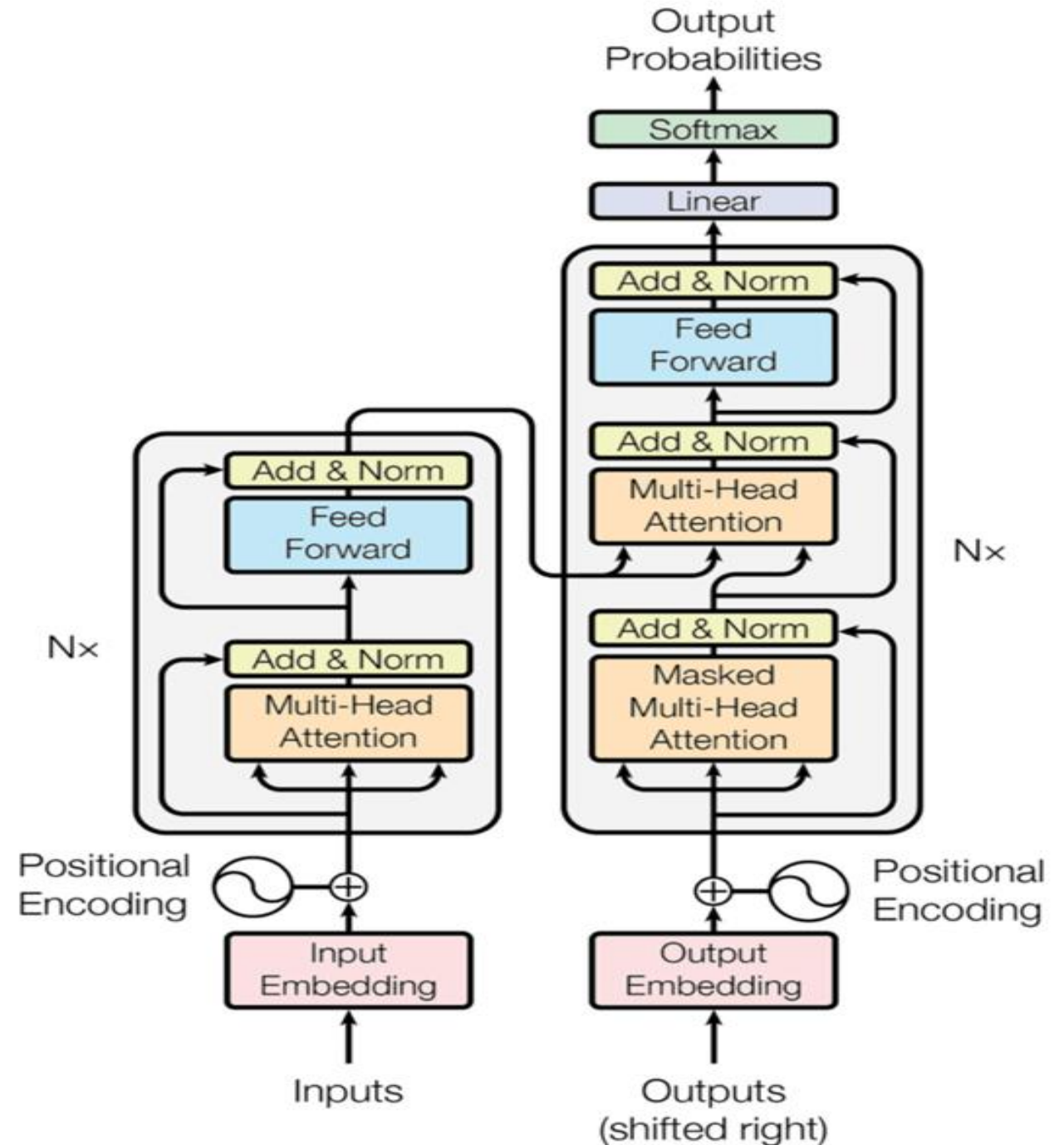


DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

**Paper Explained By Ali Reza Torabi,
Amir Hossein Aghazadeh , Alireza Nosrati**

Understanding The Transformer Architecture :

1. Input Embeddings and Tokens and Positional Encoding
2. Attention Mechanism
3. Feed Forward Neural Network
4. Output Probabilities



To process text effectively, the first step is to represent words in a way that models can understand.

1.1 One Hot Encoding:

- The length of the one-hot vector depends on the number of unique words in the vocabulary

