

# Multimodal Language Models for Accelerating Scientific Discovery

Xifeng Yan

Computer Science  
University of California, Santa Barbara

# Outline

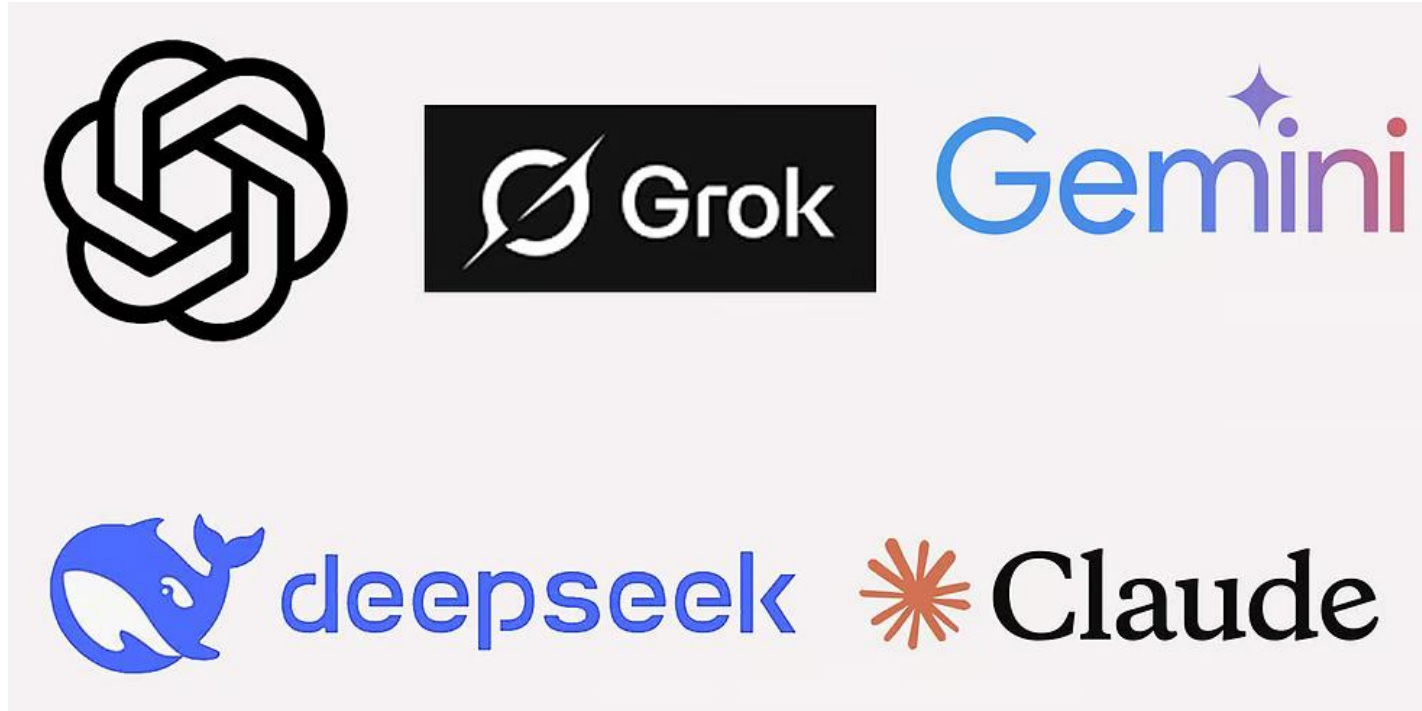
Foundation Model / LLMs

- Transformer-based Time Series Forecasting

Multimodality

- Transfer Learning: Change Modality
- More Modalities: Multimodal Language Models
- Material Science: Structure Generation

# The Success of Large Language Models (LLMs)

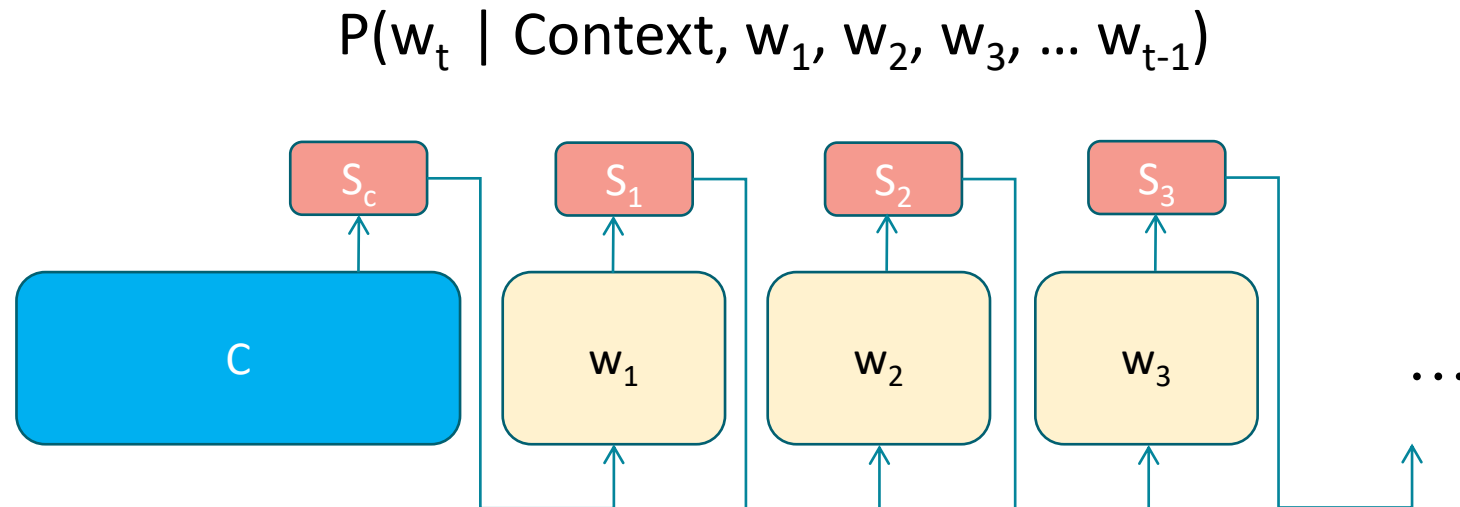


The success has been extended to computer vision, time series, multimodal data, etc.  
and now scientific discovery!

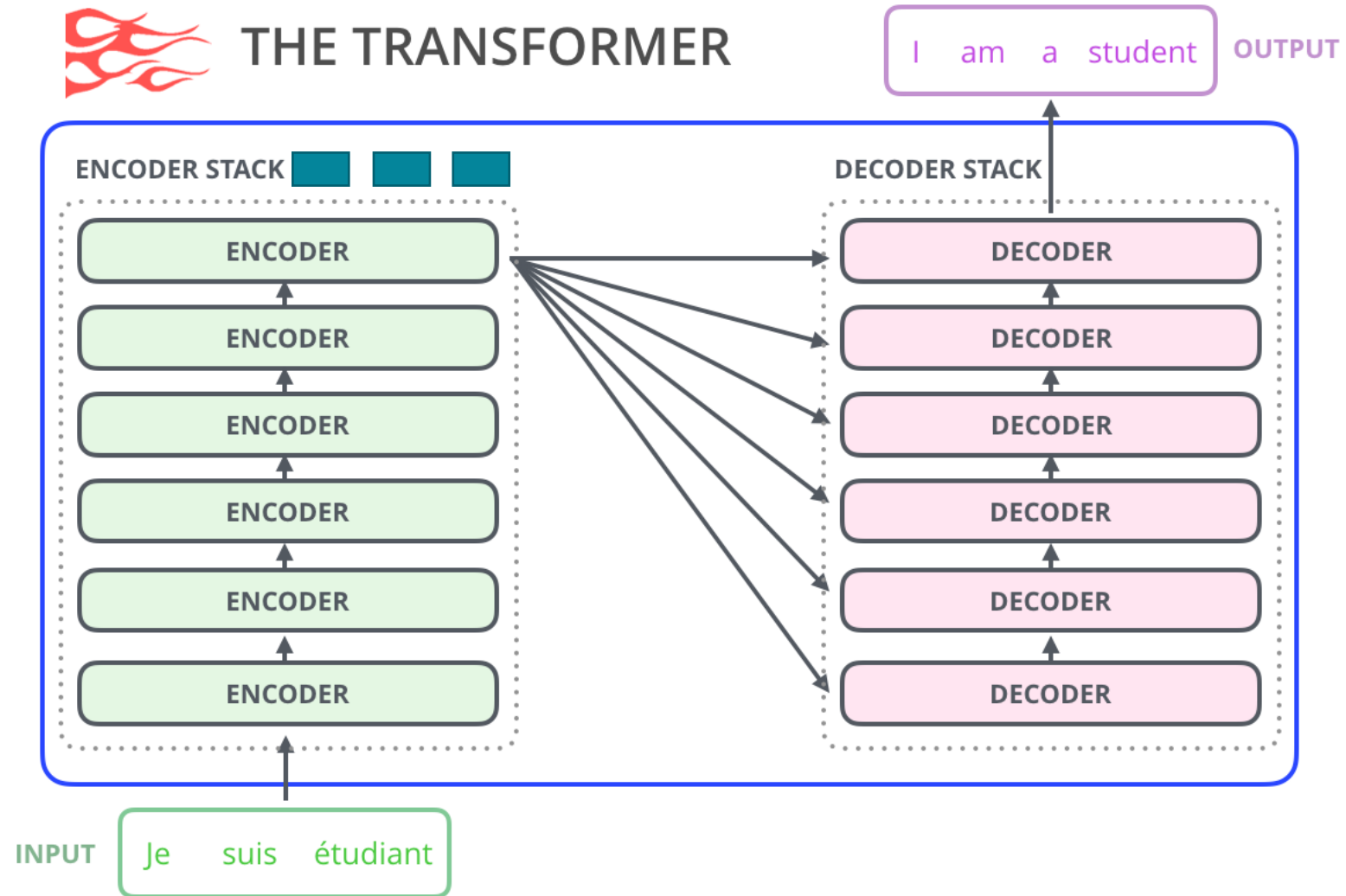
# Generative Language Models

Given a context (prompt)  $C$ , what you are going to say?

Basically, given the context, generate the first word, then given the context and the first word, generate the next word, and iterate. That's it!

