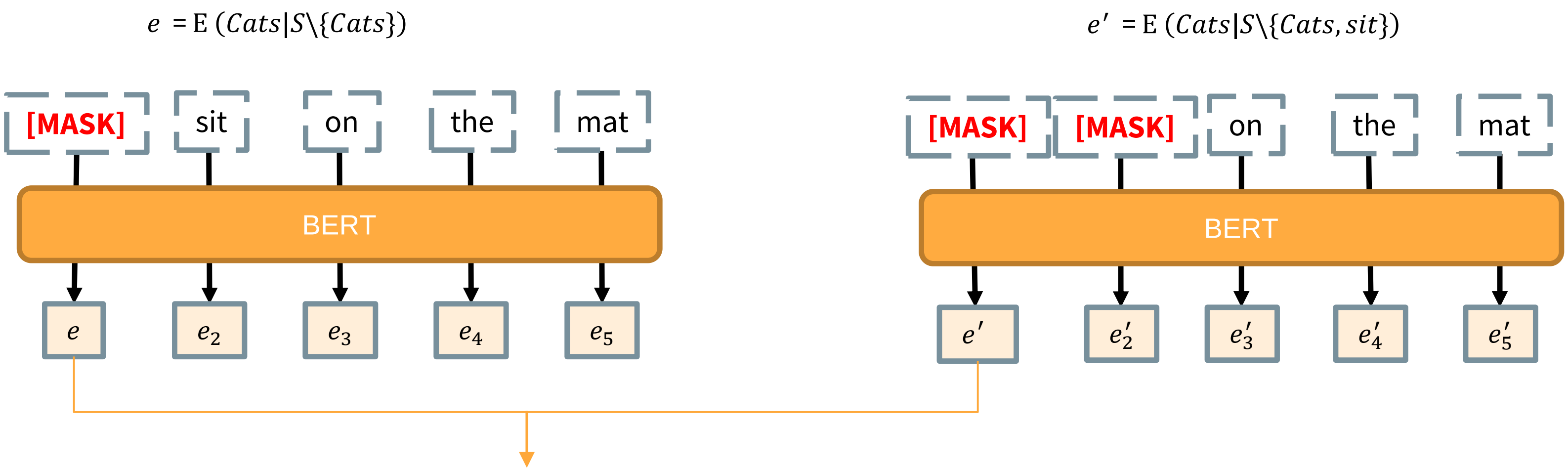
$e = E(Cats|S\backslash\{Cats\})$

$e' = E(Cats|S\backslash\{Cats, sit\})$



Perturbed Masking

Example: Calculate impact *sit* has on *Cats:* $f\,(Cats, sit) = d(e, e')$



$f\,(Cats, sit) = d(e, e')$ = Distance between e and e’

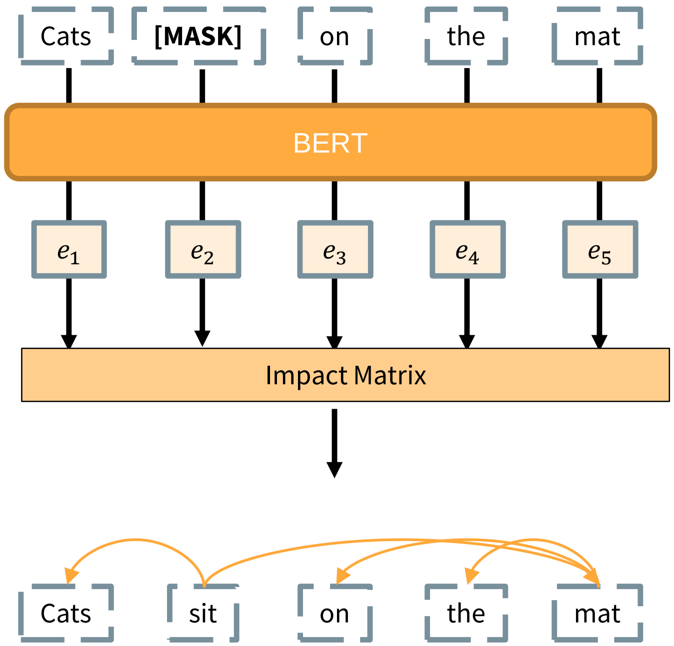
Impact Matrix

	Cats	sit	on	the	mat
Cats	-	$f(Cat, sit)$			
sit	$f(sit, Cats)$	-			
on			-		
the				-	
mat					-

Supervised Probe: learning to map representations to task

Ours: Impact Matrix + task specific algo => task

Unsupervised Probing with Perturbed Masking



1. **Perturb input sentence and extract an impact matrix.**

2. **Use task-specific algorithm to extract task-related knowledge from the impact matrix**

Application 1: Dependency probe

- Using graph-based dependency parsing algorithm to extract dependency trees out of impact matrixes.

Model	Parsing UAS	
	WSJ10-U	PUD
Right-chain	49.5	35.0
Left-chain	20.6	10.7
Random BERT	16.9	10.2
Eisner+Dist	**58.6**	**41.7**
Eisner+Prob	52.7	34.1
CLE+Dist	51.5	33.2

Application 1: Dependency probe

1. Despite its parameter-free nature, our probe corroborates findings from previous studies
2. However, we also observe that the structures induced from BERT only correlate with human-designed syntax weakly

Model	**Parsing UAS**	
	WSJ10-U	**PUD**
Right-chain	49.5	35.0
Left-chain	20.6	10.7
Random BERT	16.9	10.2
Eisner+Dist	**58.6**	**41.7**
Eisner+Prob	52.7	34.1
CLE+Dist	51.5	33.2

Application 1: Dependency probe

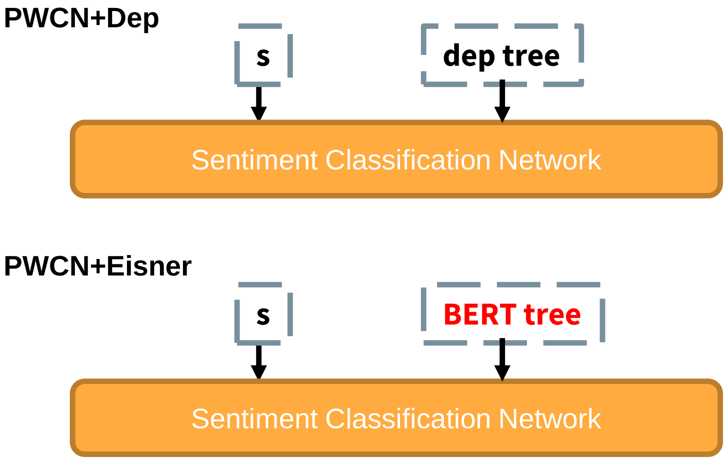
1. Despite its parameter-free nature, our probe corroborates findings from previous studies
2. However, we also observe that the structures induced from BERT only correlate with human-designed syntax weakly

Would BERT learn **better** dependency structures?

