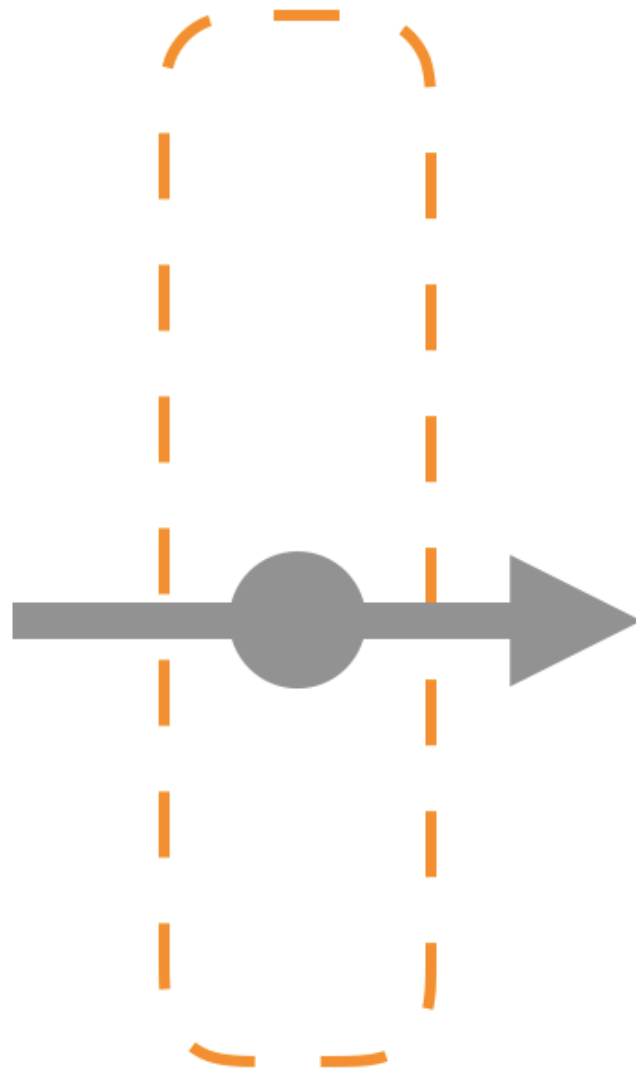
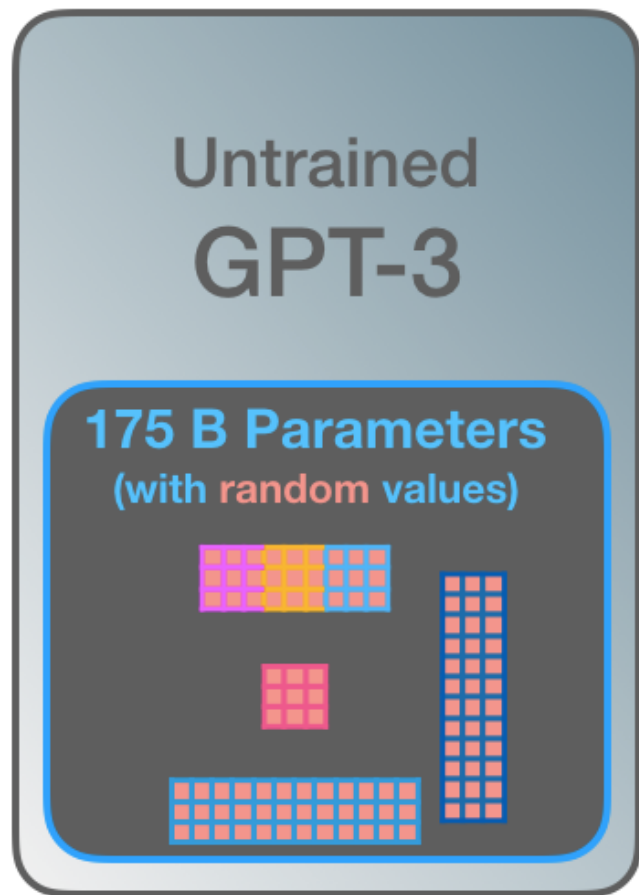


# The Massive GPT-3

## Unsupervised Pre-training



**Input Prompt:**

Recite the first law of robotics



**Output:**



## Unsupervised Pre-training

Untrained  
GPT-3

**Expensive training on massive datasets**

**Dataset:** 300 billion tokens of text

**Objective:** Predict the next word

**Example:**

a

robot

must

?

Training is the process of exposing the model to lots of text. That process has been completed. All the experiments you see now are from that one trained model. It was estimated to cost 355 GPU years and cost \$4.6m.

