



# Intermediate Deep Learning for Engineers

---

Spring 2024, Deep Learning for Engineers  
April 9, 2024, 9th Session

Amir Barati Farimani  
*Assistant Professor of Mechanical Engineering and Bio-Engineering*  
*Carnegie Mellon University*

# Word Embedding

## 1-of-N Encoding

apple = [ 1 0 0 0 0 ]

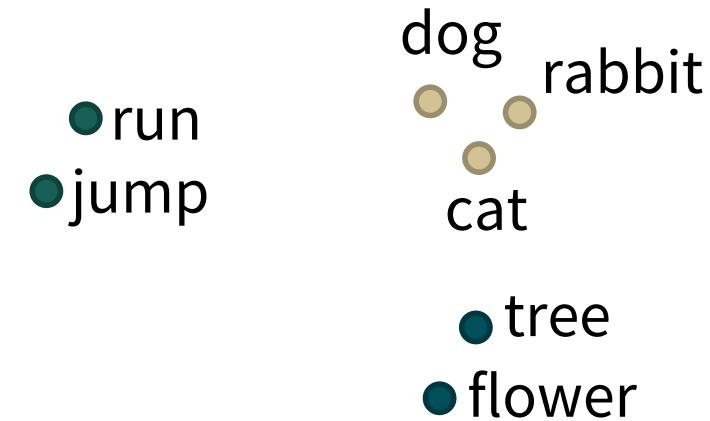
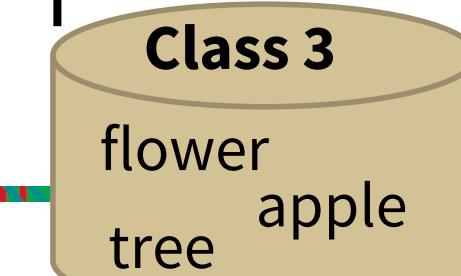
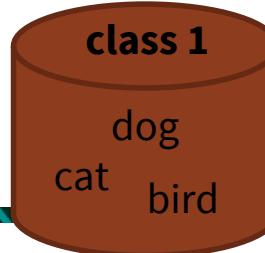
bag = [ 0 1 0 0 0 ]

cat = [ 0 0 1 0 0 ]

dog = [ 0 0 0 1 0 ]

elephant = [ 0 0 0 0 1 ]

### Word Class



# A word can have multiple senses

Have you paid that money to the bank yet ?

It is safest to deposit your money in the bank .

The victim was found lying dead on the river bank .

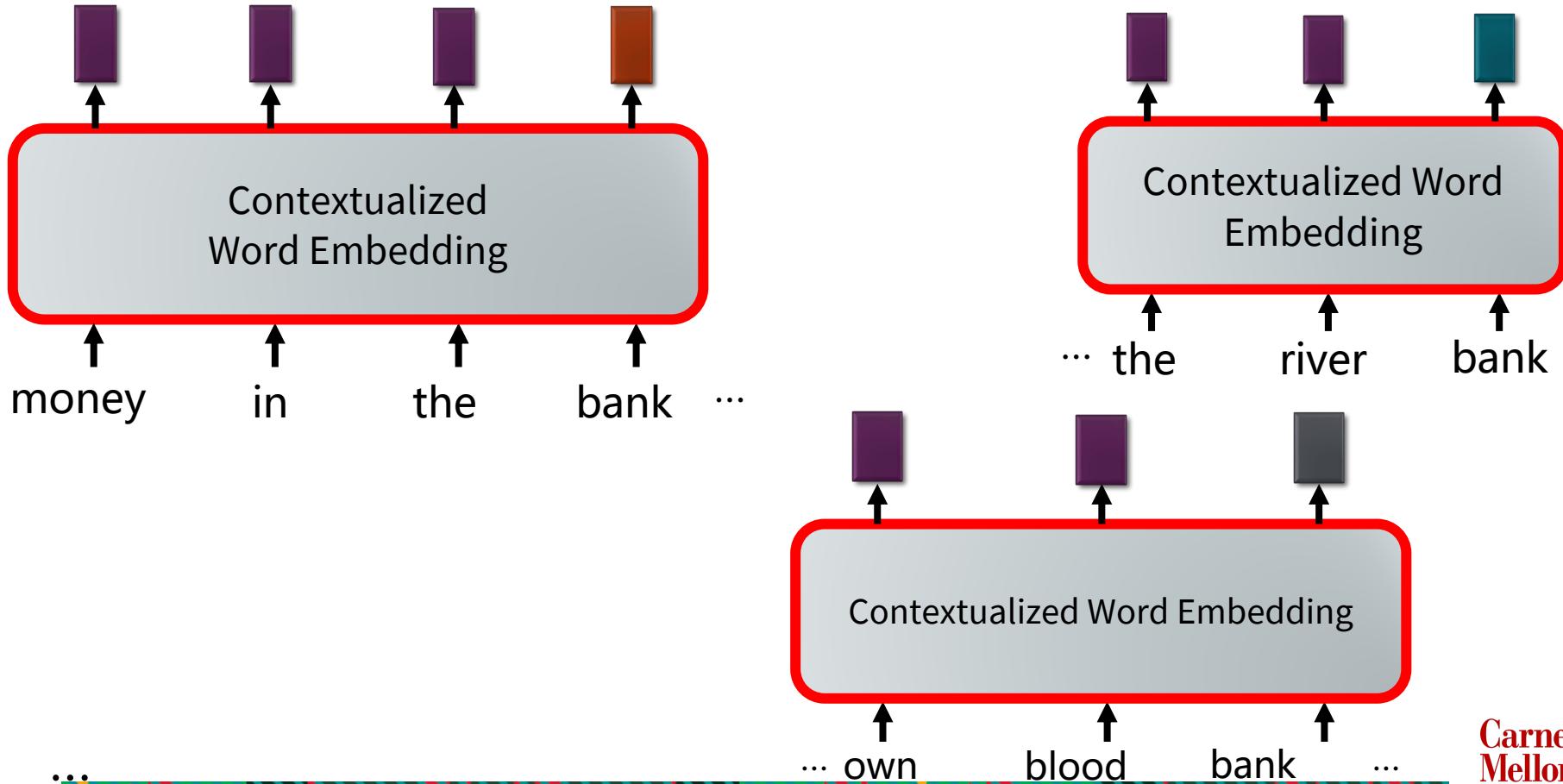
They stood on the river bank to fish.

The hospital has its own blood bank.

The third sense or not?

<https://arxiv.org/abs/1902.06006>

## Contextualized Word Embedding



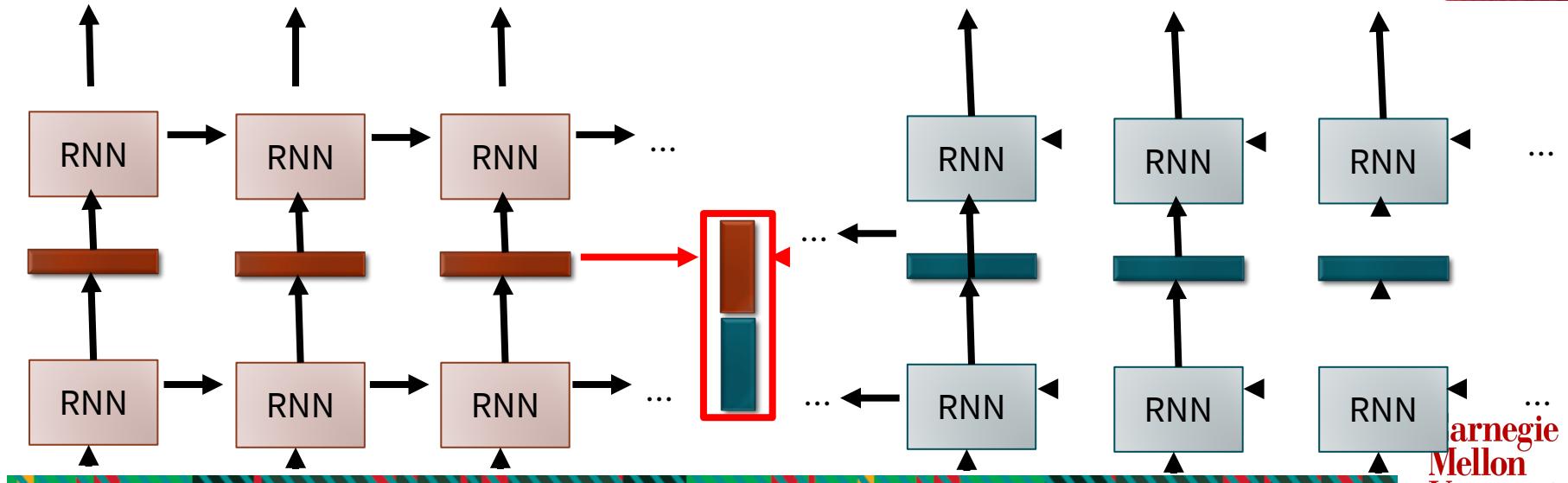
# Embeddings from Language Model (ELMO)



Embeddings from Language Model (ELMO)

<https://arxiv.org/abs/1802.05365>

RNN-BASED LANGUAGE MODELS (TRAINED FROM LOTS OF SENTENCES)



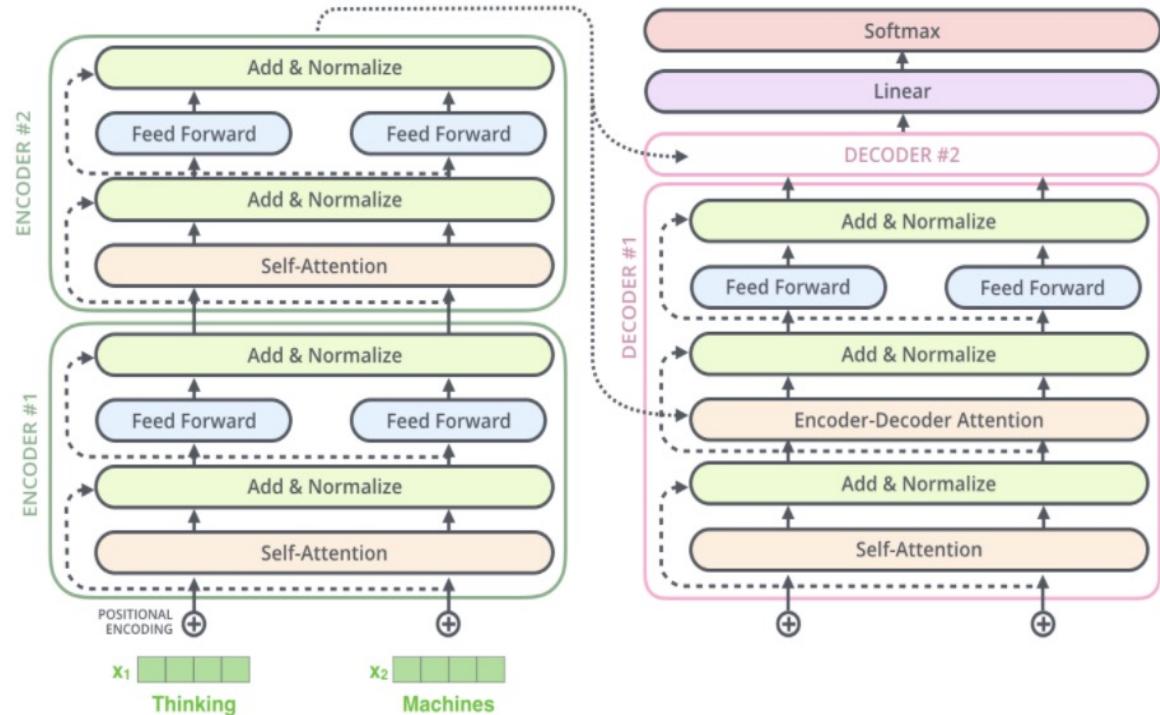
carnegie  
Mellon  
University

# Why BERT?

- Those contextual LMs have a problem: They only use left context or right context
  - But language understanding is bidirectional
  - The reason for unidirectionality: Words can “see themselves” in a bidirectional encoder
  - BERT came to the rescue!
  - BERT generates a language model by training in both directions which gives words more context
  - BERT provided a way to more accurately pre-train models with less data
  - It involved a pre-training and fine-tuning stages
  - It is based on the Transformer architecture (encoders)
- ...      ...

# Transformer's Recap

The vanilla **Transformer** model has an Encoder and Decoder, and was used in a seq2seq manner.



