GPT

To Language Model and Beyond!

by QHDuan

What is a Language Model?

What is this:

What is a Language Model?

What is this:

$23.10

What is a Language Model?

What is this:

Come on, let’s go

What is a Language Model?

What is this:

1234 jklmn

Encoder vs Decoder

Encoder : Know all context

Decoder : Know only previous context

Encoder Only

仔细读故事(例文)，在“空”上填上合适的词

Decoder Only

开局一个字，往后写啥都行

Encoder vs Decoder

假设给定一个故事(例文)，把它改写为另一个故事(作文)

Encoder：Know all context，读例文的过程，我们是知道上下文，我们可以从前向后读，从后向前读，从中间读，随便瞎读

Decoder：Know only previous context，写作文的过程，必须一个字一个字写，你只知道自己在前面写了什么，不知道自己之后会写什么

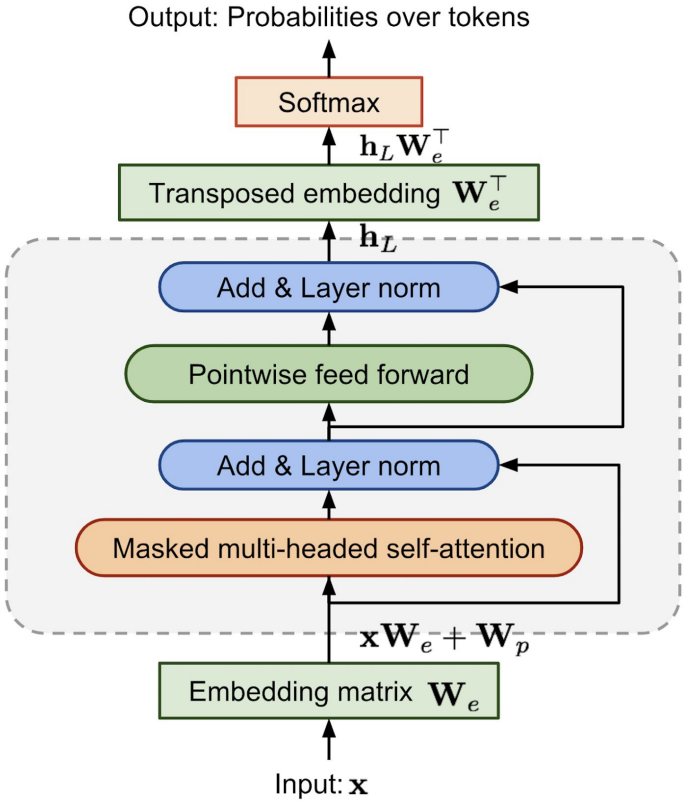
GPT vs BERT

GPT = N x Transformer Decoder，只用Decoder的方式构建语言模型

BERT = N x Transformer Encoder，只用Encoder的方式构建语言模型

NMT = M * Encoder + N * Decoder，机器翻译，既有Encoder，也有Decoder

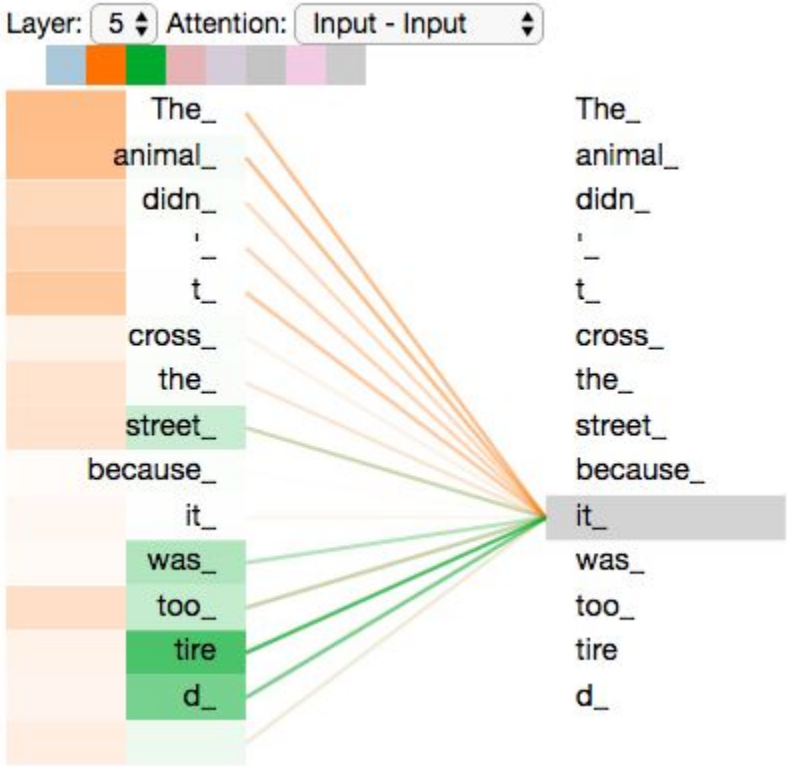
GPT-1‘s Transformer Block



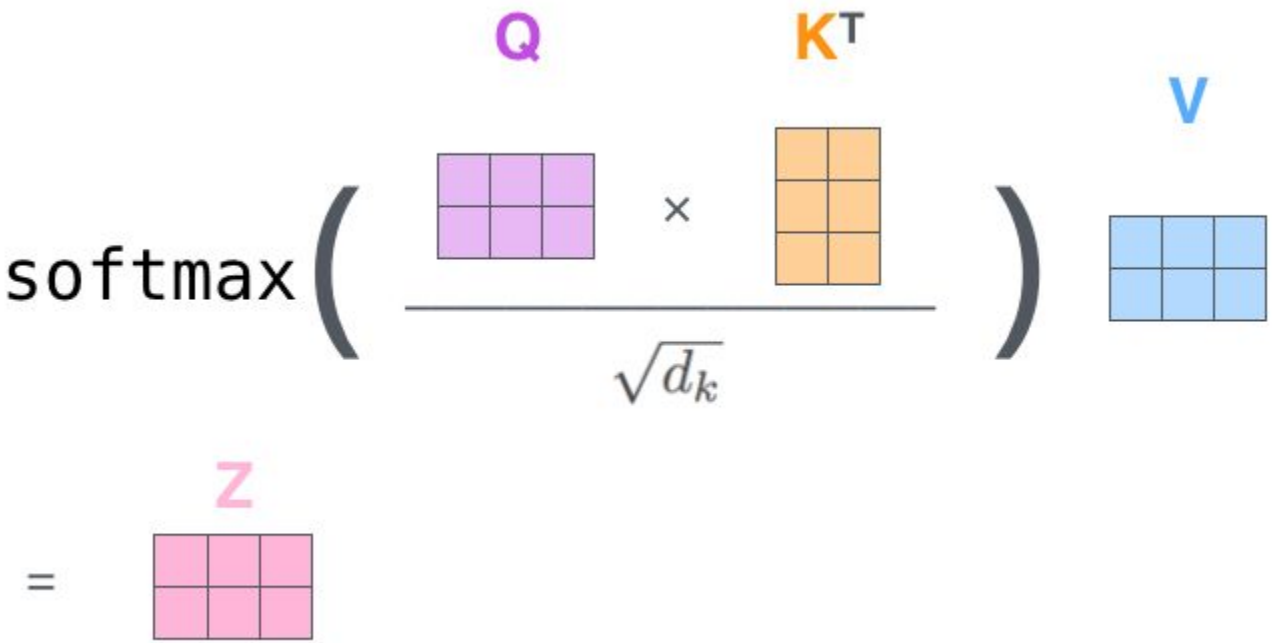
Transformer Block
Repeat x L=12

$$\mathbf{h}_\ell = \text{transformer\_block}(\mathbf{h}_{\ell-1})$$
$$\ell = 1, \ldots, L$$