Model	**Parsing UAS**	
	WSJ10-U	**PUD**
Right-chain	49.5	35.0
Left-chain	20.6	10.7
Random BERT	16.9	10.2
Eisner+Dist	**58.6**	**41.7**
Eisner+Prob	52.7	34.1
CLE+Dist	51.5	33.2

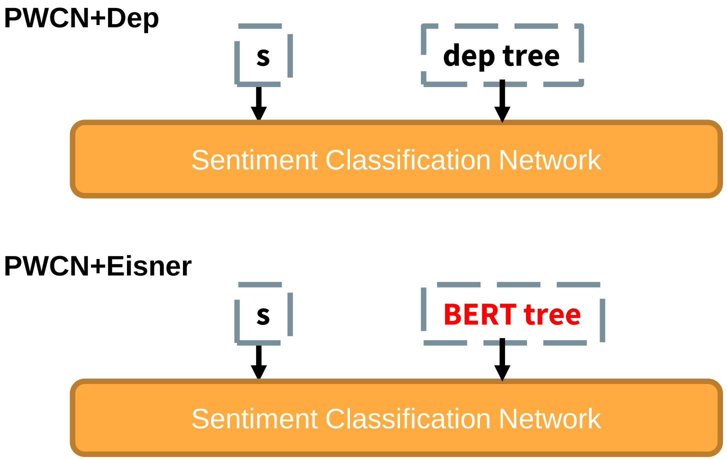
Empirical usefulness of the induced structure

Input sentence: **s**
Parser generated dep tree for s: **dep tree**
BERT generated dep tree for s: **BERT tree**



Empirical usefulness of the induced structure

Input sentence: **s**
Parser generated dep tree for s: **dep tree**
BERT generated dep tree for s: **BERT tree**



Model	Laptop		Restaurant	
	Acc	Macro-F1	Acc	Macro-F1
LSTM	69.63	63.51	77.99	66.91
PWCN				
+Pos	75.23	71.71	81.12	71.81
+Dep	76.08	72.02	80.96	72.21
+Eisner	75.99	72.01	**81.21**	**73.00**
+right-chain	75.64	71.53	81.07	72.51
+left-chain	74.39	70.78	80.82	72.71

Other applications

- Other probes (refer to paper)
- Unsupervised syntactic parsing (Kim et al., 2020; Li et al., 2020)
- Chinese word segmentation
- LM pre-training

UNC NLP

EMNLP 2020
The 2020 Conference on Empirical Methods in Natural Language Processing
16th – 20th November 2020

Vokenization:

Improving Language Understanding with Contextualized, Visual-Grounded Supervision

Hao Tan, Mohit Bansal

UNC Chapel Hill
haotan, mbansal@cs.unc.edu

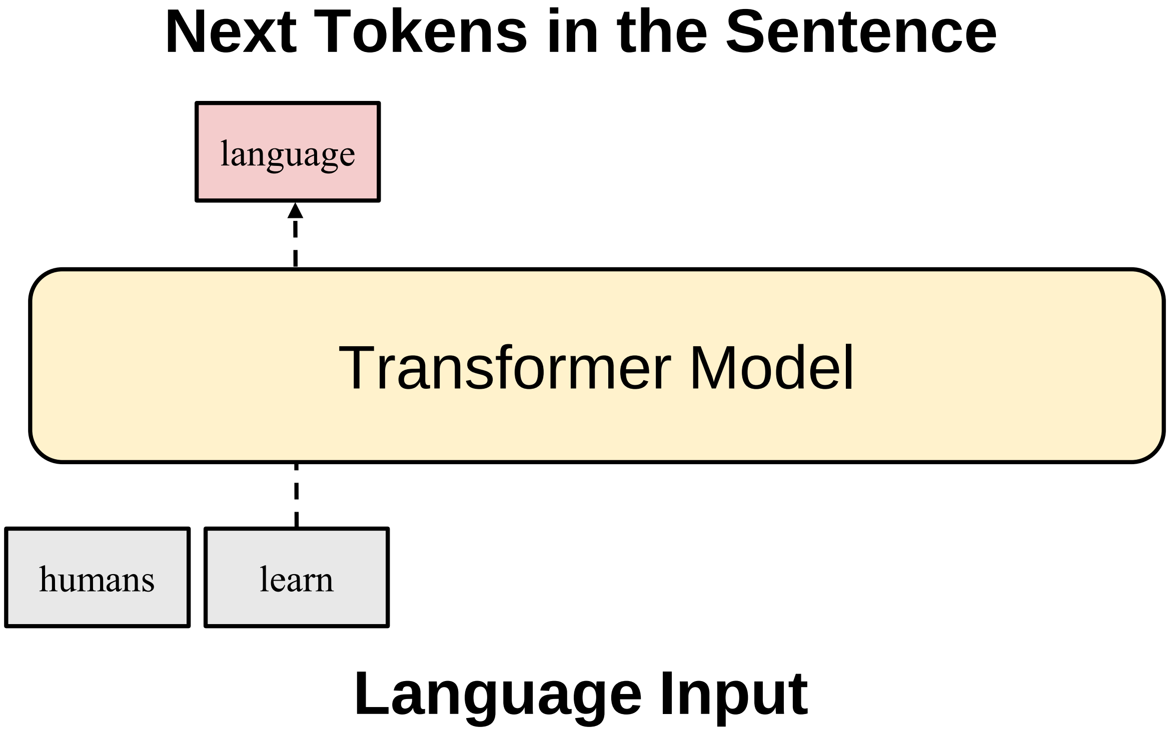
thetimes.co.uk/article/feline-philosophy-by-john-gray-review-what-cats-teach-us-about-the-meaning-of-life-bgzwkqpqm

Visual Supervision to Language

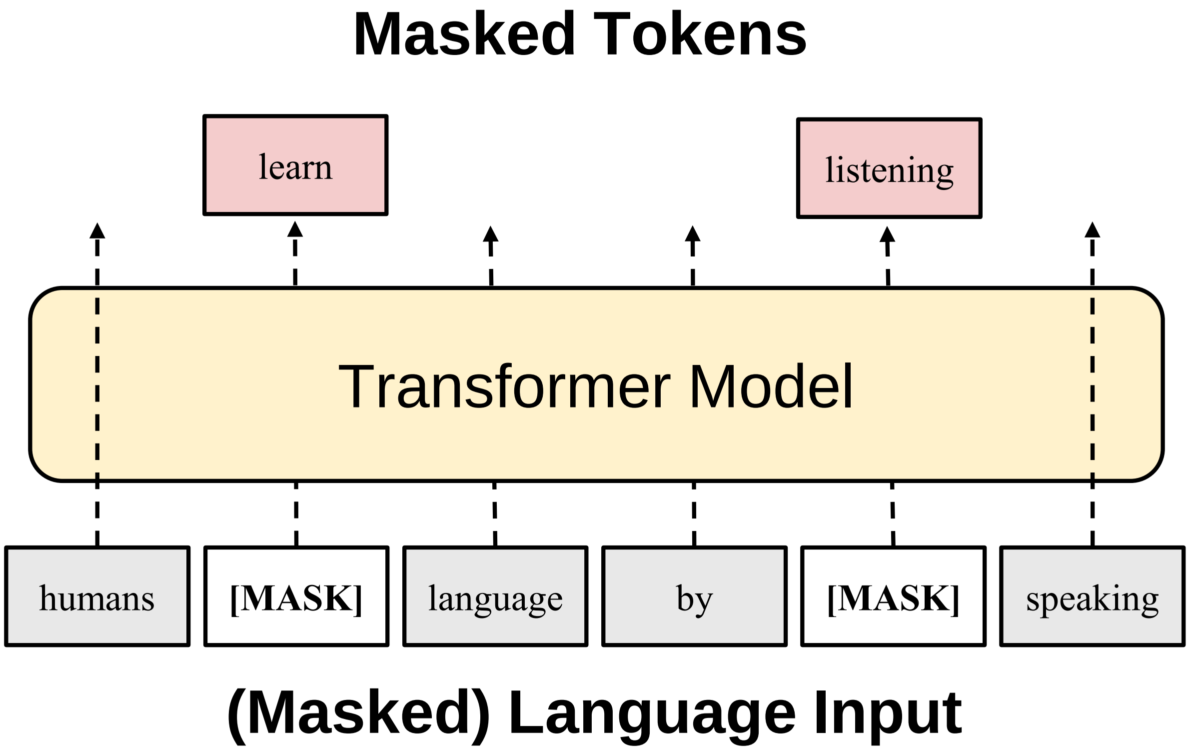
Look, this is a “cat”!

