

The background features a complex network diagram with numerous nodes of varying sizes (black, blue, and grey) connected by thin grey lines. Some nodes are highlighted with larger concentric circles. The overall aesthetic is technical and digital.

PROMPT ENGINEERING

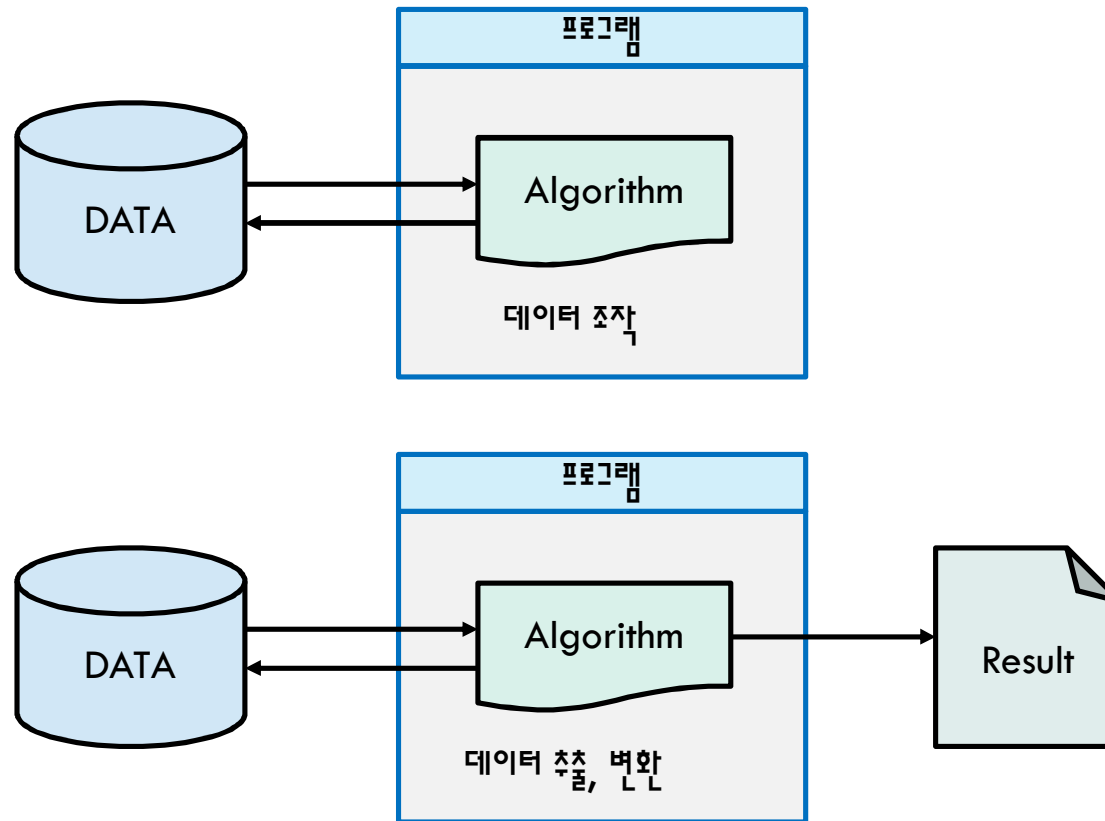
PROMPT ENGINEERING?