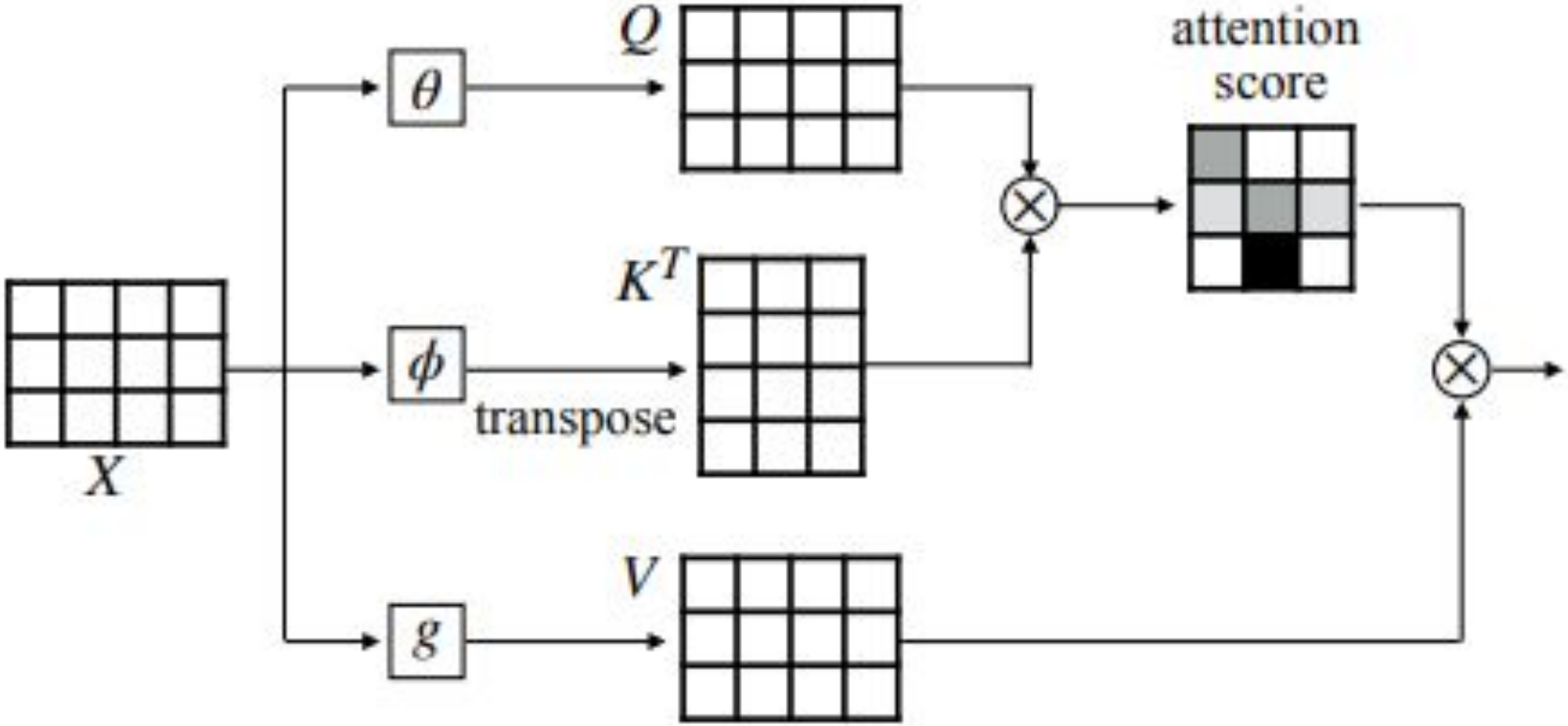
Attention


Layer: 5
Attention: Input - Input
The_
animal_
didn_
'_
t_
cross_
the_
street_
because_
it_
was_
too_
tire
d_
The_
animal_
didn_
'_
t_
cross_
the_
street_
because_
it_
was_
too_
tire
d_

