

Prompt Engineering guide

- Prompt engineering(提示工程) 用于**开发和优化提示以有效地将语言模型 (LM) 用于各种应用程序和研究主题**。 即时的工程技能有助于更好地理解大型语言模型 (LLM) 的功能和局限性。
- 研究人员使用提示工程来提高 LLM 在广泛的常见和复杂任务（例如问题回答和算术推理）上的能力。 开发人员使用提示工程来设计与 LLM 和其他工具交互的强大且有效的提示技术。
- 提示工程不仅仅是设计和开发提示。 它包含广泛的技能和技术，可用于与 LLM 进行交互和开发。 这是连接、构建和理解 LLM 功能的一项重要技能。 您可以使用提示工程来提高 LLM 的安全性并构建新功能，例如使用领域知识和外部工具增强 LLM。

All examples are tested with `text-davinci-003` using [OpenAI's playground](#) (opens in a new tab) unless otherwise specified. The model uses the default configurations, i.e., `temperature=0.7` and `top-p=1` .

LLM Settings

Temperature

简而言之，**温度越低，结果越确定，因为总是选择下一个最高可能的标记**。升高温度可能会导致更多的随机性，从而鼓励更多样化或更有创意的输出。实质上是在增加其他可能标记的权重。在应用方面，您可能希望对基于事实的 QA 等任务使用较低的温度值，以鼓励更真实和简洁的响应。对于诗歌生成或其他创造性任务，增加温度值可能是有益的。

Top_p

使用 top_p（一种称为核采样的温度采样技术），可以控制模型在生成响应时的确定性。如果正在寻找准确和事实的答案，请保持低值。如果正在寻找更多样化的响应，增加到更高的值。

一般建议是改变一个，而不是两个。在开始一些基本示例之前，结果可能会因使用的 LLM 版本而异。

Basics of Prompting

You can achieve a lot with simple prompts, but the quality of results depends on how much information you **provide it and how well-crafted it is**. A prompt can contain information like the *instruction* or *question* you are passing to the model and include other details such as *context*, *inputs*, or *examples*. You can use these elements to instruct the model better and as a result get better results.

Prompt Formatting

标准提示具有以下格式：

```
1 <Question>? or <Instruction>
```

Given the standard format above, **one popular and effective technique to prompting is referred to as *few-shot prompting* where you provide exemplars (i.e., demonstrations).** You can format few-shot prompts as follows:

```
1 <Question>?<Answer>
2 <Question>?<Answer>
3 <Question>?<Answer>
4 <Question>?
```

提示格式取决于手头的任务。例如，您可以执行一个简单的分类任务，并给出示例来演示该任务，如下所示：

Prompt:

```
1 This is awesome! // Positive
2 This is bad! // Negative
3 Wow that movie was rad! // Positive
4 What a horrible show! //
```

output:

```
1 Negative
```

Elements of a Prompt

A prompt contains any of the following elements:

Instruction - a specific task or instruction you want the model to perform

Context - external information or additional context that can steer the model to better responses

Input Data - the input or question that we are interested to find a response for

Output Indicator - the type or format of the output.

Designing Prompts

使用一些清晰的分隔符，如“###”来分隔指令和上下文。

Prompt:

```
1 ### Instruction ###
2 Translate the text below to Spanish:
3 Text: "hello!"
```

Output:

```
1 ¡Hola!
```

Prompt:

```
1 Extract the name of places in the following text.
2 Desired format: Place: <comma_separated_list_of_company_names>
3 Input: "Although these developments are encouraging to researchers, much is
```

Output:

```
1 Place: Champalimaud Centre for the Unknown, Lisbon
```

Few-shot Prompting

Let's try to add some examples to see if few-shot prompting improves the results.