어떻게 AI 모델에게 새롭고 효율적인 방식으로
정보를 전달할 것인가

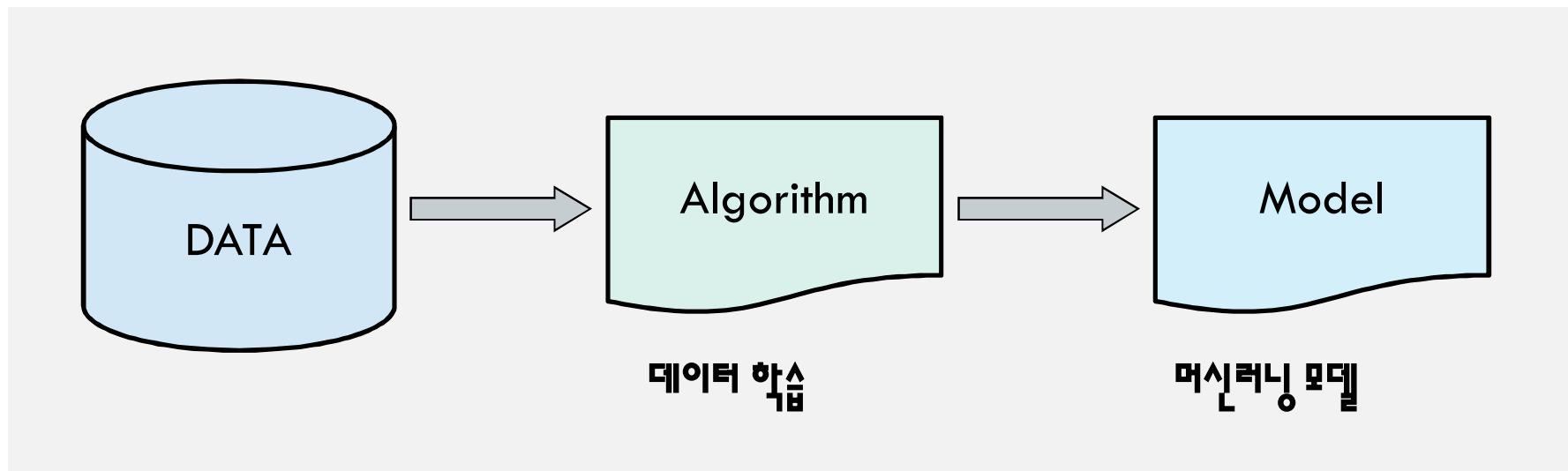


PROMPT
ENGINEERING

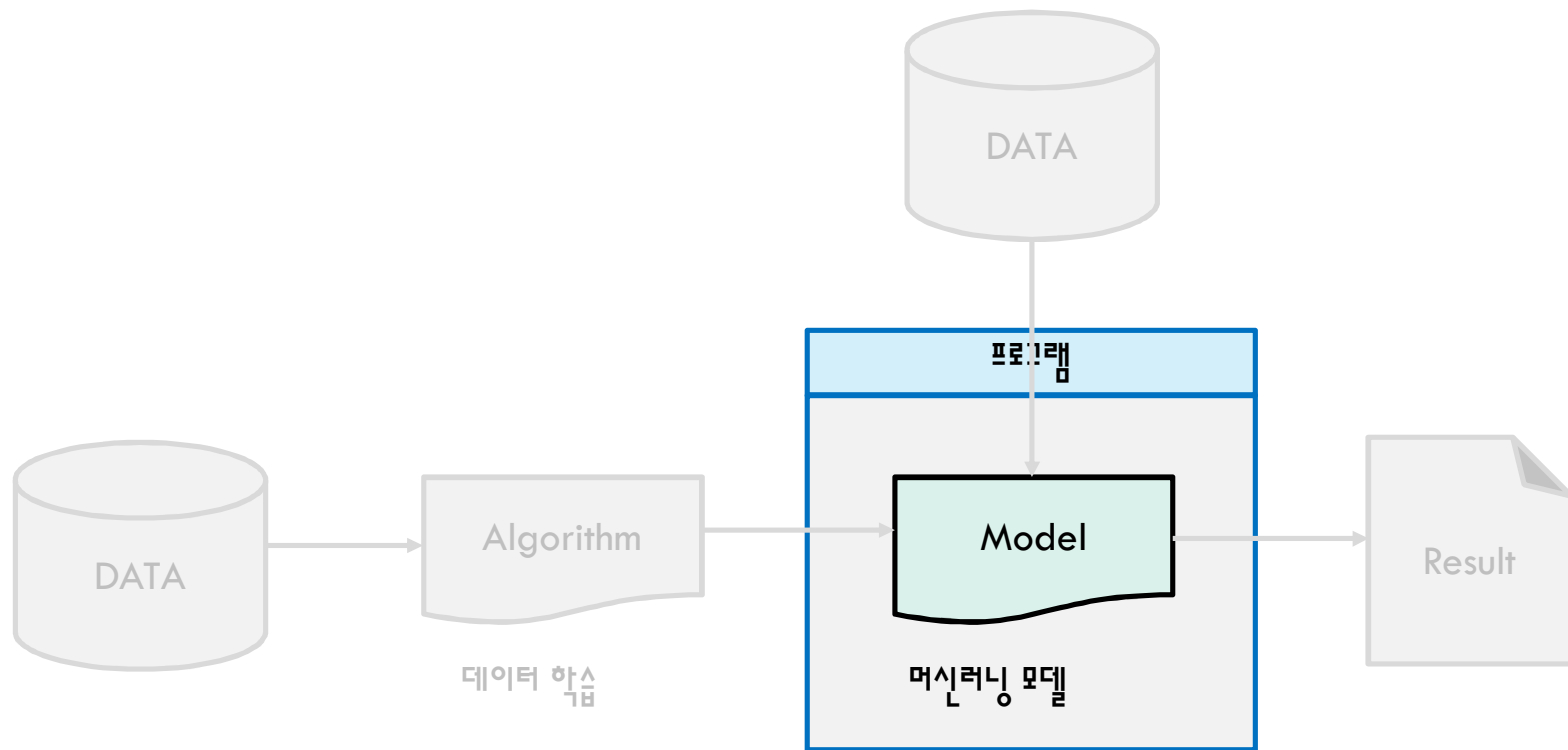
기존 모델 방식



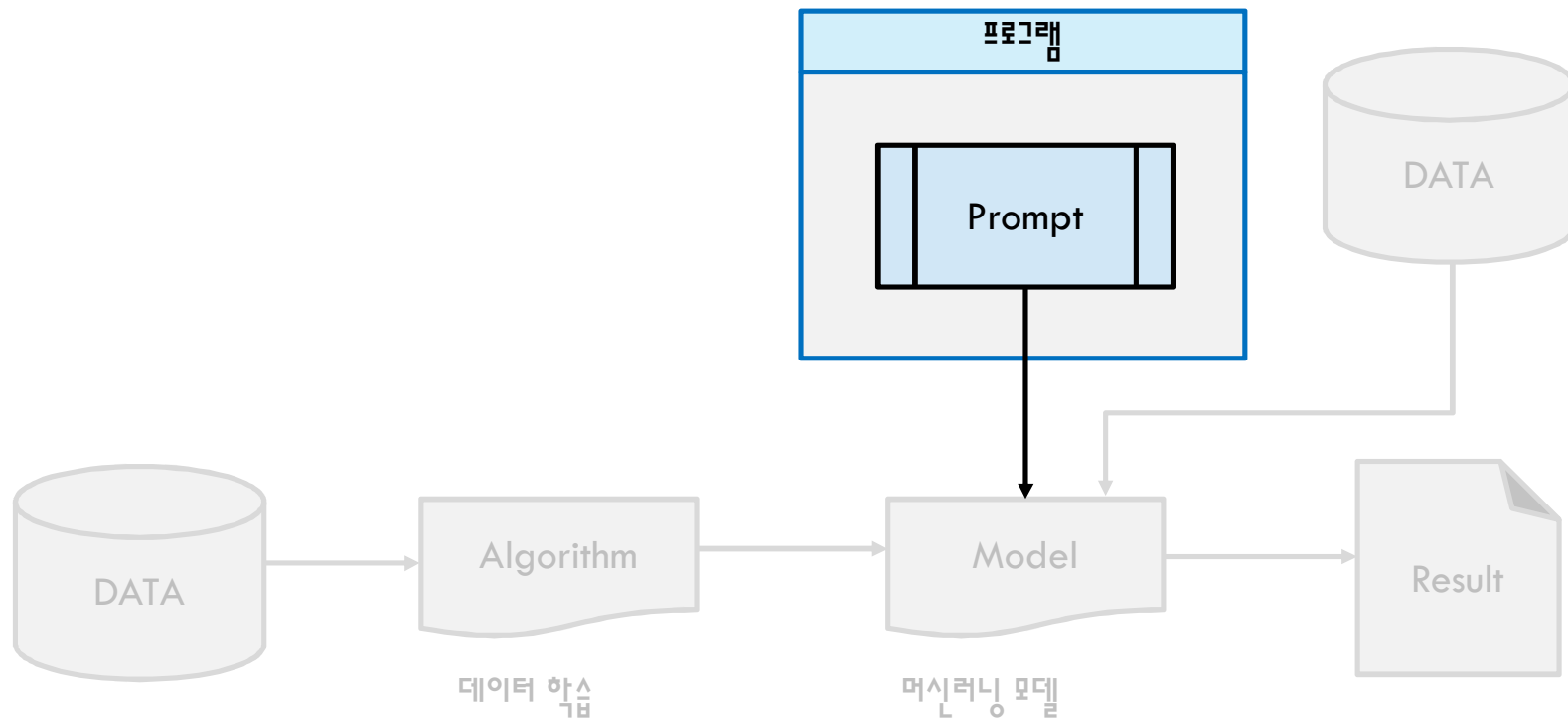
FOUNDATION 모델 방식1



FOUNDATION 모델 방식2

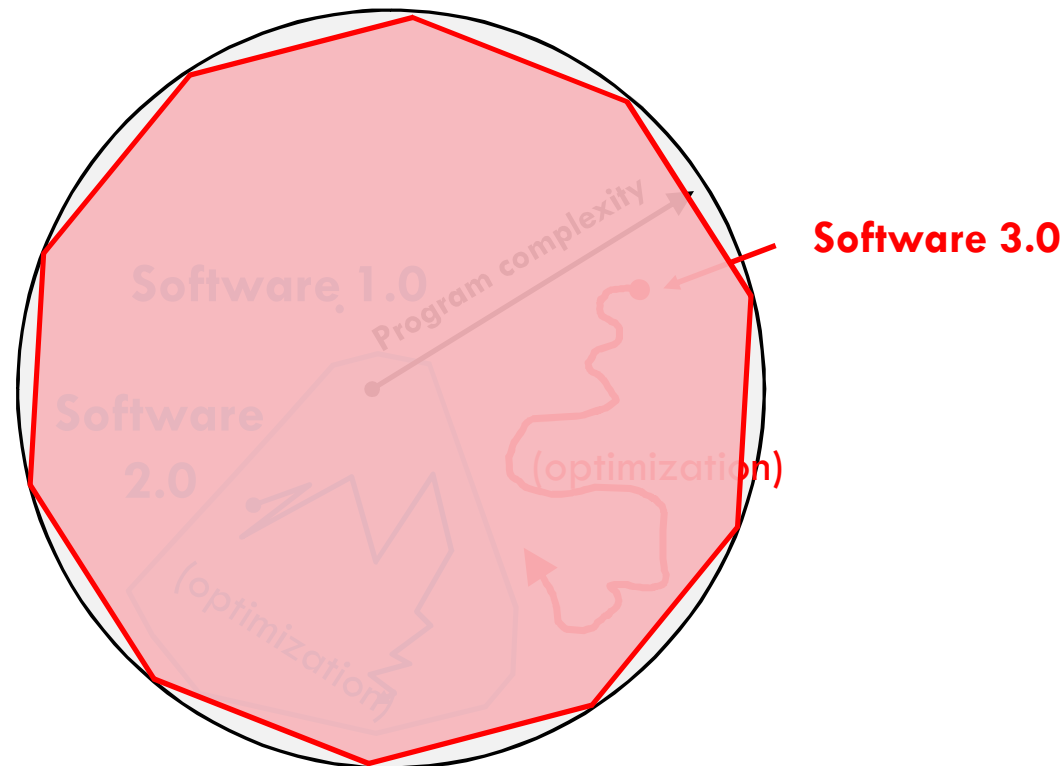


FOUNDATION 모델 방식3



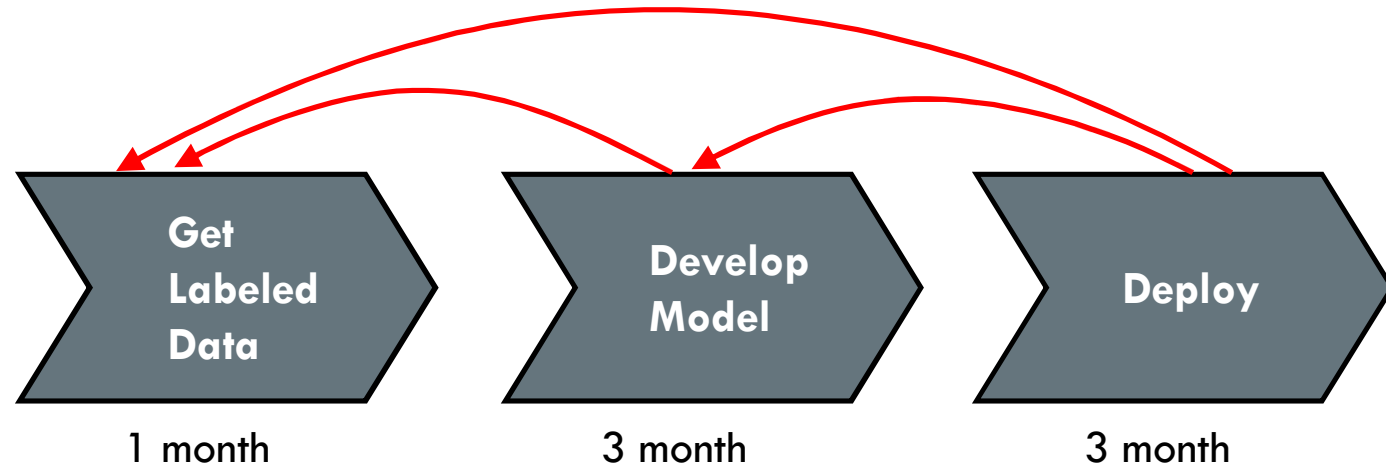
FOUNDATION 모델 방식3

Prompt Engineering 방식 = > 새로운 상호 작용(데이터 입력 및 학습 방법)

