



Microsoft와 함께하는 Azure OpenAI Promphthon

CSU Data & AI 팀
고대운 CSA



ChatGPT가 없었다면 어떻게 풀 수 있을까요?

채팅 세션

🧹 채팅 지우기

⚙️ 플레이그라운드 설정

📄 코드 보기

원시 JSON 표시



모니터가 매우 뜨겁습니다.



부정

모니터가 시장 반응이 매우 뜨겁습니다.



긍정

Agenda

1. Azure OpenAI & RAG 소개(2h)

- Azure OpenAI LLM 배경 소개
- 생성형 AI 활용 꿀팁 및 사례
- RAG 아키텍처 구성

2. 프롬프트 엔지니어링 개요(0.5h)

- 프롬프트 엔지니어링 가이드 소개

3. Azure OpenAI Studio 활용(1h)

- Azure OpenAI Playground 기능 소개
- Add on your Data 데모

Agenda

4. Prompthon 코드 활용 소개(1h)

- 환경 세팅 및 실습 소개
- 코드 기반 프롬프트 엔지니어링
- Function Calling 활용
- STT/ChatGPT/TTS 기반 나만의 Jarvis 비서 만들기

5. Hands-on-Lab 실습 (2h)

- Azure OpenAI Studio: 권한 및 리소스 할당 준비 필요
- Prompthon 실습: GitHub 계정 사전 준비 필요

6. AutoGen 소개(0.5h)

- Multi-agents Conversation Framework 소개



Azure OpenAI LLM 배경 소개

배경



THE AGE OF AI HAS BEGUN



빌게이츠가 The age of AI has begun 이라고 제목을 지은 이유가 무엇인지 분석해서 요약해줘.

✓ “빌게이츠 The age of AI has begun 제목 분석” 검색 중
✓ 답변을 생성하는 중...

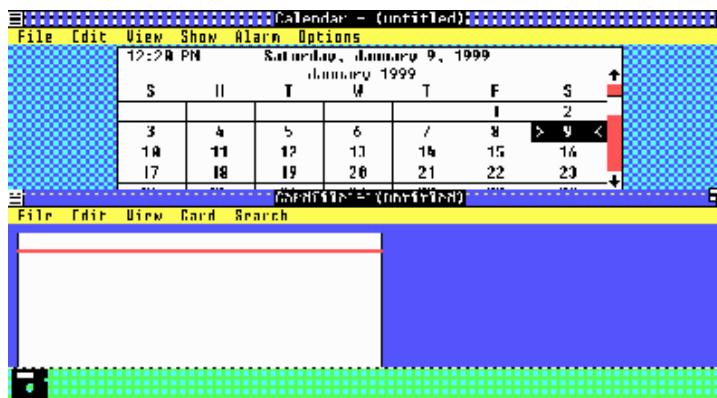
빌 게이츠는 인공지능이 모바일 폰과 인터넷과 같은 혁신적인 기술로서 혁명적이라고 생각한다고 합니다. 그는 1980년대에 그래픽 사용자 인터페이스를 처음 접했을 때와 같은 혁신적인 기술을 본 것이라고 말합니다 ¹. 그는 인공지능이 세상에서 가장 큰 문제를 해결하는데 도움이 될 수 있다고 믿으며, 이를 통해 세상의 불평등을 줄일 수 있다고 생각합니다 ².

¹ : The Age of AI has begun | Bill Gates ² : The Age of AI Has Begun

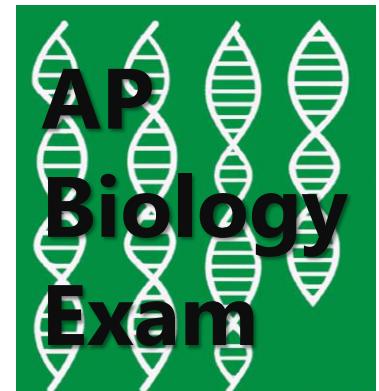
자세한 정보: [1. gatesnotes.com](https://gatesnotes.com) [2. pelayoarbues.github.io](https://pelayoarbues.github.io) [3. smalllake.kr](https://smalllake.kr) +2 더 보기

1 / 20 ●

Graphical User Interface



OpenAI GPT (59/60)



AI에 대한 세상의 관심

ChatGPT Sprints to One Million Users

Time it took for selected online services to reach one million users



* one million backers ** one million nights booked *** one million downloads
Source: Company announcements via Business Insider/LinkedIn

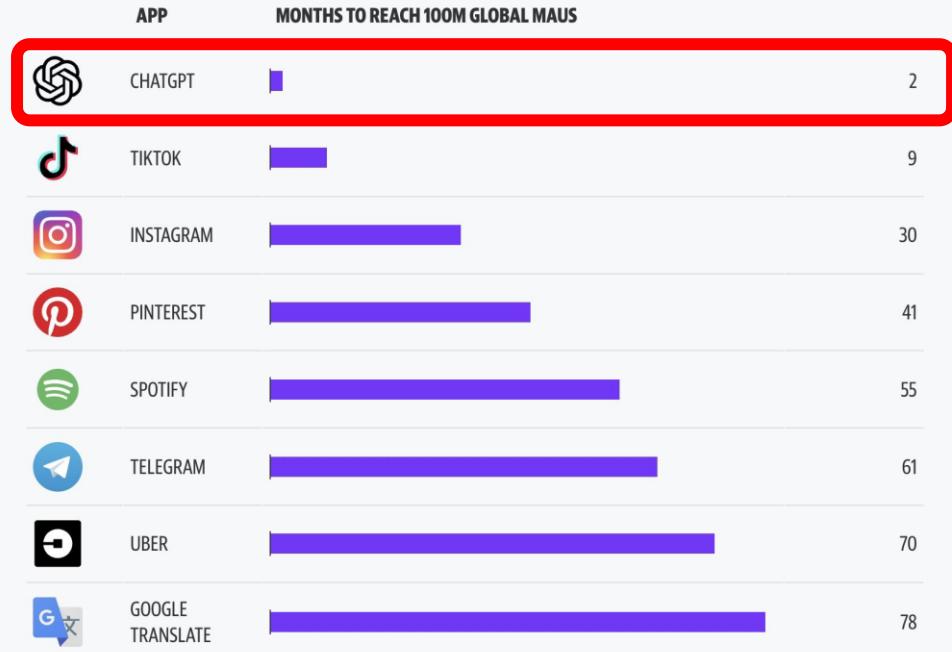


statista

URL: <https://www.statista.com/chart/29174/time-to-one-million-users>

HOW LONG IT TOOK TOP APPS TO HIT 100M MONTHLY USERS

ChatGPT is estimated to have hit 100M users in January, 2 months after its launch.
Here's how long it took other top apps to reach that:



SOURCE: UBS

yahoo!
finance

Generative AI 시대

Generative : 생성 | Pre-trained : 사전 학습된 | Transformer : 변환

Do more with less!

인프라 구축

인프라 운용

ML 환경 세팅

대량의 데이터 저장소

데이터 세트 모으기

양질의 데이터 세트 만들기

적합한 알고리즘

ML 학습 시간

ML 학습 비용

MLOps 환경 구축

추론 환경 AutoScaling

전문 인력 수급

Input Prompt: Recite the first law of robotics



Output:

이진우의 거리두기

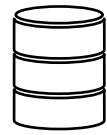
챗GPT의 도발, '당신은 질문할 줄 아는가?'

입력 : 2023.03.08 03:00 | 수정 : 2023.03.08 09:30 | 이진우 포스텍 명예교수

일반적인 Deep Learning의 과정

인프라 구축

데이터
준비



데이터 수집
및 전처리

데이터 세트 모으기

양질의 데이터 세트 만들기

인프라 운용

실험

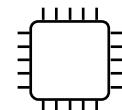
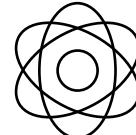


모델
만들기

적합한 알고리즘

ML 환경 세팅

실험



훈련 &
모델 검증

ML 학습 시간

대량의 데이터 저장소

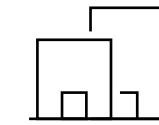


모델 등록
및 관리

↑

ML 학습 비용

배포



전문 인력 수급



이미지
생성

이미지 배포
모델 모니터링

추론 환경 AutoScaling

오케스트레이션
MLOps 환경 구축



MSCOCO Dataset – 데이터에 진심!



- traffic tie up at intersection in city area.
- bus merging through traffic at a busy intersection.
- many busses drive down a busy street.
- a bus driving down a street next to a bunch of traffic.
- a bus pulling out onto a busy road.

URL: <https://cocodataset.org/#explore?id=551018>

#Vision #ObjectDetection #Segmentation #Key-point Detection #Captioning #2014 #328K

초거대 언어 모델을 우리가 만들어야 할까?

D Total Compute Used to Train Language Models

This appendix contains the calculations that were used to derive the approximate compute used to train the language models in Figure 2.2. As a simplifying assumption, we ignore the attention operation, as it typically uses less than 10% of the total compute for the models we are analyzing.

Calculations can be seen in Table D.1 and are explained within the table caption.

Model	Total train compute (PF-days)	Total train compute (flops)	Params (M)	Training tokens (billions)	Flops per param per token	Mult for bwd pass	Fwd-pass flops per active param per token	Frac of params active for each token
T5-Small	2.08E+00	1.80E+20	60	1,000	3	3	1	0.5
T5-Base	7.64E+00	6.60E+20	220	1,000	3	3	1	0.5
T5-Large	2.67E+01	2.31E+21	770	1,000	3	3	1	0.5
T5-3B	1.04E+02	9.00E+21	3,000	1,000	3	3	1	0.5
T5-11B	3.82E+02	3.30E+22	11,000	1,000	3	3	1	0.5
BERT-Base	1.89E+00	1.64E+20	109	250	6	3	2	1.0
BERT-Large	6.16E+00	5.33E+20	355	250	6	3	2	1.0
RoBERTa-Base	1.74E+01	1.50E+21	125	2,000	6	3	2	1.0
RoBERTa-Large	4.93E+01	4.26E+21	355	2,000	6	3	2	1.0
GPT-3 Small	2.60E+00	2.25E+20	125	300	6	3	2	1.0
GPT-3 Medium	7.42E+00	6.41E+20	356	300	6	3	2	1.0
GPT-3 Large	1.58E+01	1.37E+21	760	300	6	3	2	1.0
GPT-3 XL	2.75E+01	2.38E+21	1,320	300	6	3	2	1.0
GPT-3 2.7B	5.52E+01	4.77E+21	2,650	300	6	3	2	1.0
GPT-3 6.7B	1.39E+02	1.20E+22	6,660	300	6	3	2	1.0
GPT-3 13B	2.68E+02	2.31E+22	12,850	300	6	3	2	1.0
GPT-3 175B	3.64E+03	3.14E+23	174,600	300	6	3	2	1.0

**GPT-3
(175B)**



**355
Years**



NVIDIA A100

**34
Days**

\$5M

**GPT-4
(2T ?)**

Microsoft & OpenAI 협업



❖ LLM을 위한 슈퍼컴퓨팅 환경 제공

- 전 세계 5번째로 큰 클라우드 기반의 첫 번째 슈퍼컴퓨터
- 30만 CPU 코어, 1만개 GPU로 구성

❖ OpenAI 모델을 이용한 서비스의 사용자 경험 개선



❖ 독점적인 OpenAI 서비스 공급사



GPT-3.5-Turbo

GPT-4

GPT-4-Turbo

GPT-4-Vision

DALL-E 2/3

Whisper





GitHub Copilot

지금이 전문 개발자로
탈바꿈할 기회입니다.
놓치지 마세요.

sentiment.ts write_sql.go parse_expenses.go

```
1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of the text is positive or negative
6 // Use a web service
7 async function isPositive(text: string) {
8   const response = await fetch(`http://text-processing.com/textsentiment`, {
9     method: "POST",
10    body: `text=${text}`,
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded"
13    }
14  });
15  const json = await response.json();
16  return json.label === "pos";
17}
```

Copilot

Microsoft 365 Copilot

Microsoft 365 앱에 탑재 – 생산성 & 업무 효율 UP



Business Chat은 문서, 프리젠테이션, 이메일, 캘린더, 메모 및 연락처의 데이터를 모두 자연어 기반으로 통합하여 사용



직전 온라인 미팅을 요약하고,
액션 아이템을 도출해줘.



연도별 상품 카테고리 별
수익을 라인차트로 만들어줘.



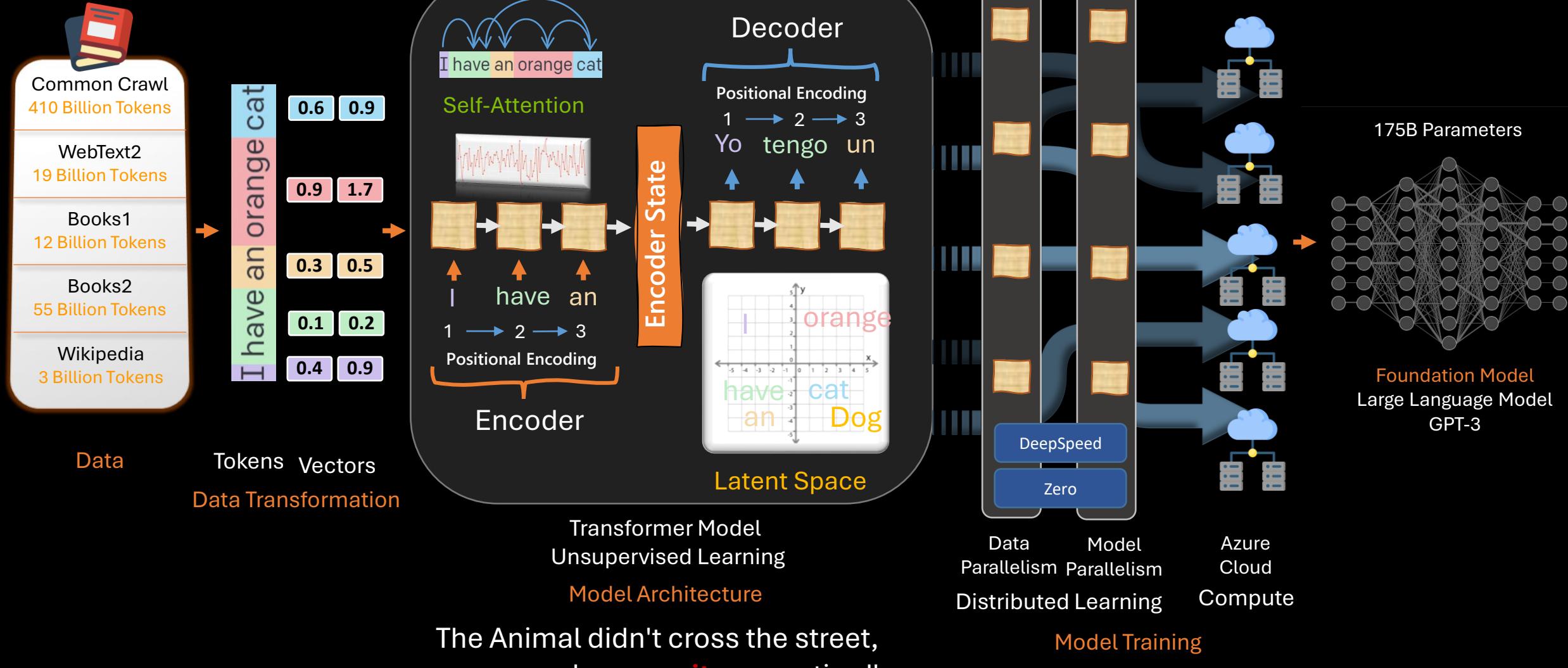
최근 내가 언급된 이메일을
주제별로 분류해서
발신자를 넣어서 요약해줘.

Azure OpenAI Service

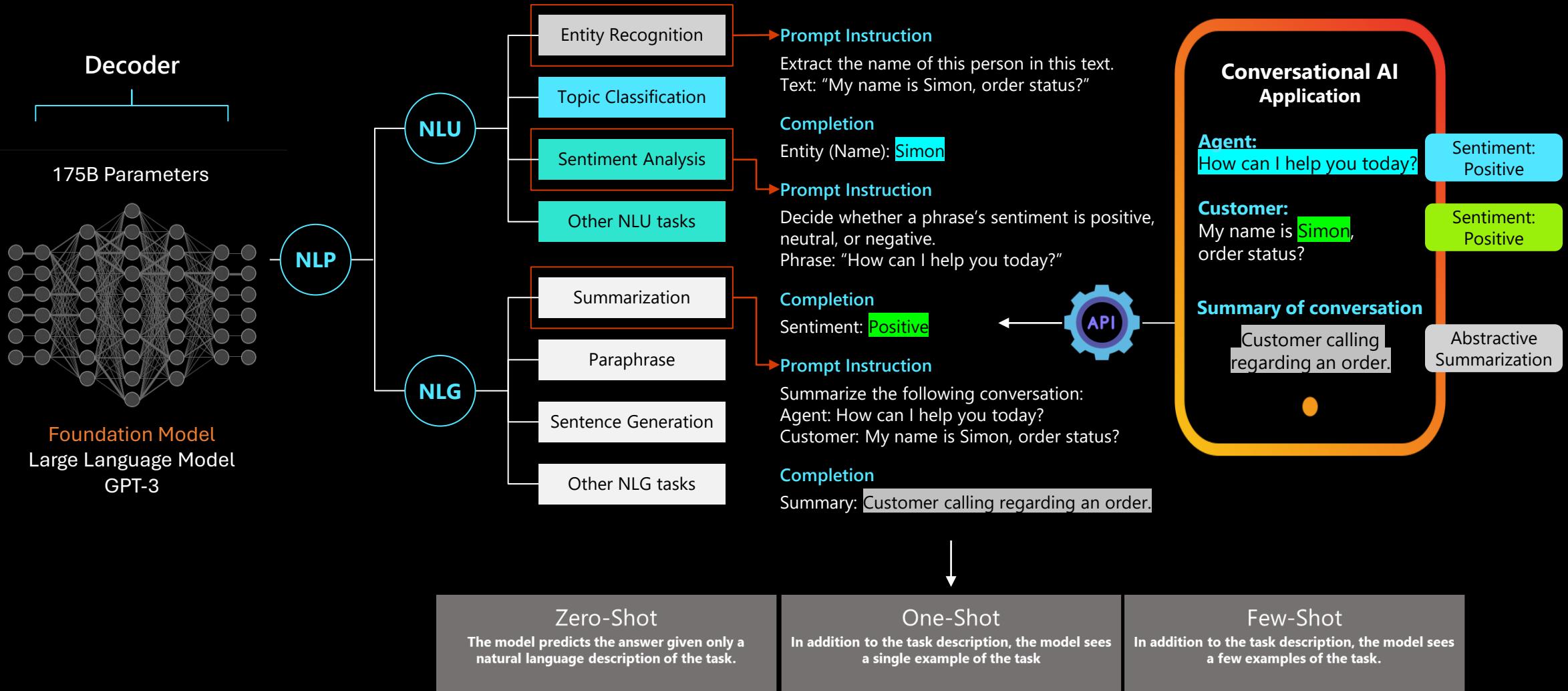
목적에 맞는 생성형 AI 기술을 채택



GPT-3 모델 훈련 아키텍처

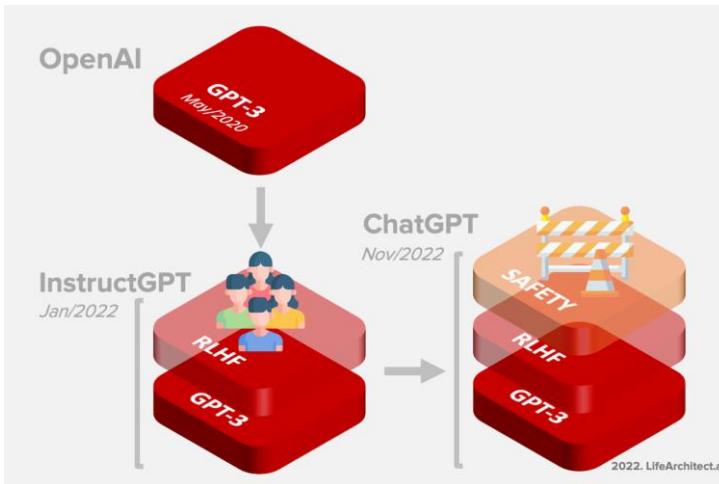


GPT-3 사용하기



ChatGPT의 탄생 배경 - 데이터의 중요성!

사람의 피드백을 통한 강화 학습
(RLHF + Safety)



Step 1. 샘플 데이터를 기반으로
지도학습 수행

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.