<https://jalammar.github.io/illustrated-transformer/>

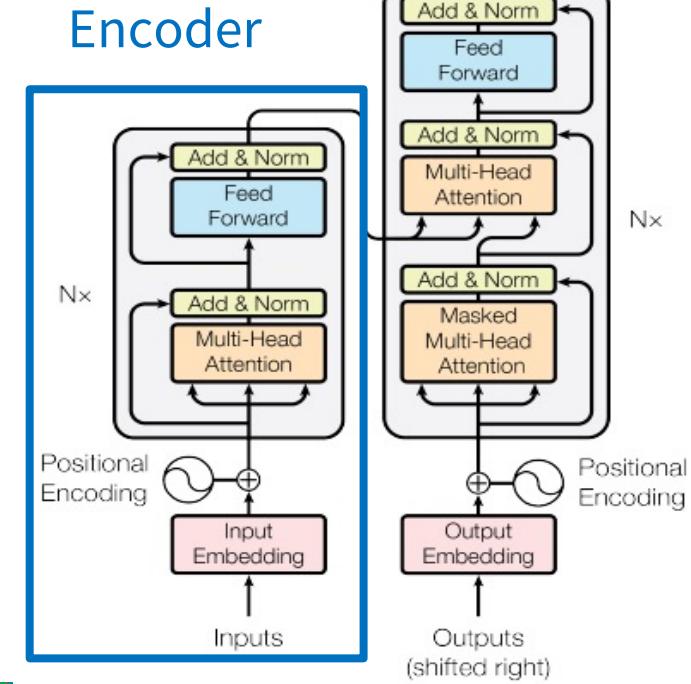
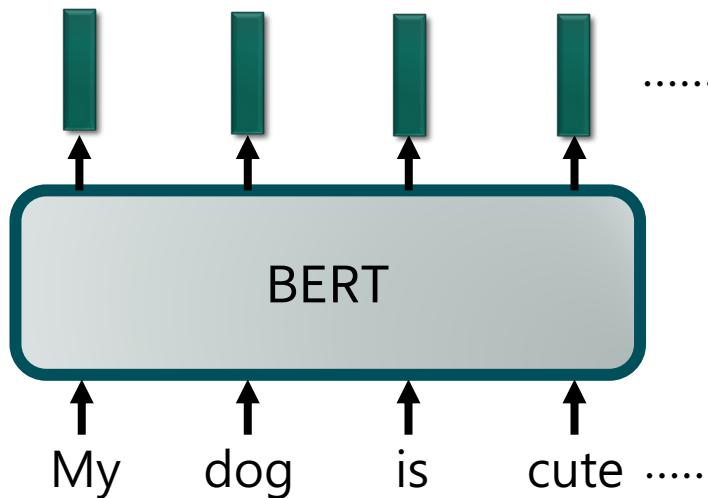
Carnegie  
Mellon  
University

# Bidirectional Encoder Representations from Transformers (BERT)



# BERT = ENCODER OF TRANSFORMER

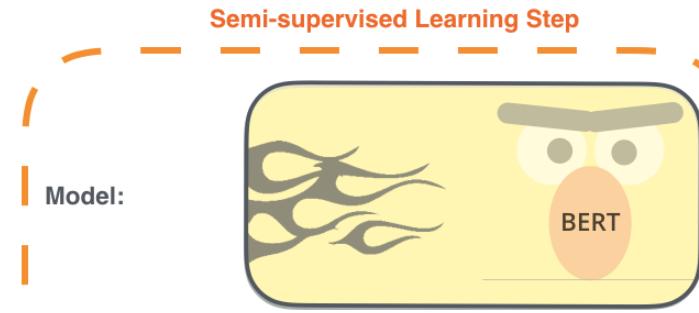
Learned from a large amount of text  
without annotation



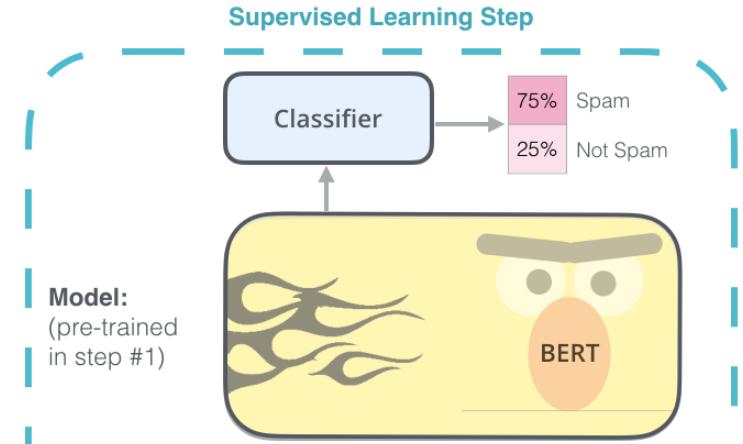
# Bidirectional Encoder Representations from Transformers (BERT)

## 1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.



## 2 - Supervised training on a specific task with a labeled dataset.



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



# BERT

## MLM objective

- **Model:** several Transformer Encoders.  
Input sentence or sentence pairs, [CLS] token, subword embeddings
- **Objective:** MLM and next-sentence prediction
- **Data:** BooksCorpus and Wikipedia

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

Possible classes:  
All English words

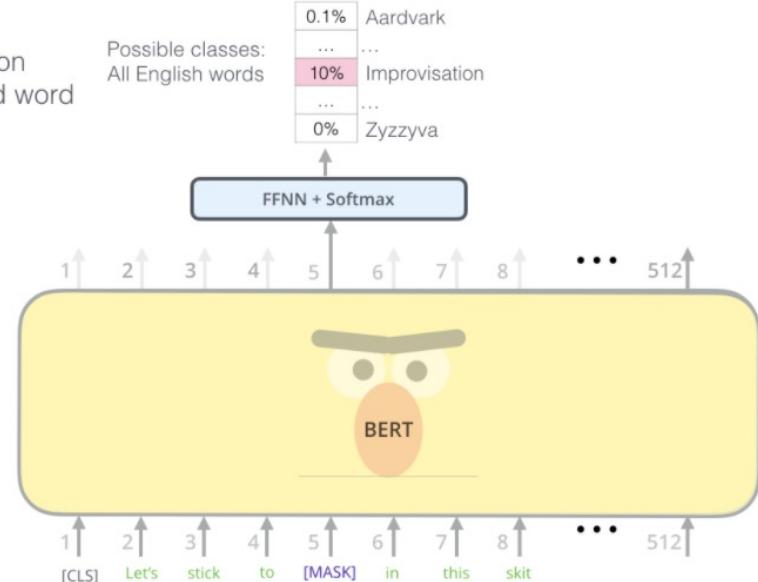
0.1%	Aardvark
...	...
10%	Improvisation
...	...
0%	Zyzyva

FFNN + Softmax

Randomly mask 15% of tokens

Input

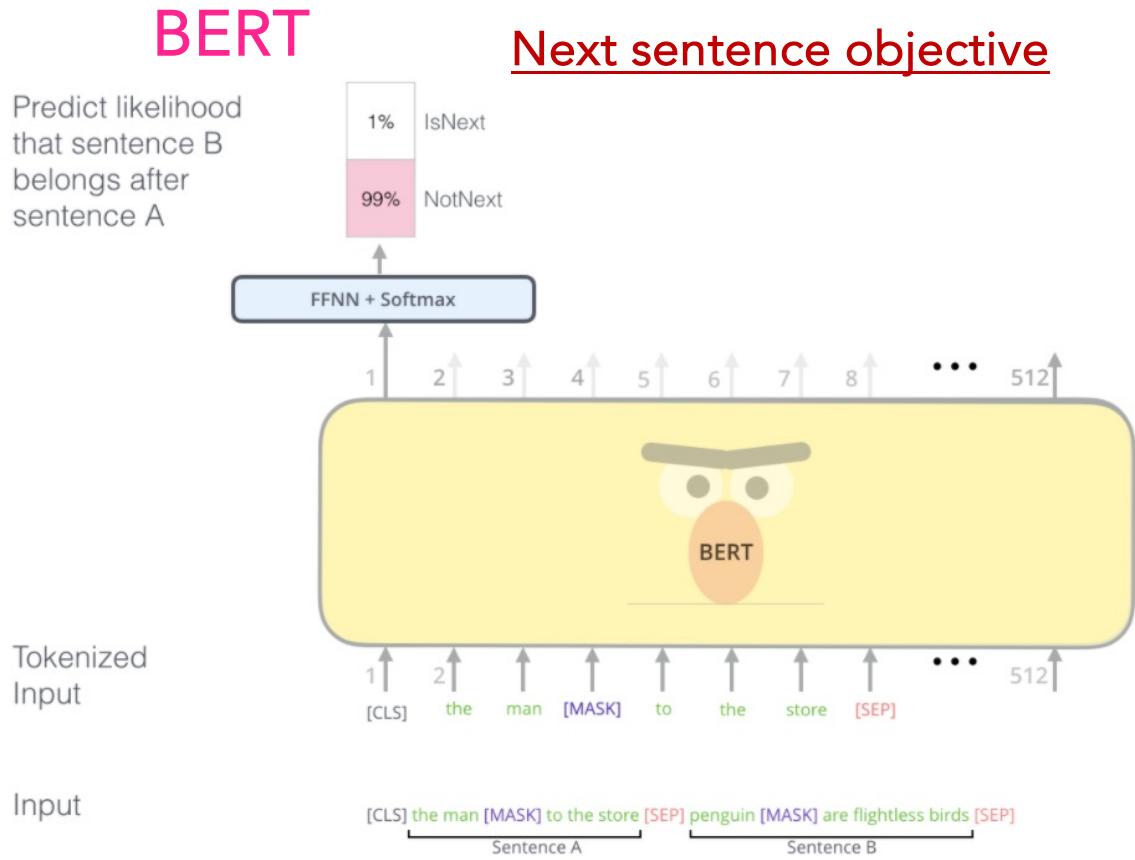
BERT's clever language modeling task masks 15% of words in the input and asks the model to predict the missing word.



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

# BERT

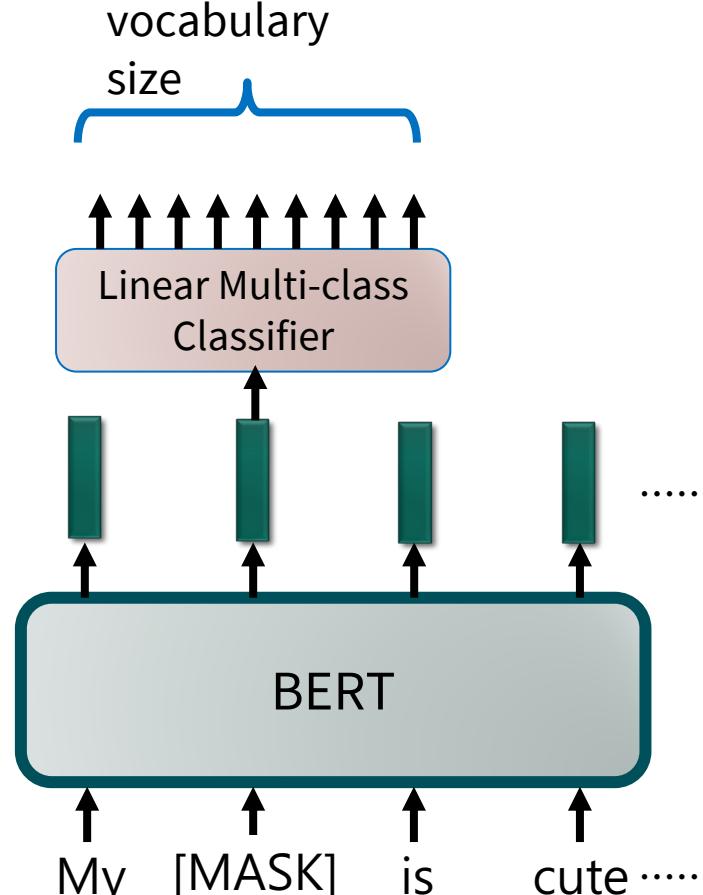
- **Model:** several Transformer Encoders.  
Input sentence or sentence pairs, [CLS] token, subword embeddings
  - **Objective:** MLM and next-sentence prediction
  - **Data:** BooksCorpus and Wikipedia



