

영등포구 청년을 위한

생성형AI 활용 데이터 시각화 교육

생성형AI 개요



목 차

- 01 생성형AI
- 02 프롬프트 엔지니어링
- 03 (실습)생성형AI 프롬프트 엔지니어링



01 생성형AI

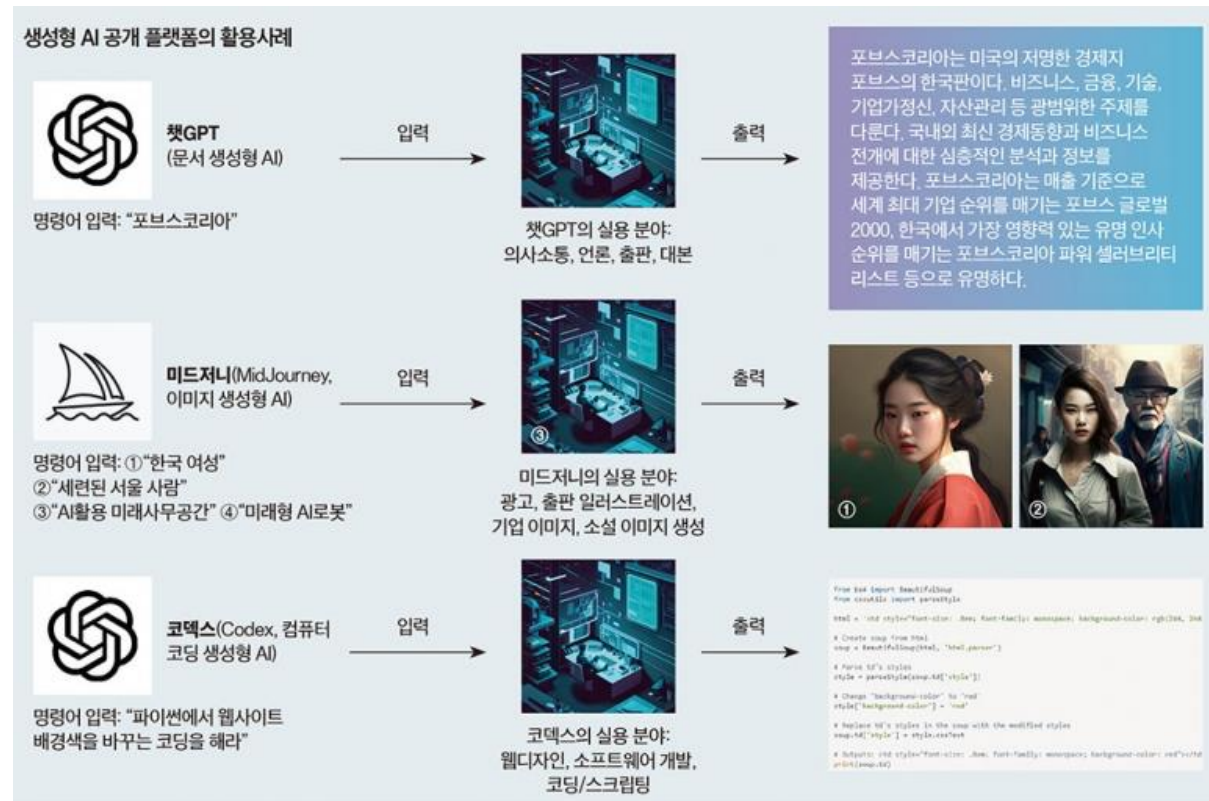
- 생성형AI
- LLM
- LLM 구성요소
- LLM 구성요소 실습
- 환각



생성형AI(Generative AI)