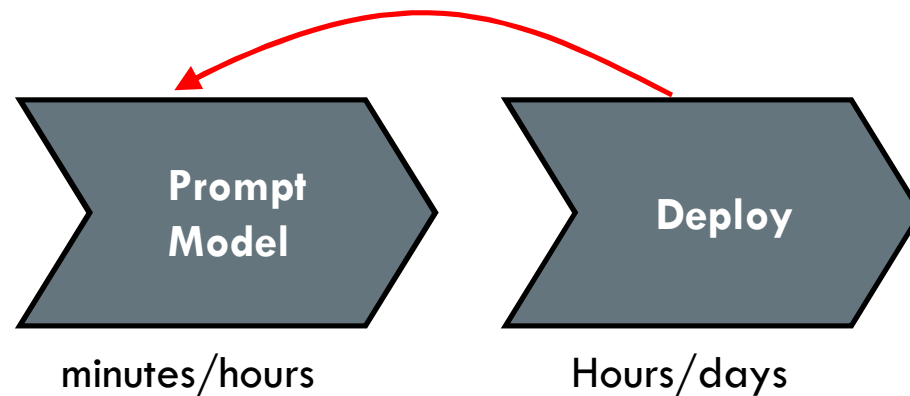


PROMPT ENGINEERING

Traditional ML



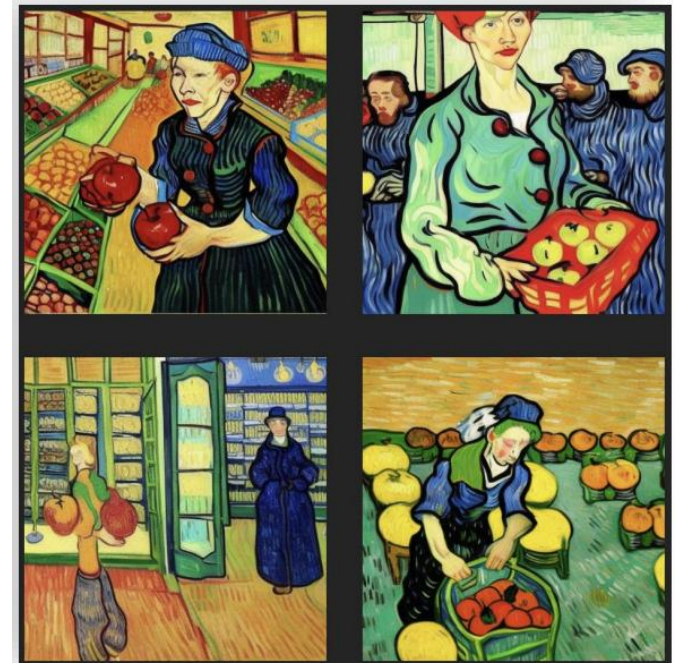
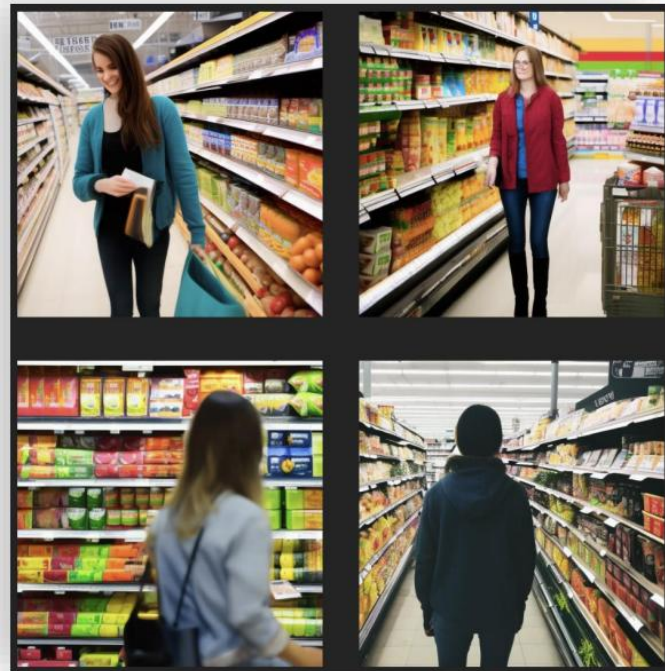
Prompt-based ML



PROMPT ENGINEERING DESIGN

- 자연어로 컴퓨터와 상호작용하는 방법임
- 실험적인 방법론에 기반하여, 원하는 결과를 정확히 의도한대로 일관성 있게 만드는 것

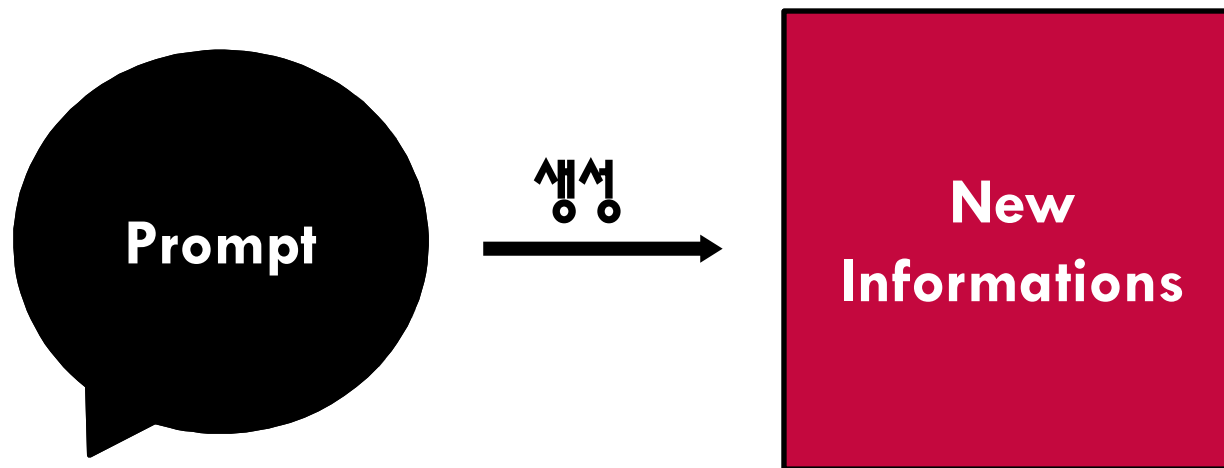
PROMPT ENGINEERING



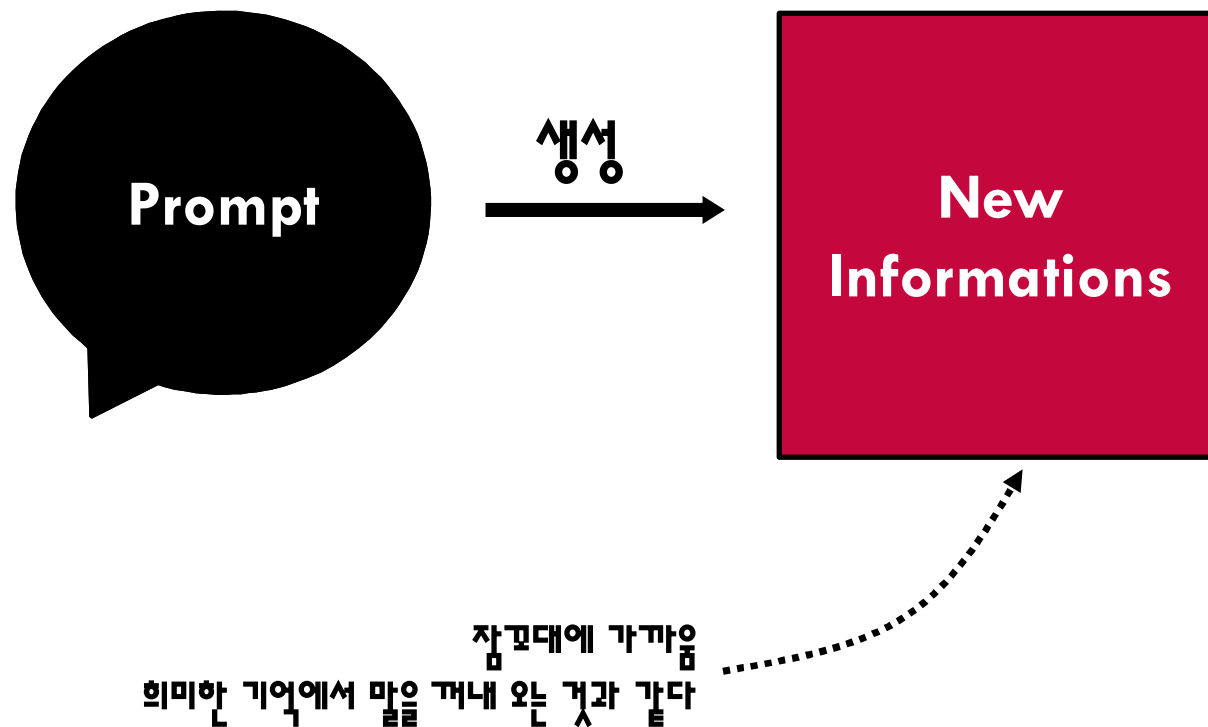
PROMPT ENGINEERING DESIGN

- 답변을 위해 필요한 적절한 컨텍스트 제공(incontext learning)
- 원하는 결과를 추출하기 위한 프롬프트 작성
- 원하는 포맷의 출력을 위한 프롬프트 작성