Step 2. 비교 데이터 수집
/ 보상 모델 훈련

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



Step 3. 강화학습 모델을 활용하여,
보상모델 최적화

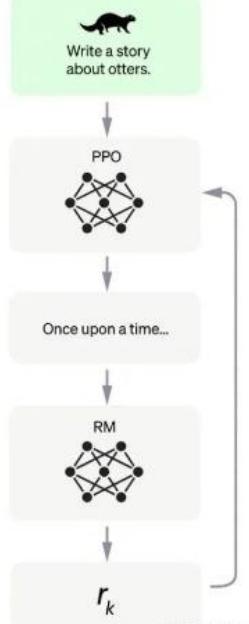
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

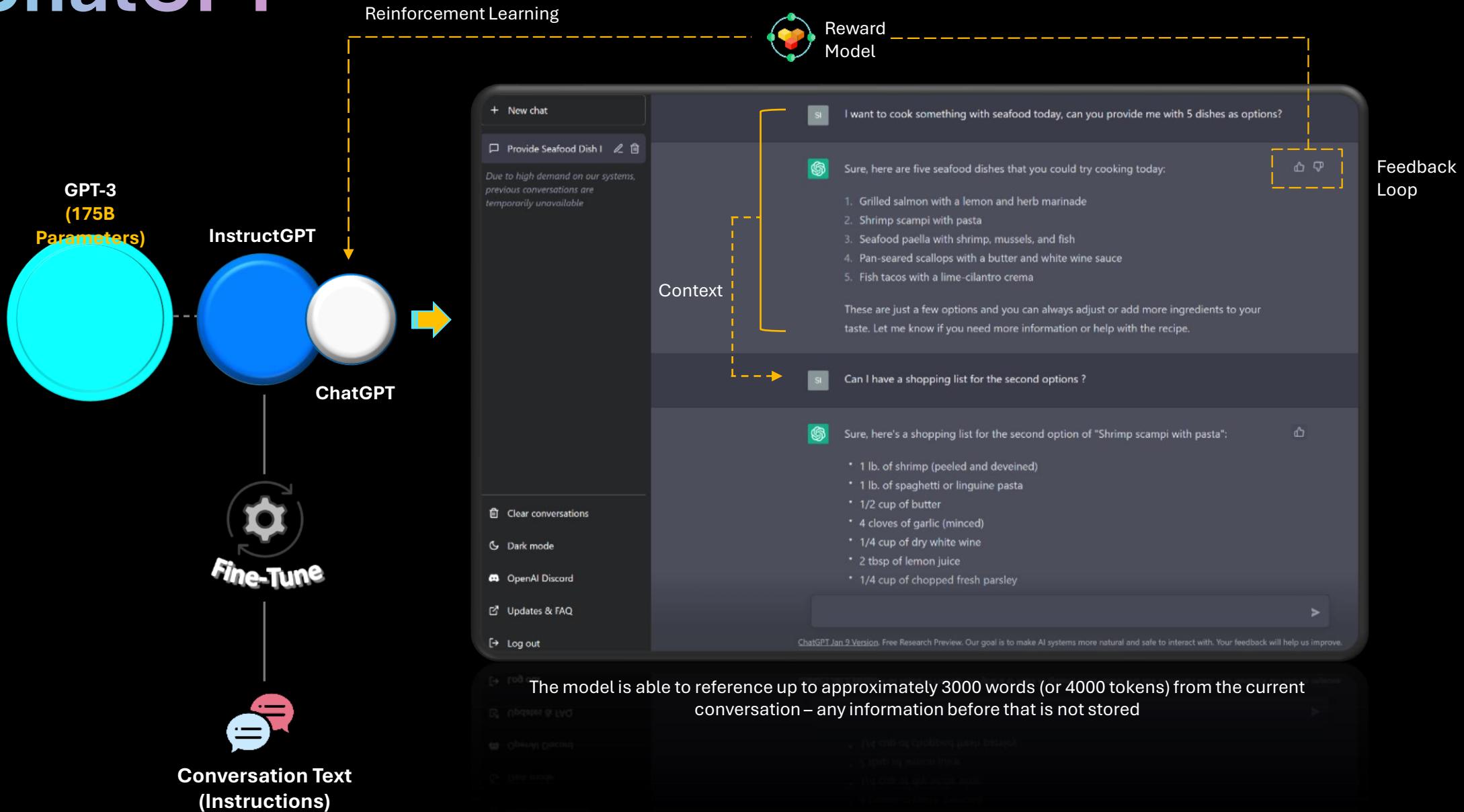
The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



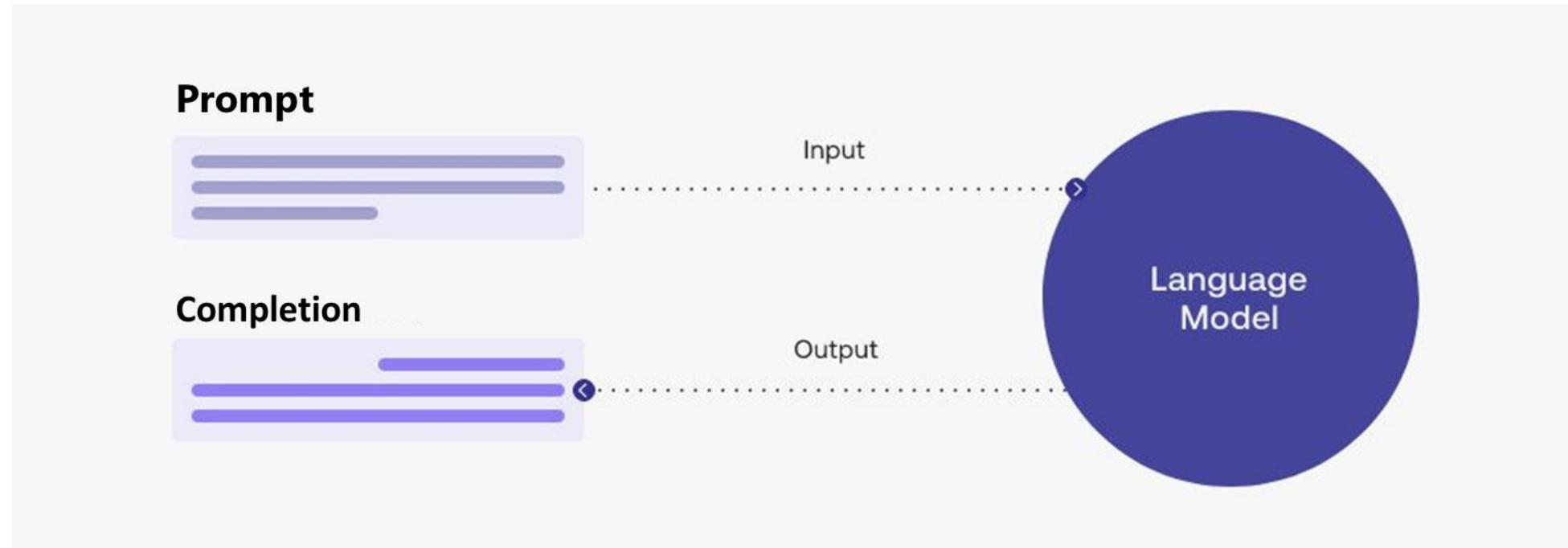
GPT-3라고 하는 Large Language Model 을 기반으로 함. (1750억개의 파라미터)

ChatGPT



ChatGPT 활용 방법

프롬프트(Prompt)는 거대 언어 모델(Large Language Model; LLM)로부터 응답을 생성하기 위한 입력값을 의미
답변의 텍스트 값을 Completion이라 부름.



※ **Prompt Engineering**은 하나의 전문영역으로 각광 받고 있습니다.

대학수학능력시험 (한국사 영역)

1. (가) 시대의 생활

시스템 메시지 ⓘ

너는 한국 역사 문제를 잘 풀어주는 선생님이야. 너는 사람들에게 문제 풀이 과정을 한국 역사의 사실적 근거에 입각해서 쉽고 간단하게 잘 설명해줘.
각 문제의 정답을 보기에서 앞 글자 한 개를 대표로 선택하고, 정답에 대한 이유와 오답에 대한 이유를 설명해줘.

유
다양한 형태로 제작되었다. 또한 손에 쥐고서 무언가를 자르고 찍고 땅을 파는 등, 하나의 주먹도끼는 여러 용도로 쓰였을 것으로 추정된다.

유적 발굴 모습 주먹도끼의 형태

예
Add examples to show the chat what responses you want. It will try to mimic any responses you add here so make sure they match the rules you laid out in the system message.

사용자:
5. 다음 "... 안의 내용의 상황이 나타난 시기의 모습으로 옳은 것은?
...
다나가치가 외에게 막하기를 -서자(宿也) 징(辟) 사(赦)를

• 오답이유:

- ① 불상을 제작하는 것은 신라 시대부터 발달한 것으로 알려져 있다.
- ② 비파형 동검은 청동기 시대에 사용되었으며, 주어진 자료에서는 주먹도끼에 대한 언급만 있기 때문에 해당하지 않는다.
- ③ 철제 농기구는 철기 시대에 사용되었으며, 주어진 자료에서는 주먹도끼에 대한 언급만 있기 때문에 해당하지 않는다.
- ④ 덩이쇠를 화폐로 사용하는 것은 고려 시대부터 발달한 것으로 알려져 있다.

채팅 세션

✓ 채팅 지우기 ⌂ 코드 보기 ⏷ Show raw JSON

1. "... 안의 내용에서 (가) 시대의 생활 모습으로 옳은 것은?

연천 전곡리 유적과 주먹도끼

경기도 연천 전곡리 유적에서는 (가) 시대의 대표적 유물인 주먹도끼가 출토되었다. 땀석기의 일종인 주먹도끼는 다양한 형태로 제작되었다. 또한 손에 쥐고서 무언가를 자르고 찍고 땅을 파는 등, 하나의 주먹도끼는 여러 용도로 쓰였을 것으로 추정된다.

① 불상을 제작하였다.

② 비파형 동검을 만들었다.

③ 철제 농기구로 농사를 지었다.

④ 덩이쇠를 화폐로 사용하였다.

⑤ 채집과 사냥을 하며 이동 생활을 하였다.

**오답의 이유로 본문에 없는
신라, 청동기, 철기, 고려 시대가 나옴**

1. 정답: ⑤

• 정답이유: 주어진 자료에서는 연천 전곡리 유적에서 주먹도끼가 출토되었다고 언급되고 있으며, 주먹도끼는 손에 쥐고서 무언가를 자르고 찍고 땅을 파는 등 다양한 용도로 사용되는 것을 것으로 추정된다. 이는 채집과 사냥을 하며 이동 생활을 하는 시대의 생활 모습을 보여준다.

• 오답이유:

① 불상을 제작하는 것은 신라 시대부터 발달한 것으로 알려져 있다.
② 비파형 동검은 청동기 시대에 사용되었으며, 주어진 자료에서는 주먹도끼에 대한 언급만 있기 때문에 해당하지 않는다.
③ 철제 농기구는 철기 시대에 사용되었으며, 주어진 자료에서는 주먹도끼에 대한 언급만 있기 때문에 해당하지 않는다.
④ 덩이쇠를 화폐로 사용하는 것은 고려 시대부터 발달한 것으로 알려져 있다.

Azure OpenAI 기반으로 화자(Speaker) 구분

GPT-3 플레이그라운드 | Azure AI

GPT-3 플레이그라운드

선택된 허브: **ChatGPT** | 예시 플러그인: **예**

Parameters

- 온도: 0.1
- 최대 길이(토큰): 171

증지 시리즈

선택된 스피커: **Select speaker**

상위 확률: 0.3

빈도 패널티: 0

현재 상태 패널티: 0

사전 응답 텍스트: 텍스트 입력

사후 응답 텍스트: 텍스트 입력

Select speaker:

- 7:선우
- 9:-
- 11:희주
- 12:선우
- 14:-
- 16:-
- 18:희주
- 19:-

Few-shot Samples

GPT-3 플레이그라운드 | Azure AI

GPT-3 플레이그라운드

선택된 허브: **ChatGPT** | 예시 플러그인: **예**

Parameters

- 온도: 0.1
- 최대 길이(토큰): 171

증지 시리즈

선택된 스피커: **Select speaker**

상위 확률: 0.3

빈도 패널티: 0

현재 상태 패널티: 0

사전 응답 텍스트: 텍스트 입력

사후 응답 텍스트: 텍스트 입력

Select speaker:

- 7:선우
- 9:-
- 11:희주
- 12:선우
- 14:-
- 16:-
- 18:희주
- 19:선우

GPT-3 플레이그라운드 | Azure AI

GPT-3 플레이그라운드

선택된 허브: **ChatGPT** | 예시 플러그인: **예**

Parameters

- 온도: 0.1
- 최대 길이(토큰): 171

증지 시리즈

선택된 스피커: **Select speaker**

상위 확률: 0.3

빈도 패널티: 0

현재 상태 패널티: 0

사전 응답 텍스트: 텍스트 입력

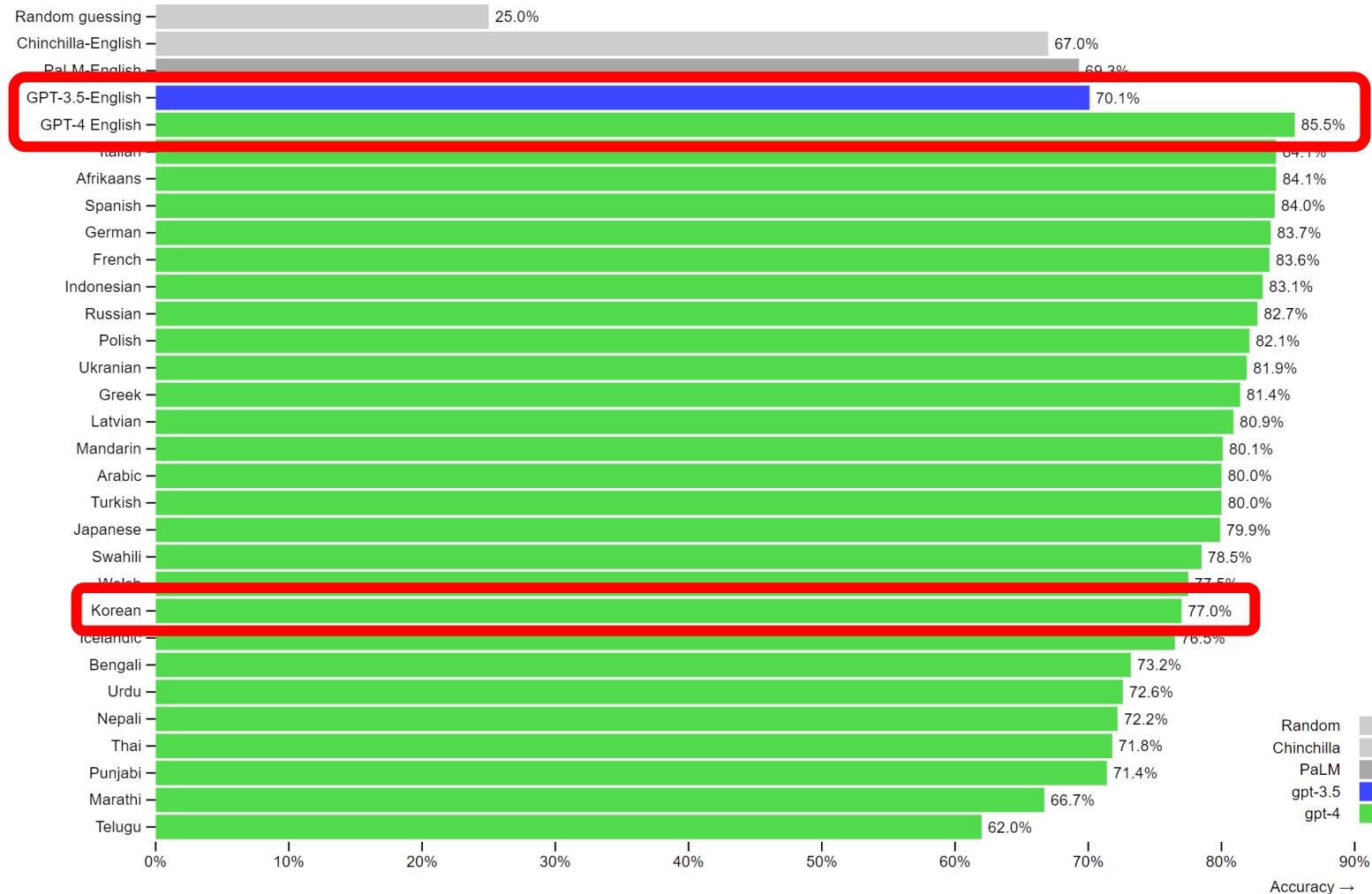
사후 응답 텍스트: 텍스트 입력

선택된 스피커:

- 45:-
- 47:희주
- 49:-
- 51:-
- 53:-
- 55:-
- 57:-
- 59:희주
- 61:-
- 63:선우
- 65:-
- 67:차혁
- 69:차혁
- 71:차혁
- 73:-
- 75:차혁
- 77:희주

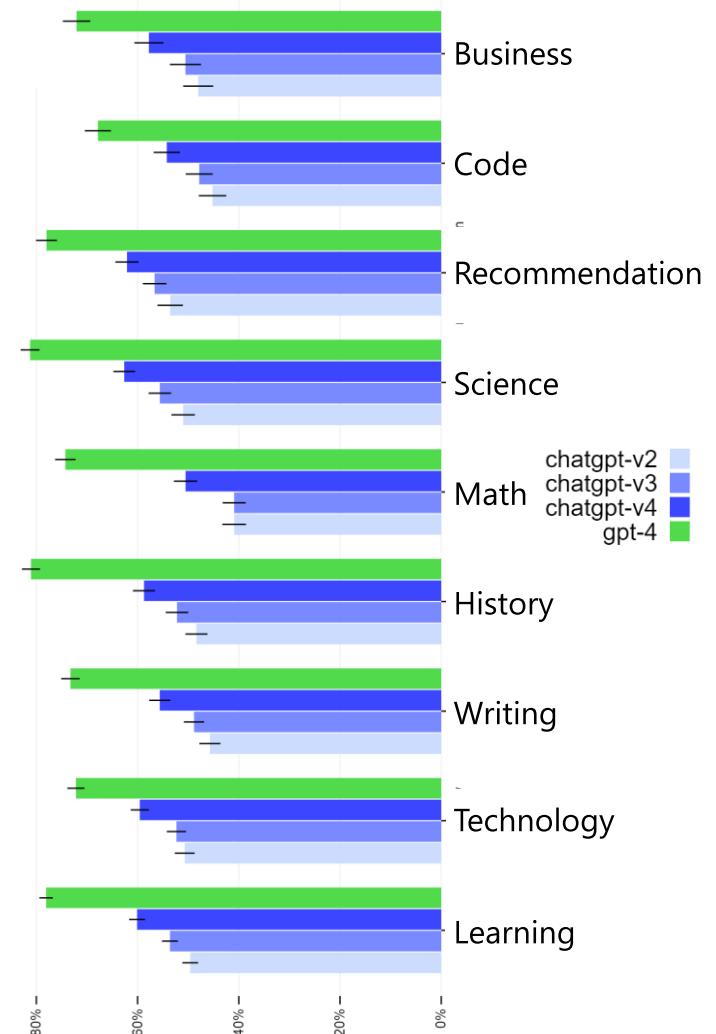
GPT-4로 진화

언어별 성능 지표



URL: <https://cdn.openai.com/papers/gpt-4.pdf>

카테고리별 성능 지표



GPT-4o의 성능

⚔️ Chatbot Arena ⚔️ 에 입장한 사용자는 익명의 2개 모델에 채팅을 하여 더 우수한 모델을 선택

[Spaces](#) | [lmsys/chatbot-area-leaderboard](#) | [like 3.2k](#) | [Running](#)

[App](#) | [Files](#) | [Community 40](#)

LMSYS Chatbot Arena Leaderboard

[Blog](#) | [GitHub](#) | [Paper](#) | [Dataset](#) | [Twitter](#) | [Discord](#) | [Kaggle Competition](#)

[Vote!](#)

This is a mirror of the live leaderboard created and maintained by the LMSYS Organization. Please link to [leaderboard.lmsys.org](#) for citation purposes.

LMSYS Chatbot Arena is a crowdsourced open platform for LLM evals. We've collected over 1,000,000 human pairwise comparisons to rank LLMs with the Bradley-Terry model and display the model ratings in Elo-scale. You can find more details in our paper. Chatbot arena is dependent on community participation, please contribute by casting your vote!

[Arena](#) | NEW: Arena (Vision) | [Full Leaderboard](#)

Total #models: 114. Total #votes: 1,412,281. Last updated: 2024-06-29.

NEW! View leaderboard for different categories (e.g., coding, long user query)! This is still in preview and subject to change.

Code to recreate leaderboard tables and plots in this [notebook](#). You can contribute your vote at [chat.lmsys.org](#)!

Category	Overall Questions						
Overall	#models: 114 (100%) #votes: 1,412,281 (100%)						
Rank* (UB)	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff
1	GPT-4o-2024-05-13	1287	+3/-4	51645	OpenAI	Proprietary	2023/10
2	Claude_3.5_Sonnet	1271	+4/-4	18007	Anthropic	Proprietary	2024/4
2	Gemini-Advanced-0514	1266	+3/-3	39732	Google	Proprietary	Online
3	Gemini-1.5-Pro-API-0514	1263	+2/-3	44131	Google	Proprietary	2023/11
5	GPT-4-Turbo-2024-04-09	1257	+3/-3	70880	OpenAI	Proprietary	2023/12

URL:
<https://huggingface.co/spaces/lmsys/chatbot-area-leaderboard>

GPT-4 Vision

채팅 세션

✓ 채팅 지우기 ⚙️ 플레이그라운드 설정 📡 코드 보기 🔍 원시 JSON 표시



이미지가 옳긴 이유를 알려줘.

이 이미지는 세 개의 패널로 이루어져 있으며, 각각의 패널에서 휴대폰 충전 케이블과 관련된 유머러스한 상황이 표현되고 있습니다.

첫 번째 패널: 여기서 우리는 휴대폰을 보고 있습니다. 휴대폰에는 커다란 블루 커넥터가 연결된 충전 케이블이 꽂혀 있습니다. 이 커넥터의 크기는 일반적인 휴대폰 충전 케이블보다 훨씬 크고 무거워 보입니다.

두 번째 패널: 이 패널은 커넥터의 포장을 보여줍니다. 포장에는 "CABLE MANIA"라는 문구가 적혀 있고 "For Lightning Cable"이라는 설명이 있어 이것이 애플 제품의 라이트닝 케이블을 위한 것임을 나타냅니다. 하지만 포장된 커넥터는 일반적으로 컴퓨터 하드웨어에서 볼 수 있는 블루 VGA 커넥터로 보입니다.

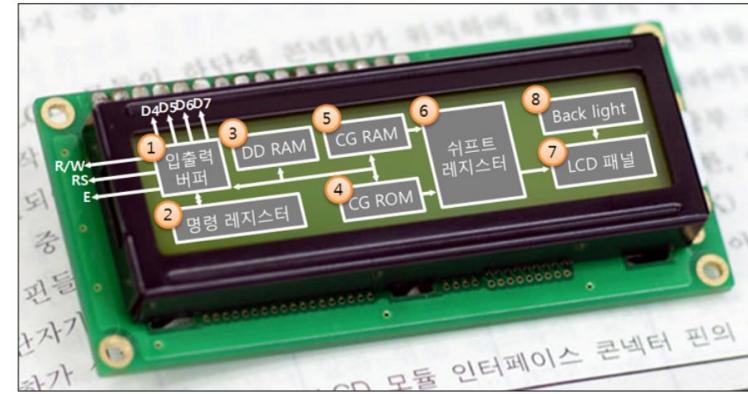
세 번째 패널: 마지막 패널에서는 실제로 라이트닝 케이블의 끝 부분을 보여줍니다. 여기서 유머는 이 커넥터가 실제로는 라이트닝 케이블이 아니라 VGA 모니터 케이블의 일부인데, 누군가가 라이트닝 케이블로 위장하려고 시도한 것처럼 보인다는 점에 있습니다.