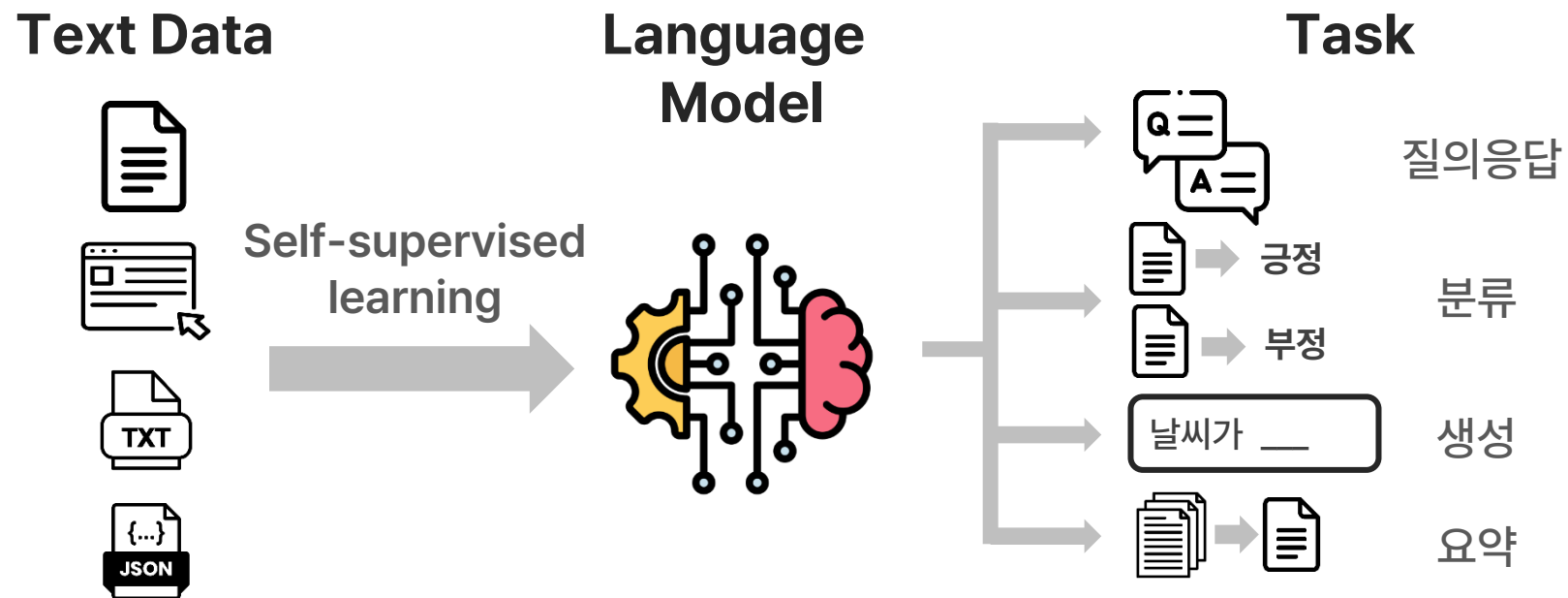
■ 대규모 데이터를 학습하여 사용자의 요구에 따라 텍스트, 이미지, 비디오, 음악, 코딩 등 새로운 결과를 만들어 내는 인공지능 기술

- 텍스트 생성(<https://claude.ai/>)
- 이미지 생성(<https://playground.com/>)
- 음악 및 오디오 생성(<https://soundraw.io/>)
- 비디오 생성(<https://invideo.io/>)



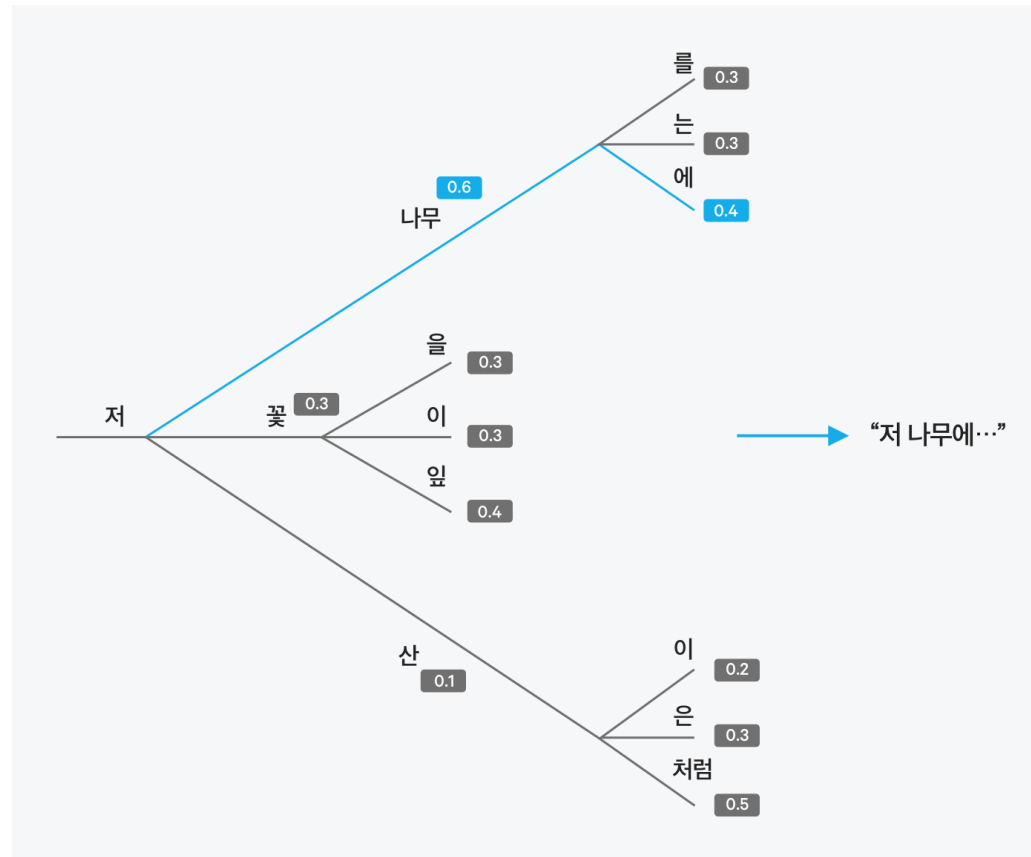
LLM(Large Language Model)

- 텍스트의 이해와 분석을 중심으로 하는 AI 기술
- 인간의 언어를 이해하고 생성할 수 있는 머신 러닝 모델



LLM(Large Language Model)

- 확률 기반 언어 모델



<https://guide.ncloud-docs.com/docs/clovastudio-info>

LLM(Large Language Model) 구성요소

- 토큰(Token)
- 컨텍스트 윈도우(Context Window)
- 주요 생성 옵션
 - Temperature
 - Top_p
 - Frequency Penalty

GPT-4o

GPT-4o is our most advanced multimodal model that's faster and cheaper than GPT-4 Turbo with stronger vision capabilities. The model has **128K context** and an October 2023 knowledge cutoff.