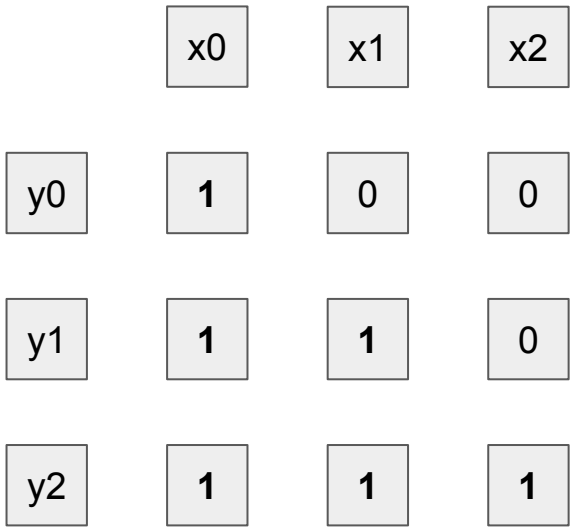
Self Attention



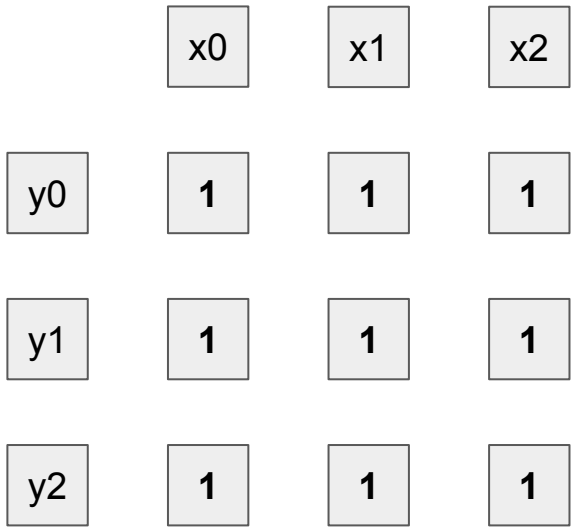
Self Attention



Attention Mask



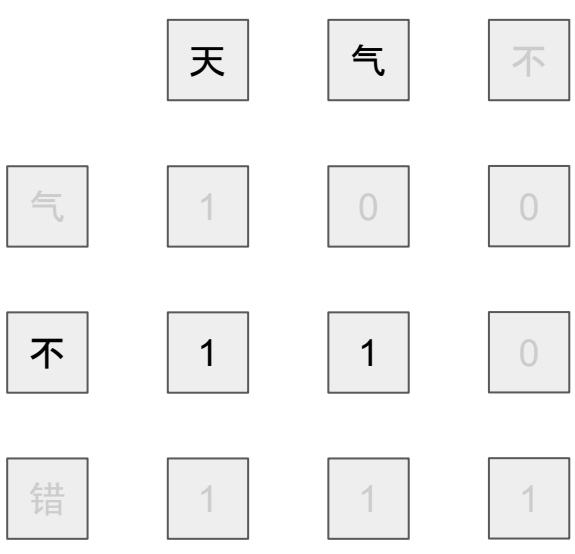
Decoder / GPT