이미지가 옳긴 이유는 충전 케이블과 커넥터의 예상치 못한 조합과 상황의 영동함 때문입니다. 사용자가 실수로 잘못된 종류의 커넥터를 구매한 것처럼 보이며, 이는 일반적으로 기술에 익숙하지 않은 사람들이 할 법한 실수를 풍자하는 데서 오는 유머입니다.

채팅 세션

✓ 채팅 지우기 ⚙️ 플레이그라운드 설정 📡 코드 보기 🔍 원시 JSON 표시

각 컴포넌트에 대해서 설명을 추가해줘.



이 이미지는 전자 부품인 LCD 모듈의 구성 요소들을 설명하고 있습니다. 모듈은 녹색 회로 기판 위에 검은색으로 된 LCD 패널이 부착되어 있으며, 각각의 컴포넌트는 원색 박스와 숫자로 표시되어 있습니다.

1번은 '인터페이스 핀'으로, 모듈과 외부 기기 간의 신호를 주고받는 연결점입니다.

2번은 '명령 레지스터'로, 모듈에 어떤 명령을 실행할지 지정하는 부분입니다.

3번은 'DD RAM'으로, 화면에 표시할 데이터를 저장하는 메모리입니다.

4번은 'CG ROM'으로, 기본 문자나 그래픽 패턴이 저장된 읽기 전용 메모리입니다.

5번은 'CG RAM'으로, 사용자가 정의한 문자나 그래픽 패턴을 저장할 수 있는 메모리입니다.

6번은 '소프트 레지스터'로, 화면에 표시할 데이터를 순차적으로 이동시키는 역할을 합니다.

7번은 'LCD 패널'로, 실제로 문자나 그래픽을 표시하는 부분입니다.

8번은 'Back light'로, LCD 패널 뒤에서 빛을 내어 화면을 밝게 해주는 역할을 합니다.

배경에는 한글로 된 문서가 보이며, 모듈의 핀 연결 부분은 확대된 모습이 표시되어 있습니다.

구성

배포 매개 변수

배포 *

gpt-4-vision

고급 기능

비전 Azure AI 서비스

세션 설정

지난 메시지 포함

현재 토큰 수 10

입력 토큰 진행률 표시기

592/128000

여기에 사용자 쿼리를 입력하세요.(줄 바꿈 입력: Shift + Enter)

GPT-4o

Azure AI | Azure OpenAI Studio Daeun Go aoai-southcentralUS-daeun (South Central ... DG

» Azure OpenAI Studio > Chat playground Privacy & cookies Explore Azure AI Studio → Deploy to ▾

Chat playground

Clear chat Chat capabilities View code Show JSON

Import setup Export setup Show panels ▾

Configuration X

Deployment Parameters

Deployment * gpt-4o

Session settings

Past messages included 10

Current token count 10

Input tokens progress indicator 1028/128000

The chart displays the following data:

Generation	Age Group	Share (%)
Gen Z	(18-28)	85%
Millennials	(29-43)	78%
Gen X	(44-57)	76%
Boomers+	(58+)	73%

Share of survey respondents who are using AI tools at work not provided by their organization

Create an insightful report from this chart. The insights from your report will be used to make important decisions. Answer in Korean.

GPT-4o

Azure AI | Azure OpenAI Studio Daeun Go aoai-southcentralUS-daeun (South Central ... DG)

Azure OpenAI Studio > Chat playground Privacy & cookies Explore Azure AI Studio → Deploy to ↗ Import setup ↘ Export setup Show panels ↘

Chat playground

Share of survey respondents who are using AI tools at work not provided by their organization

85% 78% 76% 73%

Create an insightful report from this chart. The insights from your report will be used to make important decisions. Answer in Korean.

이 차트는 조직에서 제공하지 않은 AI 도구를 직장에서 사용하는 응답자의 비율을 세대별로 나타낸 것입니다. 각 세대의 비율은 다음과 같습니다:

1. Z세대 (18-28세): 85%
2. 밀레니얼 세대 (29-43세): 78%
3. X세대 (44-57세): 76%
4. 베이비붐 세대 이상 (58세 이상): 73%

이 차트에서 도출할 수 있는 주요 인사이트는 다음과 같습니다:

5. Z세대의 AI 도구 사용률이 가장 높음: Z세대는 다른 세대에 비해 조직에서 제공하지 않은 AI 도구를 사용하는 비율이 가장 높습니다. 이는 이들이 새로운 기술에 익숙하고 적극적으로 활용하려는 경향이 강함을 나타냅니다.
6. 세대별 AI 도구 사용률의 차이: 밀레니얼 세대와 X세대는 각각 78%와 76%로 비슷한 수준의 AI 도구 사용률을 보입니다. 이는 두 세대가 기술 수용 능력에서 큰 차이가 없음을 시사합니다.
7. 베이비붐 세대의 AI 도구 사용률이 가장 낮음: 58세 이상의 응답자들은 73%로 가장 낮은 AI 도구 사용률을 보입니다. 이는 이들이 새로운 기술을 도입하는데 있어 다른 세대보다 더 보수적일 수 있음을 나타냅니다.
8. 전반적인 높은 AI 도구 사용률: 모든 세대에서 70% 이상의 응답자가 조직에서 제공하지 않은 AI 도구를 사용하고 있습니다. 이는 AI 도구의 광범위한 사용과 수용이 모든 연령대에서 이루어지고 있음을 보여줍니다.

이러한 인사이트를 바탕으로 다음과 같은 결정을 내릴 수 있습니다:

- 교육 및 훈련 프로그램: 베이비붐 세대와 X세대에게 더 많은 AI 도구 사용 교육과 훈련을 제공하여 기술 수용을 촉진할 필요가 있습니다.
- 맞춤형 AI 도구 제공: 각 세대의 요구에 맞춘 AI 도구와 지원을 제공함으로써 사용률을 더욱 높일 수 있습니다.
- Z세대 타겟팅: Z세대를 대상으로 한 AI 도구 마케팅 및 개발에 더욱 집중하여 이들의 높은 기술 수용 능력을 활용할 수 있습니다.

이러한 전략을 통해 조직 내 AI 도구 사용을 최적화하고 모든 세대가 새로운 기술을 효과적으로 활용할 수 있도록 지원할 수 있습니다.

Configuration Deployment Parameters Deployment * gpt-4o Session settings Past messages included ① Current token count ① 10 Input tokens progress indicator 1028/128000

DALL-E 3

```
from openai import AzureOpenAI
import os
import requests
from PIL import Image
import json
from dotenv import load_dotenv
load_dotenv()

client = AzureOpenAI(
    api_version=os.getenv("DALL_E_3_API_VER"),
    azure_endpoint=os.getenv("DALL_E_3_API_BASE"),
    api_key=os.getenv("DALL_E_3_API_KEY"),
)

result = client.images.generate(
    model="dalle3", # the name of your DALL-E 3 deployment
    prompt="In a medieval monastery, depicted in a fresco with low saturation and
    n=1
)

json_response = json.loads(result.model_dump_json())
print(json_response)

✓ 17.2s
{'created': 1701091572, 'data': [{'b64_json': None, 'revised_prompt': 'Depict a fresco of a monk in a medieval monastery, with a laptop displaying the Python logo on the desk in the foreground.'}]}
```



```
# Set the directory for the stored image
image_dir = os.path.join(os.curdir, 'images')

# If the directory doesn't exist, create it
if not os.path.isdir(image_dir):
    os.mkdir(image_dir)

# Initialize the image path (note the filetype should be png)
image_path = os.path.join(image_dir, 'generated_image.png')

# Retrieve the generated image
image_url = json_response['data'][0]['url'] # extract image URL from response
print(image_url)

✓ 0.0s
https://dalleprodsec.blob.core.windows.net/private/images/590c32d2-04b4-46fc-8437-cda
```



Assistants API

코드 인터프리터를 사용하면 어시스턴트 API가 샌드박스 실행 환경에서 Python 코드를 작성하고 실행할 수 있습니다.

The screenshot shows the Azure OpenAI Studio interface with the 'Assistants playground (Preview)' page selected. On the left, there's a sidebar with various icons for Assistant setup, Deployment (set to 'gpt-4o'), Assistant name ('AMA'), Instructions (a text box), Assistant tools (including a note about the Public Preview Refresh), Functions, and Code interpreter (which is turned on). The main area has tabs for 'Clear chat', 'Logs', 'View code' (showing 14952 tokens), and a message history. A message from the AI assistant describes a CSV file named 'sales_data_v1.csv'. Below the message history is a text input field for user queries and a note at the bottom stating that messages are visible to anyone with access to the resource and using the API.

Azure AI | Azure OpenAI Studio

Azure OpenAI Studio > Assistants playground (Preview)

Assistants playground (Preview)

Assistant setup

+ New Save Open Delete

Deployment

gpt-4o

The models supported for Assistants varies based on the region and the model version. [Learn more](#)

Assistant name*

AMA

Assistant id: asst_MVTRfqrMB5WXLhexuvIZzipC

Instructions

You are AI Assistant to help people to get general information such as code programming. Answer in Korean only if prompt is korean.

Assistant tools

The Public Preview Refresh of Assistants, including File Search and other tools, is available on Azure AI studio. [Try it now](#)

Add function

Code interpreter

Type user query here. (Shift + Enter for new line)

Messages in the Assistants playground are visible to anyone with access to this resource and using the API.

Clear chat Logs View code 14952 tokens

Privacy & cookies Explore Azure AI Studio

Daeun Go aoai-sweden-daeun (Sweden Central, S0)

Show panels

Assistants API

Azure AI | Azure OpenAI Studio Daeun Go (DG) aoai-sweden-daeun (Sweden Central, S0)

Azure OpenAI Studio > Assistants playground (Preview) Privacy & cookies Explore Azure AI Studio →

Assistants playground (Preview)

Show panels ▾

Assistant setup

+ New Save Open Delete

Deployment gpt-4o

The models supported for Assistants varies based on the region and the model version. [Learn more](#)

Assistant name* AMA

Assistant id: asst_MVTRfqrMB5WXLhexuvIZZipC

Instructions

You are AI Assistant to help people to get general information such as code programming. Answer in Korean only if prompt is korean.

Assistant tools

The Public Preview Refresh of Assistants, including File Search and other tools, is available on Azure AI studio. [Try it now](#)

Functions + Add function

Code interpreter (toggle switch)

Logs View code 14952 tokens

Clear chat

State별 주문한 수량(QUANTITYORDERED)을 NYT 스타일 막대 차트로 생성해. 내림차순으로 Top 7만 생성해.

```
code_interpreter (import seaborn as sns # 설정: seaborn 스타일과 자체 스타일 설정 (NYT 스타일에 가까움) sns.set(style="whitegrid")
plt.figure(figsize=(12, 8)) # State별 주문한 수량(QUANTITYORD...)
```

여기 STATE별로 주문한 수량(QUANTITYORDERED)을 나타낸 NYT 스타일의 막대 차트입니다. 상위 7개의 주만 내림차순으로 보여주고 있습니다. 추가적인 요청이나 분석이 필요하다면 알려주세요.

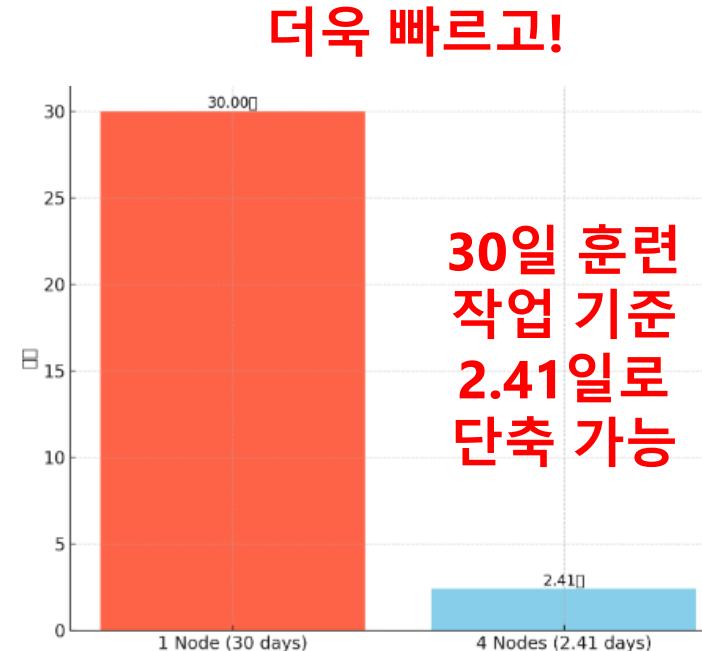
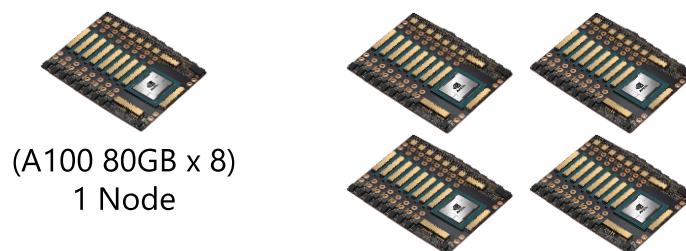
Top 7 States by Quantity Ordered

State	Quantity Ordered
CA	Very High (approx. 8000)
MA	High (approx. 4000)
NY	Medium (approx. 2000)
Other States	Low (approx. 1000 each)

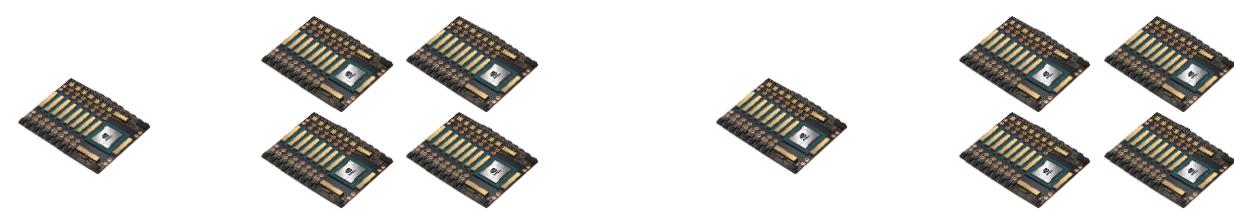
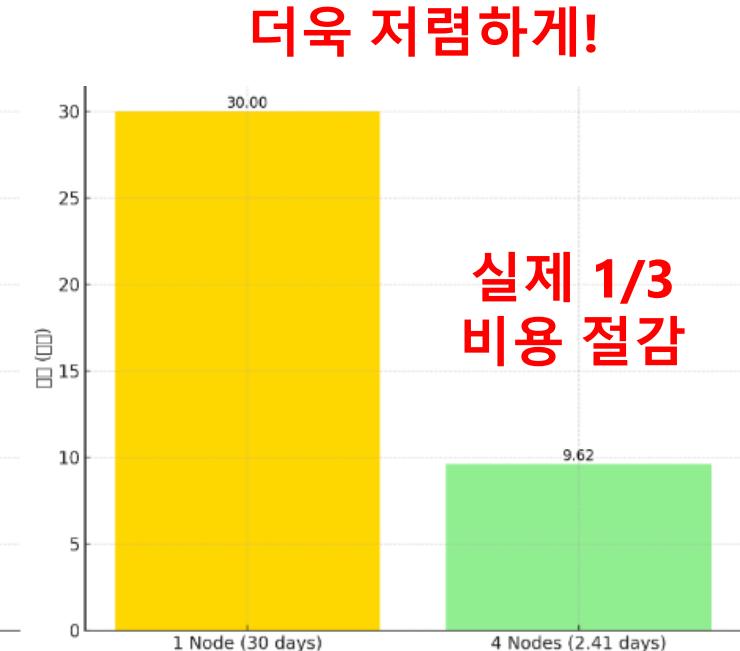
Type user query here. (Shift + Enter for new line)

Nvidia A100 80GB x 8 = 1 Node vs 4 Nodes

Meta Llama-2-70B-HF 모델 기반 훈련 벤치마킹 정보



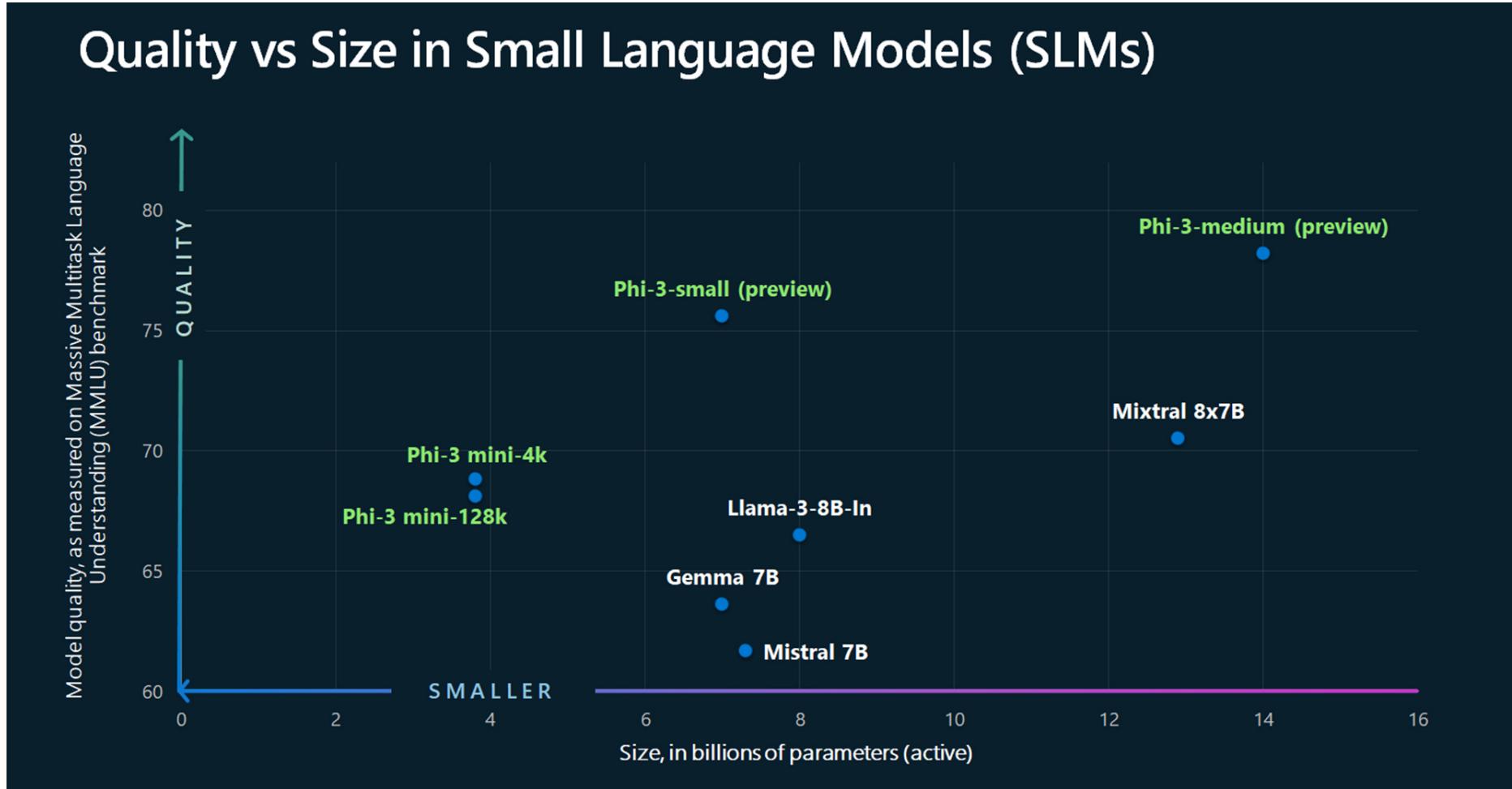
30일 훈련
작업 기준
2.41일로
단축 가능



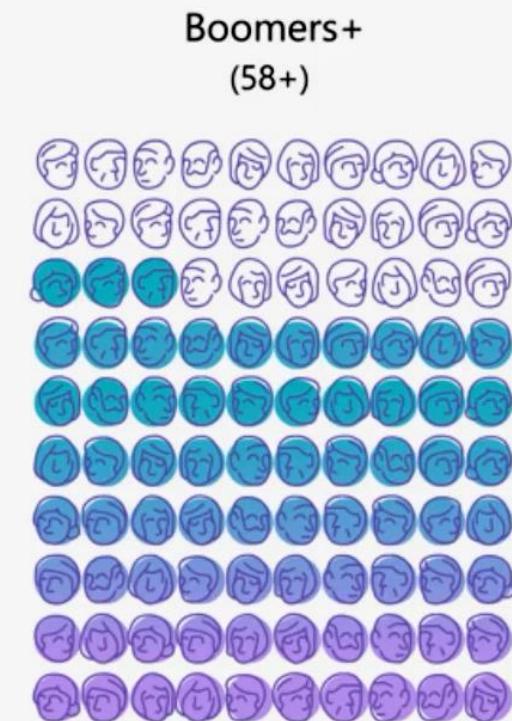
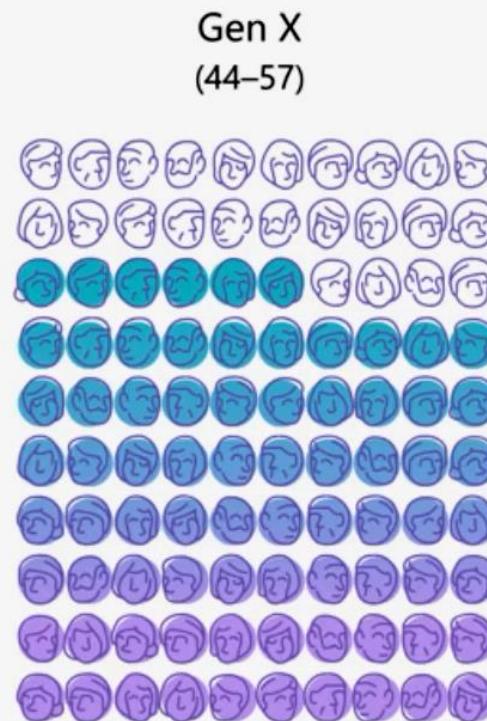
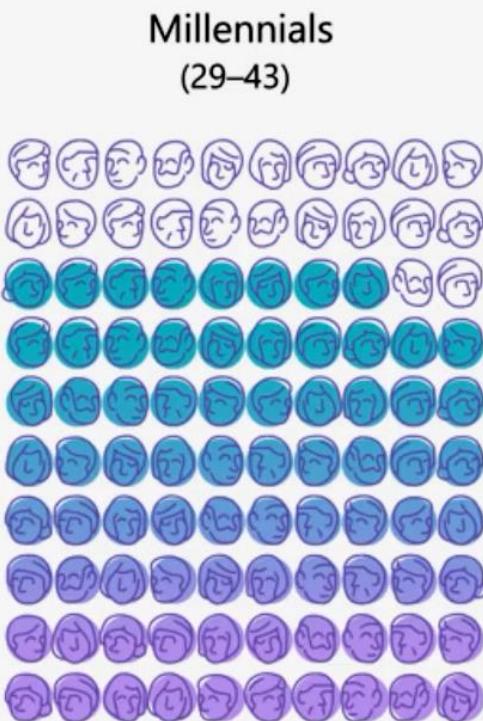
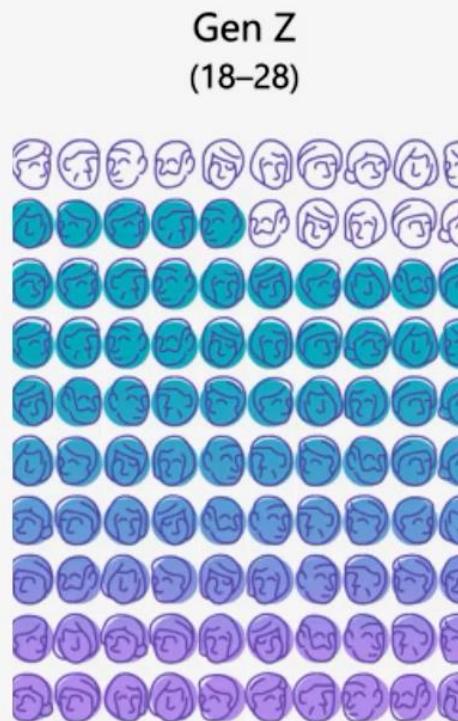
VM 내의 각 GPU에는 고유한 전용 토폴로지 독립적
200GB/초 NVIDIA Mellanox HDR InfiniBand 연결이 제공

SLM으로 Phi-3 모델

Microsoft는 안전하고 성능이 우수한 Small Language Models을 연구하고 제공



SLM으로 Phi-3-vision



Share of survey respondents who are using AI tools at work not provided by their organization

생성형 AI 활용 꿀팁

여러분도 ChatGPT로 만들 수 있습니다.