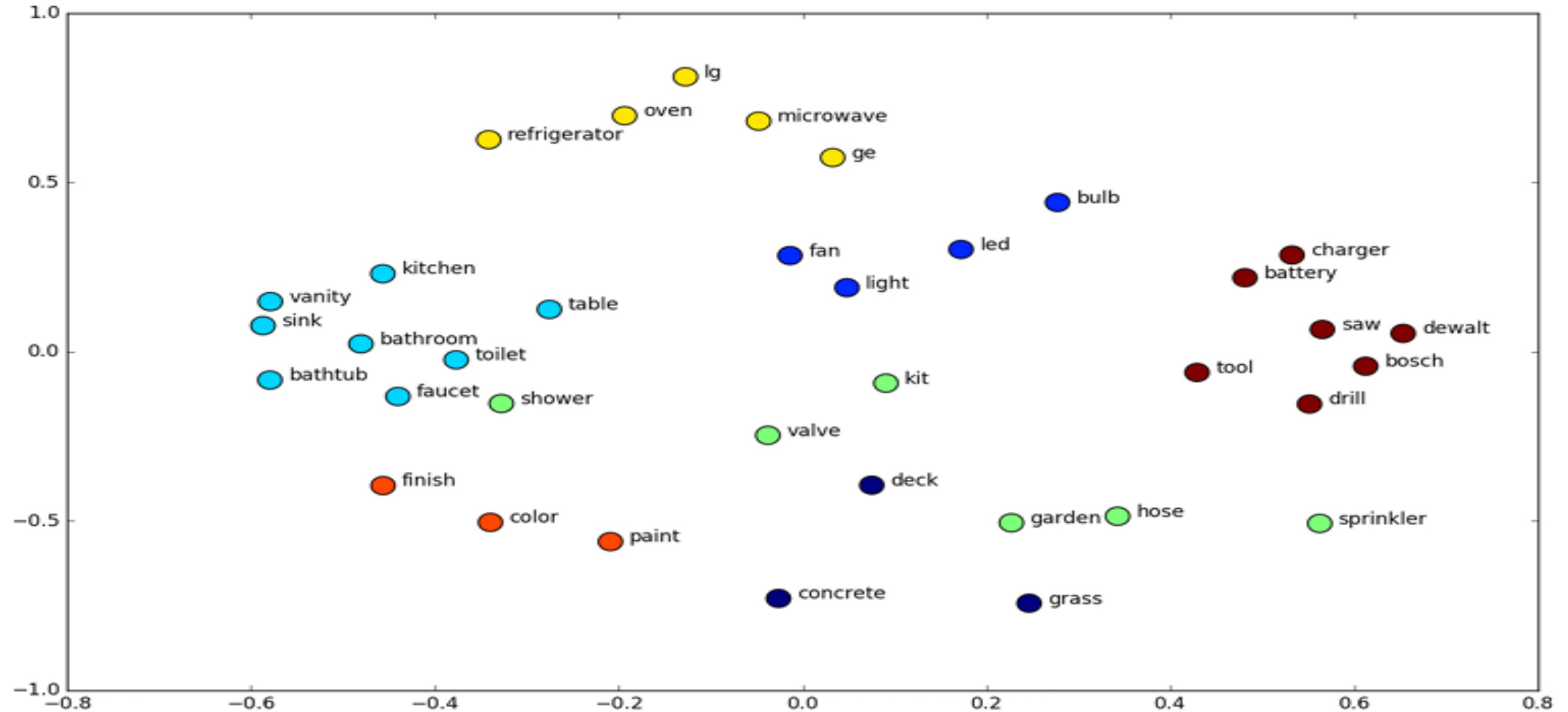
Hello, I'm Ironman. I have Friday AI

One Hot Encoding

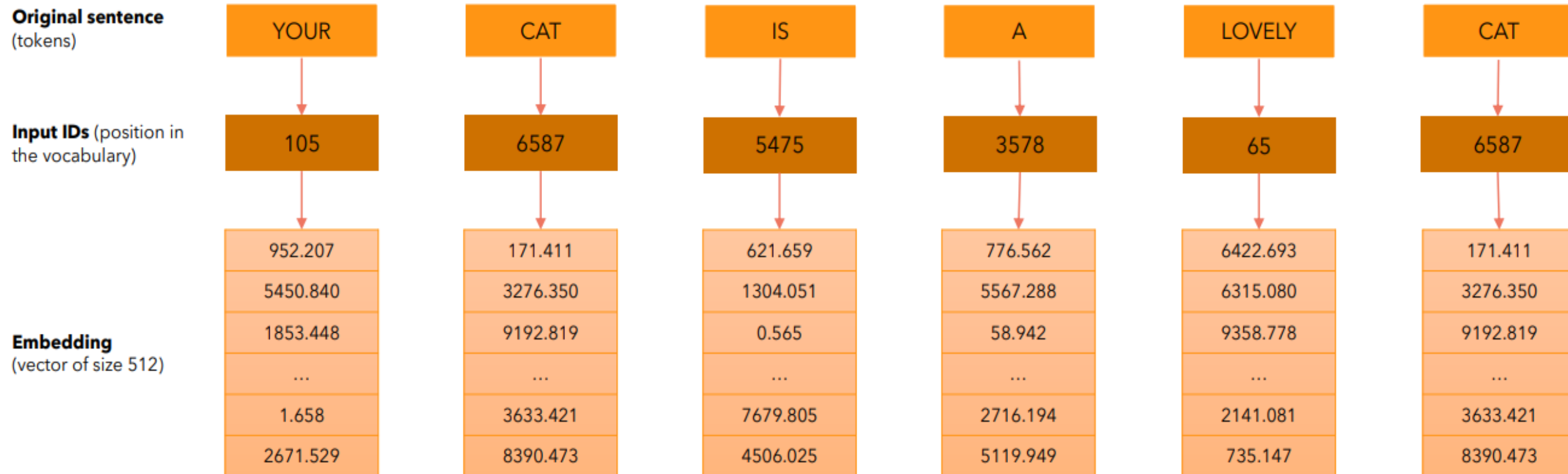


AI	1	0	0	0	0	0	0
Ironman	0	1	0	0	0	0	0
Friday	0	0	1	0	0	0	0
have	0	0	0	1	0	0	0
Hello	0	0	0	0	1	0	0
I	0	0	0	0	0	1	0
I'm	0	0	0	0	0	0	1