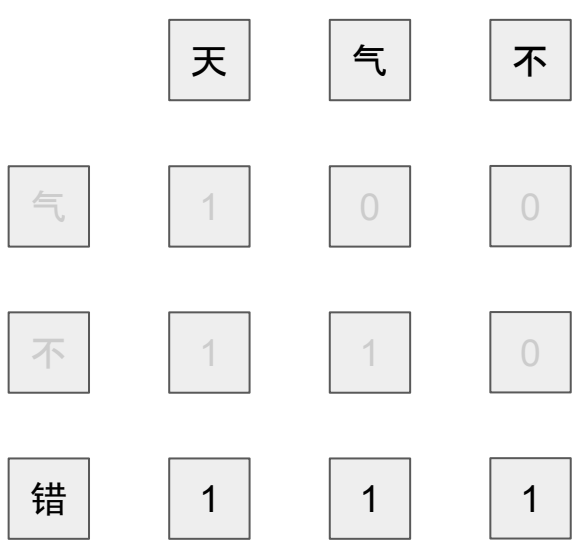
Encoder / BERT

Attention Mask Decoder

预测：气

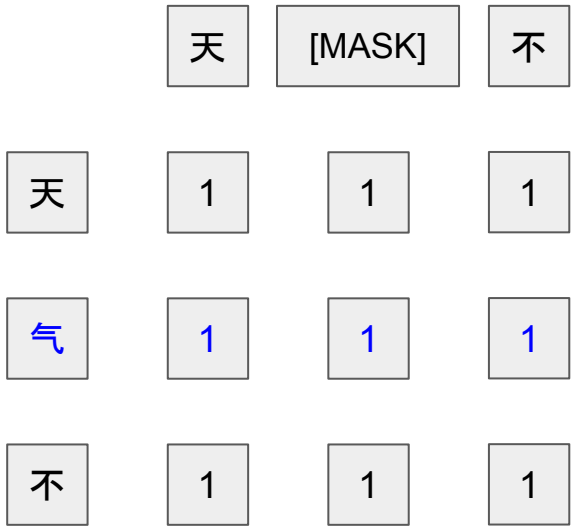


预测：不

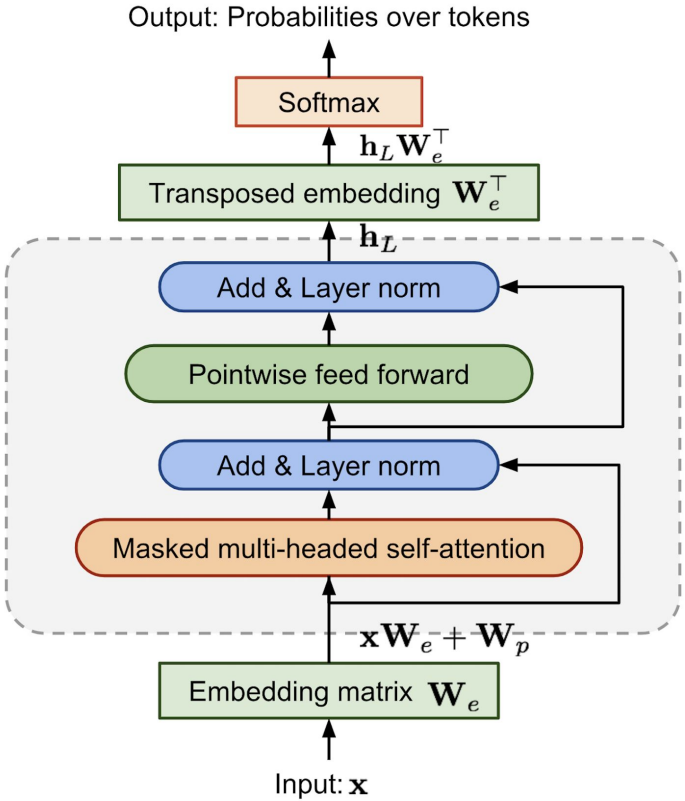


预测：错

Attention Mask Encoder



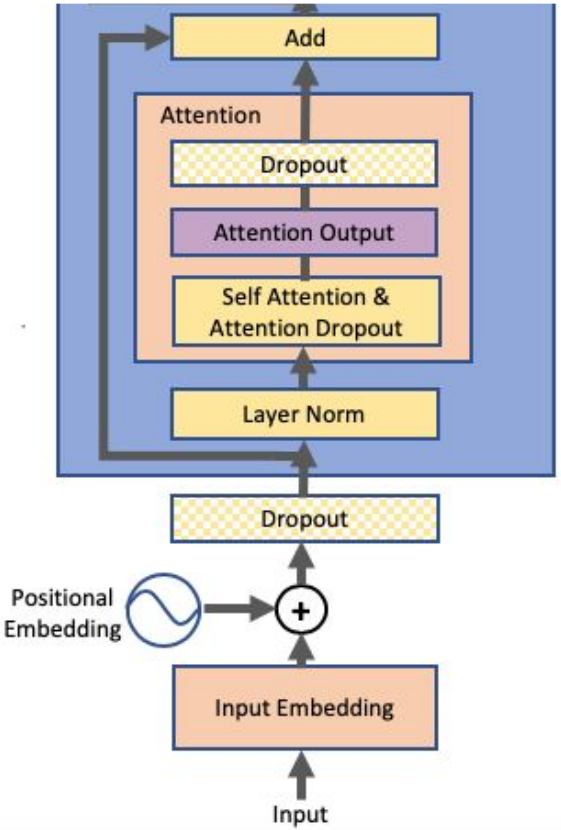