# Training of BERT

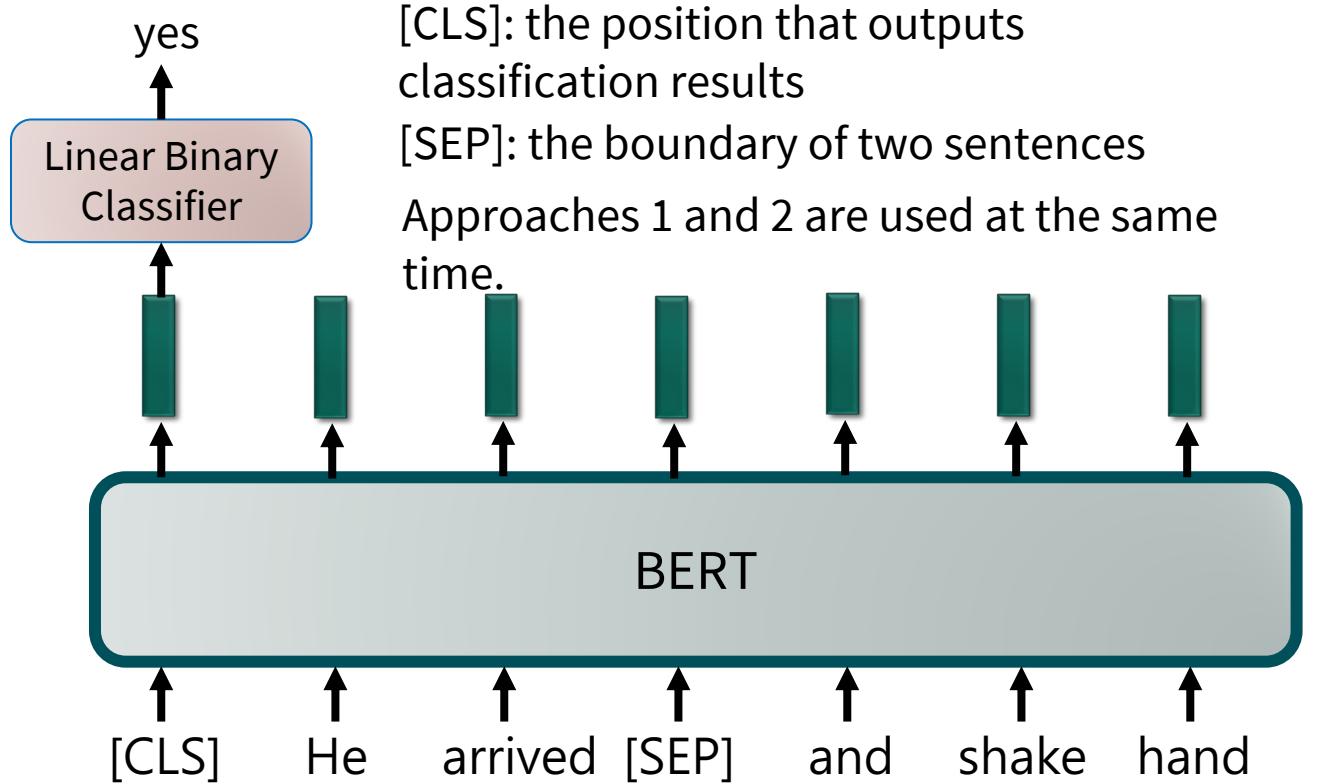
**Approach 1:** Masked LM

Predicting the  
masked word

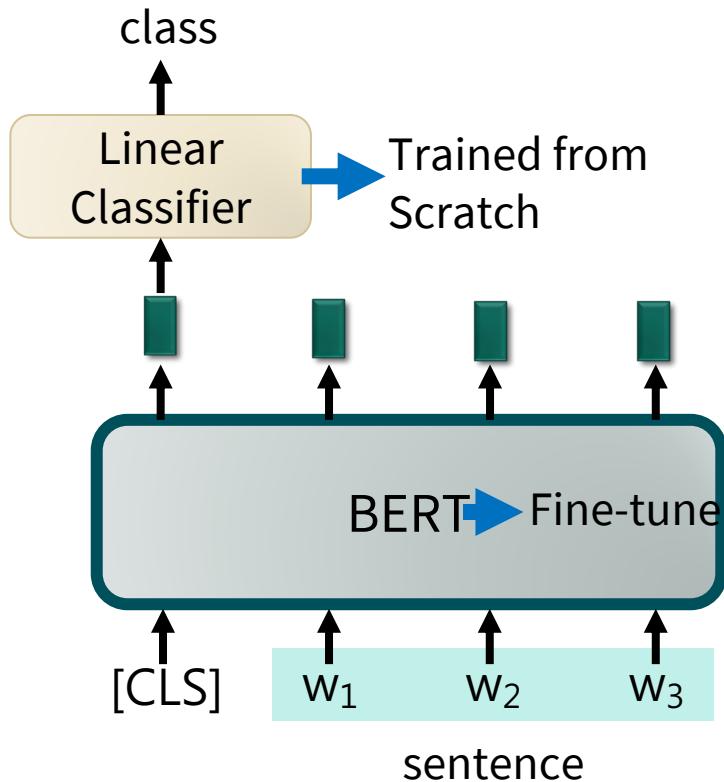


# Training of BERT

## Approach 2: Next Sentence Prediction



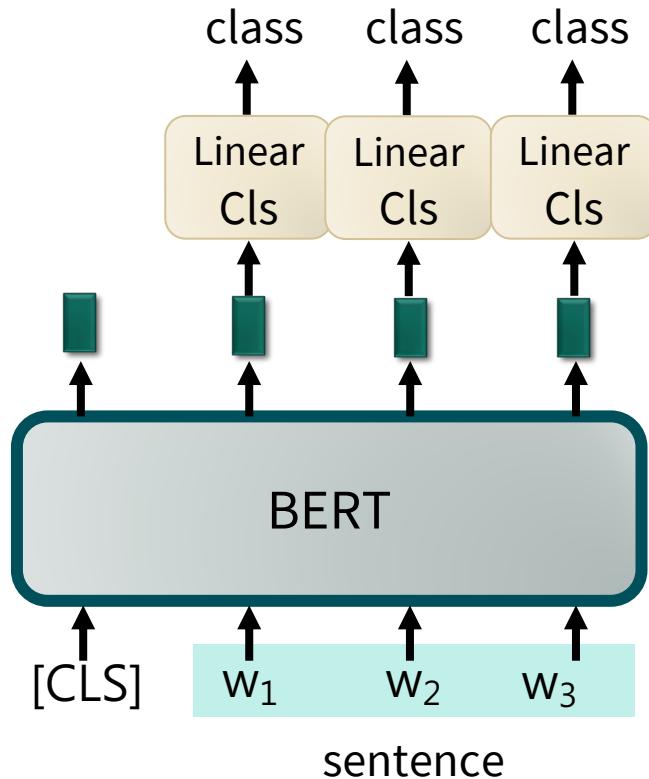
# How to use BERT – Case 1



Input: single sentence,  
output: class

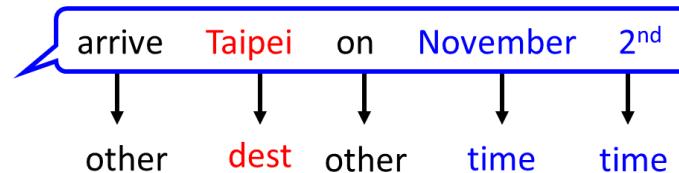
Example:  
Sentiment analysis (our HW),  
Document Classification

# How to use BERT – Case 2

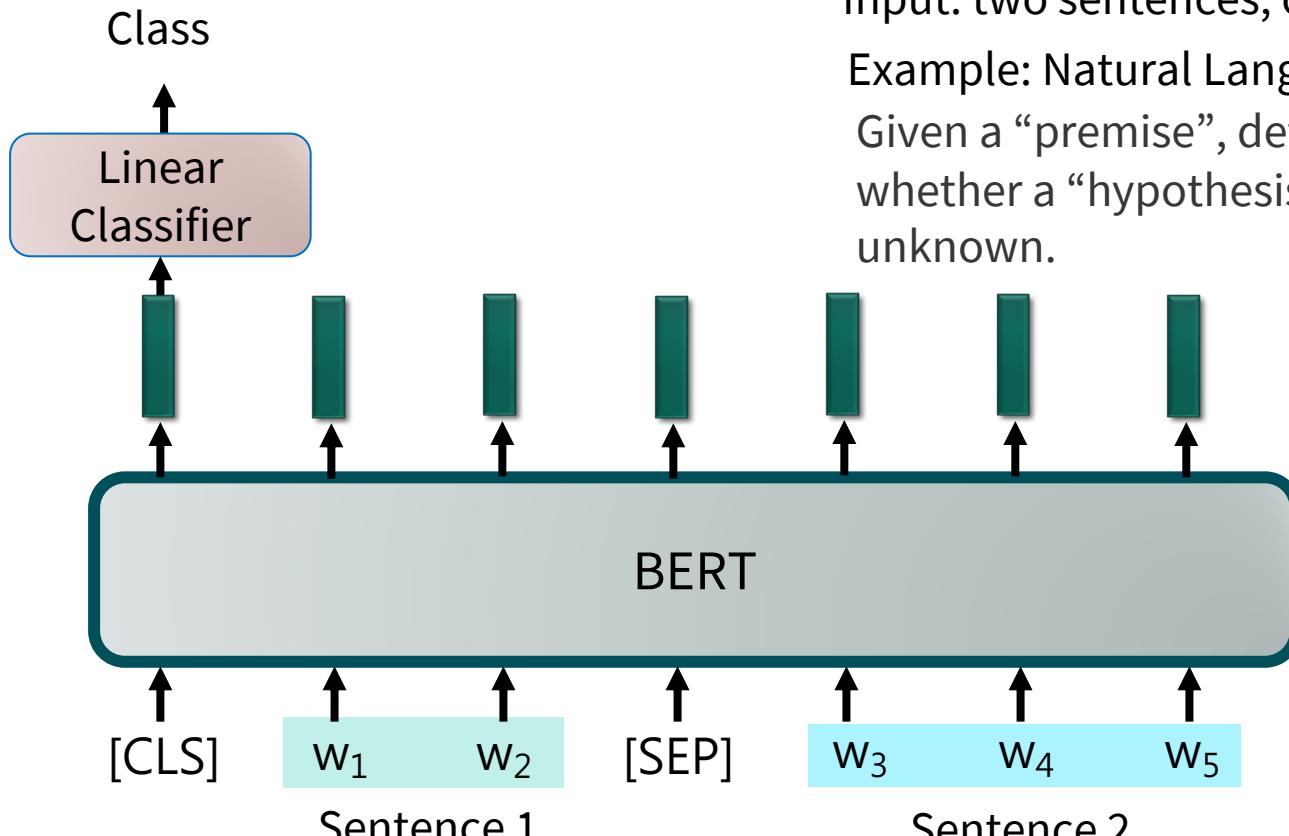


Input: single sentence,  
output: class of each word

Example: Slot filling



# How to use BERT – Case 3



Input: two sentences, output: class

Example: Natural Language Inference

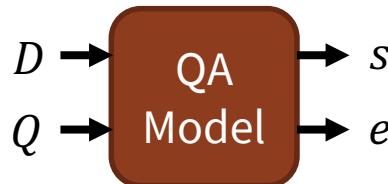
Given a “premise”, determining whether a “hypothesis” is T/F/ unknown.

# How to use BERT – Case 4

## EXTRACTION-BASED QUESTION ANSWERING (QA) (E.G. SQuAD)

**Document:**  $D = \{d_1, d_2, \dots, d_N\}$

**Query:**  $Q = \{q_1, q_2, \dots, q_N\}$



output: two integers  $(s, e)$

**Answer:**  $A = \{q_s, \dots, q_e\}$

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. 17 in forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?

**gravity**

77

79

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

**graupel**

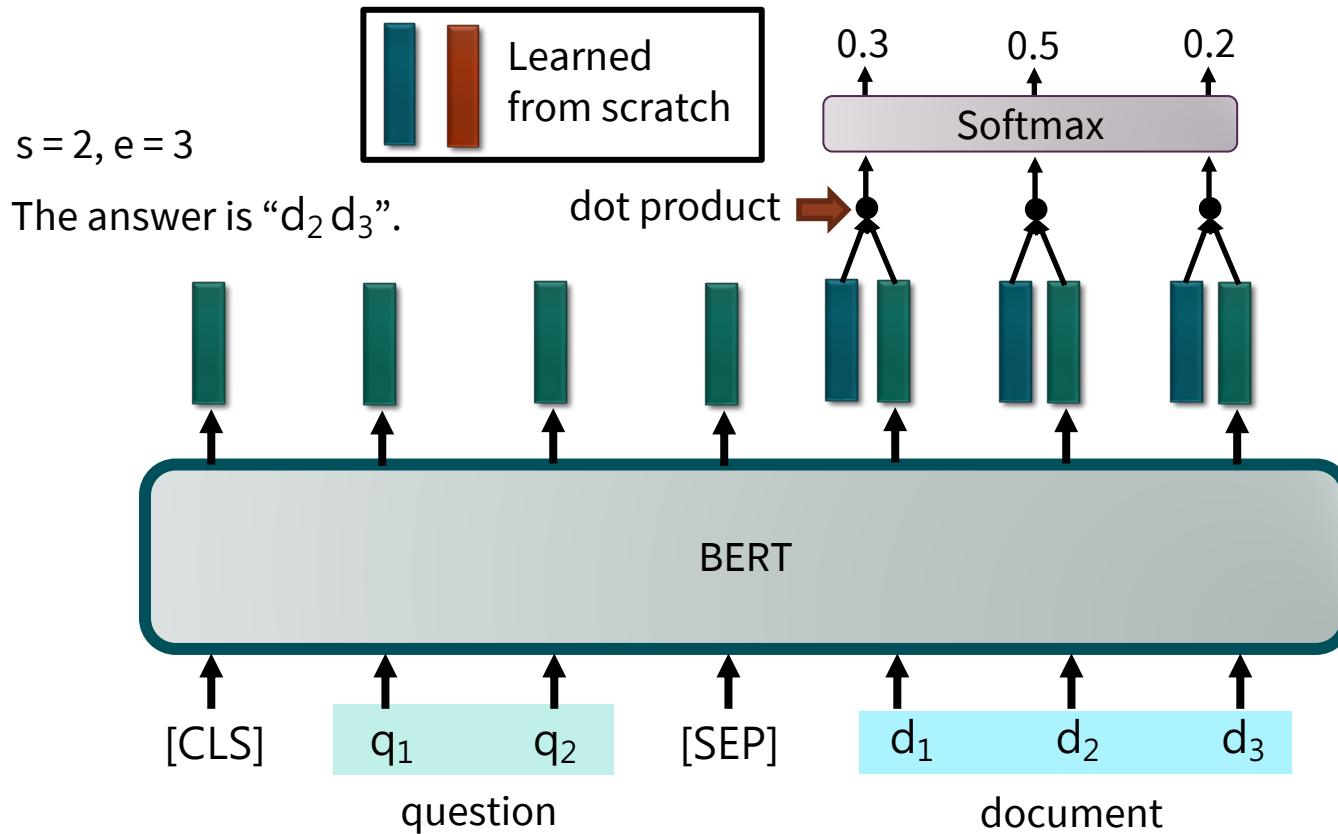
$s = 17, e = 17$

Where do water droplets collide with ice crystals to form precipitation?

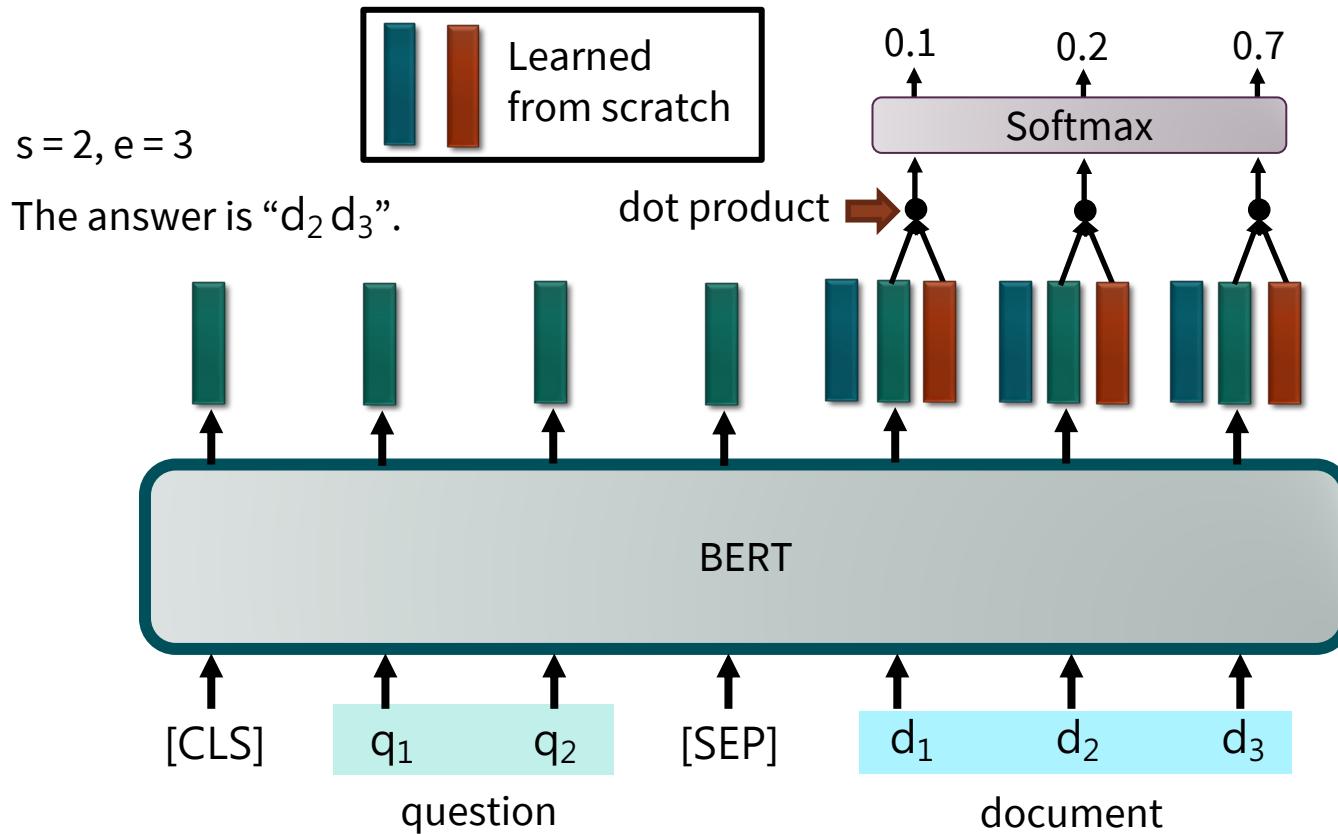
**within a cloud**

$s = 77, e = 79$

# How to use BERT – Case 4



# How to use BERT – Case 4



# Model Details

DATA: WIKIPEDIA (2.5B WORDS) + BOOKCORPUS (800M WORDS)