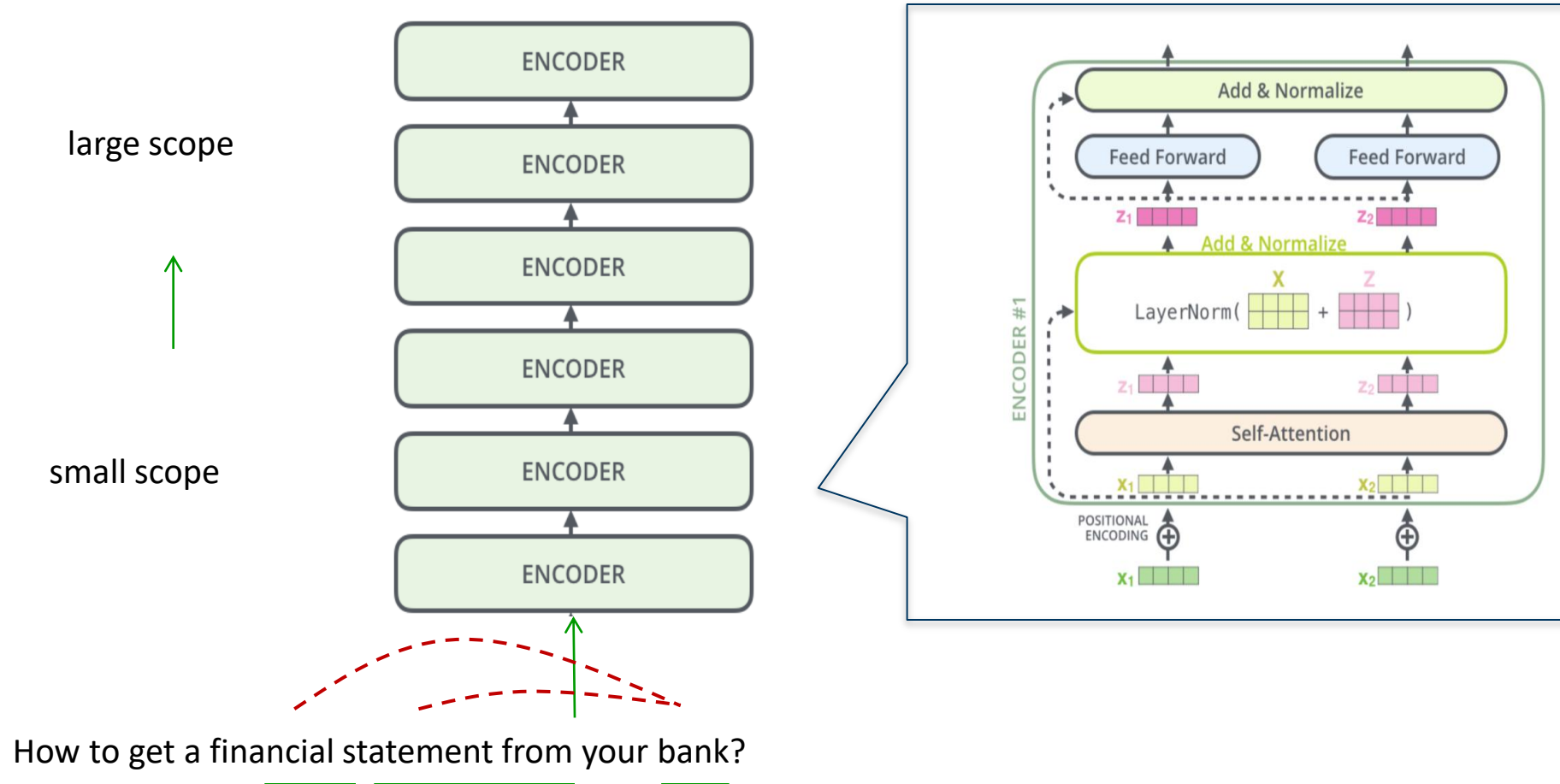


# Transformer: The Foundation Model Behind LLMs



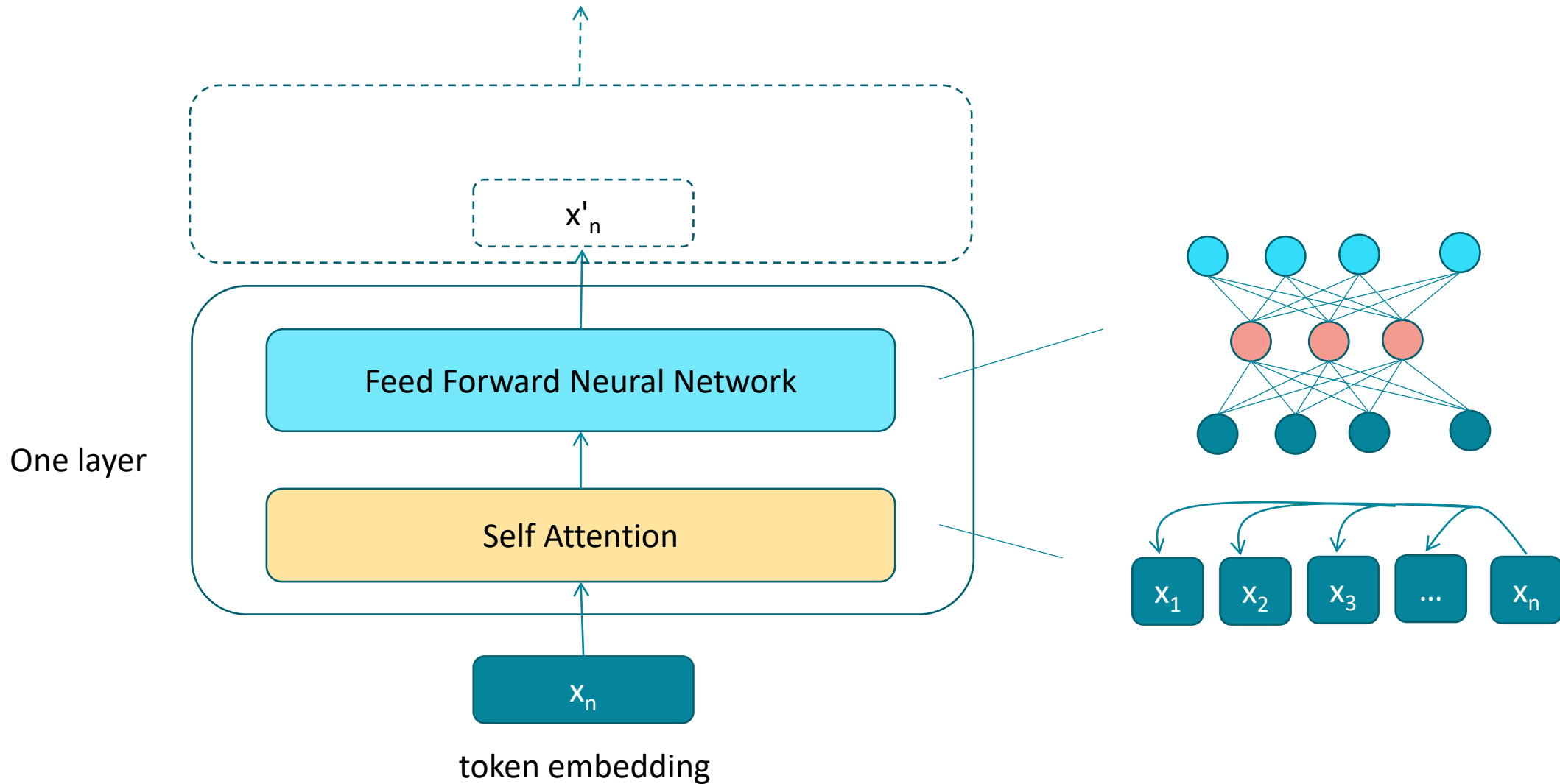
from The Illustrated GPT-2 by Jay Alammar

# Transformer: Attention is All You Need

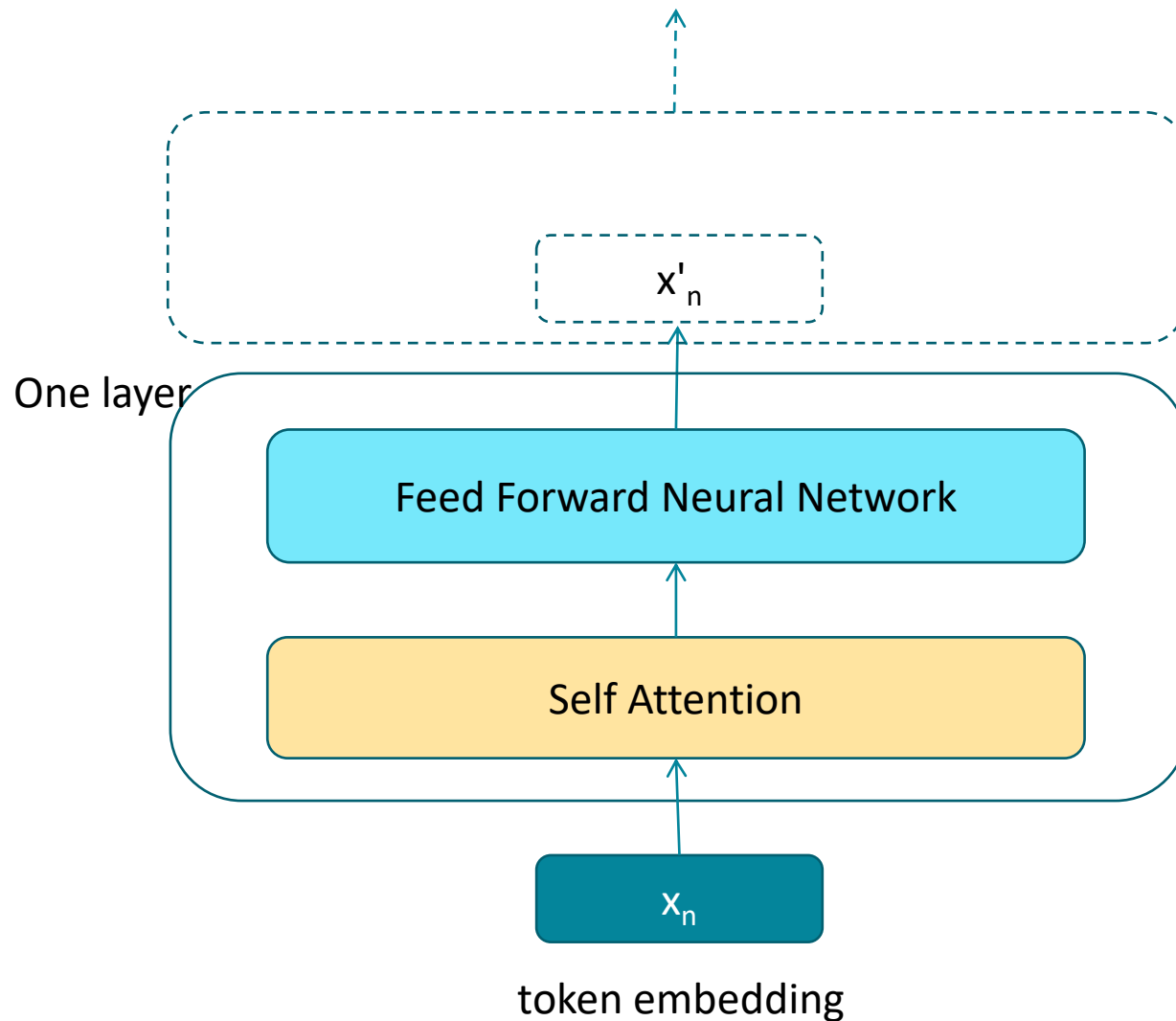


Vaswani et al., Transformer: Attention is All You Need (2017)

# Transformer Layer



# A Lot of Questions Remain



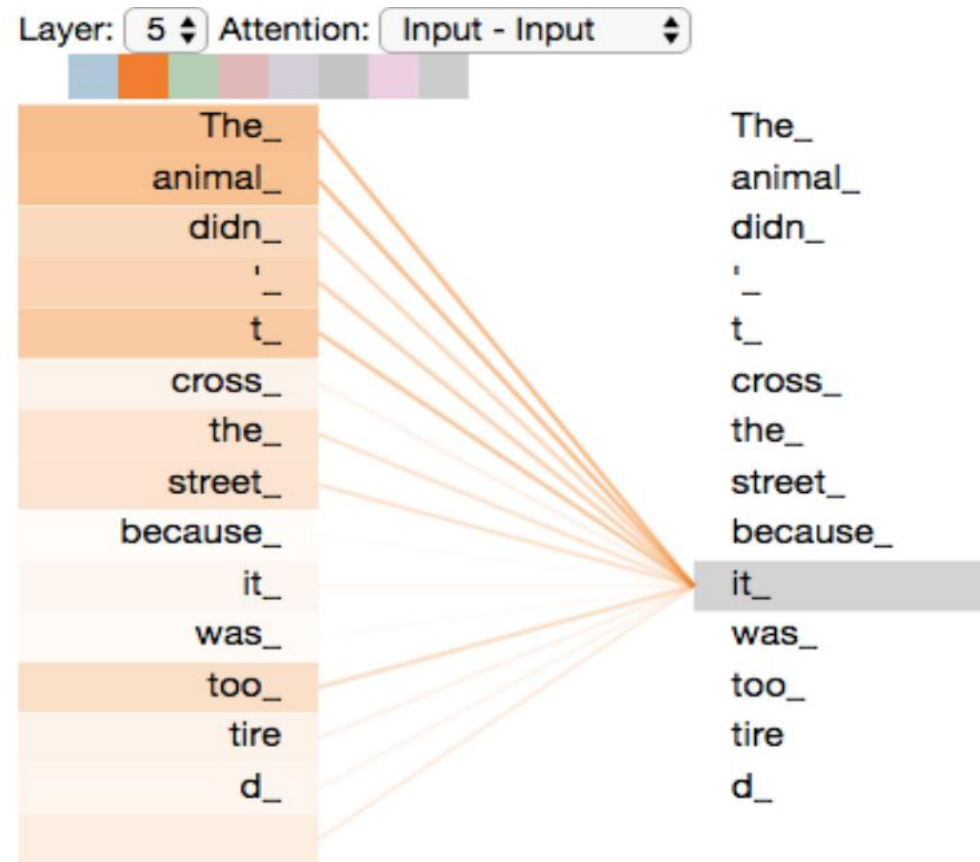
The Transformer is one of the first models in the history of artificial intelligence that has achieved remarkable success despite our limited understanding of its inner workings.

1. What are these layers really doing?
2. How important are these layers?
3. Are they really needed?

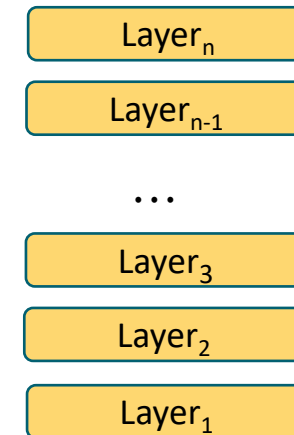


# Classical View: Attention

The animal didn't cross the street because **it** was too tired.  
The animal didn't cross the street because **it** was too busy.

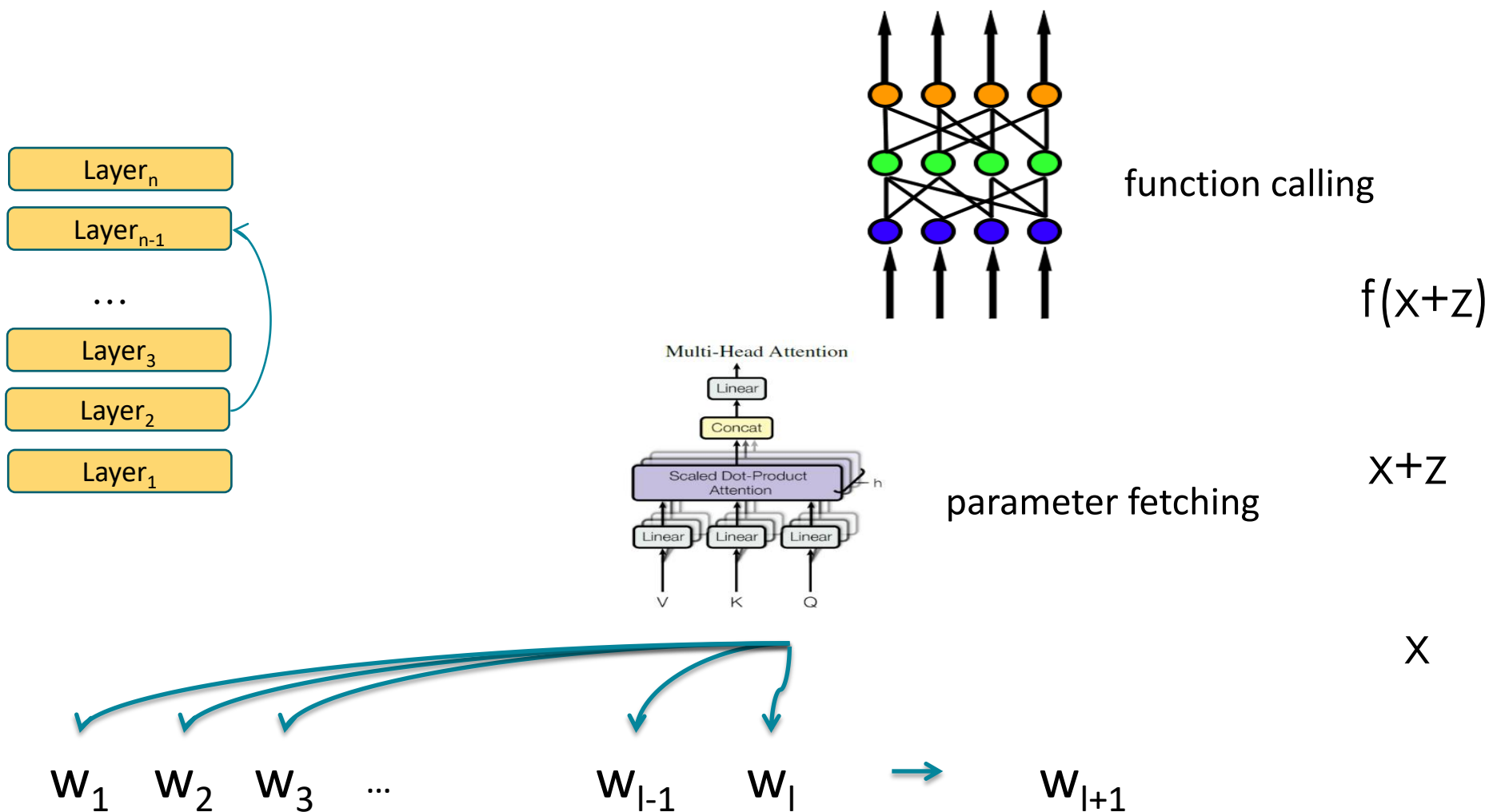


GPT decoder only model



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

# My Immature View



# # of Layers needed

[Context] The grey seal was found at Cruden Bay on Tuesday. The Scottish SPCA removed the seal from beach and took it to the charity's rescue centre...

Thankfully there were no injuries in this case and we were able to release him relatively quickly.

[Prompt] Where the grey seal was found?

**The grey seal was found at Cruden Bay.**

[Prompt] Please summarize this paragraph into a single sentence:

**A grey seal was found tangled in netting at Cruden Bay , rescued by the Scottish SPCA , and released back into the water with no injuries.**

[Prompt] Please calculate the sum of the 8 numbers in the list: [99, 45, 12, 78, 33, 66, 21, 54]

**To calculate the sum of all 8 numbers in the list, I'll add each number one by one:**

$$99 + 45 = 144$$

$$144 + 12 = 156$$

$$156 + 78 = 234$$

$$234 + 33 = 267$$

$$267 + 66 = 333$$

$$333 + 21 = 354$$

$$354 + 54 = 408$$

**The sum of all 8 numbers in the list is 408.**

Layer-skipping patterns of FlexiDepth (Llama-3-8B-Instruct) for language (left) and math (right) tasks. The light-to-dark blue gradient represents layer usage from 16 to 32.