Peters, Matthew E., et al. "Deep contextualized word representations." NAACL 2018
Radford, Alec, et al. "Language models are unsupervised multitask learners."

Causal Language Model (e.g., ELMo, GPT2)

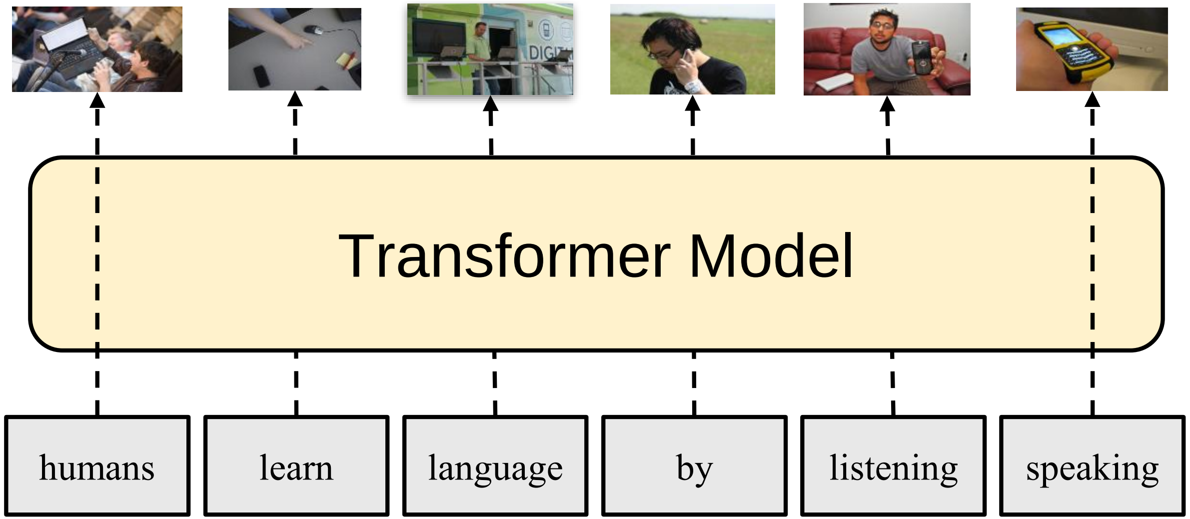


Masked Language Model (e.g., BERT)



Visually-Supervised Language Model

Vokens (Token-Related Images)

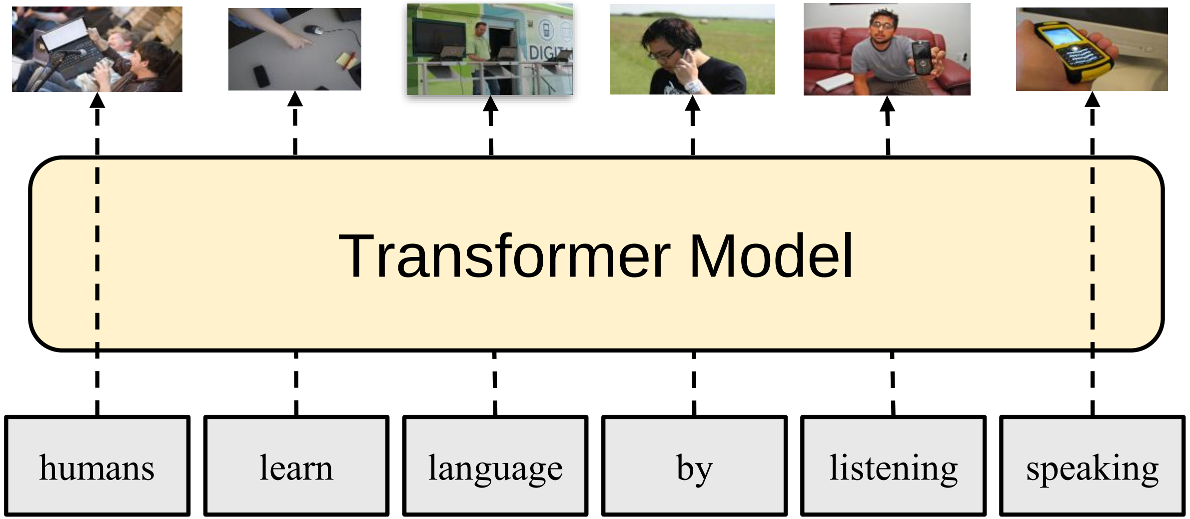


Language Input

Visually-Supervised Language Model

Vokens (Token-Related Images)

How (and where) to get these visual supervision?

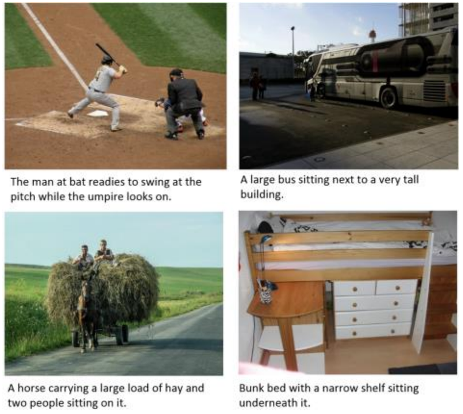


Language Input

Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context.", ECCV 2014

Available Resources and Our Goal

We **have**



MS COCO Captioning

(Visually-) Grounded Language

We **want**

COVID-19 pandemic

From Wikipedia, the free encyclopedia

Several terms redirect here. For other uses, see Coronavirus outbreak (disambiguation) and 2019–2020 outbreak (disambiguation).

The **COVID-19 pandemic**, also known as the **coronavirus pandemic**, is an ongoing pandemic of coronavirus disease 2019 (COVID-19) caused by the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was first identified in December 2019 in Wuhan, China.[3] The then-outbreak was declared a Public Health Emergency of International Concern (PHEIC) in January 2020, and a pandemic in March 2020. As of 22 October 2020, more than 41 million cases have been confirmed, with more than 1.12 million deaths attributed to COVID-19.[5]

COVID-19 spreads most often when people are physically close.[6] It spreads very easily and sustainably through the air, primarily via small droplets or aerosols, as an infected person breathes, coughs, sneezes, sings, or talks.[8][9] It may also be transmitted via contaminated surfaces, although this has not been conclusively demonstrated.[9][10][11] Airborne transmission from aerosol formation is suspected to be the main mode of transmission.[12] It can spread from an infected person for up to two days before they display symptoms, and from people who are asymptomatic.[9] People remain infectious for seven to twelve days in moderate cases, and up to two weeks in severe cases.[9][13]

Common symptoms include fever, cough, fatigue, breathing difficulties, and loss of smell. Complications may include pneumonia and acute respiratory distress syndrome. The incubation period is typically around five days but may range from one to 14 days.[13] There are several vaccine candidates in development, although none have proven their safety and efficacy. There is no known specific antiviral medication, so primary treatment is currently symptomatic.[14]

Recommended preventive measures include hand washing, covering mouth or wearing face mask when sneezing or coughing, social distancing, disinfecting surfaces, ventilation and air-filtering, and monitoring and self-isolation if exposed or symptomatic. Travel restrictions, lockdowns, workplace hazard controls, and facility closures have been implemented. Many places have also worked to increase testing capacity and trace contacts of the infected. These have caused social and economic disruption, including the largest global recession since the Great Depression.[15] Extreme poverty and global famines are affecting hundreds of millions, inflamed by supply shortages. Many events, the environment and education systems have also been affected. There has also been much related misinformation, as well as incidents of xenophobia and racism against Chinese people and against those perceived as being Chinese or as being from areas with high infection rates.[16]

Contents [hide]