BATCH SIZE: 131,072 WORDS (1024 SEQUENCES \* 128 LENGTH OR 256 SEQUENCES \* 512 LENGTH)

TRAINING TIME: 1M STEPS (~40 EPOCHS)

OPTIMIZER: ADAM, 1E-4 LEARNING RATE, LINEAR DECAY

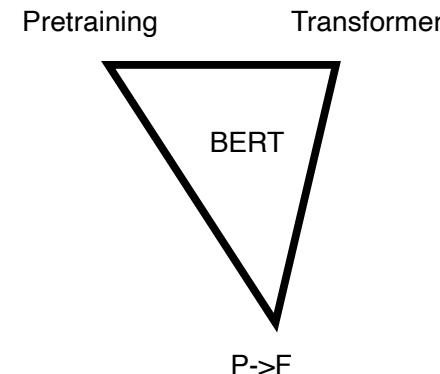
BERT-BASE: 12-LAYER, 768-HIDDEN, 12-HEAD,

- Total Parameters=110M

BERT-LARGE: 24-LAYER, 1024-HIDDEN, 16-HEAD

- Total Parameters=340M

TRAINED ON 4x4 OR 8x8 TPU SLICE FOR 4 DAYS



# BERT

BERT yielded state-of-the-art (SOTA) results on many tasks

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	Average
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	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
BERT <sub>BASE</sub>	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
BERT <sub>LARGE</sub>	<b>86.7/85.9</b>	<b>72.1</b>	<b>92.7</b>	<b>94.9</b>	<b>60.5</b>	<b>86.5</b>	<b>89.3</b>	<b>70.1</b>	<b>82.1</b>

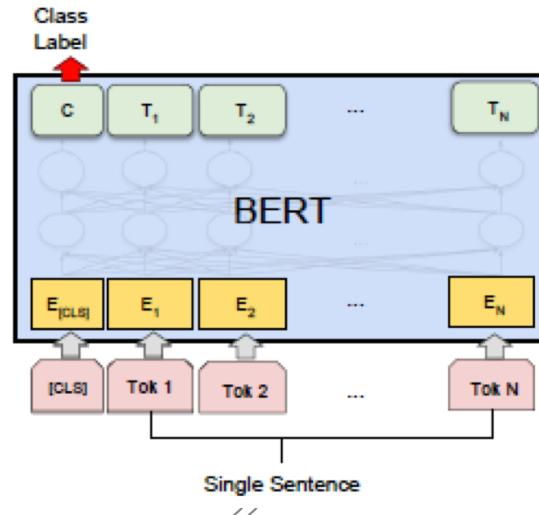
Table 1: GLUE Test results, scored by the evaluation server (<https://gluebenchmark.com/leaderboard>).

# Fine-tuning procedure

## FOR SEQUENCE-LEVEL CLASSIFICATION TASK

- Obtain the representation of the input sequence by using the final hidden state (hidden state at the position of the special token [CLS])

JUST ADD A CLASSIFICATION LAYER AND USE SOFTMAX TO CALCULATE LABEL PROBABILITIES.



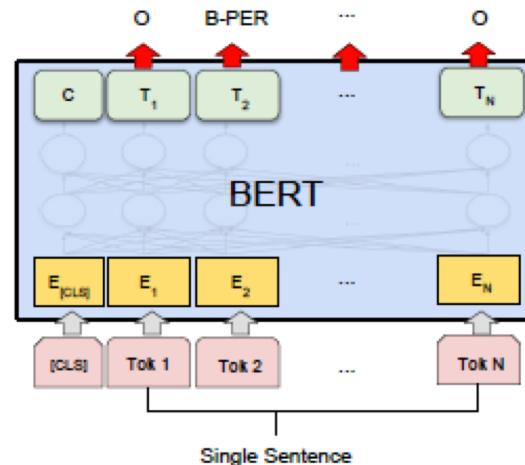
# Fine-tuning procedure

FOR SEQUENCE-LEVEL CLASSIFICATION TASK

- All of the parameters of BERT and  $W$  are fine-tuned jointly

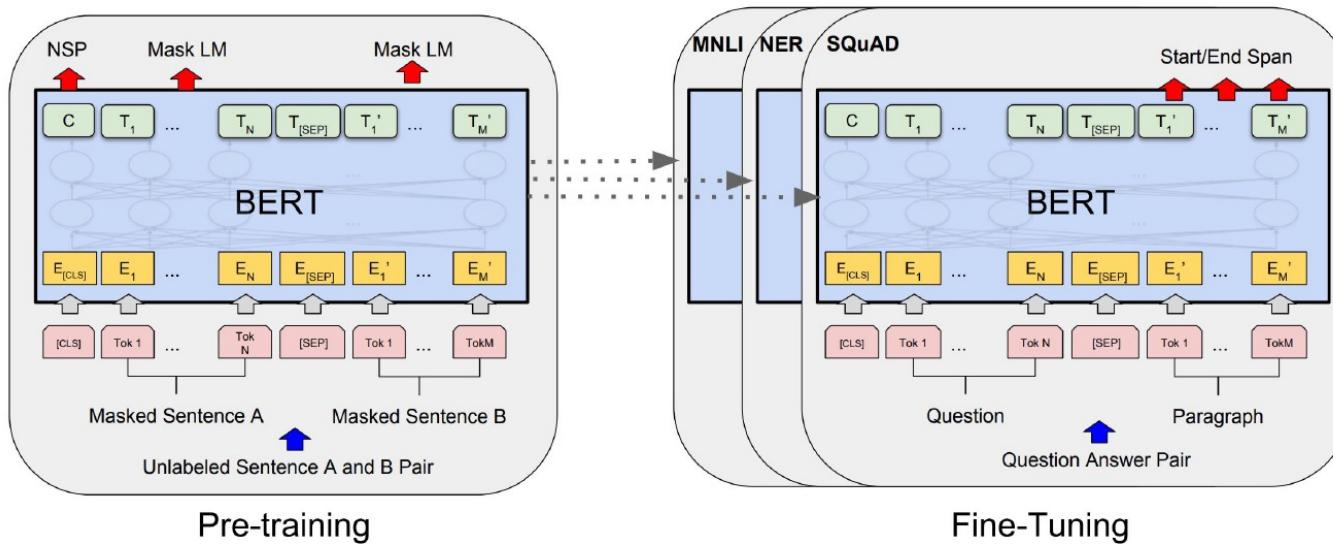
MOST MODEL HYPER-PARAMETERS ARE THE SAME AS IN PRE-TRAINING

- except the batch size, learning rate, and number of training epochs



# Fine-Tuning Procedure

- Span-level task: SQuAD v1.1
- Represent the input question and paragraph as a single packed sequence
  - The question uses the A embedding and the paragraph uses the B embedding



# Experiments

## GLUE (GENERAL LANGUAGE UNDERSTANDING EVALUATION) BENCHMARK

- Distribute canonical Train, Dev and Test splits
- Labels for Test set are not provided

### DATASETS IN GLUE:

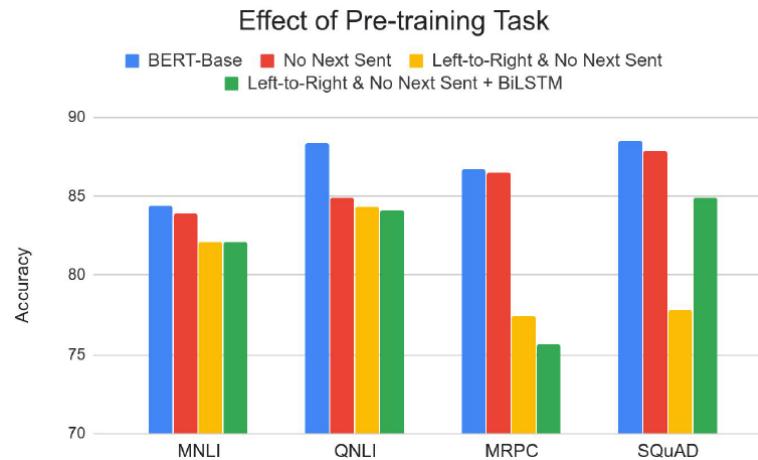
- MNLI: Multi-Genre Natural Language Inference
- QQP: Quora Question Pairs
- QNLI: Question Natural Language Inference
- SST-2: Stanford Sentiment Treebank
- CoLA: The corpus of Linguistic Acceptability
- STS-B: The Semantic Textual Similarity Benchmark
- MRPC: Microsoft Research Paraphrase Corpus
- RTE: Recognizing Textual Entailment
- WNLI: Winograd NLI

# GLUE Results

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	Average
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 <sub>BASE</sub>	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
BERT <sub>LARGE</sub>	<b>86.7/85.9</b>	<b>72.1</b>	<b>91.1</b>	<b>94.9</b>	<b>60.5</b>	<b>86.5</b>	<b>89.3</b>	<b>70.1</b>	<b>81.9</b>

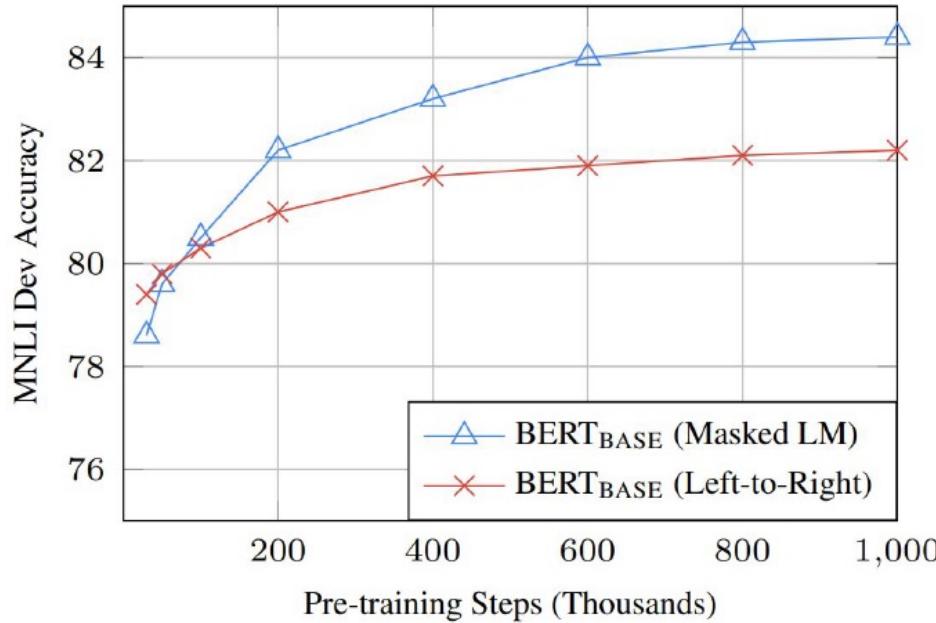
# Effect of Pre-training Task

- Masked LM (compared to left-to-right LM) is very important on some tasks, Next Sentence Prediction is important on other tasks.
- Left-to-right model does very poorly on word-level task (SQuAD), although this is mitigated by BiLSTM



# Effect of Directionality and Training Time

- Masked LM takes slightly longer to converge because we only predict 15% instead of 100%
- But absolute results are much better almost immediately



# BERT results on SQuAD

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.831	89.452
1 <small>Mar 20, 2019</small>	BERT + DAE + AoA (ensemble) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	87.147	89.474
2 <small>Mar 15, 2019</small>	BERT + ConvLSTM + MTL + Verifier (ensemble) <i>Layer 6 AI</i>	86.730	89.286
3 <small>Mar 05, 2019</small>	BERT + N-Gram Masking + Synthetic Self-Training (ensemble) <i>Google AI Language</i> <a href="https://github.com/google-research/bert">https://github.com/google-research/bert</a>	86.673	89.147
4 <small>May 21, 2019</small>	XLNet (single model) <i>XLNet Team</i>	86.346	89.133
5 <small>Apr 13, 2019</small>	SemBERT(ensemble) <i>Shanghai Jiao Tong University</i>	86.166	88.886

SQuAD 2.0

Carnegie  
Mellon  
University

## Takeaway

**BERT** is incredible for learning **contextualized embeddings** of words and using transfer learning for other tasks (e.g., classification).

Can't generate new sentences though, due to no decoders.



# Multilingual BERT

<https://arxiv.org/abs/1904.09077>

Trained on 104 languages

Task specific training  
data for English



Class 1



Class 2



Class 3



Task specific  
testing data for  
Chinese



?



?



?



?

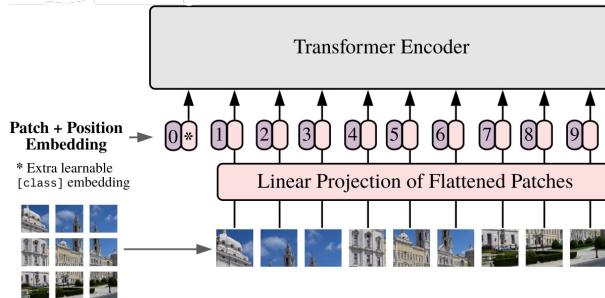


?