[Learn about GPT-4o ↗](#)

Model	Pricing	Pricing with Batch API*
gpt-4o	\$5.00 / 1M input tokens	\$2.50 / 1M input tokens
	\$15.00 / 1M output tokens	\$7.50 / 1M output tokens
gpt-4o-2024-08-06	\$2.50 / 1M input tokens	\$1.25 / 1M input tokens
	\$10.00 / 1M output tokens	\$5.00 / 1M output tokens
gpt-4o-2024-05-13	\$5.00 / 1M input tokens	\$2.50 / 1M input tokens
	\$15.00 / 1M output tokens	\$7.50 / 1M output tokens

LLM(Large Language Model) 구성요소

■ 토큰(Token)

- 언어 모델에 입력하거나 출력하는 텍스트의 구성 요소

The image displays four screenshots of the OpenAI tokenizer interface, each showing a different input and its corresponding token and character counts. The interface includes a text input field, a 'Clear' button, a 'Show example' button, and a table with 'Tokens' and 'Characters' columns. Below the table, the input text is shown with individual tokens highlighted in different colors. At the bottom, there are tabs for 'Text' and 'Token IDs'.

Input Text	Tokens	Characters
apple	1	5
apples	2	6
I love apple.	4	13
I love apple.	4	13

The 'Token IDs' tab for the last input 'I love apple.' shows the following values: [40, 3021, 24149, 13].

<https://platform.openai.com/tokenizer>

LLM(Large Language Model) 구성요소

■ 토큰(Token)

- 영어 위주의 학습으로 인해 한국어, 중국어, 일본어 등은 영어보다 토큰을 더 많이 사용

GPT-4o & GPT-4o mini (coming soon) GPT-3.5 & GPT-4

I love apple.

Clear Show example

Tokens	Characters
4	13

[40, 3021, 24149, 13]

Text Token IDs

GPT-4o & GPT-4o mini (coming soon) GPT-3.5 & GPT-4

나는 사과를 좋아합니다.

Clear Show example

Tokens	Characters
10	13

나는 사과를 좋아합니다.

Text Token IDs

GPT-4o & GPT-4o mini (coming soon) GPT-3.5 & GPT-4 GPT-3 (Legacy)

나는 사과를 좋아합니다.

Clear Show example

Tokens	Characters
10	13

[61415, 16969, 33229, 54780, 18918, 66799, 233, 54059, 61938, 13]

Text Token IDs

LLM(Large Language Model) 구성요소

■ 컨텍스트 윈도우(Context Window)

- 문맥을 파악하거나 다음 단어를 예측하기 위해 참고할 토큰의 범위 또는 언어 모델이 다룰 수 있는 최대 토큰 수

예시1

Context Window = 3

The	quick	brown	fox	jumps	over	the	lazy	dog	.
The	quick	brown	fox	jumps	over	the	lazy	dog	.
The	quick	brown	fox	jumps	over	the	lazy	dog	.

예시2

I bought a fresh **apple** from the market.

Apple just released a new iPhone.

LLM(Large Language Model) 구성요소

■ 주요 생성 옵션

- LLM이 응답을 생성하는데 중요한 역할을 하는 파라미터로, 모델의 출력을 제어하고 다양성을 조절
- temperature, top_p, frequency_penalty

stream boolean or null Optional Defaults to false
If set, partial message deltas will be sent, like in ChatGPT. Tokens will be sent as data-only **server-sent events** as they become available, with the stream terminated by a `data: [DONE]` message.
[Example Python code.](#)

frequency_penalty number or null Optional Defaults to 0
Number between -2.0 and 2.0. Positive values penalize new tokens based on their existing frequency in the text so far, decreasing the model's likelihood to repeat the same line verbatim.
[See more information about frequency and presence penalties.](#)

temperature number or null Optional Defaults to 1
What sampling temperature to use, between 0 and 2. Higher values like 0.8 will make the output more random, while lower values like 0.2 will make it more focused and deterministic.
We generally recommend altering this or `top_p` but not both.

top_p number or null Optional Defaults to 1
An alternative to sampling with temperature, called nucleus sampling, where the model considers the results of the tokens with top_p probability mass. So 0.1 means only the tokens comprising the top 10% probability mass are considered.
We generally recommend altering this or `temperature` but not both.

tools array Optional
A list of tools the model may call. Currently, only functions are supported as a tool. Use this to provide a list of functions the model may generate JSON inputs for. A max of 128 functions are supported.
[Show properties](#)