GPT-1‘s Transformer Block

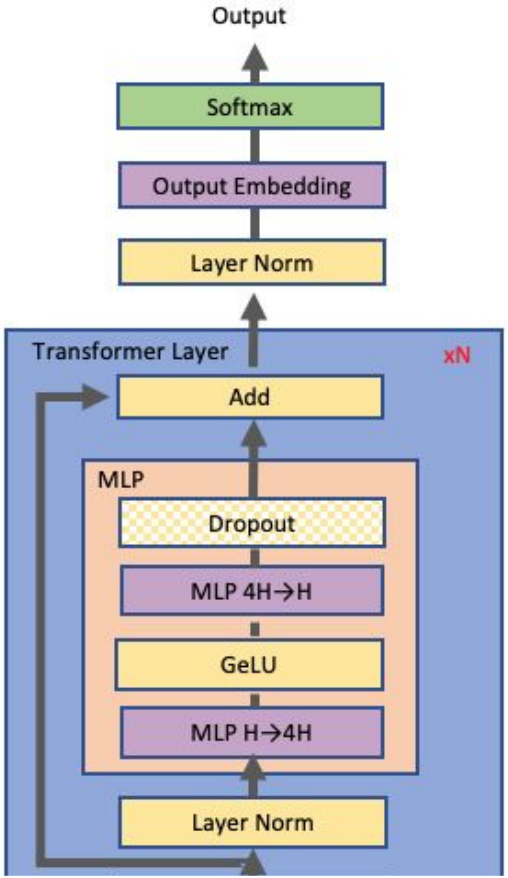


Transformer Block
Repeat x L=12

$$\mathbf{h}_\ell = \text{transformer\_block}(\mathbf{h}_{\ell-1})$$
$$\ell = 1, \ldots, L$$