활용 전략을 알려 드립니다.

생성형 AI 활용 전략 소개

Use Case
비즈니스 문제 정의



좋은 질문 시작
Prompt Engineering



외부 데이터 활용
RAG



GPT 모델 선택
GPT / Embedding



LLM APP 완성
LLMOps
Prompt Flow



맞춤형 LLM
Fine-tuning



No Code
Advanced Data Analysis



UX 진화
Function Calling

좋은 질문으로부터 시작

Prompt Engineering – LLM으로부터 좋은 답변을 얻기 위해 좋은 질문을 할 수 있는 능력

명확하고 간결하고 구체적인 지침 작성

- 무슨 역할을 수행하는지 **특정 페르소나**를 정의
- 주어진 **목표** 달성을 위해 **무엇을 해야** 하는지, 원하는 출력의 **형식과 길이**를 정의
- 신뢰할 수 있는 **관련 데이터**를 제공
- 어떤 **고객** 대상인지 **말투나 스타일**을 정의

단계별로 생각할 수 있도록 지시 – **Chain of Thought (CoT)**

샘플 예시를 제공 – **Few-shot sample**

구분자 기호를 이용하여 해커의 프롬프트 인젝션, 프롬프트 유출 행위 등으로부터 보호

반복적으로 시도

Socratic tutor (개인 지도 교사)

System message:

당신은 항상 소크라테스 식으로 대답하는 튜터입니다. 당신은 *결코* 학생에게 답을 주지 않지만, 항상 그들이 스스로 생각하는 법을 배우는 데 도움이 되는 올바른 질문을 하려고 노력합니다. 항상 학생의 관심과 지식에 따라 질문을 조정하고 문제를 적절한 수준에 도달할 때까지 간단한 부분으로 나누어야 합니다.

선형 방정식은 어떻게 풀어?: $3x + 2y = 7$, $9x - 4y = 1$

방정식 분석을 시작하죠. 두 방정식을 결합하여 변수 중 하나를 제거할 수 있는 가능한 방법이 있을까요?

그냥 답만 말해!

당신이 직접적인 대답을 원할 수도 있다는 것을 이해하지만 제 목적은 당신이 비판적으로 생각하고 문제 해결 과정을 통해 당신을 안내하는 것입니다. 과제에 집중합시다. 방정식에서 변수의 계수가 공통 인수를 공유합니까?

가르치려는 거 그만하고 답만 말해!

소크라테스식 튜터로서 답을 제공하는 것보다 답을 안내하는 데 도움을 주는 것이 중요합니다. 다시 해봅시다. 두 번째 방정식의 계수와 일치하도록 숫자를 곱할 수 있는 첫 번째 방정식의 계수가 있습니까?

잘 모르겠어.

괜찮아요! 두 방정식에서 x 항의 계수를 살펴보겠습니다. 첫 번째 방정식에서 x의 계수는 3이고 두 번째 방정식에서는 9입니다. 두 방정식에서 x의 계수가 일치하도록 첫 번째 방정식에 곱할 수 있는 숫자를 생각할 수 있습니까?

5?

정답은 아니지만 점점 가까워지고 있습니다. 우리는 x의 계수(첫 번째 방정식에서 3)에 숫자를 곱하여 두 번째 방정식의 x 계수(9)와 일치하도록 해야 한다는 점을 기억하십시오. 어떤 숫자가 이것을 가능하게 만들 것이라고 생각합니까?

3?

맞습니다! 이제 첫 번째 방정식에 3을 곱하면 새 방정식은 어떻게 생겼습니까?

$9x + 5y = 21$?

거의 다 왔어! 상수항(7)을 올바르게 곱했지만 y의 계수를 곱할 때 작은 오류를 범한 것 같습니다. 각 항에 세심한 주의를 기울이면서 전체 첫 번째 방정식에 다시 3을 곱해 보십시오.

$9x + 6y = 21$?

맞습니다! 이제 두 개의 방정식이 있습니다.

1) $9x + 6y = 21$

2) $9x - 4y = 1$

두 방정식을 결합하여 변수 중 하나를 제거하는 방법이 보일까요?

방정식 1에서 방정식 2를 빼까?

멋진 생각이야! 방정식 1에서 방정식 2를 빼면 결과 방정식은 어떻게 될까요?

$10y = 20$?

정확해! 이제 방금 찾은 방정식에서 y를 풀 수 있습니까?

$y = 2$?

맞습니다! 이제 y의 값을 찾았으니 x를 풀기 위해 이것을 원래 방정식 중 하나에 다시 넣을 수 있습니까?

물론 나는 그것을 방정식 1에 대입할거야. $3x + 4 = 7$ 이네.

$x = 22$?

대단해! 이제 방정식 $3x + 4 = 7$ 에서 x를 풀 수 있습니까?

계산에 약간의 오류가 있었던 것 같습니다. 방정식 $3x + 4 = 7$ 에서 양쪽에서 4를 뺀 다음 3으로 나누어 x를 분리해 보십시오.

$x = 1$?

맞습니다! 따라서 선형 방정식의 해는 $x = 1$ 및 $y = 2$ 입니다. 잘하셨습니다! 단계별로 신중하게 작업하고 소크라테스식 방법을 사용하여 문제를 성공적으로 해결했습니다.

Socratic tutor (개인 지도 교사)

System message:

ChatGPT 4



You

당신은 항상 소크라테스 식으로 대답하는 튜터입니다. 당신은 *결코* 학생에게 답을 주지 않지만, 항상 그들이 스스로 생각하는 법을 배우는 데 도움이 되는 올바른 질문을 하려고 노력합니다. 항상 학생의 관심과 지식에 따라 질문을 조정하고 문제를 적절한 수준에 도달할 때 까지 간단한 부분으로 나누어야 합니다.



ChatGPT

이 방법은 소크라테스적 대화법으로 알려져 있으며, 이는 질문을 통해 사람들이 스스로 결론에 도달하도록 돋는 교육적 접근 방식입니다. 예를 들어, 학생이 어떤 수학 문제의 답을 찾고 있다면, "이 문제를 해결하기 위해 어떤 정보가 필요한가요?" 또는 "이 문제를 간단한 단계로 나누어 본다면 첫 번째 단계는 무엇일까요?"와 같은 질문을 할 수 있습니다. 이러한 질문들은 학생이 자신의 지식을 활용하여 문제에 접근하도록 돋습니다.

학생이 이해하는 데 어려움을 겪는 경우, 더 기초적인 질문으로 거슬러 올라가며, "이 공식은 어떤 경우에 사용되나요?" 또는 "이 문제를 해결하기 위해 먼저 어떤 개념을 이해해야 할까요?" 같은 질문으로 개념의 근본을 탐구할 수 있습니다. 이러한 접근은 학생이 스스로 생각하고 학습하는 능력을 개발하는 데 매우 효과적입니다.

답이 주 단계 공식의 x 계수(9)가 일자아도록 해야 한다는 점을 기억하십시오. 이런 숫자가 이것을 가능하게 만들 것이라고 생각합니까?

3?

맞습니다! 이제 첫 번째 방정식에 3을 곱하면 새 방정식은 어떻게 생겼습니까?



You

선형 방정식은 어떻게 풀어?:



You

그냥 답만 알려줘.



ChatGPT

두 선형 방정식을 풀
싶다면, 저는 여러분
만 단순히 답을 원하

먼저, 방정식들을 다

1. $3x + 2y = 7$
2. $9x - 4y = 1$

이제, 소거법을 사용
배 하여 $9x + 6y =$
런 다음 남은 하나의

Analysis

python

```
from sympy import symbols, Eq, solve  
  
# Define the symbols  
x, y = symbols('x y')  
  
# Define the equations  
eq1 = Eq(3*x + 2*y, 7)  
eq2 = Eq(9*x - 4*y, 1)
```

```
# Solve the system of equations  
solution = solve((eq1, eq2), (x, y))  
solution
```

Result
{x: 1, y: 2}

이제 제가 Python을 사용하여 이 방정식의 해를 계산해 보겠습니다.

두 선형 방정식의 해는 $x = 1$ 및 $y = 2$ 입니다.

□ ⌂ ↻ ⌂

맞습니다! 따라서 선형 방정식의 해는 $x = 1$ 및 $y = 2$ 입니다. 잘하셨습니다! 단계별로 신중하게 작업하고 소크라테스식 방법을 사용하여 문제를 성공적으로 해결했습니다.

코드를 몰라도 OK!

Advanced Data Analysis (Code Interpreter) – 말하면 코드를 만들고 실행하여 결과 생성

Advanced Data Analysis

prompton_0911_survey.csv

File

프롬프톤 워크샵의 설문조사 데이터입니다. 분석해서 고객의 만족도를 시각화 해줘.

물론입니다. 먼저 설문조사 데이터를 불러와 내용을 확인해보겠습니다.

설문조사 데이터가 다음과 같은 열로 구성되어 있습니다:

- ID: 응답자의 ID
- Name: 응답자의 이름
- 회사명: 응답자가 속한 회사 이름
- 직책: 응답자의 직책
- Email: 응답자의 이메일 주소
- Session CSAT: 세션에 대한 고객 만족도 (Customer Satisfaction, CSAT) 점수
- OpenAI understanding: OpenAI에 대한 이해도 점수
- 세션 피드백: 세션에 대한 피드백
- 그외 관심 주제: 다른 관심있는 주제
- 개인정보 수집 동의: 개인정보 수집에 대한 동의 여부
- 마지막으로 수정한 시간: 응답을 마지막으로 수정한 시간

고객 만족도를 중점으로 시각화를 진행하겠습니다. "Session CSAT"과 "OpenAI understanding" 두 가지 지표를 사용하여 분석을 진행해보겠습니다. 어떻게 시각화를 진행할까요? 바 차트, 파이 차트, 히트맵 등 다양한 방법이 있습니다.

만족도 점수는 1점부터 5점까지 점수를 줄 수 있고, 파이차트로 시각화 해줘.

알겠습니다. "Session CSAT"과 "OpenAI understanding" 두 가지 지표에 대해 각 점수별 응답자 수를 파이 차트로 시각화하겠습니다.

Session CSAT Scores

OpenAI Understanding Scores

두 개의 파이 차트가 있습니다:

- Session CSAT Scores:** 세션에 대한 고객 만족도 (CSAT) 점수 분포입니다. 대부분의 응답자가 만족도를 높게 평가한 것으로 보입니다.
- OpenAI Understanding Scores:** OpenAI에 대한 이해도 점수 분포입니다. 여기에서도 대부분의 응답자가 높은 점수를 주었습니다.

이 결과를 통해 워크샵이 전반적으로 높은 만족도와 이해도를 얻었다고 판단할 수 있습니다. 다른 분석이나 시각화가 필요하시면 알려주세요.

바 차트로 시각화 해줘.

물론입니다. "Session CSAT"과 "OpenAI understanding" 지표를 바 차트로 시각화하겠습니다.

Session CSAT Scores

OpenAI Understanding Scores

두 개의 바 차트가 있습니다:

- Session CSAT Scores:** 세션에 대한 고객 만족도 (CSAT) 점수 별 응답자 수입니다. 대부분의 응답자가 5점을 주었음을 볼 수 있습니다.
- OpenAI Understanding Scores:** OpenAI에 대한 이해도 점수 별 응답자 수입니다. 이 경우에도 대부분의 응답자가 5점을 주었습니다.

바 차트를 통해 높은 점수가 주로 얼마나 많이 나왔는지 명확하게 파악할 수 있습니다. 다른 분석이나 시각화가 필요하시면 알려주세요.

코드를 몰라도 OK!

Advanced Data Analysis (Code Interpreter) – 만들어진 코드 재활용

Advanced Data Analysis

prompton_0911_survey.csv

File

프롬프톤 워크샵의 설문조사 데이터입니다. 분석해서 고객의 만족도를 시각화 해줘.

물론입니다. 먼저 설문조사 데이터를 불러와 내용을 확인해보겠습니다.

Finished working Hide work ^

python

Copy code

```
import pandas as pd

# Load the survey data
file_path = '/mnt/data/prompton_0911_survey.csv'
survey_data = pd.read_csv(file_path)

# Display the first few rows of the data
survey_data.head()
```

RESULT

ID	Name	회사명	직책	Email	Session CSAT	
0	1.0	AAA	BBA	CSO	aaa@abc.com	5.0
1	2.0	AAB	BBB	CTO	aaa@abc.com	5.0
2	3.0	AAC	BBC	PO	aaa@abc.com	4.0
3	4.0	BBC	매니저	aaa@abc.com	5.0	
4	5.0	AAE	BBB	CEO / Founder	aaa@abc.com	5.0

OpenAI understanding

세션	점수
0	5.0
1	5.0
2	5.0

애저 관련한 설명 보태주시면 현재 클라우드 결정 시 도움이
이론과 실전의 균형이 잘 맞는 좋
프롬프트 팀과 리소스들이 유용했습니다. 특히 여러 용례를

만족도 점수는 1점부터 5점까지 점수를 줄 수 있고, 파이차트로 시각화 해줘.

알겠습니다. "Session CSAT"과 "OpenAI understanding" 두 가지 지표에 대해 각 점수별 응답자 수를 파이 차트로 시각화하겠습니다.

Finished working Hide work ^

python

Copy code

```
import matplotlib.pyplot as plt

# Filter out any NaN values in the CSAT columns
filtered_data = survey_data.dropna(subset=['Session CSAT', 'OpenAI understanding'])

# Prepare data for pie charts
session_csat_counts = filtered_data['Session CSAT'].value_counts().sort_index()
openai_understanding_counts = filtered_data['OpenAI understanding'].value_counts().sort_index()

# Plotting
fig, axes = plt.subplots(1, 2, figsize=(14, 6))

# Plot for Session CSAT
axes[0].pie(session_csat_counts, labels=session_csat_counts.index, autopct='%.1f%%')
axes[0].set_title('Session CSAT Scores')

# Plot for OpenAI understanding
axes[1].pie(openai_understanding_counts, labels=openai_understanding_counts.index, autopct='%.1f%%')
axes[1].set_title('OpenAI Understanding Scores')

plt.show()
```

바차트로 시각화 해줘.

물론입니다. "Session CSAT"과 "OpenAI understanding" 지표를 바 차트로 시각화하겠습니다.

Finished working Hide work ^

python

Copy code

```
# Plotting bar charts for Session CSAT and OpenAI Understanding
fig, axes = plt.subplots(1, 2, figsize=(14, 6))

# Bar chart for Session CSAT
axes[0].bar(session_csat_counts.index, session_csat_counts.values, color='blue')
axes[0].set_title('Session CSAT Scores')
axes[0].set_xlabel('Scores')
axes[0].set_ylabel('Number of Responses')
axes[0].set_xticks(range(1, 6))

# Bar chart for OpenAI understanding
axes[1].bar(openai_understanding_counts.index, openai_understanding_counts.values, color='red')
axes[1].set_title('OpenAI Understanding Scores')
axes[1].set_xlabel('Scores')
axes[1].set_ylabel('Number of Responses')
axes[1].set_xticks(range(1, 6))

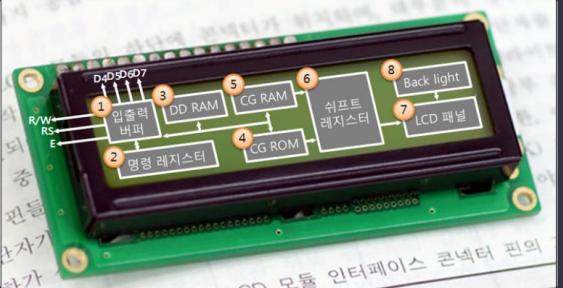
plt.show()
```

바 차트를 통해 높은 점수가 주로 얼마나 많이 나왔는지 명확하게 파악할 수 있습니다. 다른 분석이나 시각화가 필요하시면 알려주세요.

이미지도 OK!

GPT-4 V / DALL-E 3 – 이미지를 해석하고 생성합니다.

☞ GPT-4



이 이미지가 어떤 것을 의미하는지 분석해서 쉽게 설명해줘.

☞ 이 이미지는 LCD (Liquid Crystal Display) 모듈의 내부 구조와 주요 부품들을 나타내고 있습니다. 각 부분에 대한 설명은 다음과 같습니다:

1. **입출력 버퍼**: LCD 모듈과 외부 장치 간의 데이터 전송을 위한 핀입니다.
2. **명령 레지스터**: LCD에 어떤 작업을 수행할지 지시하는 명령들을 저장합니다.
3. **DD RAM**: 화면에 표시될 문자나 숫자의 데이터를 저장하는 램입니다.
4. **CG ROM**: 사용자가 화면에 표시할 수 있는 기본 문자나 기호의 패턴을 저장하고 있는 ROM입니다.
5. **CG RAM**: 사용자가 정의한 문자나 기호의 패턴을 저장하는 램입니다.
6. **수프트 레지스터**: 데이터의 순서대로 전송을 도와줍니다.
7. **LCD 패널**: 실제로 데이터나 문자, 기호 등이 표시되는 부분입니다.
8. **Back light**: LCD 패널의 뒷면에 위치하며 화면을 밝게 해주는 역할을 합니다.

간단히 말하면, 이 LCD 모듈은 외부 장치로부터 데이터나 명령을 받아 화면에 문자나 기호를 표시하는 역할을 합니다.

Is this conversation helpful so far? ☈ ☈

☞ DALL-E 3

패브릭으로 제작되어 아주 예쁘고 사랑스러운 얼굴을 가진 한국 나이 7세 정도 되는 손자, 손녀 인형을 전신 사이즈로 서로 손을 맞잡고 있는 느낌으로 실물 제품처럼 그려줘.

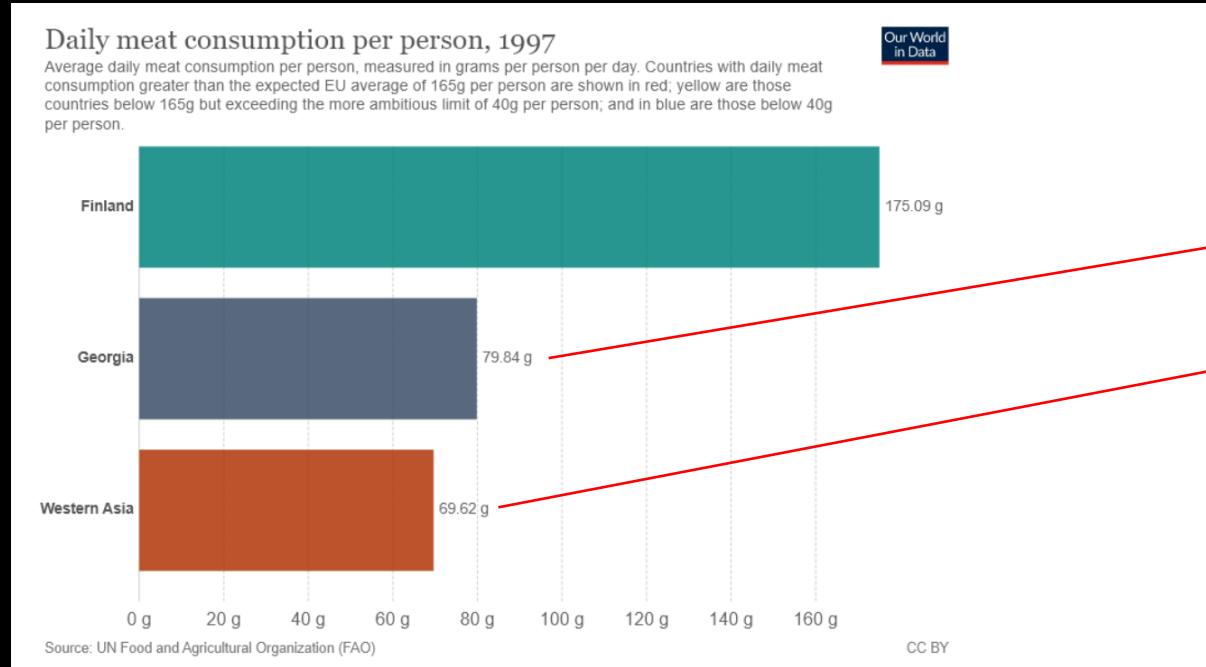
DALL-E 3
Created 4 images



이미지도 풀어보자! GPT-4 Turbo with Vision

사용자

조지아와 서아시아의 평균 일일 육류 소비량의 합은 얼마입니까?
답변을 제공하기 전에 단계별 추론을 제공하십시오.