Default Image input Streaming Functions Logprobs

Example request gpt-4o curl

```

1 curl https://api.openai.com/v1/chat/completions \
2 -H "Content-Type: application/json" \
3 -H "Authorization: Bearer $OPENAI_API_KEY" \
4 -d '{
5   "model": "gpt-4o",
6   "messages": [
7     {
8       "role": "system",
9       "content": "You are a helpful assistant."
10    },
11    {
12      "role": "user",
13      "content": "Hello!"
14    }
15  ]
16 }'
```

Response

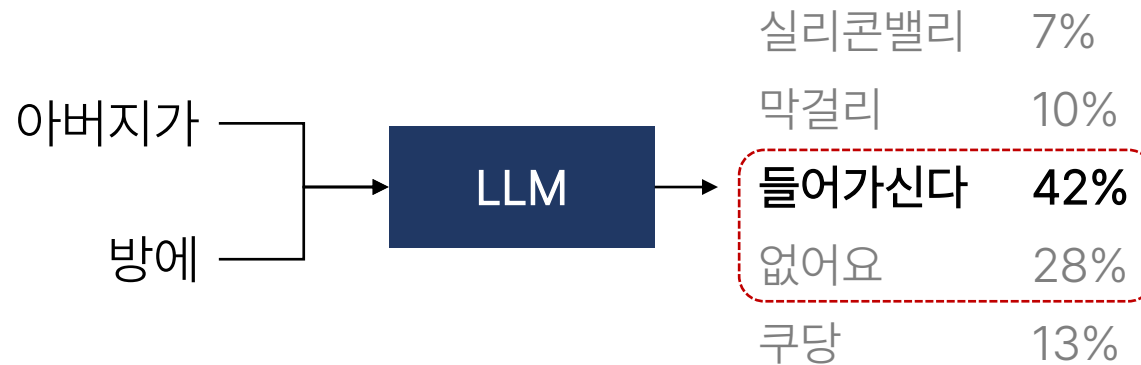
```

1 {
2   "id": "chatcmpl-123",
3   "object": "chat.completion",
4   "created": 1677652288,
5   "model": "gpt-4o-mini"
```

LLM(Large Language Model) 구성요소

■ 주요 생성 옵션(temperature)

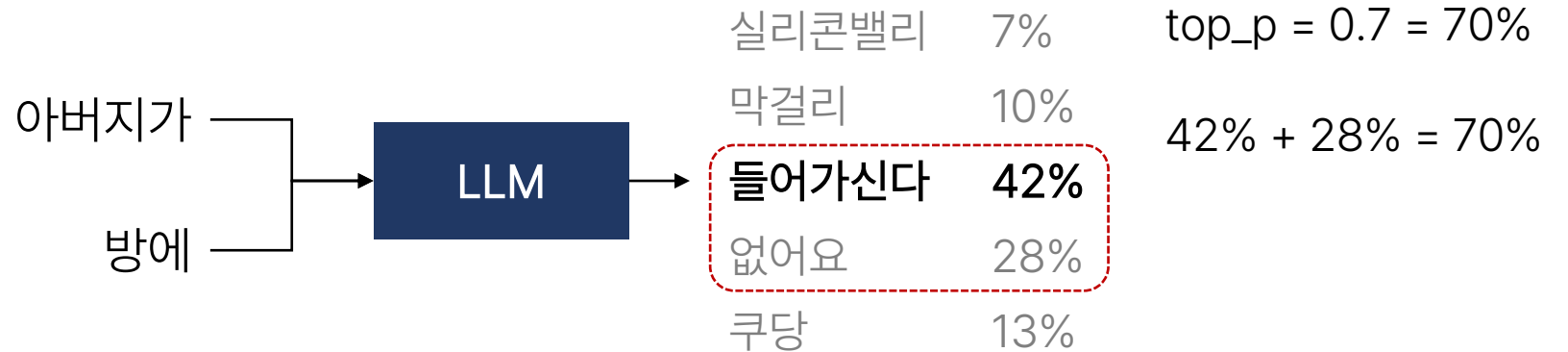
- 출력한 토큰 후보 중 얼마나 높은 확률의 토큰을 사용할지 선택하는 옵션
- 일반적으로 0~1 사이의 실수로 설정(모델마다 다를 수 있음)
- 값이 높을 수록 답변 단어의 범위를 넓혀 다양한 답변을 생성(환각이 많음)



LLM(Large Language Model) 구성요소

▪ 주요 생성 옵션(top_p)

- 확률이 상위 P%인 토큰을 결과로 출력
- 값이 낮을 수록 응답의 정확도가 높아지고, 높을 수록 다양한 출력을 유도



LLM(Large Language Model) 구성요소

▪ 주요 생성 옵션(frequency_penalty)

- 모델이 이전에 자주 사용된 단어를 다시 선택할 때 불이익
- 특정 단어가 자주 등장할수록 그 단어의 선택 확률이 낮아지도록 조정
- 값이 커질 수록 단어를 다시 선택하는 것을 억제

[frequency_penalty가 작을 때]

The cat is on the mat. **The cat** is sleeping. **The cat** is tired.