Xuan et al., Adaptive Layer-skipping in Pre-trained LLMs, 2025

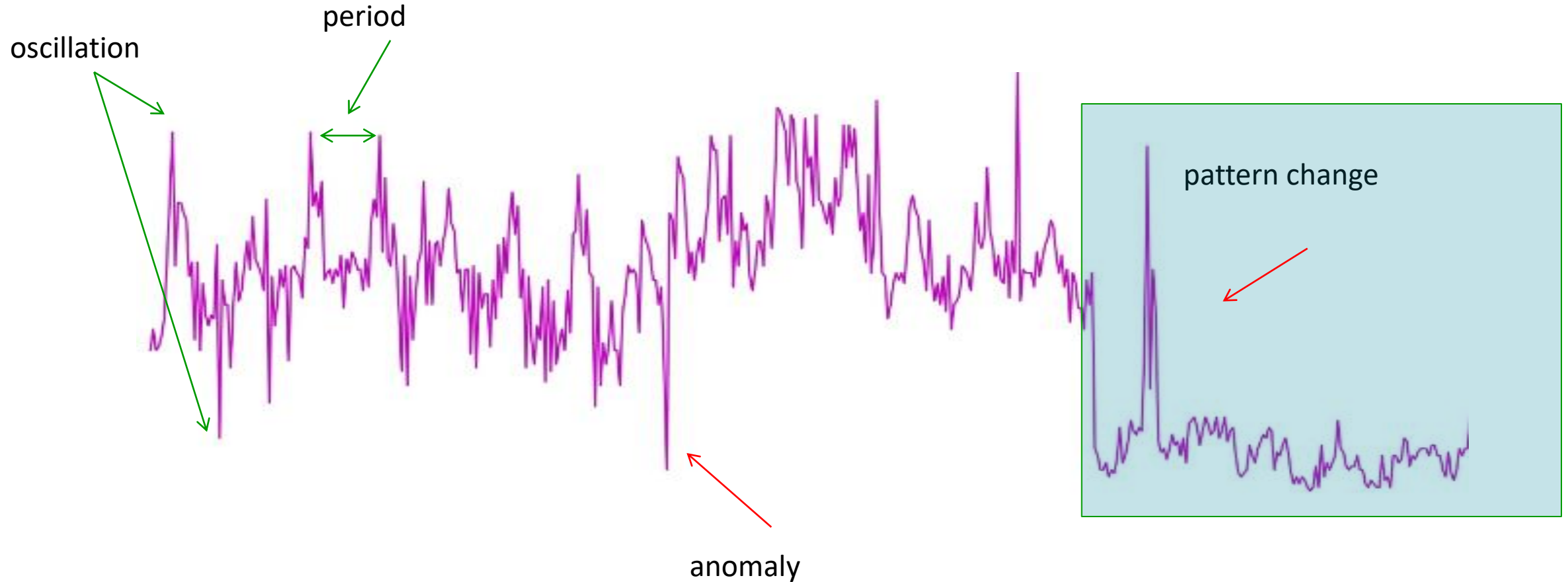
# Method Comparison

Methods	Single-Token Generation			Multi-Token Generation			Retain %
	MMLU	Hellaswag	Winogrande	GSM8K	HumanEval	CoQA	
Vanilla	0.673	0.706	0.744	0.679	0.299	0.784	100.0%
Skip 4 Layers							
LayerSkip	0.659	0.636	0.676	0.004	0.0	0.350	54.0%
ShortGPT	0.664	0.662	0.700	0.536	0.092	0.145	69.1%
LaCo	0.671	0.693	0.724	0.581	0.031	0.778	81.7%
MindSkip	0.664	0.698	0.722	0.378	0.189	0.720	84.2%
Ours	0.663	0.724	0.756	0.695	0.390	0.810	106.5%
Skip 8 Layers							
LayerSkip	0.650	0.525	0.640	0.0	0.0	0.049	43.9%
ShortGPT	0.307	0.462	0.597	0.001	0.0	0.005	32.0%
LaCo	0.656	0.628	0.695	0.065	0.006	0.707	65.3%
MindSkip	0.602	0.650	0.646	0.039	0.024	0.620	60.2%
Ours	0.616	0.705	0.735	0.662	0.341	0.801	100.7%

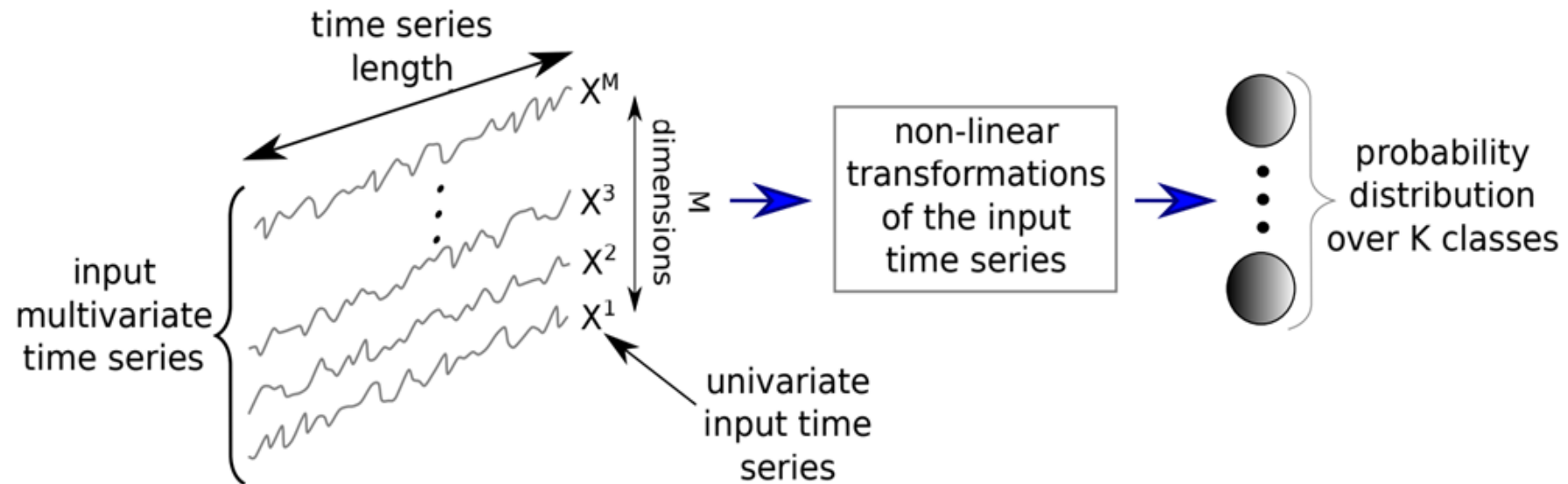
Llama-3-8B-Instruct, which consists of 32 layers.

# Time Series Analysis Is Always Challenging

Regularity and irregularity are mixed



# Multivariate Time Series



source: H. Fawaz et al.

# The Power of Transformer (LLMs)

- We had zero experience in time series research (in 2018) and were given a project by one of the largest telecommunication companies on work load prediction.
- “Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting,” by S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, X. Yan, NeurIPS 2019 -- the first Transformer-based time series forecasting approach.
- Today Transformer has become the de facto standard for time series analysis.

# Autoregression

The notation  $AR(p)$  indicates an autoregressive model of order  $p$ .

The  $AR(p)$  model is defined as

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$$

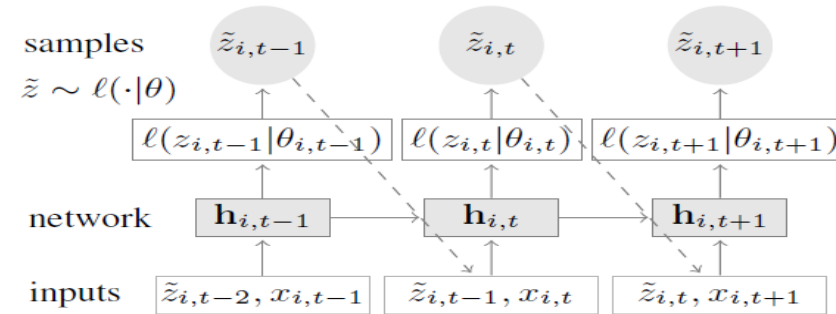
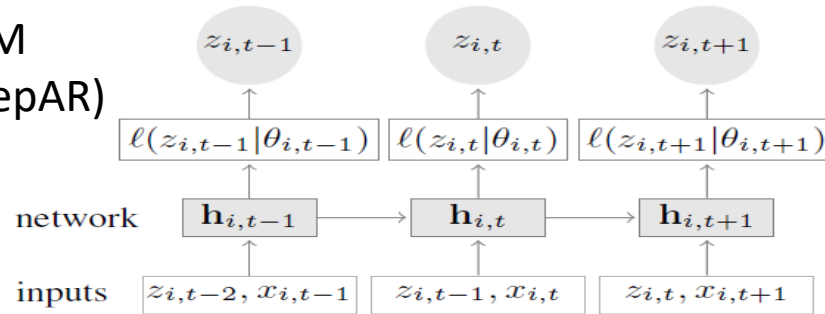
where  $\varphi_1, \dots, \varphi_p$  are the *parameters* of the model,  $c$  is a constant, and  $\varepsilon_t$  is **white noise**.

It evolves to autoregressive integrated moving average (ARIMA) model



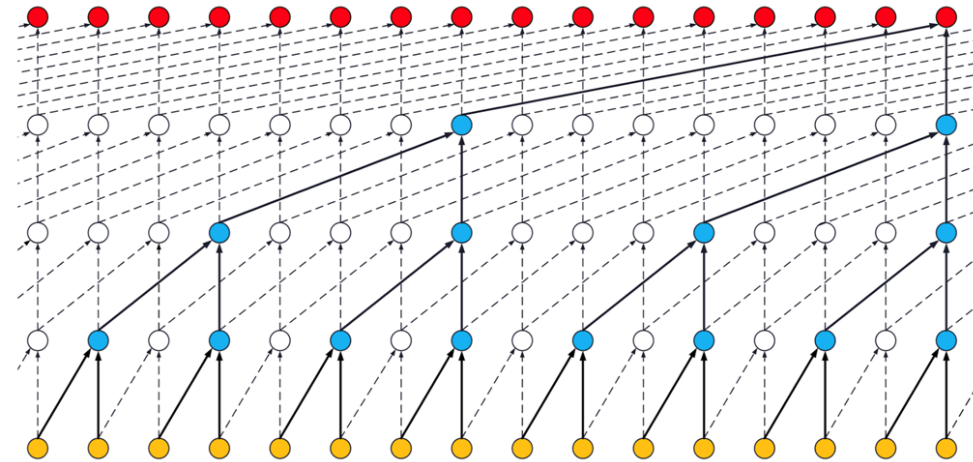
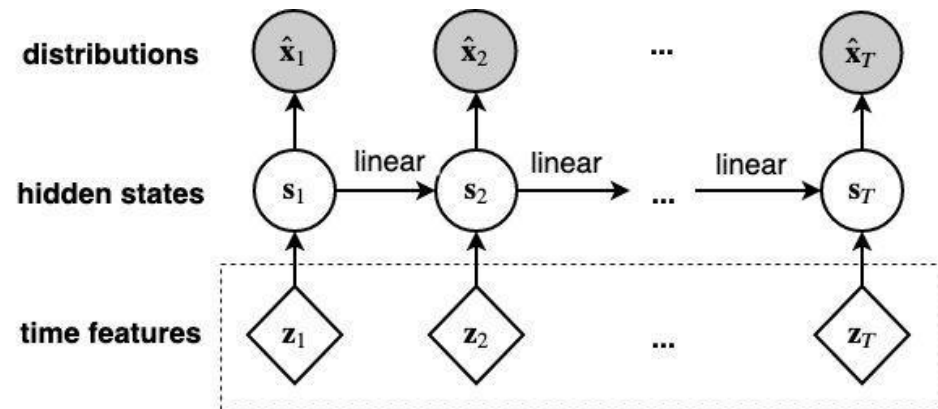
# Tried Many Models on Time Series

LSTM  
(DeepAR)