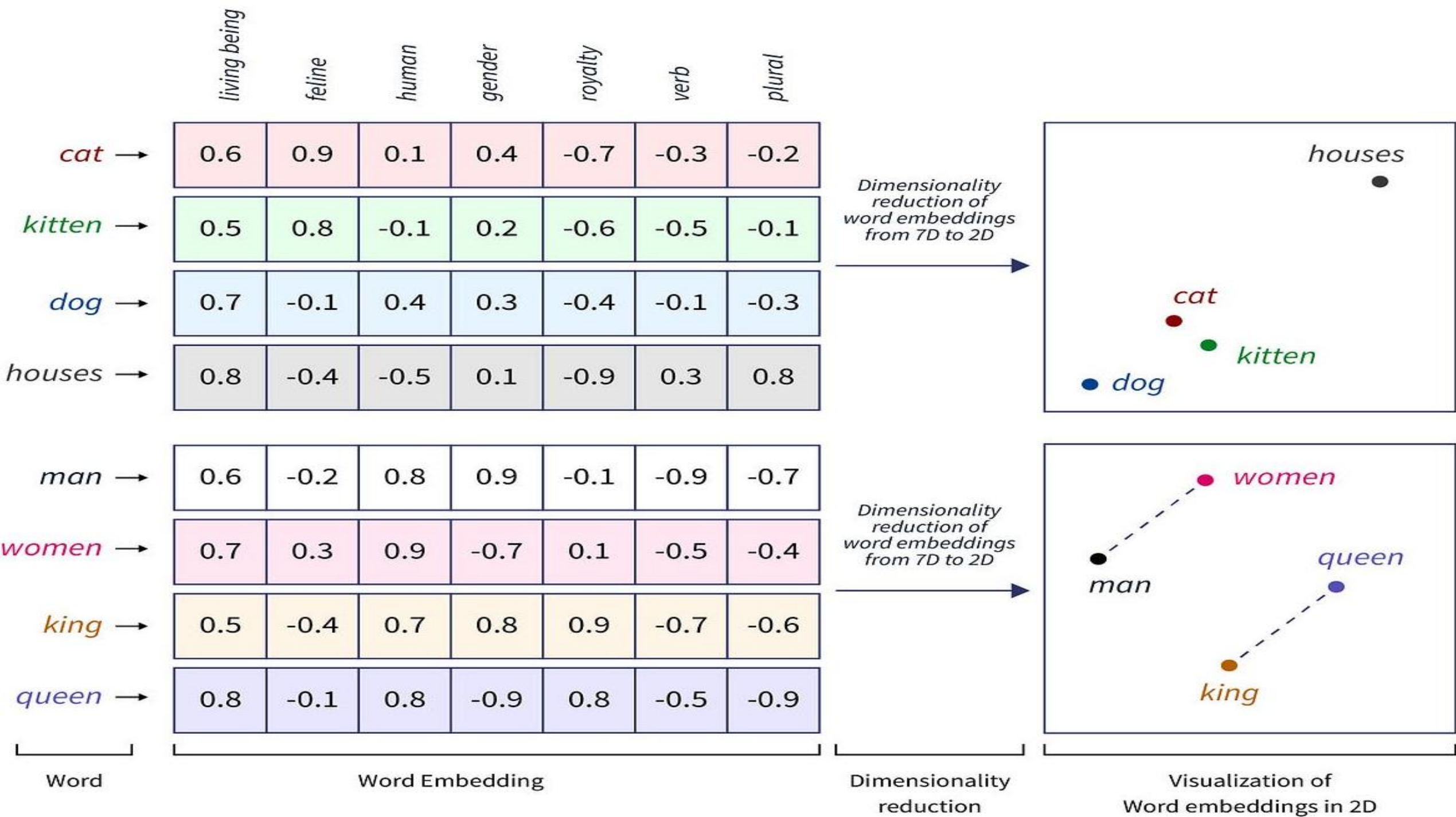
1.2- Input Embedding



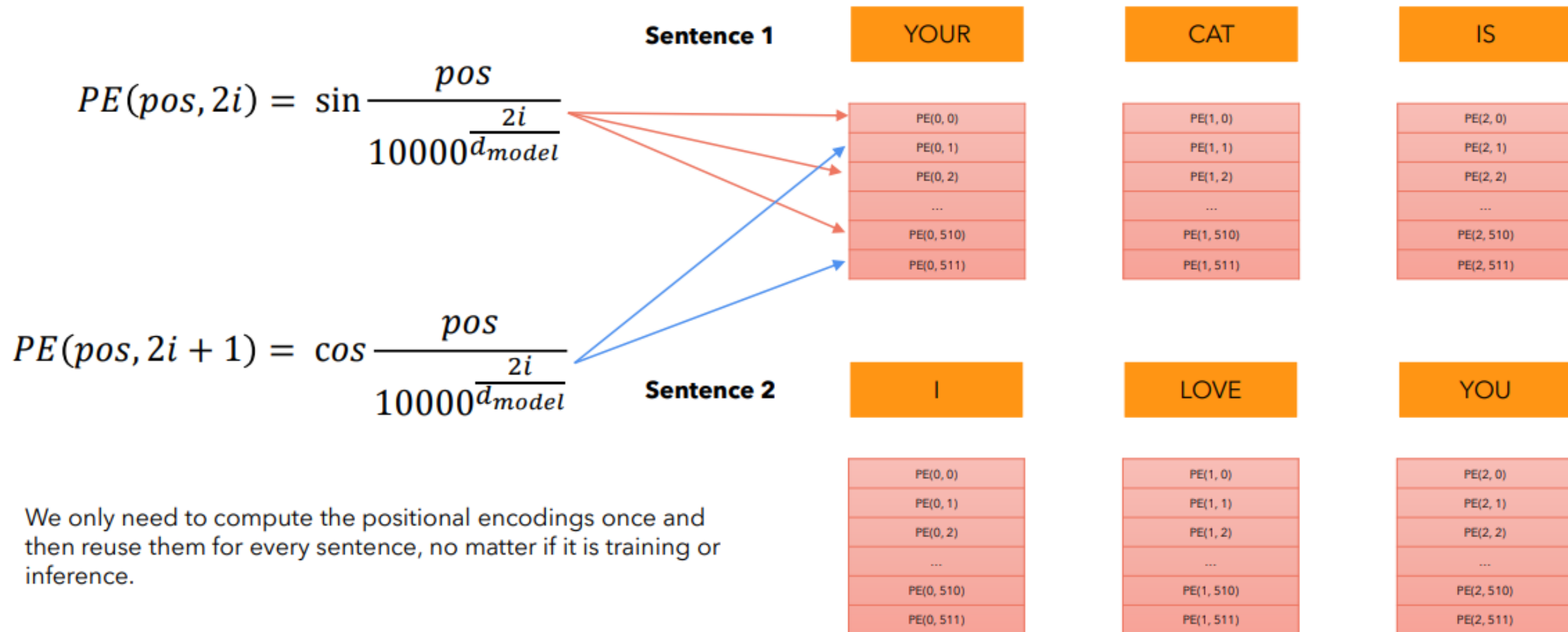
What is an Input Embedding ?