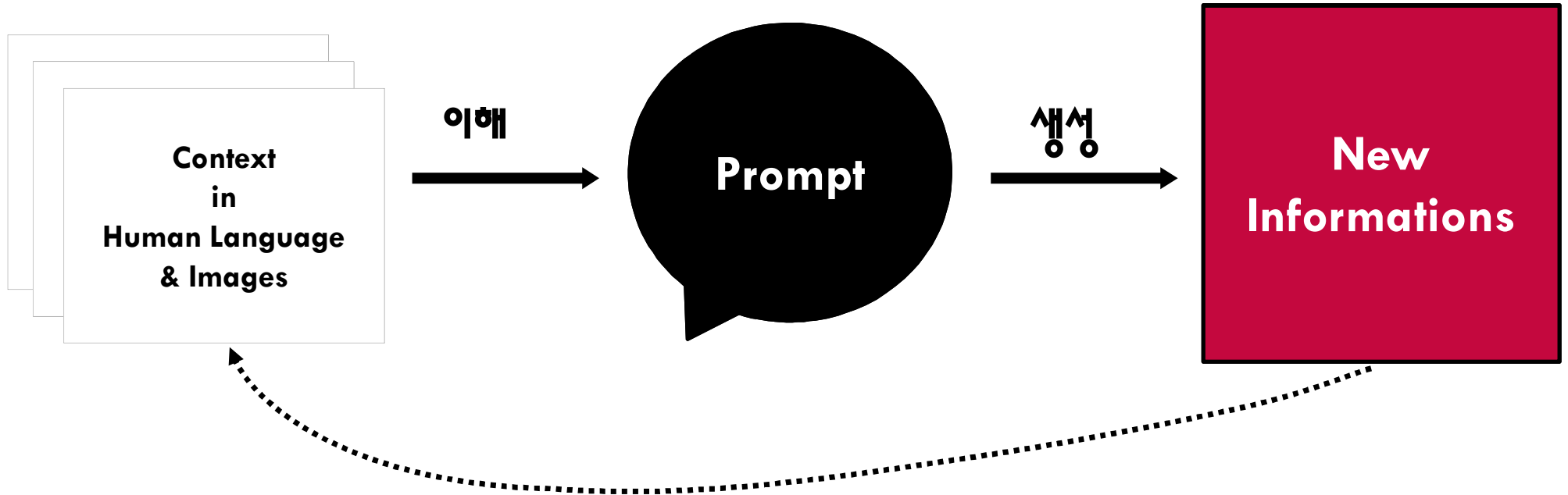
PROMPT ENGINEERING 개념



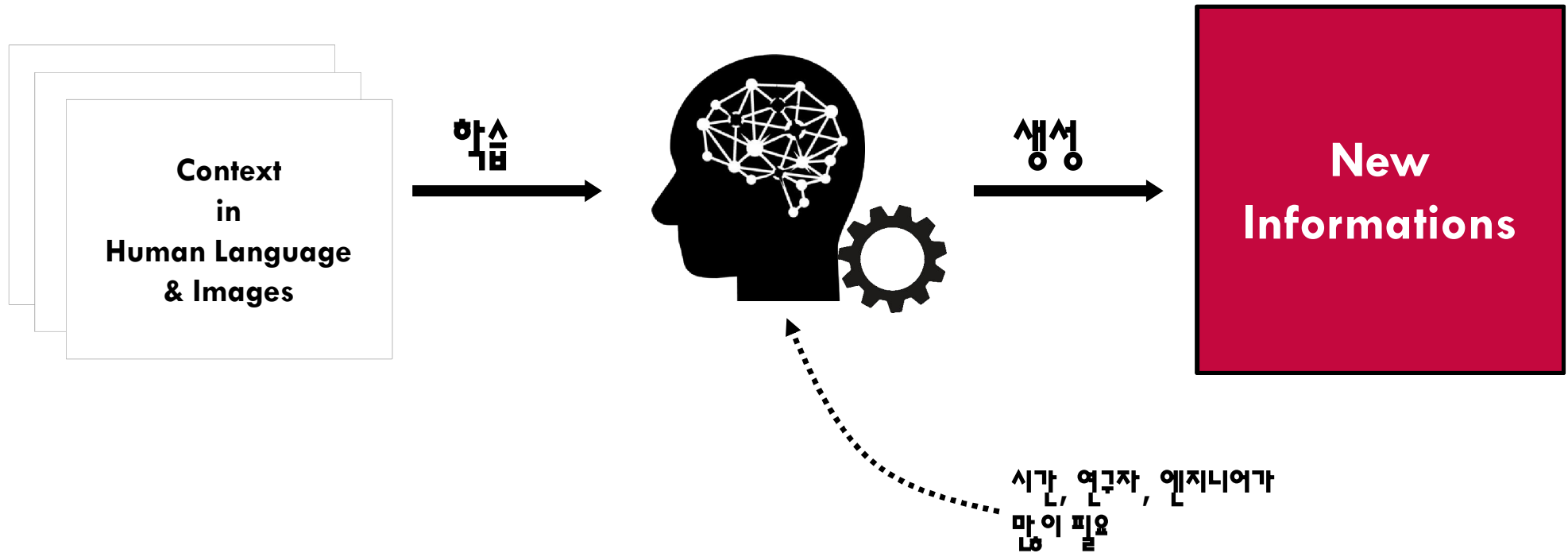
PROMPT ENGINEERING 개념



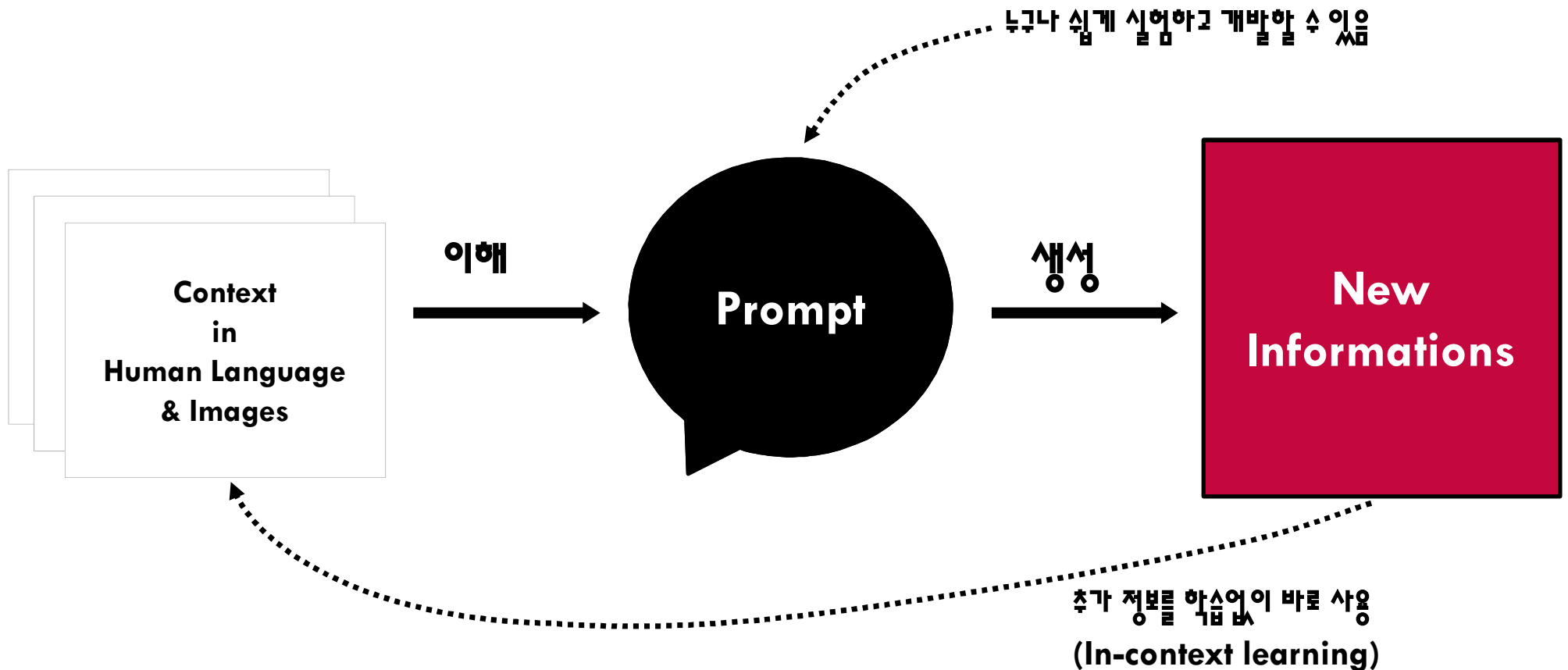
PROMPT ENGINEERING 개념



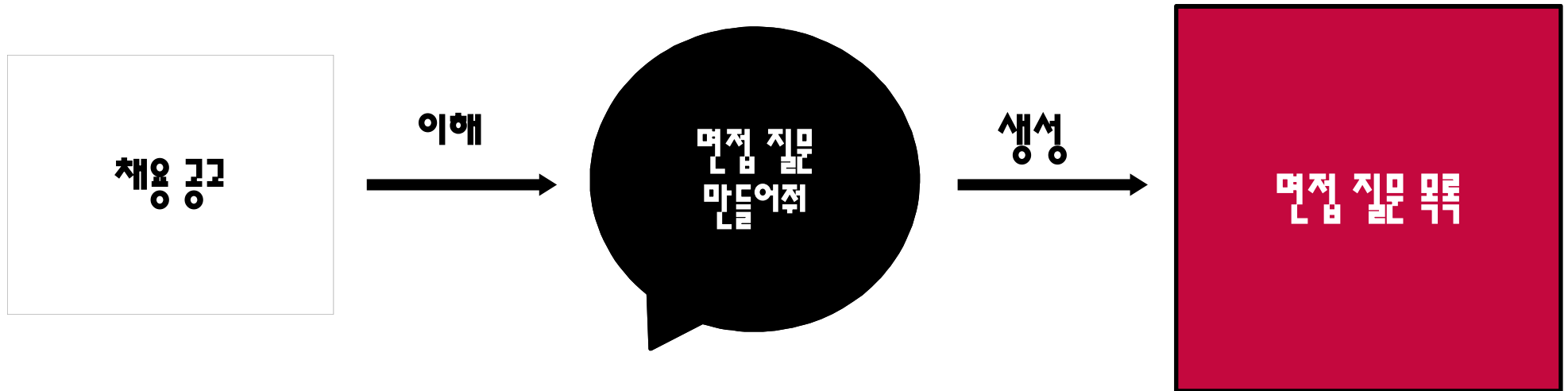
PROMPT ENGINEERING 개념



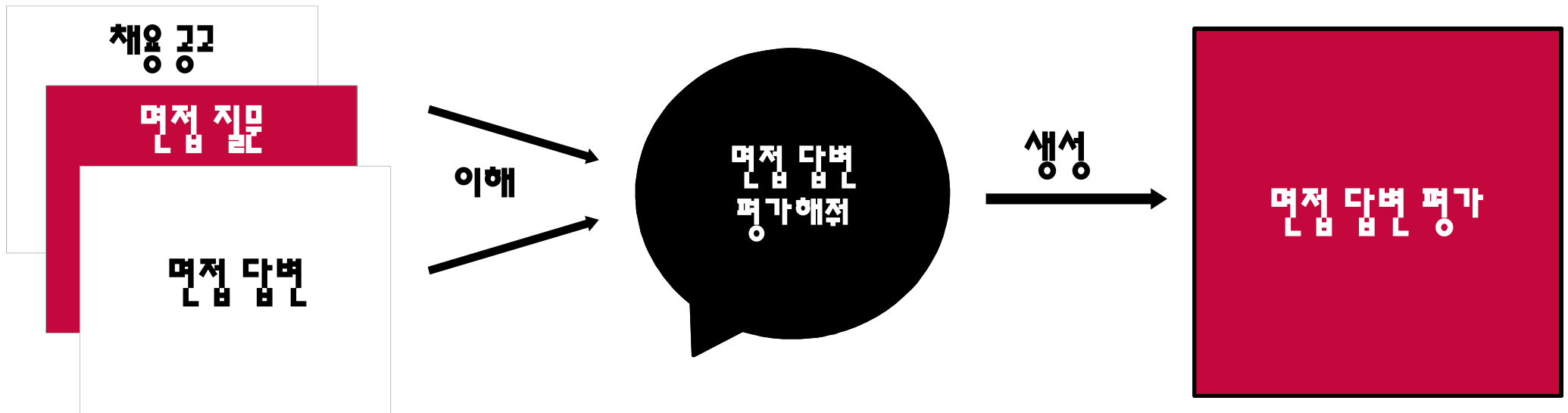
PROMPT ENGINEERING 개념



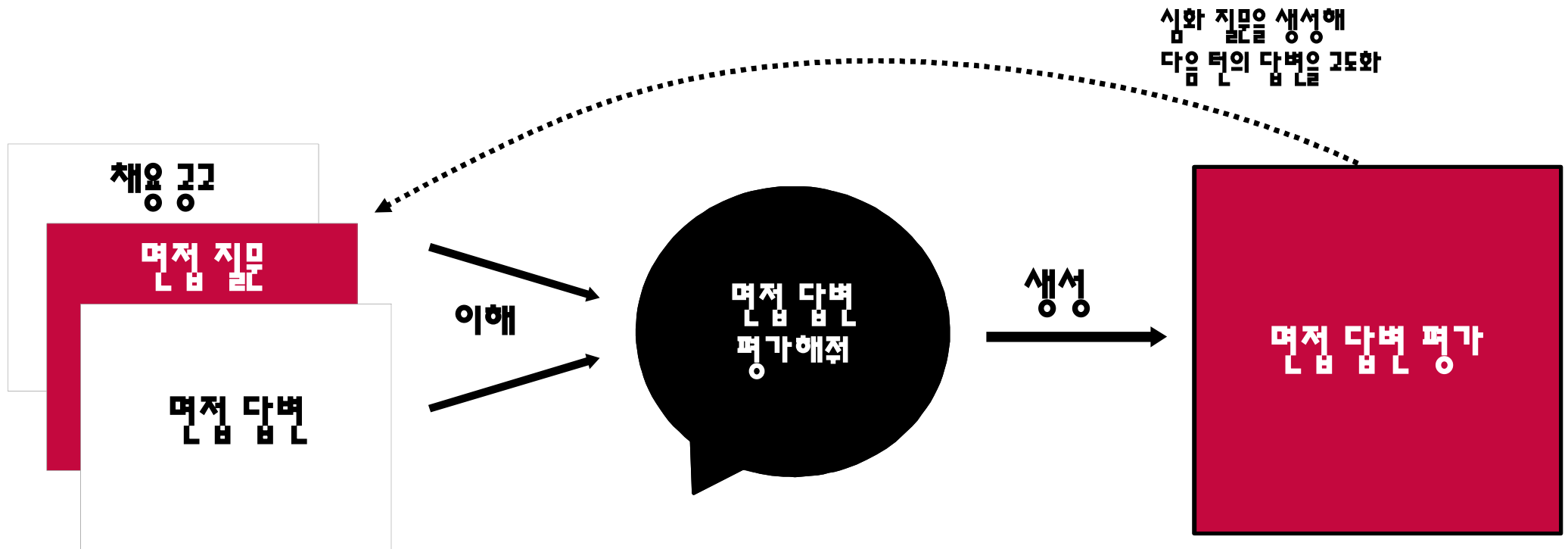
PROMPT ENGINEERING 개념



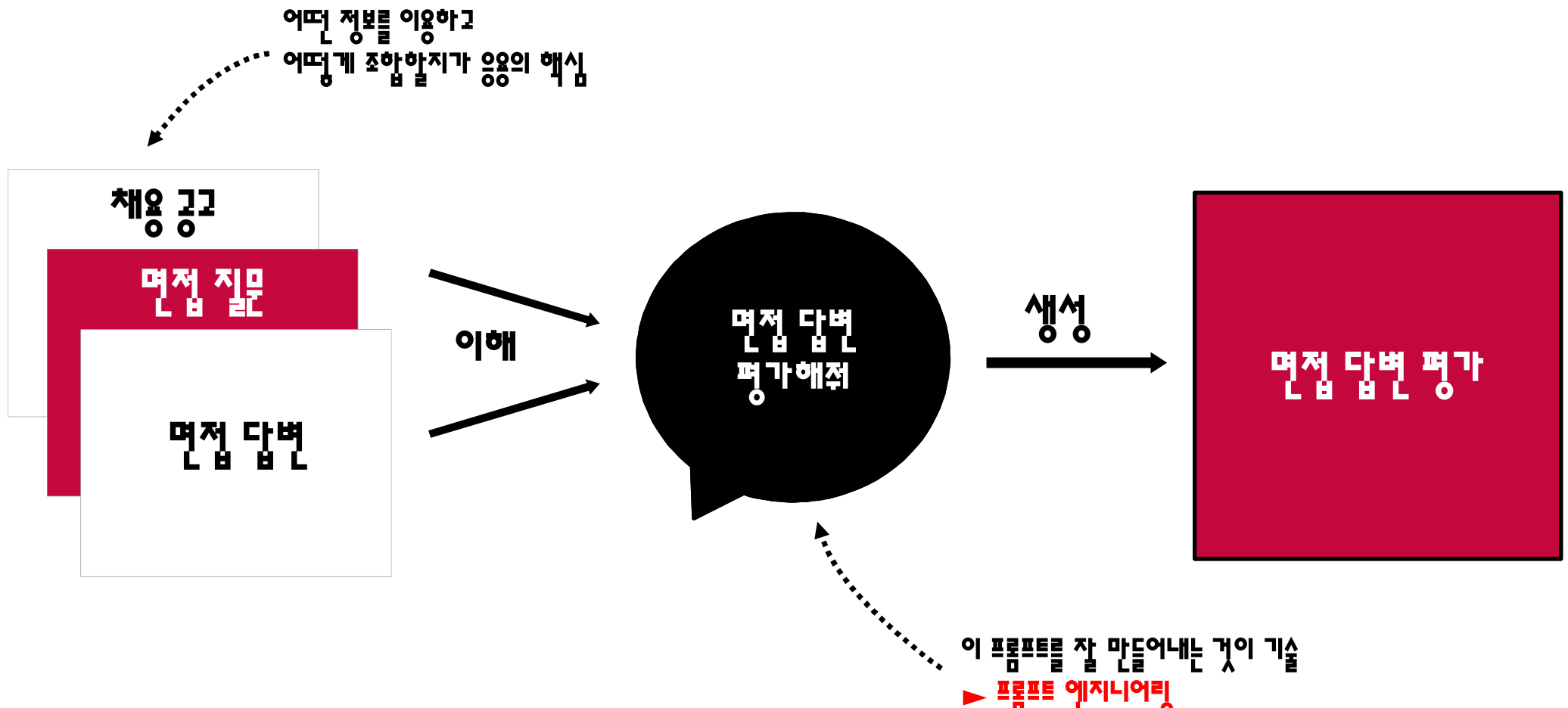
PROMPT ENGINEERING 개념



PROMPT ENGINEERING 개념



PROMPT ENGINEERING 개념



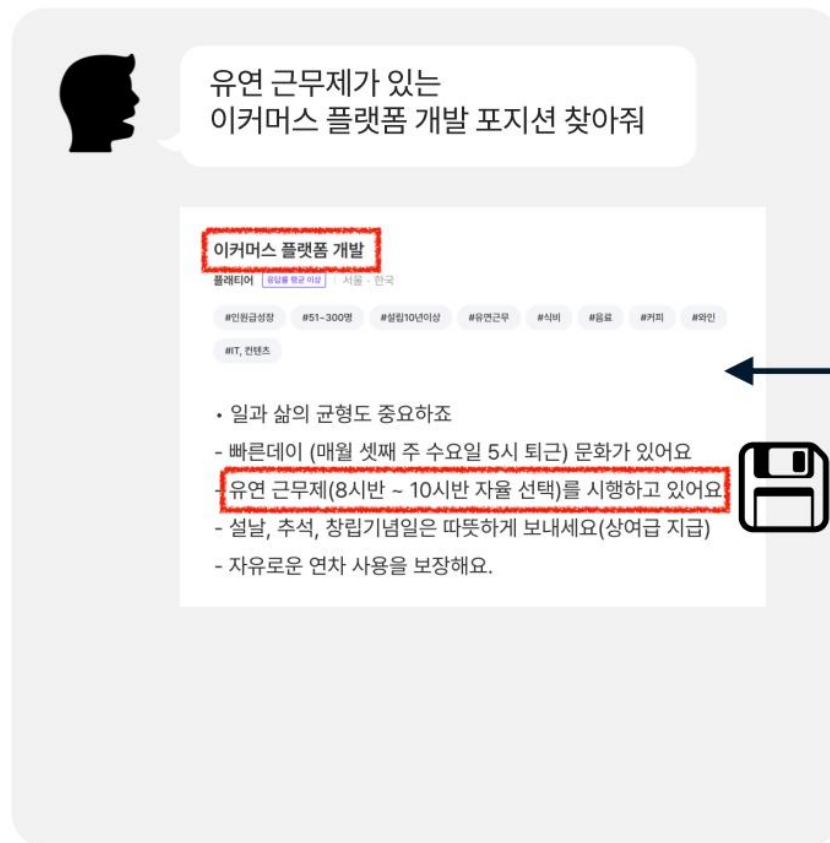
PROMPT ENGINEERING 개념



유연 근무제가 있는
이커머스 플랫폼 개발 포지션 찾아줘

← User Prompt

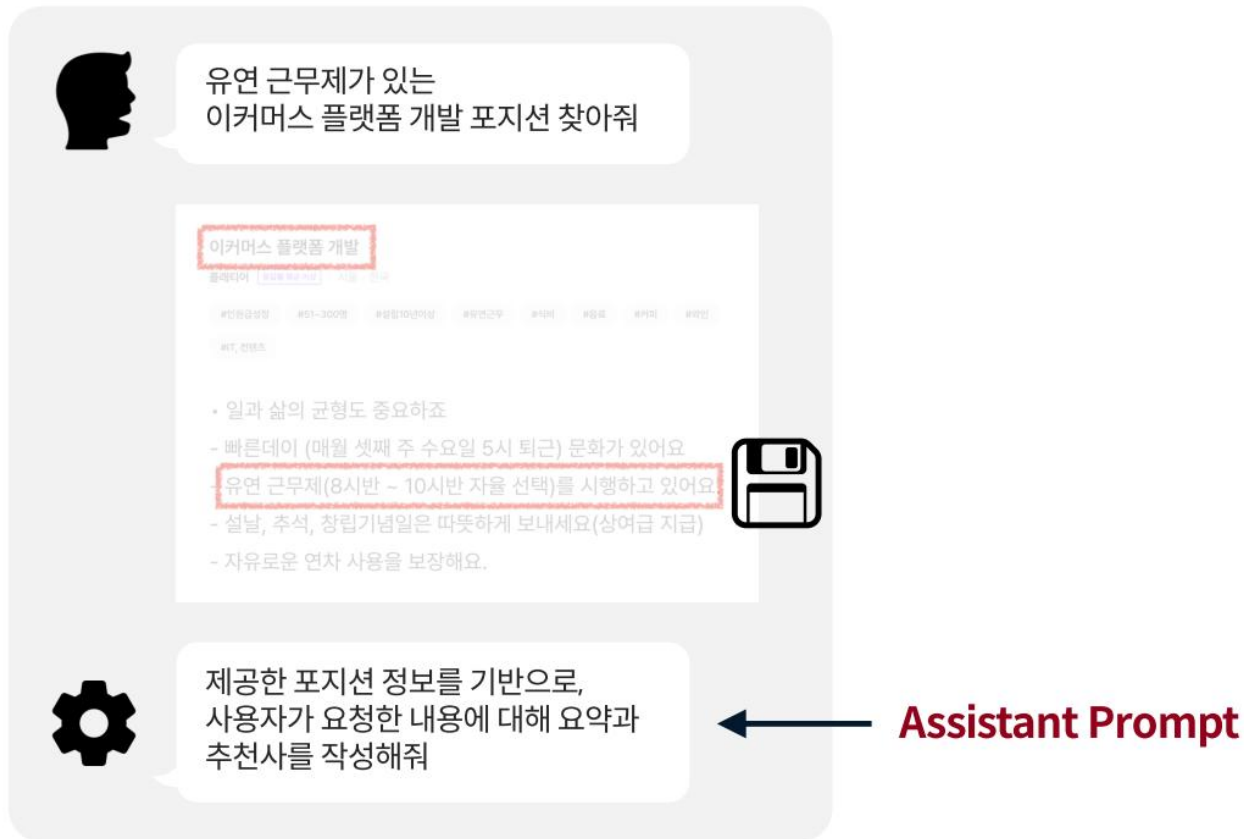
PROMPT ENGINEERING 개념



Context from Vector Search



PROMPT ENGINEERING 개념



PROMPT ENGINEERING 개념



유연 근무제가 있는
이커머스 플랫폼 개발 포지션 찾아줘

이커머스 플랫폼 개발

플래티어 | 응답률 100% | 서울 · 한국

#인원급상장 #51-300명 #설립10년이상 #유연근무 #식비 #음료 #커피 #와인
#IT, 컨텐츠

- 일과 삶의 균형도 중요하죠
- 빠른데이 (매월 셋째 주 수요일 5시 퇴근) 문화가 있어요
- 유연 근무제(8시반 ~ 10시반 자율 선택)를 시행하고 있어요
- 설날, 추석, 창립기념일은 따뜻하게 보내세요(상여급 지급)
- 자유로운 연차 사용을 보장해요.



제공한 포지션 정보를 기반으로,
사용자가 요청한 내용에 대해 요약과
추천사를 작성해줘



유연 근무제가 있는 이커머스 플랫폼 회사로
플래티어의 플랫폼 개발자 포지션을 찾았습니다.

플래티어는 8시반~10시반 자율 선택으로
유연 근무제를 시행하고 있으며, 자유로운 연차