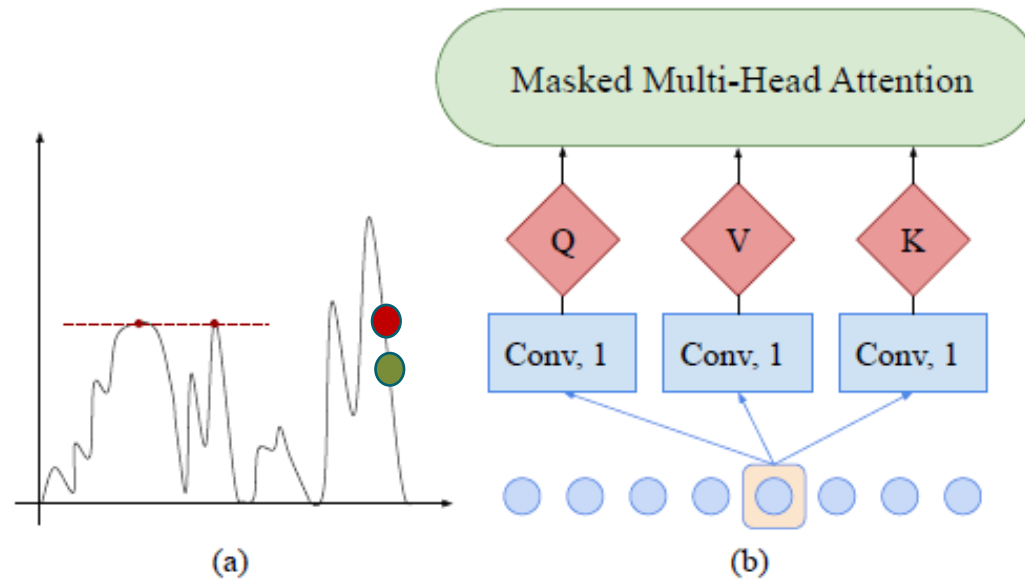
Temporal Convolutional Networks

State Space Model



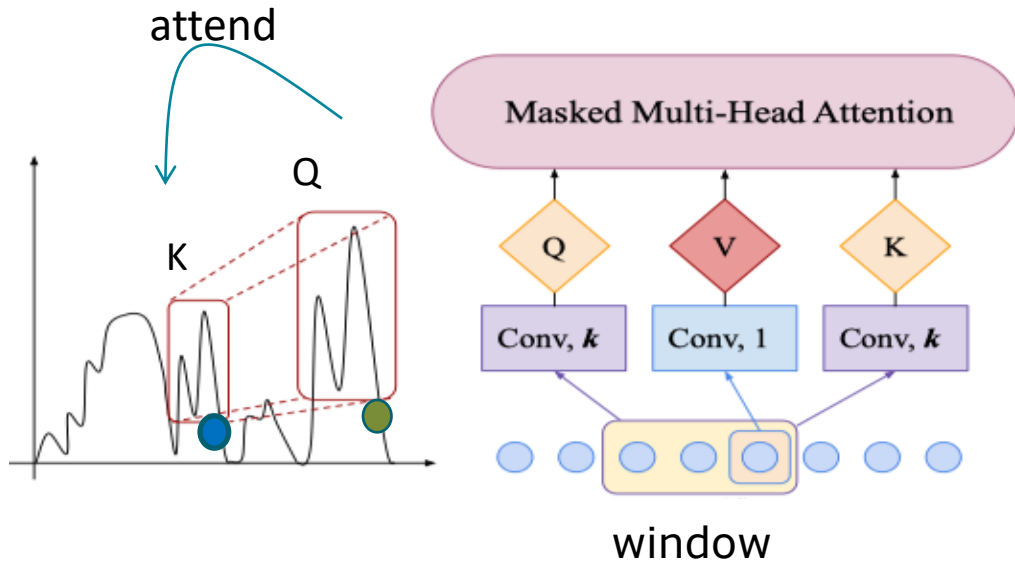
From: Bai et al. 2018

# Why Not Directly Try Transformer: Our First Try



Check the follow-up of all the previous data points that share the same value

# Our Intuition: It Is All About Patterns



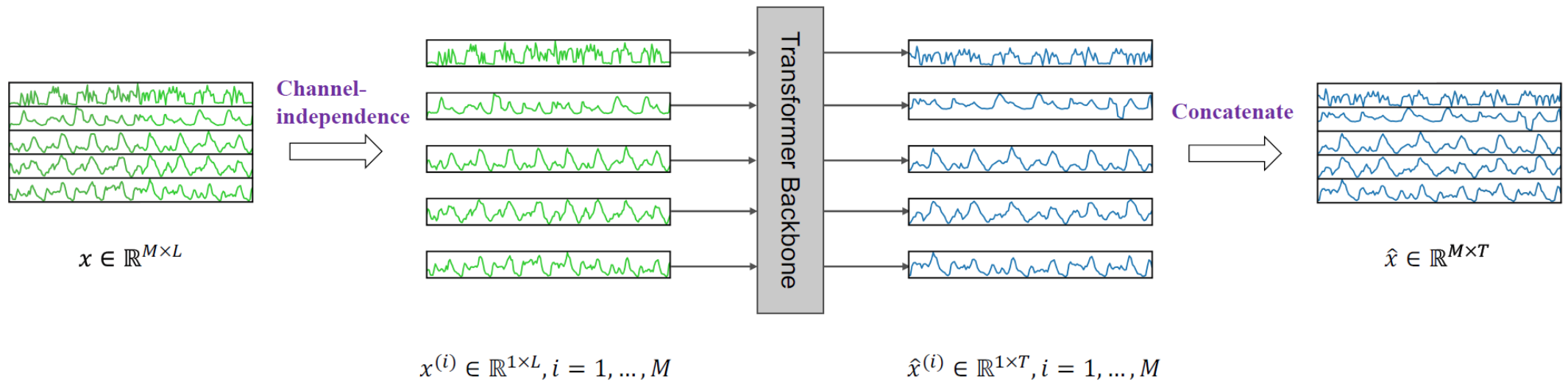
Use convolution to summarize local patterns in order to produce queries and keys.

# Experimental Results

Table 1: Results summary ( $R_{0.5}/R_{0.9}$ -loss) of all methods. `e-c` and `t-c` represent `electricity-c` and `traffic-c`, respectively. In the 1st and 3rd row, we perform rolling-day prediction of 7 days while in the 2nd and 4th row, we directly forecast 7 days ahead. TRMF outputs points predictions, so we only report  $R_{0.5}$ .  $\diamond$  denotes results from [6].

	ARIMA	ETS	TRMF	DeepAR	DeepState	Ours
<code>e-c</code> <sub>1d</sub>	0.154/0.102	0.101/0.077	0.084/-	0.075 $\diamond$ /0.040 $\diamond$	0.083 $\diamond$ /0.056 $\diamond$	<b>0.059/0.034</b>
<code>e-c</code> <sub>7d</sub>	0.283 $\diamond$ /0.109 $\diamond$	0.121 $\diamond$ /0.101 $\diamond$	0.087/-	0.082/0.053	0.085 $\diamond$ /0.052 $\diamond$	<b>0.070/0.044</b>
<code>t-c</code> <sub>1d</sub>	0.223/0.137	0.236/0.148	0.186/-	0.161 $\diamond$ /0.099 $\diamond$	0.167 $\diamond$ /0.113 $\diamond$	<b>0.122/0.081</b>
<code>t-c</code> <sub>7d</sub>	0.492 $\diamond$ /0.280 $\diamond$	0.509 $\diamond$ /0.529 $\diamond$	0.202/-	0.179/0.105	0.168 $\diamond$ /0.114 $\diamond$	<b>0.139/0.094</b>

# Multiple Time Series

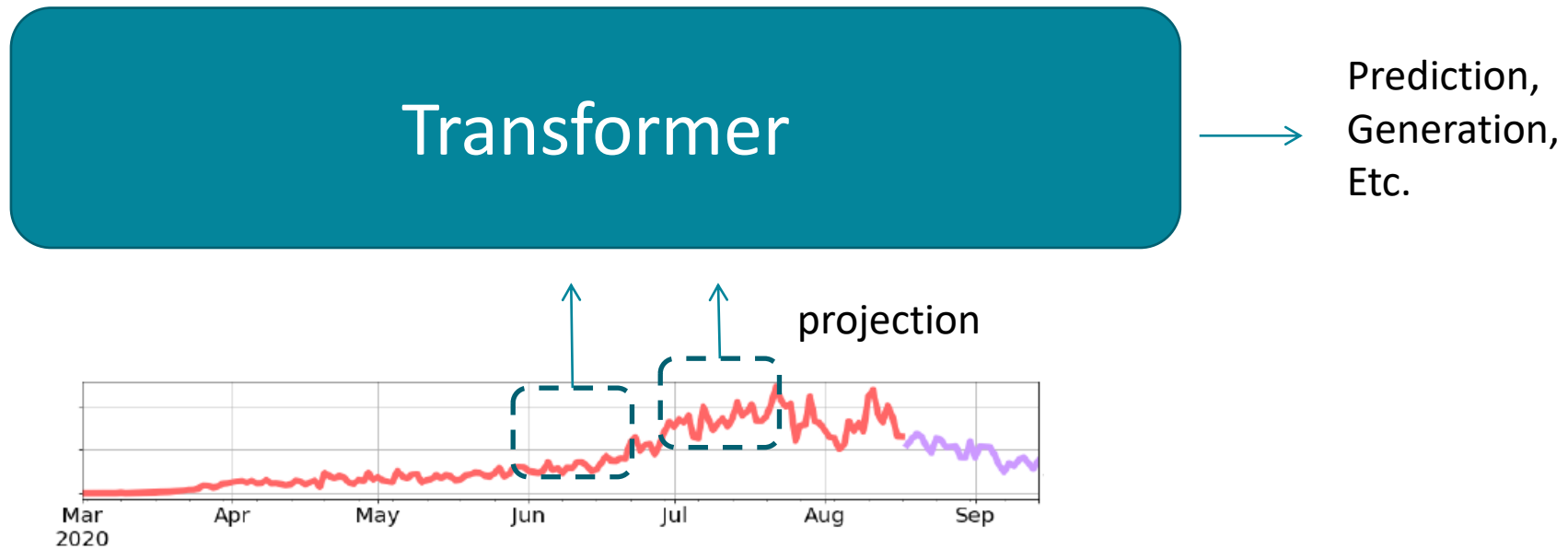


## PatchTST Model Overview

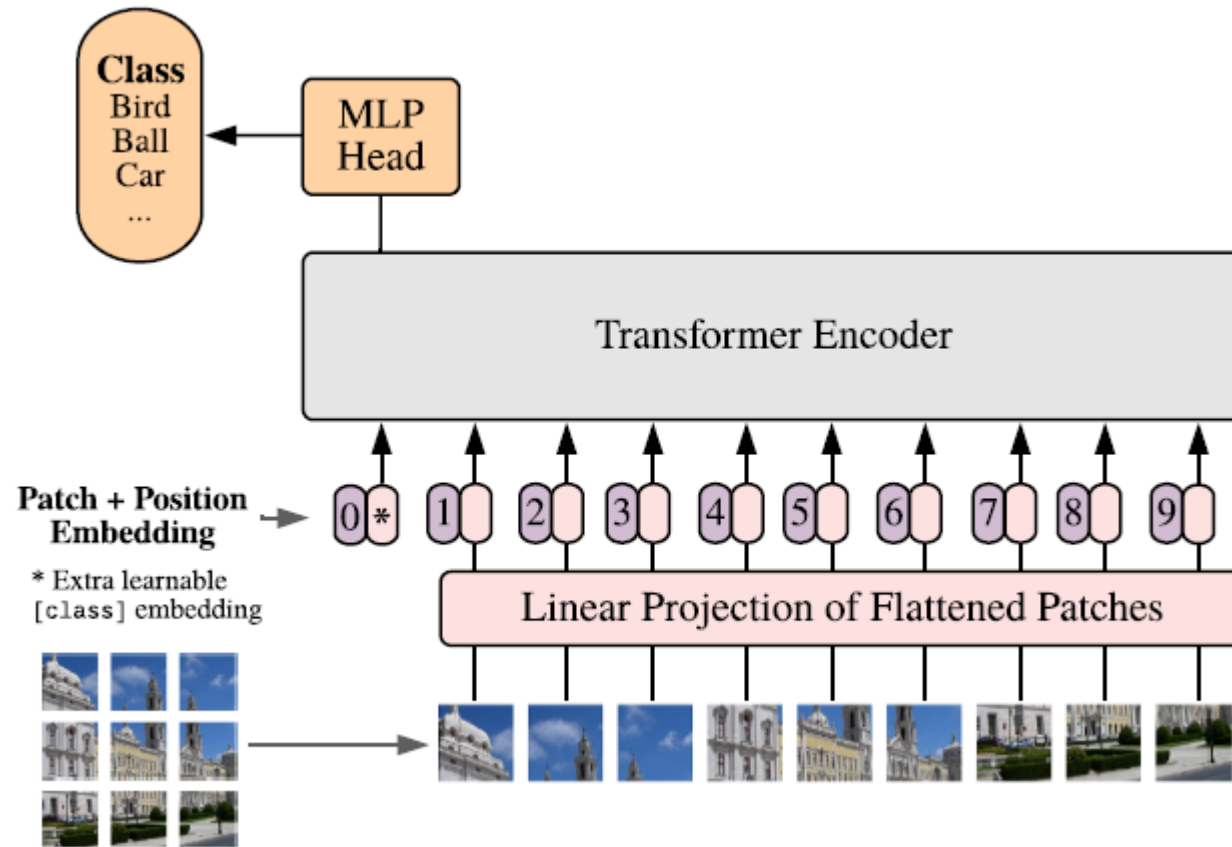
A Time Series is Worth 64 Words: Long-term Forecasting with Transformers, 2022, [Yugui Nie](#), et al., 2022

# Recap 1: Try Transformer First

Let's draw the architecture in another way



# ViT: Vision Transformer

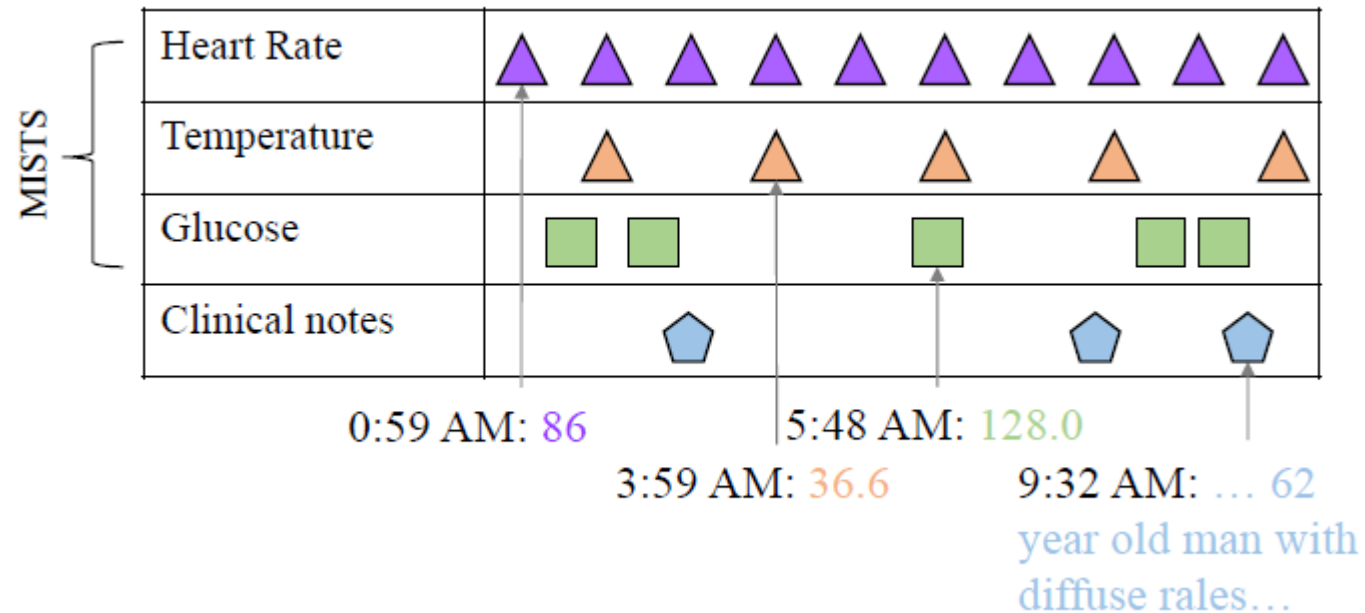


Alexey Dosovitskiy et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale 2020

# Change Modality: Transfer Learning

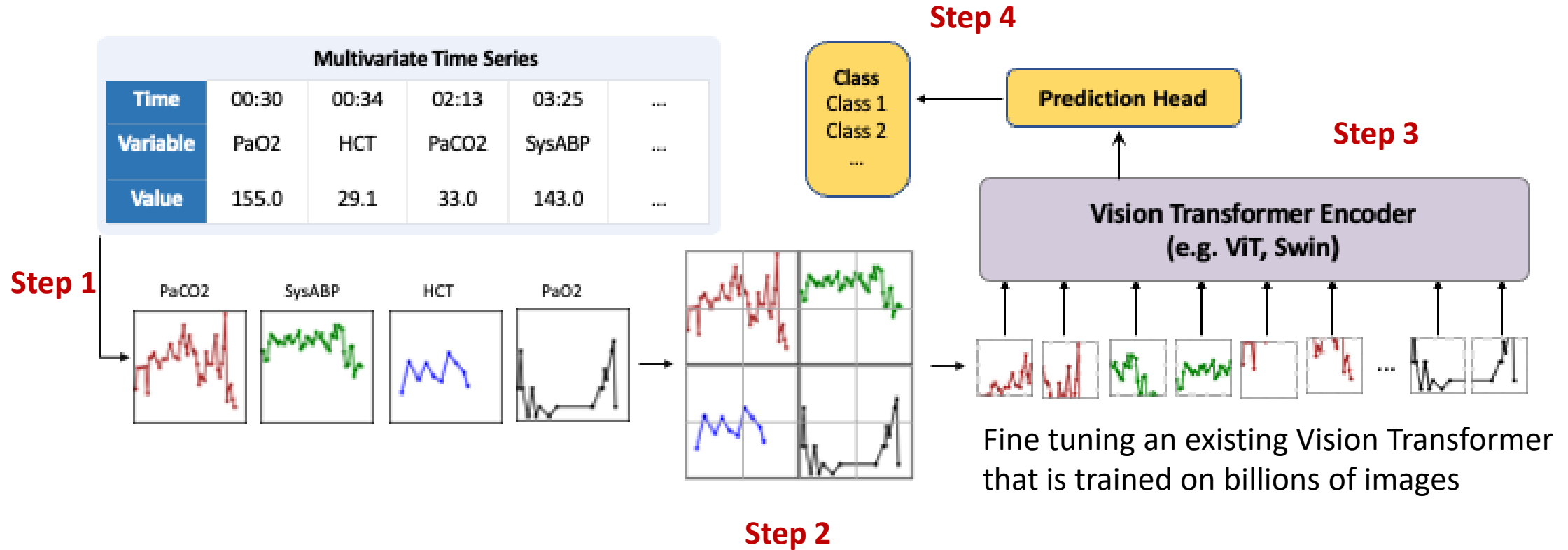
Irregularly Sampled Time Series: Medical Domain