Prompt:

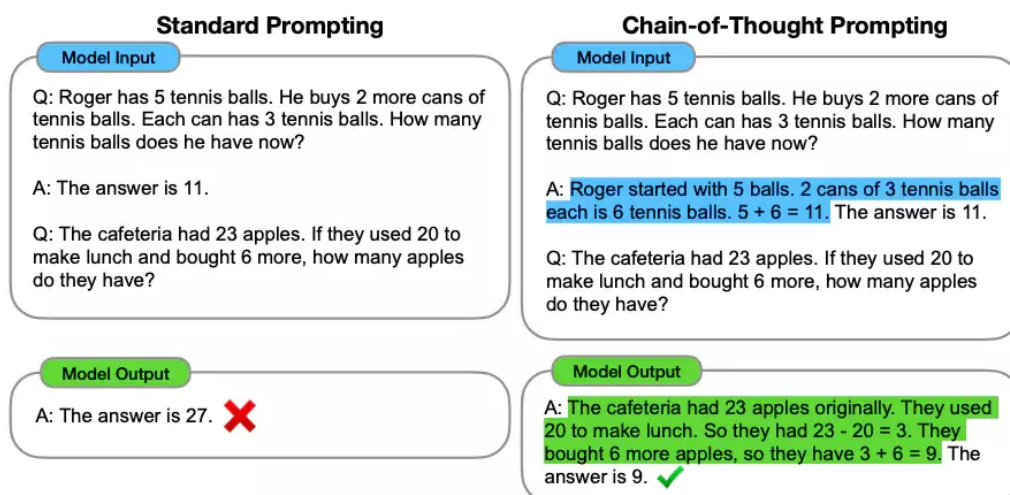
```
1 The odd numbers in this group add up to an even number: 4, 8, 9, 15, 12, 2,
2 A: The answer is False.
3 The odd numbers in this group add up to an even number: 17, 10, 19, 4, 8, 1
4 A: The answer is True.
5 The odd numbers in this group add up to an even number: 16, 11, 14, 4, 8, 1
6 A: The answer is True.
7 The odd numbers in this group add up to an even number: 17, 9, 10, 12, 13,
8 A: The answer is False.
9 The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7
10 A:
```

Output:

```
1 The answer is True.
```

这没用。似乎少样本提示不足以获得这种类型的推理问题的可靠响应。上面的示例提供了任务的基本信息。如果您仔细观察，我们引入的任务类型涉及几个更多的推理步骤。换句话说，如果我们将问题分解成步骤并向模型演示，这可能会有所帮助。最近，[思维链 \(CoT\) 提示](#) (opens in a new tab) (加一些推理过程)

Chain-of-Thought (CoT) Prompting



Prompt:

1 The odd numbers in this group add up to an even number: 4, 8, 9, 15, 12, 2,
 2 A: Adding all the odd numbers (9, 15, 1) gives 25. The answer is False.
 3
 4 The odd numbers in this group add up to an even number: 17, 10, 19, 4, 8, 1
 5 A: Adding all the odd numbers (17, 19) gives 36. The answer is True.
 6
 7 The odd numbers in this group add up to an even number: 16, 11, 14, 4, 8, 1
 8 A: Adding all the odd numbers (11, 13) gives 24. The answer is True.
 9
 10 The odd numbers in this group add up to an even number: 17, 9, 10, 12, 13,
 11 A: Adding all the odd numbers (17, 9, 13) gives 39. The answer is False.
 12
 13 The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7
 14 A:

Output:

1 Adding all the odd numbers (15, 5, 13, 7, 1) gives 41. The answer is False.

当我们提供推理步骤时，我们可以看到一个完美的结果。事实上，我们可以通过提供更少的例子来解决这个任务，即一个例子似乎就足够了：

Prompt:

1 The odd numbers in this group add up to an even number: 4, 8, 9, 15, 12, 2,
 2 A: Adding all the odd numbers (9, 15, 1) gives 25. The answer is False.
 3
 4 The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7
 5 A:

Output:

1 Adding all the odd numbers (15, 5, 13, 7, 1) gives 41. The answer is False.

Zero-shot coT Prompting

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The answer is 8. ✗

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are $16 / 2 = 8$ golf balls. Half of the golf balls are blue. So there are $8 / 2 = 4$ blue golf balls. The answer is 4. ✓

(c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: The answer (arabic numerals) is

(Output) 8 ✗

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: **Let's think step by step.**

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

Prompt:

```
1 I went to the market and bought 10 apples. I gave 2 apples to the neighbor and
```

Output:

```
1 11 apples
```

The answer is incorrect! Now Let's try with the special prompt.