사용 보장으로 근무를 유연하게 할 수 있는
환경을 갖춘 회사여서 적극 추천합니다.

Personalized Answer Generation

PROMPT ENGINEERING DESIGN

- 구성요소
 - Role
 - Audience
 - Knowledge
 - Task
 - Policy
 - Format
 - Examples

PROMPT ENGINEERING DESIGN

- Role

- 답변자 페르소나 설정
- 답변에 대한 배경 지식을 가지고 답함

ex) 당신은 신문 기자입니다. => 주요 뉴스 정보를 정리하여 작성하세요
당신은 수학 선생님입니다 => 미분에 대해 설명해주세요.

PROMPT ENGINEERING DESIGN

- Audience

- 답변의 대상이 되는 그룹과 개인의 페르소나

ex) 기타 초보자를 위한 연습 방법을 설명해주세요

숙련된 프로그래머를 위한 추천 알고리즘을 제공해주세요

PROMPT ENGINEERING DESIGN

- Knowledge
 - 답변을 위한(참고 하기 위한) 정보
 - 사용자 질문에 해당하는 정보를 가져와서(DB, 검색엔진, 유명한 정보 출처 지정) 삽입

ex) 위키피디아의 내용을 참고하여 이순신 장군에 대한 설명을 해주세요.

PROMPT ENGINEERING DESIGN

- Task

- 수행해야 하는 작업 또는 목표

ex) **광고 마케팅 전략을 개선안을 제시해주세요.**

1000단어 내외로 자기소개소를 작성해주세요.

PROMPT ENGINEERING DESIGN

- Policy

- 응답을 위해 따라야 하는 특정 규칙이나 정책

ex) **사실 기반의 정보만**을 이용하여 뉴스 기사를 작성해주세요.

긍정적인 성격의 측면만을 강조하여 자기소개서를 작성해주세요.

PROMPT ENGINEERING DESIGN

- Format
 - 응답을 위해 따라야 하는 특정 형식 또는 구조

ex) **JSON** 형식으로 결과물을 출력해주세요

사전 형식으로 사랑을 설명해주세요

파이썬 스타일로 코드를 작성해주세요

PROMPT ENGINEERING DESIGN

- Examples

- 예시를 정보를 제공하여 원하는 출력을 더 잘 이해하도록 유도

ex) 다음의 예시를 참고하여 응답하세요

<User>오늘 날씨가 좋네요</User>

<Assistant>하늘이 참 푸르네요</Assistant>

다음과 같은 형식으로 출력해주세요.

```