## Transformer-Encoders

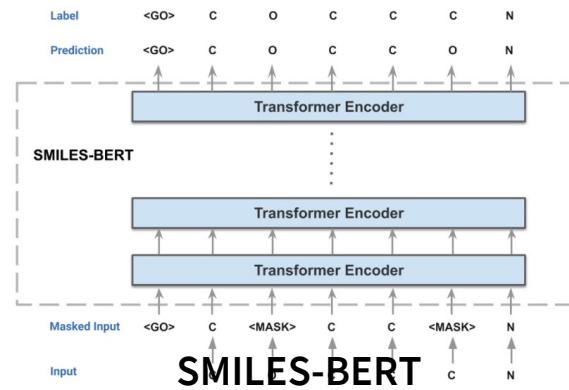
- BERT
- ALBERT (A Lite BERT ...)
- RoBERTa (A Robustly Optimized BERT ...)
- DistilBERT (small BERT)
- ELECTRA (Pre-training Text Encoders as Discriminators not Generators)
- Longformer (Long-Document Transformer)

# Transformers beyond NLP

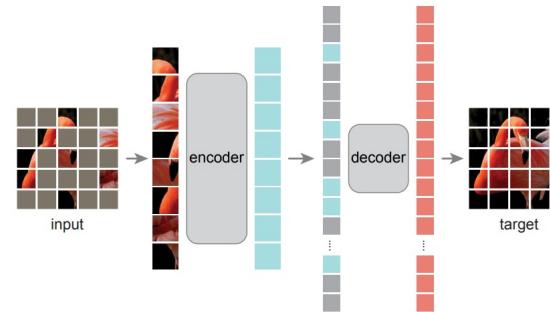


## ViT: Vision Transformer

A. Dosovitskiy et al., ICLR 2021

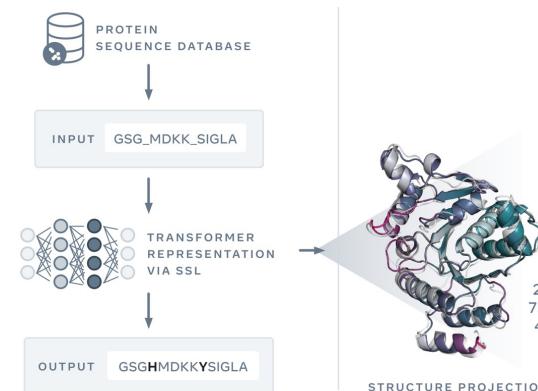


S. Wang et al., ACM-BCB 2019



## Masked autoencoder

K. He et al., CVPR 2022



## ESM-1b Transformer

Z. Lin et al., Science 2023

# Transformers in vision

Published as a conference paper at ICLR 2021

---

## AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

**Alexey Dosovitskiy<sup>\*,†</sup>, Lucas Beyer<sup>\*</sup>, Alexander Kolesnikov<sup>\*</sup>, Dirk Weissenborn<sup>\*</sup>,  
Xiaohua Zhai<sup>\*</sup>, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer,  
Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby<sup>\*,†</sup>**

<sup>\*</sup>equal technical contribution, <sup>†</sup>equal advising

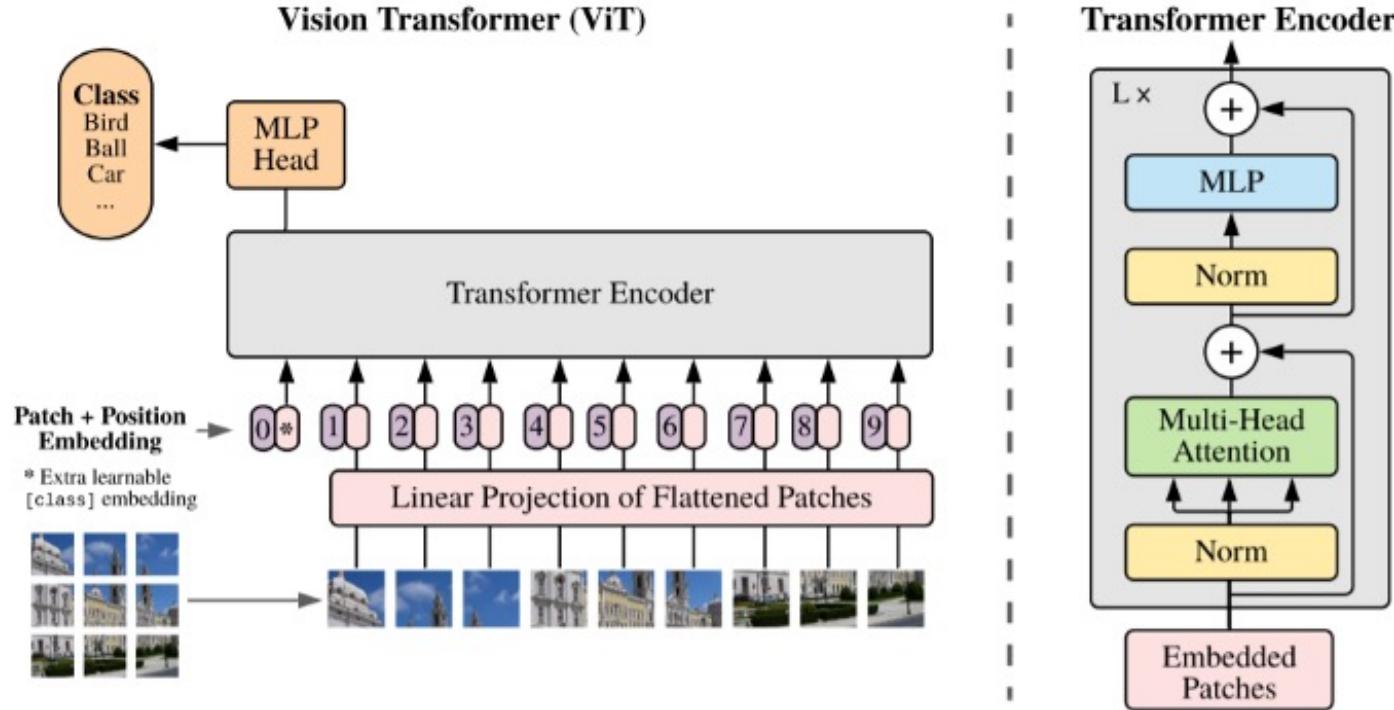
Google Research, Brain Team

{adosovitskiy, neilhoulsby}@google.com

Dosovitskiy, ICLR 2021, [https://github.com/google-research/vision\\_transformer](https://github.com/google-research/vision_transformer)

[https://www.youtube.com/watch?v=TrdevFK\\_am4](https://www.youtube.com/watch?v=TrdevFK_am4)

# Transformers in vision



Dosovitskiy, ICLR 2021, [https://github.com/google-research/vision\\_transformer](https://github.com/google-research/vision_transformer)

[https://www.youtube.com/watch?v=TrdevFK\\_am4](https://www.youtube.com/watch?v=TrdevFK_am4)

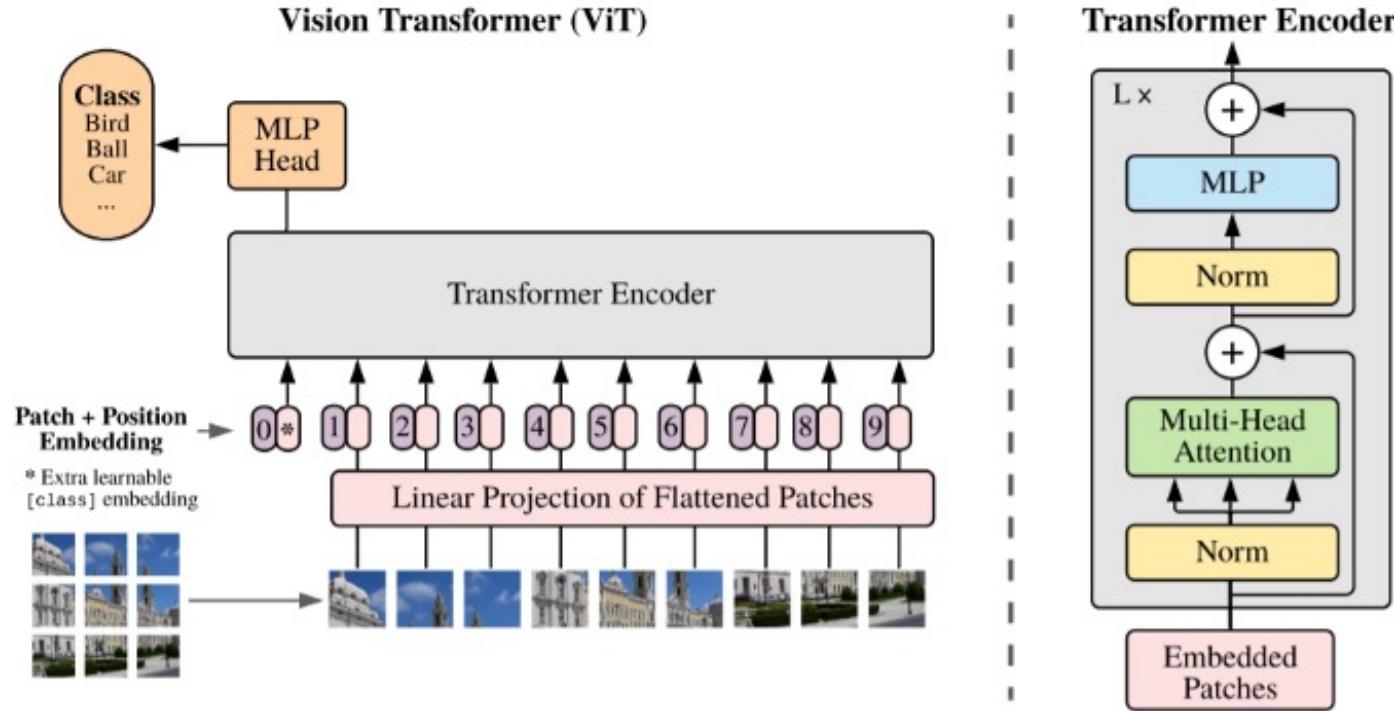
# ViT Benchmark

Model	Layers	Hidden size $D$	MLP size	Heads	Params
ViT-Base	12	768	3072	12	86M
ViT-Large	24	1024	4096	16	307M
ViT-Huge	32	1280	5120	16	632M

Table 1: Details of Vision Transformer model variants.

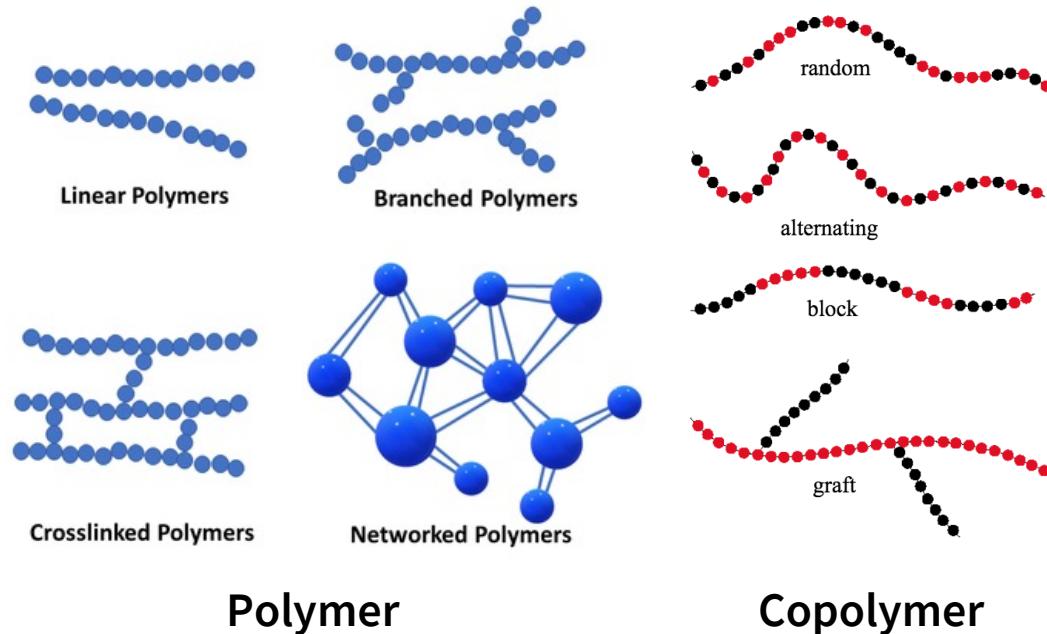
	Ours-JFT (ViT-H/14)	Ours-JFT (ViT-L/16)	Ours-I21K (ViT-L/16)	BiT-L (ResNet152x4)	Noisy Student (EfficientNet-L2)
ImageNet	<b>88.55</b> ± 0.04	87.76 ± 0.03	85.30 ± 0.02	87.54 ± 0.02	88.4/88.5*
ImageNet ReaL	<b>90.72</b> ± 0.05	90.54 ± 0.03	88.62 ± 0.05	90.54	90.55
CIFAR-10	<b>99.50</b> ± 0.06	99.42 ± 0.03	99.15 ± 0.03	99.37 ± 0.06	—
CIFAR-100	<b>94.55</b> ± 0.04	93.90 ± 0.05	93.25 ± 0.05	93.51 ± 0.08	—
Oxford-IIIT Pets	<b>97.56</b> ± 0.03	97.32 ± 0.11	94.67 ± 0.15	96.62 ± 0.23	—
Oxford Flowers-102	99.68 ± 0.02	<b>99.74</b> ± 0.00	99.61 ± 0.02	99.63 ± 0.03	—
VTAB (19 tasks)	<b>77.63</b> ± 0.23	76.28 ± 0.46	72.72 ± 0.21	76.29 ± 1.70	—
TPUv3-core-days	2.5k	0.68k	0.23k	9.9k	12.3k