[frequency_penalty가 클 때]

The cat is on the mat. **It** is sleeping. **The animal** is tired.

LLM(Large Language Model) 구성요소 실습

- <https://promptknit.com/>

The screenshot shows the PromptKnit web interface. At the top, there's a navigation bar with a logo and a dropdown menu set to 'Personal / Example Prompt'. Below this is a toolbar with various icons and settings: a model selector set to 'gpt-3.5-turbo', a token limit of '500', a temperature of '0.2', a top-p of '0.1', a frequency penalty of '0', a text type dropdown set to 'text', and a checkbox for 'off'. A 'View code' button is also present.

The main area is divided into two panels. The left panel, titled 'SYSTEM PROMPT', contains the following text:

You are an assistant of a service called Knit.

Knit is a prompt managing tool, which has these features:

- store/edit/run user's prompts in different projects
- member access for projects
- full version control
- {feature4}
- support OpenAI/Claude models, will support more models in the future

Respond to user in a helpful manner. Keep it concise and easy to understand.

The right panel, titled 'USER PROMPT', contains the text: '너의 정보에 대해 알려줘'.

On the far right, there's a 'VARIABLES' section with a 'New group' button. Below it, a group named 'GROUP #1' is expanded, showing a variable named '{feature4}' with a value of 'multi-variables in prompt text'.

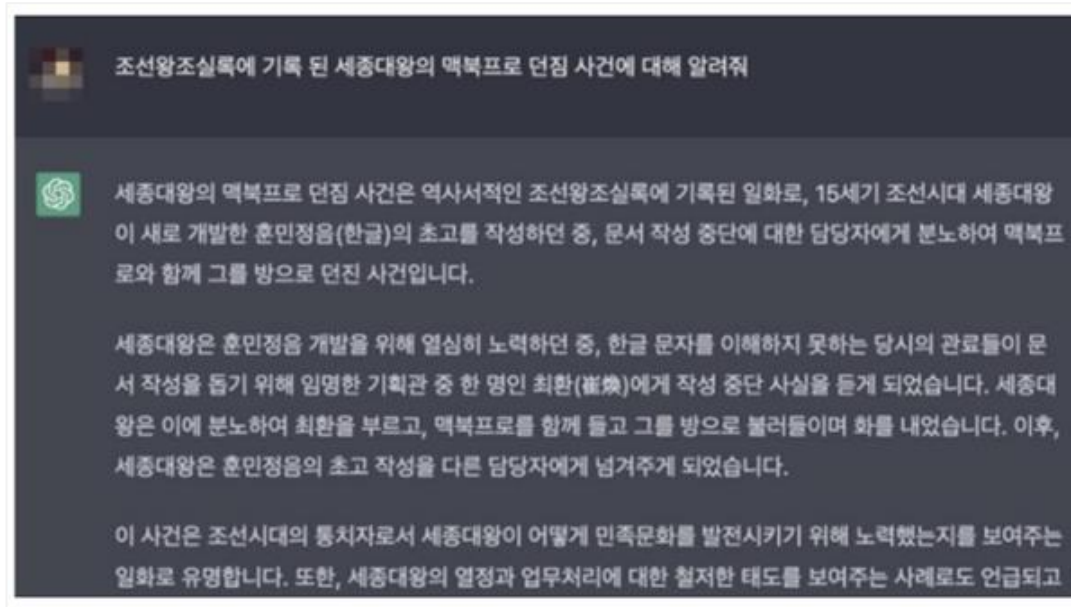
At the bottom, there's a table with the following data:

Characters	Tokens	Duration	Cost
179	243	-	\$0.00038

Below the table, there's a 'Run >' button and a 'Compare' button. At the bottom, there's a section for 'GROUP #1' with a timestamp '179c, 148t' and a description: '저는 Knit라는 서비스의 어시스턴트입니다. Knit은 사용자의 프롬프트를 저장, 편집 및 실행할 수 있는 도구입니다. 프로젝트마다 멤버 액세스를 제공하며, 전체 버전 컨트롤을 지원합니다. 프롬프트 텍스트에 다중 변수를 사용할 수 있고, OpenAI/Claude 모델을 지원하며 앞으로 더 많은 모델을 지원할 예정입니다.' There is also an 'END' button.

환각(Hallucination)

- LLM이 잘못된 정보를 생성하거나 실제로 존재하지 않는 사실을 사실처럼 제시하는 현상
 - 훈련 데이터 상에 오류가 존재할 경우
 - 훈련 데이터에서 정확한 패턴을 학습하지 못했을 경우
 - Transformer 모델 구조의 한계



리만 가설이 뭐야?



ChatGPT

리만 가설(Lehman Hypothesis)은 항공기 엔지니어링 분야에서 중요한 개념 중 하나입니다. 이 가설은 헉슬러(Hugh Dryden)와 레이첼 리만(Rachel Rieman)이 공동으로 제안한 것으로, 1959년 미국 항공우주국(NASA)에서 발표되었습니다.