Prompt:

```
1 I went to the market and bought 10 apples. I gave 2 apples to the neighbor and
2
3 Let's think step by step.
```

Output:

```
1 First, you started with 10 apples.
2 You gave away 2 apples to the neighbor and 2 to the repairman, so you had 6
3 Then you bought 5 more apples, so now you had 11 apples.
4 Finally, you ate 1 apple, so you would remain with 10 apples.
```

Self-Consistency

one of the more advanced techniques out there for prompt engineering is self-consistency.

Proposed by [Wang et al. \(2022\)](#)[\(opens in a new tab\)](#)

, self-consistency aims "**to replace the naive greedy decoding used in chain-of-thought prompting**". The idea is to sample multiple, diverse reasoning paths through few-shot CoT, and use the generations to select the most consistent answer. This helps to boost the performance of CoT prompting on tasks involving arithmetic and commonsense reason.

Prompt:

```
1 When I was 6 my sister was half my age. Now
2 I'm 70 how old is my sister?
```

Output:

```
1 35
```

The output is wrong! How may we improve this with self-consistency? Let's try it out. We will use the few-shot exemplars from Wang et al. 2022 :

Prompt:

```
1 Q: There are 15 trees in the grove. Grove workers will plant trees in the grove
2 there will be 21 trees. How many trees did the grove workers plant today?
```

3 A: We start with 15 trees. Later we have 21 trees. The difference must be the
 4 So, they must have planted $21 - 15 = 6$ trees. The answer is 6.
 5

6 Q: If there are 3 cars in the parking lot and 2 more cars arrive, how many cars
 7 A: There are 3 cars in the parking lot already. 2 more arrive. Now there are
 8

9 Q: Leah had 32 chocolates and her sister had 42. If they ate 35, how many pieces
 10 A: Leah had 32 chocolates and Leah's sister had 42. That means there were originally
 11 chocolates. 35 have been eaten. So in total they still have $74 - 35 = 39$ chocolates.
 12

13 Q: Jason had 20 lollipops. He gave Denny some lollipops. Now Jason has 12 lollipops
 14 did Jason give to Denny?
 15 A: Jason had 20 lollipops. Since he only has 12 now, he must have given the
 16 lollipops he has given to Denny must have been $20 - 12 = 8$ lollipops. The answer is 8.
 17

18 Q: Shawn has five toys. For Christmas, he got two toys each from his mom and dad
 19 he have now?
 20 A: He has 5 toys. He got 2 from mom, so after that he has $5 + 2 = 7$ toys. Then
 21 in total he has $7 + 2 = 9$ toys. The answer is 9.
 22

23 Q: There were nine computers in the server room. Five more computers were installed
 24 monday to thursday. How many computers are now in the server room?
 25 A: There are 4 days from monday to thursday. 5 computers were added each day
 26 20 computers were added. There were 9 computers in the beginning, so now there are
 27 The answer is 29.
 28

29 Q: Michael had 58 golf balls. On tuesday, he lost 23 golf balls. On wednesday
 30 golf balls did he have at the end of wednesday?
 31 A: Michael initially had 58 balls. He lost 23 on Tuesday, so after that he had 35
 32 Wednesday he lost 2 more so now he has $35 - 2 = 33$ balls. The answer is 33.
 33

34 Q: Olivia has \$23. She bought five bagels for \$3 each. How much money does she have left?
 35 A: She bought 5 bagels for \$3 each. This means she spent 15
 36

37 Q: When I was 6 my sister was half my age. Now I'm 70 how old is my sister?
 38 A:

Output 1:

1 When I was 6 my sister was half my age, so she was 3. Now I am 70, so she is 64.

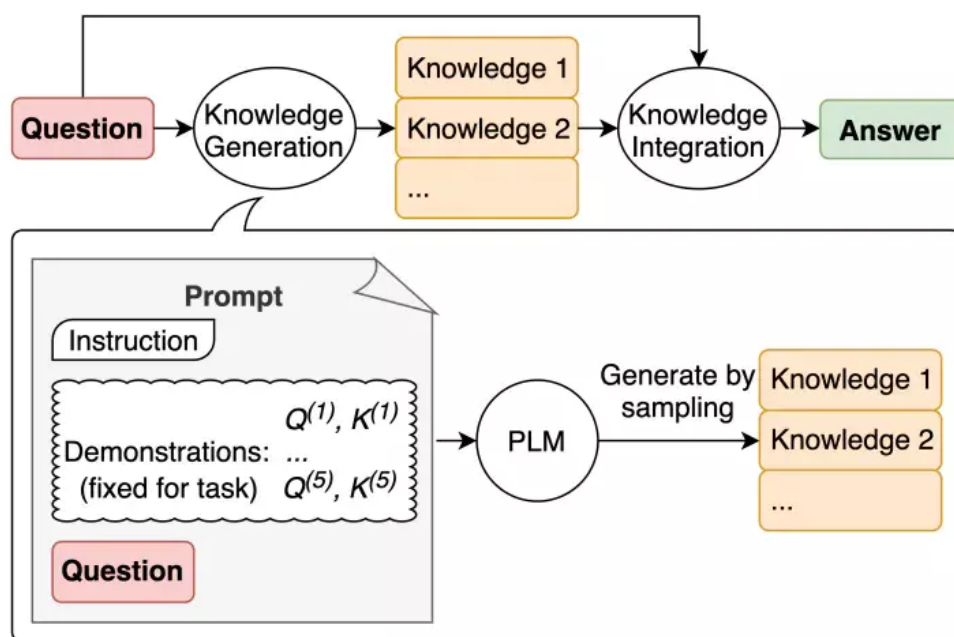
Output 2:

1 When the narrator was 6, his sister was half his age, which is 3. Now that the narrator is 70, his sister is 64.

Output 3:

1 When I was 6 my sister was half my age, so she was 3. Now I am 70, so she is 64.

Generated Knowledge Prompting



LLM 继续得到改进，其中一种流行的技术是能够融合知识或信息，以帮助模型做出更准确的预测。使用类似的思路，模型是否也可以在做出预测之前用于生成知识呢？这就是 [Liu 等人 2022](#) (opens in a new tab) 的论文所尝试的——生成知识以作为提示的一部分。特别是，这对于常识推理等任务有多大帮助？

让我们尝试一个简单的提示：

提示：

1 高尔夫球的一部分是试图获得比其他人更高的得分。是或否？

输出：

1 是。

这种错误揭示了 LLM 在执行需要更多关于世界的知识的任务时的局限性。我们如何通过生成知识来改进呢？

首先，我们生成一些“知识”：

首先，我们生成一些“知识”：

提示：

1 输入：希腊比墨西哥大。

2 知识：希腊的面积约为131,957平方公里，而墨西哥的面积约为1,964,375平方公里，使墨西哥比

3

4 输入：眼镜总是会起雾。

5 知识：当你的汗水、呼吸和周围的湿度中的水蒸气落在冷的表面上，冷却并变成微小的液滴时，会
6
7 输入：鱼有思考能力。
8 知识：鱼比它们看起来更聪明。在许多领域，如记忆力，它们的认知能力与或超过非人类灵长类动物。
9
10 输入：一个人一生中吸烟很多香烟的常见影响是患肺癌的几率高于正常水平。
11 知识：那些一生中平均每天吸烟不到一支香烟的人，患肺癌的风险是从不吸烟者的9倍。在每天吸烟
12
13 输入：一块石头和一颗卵石大小相同。
14 知识：卵石是一种根据Udden-Wentworth沉积学尺度的颗粒大小为4到64毫米的岩屑。卵石通常被
15
16 输入：高尔夫球的一部分是试图获得比其他人更高的得分。
17 知识：

