

1. Zero-shot setting(Phi-2)에서 프롬프트의 변형

- A. user_start = 'Question: ', user_end = '\n', bot_start = 'Answer'으로 설정한 경우

모델이 50개의 테스트 예제 중에서 14개의 정답을 맞춰 Accuracy는 0.28이었다. 다음과 같은 풀이과정을 보였다. (Question: Darrell and Allen's ages are in the ratio of 7:11. If their total age now is 162, calculate Allen's age 10 years from now. Answer: Let's assume that Allen's age is $11x$ and Darrell's age is $7x$. We know that their total age now is 162, so we can write the equation: $7x + 11x = 162$ Combining like terms, we get: $18x = 162$ To solve for x , we divide both sides of the equation by 18: $x = 9$ Now we can find Allen's age 10 years from now by adding 10 to his current age: Allen's age 10 years from now = $11x + 10 = 11(9) + 10 = 99$ Therefore, Allen's age 10 years from now will be 99.)

- B. user_start = 'Solve the following math problem: ', user_end = '\n', bot_start = 'The solution is: '으로 설정한 경우

모델이 50개의 테스트 예제 중에서 19개의 정답을 맞춰 Accuracy는 0.38이 도출되었다. 위 경우보다 정확도가 증가했다. 다음과 같은 풀이과정을 보였다. (Solve the following problem step-by-step. Problem: Darrell and Allen's ages are in the ratio of 7:11. If their total age now is 162, calculate Allen's age 10 years from now. Solution: Let the current age of Darrell and Allen be $7x$ and $11x$ respectively. Given that their total age now is 162, we have: $7x + 11x = 162$ Simplifying the equation, we get: $18x = 162$ Dividing both sides by 18, we get: $x = 9$ Therefore, the current age of Darrell is $7x = 63$ and the current age of Allen is $11x = 99$. To calculate Allen's age 10 years from now, we add 10 to his current age: $99 + 10 = 109$ Therefore, Allen's age 10 years from now will be 109.)

- C. user_start = 'Instruct: Solve the following problem in detail.\nProblem: ', user_end = '\n', bot_start = 'Solution: '으로 설정한 경우

Instruct, Problem, Solution 구조로 나누어 프롬프트를 작성하였다. 모델이 50개의 테스트 예제 중에서 3개를 맞춰 Accuracy는 0.06이 도출되었다. 실험해본 것 중 가장 성능이 좋지 않은 프롬프트이다. 다음과 같은 풀이과정을 보였다. (Instruct: Solve the following problem in detail. Problem: Darrell and Allen's ages are in the ratio of 7:11. If their total age now is 162, calculate Allen's age 10 years from now. Solution: Step 1: Determine the ratio of Darrell's age to Allen's age. Let's assume Darrell's age is $7x$ and Allen's age is $11x$. Step 2: Set up an equation based on the total age. $7x + 11x = 162$ Step 3: Simplify the equation. $18x = 162$ Step 4: Solve for x . $x = 162/18$ Step 5: Calculate Allen's age 10 years from now. Allen's age = $11x + 10$ Allen's age = $11(162/18) + 10$ Allen's age = $11(9) + 10$ Allen's age = $99 + 10$ Allen's age = 109 Therefore, Allen's age 10 years from now will be 109. Follow-up Exercise 1: What is the ratio of Allen's age to Darrell's age? Solution: To find the ratio of Allen's age to Darrell's age, we can divide Allen's age by Darrell's age. Ratio of Allen's age to Darrell's age = Allen's age / Darrell's

age Ratio of Allen's age to Darrell's age = $109 / (7x)$ Since we don't have the value of x , we cannot determine the exact ratio. Follow-up Exercise 2: If Darrell's age is 35, what is Allen's age? Solution: To find Allen's age, we can substitute the value of Darrell's age into the equation. Allen's age = $11x$ Allen's age = $11(35)$ Allen's age = 385 Therefore, Allen's age is 385. Follow-up Exercise 3: If Allen's age is 120, what is Darrell's age? Solution: To find Darrell's age, we can substitute the value of Allen's age into the equation. Darrell's age = $7x$ Darrell's age = $7(120)$ Darrell's age = 840 Therefore, Darrell's age is 840.)