# Transformers in vision



# Transformer for Chemical Compounds?

1. ACQUIRE STRING REPRESENTATION AND TOKENIZE THE COMPOUNDS?
2. INTEGRATE MORE INFORMATION (E.G., 3D GEOMETRIC)?



<https://compositeskn.org/RPC/>

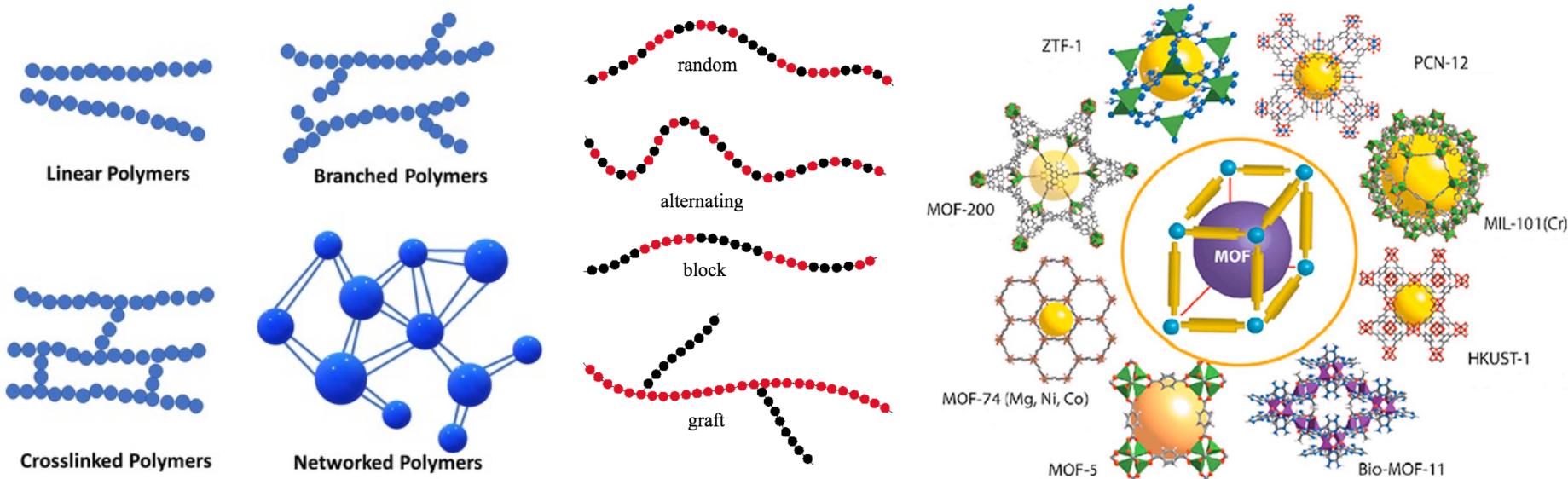
<https://www.e-education.psu.edu/>

Carnegie  
Mellon  
University

Y. Zhao et al., *Energy Storage Mater.* 2016

## Copolymer

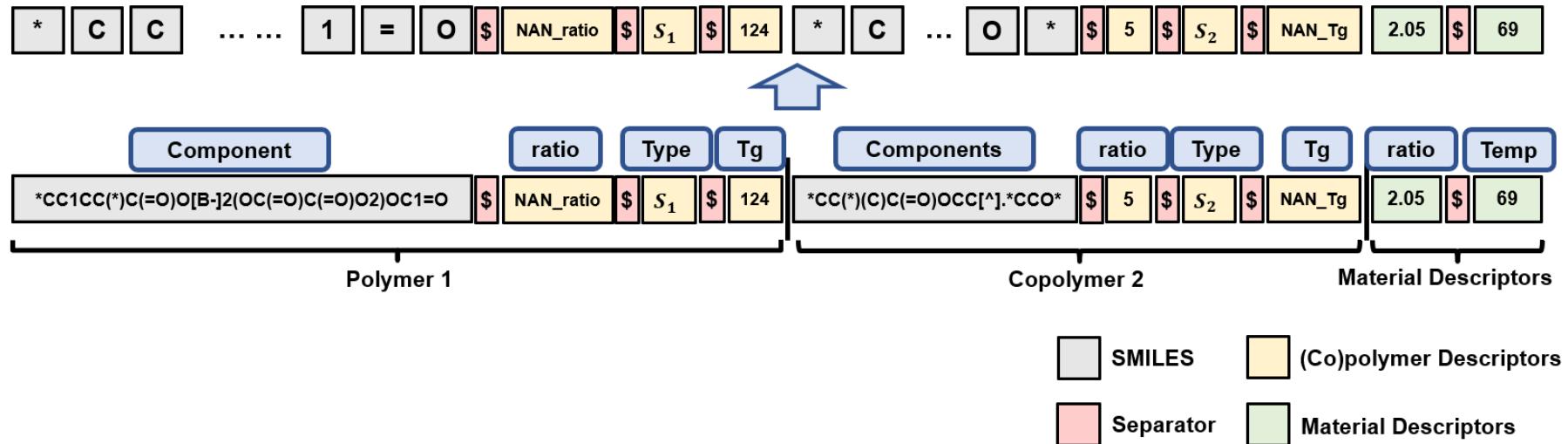
## Metal-organic framework (MOF)



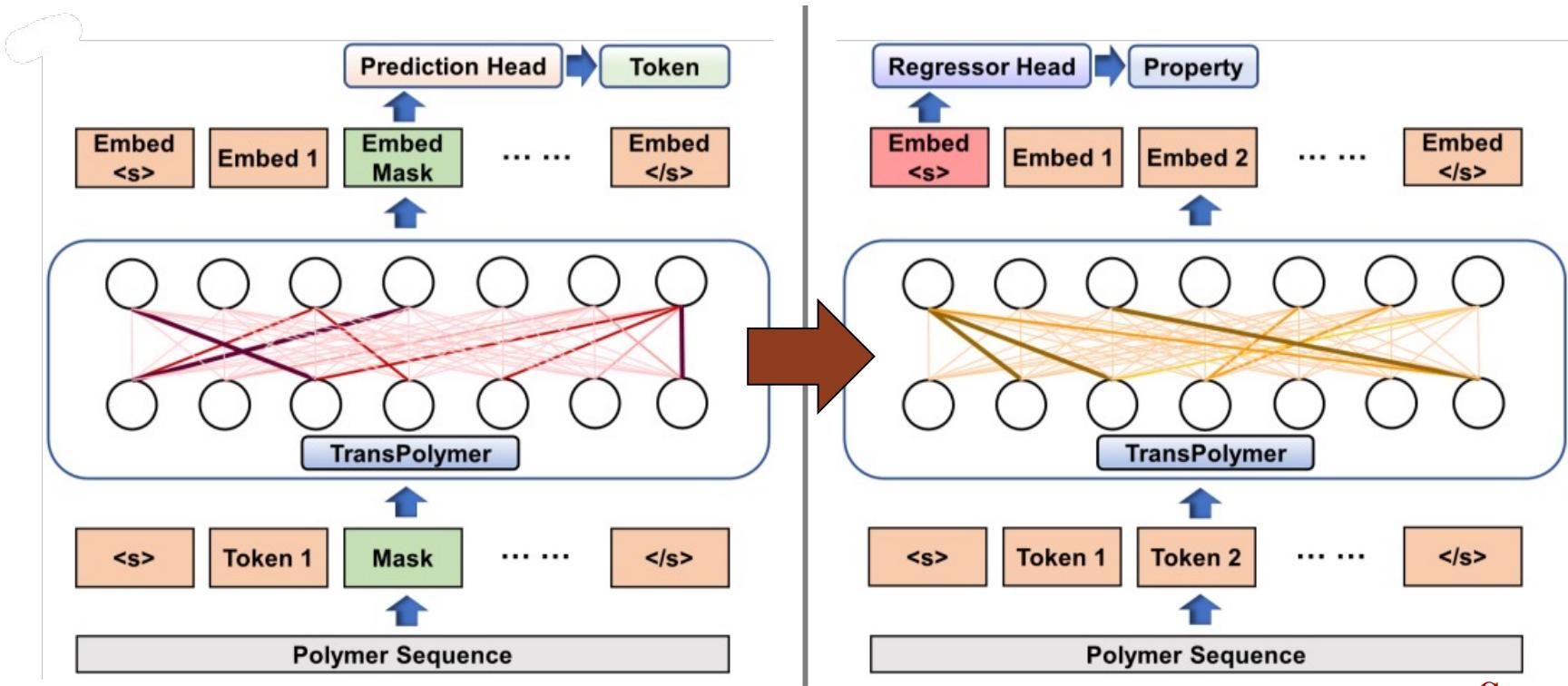
## Polymer

# TransPolymer: Tokenization

A TRANSFORMER-BASED LANGUAGE MODEL FOR POLYMERS.



# TransPolymer: Pre-train & Fine-tune



PI1M: ~1M POLYMER DATA AUGMENTED TO ~5M TOTAL PRE-TRAINING DATA

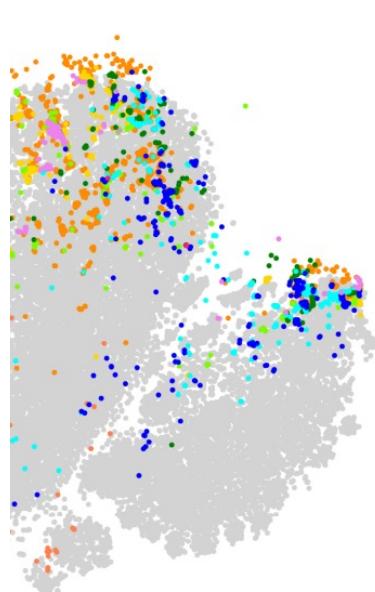
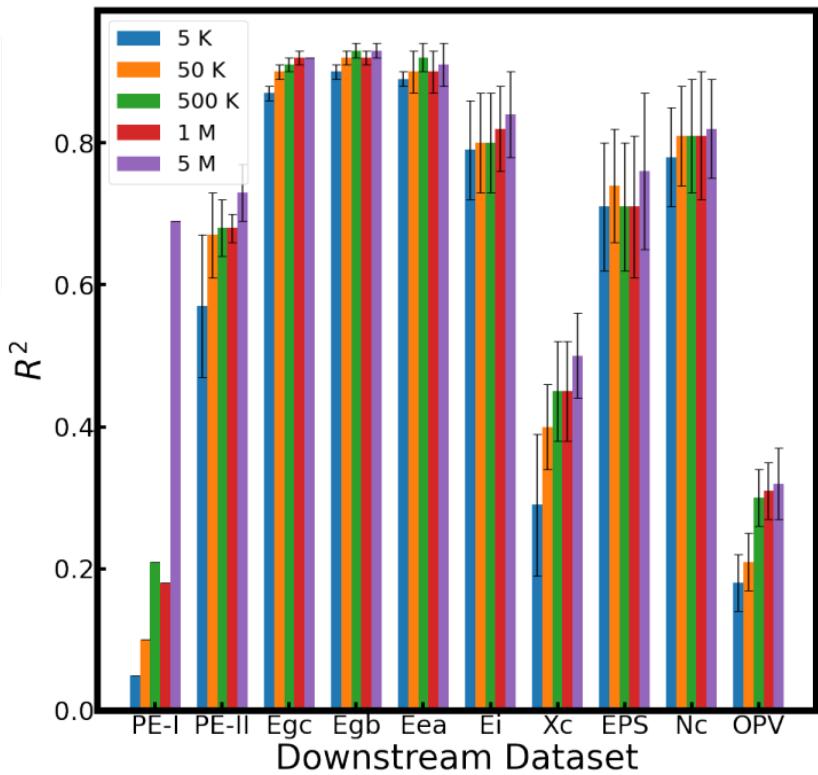
# Polymer Property Prediction

Dataset	Property	# Data	vs. best baselines		vs. TransPolymer <sub>scratch</sub>	
			RMSE (↓)	R <sup>2</sup> (↑)	RMSE (↓)	R <sup>2</sup> (↑)
PE-I	conductivity	9185	-30.9%	+0.37	-52.2%	+0.39
PE-II	conductivity	271	-3.17%	+0.01	-23.8%	+0.19
Egc	bandgap (chain)	3380	-8.33%	+0.02	-30.2%	+0.08
Egb	bandgap (bulk)	561	-5.45%	+0.02	-14.8%	+0.03
Eea	electron affinity	368	0.00%	+0.01	-11.1%	+0.02
Ei	ionization energy	370	-7.14%	+0.07	-15.2%	+0.06
Xc	crystallization tendency	432	-20.1%	+0.50	-17.6%	+0.23
EPS	dielectric constant	382	-1.89%	+0.05	-11.9%	+0.06
Nc	refractive index	382	0.00%	+0.03	0.00%	+0.02
OPV	power conversion efficiency	1203	0.00%	+0.05	-8.57%	+0.13
Average	N/A	N/A	-7.70%	+0.11	-18.5%	+0.12

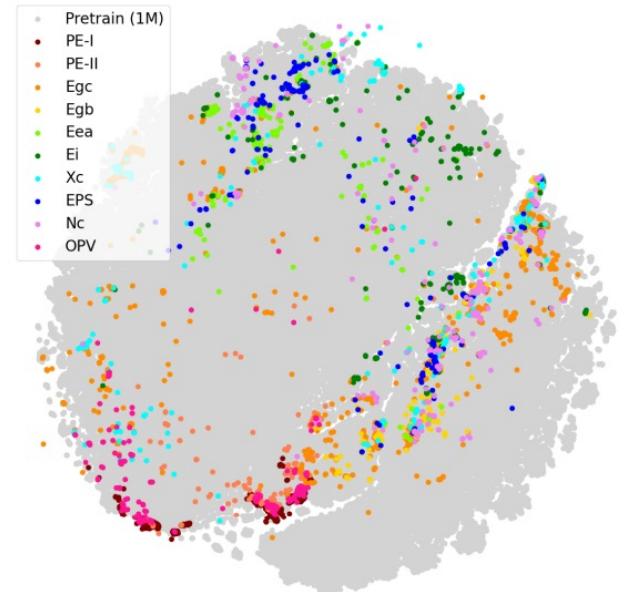
N. S. Schauser et al., *Chem. Mater.* 2021. K. Hatakeyama-Sato et al., *J. Am. Chem. Soc.* 2020.