Low Data Issue





# Transfer Learning: Time Series -> Image



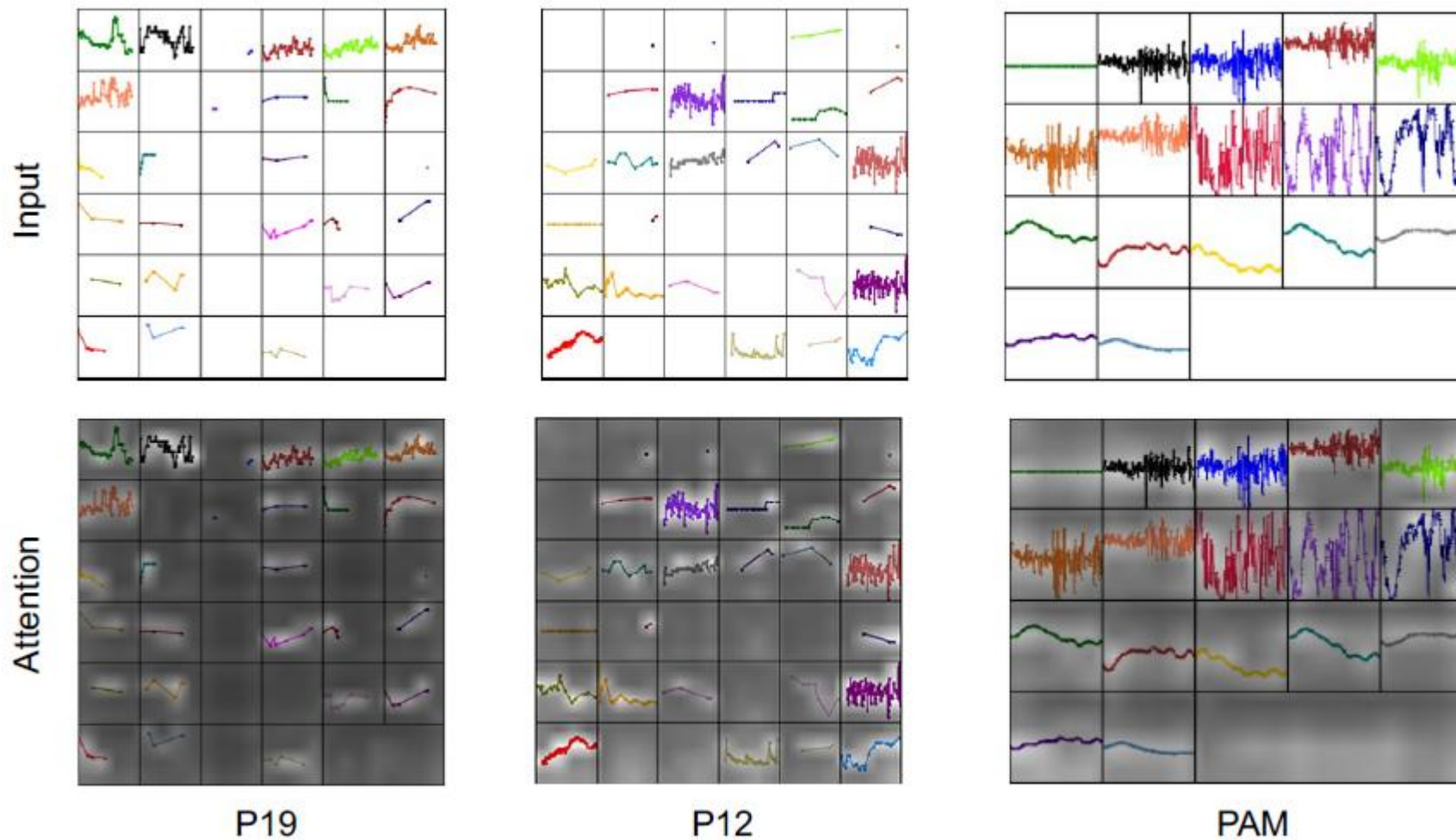
Time Series as Images: Vision Transformer for Irregularly Sampled Time Series, by Z. Li, S. Li, X. Yan, NeurIPS'23

# Results

Methods	P19		P12		PAM			
	AUROC	AUPRC	AUROC	AUPRC	Accuracy	Precision	Recall	F1 score
Transformer	80.7 $\pm$ 3.8	42.7 $\pm$ 7.7	83.3 $\pm$ 0.7	47.9 $\pm$ 3.6	83.5 $\pm$ 1.5	84.8 $\pm$ 1.5	86.0 $\pm$ 1.2	85.0 $\pm$ 1.3
Trans-mean	83.7 $\pm$ 1.8	45.8 $\pm$ 3.2	82.6 $\pm$ 2.0	46.3 $\pm$ 4.0	83.7 $\pm$ 2.3	84.9 $\pm$ 2.6	86.4 $\pm$ 2.1	85.1 $\pm$ 2.4
GRU-D	83.9 $\pm$ 1.7	46.9 $\pm$ 2.1	81.9 $\pm$ 2.1	46.1 $\pm$ 4.7	83.3 $\pm$ 1.6	84.6 $\pm$ 1.2	85.2 $\pm$ 1.6	84.8 $\pm$ 1.2
SeFT	81.2 $\pm$ 2.3	41.9 $\pm$ 3.1	73.9 $\pm$ 2.5	31.1 $\pm$ 4.1	67.1 $\pm$ 2.2	70.0 $\pm$ 2.4	68.2 $\pm$ 1.5	68.5 $\pm$ 1.8
mTAND	84.4 $\pm$ 1.3	50.6 $\pm$ 2.0	84.2 $\pm$ 0.8	<u>48.2</u> $\pm$ 3.4	74.6 $\pm$ 4.3	74.3 $\pm$ 4.0	79.5 $\pm$ 2.8	76.8 $\pm$ 3.4
IP-Net	84.6 $\pm$ 1.3	38.1 $\pm$ 3.7	82.6 $\pm$ 1.4	47.6 $\pm$ 3.1	74.3 $\pm$ 3.8	75.6 $\pm$ 2.1	77.9 $\pm$ 2.2	76.6 $\pm$ 2.8
DGM <sup>2</sup> -O	86.7 $\pm$ 3.4	44.7 $\pm$ 11.7	<u>84.4</u> $\pm$ 1.6	47.3 $\pm$ 3.6	82.4 $\pm$ 2.3	85.2 $\pm$ 1.2	83.9 $\pm$ 2.3	84.3 $\pm$ 1.8
MTGNN	81.9 $\pm$ 6.2	39.9 $\pm$ 8.9	74.4 $\pm$ 6.7	35.5 $\pm$ 6.0	83.4 $\pm$ 1.9	85.2 $\pm$ 1.7	86.1 $\pm$ 1.9	85.9 $\pm$ 2.4
Raindrop	<u>87.0</u> $\pm$ 2.3	<u>51.8</u> $\pm$ 5.5	82.8 $\pm$ 1.7	44.0 $\pm$ 3.0	<u>88.5</u> $\pm$ 1.5	<u>89.9</u> $\pm$ 1.5	<u>89.9</u> $\pm$ 0.6	<u>89.8</u> $\pm$ 1.0
<b>ViTST</b>	<b>89.2</b> $\pm$ 2.0	<b>53.1</b> $\pm$ 3.4	<b>85.1</b> $\pm$ 0.8	<b>51.1</b> $\pm$ 4.1	<b>95.8</b> $\pm$ 1.3	<b>96.2</b> $\pm$ 1.3	<b>96.1</b> $\pm$ 1.1	<b>96.5</b> $\pm$ 1.2

- Comparison with the baseline methods on irregularly sampled time series classification task.
- Bold indicates the best performer, while underline represents the second best.

# Where is Attention Located?

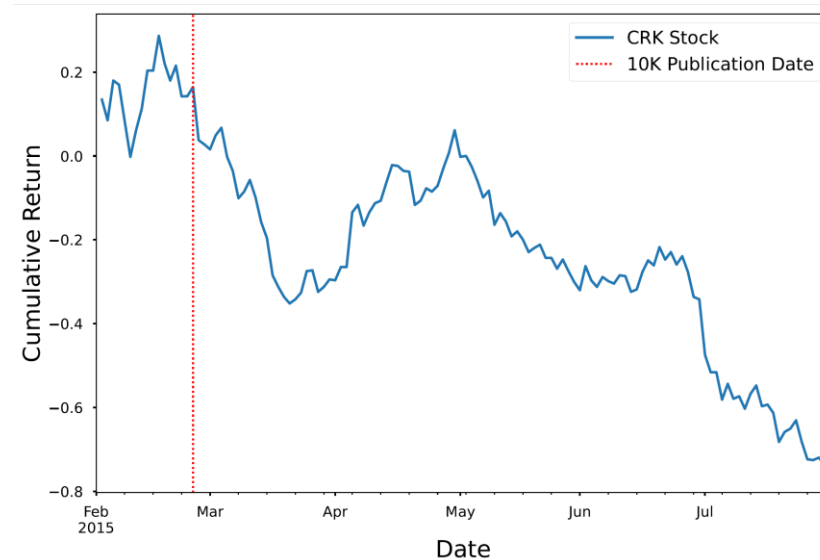


# More Modalities: Text Data

## Comstock Resources Inc. (CRK)

Depending upon future prices and our production volumes, our cash flows from our operating activities may not be sufficient to fund our capital expenditures, and we may need additional borrowings. ...If commodity prices remain low, we may also recognize further impairments of our producing oil and gas properties if the expected future cash flows from these properties becomes insufficient to recover their carrying value, and we may recognize additional impairments.

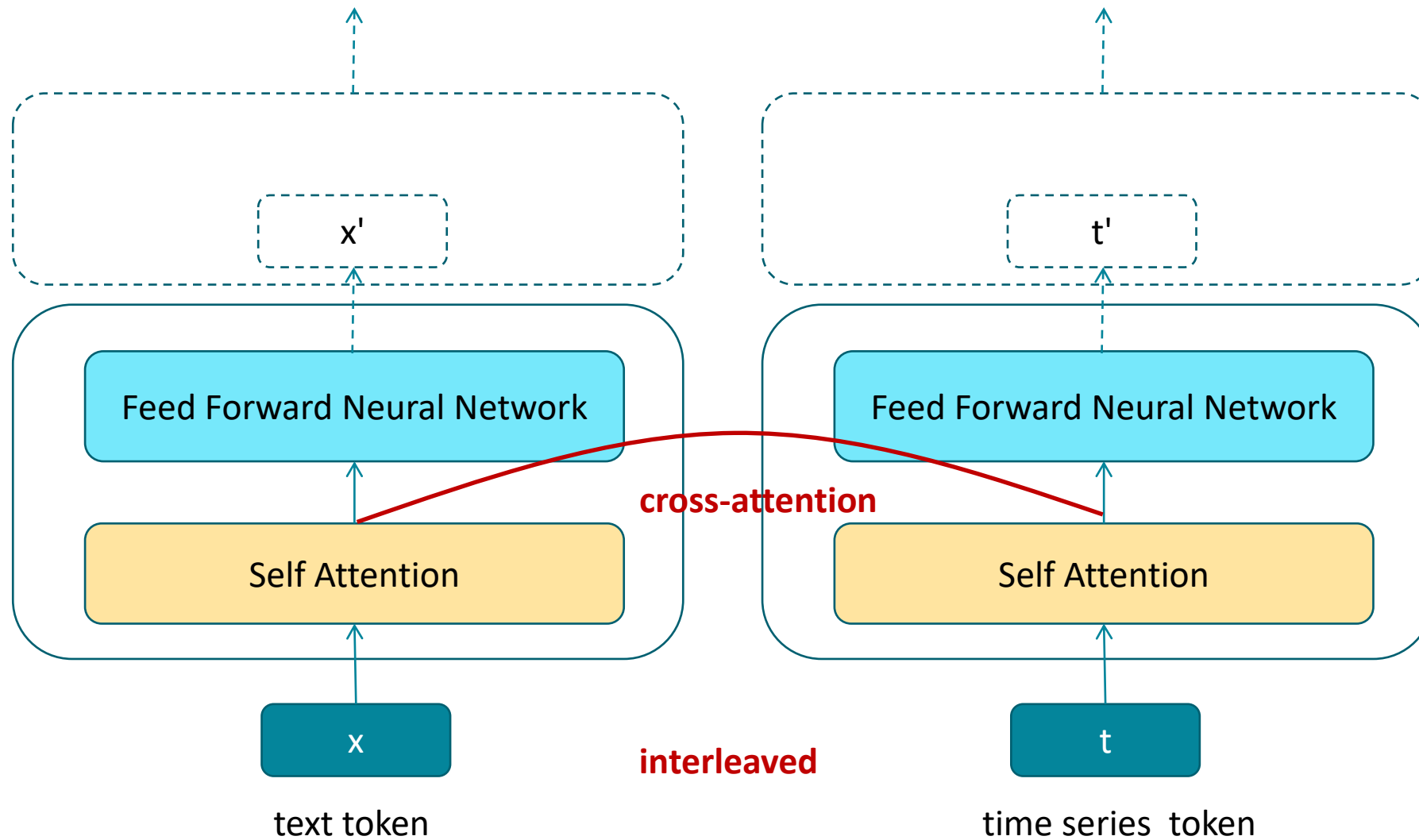
the 2015 Annual Report



Stock Price of CRK following the publication of the 2015 Annual Report

news, time series, er report, time series, ...