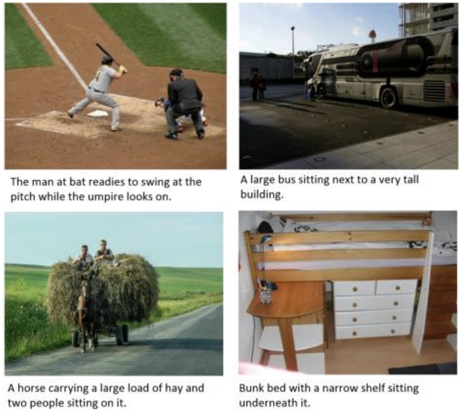
English Wikipedia

Plain Language

Available Resources and Our Goal

Challenges 2:
Low (Visual) Grounding Ratio

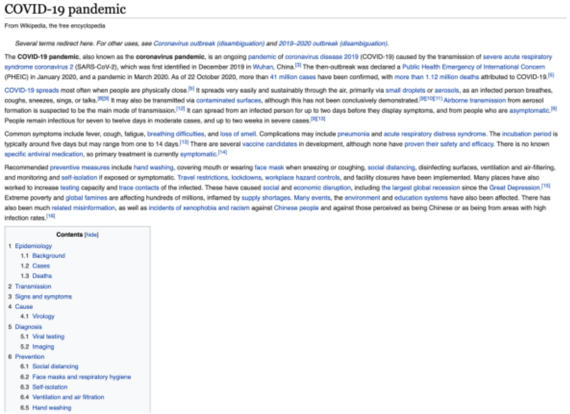
We have



MS COCO Captioning

Challenges 1:
Data
Divergence

We want


COVID-19 pandemic

From Wikipedia, the free encyclopedia

Several terms redirect here. For other uses, see Coronavirus outbreak (disambiguation) and 2019–2020 outbreak (disambiguation).

The **COVID-19 pandemic**, also known as the **coronavirus pandemic**, is an ongoing pandemic of coronavirus disease 2019 (COVID-19) caused by the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was first identified in December 2019 in Wuhan, China.[3] The then-outbreak was declared a Public Health Emergency of International Concern (PHEIC) in January 2020, and a pandemic in March 2020. As of 22 October 2020, more than 41 million cases have been confirmed, with more than 1.12 million deaths attributed to COVID-19.[5]

COVID-19 spreads most often when people are physically close.[5] It spreads very easily and sustainably through the air, primarily via small droplets or aerosols, as an infected person breathes, coughs, sneezes, sings, or talks.[8][9] It may also be transmitted via contaminated surfaces, although this has not been conclusively demonstrated.[9][10][11] Airborne transmission from aerosol formation is suspected to be the main mode of transmission.[12] It can spread from an infected person for up to two days before they display symptoms, and from people who are asymptomatic.[9] People remain infectious for seven to twelve days in moderate cases, and up to two weeks in severe cases.[9][13]

Common symptoms include fever, cough, fatigue, breathing difficulties, and loss of smell. Complications may include pneumonia and acute respiratory distress syndrome. The incubation period is typically around five days but may range from one to 14 days.[13] There are several vaccine candidates in development, although none have proven their safety and efficacy. There is no known specific antiviral medication, so primary treatment is currently symptomatic.[14]

Recommended preventive measures include hand washing, covering mouth or wearing face mask when sneezing or coughing, social distancing, disinfecting surfaces, ventilation and air-filtering, and monitoring and self-isolation if exposed or symptomatic. Travel restrictions, lockdowns, workplace hazard controls, and facility closures have been implemented. Many places have also worked to increase testing capacity and trace contacts of the infected. These have caused social and economic disruption, including the largest global recession since the Great Depression.[15] Extreme poverty and global famines are affecting hundreds of millions, inflamed by supply shortages. Many events, the environment and education systems have also been affected. There has also been much related misinformation, as well as incidents of xenophobia and racism against Chinese people and against those perceived as being Chinese or as being from areas with high infection rates.[16]

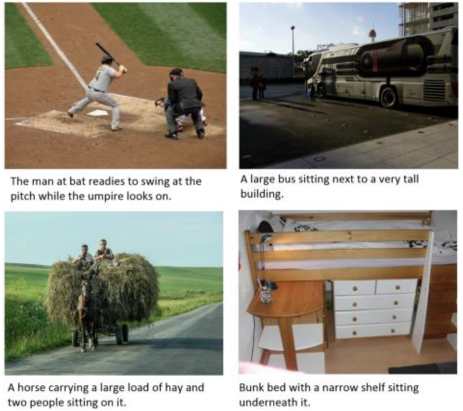
Contents [hide]

English Wikipedia

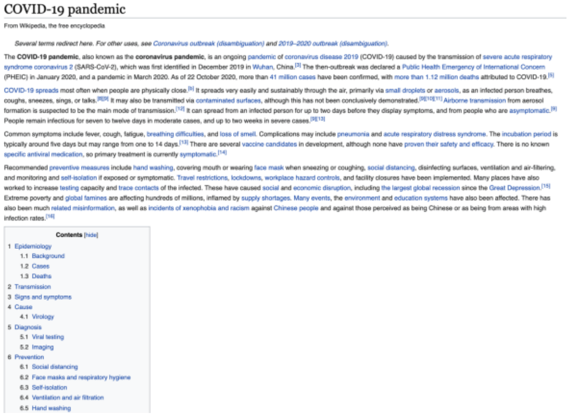
(Visually-) Grounded Language

Plain Language

Challenges 1: Data Divergence



MS COCO Captioning:
7M Tokens



COVID-19 pandemic

From Wikipedia, the free encyclopedia

Several terms redirect here. For other uses, see Coronavirus outbreak (disambiguation) and 2019–2020 outbreak (disambiguation).

The **COVID-19 pandemic**, also known as the **coronavirus pandemic**, is an ongoing pandemic of coronavirus disease 2019 (COVID-19) caused by the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was first identified in December 2019 in Wuhan, China.[3] The then-outbreak was declared a Public Health Emergency of International Concern (PHEIC) in January 2020, and a pandemic in March 2020. As of 22 October 2020, more than 41 million cases have been confirmed, with more than 1.12 million deaths attributed to COVID-19.[5]

COVID-19 spreads most often when people are physically close.[9] It spreads very easily and sustainably through the air, primarily via small droplets or aerosols, as an infected person breathes, coughs, sneezes, sings, or talks.[8][9] It may also be transmitted via contaminated surfaces, although this has not been conclusively demonstrated.[9][10][11] Airborne transmission from aerosol formation is suspected to be the main mode of transmission.[12] It can spread from an infected person for up to two days before they display symptoms, and from people who are asymptomatic.[9] People remain infectious for seven to twelve days in moderate cases, and up to two weeks in severe cases.[9][13]

Common symptoms include fever, cough, fatigue, breathing difficulties, and loss of smell. Complications may include pneumonia and acute respiratory distress syndrome. The incubation period is typically around five days but may range from one to 14 days.[13] There are several vaccine candidates in development, although none have proven their safety and efficacy. There is no known specific antiviral medication, so primary treatment is currently symptomatic.[14]

Recommended preventive measures include hand washing, covering mouth or wearing face mask when sneezing or coughing, social distancing, disinfecting surfaces, ventilation and air-filtering, and monitoring and self-isolation if exposed or symptomatic. Travel restrictions, lockdowns, workplace hazard controls, and facility closures have been implemented. Many places have also worked to increase testing capacity and trace contacts of the infected. These have caused social and economic disruption, including the largest global recession since the Great Depression.[15] Extreme poverty and global famines are affecting hundreds of millions, inflamed by supply shortages. Many events, the environment and education systems have also been affected. There has also been much related misinformation, as well as incidents of xenophobia and racism against Chinese people and against those perceived as being Chinese or as being from areas with high infection rates.[16]

Contents [hide]

English Wikipedia:
~2800M Tokens

The **amount** of grounded language is much less than plain language.