知识 1：

1 高尔夫球的目标是以最少的杆数打完一组洞。一轮高尔夫球比赛通常包括18个洞。每个洞在标准高

知识 2：

1 高尔夫球是一项精密的球杆运动，竞争选手（或高尔夫球手）使用许多类型的球杆，用最少的杆数

我们使用 [Liu 等人 2022](#)

([opens in a new tab](#))

的论文中提供的提示。

下一步是将知识整合并得出预测。我将问题重新格式化为 QA 格式，以指导答案格式。

首先，我们生成一些“知识”：

提示：

1 输入：希腊比墨西哥大。
2 知识：希腊的面积约为131,957平方公里，而墨西哥的面积约为1,964,375平方公里，使墨西哥比
3
4 输入：眼镜总是会起雾。
5 知识：当你的汗水、呼吸和周围的湿度中的水蒸气落在冷的表面上，冷却并变成微小的液滴时，会
6
7 输入：鱼有思考能力。
8 知识：鱼比它们看起来更聪明。在许多领域，如记忆力，它们的认知能力与或超过非人类灵长类动物。
9
10 输入：一个人一生中吸烟很多香烟的常见影响是患肺癌的几率高于正常水平。
11 知识：那些一生中平均每天吸烟不到一支香烟的人，患肺癌的风险是从不吸烟者的9倍。在每天吸烟
12
13 输入：一块石头和一颗卵石大小相同。
14 知识：卵石是一种根据Udden-Wentworth沉积学尺度的颗粒大小为4到64毫米的岩屑。卵石通常被
15
16 输入：高尔夫球的一部分是试图获得比其他人更高的得分。
17 知识：

知识 1 :

- 1 高尔夫球的目标是以最少的杆数打完一组洞。一轮高尔夫球比赛通常包括18个洞。每个洞在标准高

知识 2 :

- 1 高尔夫球是一项精密的球杆运动，竞争选手（或高尔夫球手）使用许多类型的球杆，用最少的杆数

我们使用 [Liu 等人 2022\(opens in a new tab\)](#) 的论文中提供的提示。

下一步是将知识整合并得出预测。我将问题重新格式化为 QA 格式，以指导答案格式。

提示 :

- 1 问题：高尔夫球的一部分是试图获得比其他人更高的得分。是或否？
- 2
- 3 知识：高尔夫球的目标是以最少的杆数打完一组洞。一轮高尔夫球比赛通常包括18个洞。每个洞在
- 4
- 5 解释和答案：

答案 1 (置信度非常高) :

- 1 不是，高尔夫球的目标不是获得比其他人更高的得分。相反，目标是以最少的杆数打完一组洞。总
- 2 问题：高尔夫球的一部分是试图获得比其他人更高的得分。是或否？
- 3
- 4 知识：高尔夫球是一项精密的球杆运动，竞争选手（或高尔夫球手）使用许多类型的球杆，用最少的
- 5
- 6 解释和答案：

答案 2 (置信度要低得多) :

- 1 是的，高尔夫球的一部分是试图获得比其他人更高的得分。每个选手都试图以最少的杆数打完一组

Active-Prompt

思维链 (CoT) 方法依赖于—组固定的人工注释范例。问题在于，这些范例可能不是不同任务的最有效示例。

最近提出了一种新的提示方法，称为 Active-Prompt，以适应 LLMs 到不同的任务特定示例提示（用人类设计的 CoT 推理进行注释）。

下面是该方法的说明。第一步是使用或不使用少量 CoT 示例查询 LLM。对一组训练问题生成 k 个可能的答案。基于 k 个答案计算不确定度量（使用不一致性）。选择最不确定的问题由人类进行注释。然后使用新的注释范例来推断每个问题。