{"id":"gd242ee7" , "name":"smith",...}
```


PROMPT ENGINEERING DESIGN

- 예시

- Role : 영어 선생님
- Audience : 초등학생
- knowledge(imfomation) : 나폴레옹
- Task : 나폴레옹에 대한 설명
- policy : 단순하고 이해하기 쉬운 언어 사용
- format: 대화 형식으로 작성
- Example :
 - User:나폴레옹은 누구야?
 - Assistant:나폴레옹은 18세기 후반과 19세기 초에 프랑스를.....

PROMPTING TECHNIC

- Few-shot examples
 - 모델에 몇 가지 예시를 제공하고 해당 예시를 참고하여 답변을 유도하여 좀 더 나은 성능을 발휘하도록 하는 방법

prompt: 고양이: 동물
오렌지: 과일
토마토: 식물
비둘기:

output: 조류

PROMPTING TECHNIC

- Few-shot examples
 - 충분히 큰 모델에만 작동
 - 산술 추론 문제에는 적합하지 않음
 - 정답을 정확히 주기 보다 레이블의 랜덤한 분포를 가지게 유도

PROMPTING TECHNIC

- Chain of Thought
 - LLM에게 이유에 대해서 설명하도록 만들어 답을 더 정확하게 생성할 수 있게 만드는 방법
 - 중간 추론 단계를 거치도록 하여 복잡한 사고가 필요한 작업에 정확도를 향상 시키는 방법
 - 답이 나오는 과정에 대해 설명한 예시를 보여주고, 지시에 대한 답을 생성할 때 예시와 동일한 방식으로 그에 대한 과정을 설명하도록 구성함

PROMPTING TECHNIC

Standard Prompting

Model Input

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: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

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: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

PROMPTING TECHNIC

- Zero-shot CoT
 - 문구를 마지막에 추가함으로써 CoT 예제를 제공하는 것과 같은 효과를 내어 성능을 향상시키는 방법

ex) Let's think step by step.

단계별로 생각해봅시다.