**Text:** Second Law of Robotics: A robot must obey the orders given it by human beings



**Generated training examples**

**Example #**

**Input (features)**

**Correct output (labels)**

**1**

Second law of robotics :

a

**2**

Second law of robotics : a

robot

**3**

Second law of robotics : a robot

must

...

## Unsupervised Pre-training

Input (features)

a

robot

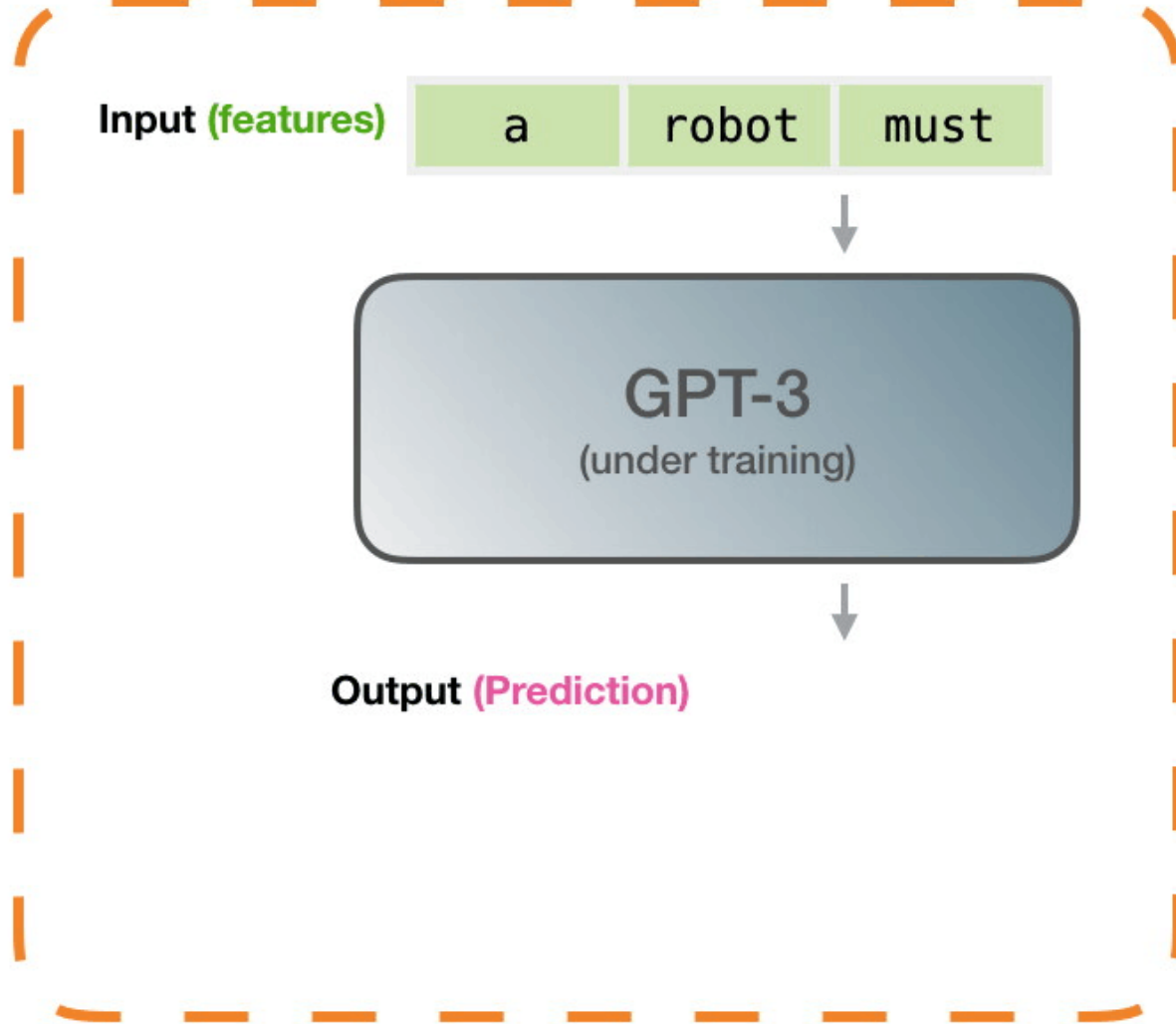
must

GPT-3  
(under training)

Output (Prediction)

Correct output (label):

obey



1 2 3 4 5 6

... 2048

Input

Recite the first law of robotics

GPT-3

Output:



GPT-3

robotics

1

2

3

1- Convert word  
into vector

2- Magic

3- Convert vector  
into word



**Vector** (I think of size 12,288)

**Embedding of** `robotics`  
+ positional encoding for position #6



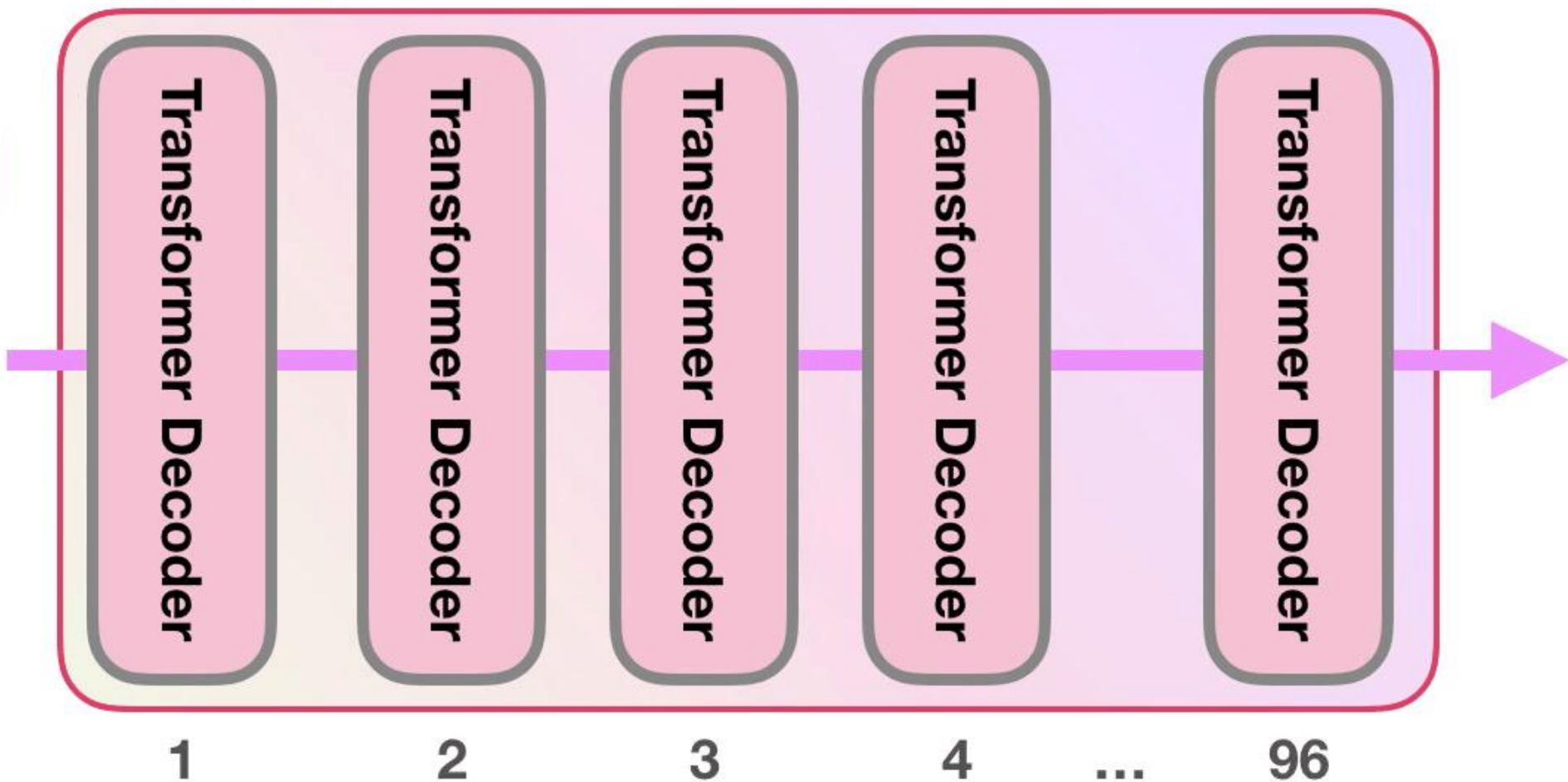
**Vector** (I think of size 12,288)

**Prediction result**



GPT-3

robotics



A

# GPT-3 vs GPT-2

- Version 3 takes the GPT model to a whole new level as it's trained on a whopping 175 billion parameters (which is over 10x the size of its predecessor, GPT-2).
- GPT-3 was created to be more robust than GPT-2 in that it is capable of handling more niche topics. GPT-2 was known to have poor performance when given tasks in specialized areas such as music and storytelling. GPT-3 can now go further with tasks such as **answering questions, writing essays, text summarization, language translation, and generating computer code**. The ability for it to be able to generate computer code is already a major feat unto itself.

# References

- Animations: <https://jalammar.github.io/how-gpt3-works-visualizations-animations/>