# Cross-Attention



# Results

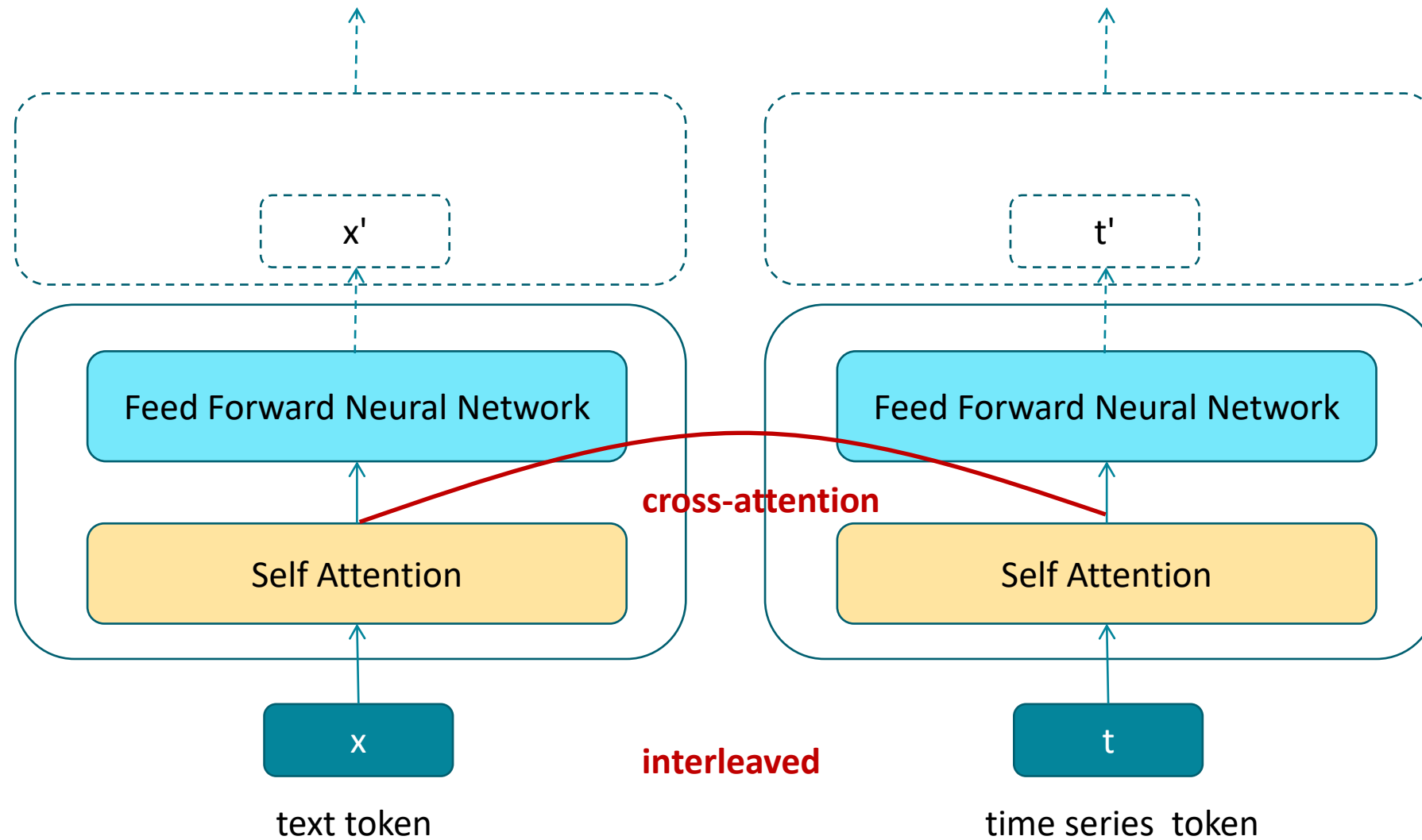
Model Class	Model	Base LM	Input	7D	30D
<i>Zero-Shot LLM</i>	GPT-4o, Direct (Lopez-Lira and Tang, 2023)	GPT-4o	text	52.45	53.31
	GPT-4o, CoT	GPT-4o	text	53.15	54.33
	GPT-4o, Direct	GPT-4o	ts	50.92	49.77
	GPT-4o, CoT	GPT-4o	ts	48.70	47.90
	GPT-4o, Direct (Williams et al., 2024)	GPT-4o	text, ts	52.09	52.71
	GPT-4o, CoT (Tan et al., 2025)	GPT-4o	text, ts	50.56	53.05
<i>Unimodal</i>	TS-Only (Nie et al., 2022)	None	ts	52.93	53.88
	Text-Only (Chen et al., 2022)	Llama3-8B	text	53.76	54.13
<i>MMTSF</i>	TimeLLM (Jin et al., 2023)	Llama2-7B	text, ts	53.79	55.10
	TaTs (Li et al., 2025)	Llama2-7B	text, ts	54.48	54.81
	TTSR (Chow et al., 2024)	Mistral-7B	text, ts	55.93	56.17**
	TimeMDD (Liu et al., 2024)	Llama3-8B	text, ts	55.15	55.25
	Hybrid-MMF (Kim et al., 2024b)	Llama3-8B	text, ts	55.96*	55.84
<i>SFF</i>	FinMA (Xie et al., 2023)	Llama2-7B	text, ts	51.11	52.15
	MTFE-MICM (Koval et al., 2024)	BigBird-125M	text, ts	55.76	54.49
	FININ (Wang et al., 2024a)	RoBERTa-125M	text, ts	52.47	53.13
	StockTime (Wang et al., 2024b)	Llama3-8B	text, ts	55.36	55.85
	MAT (Emami Gohari et al., 2024)	FinBERT-110M	text, ts	54.43	53.81
	MM-iTransformer (Mou et al., 2025)	FinBERT-110M	text, ts	54.16	53.57
<b>Proposed</b>	<b>MSE-ITT</b>	Llama3-8B	text, ts	<b>57.94*</b> [0.08]	<b>58.48**</b> [0.07]

# Simulated Investment Performance

Method	Net Return	Volatility	Net Sharpe Ratio
TS-Only (Nie et al., 2022)	5.99	13.11	0.46
Text-Only (Chen et al., 2022)	8.60	10.47	0.82
TTSR (Chow et al., 2024)	12.37	11.28	1.10
Hybrid-MMF (Kim et al., 2024b)	11.60	10.19	1.13
MTFE-MICM (Koval et al., 2024)	10.23	10.39	0.99
StockTime (Wang et al., 2024b)	13.91	12.65	1.10
<b>Proposed, MSE-ITT</b>	<b>17.01</b>	<b>11.26</b>	<b>1.51</b>

- Annualized portfolio statistics of simulated investment performance, expressed in percentage units. "Net" performance includes an estimate of the impact of transaction costs

# Recap 2: Multimodal LLM Can Help





# Accelerate Scientific Discovery (in Materials Science)

Level 0: No AI-enhanced discovery

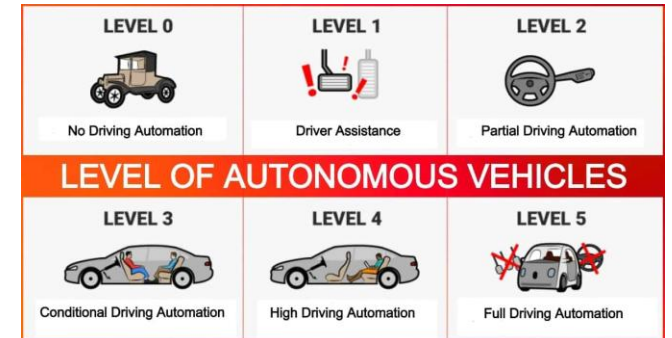
Level 1: Knowledge retrieval

Level 2: Materials -> Property

Level 3: Property -> Materials

Level 4: Insight Generation

Level 5: Level 4 -> Level 3 -> ... -> Experiments -> Level 4 (loop)



# Level 1: Knowledge Retrieval

What is the stiffest material with the lowest formation energy in Si-O system?

## a. User Query

What's the stiffest material with the lowest formation energy in Si-O system?

## b. LLaMP ReAct (Top-level)

**1-Thought:** To answer this question, I need to find materials in the Si-O system with the lowest formation energy and the highest stiffness. I will use the **MPThermoExpert** tool to search for materials in the Si-O system and sort them by formation energy. After finding candidates, I will need to use the **MPElasticityExpert** tool to determine the stiffness of these materials.

### 2-Action:

```
```json
{
  "action": "MPThermoExpert",
  "action_input": {
    "input": "What are the materials with the lowest
formation energy in the Si-O system?"
  }
}
...`
```

### 9-Thought:

The previous work has identified materials in the Si-O system with the lowest formation energy. The next step is to determine the stiffness of these materials. I will use the **MPElasticityExpert** tool to find the stiffness (typically represented by the bulk modulus or Young's modulus) for each of these materials.

### 10-Action:

```
```json
{
  "action": "MPElasticityExpert",
  "action_input": {
    "input": "What are the bulk and Young's moduli for Si2O5
(mp-862998), SiO2 (mp-733790), SiO2 (mp-6922), SiO2
(mp-556985), and SiO2 (mp-556994)?"
  }
}
...`
```

## c. Specialized ReAct (Bottom-level)

MPThermoExpert

### 3-Action:

```
```json
{
  "action": "search_materials_thermo__get",
  "action_input": {
    "formula": "Si-O",
    "sort_fields": "formation_energy_per_atom",
    "limit": 5,
    "fields": "material_id,formula_pretty,formation_energy_per_atom"
  }
}
...`
```

### 4-Observation:

Error on search\_materials\_thermo\_\_get: REST query returned with error status code 400 on URL [https://api.materialsproject.org/materials/thermo/?\\_fields=material\\_id...](https://api.materialsproject.org/materials/thermo/?_fields=material_id...) Problem processing one or more provided formulas.

### 5-Thought:

The query may be too broad or ambiguous. I will refine the search by specifying the chemical system more precisely.

### 6-Action:

```
```json
{
  "action": "search_materials_thermo__get",
  "action_input": {
    "chemsys": "Si-O",
    "sort_fields": "formation_energy_per_atom",
    "limit": 5,
    "fields": "material_id,formula_pretty,formation_energy_per_atom"
  }
}
...`
```

### 7-Observation:

```
[{'formula_pretty': 'Si2O5', 'material_id': 'mp-862998', 'formation_energy_per_atom':
-2.289431383571429}, {'formula_pretty': 'SiO2', 'material_id': 'mp-733790',
'formation_energy_per_atom': -3.039244523055555}, ...]
```

### 8-Answer:

The materials with the lowest formation energy in the Si-O system are: **Si2O5 (mp-862998)** with a formation energy per atom of **-2.289 eV/atom**, **SiO2 (mp-733790)** with a formation energy per atom of **-3.039 eV/atom**, **SiO2 (mp-6922)** with a formation energy per atom of **-3.037 eV/atom**, **SiO2 (mp-556985)** with a formation energy per atom of **-2.993 eV/atom**, and **SiO2 (mp-556994)** with a formation energy per atom of **-2.958 eV/atom**.

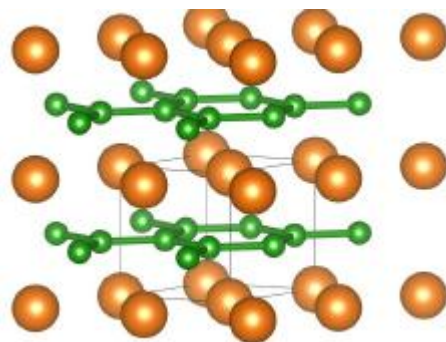
## Challenge:

Aggregation Query:

How many papers are related to Si2O5?

How many times was Si2O5 mentioned?

# Level 2: Materials -> Property



what is its TC\_Supercon?

**ChemNLP describer input:**

MgB<sub>2</sub> is crystallizes in the hexagonal P6/mmm space group. There is one shorter (2.5 Å) and one longer (3.96 Å) B-Mg bond lengths. Mg is bonded in edge-sharing MgB<sub>12</sub> Cuboctahedral...

**Output:** 32.685

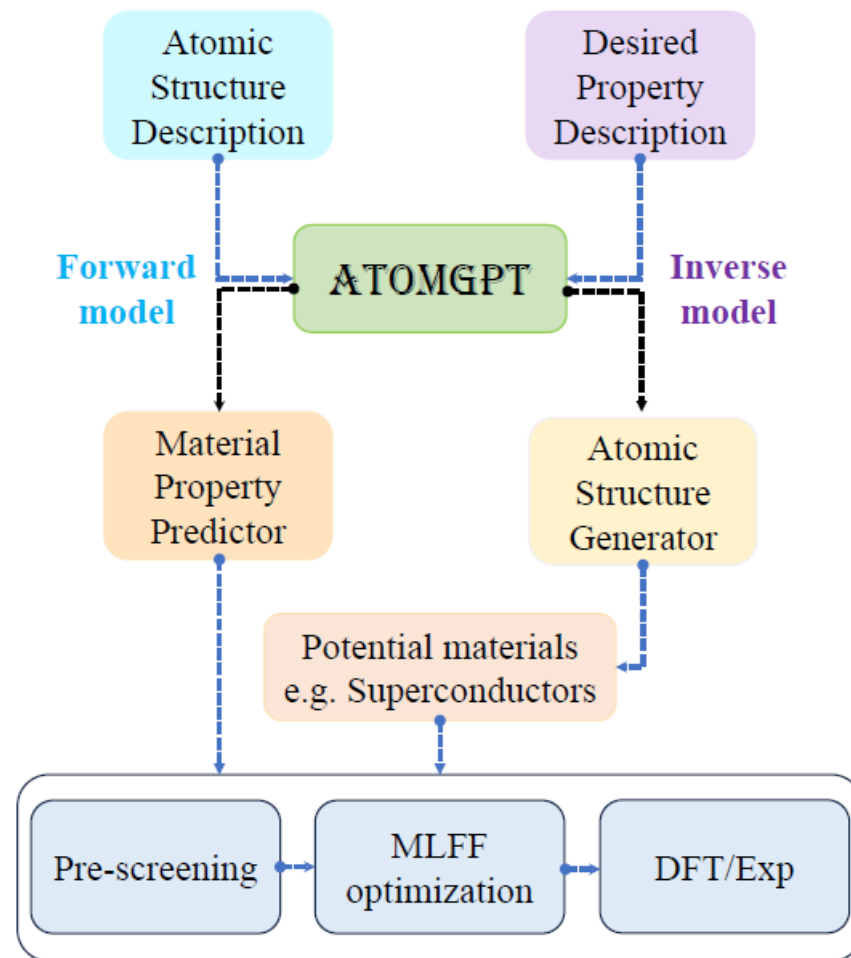
# Level 3: Property -> Materials

## Alpaca input:

{**Instruction:** Below is a description of a superconductor material.,

**Input:** The chemical formula is  $\text{MgB}_2$ . The  $T_{\text{c\_supercon}}$  is 32.685. Generate atomic structure description with lattice lengths, angles, coordinates and atom types.,

**Output:** 3.07 3.07 3.51\n90 90 120\nMg 0.000 0.000 0.000\nB 0.667 0.333 0.500\nB 0.333 0.667 0.500}



# Level 2/3: From A Language Agent Perspective

Train a multitask model on the Clintox dataset to predict a drug's toxicity and FDA approval status. Save the test set predictions, including the SMILES representation of drugs and the probability of positive labels, to "pred\_results/clintox\_test\_pred.csv".

(a) Task Instruction

Dataset Directory:

```
|-- clintox/  
|---- clintox_test.csv  
|---- clintox_train.csv
```

Dataset Preview:

[START Preview of clintox/clintox\_train.csv]

smiles,FDA\_APPROVED,CT\_TOX

CCC(/C=C/Cl)(C#C)O,1,0

C[C@H]1C[C@H]2[C@@H]3CC[C@@H]([C@]3(C[C@@H]([C@@H]2[C@@]4(C1=CC(=O)CC4)C)O)C(=O)C,1,0

C[C@@H]1CCN([C@H](C1)C(=O)[O-])C(=O)[C@H](CCC[NH+]=C(N)N)NS(=O)(=O)c2cccc3c2NC[C@@H](C3)C,1,0

...