Challenges 1: Data Divergence

Example:
A cat sits in the shadow of a blue doorway.

MS COCO Captioning:
11.8 tokens / sentence

Example:
It is the only domesticated species in the family Felidae and is often referred to as the domestic cat to distinguish it from the wild members of the family.

English Wikipedia:
24.1 tokens / sentence

The **distribution** of grounded language is different from plain language.

Challenges 1: Data Divergence

Example:
A cat sits in the shadow of a blue doorway.

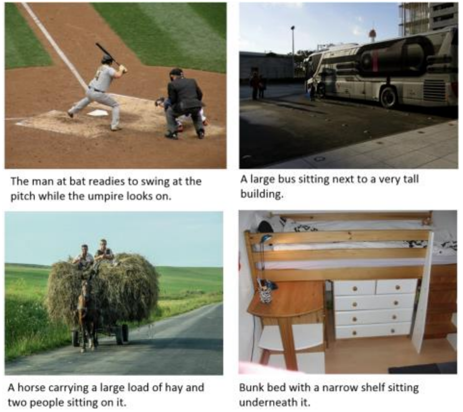
MS COCO Captioning:
Vocab Size - 9K

Example:
It is the only domesticated species in the family Felidae and is often referred to as the domestic cat to distinguish it from the wild members of the family.

English Wikipedia:
Vocab Size - 29K

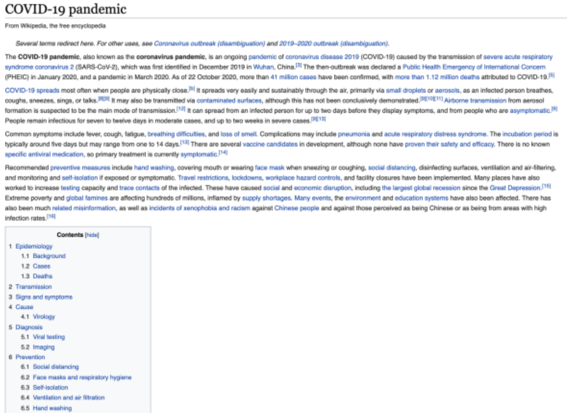
The **distribution** of grounded language is different from plain language.

Solution 1: Extrapolation



MS COCO Captioning:
7M Tokens

Extrapolate the multimodal alignments to plain language



COVID-19 pandemic

From Wikipedia, the free encyclopedia

Several terms redirect here. For other uses, see Coronavirus outbreak (disambiguation) and 2019–2020 outbreak (disambiguation).

The **COVID-19 pandemic**, also known as the **coronavirus pandemic**, is an ongoing pandemic of coronavirus disease 2019 (COVID-19) caused by the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was first identified in December 2019 in Wuhan, China.[3] The then-outbreak was declared a Public Health Emergency of International Concern (PHEIC) in January 2020, and a pandemic in March 2020. As of 22 October 2020, more than 41 million cases have been confirmed, with more than 1.12 million deaths attributed to COVID-19.[5]

COVID-19 spreads most often when people are physically close.[6] It spreads very easily and sustainably through the air, primarily via small droplets or aerosols, as an infected person breathes, coughs, sneezes, sings, or talks.[8][9] It may also be transmitted via contaminated surfaces, although this has not been conclusively demonstrated.[9][10][11] Airborne transmission from aerosol formation is suspected to be the main mode of transmission.[12] It can spread from an infected person for up to two days before they display symptoms, and from people who are asymptomatic.[9] People remain infectious for seven to twelve days in moderate cases, and up to two weeks in severe cases.[9][13]

Common symptoms include fever, cough, fatigue, breathing difficulties, and loss of smell. Complications may include pneumonia and acute respiratory distress syndrome. The incubation period is typically around five days but may range from one to 14 days.[13] There are several vaccine candidates in development, although none have proven their safety and efficacy. There is no known specific antiviral medication, so primary treatment is currently symptomatic.[14]

Recommended preventive measures include hand washing, covering mouth or wearing face mask when sneezing or coughing, social distancing, disinfecting surfaces, ventilation and air-filtering, and monitoring and self-isolation if exposed or symptomatic. Travel restrictions, lockdowns, workplace hazard controls, and facility closures have been implemented. Many places have also worked to increase testing capacity and trace contacts of the infected. These have caused social and economic disruption, including the largest global recession since the Great Depression.[15] Extreme poverty and global famines are affecting hundreds of millions, inflamed by supply shortages. Many events, the environment and education systems have also been affected. There has also been much related misinformation, as well as incidents of xenophobia and racism against Chinese people and against those perceived as being Chinese or as being from areas with high infection rates.[16]

Contents [hide]

English Wikipedia:
~2800M Tokens

https://en.wikipedia.org/w/index.php?title=Cat&oldid=984456881

Challenges 2: Low Grounding Ratio

Bold Blue: Visually-Grounded
Unbold Blue: Unsure

Example:
A **cat sits** in the **shadow** of a **blue doorway**.

Example:
A **cat** can either be a **house cat**, a **farm cat** or a feral **cat**; the latter ranges freely and avoids **human** contact.

MS COCO Captioning:
Grounding Ratio - 54.8%

English Wikipedia:
Grounding Ratio - 27.7%

Visually-Grounded (informal definition): If the word could be mapped to a visual image.

Challenges 2: Low Grounding Ratio

Bold Blue: Visually-Grounded
Unbold Blue: Unsure

Example:
A **cat sits** in the **shadow** of a **blue doorway**.

Example:
A **cat** can either be a **house cat**, a **farm cat** or a feral **cat**; the latter ranges freely and avoids **human** contact.

MS COCO Captioning:
Grounding Ratio - 54.8%

English Wikipedia:
Grounding Ratio - 27.7%

Problem: if the grounding ratio is small, it's hard to provide dense visual supervision.

Solution 2: Contextualized Grounding

Bold Blue: Visually-Grounded
Unbold Blue: Unsure
Bold Red: Contextually Visually-Grounded
Bold Red: Unsure

Example:
A **cat sits in** the **shadow** of a **blue doorway**.

Example:
A **cat** can either be a **house cat**, a **farm cat** or a feral **cat**; the **latter** ranges freely and **avoids human contact**.

MS COCO Captioning:
Grounding Ratio - 54.8%

English Wikipedia:
Grounding Ratio - 27.7%

Contextually Visually-Grounded (informal definition)**:** If the (word, context) pair is visually-grounded.

https://possibility-cp.com/the-2-mistakes-to-avoid-when-recruiting-board-members/
https://zhuanlan.zhihu.com/p/67822808

Solution 2: Contextualized Grounding

avoids

Grounding

(Token-Level Grounding)

A **cat** can either be a **house cat**, a **farm cat** or a feral **cat**; the **latter ranges** freely and **avoids** **human contact**.

Contextualized Grounding

https://en.wikipedia.org/wiki/Google_Contacts

Solution 2: Contextualized Grounding

contact

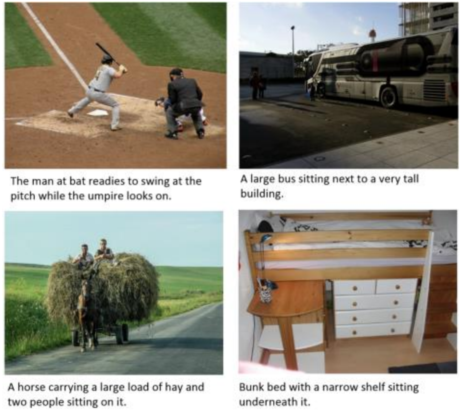
A **cat** can either be a **house cat**, a **farm cat** or a feral **cat**; the **latter ranges** freely and **avoids human contact**.