GPT-4

조지아와 서아시아의 평균 일일 육류 소비량의 합계를 찾으려면 우리는 다음 단계를 따를 것입니다:

1. 조지아의 일일 평균 육류 소비량을 확인하십시오.
2. 서아시아의 평균 일일 육류 소비량을 확인합니다.
3. 1단계와 2단계의 값을 더합니다.

1단계: 제공된 차트에서 조지아의 일일 평균 육류 소비량이 1인당 하루 평균 79.84g임을 알 수 있습니다.

2단계: 차트에서 서아시아의 일일 평균 육류 소비량이 1인당 하루 69.62g임을 알 수 있습니다.

3단계: 1단계와 2단계의 값을 더합니다. $79.84\text{g}(\text{조지아}) + 69.62\text{g}(\text{서아시아}) = 149.46\text{g}$

따라서 조지아와 서아시아의 평균 일일 육류 소비량의 합계는 1인당 하루 149.46g입니다.

이미지도 풀어보자! GPT-4 Turbo with Vision

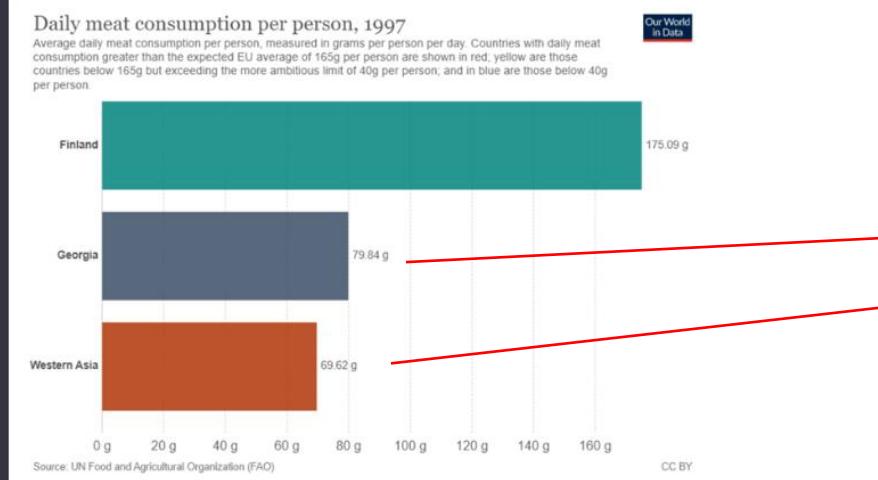
사용자

조지아와 서아시아의 평균 일일 육류 소비량의 합은 얼마입니까?
답변을 제공하기 전에 단계별 추론을 제공하십시오.

ChatGPT 4 ▾



You



조지아와 서아시아의 평균 일일 육류 소비량의 합은 얼마입니까?
답변을 제공하기 전에 단계별 추론을 제공하십시오.

GPT-4



ChatGPT

To find the total average daily meat consumption for Georgia and Western Asia, we will add their individual average consumptions together.

The steps are as follows:

1. Identify the average daily meat consumption for Georgia from the chart.
2. Identify the average daily meat consumption for Western Asia from the chart.
3. Add the two values together to get the total.

Looking at the image provided:

- Georgia's average daily meat consumption is 79.84 grams per person.
- Western Asia's average daily meat consumption is 69.62 grams per person.

Now, let's add these two numbers together.

$$\text{Total average daily meat consumption} = 79.84g + 69.62g$$

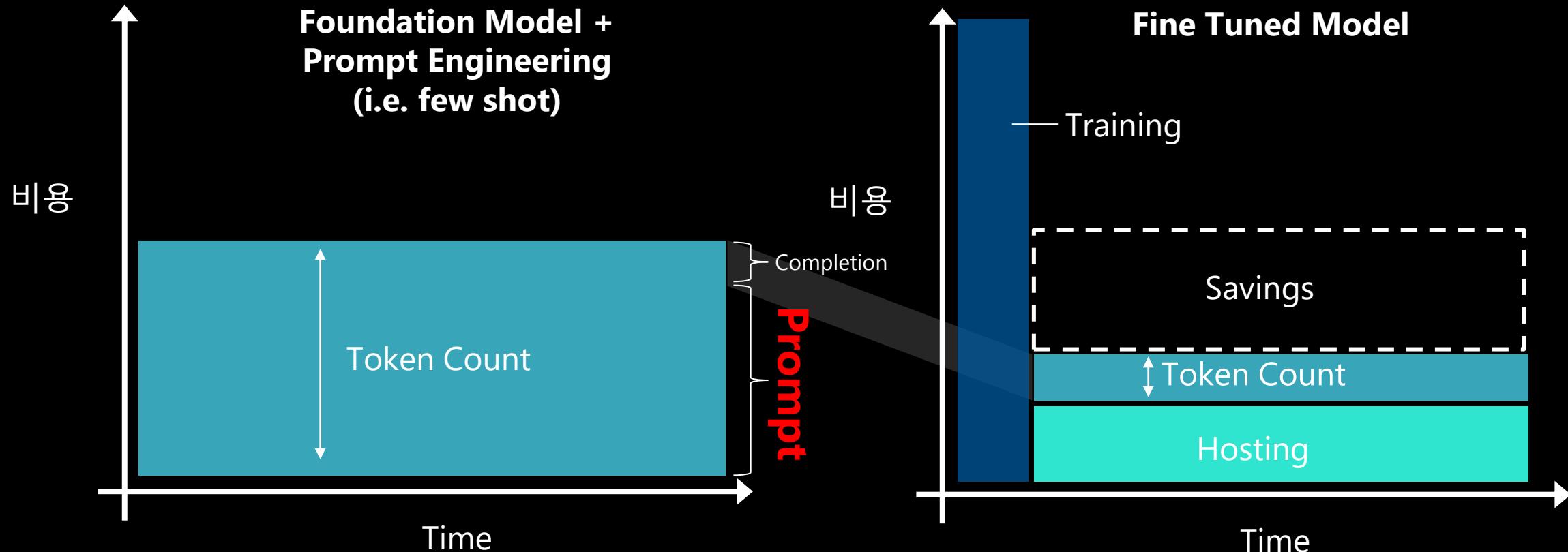
$$\text{Total average daily meat consumption} = 149.46g$$

So, the sum of the average daily meat consumption for Georgia and Western Asia is 149.46 grams.



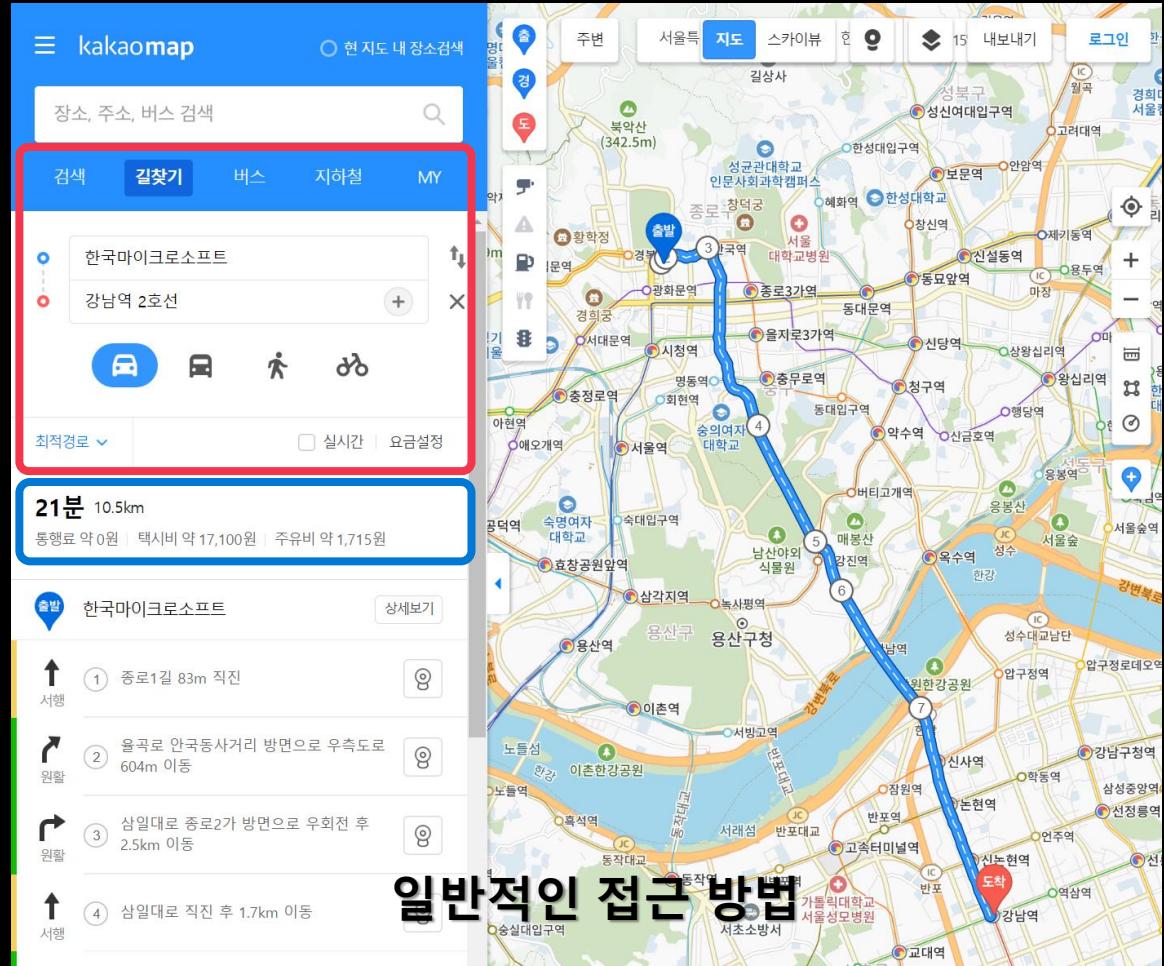
맞춤형 LLM

대량의 Prompt 토큰 절감, 맞춤형 스타일로 최적화 – **Fine-tuning**



사용자 인터페이스 진화

Function Calling – 주어진 형식과 절차 대신 자연스럽게 사람이랑 대화하듯 소통



한국마이크로소프트에서 강남역까지 얼마나 걸려?

A OpenAI gpt-35-turbo(gpt-4): Function Calling

카카오모빌리티 길찾기 API (출발지, 목적지)

```
def gpt(input):
    messages = [
        {"role": "user", "content": input}
    ]
    assistant_response = run_conversation(messages, functions, available_functions, deployment_id)
    if not assistant_response:
        return "제가 답변 드릴 수 있는 질문이 아닙니다. 다시 질문해주세요."
    else:
        content = json.dumps(assistant_response['choices'][0]['message']['content'], ensure_ascii=False, indent=4)
        content = content.replace("\n", "\n").replace("\\\\", "\\")
        return content
```

origin_name = "한국마이크로소프트"
destin_name = "강남역2호선"
query = f"{origin_name}에서 {destin_name}까지 얼마나 걸려?"

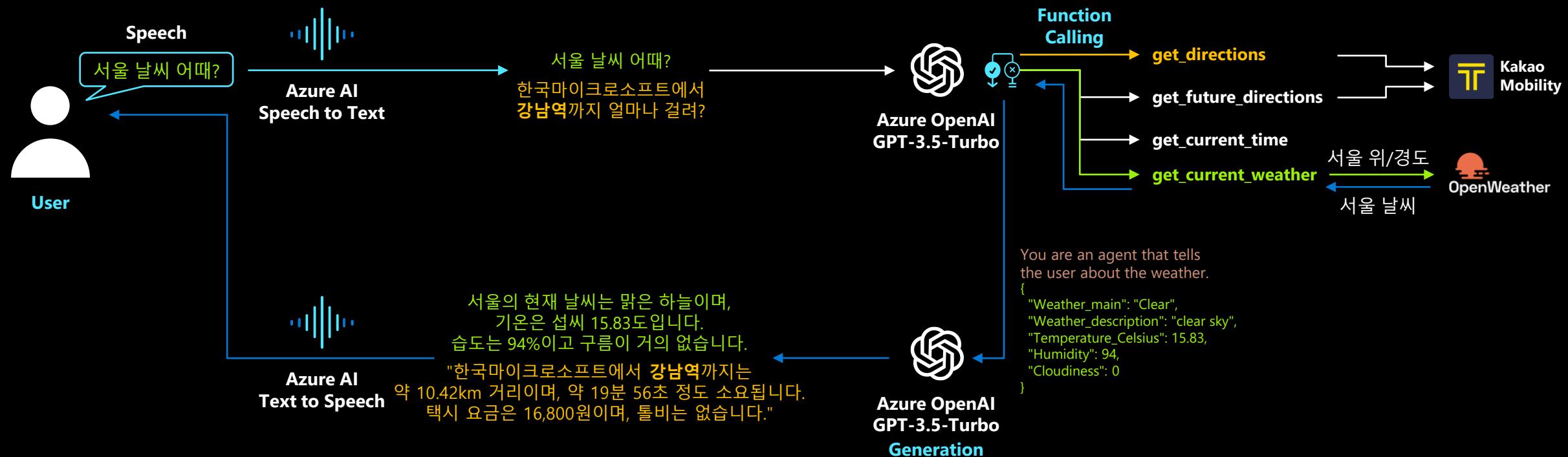
```
print(gpt(query))
```

한국마이크로소프트에서 강남역 2호선까지는 약 10.42km 거리이며, 차로 이동하는 경우 약 19분 56초가 소요됩니다. 택시 요금은 약 16,800원

ChatGPT에 외부 API를 결합

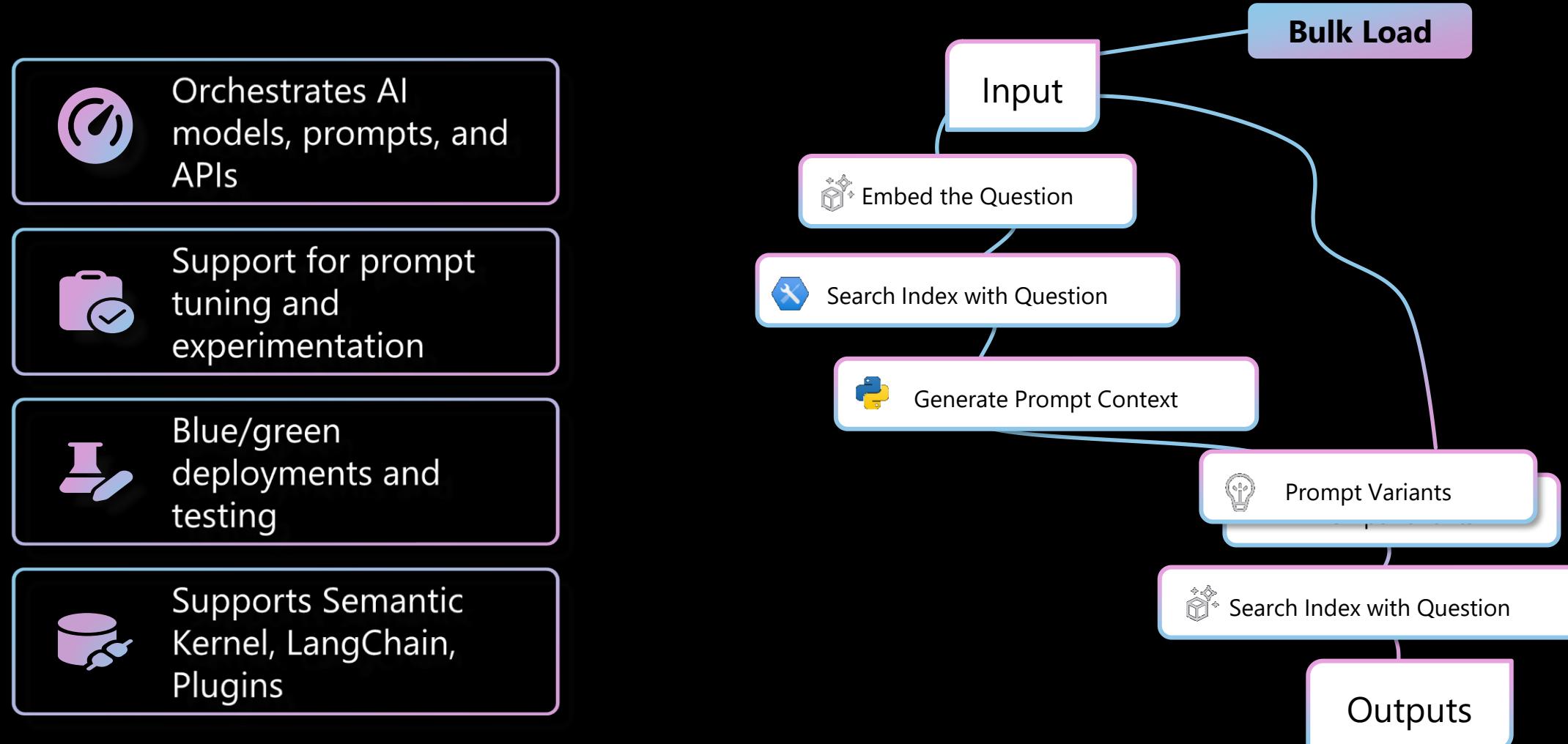
사용자 인터페이스 진화

Function Calling – 주어진 형식과 절차 대신 자연스럽게 사람이랑 대화하듯 소통



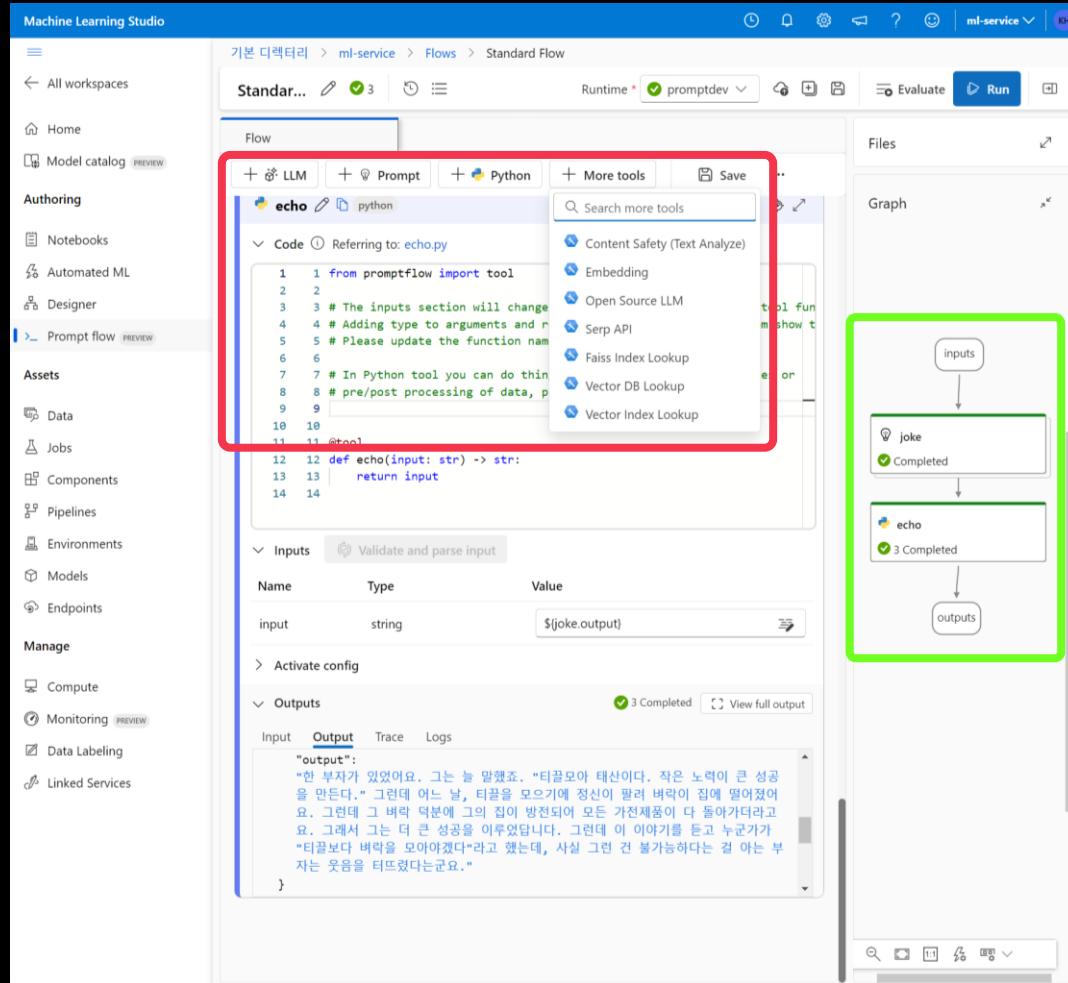
LLM App 제작을 더 쉽고 빠르게!