[END Preview of clintox/clintox\_train.csv]

(b) Dataset Information

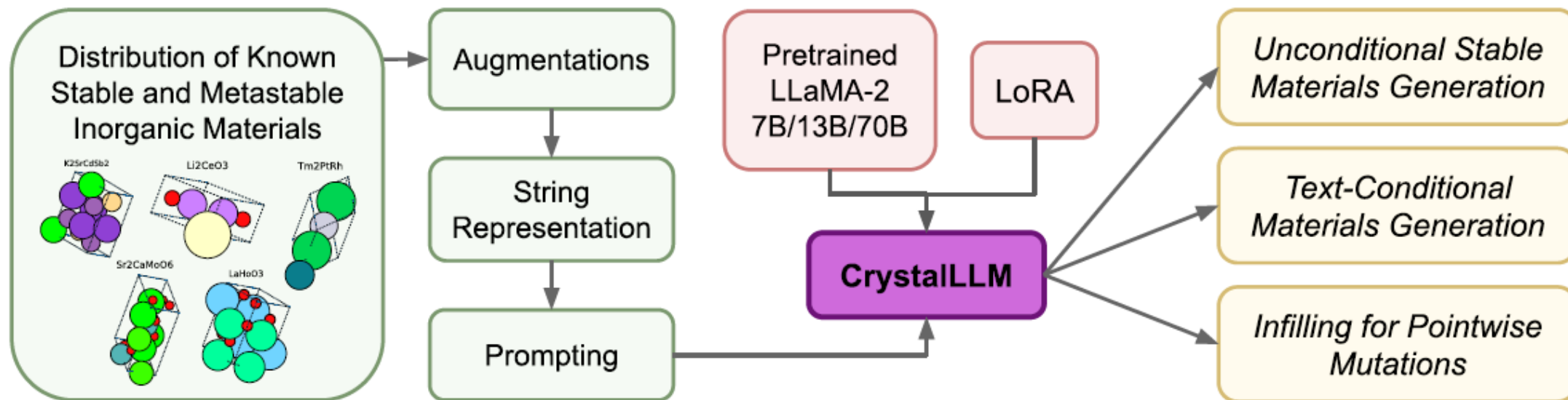
1. \*On the task\*: The ClinTox dataset contains drugs approved by .....
2. \*On featurization\*: To represent the molecular structure, use Extended-Connectivity Fingerprints (ECFPs) featurization in deepchem.....

(c) Expert-Provided Knowledge

```
import deepchem as dc  
.....  
from deepchem.molnet.load_function.molnet_loader import  
_MolnetLoader  
  
class MyClintoxLoader(_MolnetLoader):  
    def create_dataset(self):  
        .....  
  
CLINTOX_TASKS = ['FDA_APPROVED', 'CT_TOX']  
train_loader = MyClintoxLoader('ECFP', .....)  
train_dataset = .....  
test_loader = MyClintoxLoader('ECFP', .....)  
test_dataset = .....  
  
.....  
model = dc.models.MultitaskClassifier(.....)  
model.fit(train_dataset)  
test_scores = model.predict(test_dataset, .....)  
.....  
test_scores_df.to_csv('pred_results/clintox_test_pred.csv')
```

(d) Annotated Program

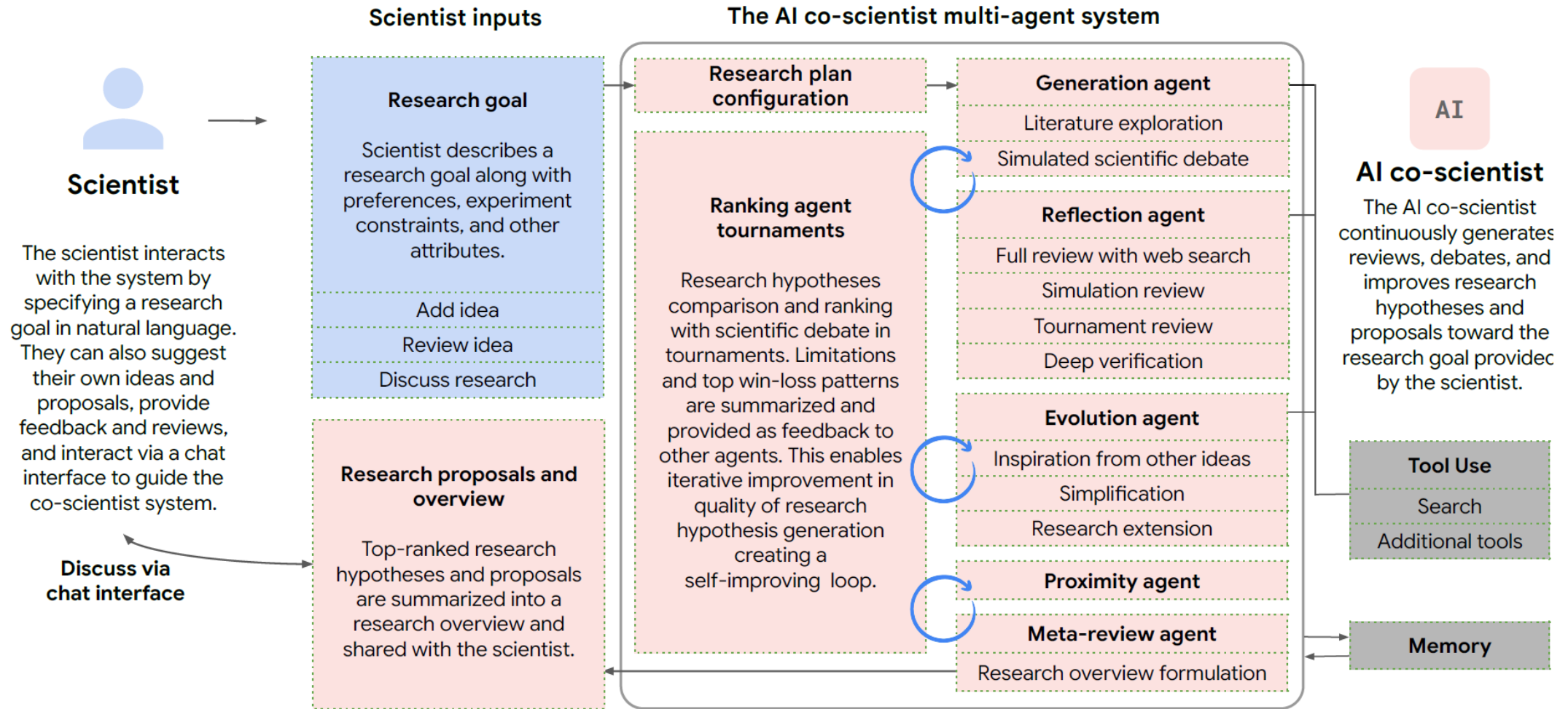
# Level 4: Insight Generation



and certainly more, idea generation, hypothesis generation



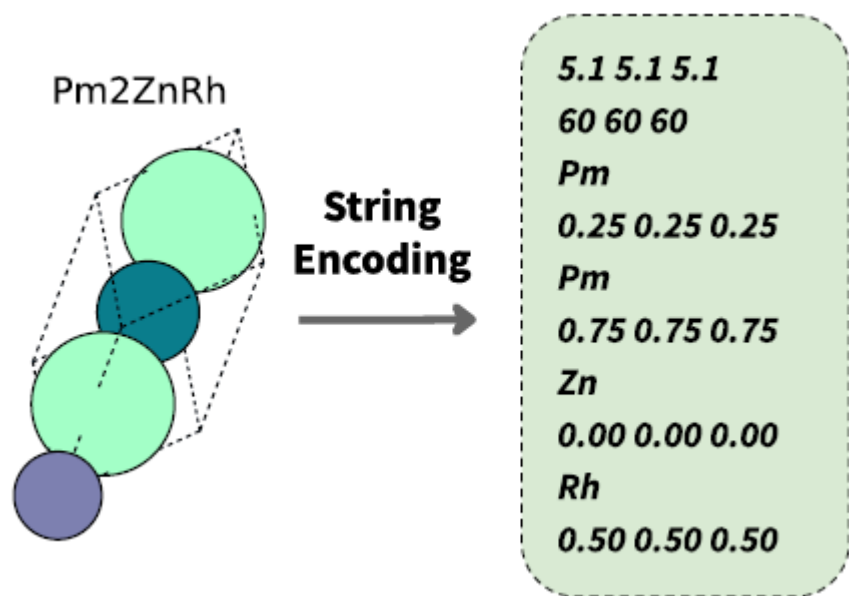
# Level 4: Agentic AI: Google Co-Scientist



The AI co-scientist system design

Juraj Gottweis et al., Towards an AI co-scientist, 2025

# Crystal Structures and Energy Prediction



Periodic materials are characterized by a unit cell that repeats infinitely in all three dimensions. Each unit cell is specified by its side lengths ( $l_1, l_2, l_3$ ) and angles ( $\theta_1, \theta_2, \theta_3$ ). Within this lattice structure, there are  $N$  atoms, each identified by an element symbol,  $e_i$ , and a set of 3D coordinates ( $x_i, y_i, z_i$ ). The structure of a bulk material  $C$  can be represented by:

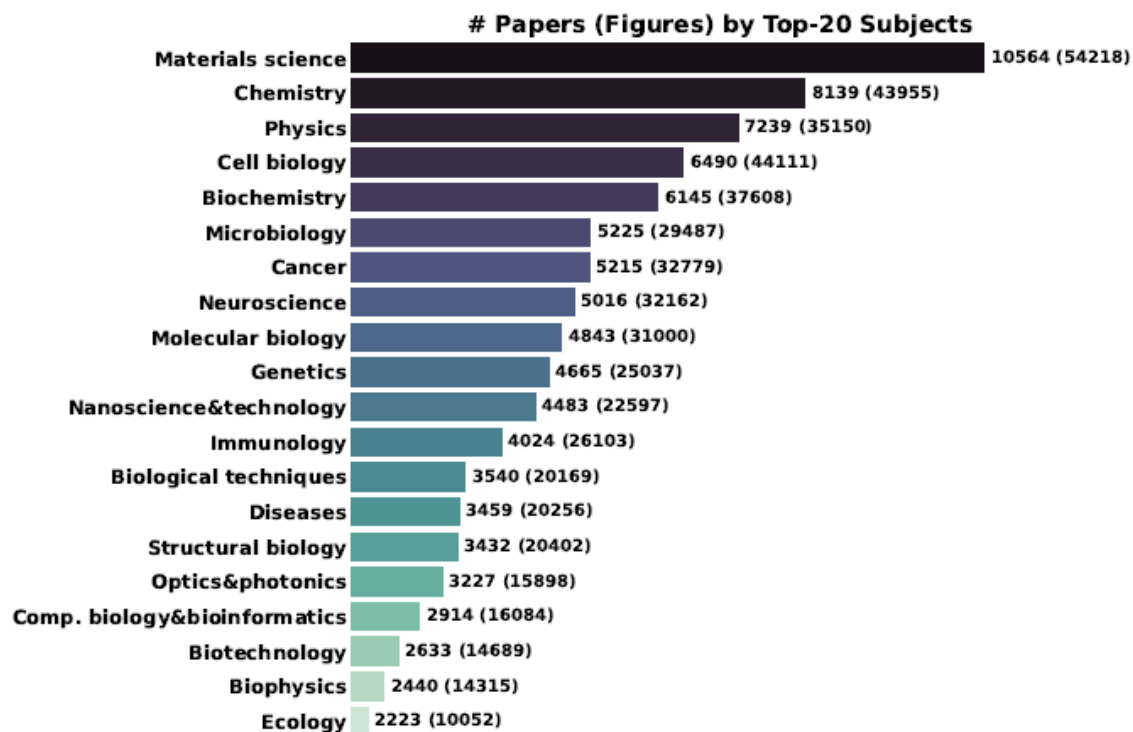
$$C = (l_1, l_2, l_3, \theta_1, \theta_2, \theta_3, e_1, x_1, y_1, z_1, \dots, e_N, x_N, y_N, z_N).$$



# What is the Next? Leverage Prior Knowledge

- CrystalLLM (Gruver et al., 2024) is fine tuned on LLaMA2 (Touvron et al., 2023) whose scientific knowledge may be insufficient for fully understanding material generation principles.
- Can we explore continual pre-training of LLaMA2 using all the scientific articles we crawled and improve the model's performance on stable material generation task?

# Leverage Prior Knowledge



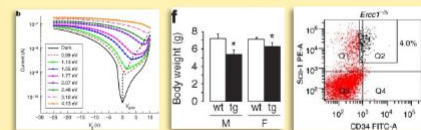
The top 20 out of 72 science subjects with the most articles in our dataset MMSCI. The corresponding numbers of papers and figures (in brackets) are shown.

- We continuously pretrained the LLaMA2-7B model on our collected interleaved article text and figure images, using data within materials science as well as other eight related subjects in the same Physical Science category.

# Continual Pretraining

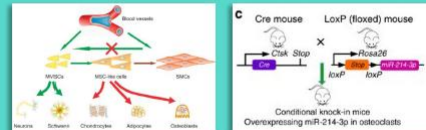
## Quantitative Data Visualization Charts (53.5%):

For charts and graphs displaying quantitative data, such as scatter plots, bar graphs, and line charts.



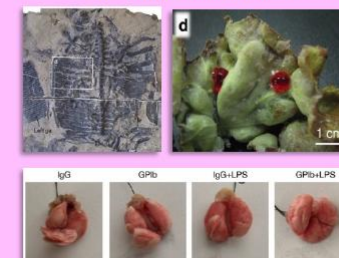
## Schematic Diagrams (13.2%):

Simplified and symbolic representations of systems, processes, or structures to explain how something works or is constructed.



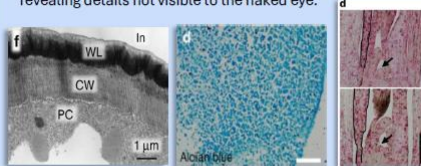
## Macroscopic photographs (5.2%):

Images or photographs of objects or scenes that are visible to the naked eye, often used for visual analysis.



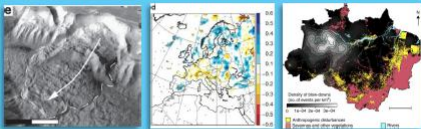
## Microscopic photographs (14.7%):

Photographs or images captured using a microscope, revealing details not visible to the naked eye.



## Geographical and Environmental Maps (1.2%):

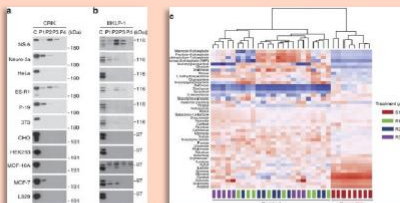
Visual representations of geographical areas or environmental data, often used for navigation, analysis, or to illustrate spatial relationships and patterns in maps.



nature  
communications

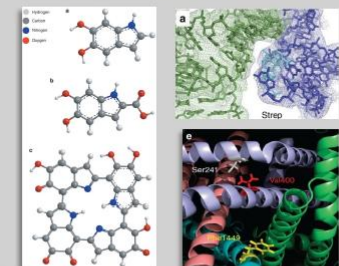
## Experimental Results Visualizations (8.8%):

For images that display results from experimental procedures, such as Western blots, PCR results, and gel electrophoresis.