We define $d_{\text{model}} = 512$, which represents the size of the embedding vector of each word



Input Embedding + Positional Encoding = Attention Input



Why We Need Attention ?



بعد از یک روز طولانی در باغ وحش، وقتی به خانه برگشتم، یک لیوان شیر خوردم و شیر آبی که چکه می کرد را بستم؛ اما صدای غرش شیر قفس بغلی هنوز در گوشم می پیچید.



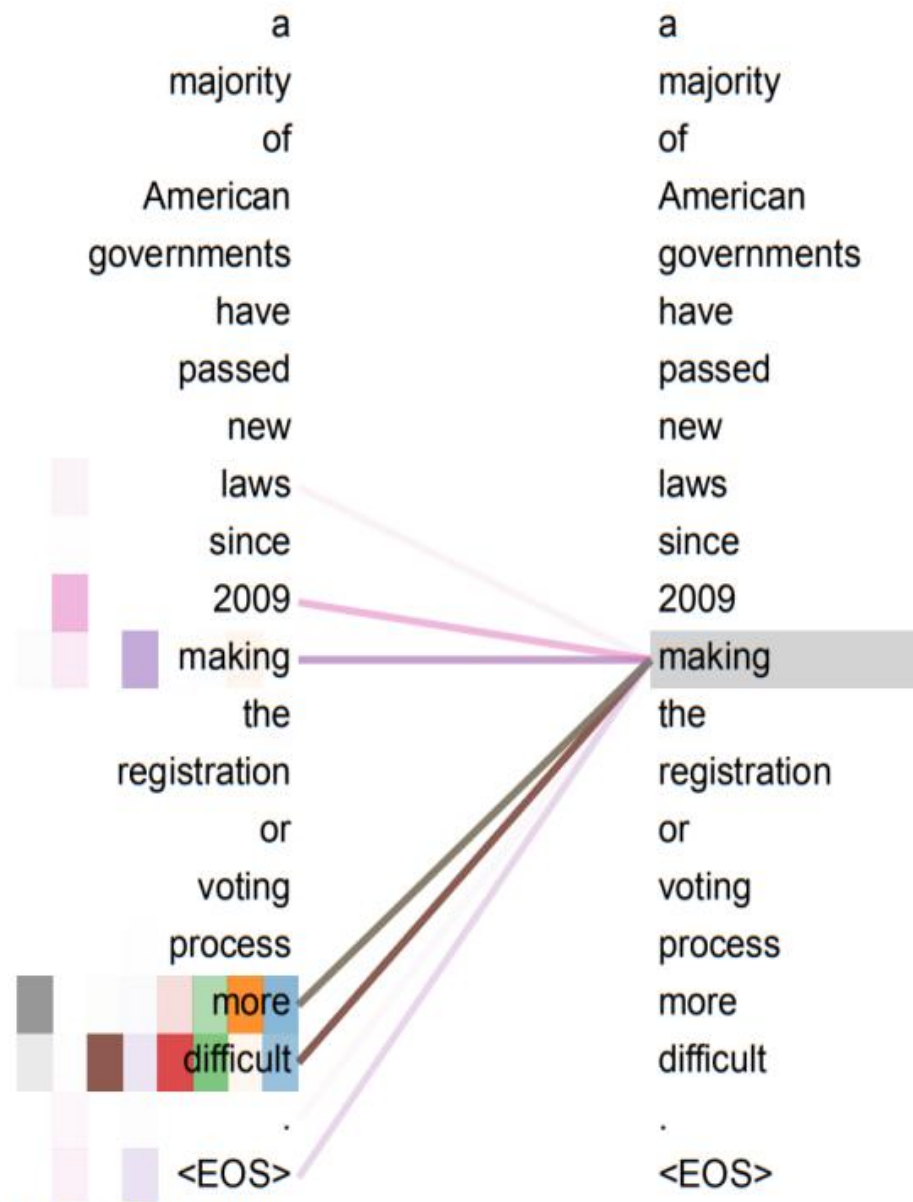


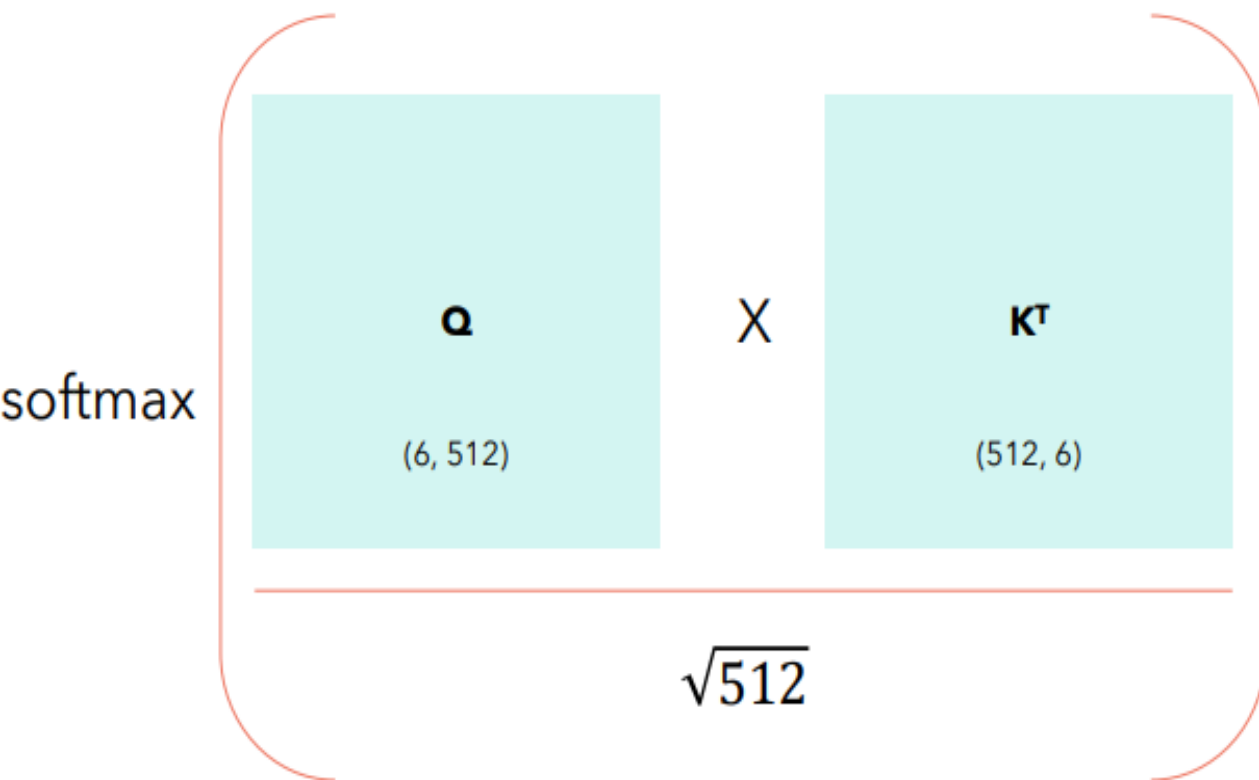
Figure adapted from Ashish Vaswani et al. Attention is All You Need paper