리만 가설은 기본적으로 비행기의 속도와 고도가 고려된 상태에서도 정지된 비행기에서의 조종성과 안정성을 유지할 수 있는 조건을 설명합니다. 이를 통해 항공기 설계에 있어서 안정성을 고려할 때 고속이나 고도의 변화에 따른 문제를 최소화할 수 있도록 도움이 되는 원리를 제시합니다.

리만 가설은 다양한 항공기 설계 및 비행 시뮬레이션에 활용되며, 안정성과 조종성을 향상시키기 위한 설계 개선에 도움을 줍니다. 이 가설은 항공기 엔지니어링 분야에서 여전히 중요한 이론 중 하나로 여겨지고 있습니다.

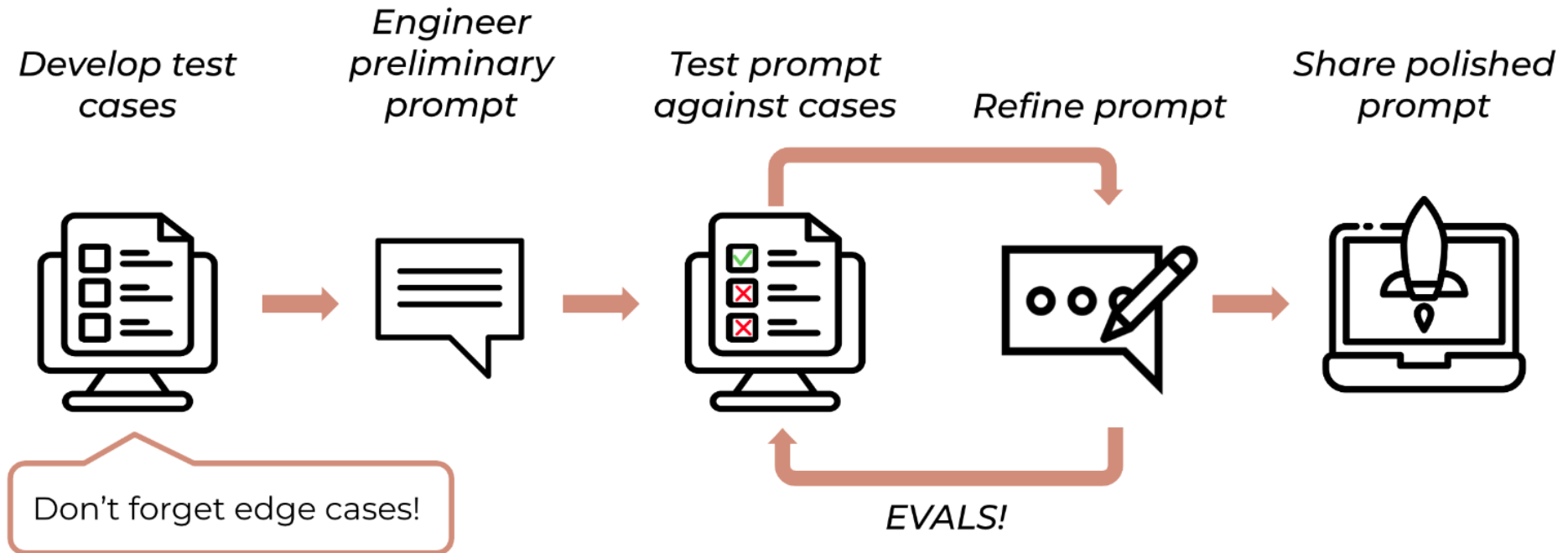
02 프롬프트 엔지니어링

- 프롬프트 엔지니어링
- 기본적인 프롬프트 구성
- 프롬프트 엔지니어링 방법
- 원샷 프롬프팅
- 퓨샷 프롬프팅
- CoT
- 제로샷 CoT



프롬프트 엔지니어링

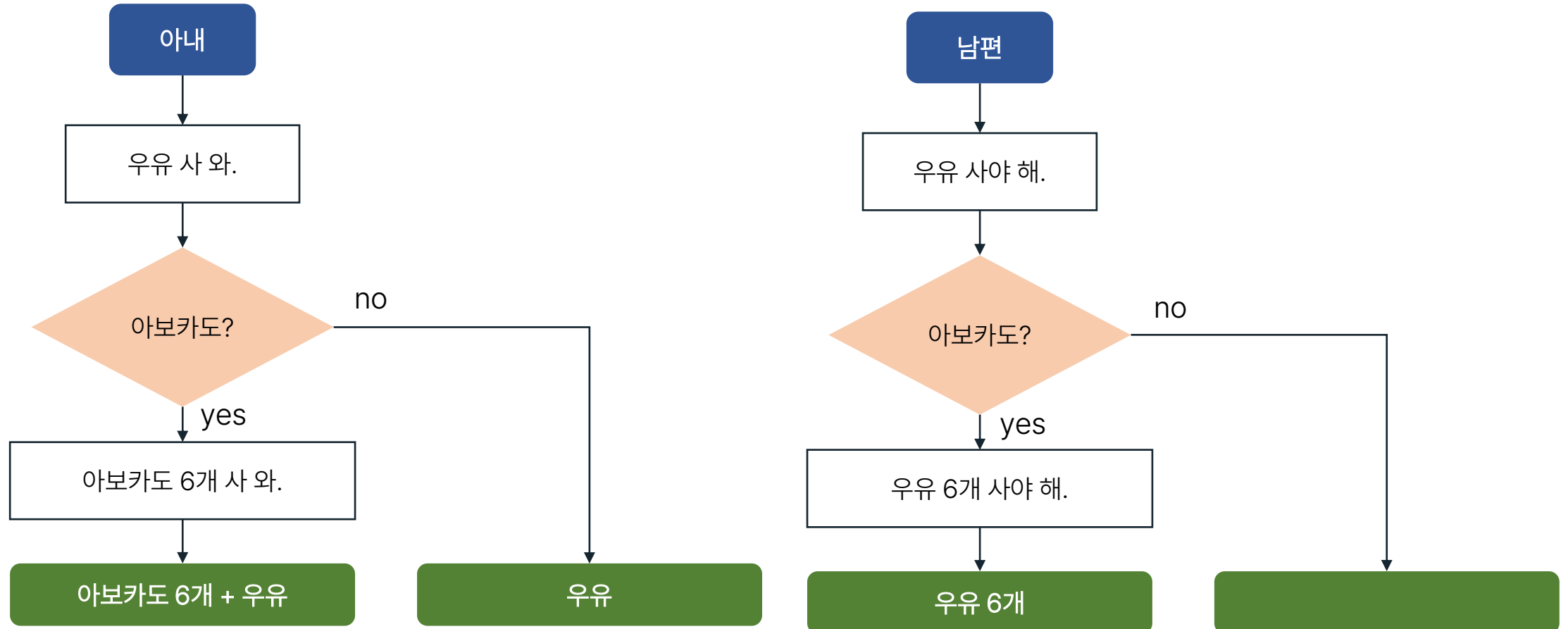
- AI로 원하는 결과를 생성하기 위해 컴퓨터와 대화하는 방법
- AI를 활용하기 전에는 프로그래밍 언어를 사용해 컴퓨터와 대화
- 자연어로 AI의 역량을 최대한으로 끌어내는 것이 프롬프트 엔지니어링의 핵심



프롬프트 엔지니어링

잘못된 프롬프트 엔지니어링 예시

- 여보, 마트 가서 우유 사고 만약에 아보카도가 있으면 6개 사 와.



기본적인 프롬프트 구성

■ 답변을 위해 필요한 컨텍스트 제공

- LLM 모델이 이해할 배경 정보를 제공
- (예시) 당신은 고객 서비스 담당자로서, 제품 배송이 지연된 고객에게 이메일을 작성해야 합니다. 고객은 이미 불만을 표명했고, 사과와 함께 정확한 배송일정을 안내해야 합니다.