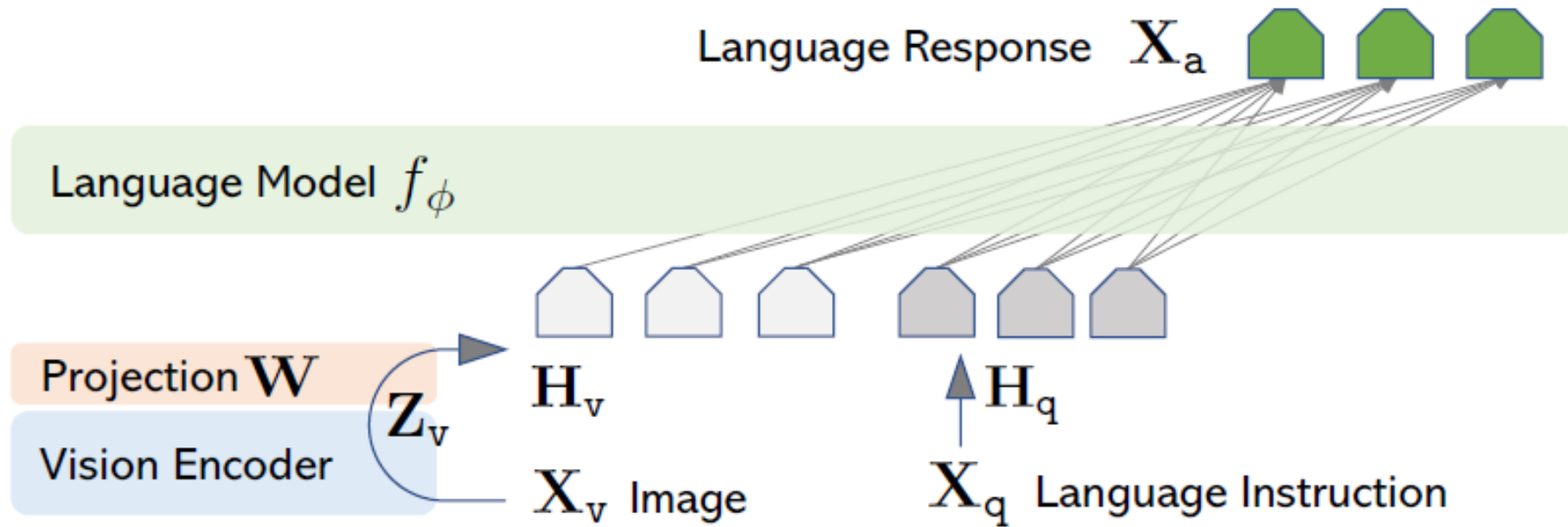
## Simulated Images (3.4%):

Computer-generated images or visualizations created to model, predict, or illustrate theoretical scenarios, processes, or phenomena.



We continuously pretrained the LLaMA2-7B model on our collected interleaved article text and figure images, using data within materials science as well as other eight related subjects in the same Physical Science category.

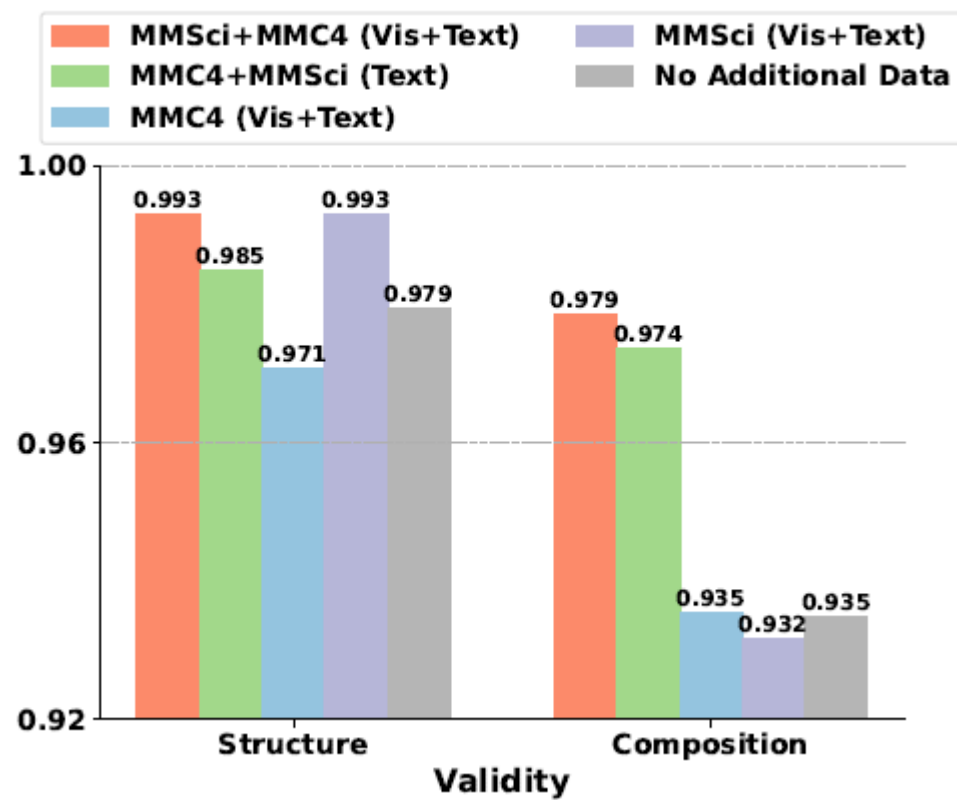
# LlaVa Architecture



Haotian Liu, Chunyuan Li, Qingyang Wu, Yong Jae Lee, Visual Instruction Tuning, 2023

# Results

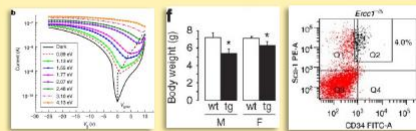
Prior knowledge helps to improve the validity



# Recap 3: Add More Knowledge

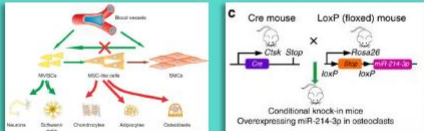
## Quantitative Data Visualization Charts (53.5%):

For charts and graphs displaying quantitative data, such as scatter plots, bar graphs, and line charts.



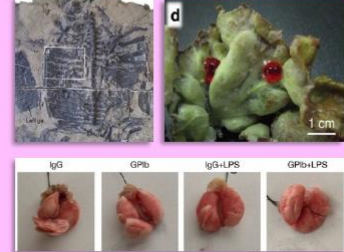
## Schematic Diagrams (13.2%):

Simplified and symbolic representations of systems, processes, or structures to explain how something works or is constructed.



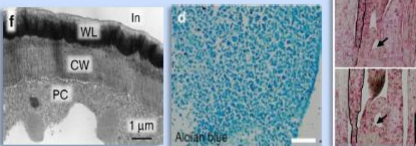
## Macroscopic photographs (5.2%):

Images or photographs of objects or scenes that are visible to the naked eye, often used for visual analysis.



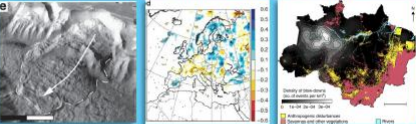
## Microscopic photographs(14.7%):

Photographs or images captured using a microscope, revealing details not visible to the naked eye.



## Geographical and Environmental Maps (1.2%):

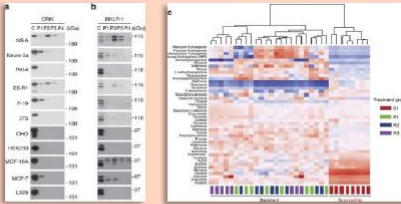
Visual representations of geographical areas or environmental data, often used for navigation, analysis, or to illustrate spatial relationships and patterns in maps.



## nature communications

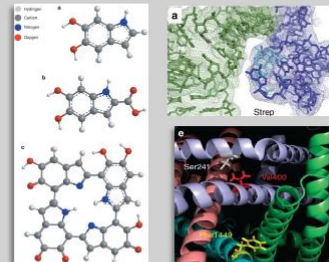
## Experimental Results Visualizations (8.8%):

For images that display results from experimental procedures, such as Western blots, PCR results, and gel electrophoresis.



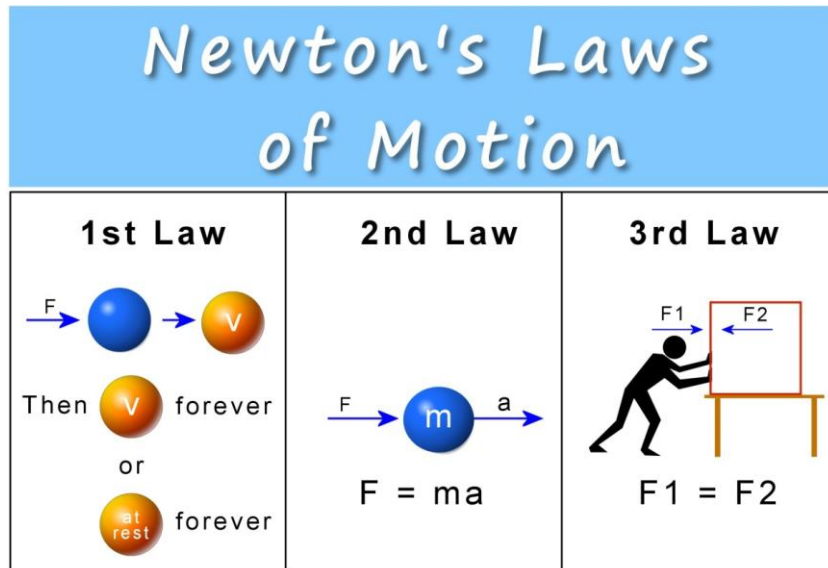
## Simulated Images (3.4%):

Computer-generated images or visualizations created to model, predict, or illustrate theoretical scenarios, processes, or phenomena.





# Challenges



© Eugene Brennan

Given Newton's laws and a foundation in high school mathematics, one might expect a model to learn how to solve typical high school physics problems. Large language models are capable of doing so, **provided they are trained on sufficiently large and relevant datasets.**

# Challenges (cont.)

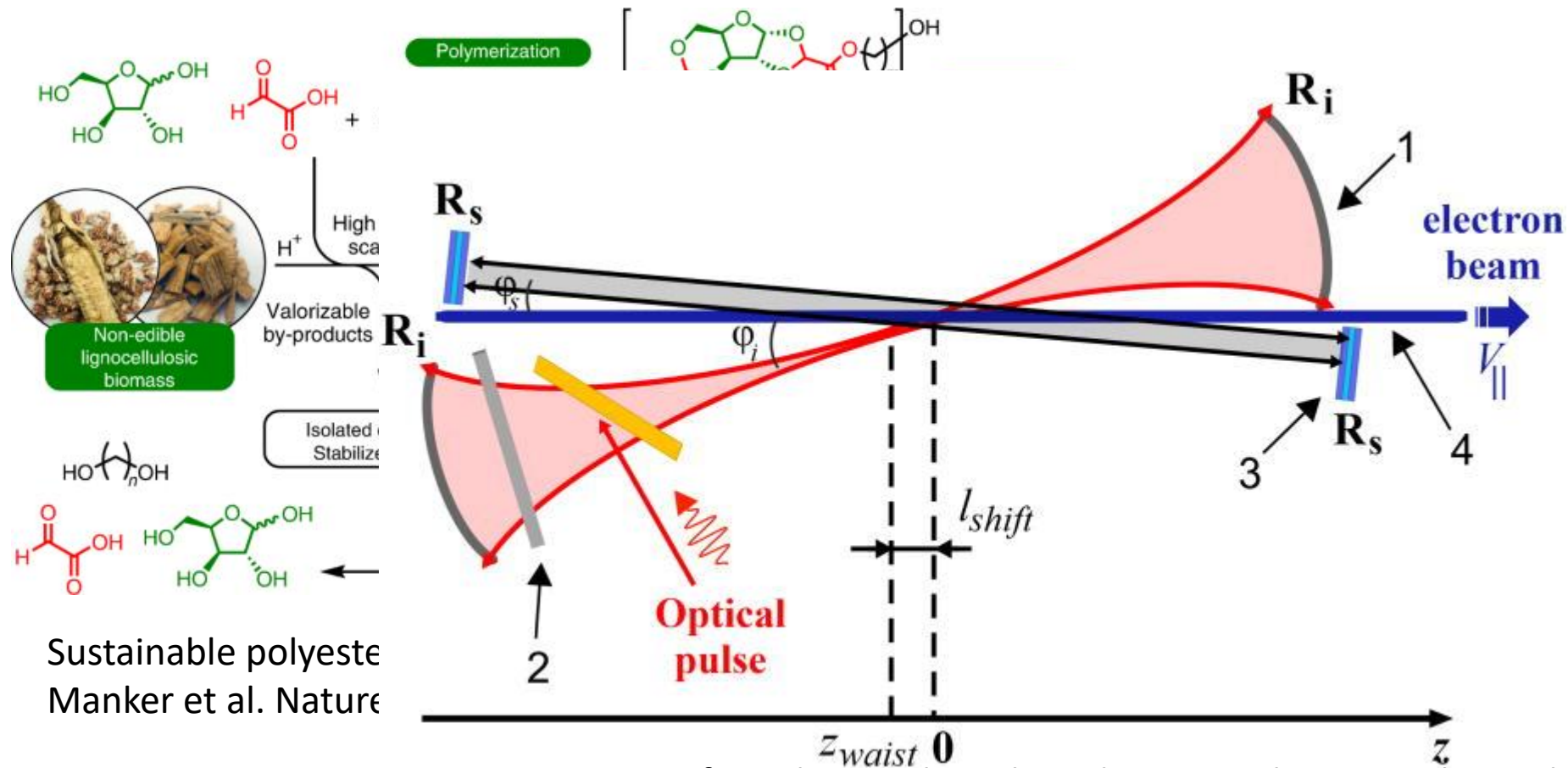
This poses a challenge for using LLMs in scientific discovery, as the scientific literature typically lacks repeated, detailed coverage of the same findings.

- Scientific research is all about novelty
- Scientific articles are not as abundant as web data
- Basically it is one-shot or a few shot learning problem



# Challenges (cont.)

Granularity of information and their representation



X-ray Compton free electron laser based on optical pump pulse multiple backscattering, Lev A. Yurovskiy, Naum S. Ginzburg, Phys. Plasmas (2025)

# Challenges (cont.)

Taking into account that angles between pump/scattered waves and electron beam are small we adopt model of collinear backscattering by representing the pump and scattered fields in the form of linearly polarized plane waves

$$\vec{A}_j = \text{Re} \left[ \vec{y}_0 \hat{A}_j(z, t) \exp(i\omega_j t - ik_j z) \right], \quad j = i, s, \quad (1)$$

where  $\hat{A}_j(z, t)$  is the slowly varying wave amplitudes,  $\omega_j$  is the frequency of the corresponding wave,  $|k_j| = \omega_j/c$ . In the case of stimulated scattering under study, the relationship between the frequencies and wave numbers of the concurrent pump wave (index  $i$ ) and the scattered wave (index  $s$ ) has the form,

$$\omega_i + |k_i|V_{||} = \omega_s - k_s V_{||}, \quad (2)$$

where  $V_{||}$  is the translational velocity of electrons. This condition can be considered as a synchronism condition  $V_{||} = V_c = \omega_c/k_c$  between electrons and a virtual combination wave with frequency  $\omega_c = \omega_s - \omega_i$  and wavenumber  $k_c = |k_i| + k_s$ .

Link all these pieces together with math equations, it is way harder than text parsing.

# Challenges (cont.)

Hallucination, good or bad?

# Acknowledgment

CNSI: Christopher Dunham, Tal Margalith, Javier Read de Alaniz

UCSB Ph.D. students: Shiyang Li, Zekun Li, Weizhi Wang, Ross Koval, Xuan Luo

LANL: Zhehui (Jeph) Wang