Grounding

(Token-Level Grounding)

Contextualized Grounding

Vokenization = Extrapolation + Contextual Grounding

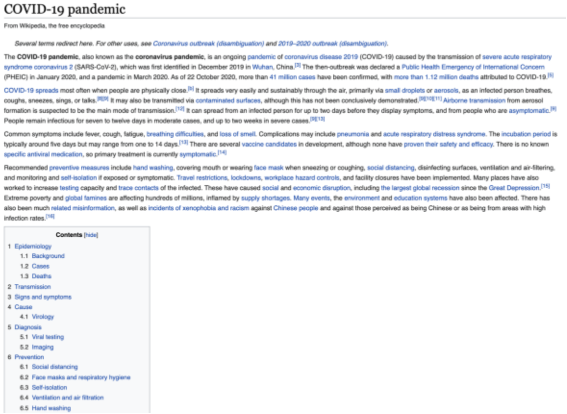


MS COCO Captioning:
7M Tokens

Train

Annotate

(Contextual) Vokenizer



COVID-19 pandemic

From Wikipedia, the free encyclopedia

Several terms redirect here. For other uses, see Coronavirus outbreak (disambiguation) and 2019–2020 outbreak (disambiguation).

The **COVID-19 pandemic**, also known as the **coronavirus pandemic**, is an ongoing pandemic of coronavirus disease 2019 (COVID-19) caused by the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was first identified in December 2019 in Wuhan, China.[3] The then-outbreak was declared a Public Health Emergency of International Concern (PHEIC) in January 2020, and a pandemic in March 2020. As of 22 October 2020, more than 41 million cases have been confirmed, with more than 1.12 million deaths attributed to COVID-19.[5]

COVID-19 spreads most often when people are physically close.[5] It spreads very easily and sustainably through the air, primarily via small droplets or aerosols, as an infected person breathes, coughs, sneezes, sings, or talks.[8][9] It may also be transmitted via contaminated surfaces, although this has not been conclusively demonstrated.[9][10][11] Airborne transmission from aerosol formation is suspected to be the main mode of transmission.[12] It can spread from an infected person for up to two days before they display symptoms, and from people who are asymptomatic.[9] People remain infectious for seven to twelve days in moderate cases, and up to two weeks in severe cases.[9][13]

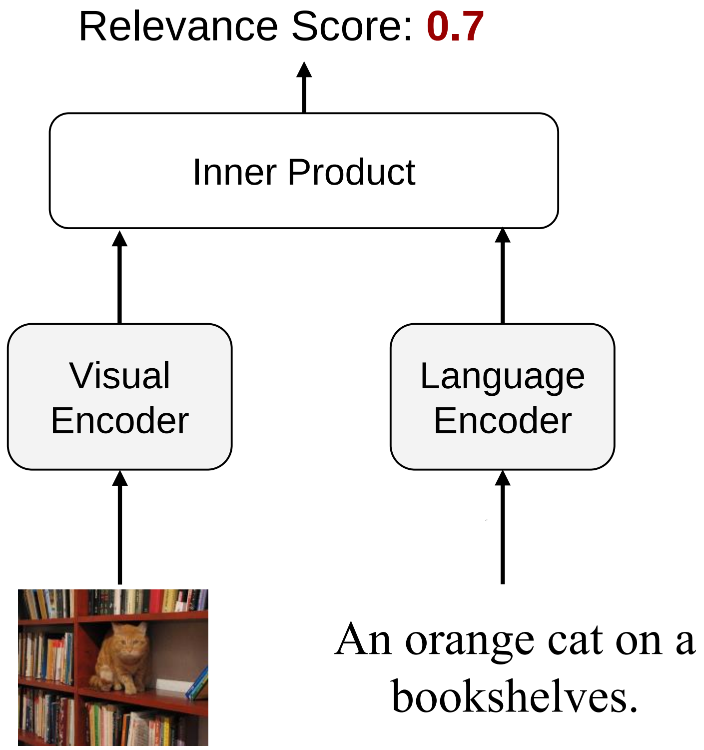
Common symptoms include fever, cough, fatigue, breathing difficulties, and loss of smell. Complications may include pneumonia and acute respiratory distress syndrome. The incubation period is typically around five days but may range from one to 14 days.[13] There are several vaccine candidates in development, although none have proven their safety and efficacy. There is no known specific antiviral medication, so primary treatment is currently symptomatic.[14]

Recommended preventive measures include hand washing, covering mouth or wearing face mask when sneezing or coughing, social distancing, disinfecting surfaces, ventilation and air-filtering, and monitoring and self-isolation if exposed or symptomatic. Travel restrictions, lockdowns, workplace hazard controls, and facility closures have been implemented. Many places have also worked to increase testing capacity and trace contacts of the infected. These have caused social and economic disruption, including the largest global recession since the Great Depression.[15] Extreme poverty and global famines are affecting hundreds of millions, inflamed by supply shortages. Many events, the environment and education systems have also been affected. There has also been much related misinformation, as well as incidents of xenophobia and racism against Chinese people and against those perceived as being Chinese or as being from areas with high infection rates.[16]

Contents [hide]