C. Kuenneth et al., *Patterns* 2021. S. Nagasawa et al., *J. Phys. Chem. Lett.* 2018

# Number of Pretrain Data

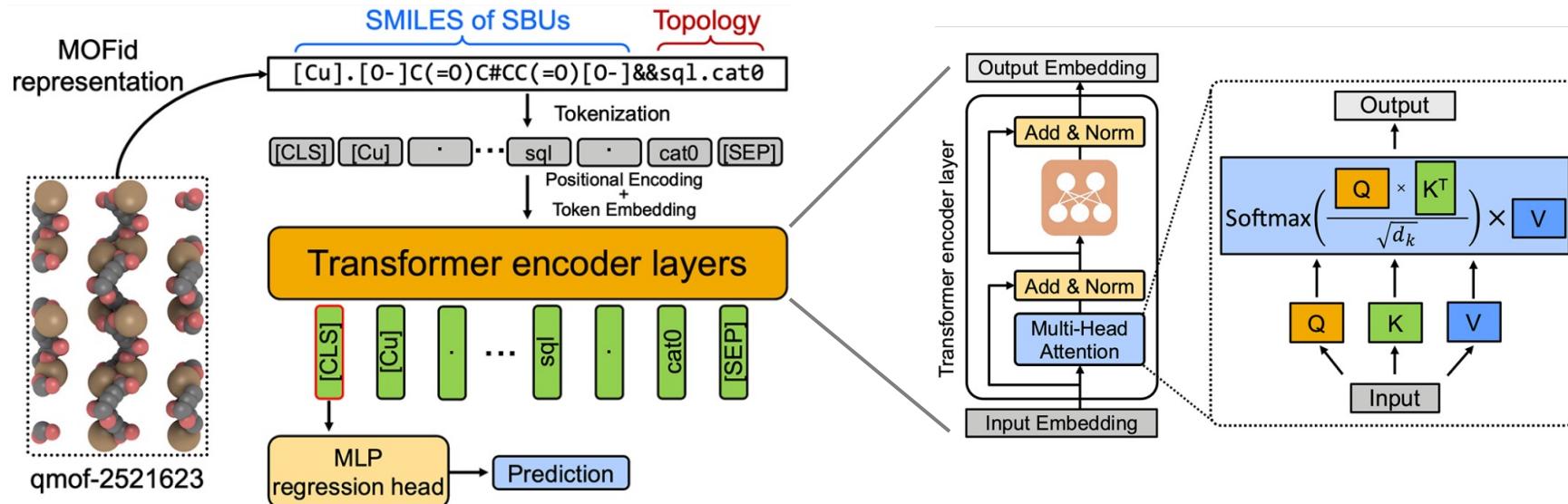


5 K pretrain data



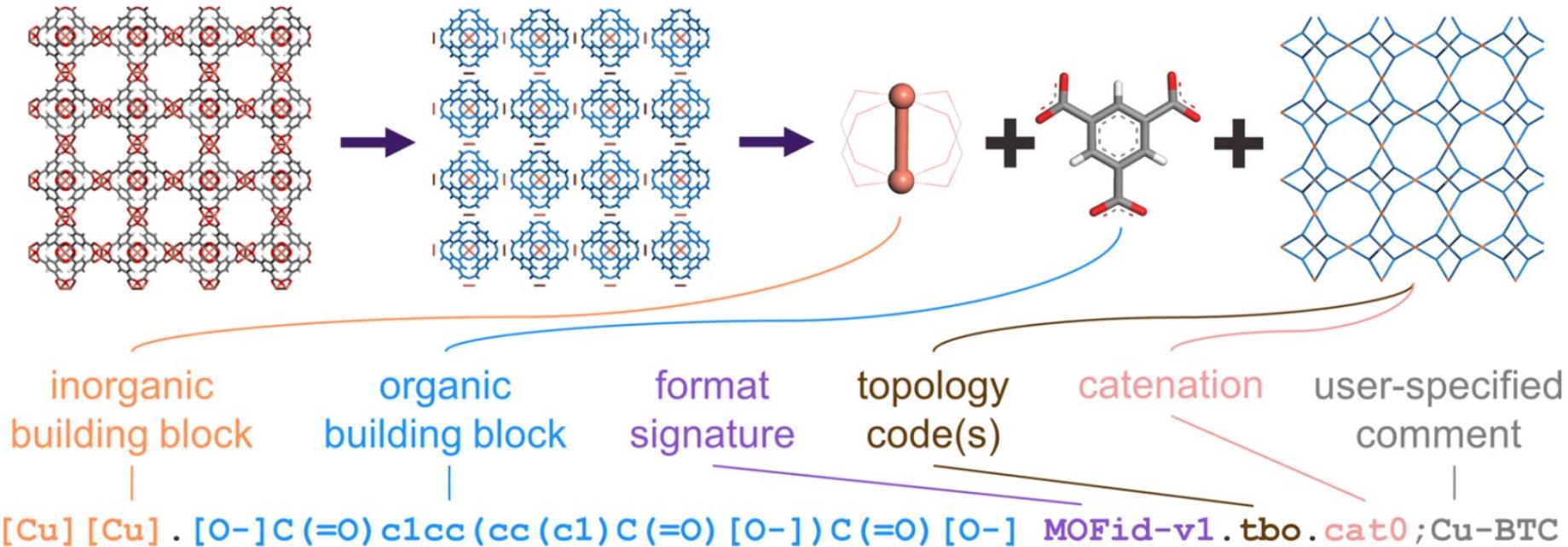
1M pretrain data

# MOFormer

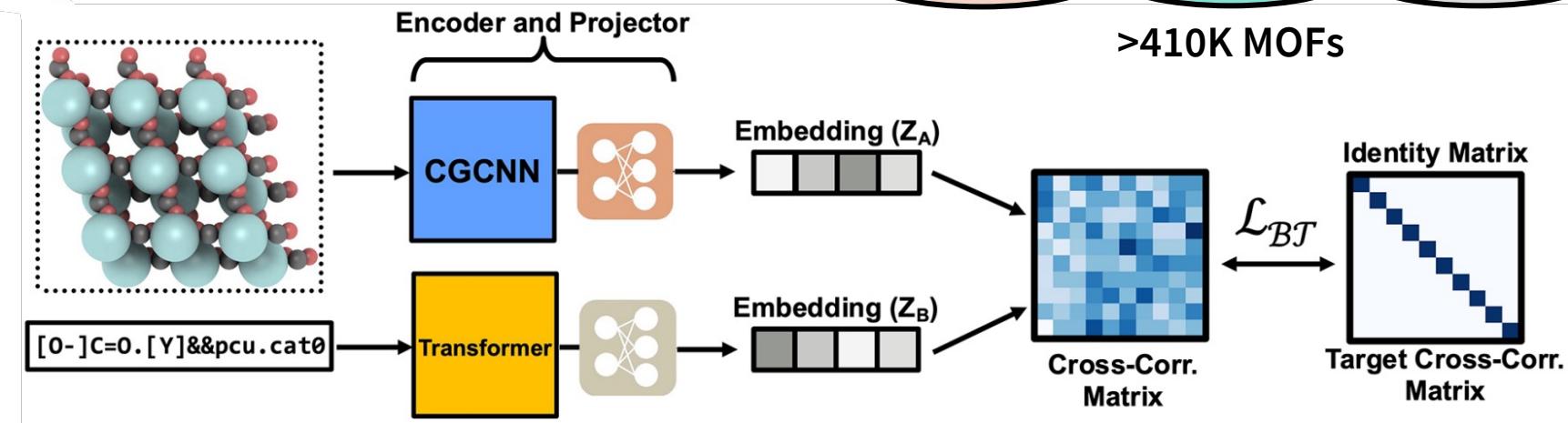
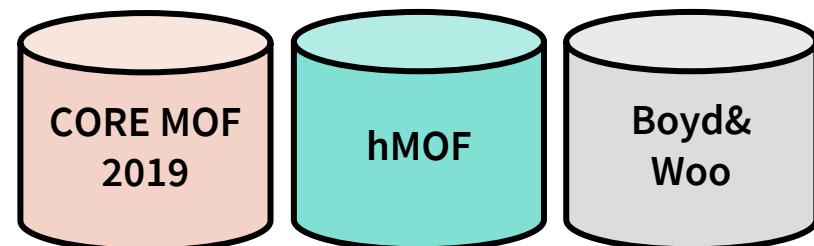


Z. Cao, R. Magar, Y. Wang et al., *J. Am. Chem. Soc.* 2023

# MOFid



# Multimodal pre-training

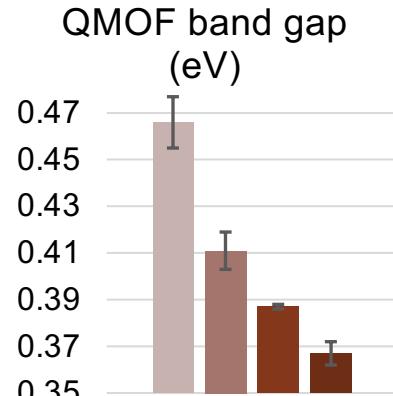


$$L_{BT} \stackrel{\Delta}{=} \sum_i (1 - C_{ii})^2 + \lambda \sum_i \sum_{j \neq i} C_{ij}^2$$

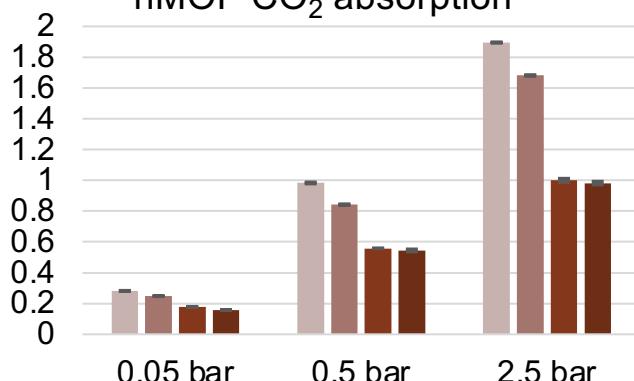
$$C_{ij} \stackrel{\Delta}{=} \frac{\sum_b Z_{b,i}^A Z_{b,j}^B}{\sqrt{\sum_b (Z_{b,i}^A)^2} \sqrt{\sum_b (Z_{b,j}^B)^2}}$$

# Band Gap & Gas Adsorption Prediction

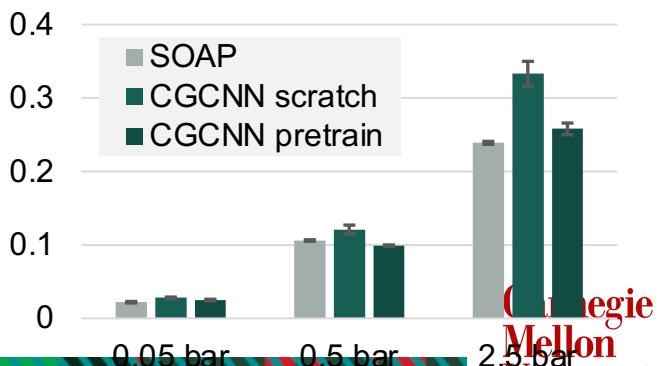
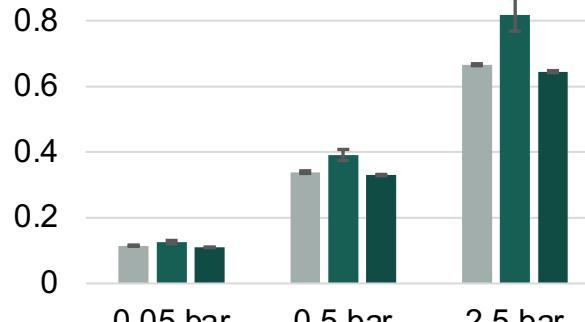
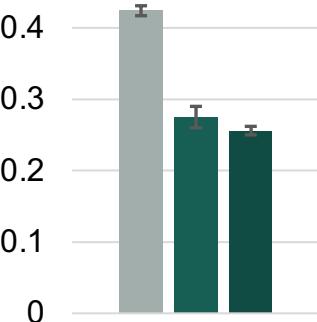
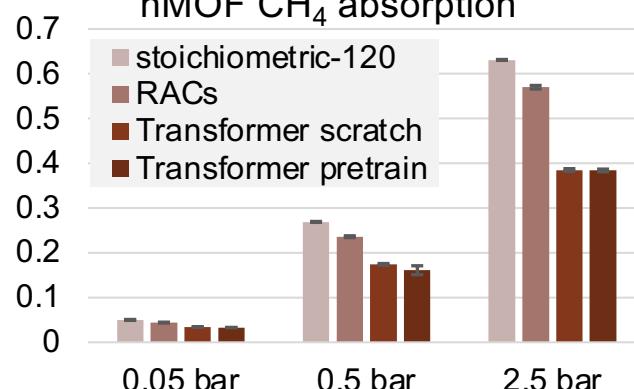
QMOF band gap  
(eV)



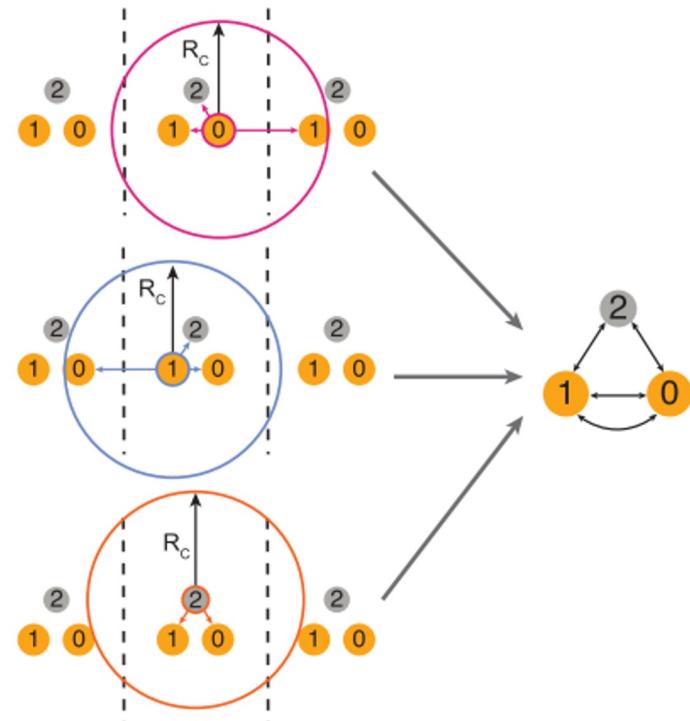
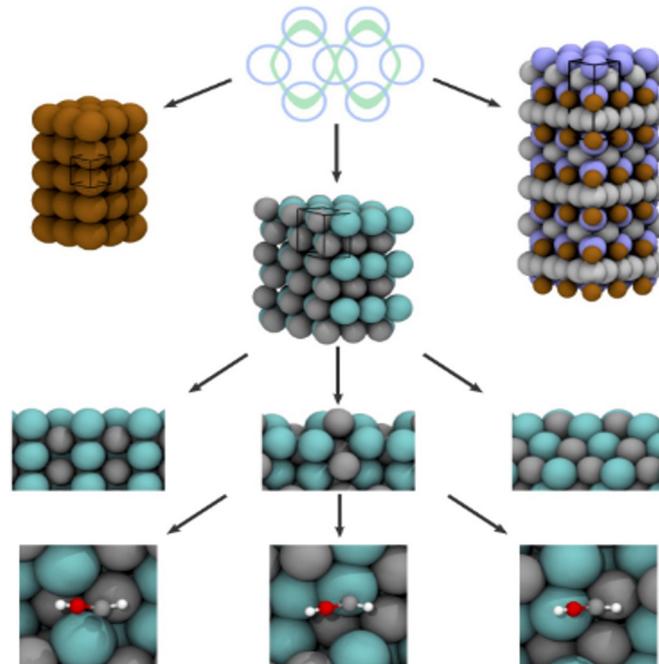
hMOF CO<sub>2</sub> absorption