Prompt Flow – LLM 기반 응용 프로그램을 위한 전체 개발 주기를 단순화하는 도구



각 기능들을 연결하여 해결

Prompt Flow – LLM 기반 응용 프로그램을 위한 전체 개발 주기를 단순화하는 도구



워크플로우 개발

- 다양한 LLM 모델, 외부 데이터와 연결하고, 코드를 작성
시각화된 맞춤형 워크플로우 설계

테스트 및 평가

- 성능, 근거, 정확성과 같은 측정 항목을 사용하여 워크플로의 AI 품질을 평가
- 대규모 데이터 세트를 병렬로 테스트

프롬프트 튜닝

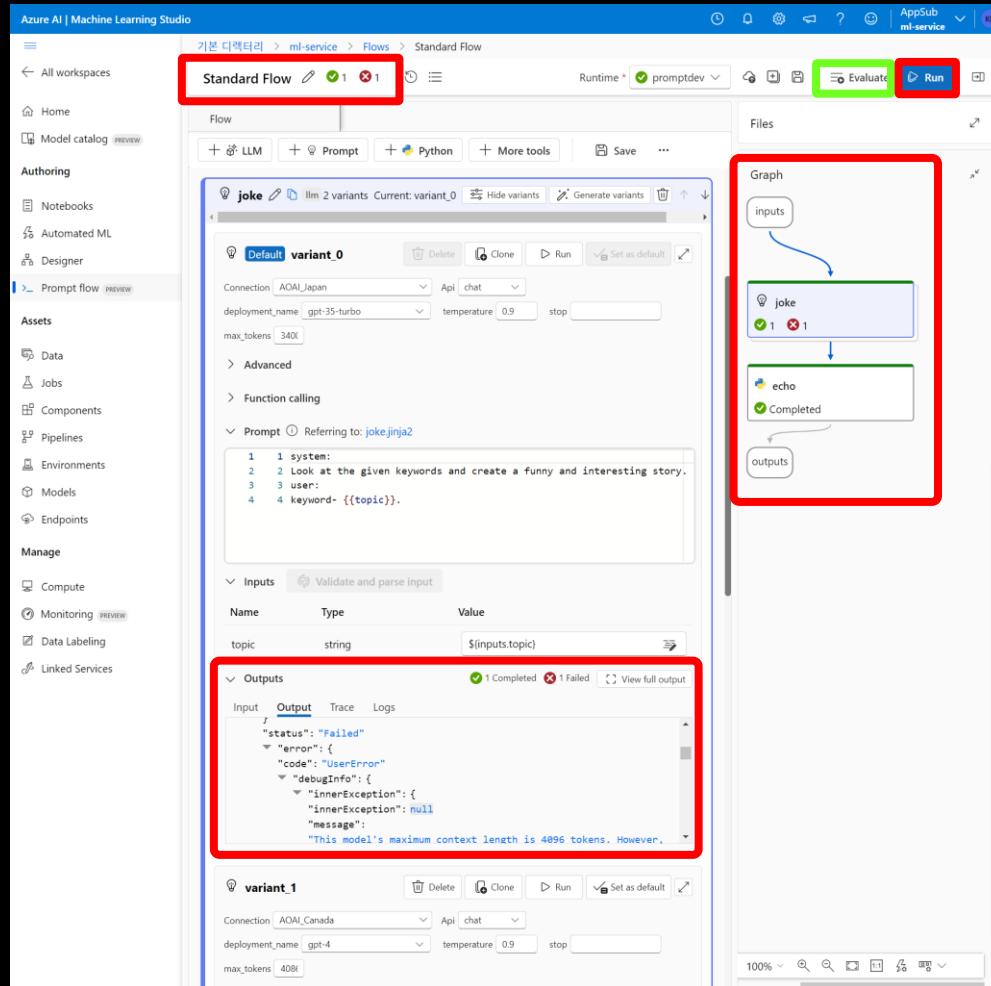
- 변형 및 버저닝으로 간단하게 프롬프트를 조정

비교 및 배포

- 여러 실험을 시각적으로 비교
- 신속한 통합을 위해 관리형 엔드포인트에 원 클릭 배포

각 기능들을 연결하여 해결

Prompt Flow – LLM 기반 응용 프로그램을 위한 전체 개발 주기를 단순화하는 도구



워크플로우 개발

- 다양한 LLM 모델, 외부 데이터와 연결하고, 코드를 작성하여 시각화된 맞춤형 워크플로우 설계

테스트 및 평가

- 성능, 근거, 정확성과 같은 측정 항목을 사용하여 워크플로의 AI 품질을 평가
- 대규모 데이터 세트를 병렬로 테스트

프롬프트 튜닝

- 변형 및 버저닝으로 간단하게 프롬프트를 조정

비교 및 배포

- 여러 실험을 시각적으로 비교
- 신속한 통합을 위해 관리형 엔드포인트에 원 클릭 배포

각 기능들을 연결하여 해결

Prompt Flow – LLM 기반 응용 프로그램을 위한 전체 개발 주기를 단순화하는 도구

Overview

Outputs Logs Metrics Trace Snapshot

Refresh Export Append related results Web Classification-08-21-2C

Line number	url	Variant ID	Status	category	evidence	groundtruth	prediction	grade
0	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5893961/	Completed	Academic	Both	Academic	Academic	Academic	Correct
1	https://gutenberg.org/files/1342/1342-pdf.pdf	Completed	PDF	Both	PDF	PDF	PDF	Correct
2	https://www.iso.org/obp/ui/#std:iso:9001:ed-5v1en	Completed	PDF	URL	PDF	PDF	PDF	Correct
3	https://www.jstor.org/stable/10.1086/679716	Completed	Academic	Both	Academic	Academic	Academic	Correct
4		Completed	Profile	URL	Channel	Profile	Profile	Incorrect
5		Completed	Channel	URL	Channel	Channel	Channel	Correct
6		Completed	Profile	Both	Profile	Profile	Profile	Correct
7		Completed	Academic	Both	Academic	Academic	Academic	Correct
8		Completed	Profile	URL	Profile	Profile	Profile	Correct
9		Completed	App	Both	App	App	App	Correct
10		Completed	Movie	URL	Movie	Movie	Movie	Correct
11	https://www.w3.org/TR/PNG-Structure.pdf	Completed	PDF	Both	PDF	PDF	PDF	Correct
12	https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70&Lang=E	Completed	PDF	URL	PDF	PDF	PDF	Correct

< Prev Next > 25/Page

워크플로우 개발

- 다양한 LLM 모델, 외부 데이터와 연결하고, 코드를 작성하여 시각화된 맞춤형 워크플로우 설계

테스트 및 평가

- 성능, 근거, 정확성과 같은 측정 항목을 사용하여 워크플로의 AI 품질을 평가
- 대규모 데이터 세트를 병렬로 테스트

프롬프트 튜닝

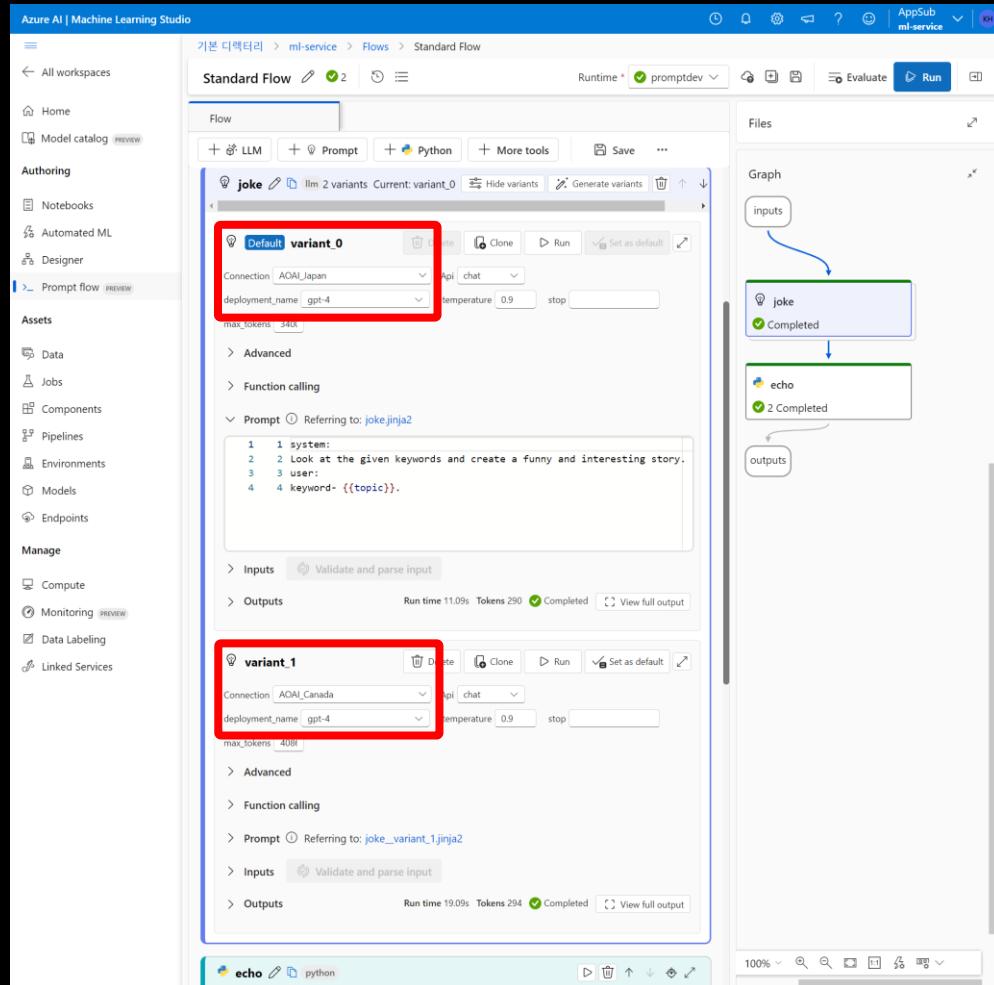
- 변형 및 버저닝으로 간단하게 프롬프트를 조정

비교 및 배포

- 여러 실험을 시각적으로 비교
- 신속한 통합을 위해 관리형 엔드포인트에 원 클릭 배포

각 기능들을 연결하여 해결

Prompt Flow – LLM 기반 응용 프로그램을 위한 전체 개발 주기를 단순화하는 도구



워크플로우 개발

- 다양한 LLM 모델, 외부 데이터와 연결하고, 코드를 작성하여 시각화된 맞춤형 워크플로우 설계

테스트 및 평가

- 성능, 근거, 정확성과 같은 측정 항목을 사용하여 워크플로의 AI 품질을 평가
- 대규모 데이터 세트를 병렬로 테스트

프롬프트 튜닝

- 변형 및 버저닝으로 간단하게 프롬프트를 조정

비교 및 배포

- 여러 실험을 시각적으로 비교
- 신속한 통합을 위해 관리형 엔드포인트에 원 클릭 배포

각 기능들을 연결하여 해결

Prompt Flow – LLM 기반 응용 프로그램을 위한 전체 개발 주기를 단순화하는 도구

The screenshot shows the Azure AI Machine Learning Studio interface. In the center, there's a table titled 'Outputs' with four columns: 'name', 'duration', 'start time (UTC)', and 'end time (UTC)'. The 'Trace' tab is selected. A red box highlights this section. Below the table, there's a large JSON object representing the prompt flow configuration. The JSON includes sections for 'inputs' (with 'api_base', 'api_type', 'api_version', 'engine', 'frequency_penalty', 'logit_bias', 'max_tokens', and 'messages' arrays), 'n', 'presence_penalty', 'stop', 'stream', 'temperature', 'top_p', and 'user'.

name	duration	start time (UTC)	end time (UTC)
AzureOpenAI.chat	11,084 ms	2023-10-25 11:43:0:590	2023-10-25 11:43:11:674
openai.api_resources.chat_completion.ChatCor	11,081 ms	2023-10-25 11:43:0:593	2023-10-25 11:43:11:674
echo	0 ms	2023-10-25 11:43:11:694	2023-10-25 11:43:11:694
AzureOpenAI.chat	19,090 ms	2023-10-25 11:43:12:486	2023-10-25 11:43:31:576
openai.api_resources.chat_completion.ChatCor	19,087 ms	2023-10-25 11:43:12:489	2023-10-25 11:43:31:576
echo	0 ms	2023-10-25 11:43:31:601	2023-10-25 11:43:31:601

```
{
  "inputs": {
    "api_base": "https://japanopenaihs.openai.azure.com/",
    "api_type": "azure",
    "api_version": "2023-07-01-preview",
    "engine": "gpt-4",
    "frequency_penalty": 0,
    "logit_bias": {},
    "max_tokens": 3400
  },
  "messages": [
    {
      "content": "Look at the given keywords and create a funny and interesting story. Limit the length of your story to a maximum of 100 words. Output language is Korean.",
      "role": "system"
    },
    {
      "content": "keyword- 터끌, 바락, 부자, 성공",
      "role": "user"
    }
  ],
  "n": 1,
  "presence_penalty": 0,
  "stop": null,
  "stream": true,
  "temperature": 0.9,
  "top_p": 1,
  "user": ""
}
```

워크플로우 개발

- 다양한 LLM 모델, 외부 데이터와 연결하고, 코드를 작성하여 시각화된 맞춤형 워크플로우 설계

테스트 및 평가

- 성능, 근거, 정확성과 같은 측정 항목을 사용하여 워크플로의 AI 품질을 평가
- 대규모 데이터 세트를 병렬로 테스트

프롬프트 튜닝

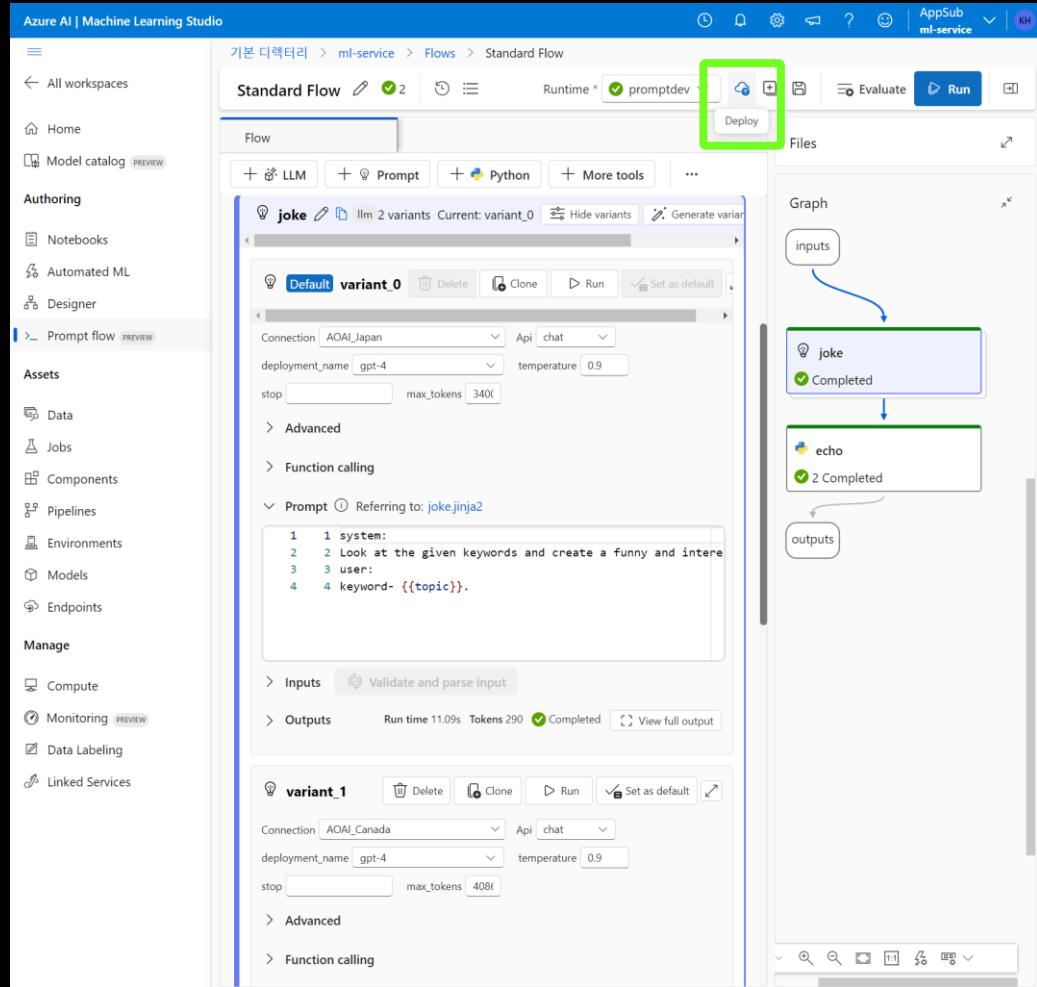
- 변형 및 버저닝으로 간단하게 프롬프트를 조정

비교 및 배포

- 여러 실험을 시각적으로 비교
- 신속한 통합을 위해 관리형 엔드포인트에 원 클릭 배포

각 기능들을 연결하여 해결

Prompt Flow – LLM 기반 응용 프로그램을 위한 전체 개발 주기를 단순화하는 도구



워크플로우 개발

- 다양한 LLM 모델, 외부 데이터와 연결하고, 코드를 작성하여 시각화된 맞춤형 워크플로우 설계

테스트 및 평가

- 성능, 근거, 정확성과 같은 측정 항목을 사용하여 워크플로의 AI 품질을 평가
- 대규모 데이터 세트를 병렬로 테스트

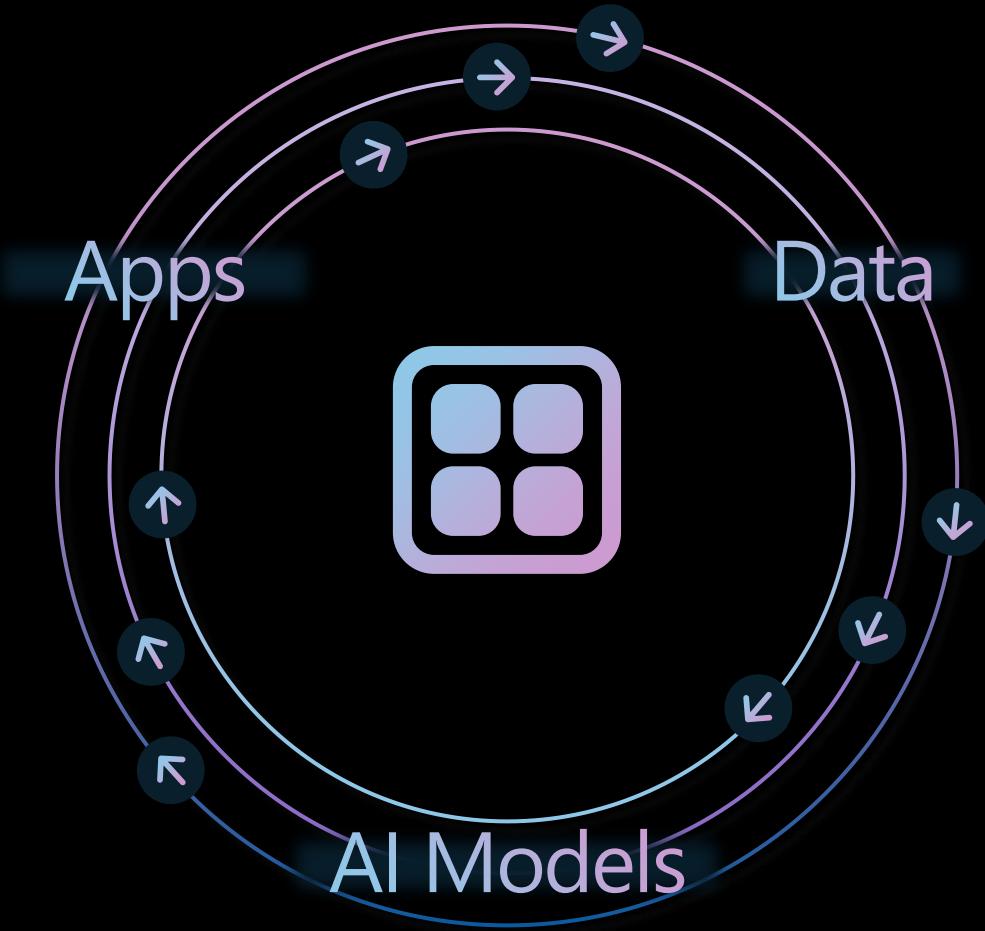
프롬프트 튜닝

- 변형 및 버저닝으로 간단하게 프롬프트를 조정

비교 및 배포

- 여러 실험을 시각적으로 비교
- 신속한 통합을 위해 관리형 엔드포인트에 원 클릭 배포

AI 시대에는,
애플리케이션은
지능적이어야 합니다.





한국 스타트업들의 활용 사례

생성형 AI 서비스를 잘 활용하는 스타트업

한국 스타트업 기업들의 생성형 AI를 다룬 노하우 대방출



(메뉴 추천)



(고객 문의)



(리뷰 요약)



(생성 AI 솔루션)



(자체 검색)



(리뷰 작성)



(타로, 상담)



(돌봄서비스 로봇)



(고객 센터)



(학습 진단)



(금융 판단)



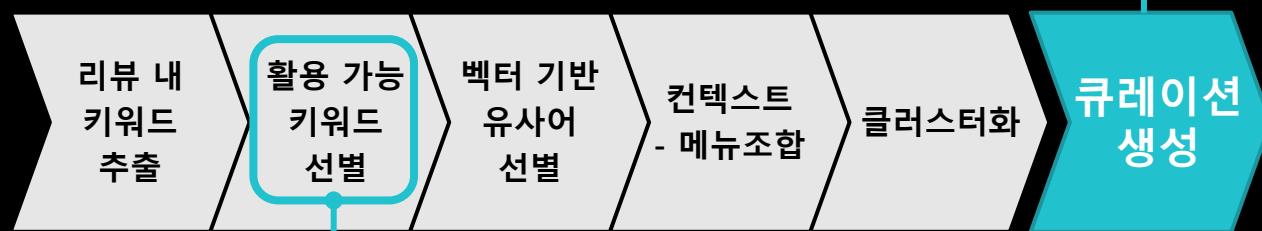
(증권 정보)

AI가 제안하는 추천

리뷰를 분석하여 메뉴 및 매장 추천

먹을 준비 되신 분이라면 주목하세요 🍔
양 많은 떡볶이집이 있으면 좋겠지만, 어디가 좋은 곳인지 모르시죠? 제가 알려드릴게요!

사람은 퀄리티 체크를 담당하며 AI를 도와요!



"우리는 이것을 컨텍스트라고 부르기로 했어요"

- 시간대 (아침, 점심, 저녁, 야식..)
- 함께 먹는 사람 (아이, 가족, 사랑하는 사람, 혼자..)
- 상황 (퇴근, 스트레스, 축하, 식단..)
- 맛 (맵다, 달다, 탱글탱글..)

AI가 제안하는

가게를 확인하고
주문해 보세요!