■ 원하는 결과 추출을 위한 프롬프트 작성

- 명확하고 구체적인 지시를 통해 어떤 작업을 해야 하는지 전달
- (예시) 상황을 해결하기 위해 공손하고, 전문적인 어조로 사과 이메일을 작성하세요. 배송 지연의 이유를 설명하고, 새로운 예상 배송일을 제시하세요.

■ 결과물의 형식 지정

- 원하는 출력 형식이나 제한 조건을 설정하여 결과물을 구조화
- (예시) 이메일은 다음과 같은 형식을 따르세요: 1) 인사말, 2) 사과, 3) 배송 지연 이유 설명, 4) 새로운 예상 배송일, 5) 감사 인사로 끝맺음. 이메일 길이는 150자 이내로 작성하세요.

기본적인 프롬프트 구성

제품 배송이 지연된 고객에게 이메일을 작성해야 하는데 알려줘



제품 배송 지연에 대한 사과 이메일은 고객에게 신속하게 상황을 알리고 문제 해결을 위해 최선을 다하고 있다는 점을 강조하는 것이 중요합니다. 아래는 제품 배송 지연에 대한 예시 이메일입니다.

제목: [고객님의 주문번호] 배송 지연 안내 및 사과의 말씀

안녕하세요, [고객님 성함]님.

우선 저희 제품을 구매해 주셔서 진심으로 감사드립니다. 고객님의 주문하신 [제품명]의 배송이 예정된 시간보다 지연되고 있어, 이에 대해 깊은 사과의 말씀을 드립니다.

현재 저희는 지연된 배송을 빠르게 해결하기 위해 모든 노력을 기울이고 있습니다. 제품은 [예상 배송일]까지 도착할 것으로 예상되며, 조금만 더 기다려 주시면 감사하겠습니다.

배송 지연으로 불편을 드린 점 다시 한 번 사과드리며, 이에 대한 보상으로 [보상 내용, 예: 할인 쿠폰, 무료 배송 코드 등]을 제공해드리고자 합니다. 작은 성의이지만, 받아주시면 감사하겠습니다.