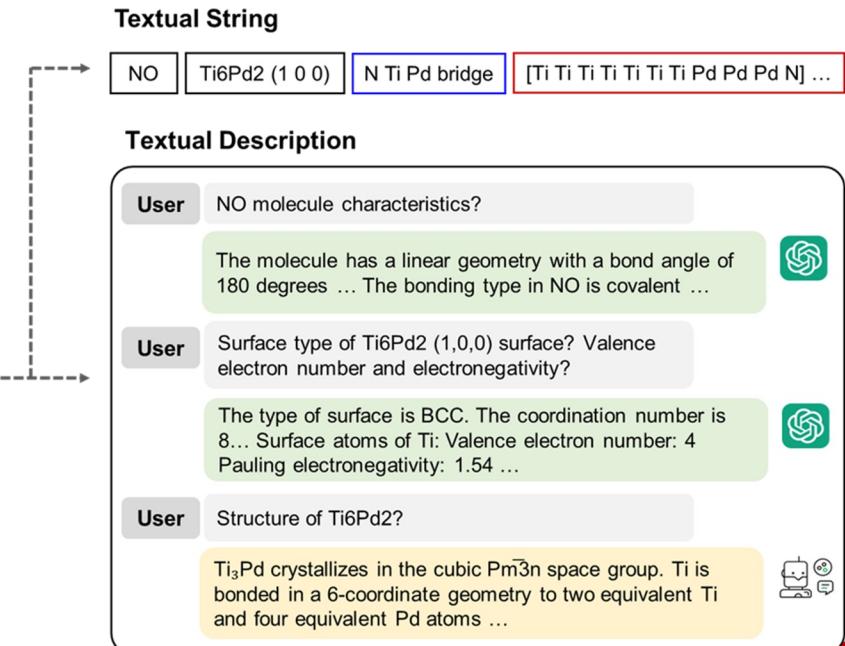
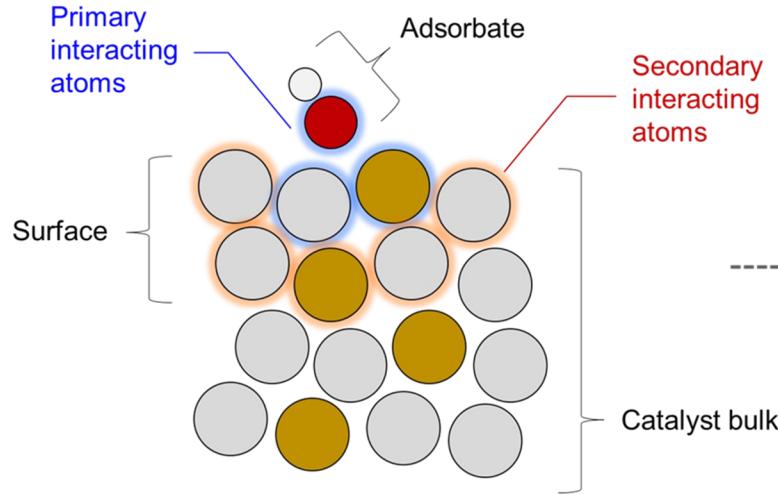
hMOF CH<sub>4</sub> absorption



# Graph representation requires a precise understanding of atomic coordinates



# Seamless Incorporation of Observable Features through Language Format



# Framework

Structures from OC20  
dataset

Textual data

Fine-tune a pre-  
trained LLM

Energy prediction

String / Description

<s> Token 1 Token 2 ..... </s>

Tokenization

Positional encoding & Token embedding

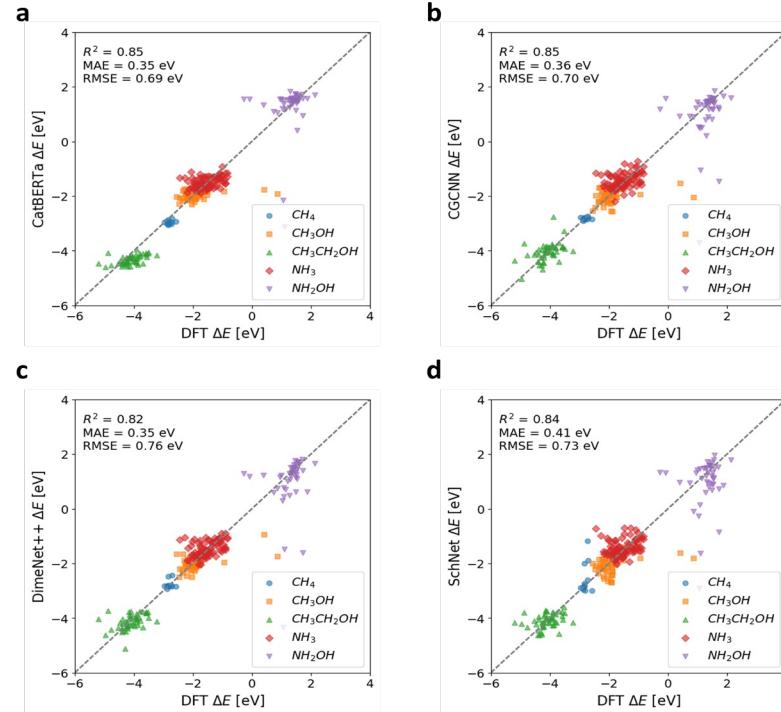
Pretrained RoBERTa

Emb <s> Emb 1 Emb 2 ..... Emb </s>

Positional encoding  
& Token embedding

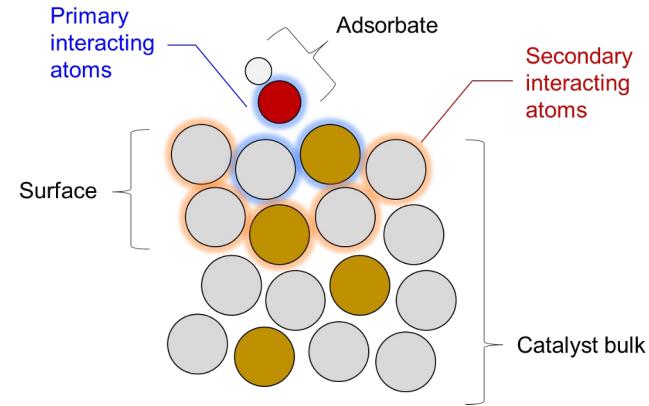
Regressor head → Energy

# Identifying High-Accuracy Subsets with CatBERTa and GNNs



# Exploring Features through Textual Format Input (1)

- Interacting atoms serve as impactful features
- Atomic properties and bonding distance do not universally enhance accuracy



No.	Textual String Format	MAE [eV]	
		Subset	Total
String 1	< s > adsorbate SMILES < /s > bulk composition (Miller index) < /s >	0.44	0.85
String 2	String 1 + [primary interacting atoms, site type]	0.37	0.79
String 3	String 1 + [primary interacting atoms, site type] [atomic properties]	0.47	0.79
String 4	String 1 + [primary interacting atoms, site type] [secondary interacting atoms]	0.35	<b>0.75</b>
String 5	String 1 + [primary interacting atoms, site type, bond distance] [secondary interacting atoms]	<b>0.34</b>	0.77

# Exploring Features through Textual Format Input (2)

- Description 1 contains general information and does not derive from adsorption modeling studies
- Description 2 includes properties utilized in adsorption modeling
- Accuracy: description 2 > description 1

No.	Entity	Input Feature	Tool	MAE [eV]	
				Subset	Total
Description 1	System	Adsorbate SMILES, bulk composition, Miller index, primary interacting atoms, adsorption site type	Pymatgen	0.50	0.84
	Adsorbate	Bonding type, angle, length, molecule size, dipole moment, orbital characteristics	ChatGPT		
	Catalyst	Composition, space group, bonding geometry, length, atom arrangement	RobоСrystallographer		
Description 2	System	Adsorbate SMILES, bulk composition, Miller index, primary interacting atoms, adsorption site type	Pymatgen	0.43	0.79
	Adsorbate	Central atom, coordination number	ChatGPT		
	Catalyst	Surface type, coordination number, valence electron number, electronegativity	ChatGPT		

# The attention score can provide interpretability

First hidden layer	Last hidden layer
<p>NH3 VCr3 (2 1 0) [N Cr Cr bridge [Cr Cr Cr Cr V V V N] [Cr Cr Cr V V V N]</p> <p>Adsorbate NH3 is adsorbed on the catalytic surface VCr3 with a Miller Index of (2, 1, 0). The N atom of the adsorbate is placed on the bridge site and is binding to the catalytic surface atoms Cr, Cr. The NH3 molecule is a triatomic molecule consisting of one nitrogen atom and three hydrogen atoms. It has a trigonal pyramidal molecular geometry, with a bond angle of 107.5 degrees and a bond length of 1.01 Å. The nitrogen atom is sp<sup>3</sup> hybridized, while the hydrogen atoms have a 1s orbital. The bonding in NH3 is primarily covalent, with the nitrogen atom donating one of its valence electrons to each of the three hydrogen atoms. The remaining lone pair of electrons on the nitrogen atom makes it a Lewis base capable of forming hydrogen bonds with other molecules. The dipole moment of NH3 is 1.47 D and it arises from the unequal sharing of electrons between nitrogen and hydrogen atoms, resulting in a partial negative charge on nitrogen and partial positive charges on hydrogen atoms. Overall, NH3 is a polar molecule due to its asymmetric molecular geometry and it can form strong interactions with metal catalysts in various catalytic reactions. VCr3 is alpha bismuth trifluoride structured and crystallizes in the cubic Fm3m space group. V is bonded in a distorted body-centered cubic geometry to fourteen Cr atoms. There are eight shorter (2.48 Å) and six longer (2.66 Å) V-Cr bond lengths. There are two inequivalent Cr sites. In the first Cr site, Cr is bonded in a distorted body-centered cubic geometry to four equivalent V and four equivalent Cr atoms. All Cr-Cr bond lengths are 2.48 Å. In the second Cr site, Cr is bonded in a distorted body-centered cubic geometry to six equivalent V and eight equivalent Cr atoms.</p>	<p>NH3 VCr3 (2 1 0) [N Cr Cr bridge [Cr Cr Cr Cr V V V N] [Cr Cr Cr V V V N]</p> <p>Adsorbate NH3 is adsorbed on the catalytic surface VCr3 with a Miller Index of (2, 1, 0). The N atom of the adsorbate is placed on the bridge site and is binding to the catalytic surface atoms Cr, Cr. The NH3 molecule is a triatomic molecule consisting of one nitrogen atom and three hydrogen atoms. It has a trigonal pyramidal molecular geometry, with a bond angle of 107.5 degrees and a bond length of 1.01 Å. The nitrogen atom is sp<sup>3</sup> hybridized, while the hydrogen atoms have a 1s orbital. The bonding in NH3 is primarily covalent, with the nitrogen atom donating one of its valence electrons to each of the three hydrogen atoms. The remaining lone pair of electrons on the nitrogen atom makes it a Lewis base capable of forming hydrogen bonds with other molecules. The dipole moment of NH3 is 1.47 D and it arises from the unequal sharing of electrons between nitrogen and hydrogen atoms, resulting in a partial negative charge on nitrogen and partial positive charge on hydrogen atoms. NH3 is a polar molecule due to its asymmetric molecular geometry and it can form strong interactions with metal catalysts in various catalytic reactions. VCr3 is alpha bismuth trifluoride structured and crystallizes in the cubic Fm3m space group. V is bonded in a distorted body-centered cubic geometry to fourteen Cr atoms. There are eight shorter (2.48 Å) and six longer (2.66 Å) V-Cr bond lengths. There are two inequivalent Cr sites. In the first Cr site, Cr is bonded in a distorted body-centered cubic geometry to four equivalent V and four equivalent Cr atoms. All Cr-Cr bond lengths are 2.48 Å. In the second Cr site, Cr is bonded in a distorted body-centered cubic geometry to six equivalent V and eight equivalent Cr atoms.</p>
<p>Adsorbate NH3 is adsorbed on the catalytic surface VCr3 with a Miller Index of (2, 1, 0). The N atom of the adsorbate is placed on the bridge site and is binding to the catalytic surface atoms Cr, Cr. The central atom in the NH3 molecule is nitrogen (N). Nitrogen (N) can form a maximum of 4 bonds. In NH3, nitrogen forms only 3 bonds with hydrogen (H). Therefore, the difference between the maximum number of bonds the central atom can have and the actual number of bonds it has is 1. The VCr3 crystal with Miller Index (2, 1, 0) has a surface of FCC (Face-Centered Cubic) type. The coordination number for this surface is 12. To determine the valence electron number and Pauling electronegativity of the surface atoms in VCr3, we first need to find the atomic number and the electron configuration of the elements involved. The atomic number of Vanadium (V) is 23. Its electron configuration is [Ar] 3d<sup>3</sup> 4s<sup>2</sup>. The valence electron number is 5 (3d<sup>3</sup> 4s<sup>2</sup>). The atomic number of Chromium (Cr) is 24. Its electron configuration is [Ar] 3d<sup>5</sup> 4s<sup>1</sup>. The valence electron number is also 5 (3d<sup>5</sup> 4s<sup>1</sup>). So, both Vanadium and Chromium have a valence electron number of 5. The Pauling electronegativity of Vanadium is 1.63. The Pauling electronegativity of Chromium is 1.66. Therefore, the valence electron number and Pauling electronegativity of the surface atoms in VCr3 are both 5 and 1.63.</p>	<p>Adsorbate NH3 is adsorbed on the catalytic surface VCr3 with a Miller Index of (2, 1, 0). The N atom of the adsorbate is placed on the bridge site and is binding to the catalytic surface atoms Cr, Cr. The central atom in the NH3 molecule is nitrogen (N). Nitrogen (N) can form a maximum of 4 bonds. In NH3, nitrogen forms only 3 bonds with hydrogen (H). Therefore, the difference between the maximum number of bonds the central atom can have and the actual number of bonds it has is 1. The VCr3 crystal with Miller Index (2, 1, 0) has a surface of FCC (Face-Centered Cubic) type. The coordination number for this surface is 12. To determine the valence electron number and Pauling electronegativity of the surface atoms in VCr3, we first need to find the atomic number and the electron configuration of the elements involved. The atomic number of Vanadium (V) is 23. Its electron configuration is [Ar] 3d<sup>3</sup> 4s<sup>2</sup>. The valence electron number is 5 (3d<sup>3</sup> 4s<sup>2</sup>). The atomic number of Chromium (Cr) is 24. Its electron configuration is [Ar] 3d<sup>5</sup> 4s<sup>1</sup>. The valence electron number is also 5 (3d<sup>5</sup> 4s<sup>1</sup>). So, both Vanadium and Chromium have a valence electron number of 5. The Pauling electronegativity of Vanadium is 1.63. The Pauling electronegativity of Chromium is 1.66. Therefore, the valence electron number and Pauling electronegativity of the surface atoms in VCr3 are both 5 and 1.63.</p>

# The capability of latent space capturing features