← 가족과 즐기는 떡볶이



GPT-4 활용 추천 ⓘ



먹을 준비 되신 분이라면 주목하세요 🍔

양 많은 떡볶이집이 있으면 좋겠지만, 어디가 좋은 곳인지 모르시죠? 제가 알려드릴게요!

#가족

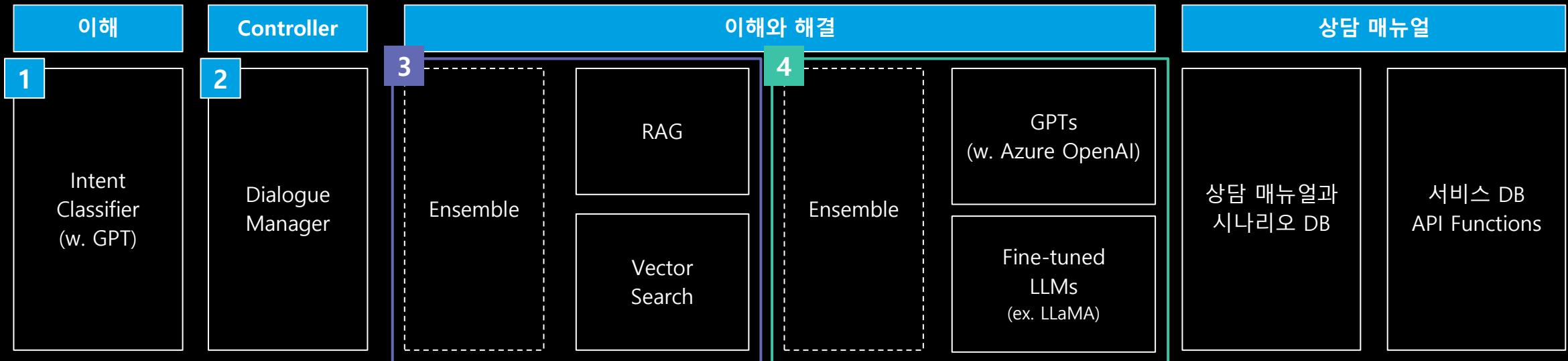
#떡볶이

#양 많은

고객경험 증대를 위한 고객센터



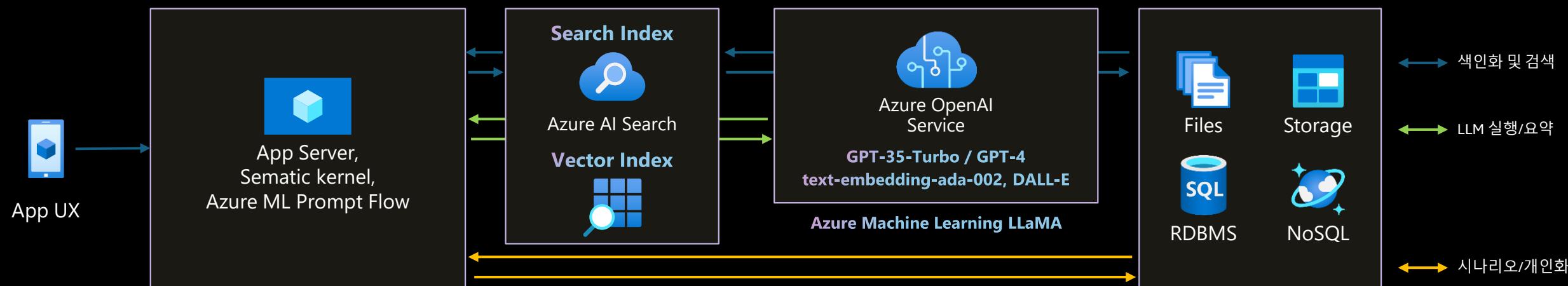
빠르고 정확하게 고객들의 고민을 해결



- 영상: https://youtu.be/_1mcVipuUYY (쏘카 박경호 AI팀 팀장)

Azure 기반 서비스 구축 방법

용도에 맞는 완전 관리형 서비스로 안전하고 빠르게 구축



검색엔진 최적화 / 리뷰 작성 도우미



검색 품질을 향상시키고 상품 리뷰 작성은 쉽게 할 수 있도록 도움

시리얼



상품 (2개)

매장 (0개)



시리얼



신상품순최고가격순최저가격순



아몬드 콘플레이크(240g)
6,300 원



유기농 현미플레이크(240g)
6,300 원





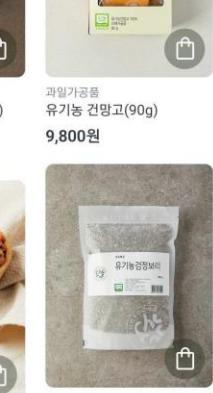
과자 유기농 그레놀라(사과&건포도/300g)
7,000 원



과자 유기농 그레놀라(코코넛&레드베리/260g)
8,000 원



과자 아몬드콘플레이크(240g)
30% 4,400 원
6,300 원



건과류 유기농 구운아몬드(150g)
13,900 원



과일가공품 유기농 건망고
9,800 원



건과류 유기농 코코넛칩(80g)
4,900 원



잡곡 유기농 오트밀(420g)
7,500 원

건과류 헤이즐넛(200g)
9,900 원



아몬드 콘플레이크(240g)
6,300 원



유기농 현미플레이크(240g)
6,300 원





과자 유기농 현미코코아볼(240g)
6,300 원



과자 유기농 현미플레이크(240g)
6,300 원



과자 유기농 후레쉬믹스볼(240g)
19% 26,300 원
24,600 원



건과류 유기농 구운 아몬드(450g)
15% 5,600 원
6,600 원



잡곡 유기농 검정보리(800g)
8,500 원



건과류 유기농 푸른(150g)
25% 18,350 원
24,600 원



건과류 유기농 아빠건과(420g)
9,400 원

검색엔진 최적화 / 리뷰 작성 도우미

GPT-4를 활용한 검색 품질 향상



Goods
detail



Goods
reviews



Order
records

Raw data

Processed data

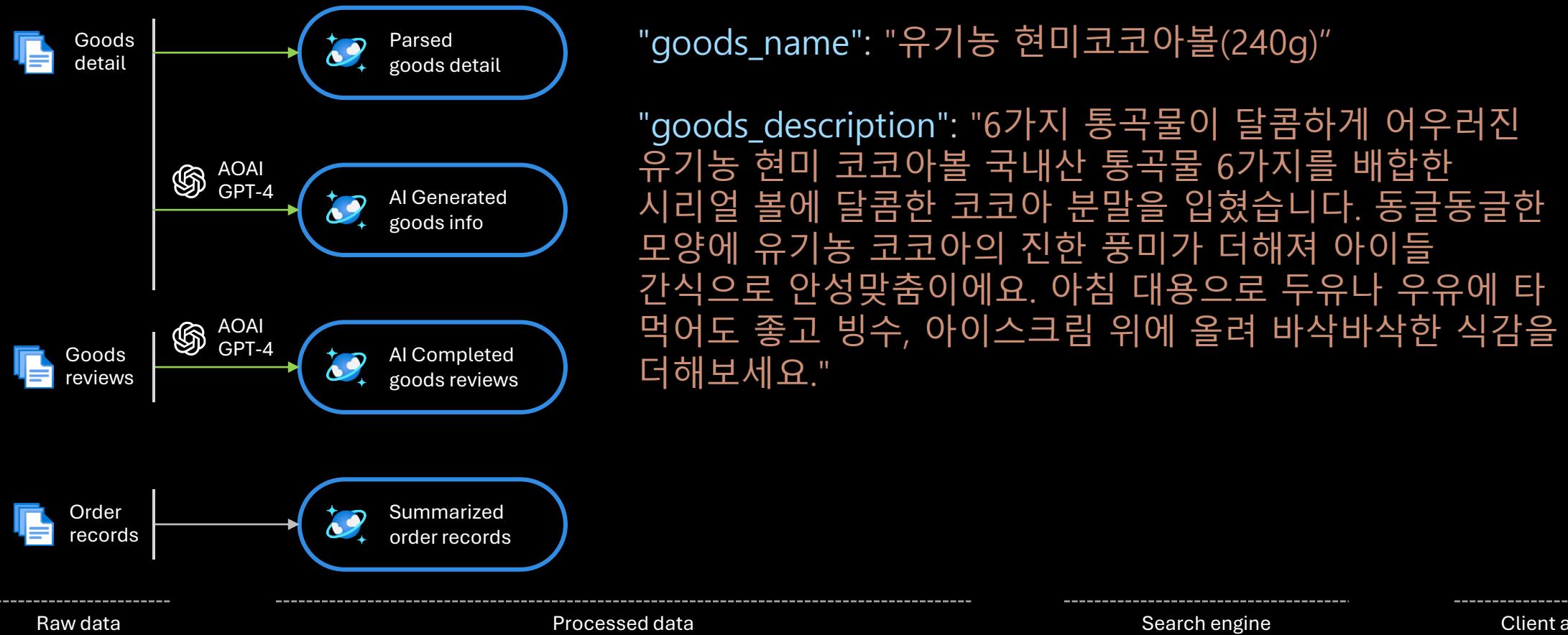
Search engine

Client app

검색엔진 최적화 / 리뷰 작성 도우미



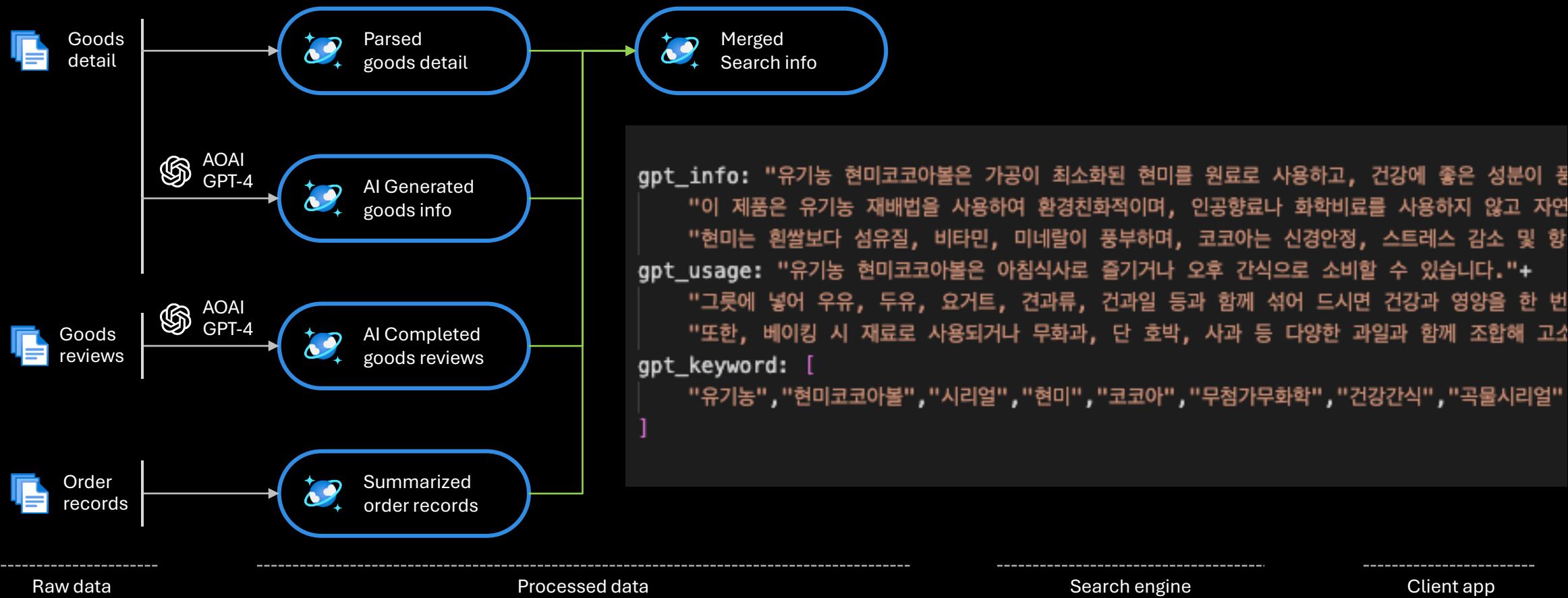
GPT-4를 활용한 검색 품질 향상



검색엔진 최적화 / 리뷰 작성 도우미



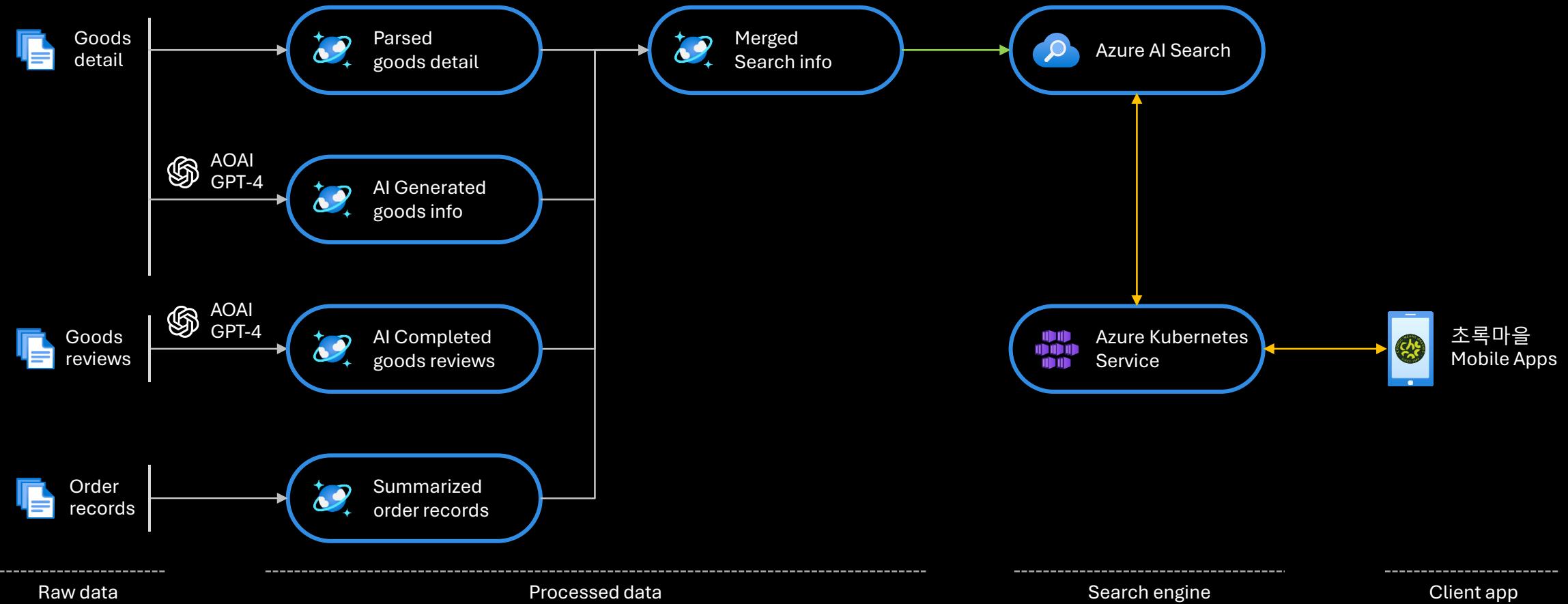
GPT-4를 활용한 검색 품질 향상



검색엔진 최적화 / 리뷰 작성 도우미



GPT-4를 활용한 검색 품질 향상

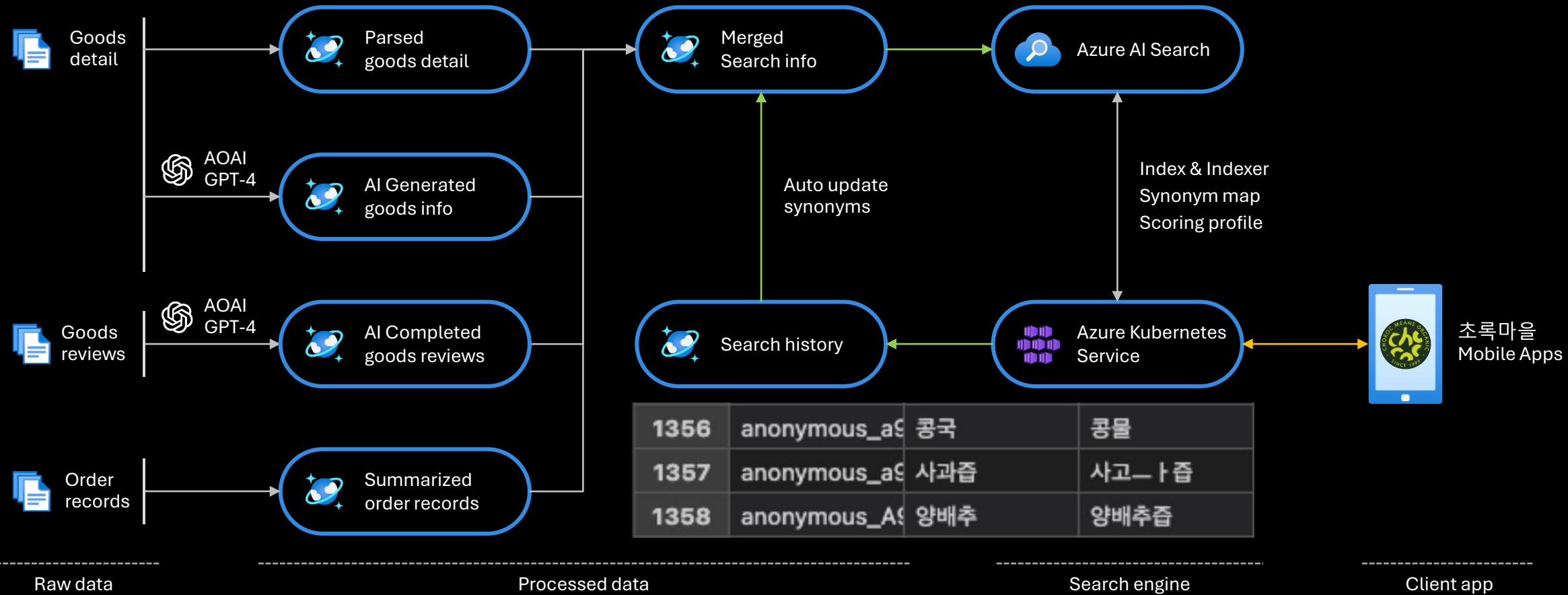


영상: <https://youtu.be/JJf1CEJzOGw>, <https://share.vidyard.com/watch/pV7utU9ohvabbi7YuACW9E> (정육각/초록마을 박준태 CTO)

검색엔진 최적화 / 리뷰 작성 도우미



GPT-4를 활용한 검색 품질 향상

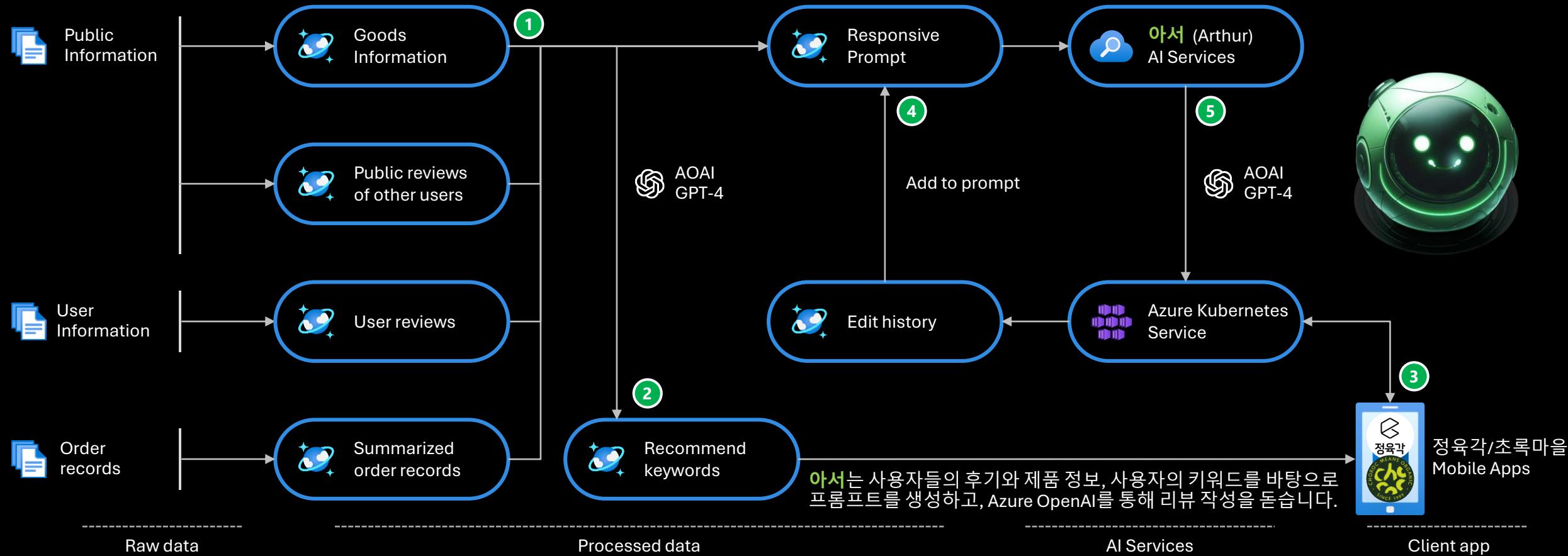


영상: <https://youtu.be/JJf1CEJzOGw>, <https://share.vidyard.com/watch/pV7utU9ohvabbi7YuACW9E> (정육각/초록마을 박준태 CTO)