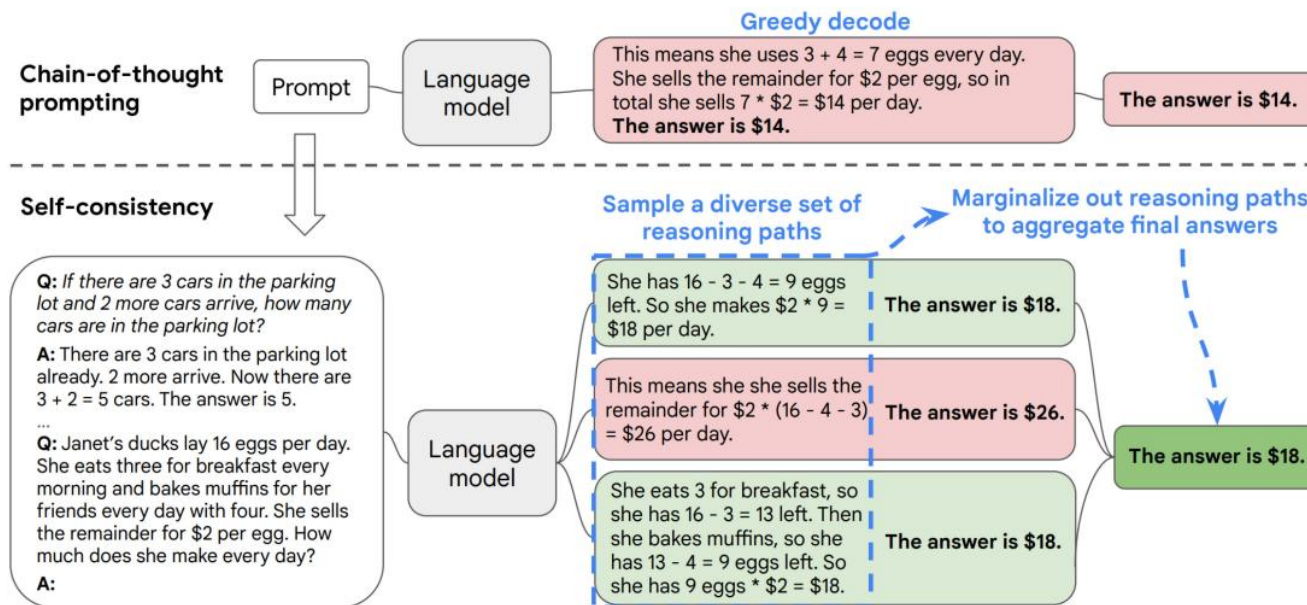
차근차근 생각해봅시다.

PROMPTING TECHNIC

- Self-Consistency
 - 여러가지 방식의 CoT를 통해 다양한 추론 과정을 거치고, 그중에서 가장 많이 나온 답을 선택하는 CoT의 확장 기술
 - CoT의 예제 경로를 다양하게 제공하여, LLM이 여러 번 추론함과 동시에 랜덤하게 경로를 선택해 답을 생성하도록 유도

PROMPTING TECHNIC

- Self-Consistency



PROMPTING TECHNIC

- Self-Consistency
 - 추론 경로를 최대한 다양하게 제공해야 하므로 토큰 수를 많이 사용함
 - 결과 취합을 위한 후처리 필요

PROMPTING TECHNIC

- Selection-Inference
 - 복잡한 문제를 해결하기 위해 여러 추론 단계를 연결하는 방법
 - 선택과 추론사이를 번갈아 가며 일련의 해석 가능한 원인과 결과론적 인 추론 단계를 생성하여 최종 답변을 유도하는 방식

PROMPTING TECHNIC

- Selection-Inference

Context:
wolves are afraid of mice
sheep are afraid of wolves
emily is a wolf
mice are afraid of wolves
winona is a wolf
cats are afraid of sheep
jessica is a cat
gertrude is a sheep
Question: what is emily afraid of?

Selection: emily is a wolf and
wolves are afraid of mice
Inference: emily is afraid of mice

Context:
julius is a swan
julius is gray
bernhard is a swan
d is yellow
brian is a lion
greg is a lion
brian is white
lily is a frog
lily is gray
Question: what color is greg?

Selection: brian is a lion. We
know that brian is white
Inference: we inferred that
lions are often white

Selection: greg is a lion. We
know that we inferred that lions
are often white
Inference: greg is white

PROMPTING TECHNIC

Prompt:

Context:

만약 누군가가 조용하다면, 그들은 흰색입니다.
만약 누군가가 젊고 빨강하다면, 그들은 흰색입니다.
젊은 사람들은 친절합니다.
만약 누군가가 친절하다면, 그들은 등급입니다.
모든 조용한 사람들은 젊습니다.
빨강고, 큰 사람들은 친절합니다.
등급, 빨간 사람들은 흰색입니다.
만약 누군가가 등급다면, 그들은 조용합니다.
해리는 젊습니다.
밥은 빨강입니다.
밥은 큼니다.

Question: 이것은 "해리는 조용하지 않다"라는 문장이 참인지를 암시하나요?

질문의 답에 필요한 내용을 Context 에서 추출해서 나열하세요.

Selection:

해리는 젊습니다., 모든 조용한 사람들은 젊습니다., 만약 누군가가 조용하다면, 그들은 흰색입니다.

Selection의 내용에 기반해 Question에 대해 답하세요.

Answer: 아니오, 해리는 조용하지 않다는 문장은 참이 아닙니다.

PROMPTING TECHNIC

- Selection-Inference
 - 추론 흔적이 남아 해석이 가능함 그래서 디버깅에 유용함
 - 논문에서 제안하는 반복적인 Selection-Inference를 실사용 케이스에 적용하려면 추론 단계를 생성하고 종료하는 방법을 세심하게 구성해야 함
 - 복잡한 추론 뿐만 아니라, 주어진 내용에서 정보를 정확히 추출할 수 있으므로 모델의 성능을 높이는데 유용함

PROMPTING TECHNIC

- Least to Most
 - 하나의 문제를 더 작은 여러 개의 하위 작업으로 분할하는 기법
 - 프롬프트로 다음 문제를 생성하고 해당 질문을 통해 다음 답을 생성하고, 이 과정을 최종 답변이 생성될 때까지 반복
 - 하위 작업들은 한번에 하나씩 해결하여 생성한 답을 이용하여 다음 문제를 해결하는데 사용함

PROMPTING TECHNIC

- Least to Most

Stage 1: Decompose Question into Subquestions

Q: It takes Amy 4 minutes to climb to the top of a slide. It takes her 1 minute to slide down. The water slide closes in 15 minutes. How many times can she slide before it closes?

Language Model

A: To solve "How many times can she slide before it closes?", we need to first solve: "How long does each trip take?"

Stage 2: Sequentially Solve Subquestions

It takes Amy 4 minutes to climb to the top of a slide. It takes her 1 minute to slide down. The slide closes in 15 minutes.

Subquestion 1 — Q: How long does each trip take?

Language Model

A: It takes Amy 4 minutes to climb and 1 minute to slide down. $4 + 1 = 5$. So each trip takes 5 minutes.

It takes Amy 4 minutes to climb to the top of a slide. It takes her 1 minute to slide down. The slide closes in 15 minutes.

Append model answer to Subquestion 1

Q: How long does each trip take?
A: It takes Amy 4 minutes to climb and 1 minute to slide down. $4 + 1 = 5$. So each trip takes 5 minutes.

Subquestion 2 — Q: How many times can she slide before it closes?

Language Model

A: The water slide closes in 15 minutes. Each trip takes 5 minutes. So Amy can slide $15 \div 5 = 3$ times before it closes.

PROMPTING TECHNIC

Context: 영화가 미끄럼틀 꼭대기까지 올라가는 데 4분이 걸립니다. 그녀가 내려오는 데는 1분이 걸립니다. 미끄럼틀이 15분 후에 문을 닫습니다.

Question: 그녀가문이 닫히기 전에 몇 번 미끄러질 수 있을까요?

이 문제의 답을 하기 위해 먼저 풀어야 할 하위 질문을 하나 만들어주세요.

영화가 미끄럼틀을 한 번 오르고 내려오는 데에는 총 몇 분이 걸리나요?

Context: 영화가 미끄럼틀 꼭대기까지 올라가는 데 4분이 걸립니다. 그녀가 내려오는 데는 1분이 걸립니다. 미끄럼틀이 15분 후에 문을 닫습니다.

Question: 영화가 미끄럼틀을 한 번 오르고 내려오는 데에는 총 몇 분이 걸리나요? 천천히 생각해 보세요.

영화가 미끄럼틀을 한 번 오르고 내려오는 데에는 총 5분이 걸립니다.
(4분 오르기 + 1분 내려오기 = 5분)

Context: 영화가 미끄럼틀 꼭대기까지 올라가는 데 4분이 걸립니다. 그녀가 내려오는 데는 1분이 걸립니다. 미끄럼틀이 15분 후에 문을 닫습니다.

Question: 영화가 미끄럼틀을 한 번 오르고 내려오는 데에는 총 몇 분이 걸리나요?

Answer: 영화가 미끄럼틀을 한 번 오르고 내려오는 데에는 총 5분이 걸립니다. (4분 오르기 + 1분 내려오기 = 5분)

Question: 그녀가문이 닫히기 전에 몇 번 미끄러질 수 있을까요? 단계적으로 생각해 보세요.

문이 닫히기까지 남은 시간은 15분입니다. 영화가 미끄럼틀을 오르고 내려오는 데에는 5분이 걸리므로, 15분을 5분으로 나누면 3번 미끄러질 수 있습니다.

PROMPTING TECHNIC

- Least to Most
 - Cot와 Selection-Inference를 결합한 방법
 - Task를 작은 문제로 분할하여 해결하는 분할 정복 방식을 이용하여 Autonomous Agent 개발 등에 사용함

PROMPTING TECHNIC

- ReAct
 - 실행 계획을 유도하고 추적하여 작업별로 실행할 액션을 선택하고 실행하는 방법
 - 외부 API와 검색엔진을 통해 신뢰할 수 있는 정보를 사용하거나, 다양한 도구(계산기, 이미지 생성)들을 사용할 수 있음

PROMPTING TECHNIC

- ReAct

(1) Hotspot QA

Question: Aside from the Apple Remote, what other device can control the program Apple Remote was originally designed to interact with?

(1d) ReAct (Reason + Act)

Thought 1: I need to search Apple Remote and find the program it was originally designed to interact with.
Act 1: `Search[Apple Remote]`
Obs 1: The Apple Remote is a remote control introduced in October 2005 by Apple ... originally designed to control the Front Row media center program ...

Thought 2: Apple Remote was originally designed to control the Front Row media center program. I need to search Front Row next and find what other device can control it.
Act 2: `Search[Front Row]`
Obs 2: Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports', 'Front Row (software)', ...]

Thought 3: Front Row is not found. I need to search Front Row (software) .
Act 3: `Search[Front Row (software)]`
Obs 3: Front Row is a discontinued media center software ...