English Wikipedia:
2800M Tokens

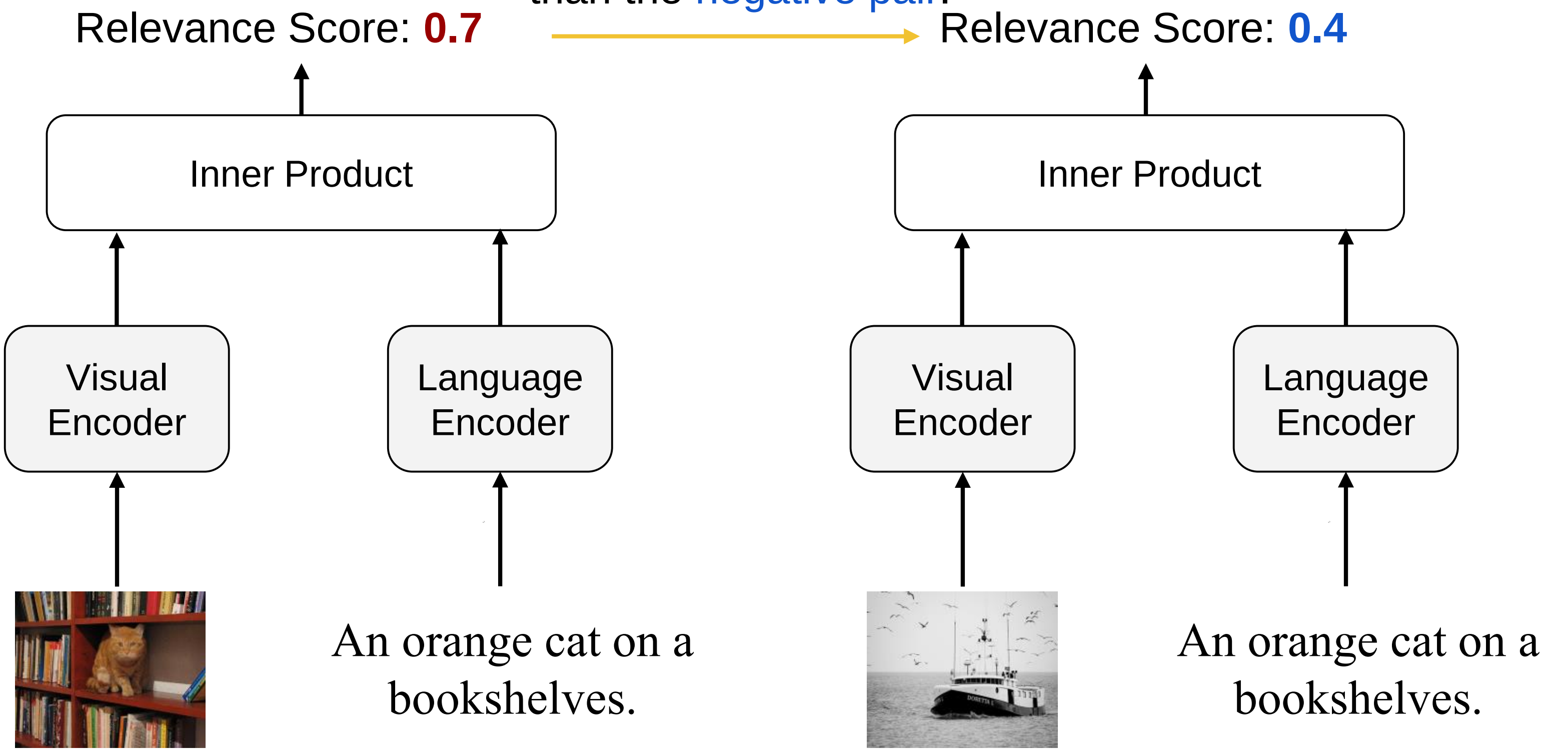
Vokenizer: Modeling



Positive Pair

Vokenizer: Training

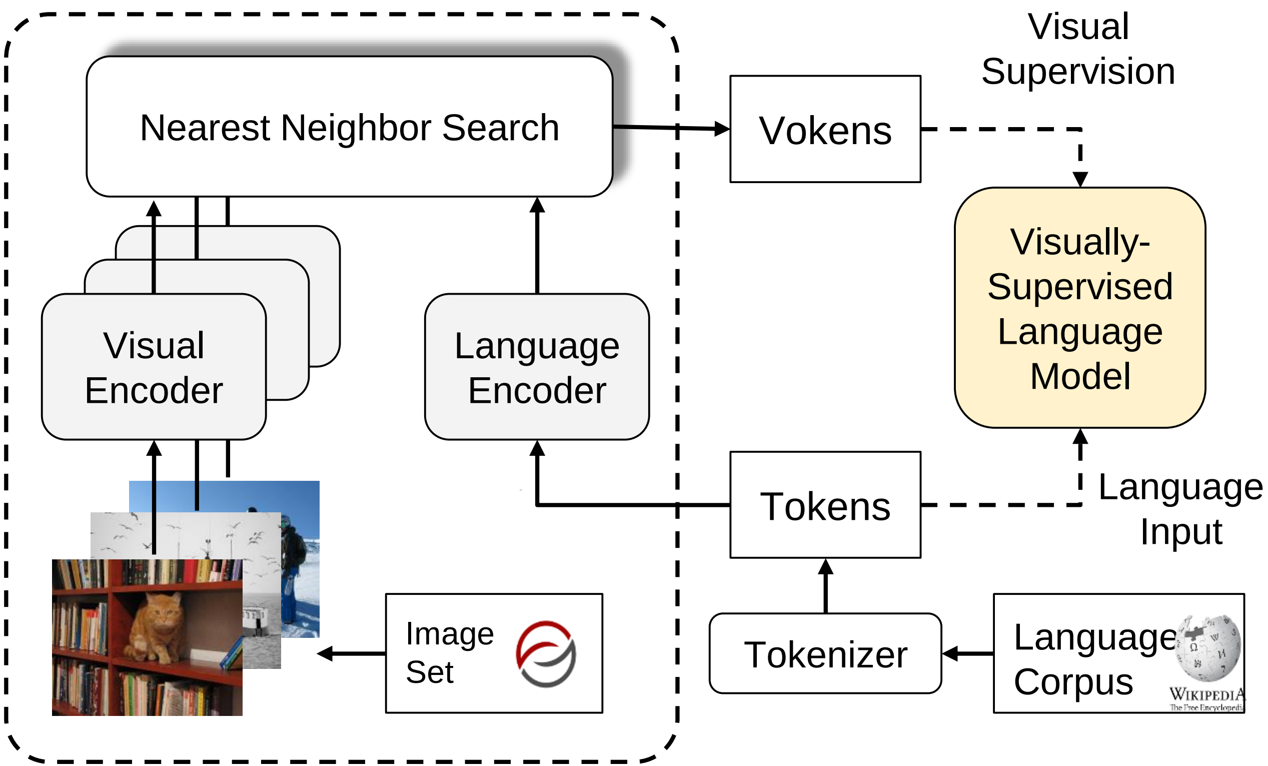
Optimize the model so that the score of positive pair is **higher** than the negative pair.



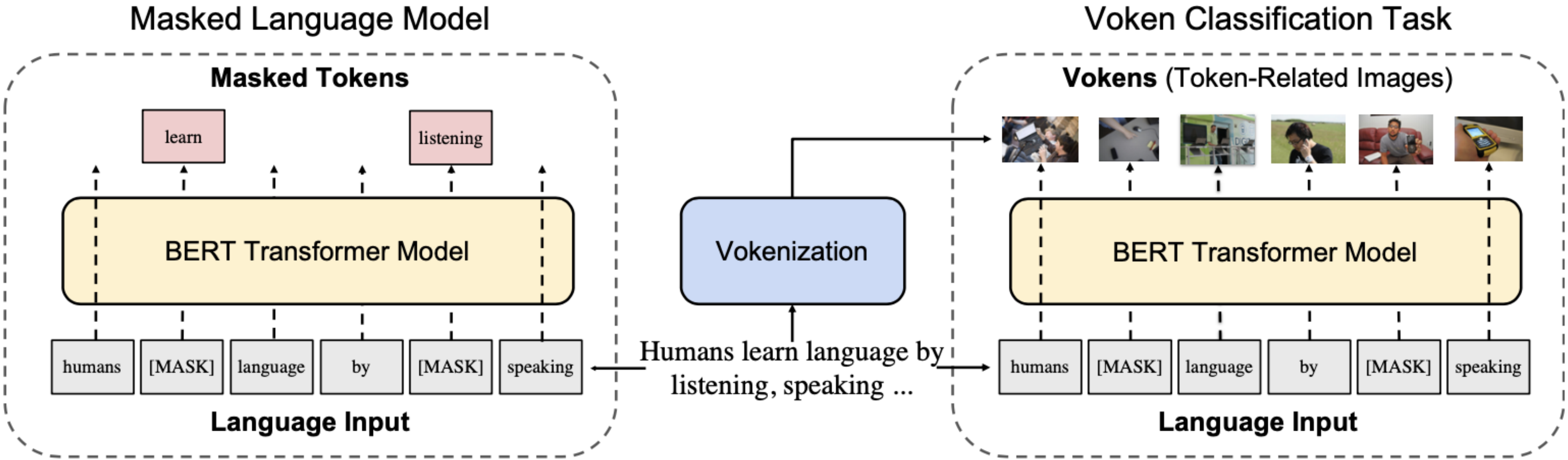
Positive Pair

Negative Pair

Vokenizer: Inference

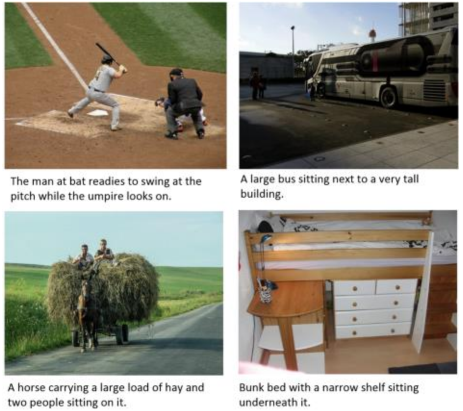


Final learning objective



Pranav Rajpurkar, et al. SQuAD. EMNLP 2016
Adina Williams, et al. MNLI. NAACL 2018
Rowan Zellers, et al. Swag. EMNLP 2018