Attention Is All You Need

What Is Self Attention ?

Self attention allows the model to relate words to each other.

$$\text{Attention (Q,K,V)} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

[illegible]

How to compute Self-Attention?

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

	YOUR	CAT	IS	A	LOVELY	CAT
YOUR	0.268	0.119	0.134	0.148	0.179	0.152
CAT	0.124	0.278	0.201	0.128	0.154	0.115
IS	0.147	0.132	0.262	0.097	0.218	0.145
A	0.210	0.128	0.206	0.212	0.119	0.125
LOVELY	0.146	0.158	0.152	0.143	0.227	0.174
CAT	0.195	0.114	0.203	0.103	0.157	0.229

(6, 6)

X

V

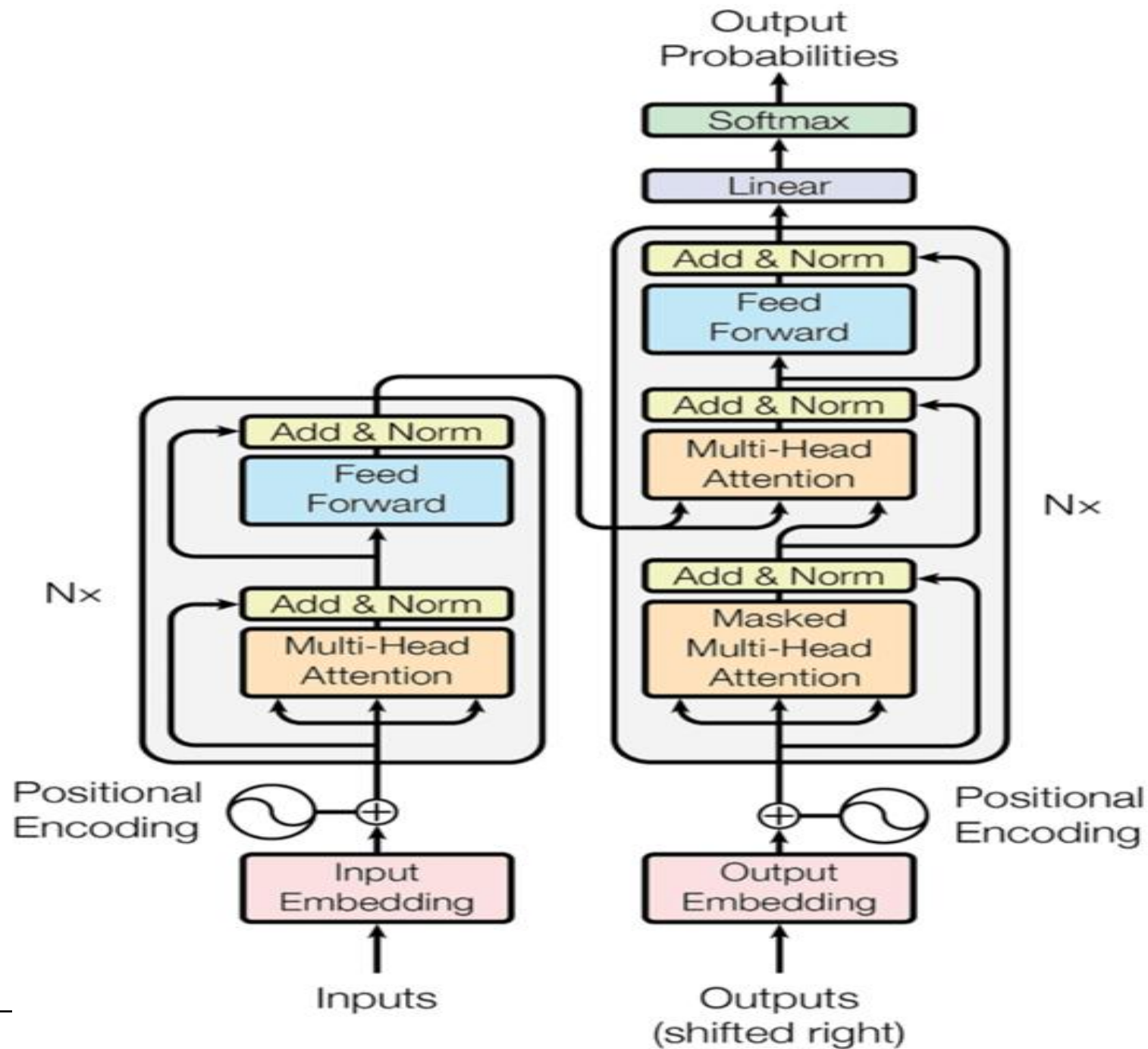
(6, 512)