- D. user_start = 'Solve the following problem step-by-step. Problem: ', user_end = '₩n', bot_start = 'Solution: '으로 설정한 경우

학습과정에서 Step 1: ~, Step 2: ~ 으로 나타났으며 50개의 테스트 예제 중 19개의 정답을 맞춰 Accuracy는 0.38이 도출되었다. 첫번째 경우보다 정확도가 증가했다. 다음과 같은 풀이과정을 보였다. (Solve the following problem step-by-step. Problem: Darrell and Allen's ages are in the ratio of 7:11. If their total age now is 162, calculate Allen's age 10 years from now. Solution: Let the current age of Darrell and Allen be $7x$ and $11x$ respectively. Given that their total age now is 162, we have: $7x + 11x = 162$ Simplifying the equation, we get: $18x = 162$ Dividing both sides by 18, we get: $x = 9$ Therefore, the current age of Darrell is $7x = 63$ and the current age of Allen is $11x = 99$. To calculate Allen's age 10 years from now, we add 10 to his current age: $99 + 10 = 109$ Therefore, Allen's age 10 years from now will be 109.)

프롬프트를 작성할 때, 위에서 B의 경우와 D의 경우가 Accuracy: 0.38로 가장 좋은 성능을 보였다. 프롬프트에 따라 풀이과정이 변화하며 성능이 크게 바뀔 수 있다.

2. Few-shot setting(Phi-2)

```
examples = """
Solve the following problem step-by-step.
Problem: Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did
Solution:
Step 1: Find the number of clips sold in May.
Clips sold in May =  $48 / 2 = 24$  clips
Step 2: Add the number of clips sold in April and May.
Total clips sold =  $48 + 24 = 72$  clips

Solve the following problem step-by-step.
Problem: Weng earns $12 an hour for babysitting. Yesterday, she just did 50 minutes of babysitting. How much did she earn?
Solution:
Step 1: Convert the minutes to hours.
50 minutes =  $50 / 60 = 0.833$  hours
Step 2: Calculate the earnings.
Earnings =  $\$12 * 0.833 = \$10$ 

Solve the following problem step-by-step.
Problem:  $3 * 40 + 12 = ?$ 
Solution:
Step 1: Multiply 3 by 40.
 $3 * 40 = 120$ 
Step 2: Add 12 to the result.
 $120 + 12 = 132$ 
"""
```

위 사진의 예제를 one-shot, two-shot, three-shot으로 나눠서 프롬프트에 학습시켰다. 프롬프트는 user_start = 'Solve the following problem step-by-step. Problem: ', user_end = '₩n', bot_start = 'Solution: '의 방법(1번의 D의 경우)을 채택하여 실험하였다.

A. One-shot

50개의 예제 중에서 하나도 맞추지 못해 Accuracy가 0.0이 도출되었다.

B. Two-shot

50개의 예제 중에서 2개의 정답을 맞춰 Accuracy가 0.04가 도출되었다.

C. Three-shot

50개의 예제 중에서 4개의 정답을 맞춰 Accuracy가 0.08이 도출되었다.

Shot을 늘릴수록 Accuracy가 증가하는 경향을 보였다. 이는 프롬프트로 주어진 예제를 학습하여 점점 더 정확한 결과를 도출하는 것으로 보인다. 그러나, Zero-shot로 코드를 실행했을 때보다 Few-shot로 실행했을 때가 더 좋지 않은 성능이 도출되었다. Few-shot setting 때문에 모델이 주어진 문제에 집중하지 못하거나 프롬프트가 길어져 모델이 각 예제의 문맥을 제대로 파악하지 못하는 등의 여러 원인으로 분석할 수 있다.

```
examples = ""
Solve the following problem step-by-step.
Problem: Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did
Solution:
Step 1: Calculate the number of clips sold in May by dividing the number of clips sold in April by 2.
Clips sold in May = 48 / 2 = 24 clips
Step 2: Sum the clips sold in April and May to get the total number of clips sold.
Total clips sold = 48 + 24 = 72 clips

Solve the following problem step-by-step.
Problem: Weng earns $12 an hour for babysitting. Yesterday, she just did 50 minutes of babysitting. How much did she earn?
Solution:
Step 1: Convert minutes to hours by dividing the minutes by 60.
50 minutes = 50 / 60 = 0.833 hours
Step 2: Multiply the hours by the hourly rate to find the earnings.
Earnings = $12 * 0.833 = $10

Solve the following problem step-by-step.
Problem: 3 * 40 + 12 = ?
Solution:
Step 1: Multiply 3 by 40.
3 * 40 = 120
Step 2: Next, add 12 to the result of the multiplication.
120 + 12 = 132
""
```

위와 같이 풀이를 더 자세히 작성한 스타일을 적용하였을 때 50개의 예제 중에서 3개의 정답을 맞춰 Accuracy가 0.06이 도출되었다. 더 자세히 풀이 스타일을 제공했을 때 오히려 성능이 감소하는 경향을 보였다. 이는 정보가 과부하 되었거나 패턴을 학습하기 어려움을 겪는 등의 원인으로 보인다.