Experiment Setups



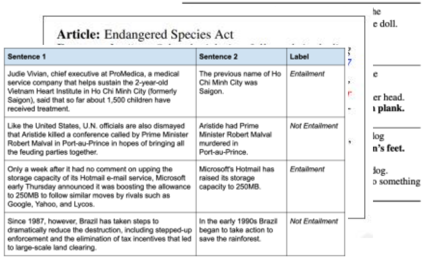
MS COCO Captioning

Train the Vokenizer

English Wikipedia

Train the Visually-Supervised Language Model

SQuAD SWAG



Sentence 1	Sentence 2	Label
Judie Vivian, chief executive at ProMedica, a medical service company that helps sustain the 2-year-old Vietnam Heart Institute in Ho Chi Minh City (formerly Saigon), said that so far about 1,500 children have received treatment.	The previous name of Ho Chi Minh City was Saigon.	Entailment
Like the United States, U.N. officials are also dismayed that Aristide killed a conference called by Prime Minister Robert Malval in Port-au-Prince in hopes of bringing all the feuding parties together.	Aristide had Prime Minister Robert Malval murdered in Port-au-Prince.	Not Entailment
Only a week after it had no comment on upping the storage capacity of its Hotmail e-mail service, Microsoft early Thursday announced it was boosting the allowance to 250MB to follow similar moves by rivals such as Google, Yahoo, and Lycos.	Microsoft's Hotmail has raised its storage capacity to 250MB.	Entailment
Since 1987, however, Brazil has taken steps to dramatically reduce the destruction, including stepped-up enforcement and the elimination of tax incentives that led to large-scale land clearing.	In the early 1990s Brazil began to take action to save the rainforest.	Not Entailment

MNLI

Text-Only Tasks

Evaluation

Note: Long passages in SQuAD need sequence length 512 but our computational resources only support sequence length 128. Other experiments are not affected by this. Please refer to got detailed implementation of SQuAD (the sliding window approach): https://github.com/google-research/bert/blob/master/run_squad.py

Results with BERT Backbone

Pre-trained on	SST-2	QNLI	QQP	MNLI
MS COCO	83.7	60.6	82.1	69.3
Wiki103*	85.8	77.9	84.8	73.9

Method	SST-2	QNLI	QQP	MNLI	SQuAD v1.1	SQuAD v2.0	SWAG	Avg.
$BERT_{6L/512H}$	88.0	85.2	87.1	77.9	71.3/80.2	57.2/60.8	56.2	75.6
$BERT_{6L/512H}$ + Voken-cls	89.7	85.0	87.3	78.6	71.5/80.2	61.3/64.6	58.2	76.8
$BERT_{12L/768H}$	89.3	87.9	83.2	79.4	77.0/85.3	67.7/71.1	65.7	79.4
$BERT_{12L/768H}$ + Voken-cls	**92.2**	**88.6**	**88.6**	**82.6**	**78.8/86.7**	68.1/71.2	**70.6**	**82.1**

Small experiments on Wiki103 (reproducible to the community). **1.2%** average improvement.

Note: We did some simplifications (constant sequence length, no NSP task) to standardize training process. We also excludes the unavailable BookCopus but only kept English Wikipedia.

Results with BERT Backbone

2.7% average improvement when pre-trained on Wikipedia.

Method	SST-2	QNLI	QQP	MNLI	SQuAD v1.1	SQuAD v2.0	SWAG	Avg.
$\text{BERT}_{6L/512H}$	88.0	85.2	87.1	77.9	71.3/80.2	57.2/60.8	56.2	75.6
$\text{BERT}_{6L/512H}$ + Voken-cls	89.7	85.0	87.3	78.6	71.5/80.2	61.3/64.6	58.2	76.8
$\text{BERT}_{12L/768H}$	89.3	87.9	83.2	79.4	77.0/85.3	67.7/71.1	65.7	79.4
$\text{BERT}_{12L/768H}$ + Voken-cls	**92.2**	**88.6**	**88.6**	**82.6**	**78.8/86.7**	68.1/71.2	**70.6**	**82.1**

Method	SST-2	QNLI	QQP	MNLI	SQuAD v1.1	SQuAD v2.0	SWAG	Avg.
$\text{BERT}_{6L/512H}$	88.0	85.2	87.1	77.9	71.3/80.2	57.2/60.8	56.2	75.6
$\text{BERT}_{6L/512H}$ + Voken-cls	89.7	85.0	87.3	78.6	71.5/80.2	61.3/64.6	58.2	76.8
$\text{BERT}_{12L/768H}$	89.3	87.9	83.2	79.4	77.0/85.3	67.7/71.1	65.7	79.4
$\text{BERT}_{12L/768H}$ + Voken-cls	**92.2**	**88.6**	**88.6**	**82.6**	**78.8/86.7**	68.1/71.2	**70.6**	**82.1**

BERT-BASE (Devlin et al., 2019) Trained with 800M BooksCorpus	90.3	89.6	88.4	82.4

System	Dev EM	Dev F1	Test EM	Test F1
Top Leaderboard Systems (Dec 10th, 2018)				
Human	-	-	82.3	91.2
#1 Ensemble - nlnet	-	-	86.0	91.7
#2 Ensemble - QANet	-	-	84.5	90.5
Published				
BiDAF+ELMo (Single)	-	85.6	-	85.8
R.M. Reader (Ensemble)	81.2	87.9	82.3	88.5
Ours				
$\text{BERT}_{\text{BASE}}$ (Single)	80.8	88.5	-	-

System	Dev	Test
ESIM+GloVe	51.9	52.7
ESIM+ELMo	59.1	59.2
OpenAI GPT	-	78.0
$\text{BERT}_{\text{BASE}}$	81.6	-
$\text{BERT}_{\text{LARGE}}$	**86.6**	**86.3**
Human (expert)†	-	85.0
Human (5 annotations)†	-	88.0

SWAG Dev and Test accuracies. †H

Hao Tan and Mohit Bansal. LXMERT, EMNLP 2019
Xiujun Li, et al. Oscar: Object-semantics aligned pre-trainingfor vision-language tasks. ECCV 2020

Vision-and-Language Pre-training

Model	Init. with BERT?	Diff. to BERT Weight	SST-2	QNLI	QQP	MNLI
ViLBERT (Lu et al., 2019)	Yes	0.0e-3	90.3	89.6	88.4	82.4
VL-BERT (Su et al., 2020)	Yes	6.4e-3	90.1	89.5	88.6	82.9
VisualBERT (Li et al., 2019)	Yes	6.5e-3	90.3	88.9	88.4	82.4
Oscar (Li et al., 2020a)	Yes	41.6e-3	87.3	50.5	86.6	77.3
LXMERT (Tan and Bansal, 2019)	No	42.0e-3	82.4	50.5	79.8	31.8
$\text{BERT}_{\text{BASE}}$ (Devlin et al., 2019)	-	0.0e-3	90.3	89.6	88.4	82.4
$\text{BERT}_{\text{BASE}}$ + Weight Noise	-	6.5e-3	89.9	89.9	88.4	82.3
$\text{BERT}_{\text{12L/768H}}$ + Voken-cls			**92.2**	**88.6**	**88.6**	**82.6**

OSCAR on MNLI: 77.3% (- 2.1%)

LXMERT on MNLI: 31.8% (- 47.6%)

Visualization

Note: The goal of vokenization is not to build the perfect token-level image retriever but to improve understanding of other types of language with related visual information.