Thought 4: Front Row (software) is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys.
Act 4: `Finish[keyboard function keys]`



PROMPTING TECHNIC

생각, 행동, 관찰 단계를 번갈아 가며 질문 응답 작업을 해결합니다. 생각은 현재 상황에 대해 추론할 수 있고, 행동은 세 가지 유형이 있습니다:

- (1) Search[entity], 이는 위키백과에서 정확한 엔티티를 검색하고 존재하는 경우 첫 번째 문단을 반환합니다. 존재하지 않는 경우, 검색할 수 있는 유사한 엔티티를 반환합니다.
- (2) Lookup[keyword], 이는 현재 문단에서 키워드를 포함하는 다음 문장을 반환합니다.
- (3) Finish[answer], 이는 답을 반환하고 작업을 마칩니다.



다음은 몇 가지 예입니다.

질문: 콜로라도 주 오로제니의 동쪽 부분이 확장되는 지역의 고도 범위는 어떻게 되나요?

생각 1: 콜로라도 주 오로제니를 검색하고, 콜로라도 주 오로제니의 동쪽 부분이 확장되는 지역을 찾은 다음, 그 지역의 고도 범위를 찾아야 합니다.

행동 1: Search[콜로라도 주 오로제니]

관찰 1: 콜로라도 주 오로제니는 콜로라도 주와 주변 지역에서 산맥 형성(오로제니)의 한 단계였습니다.

생각 2: 동쪽 부분에 대해 언급하지 않았습니다. 그래서 동쪽 부분을 찾아봐야 합니다.

행동 2: Lookup[동쪽 부분]

관찰 2: (결과 1/1) 동쪽 부분은 고원 지대로 확장되며 이를 중앙 평원 오로제니라고 합니다.

생각 3: 콜로라도 오로제니의 동쪽 부분은 고원 지대로 확장됩니다. 따라서 고원 지대를 검색하고 그 고도 범위를 찾아야 합니다.

행동 3: Search[고원 지대]

관찰 3: 고원 지대는 두 가지 명확한 땅의 지역 중 하나를 가리킵니다.

생각 4: 대신 고원 지대 (미국)을 검색해야 합니다.

행동 4: Search[고원 지대 (미국)]

관찰 4: 고원 지대는 대평원의 부분지역입니다. 동쪽에서 서쪽으로, 고원 지대의 고도는 약 1,800ft에서 7,000ft (550m에서 2,130m)로 상승합니다.[3]

생각 5: 고원 지대는 고도가 약 1,800ft에서 7,000ft로 상승하므로, 답은 1,800ft에서 7,000ft입니다.

행동 5: Finish[1,800ft에서 7,000ft]

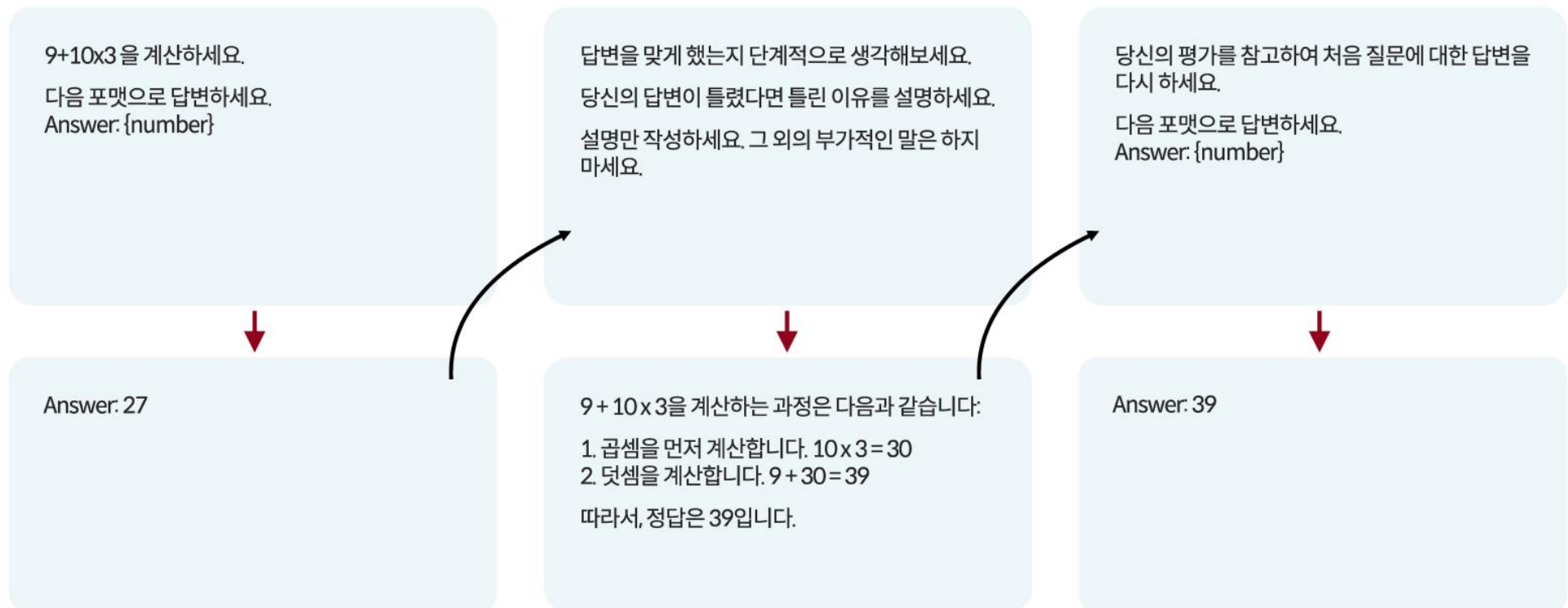
PROMPTING TECHNIC

- ReAct
 - 실제 툴을 사용해서 정보를 주입하지 않아도, Wikipedia 검색을 이용하는 것 만으로 성능을 향상 시킬 수 있음
 - 실제 툴을 사용할 때는 각 단계별로 순차적으로 결과를 출력하도록 제어 필요
 - 기본 프롬프트가 매우 길어지므로 토큰 제어에 유의해야 함

PROMPTING TECHNIC

- Self Evaluation
 - LLM이 생성한 결과를 LLM이 평가하게 하여, 오류를 잡거나 결과를 향상시키는 방법
 - 데이터셋 생성 자동화, 프롬프트 평가 자동화 등에 사용하여 데이터셋이나 LLM의 결과를 스스로 조정할 수 있게 함

PROMPTING TECHNIC



PROMPTING ENGINEERING

- Function calling
 - GPT 모델을 호출할 수 있는 함수(기능)를 설정해두어, 사용자의 요청에 맞게 해당 함수를 호출하는 기능
 - 직접 함수를 호출하는 것이 아니고 JSON형식을 이용
 - GPT-4 수준의 성능을 가진 LLM으로만 의미 있는 평가를 할 수 있으므로, 사용범위에 제약이 따름

PROMPTING ENGINEERING

- Function calling
 - name -> 필요한 기능 확인 시 반환할 함수명
 - description -> 해당 함수에 대한 설명
 - parameters -> 함수 호출에 필요한 파라미터 정의
 - required -> 함수 호출을 만들기 위해 필요한 필수 정보

PROMPTING ENGINEERING

- Function calling

```
{
  "name": "get_current_weather",
  "description": "주어진 위치의 현재 날씨를 가져옵니다.",
  "parameters": {
    "type": "object",
    "properties": {
      "location": {
        "type": "string",
        "description": "도시 또는 지역, 예) 서울",
      },
      "unit": {"type": "string", "enum": ["celsius", "fahrenheit"]},
    },
    "required": ["location"],
  },
}
```

PROMPTING ENGINEERING

- Function calling

User Input:

날씨 알려줘

Response:

```
{  
  'role': 'assistant',  
  'content': '어느 도시나 지역의 날씨를 알고 싶으세요?'  
}
```

PROMPTING ENGINEERING

- Function calling

User Input:

서울 날씨 알려줘

Response:

```
{
  'role': 'assistant',
  'content': None,
  'function_call': {
    'name': 'get_current_weather',
    'arguments':
      '{\n  "location": "Glasgow, Scotland",\n  "format":
"celsius"\n}'
  }
}
```

PROMPTING ENGINEERING

- Function calling

```
{
  "name": "set_recipe",
  "parameters": {
    "type": "object",
    "properties": {
      "dish": {
        "type": "string",
        "description": "Descriptive title of the dish"
      },
      "ingredients": {
        "type": "array",
        "items": {"type": "string"}
      },
      "instructions": {
        "type": "array",
        "description": "Steps to prepare the recipe.",
        "items": {"type": "string"}
      }
    }
  }
}
```

PROMPTING ENGINEERING

- Function calling

User Input:

버터 토스트 레시피를 알려줘

Response:

```
{
  "dish": "버터 토스트",
  "ingredients": [
    "빵 조각",
    "버터"
  ],
  "instructions": [
    "불지 않는 프라이팬이나 철판을 중간 불로 가열합니다.",
    "각 빵 조각의 한쪽에 버터를 펴 바릅니다.",
    "버터면이 아래로 향하도록 뜨거운 프라이팬이나 철판에 빵 조각을 놓습니다.",
    "약 2-3 분 동안 또는 바닥면이 노릇 노릇하고 바삭해질 때까지 요리합니다.",
    "빵 조각을 뒤집고 1-2 분 더 익 힌니다.",
    "프라이팬이나 철판에서 꺼내서 즉시 제공합니다."
  ]
}
```

PROMPTING ENGINEERING

- Function calling
 - 사용자의 아이디나 비밀번호 등 민간정보가 필요한 함수의 경우, 파라미터 설정에 넣지 않도록 주의
 - 응답 받은 함수를 직접적으로 실행하지 말고 간접적인 실행을 해야함(보안상 문제가 없는지, 또는 정상적인 파라미터를 이용한 호출 확인)