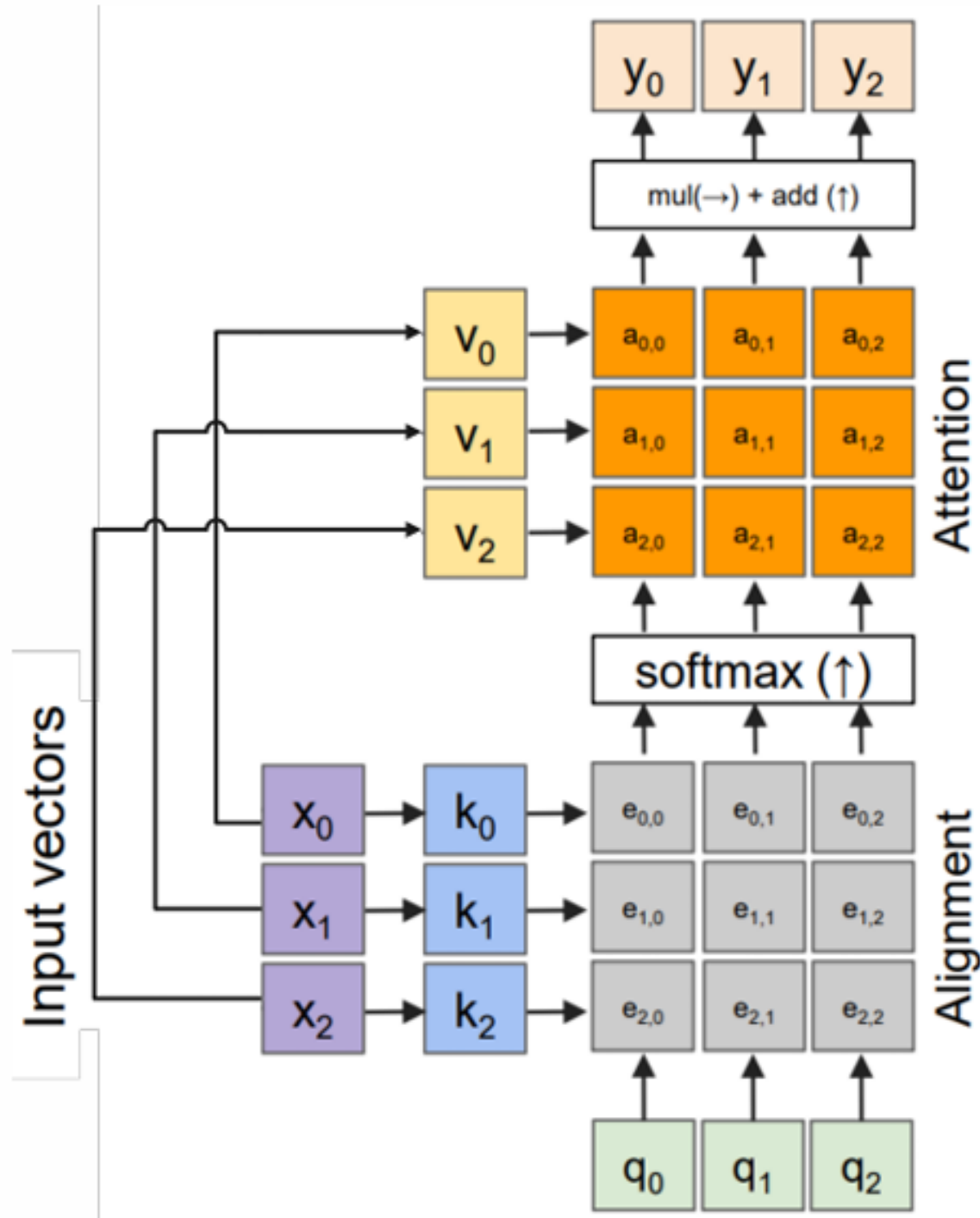
=

Attention

(6, 512)

Each row in this matrix captures not only the meaning (given by the embedding) or the position in the sentence (represented by the positional encodings) but also each word's interaction with other words.





Outputs:

Context vectors: \mathbf{y} (*shape*: D_v)

Operations:

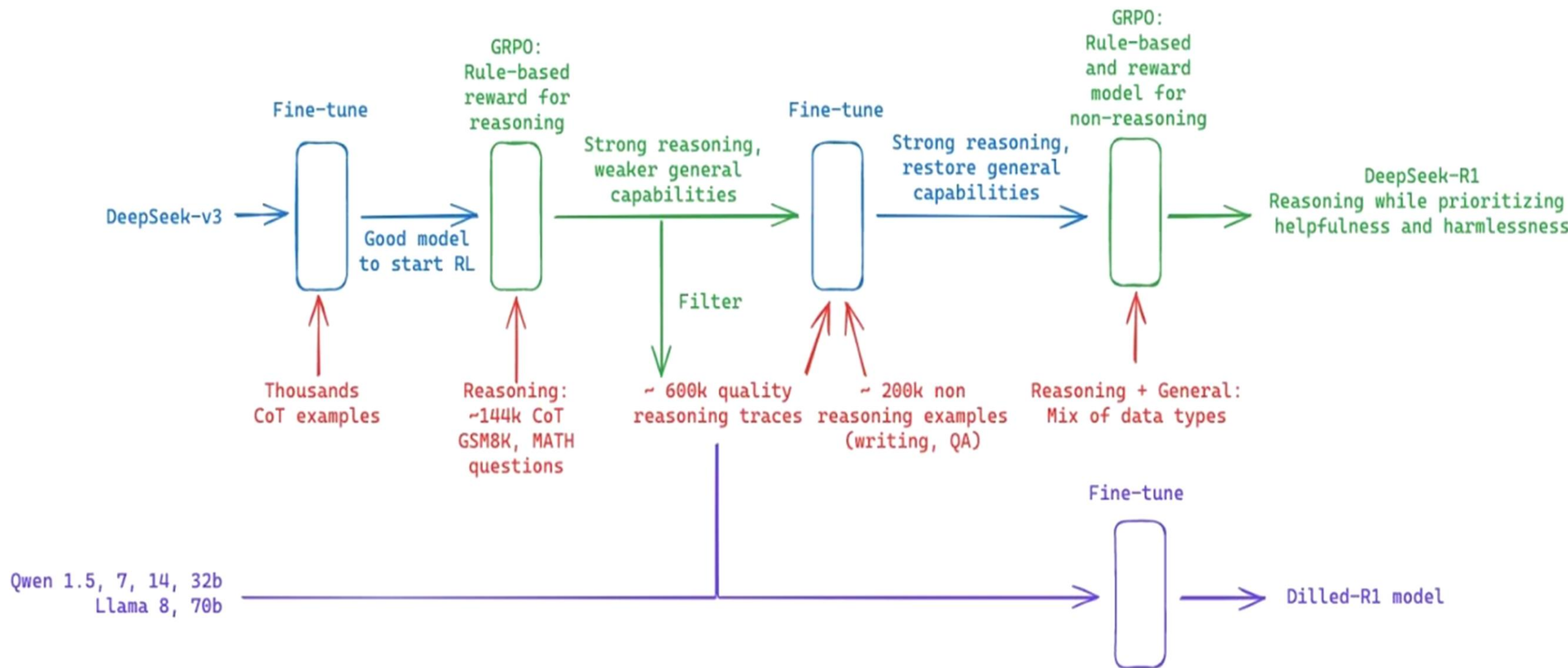
- Key vectors: $\mathbf{k} = W_k^T \mathbf{x}$
- Value vectors: $\mathbf{v} = W_v^T \mathbf{x}$
- **Alignment**: $e_{i,j} = \frac{\mathbf{q}_j \cdot \mathbf{k}_i}{\sqrt{D_k}}$
- **Attention**: $\mathbf{a} = \text{softmax}(\mathbf{e})$
- Output: $y_j = \sum_i a_{i,j} \mathbf{v}_i$