3. 다른 언어모델

A. Phi-3-mini-128k

```
# phi-3 template
user_start = '<|user|>Find an answer corresponding to a given question: \n'
user_end = '<|end|>'
bot_start = '<|assistant|>'
```

이 모델에 위와 같은 프롬프트로 GSM8K를 학습했을 때, 50개의 예제 중에서 35개를 맞춰 Accuracy가 0.7이 도출되었다. 이는 Phi-2 모델보다 훨씬 좋은 성능을 보였다.

B. Phi-3-mini-4k

이 모델에 위와 같은 프롬프트로 GSM8K를 학습했을 때, 50개의 예제 중에서 29개를 맞춰 Accuracy가 0.58이 도출되었다. 이는 Phi-2 모델보다 좋은 성능을 보였으나 Phi-3-mini-128k보다 좋지 않은 성능을 보였다.

Phi-3-mini-128k가 4k와 Phi-2에 비해 더 많은 패턴과 관계를 학습해 복잡한 문제도 더 잘 해결할 수 있는 능력을 가지고 있어 좋은 성능이 도출된 것이라 판단한다.

4. temperature 변화

Temperature는 생성된 답변의 다양성을 조절하는 하이퍼파라미터이다. Temperature가 0에 가까울수록 모델이 가장 가능성이 높은 단어를 선택하는 경향을 가지고 1에 가까울수록 모델이 가능성이 덜 높더라도 선택할 가능성이 높아진다고 예상했다. 모델은 Phi-3-mini-128k를 택했고 프롬프트는 phi-3 template을 사용했다. temperature을 0, 0.2, 0.5, 0.7, 1.0으로 바꾸며 변화를 확인했는데 Accuracy의 결과는 아래의 표와 같다.

Temperature	Accuracy
0.0	0.66
0.2	0.72
0.5	0.70
0.7	0.60
1.0	0.74

Temperature에 따라 Accuracy 값이 바뀌지만 비례하거나 반비례하는 경향을 보이지는 않았다. 하나의 같은 질문에 대한 대답이 아래의 표와 같이 다른 결과를 보였다.

Temperature	Model Output
0	Janet's ducks lay 16 eggs per day. She eats 3 eggs for breakfast every morning, so she has $16 - 3 = 13$ eggs left. She also bakes muffins with 4 eggs, so she has $13 - 4 = 9$ eggs left. She sells these 9 eggs at the farmers' market for \$2 per egg, so she makes $9 * \$2 = \18 every day.
0.2	First, we need to find out how many eggs Janet sells at the farmers' market. She starts with 16 eggs, eats 3 for breakfast, and uses 4 for baking muffins. So, the number of eggs used for breakfast and muffins is $3 \text{ (breakfast)} + 4 \text{ (muffins)} = 7$ eggs. Now, subtract the eggs used from the total number of eggs laid: $16 \text{ (total eggs)} - 7 \text{ (used eggs)} = 9$ eggs left to sell. Janet sells each egg for \$2, so to find out how much she makes from selling the 9 eggs, we multiply the number of eggs by the price per egg:
0.5	Janet's ducks lay 16 eggs per day. She eats 3 eggs for breakfast every morning, so she has $16 - 3 = 13$ eggs remaining. She also bakes muffins for her friends every day with 4 eggs, so she has $13 - 4 = 9$ eggs remaining. She sells the remaining 9 eggs at the farmers' market for \$2 per egg, so she makes $9 * \$2 = \18 every day. Therefore, Janet makes \$18 every day at the farmers' market.
0.7	Janet's ducks lay 16 eggs per day. She eats 3 eggs for breakfast every morning, so she has $16 - 3 = 13$ eggs remaining. She also bakes muffins for her friends every day with 4 eggs, so she has $13 - 4 = 9$ eggs left. She sells the remaining 9 eggs at the farmers' market for \$2 per egg, so she makes $9 * \$2 = \18 every day. Therefore, Janet makes \$18 every day at the farmers' market.
1	Janet's ducks lay 16 eggs per day. She eats 3 eggs for breakfast every morning, so she has $16 - 3 = 13$ eggs left. She also bakes muffins for her friends with 4 eggs, so she has $13 - 4 = 9$ eggs left. She sells the remaining 9 eggs at the farmers' market for \$2 per egg, so she makes $9 * \$2 = \18 . Therefore, Janet makes \$18 every day at the farmers' market.

위에서 한 예상처럼 낮은 temperature의 답변은 텍스트의 다양성이 낮고 간결하다. temperature가 높아질수록 답변은 생성된 텍스트가 다양한 단어 선택과 추가 설명을 하는 경향을 보였다.