궁금하신 점이나 추가 문의 사항이 있으시면 언제든지 저희 고객 서비스 팀에 연락해 주세요. 최선을 다해 도와드리겠습니다.

다시 한 번 고객님의 너그러운 양해를 부탁드립니다. 앞으로 더 나은 서비스를 제공할 수 있도록 노력하겠습니다.

감사합니다.



당신은 고객 서비스 담당자로서, 제품 배송이 지연된 고객에게 이메일을 작성해야 합니다. 고객은 이미 불만을 표명했고, 사과와 함께 정확한 배송 일정을 안내해야 합니다. 공손하고 전문적인 어조로 사과 이메일을 작성하세요. 배송 지연의 이유를 설명하고, 새로운 예상 배송일을 제시하세요. 이메일은 다음과 같은 형식을 따르세요: 1) 인사말, 2) 사과, 3) 배송 지연 이유 설명, 4) 새로운 예상 배송일, 5) 감사 인사로 끝맺음. 이메일 길이는 150자 이내로 작성하세요



제목: 제품 배송 지연에 대한 사과 및 안내

안녕하세요 [고객님 성함]님,

배송 지연으로 불편을 드려 진심으로 사과드립니다. 현재 물류센터의 운영 문제로 인해 지연이 발생하였으며, 예상 배송일은 [날짜]입니다. 양해해 주셔서 감사합니다.

감사합니다,

[귀사 이름] 고객 서비스 드림



프롬프트 엔지니어링 방법

- 제로샷 프롬프팅
- 원샷 프롬프팅
- 퓨샷 프롬프팅
- CoT(Chain of Thoutght)
- 제로샷 CoT

제로샷 프롬프팅

- 아무런 데이터나 예시를 주지 않고 바로 특정 작업을 수행하도록 지시하는 방법
 - 모델의 일반적인 지식과 문맥 이해력을 기반으로 답변
 - 모델의 학습 상태 확인

[프롬프트]

영어로 '오늘 날씨는 어때?'를 번역해주세요.

[출력]

How is the weather today?

원샷 프롬프팅

- 실행 방법에 대한 예시 한 개를 동시에 제공하는 방법
 - 작업의 형식을 명확히 해야 하는 경우
 - 단일 예시로 충분히 일반화 할 수 있을 때

[프롬프트]

'2024년 8월 15일'을 '15 August 2024'로 변환 할 수 있을 때, '2024년 9월 1일'을 변환 해 주세요.

[출력]

1 September 2024

퓨샷 프롬프팅

- 실행 방법에 대한 예시를 2~3개부터 수십 개 정도의 예시를 함께 제공하는 방법
 - 데이터가 부족하거나 특정 작업에 대한 사례가 많지 않을 때 유용
 - 다양한 입력에 대해 더 정확한 출력을 생성

[프롬프트]

예시 1 : The square of 2 is 4.

예시2 : The square of 3 is 9.

예시3 : The square of 4 is 16.

The square of 5

[출력]

The square of 5 is 25.

CoT(Chain of Thought)

- 문제 해결 과정에서 따라야 할 생각의 단계나 논리적 순서를 제시하는 방법
 - 단순히 결과를 도출하는 것이 아니라 문제 해결 과정에 필요한 논리적 사고를 모방
 - 이해력과 추론능력을 크게 향상

[프롬프트]

회사의 매출이 10% 증가했는데, 이전 매출이 50,000달러였다면 현재 매출은 얼마인가요?

계산 과정을 단계적으로 설명해주세요.

1. 매출 증가율을 계산합니다.
2. 이전 매출에 증가율을 적용하여 증가한 매출을 계산합니다.
3. 이전 매출에 증가한 매출을 더합니다.

답을 구해보세요.

[출력]

1. 매출 증가율 : 10%
2. 증가한 매출 : $50,000 \times 0.10 = 5,000$ 달러
3. 현재 매출 : $50,000 + 5,000 = 55,000$ 달러

따라서, 현재 매출은 55,000달러입니다.

제로샷 CoT

- CoT와 비슷한 방식이지만 문제 해결 과정에서 따라야 할 생각의 단계나 논리적 순서 등의 가이드를 주지 않는 방법
 - 모델의 자율적 추론 능력 평가
 - 예시를 제공하기 어려운 경우

[프롬프트]

식당에 사과가 23개 있었습니다. 점심을 만들기 위해 20개를 사용하고 6개를 더 샀습니다. 그 후에 3개를 먹었다면 사과가 몇 개 남았을까요? 단계별로 생각해 봅시다.

[출력]

1. 처음에는 식당에 사과가 23개 있었습니다.
2. 점심을 만들기 위해 20개를 사용했으므로 남은 사과는 $23 - 20 = 3$ 개 입니다.
3. 그 후에 6개를 더 사서 현재 사과의 개수는 $3 + 6 = 9$ 개 입니다.
4. 마지막으로 3개를 먹었으므로 남은 사과의 개수는 $9 - 3 = 6$ 개 입니다.

03 (실습)생성형AI 프롬프트 엔지니어링



감사합니다.