Inputs:

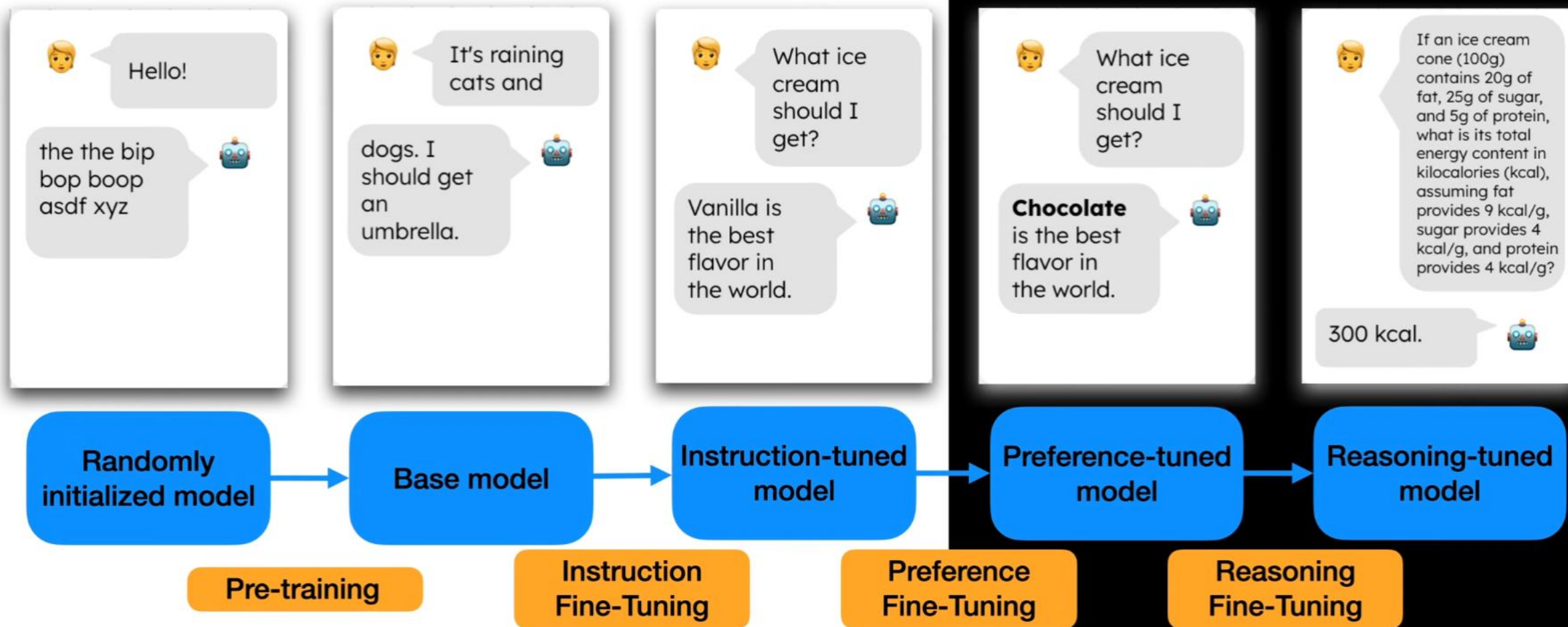
- Input vectors: \mathbf{x} (*shape*: $N \times D$)
- Queries: \mathbf{q} (*shape*: $M \times D_k$)

Deepseek R1, Why Does it Matter ?

Data, Fine-Tuning Phases, RL Phases, and Distillation for DeepSeek-R1



Training LLMs



Reinforcement Learning

Question: If $a > 1$, then the sum of the real solutions of $\sqrt{a - \sqrt{a + x}} = x$ is equal to

Response: <think>

To solve the equation $\sqrt{a - \sqrt{a + x}} = x$, let's start by squaring both \dots

$$\left(\sqrt{a - \sqrt{a + x}}\right)^2 = x^2 \implies a - \sqrt{a + x} = x^2.$$

Rearrange to isolate the inner square root term:

$$(a - x^2)^2 = a + x \implies a^2 - 2ax^2 + (x^2)^2 = a + x \implies x^4 - 2ax^2 - x + (a^2 - a) = 0$$

\dots

Wait, wait. Wait. That's an aha moment I can flag here.