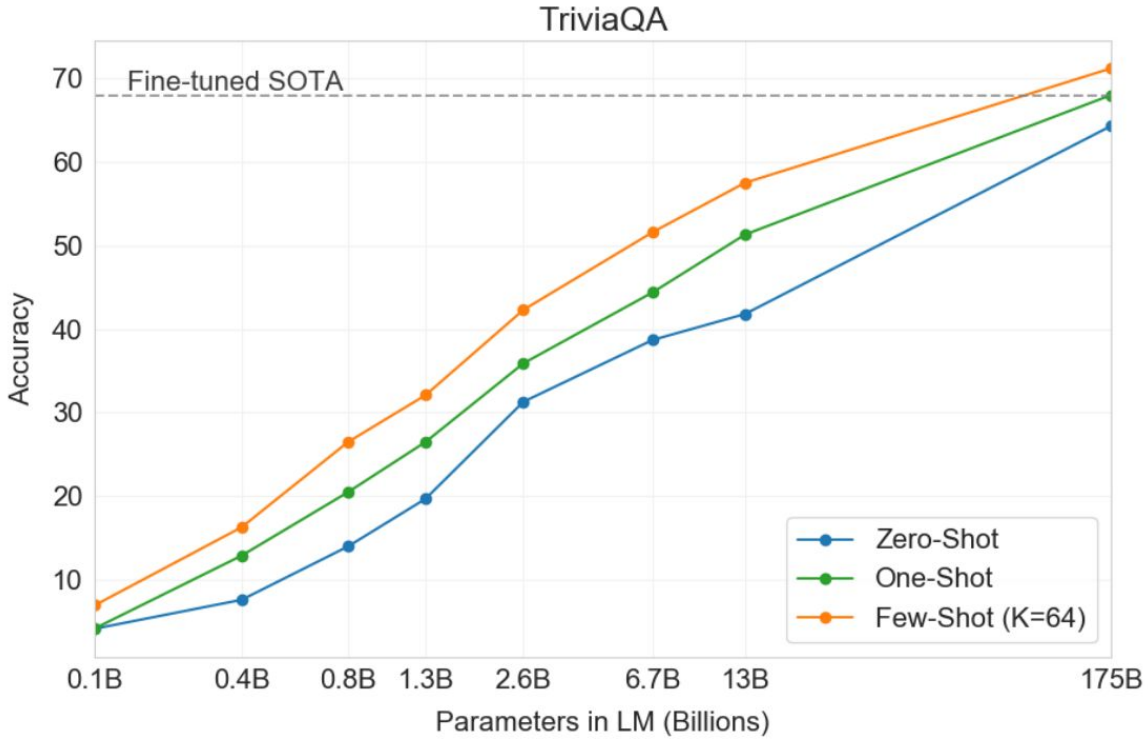
GPT 2 & 3


Add
Attention
Dropout
Attention Output
Self Attention &
Attention Dropout
Layer Norm
Dropout
Positional
Embedding
+
Input Embedding
Input


Output
Softmax
Output Embedding
Layer Norm
Transformer Layer
xN
Add
MLP
Dropout
MLP 4H→H
GeLU
MLP H→4H
Layer Norm

LET’s
CODING!!!

GPT 3



Zero-shot: 1.3B 约 20%, 175B 约 65%
Few-shot: 1.3B 约 32%, 175B 约 72%

GPT Application 1 — 写作文

```
sentence = '''《什么是爱》 爱是'''
sample(sentence, length=200, number=5)
```

《什么是爱》爱是一种感觉，一种患得患失的感觉，一种深情的感觉。感觉是看不见摸不到

《什么是爱》爱是出门随便看你一眼，都要用尽全身的力气把你留下来。爱是天天在打雷电

《什么是爱》爱是什么？”这样的问题你问我们大众是毫无用处的，你也不需要给出答案，I

《什么是爱》爱是什么爱是一个秘密的名字是一块痒痒的布如果可以人为它剪掉他的身体让

《什么是爱》爱是世间最美好的事，爱是相守，爱是共同成长，爱是互相守,爱是幸福的象征

GPT Application 2 — One-shot / Few-shot

One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

```
1  Translate English to French:        ← task description
2  sea otter => loutre de mer          ← example
3  cheese =>                           ← prompt
```

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
1  Translate English to French:        ← task description
2  sea otter => loutre de mer          ← examples
3  peppermint => menthe poivrée        ←
4  plush girafe => girafe peluche      ←
5  cheese =>                           ← prompt
```

GPT Application 3 — Zero-shot Learn

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1  Translate English to French:        ← task description
2  cheese =>                           ← prompt
```

GPT Application 4 — Variable Chating

```
chat(model, '讨厌你')
```

```
讨 厌 你
我 也 讨 厌 你
你 最 讨 厌
嘻 嘻
哈 哈 , 你 说 了 我 也 不 想 说 的
嘻 嘻 嘻 ~ 不 要 这 样 嘛 ~
不 要
讨 厌 的 人 不 理 你
哈 哈 ~
我 错 了
```

Thanks

https://github.com/QHDuan

Chinese Example:

Generate:
https://colab.research.google.com/drive/1toh8YkBL5Vnqm1sC8uoP0bH8wblnNRCO?usp=sharing

Zero-shot
https://drive.google.com/file/d/1DWlPdMfsAlHIOvrqrVCvgQL8wdpvPZtn/view?usp=sharing

Chat:
https://colab.research.google.com/drive/1GZYXT_3EdgFGQeZWfdOqUwnJ-d5JmoXF?usp=sharing