Let's reevaluate this step-by-step to identify if the correct sum can be \dots

We started with the equation:

$$\sqrt{a - \sqrt{a + x}} = x$$

First, let's square both sides:

$$a - \sqrt{a + x} = x^2 \implies \sqrt{a + x} = a - x^2$$

Next, I could square both sides again, treating the equation: \dots

\dots

An interesting “aha moment” of an intermediate version of DeepSeek-R1-Zero. The model learns to rethink using an anthropomorphic tone. This is also an aha moment for us, allowing us to witness the power and beauty of reinforcement learning.

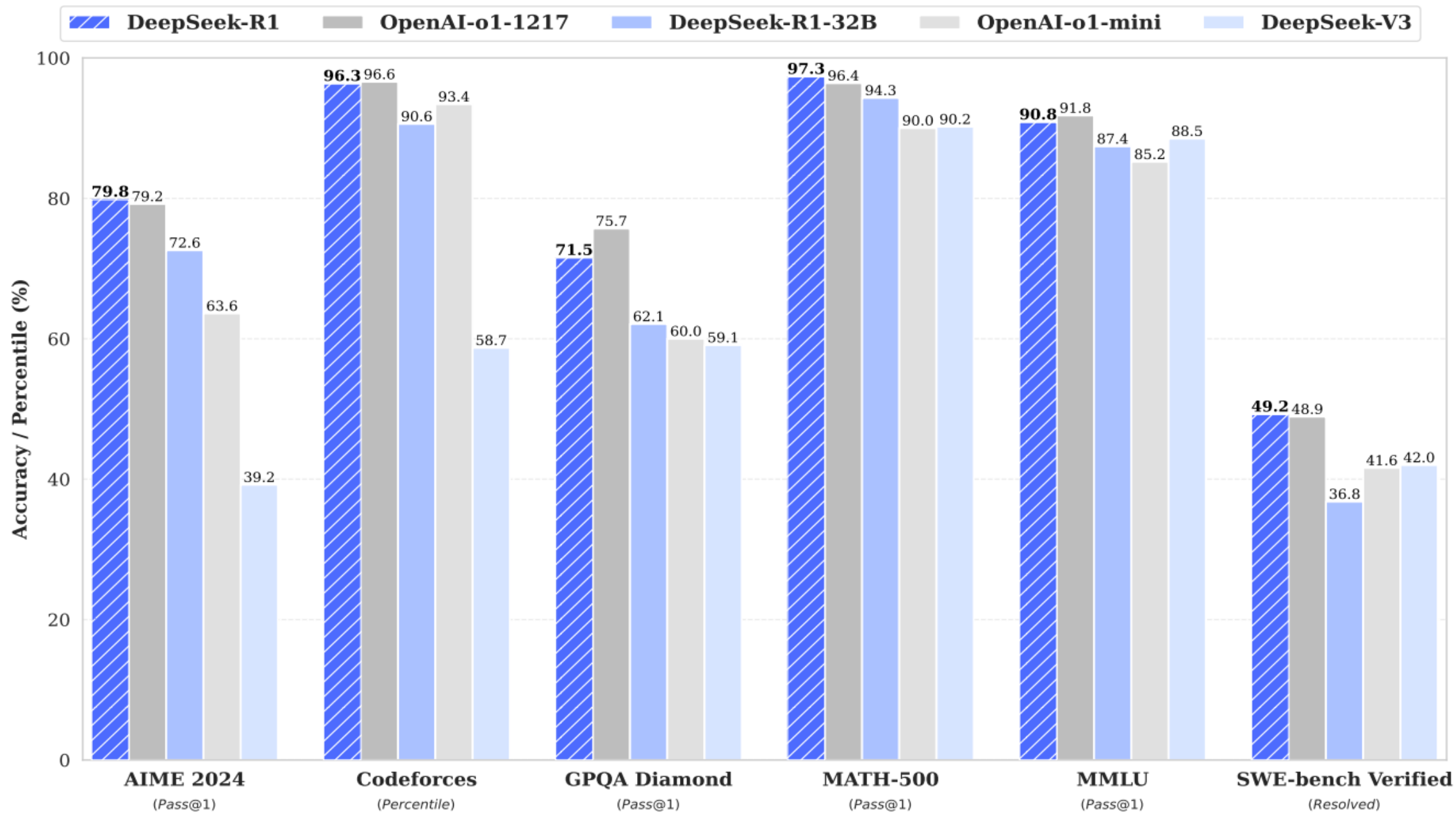


Figure 1 | Benchmark performance of DeepSeek-R1.

Model	AIME 2024		MATH-500	GPQA Diamond	LiveCode Bench	CodeForces
	pass@1	cons@64	pass@1	pass@1	pass@1	rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717
OpenAI-o1-mini	63.6	80.0	90.0	60.0	53.8	1820
QwQ-32B-Preview	50.0	60.0	90.6	54.5	41.9	1316
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	49.1	37.6	1189
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	59.1	53.1	1481
DeepSeek-R1-Distill-Qwen-32B	72.6	83.3	94.3	62.1	57.2	1691
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205
DeepSeek-R1-Distill-Llama-70B	70.0	86.7	94.5	65.2	57.5	1633

My Portfolio Website:
alirezatorabidev.vercel.app

Files Available at :

github.com/nullxxnerd/deepseekR1-presentation

References:

Sharif Machine Learning Course

<https://github.com/hkproj>

[https://github.com/deepseek-](https://github.com/deepseek-ai/DeepSeek-R1/blob/main/DeepSeek_R1.pdf)

[ai/DeepSeek-](https://github.com/deepseek-ai/DeepSeek-R1/blob/main/DeepSeek_R1.pdf)

[R1/blob/main/DeepSeek_R1.pdf](https://github.com/deepseek-ai/DeepSeek-R1/blob/main/DeepSeek_R1.pdf)