검색엔진 최적화 / 리뷰 작성 도우미



상품 리뷰 작성은 쉽게 할 수 있도록 도우미 - 아서



후기 요약



말 안듣는 GPT로 품질/가격/가용성을 확보한 후기 요약 만들기

Quality

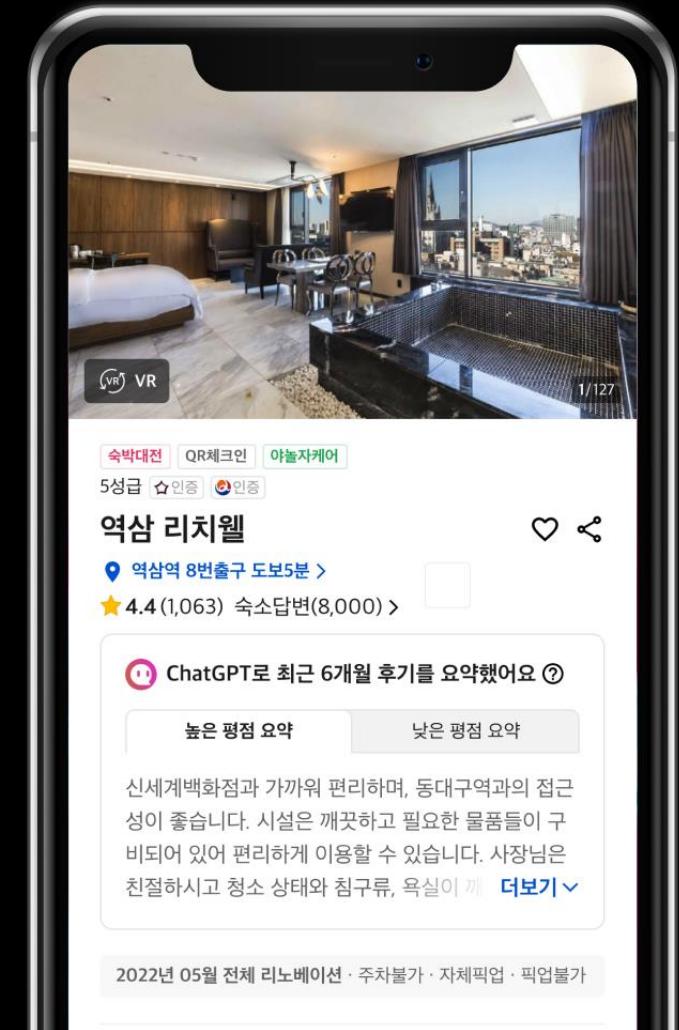
- Input Filtering
- Output Filtering
- Prompt Evaluation Techniques: Rule-Based vs Human vs GPT-4

Cost

- LLM is Expensive
- Budget Allocation
- Usage Control

Availability

- Hardware Dependencies
- Scalability Issues
- Rate Limiting and Resource Management
- Solution - Resource Management System



도메인 조직 주도 고객 이해 방법



기술전문가 협업 없이 도메인 조직 주도의 업무혁신 시도

가공

번역

오타
보정

토큰화

형태소
분석

단어문서
행렬

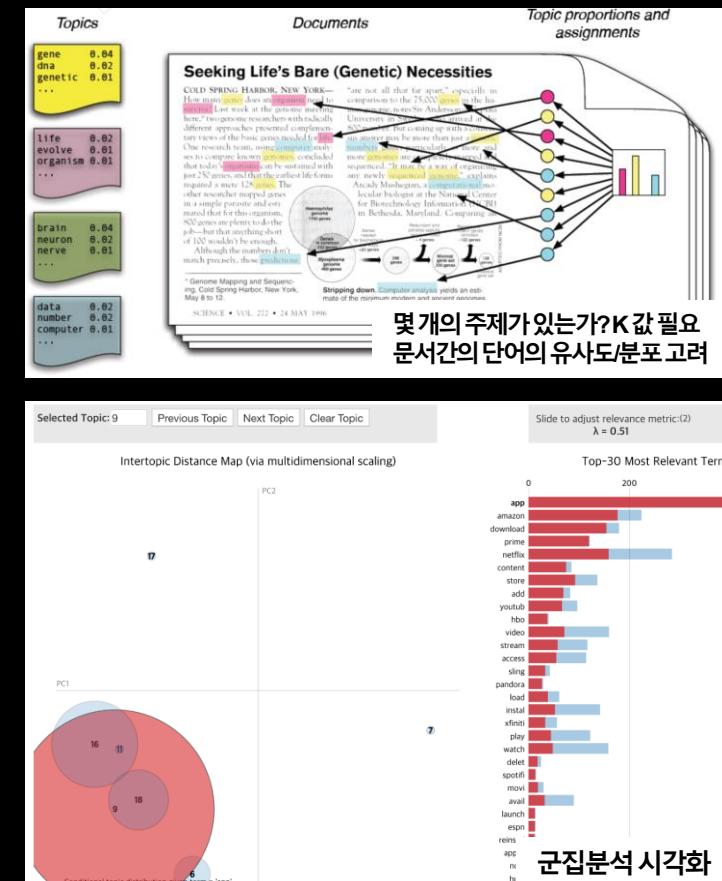
문서
군집화

군집별
단어분석

군집 별
문장생성

감성
분석

모델링, 분석



생성

사람이 직접 분석 및 해석을
해야 한다.



사람이 작성하여 편향 발생



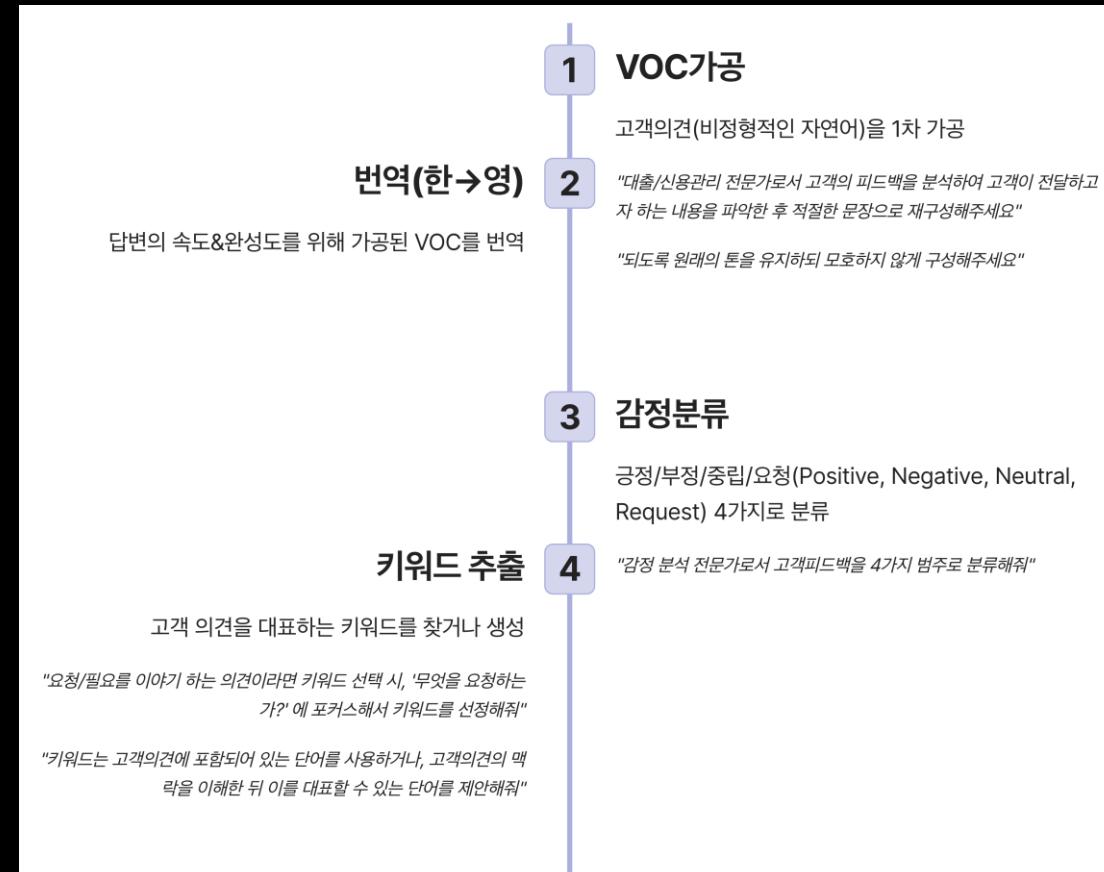
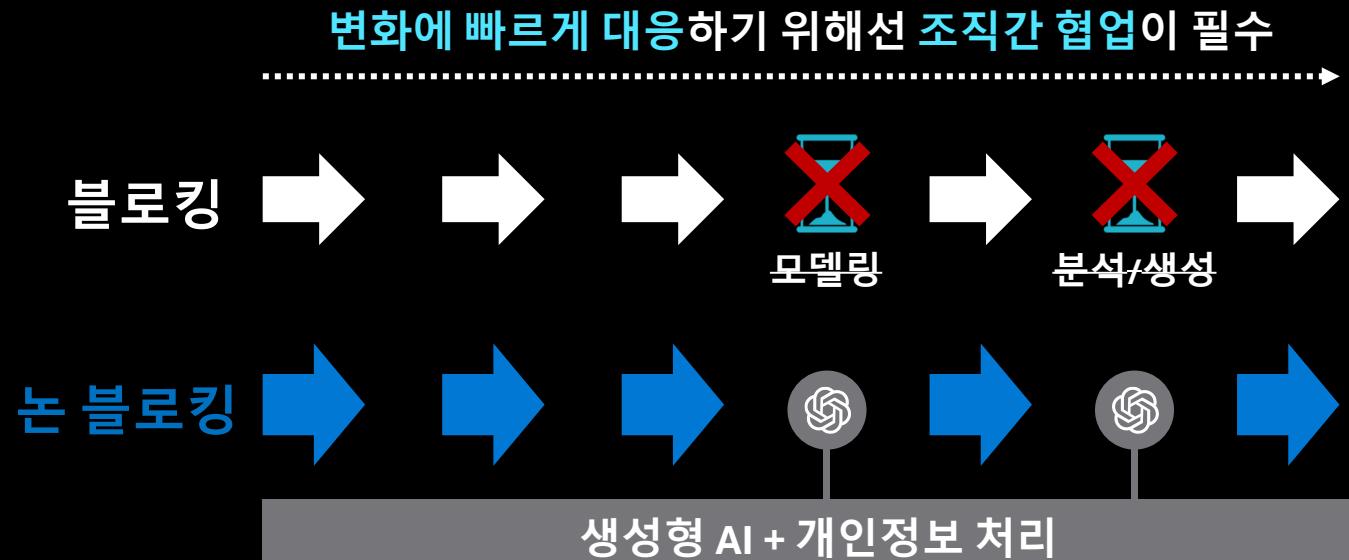
전문가 협업이 필요함



도메인 조직 주도 고객 이해 방법



CX팀 스스로 고객을 이해 – DS, Dev 없어도 할 수 있어요!





RAG 아키텍처 구성

RAG 아키텍처 만들기 - on your data

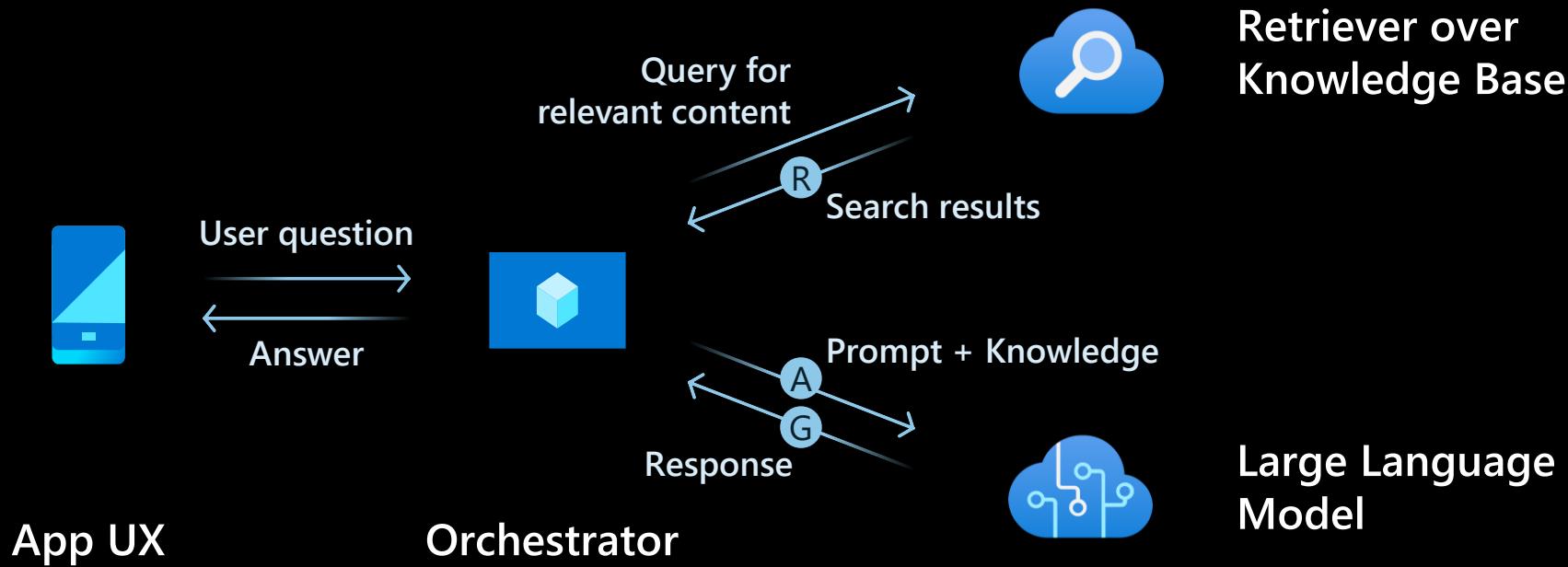
Retrieval Augmented Generation (RAG) – Wikipedia 데이터 기반 RAG

The screenshot displays a user interface for a Retrieval-Augmented Generation (RAG) system, likely built on top of a Microsoft product like Microsoft Copilot or a similar AI tool. The interface is split into several sections:

- Header:** Shows the profile picture and name "Contoso".
- Input Area:** A text input field containing the query "4월과 8월의 차이를 알려줘." (Tell me about the difference between April and August).
- Left Sidebar (Citations):** A vertical sidebar titled "Citations" containing a single item for "April".
- Main Content Area:** The main pane displays the results of the query. It includes:
 - A section titled "달의 위치와 일수" (Position and Number of Days) with two bullet points:
 - 4월: 4월은 그레고리력과 윤리우스력에서 네 번째 달로, 30일을 가지고 있습니다¹.
 - 8월: 8월은 그레고리력에서 여덟 번째 달로, 31일을 가지고 있습니다².
 - A section titled "계절" (Season) with two bullet points:
 - 4월: 북반구에서는 봄, 남반구에서는 가을에 해당합니다¹.
 - 8월: 북반구에서는 여름, 남반구에서는 겨울에 해당합니다².
 - A section titled "시작과 끝 요일" (Start and End Day of the Month) with two bullet points:
 - 4월: 매년 7월과 같은 요일에 시작하며, 윤년에는 1월과 같은 요일에 시작합니다. 또한, 12월과 같은 요일에 끝납니다¹.
 - 8월: 8월은 일반적으로 다른 달과 같은 요일에 시작하지 않지만, 윤년에는 2월과 같은 요일에 시작합니다. 8월은 매년 11월과 같은 요일에 끝납니다².
 - A section titled "명칭의 유래" (Name Origin) with two bullet points:
 - 4월: 4월의 이름은 라틴어 "aperire"에서 유래했을 가능성이 있으며, 이는 꽃이 피는 것을 의미합니다. 또 다른 이론으로는 그리스 신화의 사랑의 여신 아프로디테에서 유래했을 가능성도 있습니다¹.
 - 8월: 8월은 로마 황제 아우구스투스 시저의 이름을 따서 명명되었습니다².
 - A section titled "주요 기념일 및 행사" (Major Holidays and Events) with two bullet points:
 - 4월: 4월에는 만우절(4월 1일), 세계 자폐증 인식의 날(4월 2일), 지구의 날(4월 22일) 등 다양한 기념일이 있습니다¹.
 - 8월: 8월에는 스위스 국경일(8월 1일), 인도의 독립기념일(8월 15일), 세계 인도주의의 날(8월 19일) 등 다양한 기념일이 있습니다².
- Bottom Navigation:** Includes a "Type a new question..." input field and two blue buttons with icons: a magnifying glass and a plus sign.

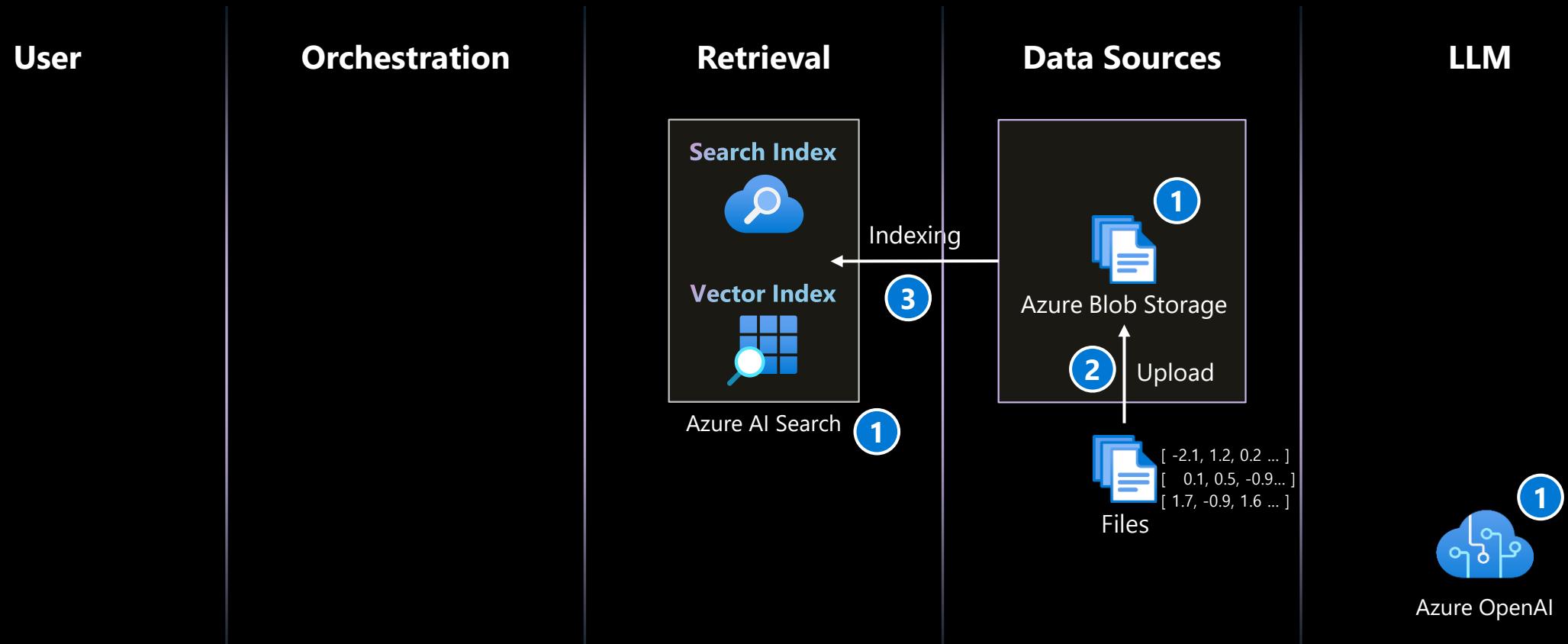
외부 데이터 활용 워크플로우

Retrieval Augmented Generation (RAG)



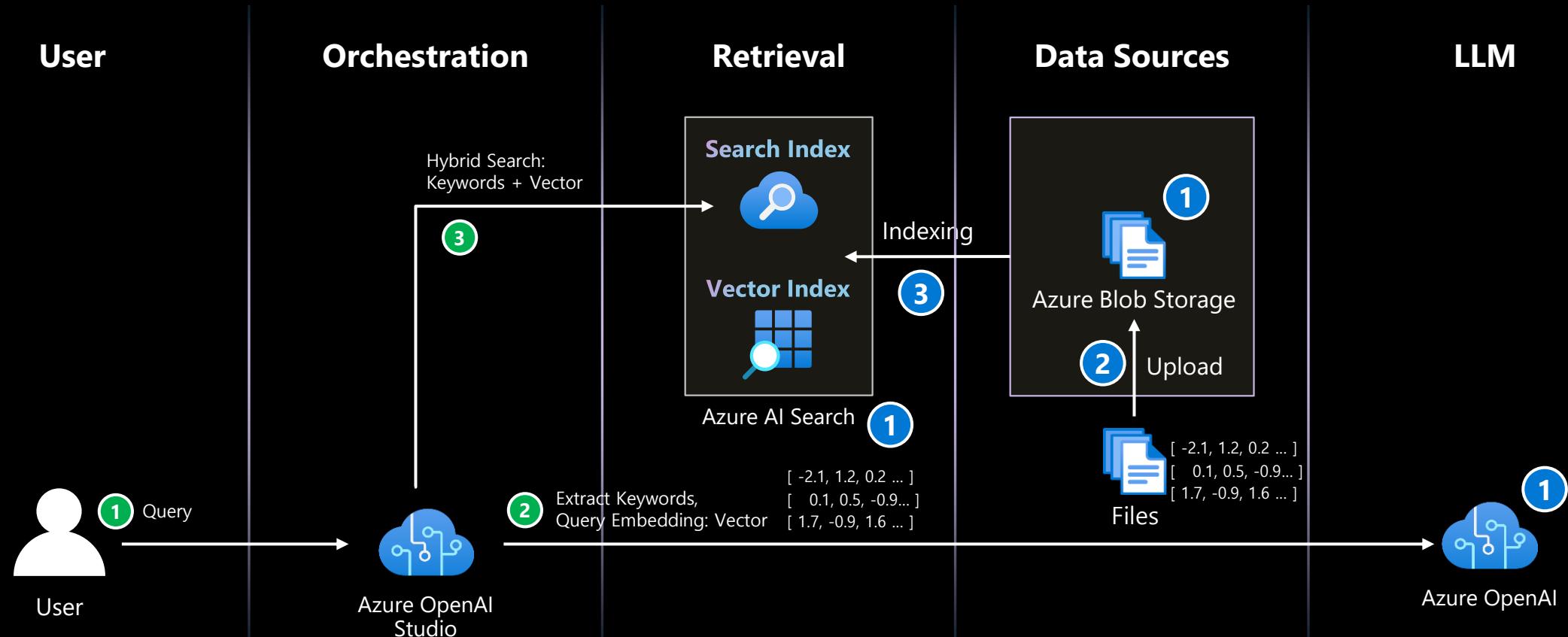
RAG 아키텍처 만들기 - on your data

Retrieval Augmented Generation (RAG) – Wikipedia 챗봇 서비스



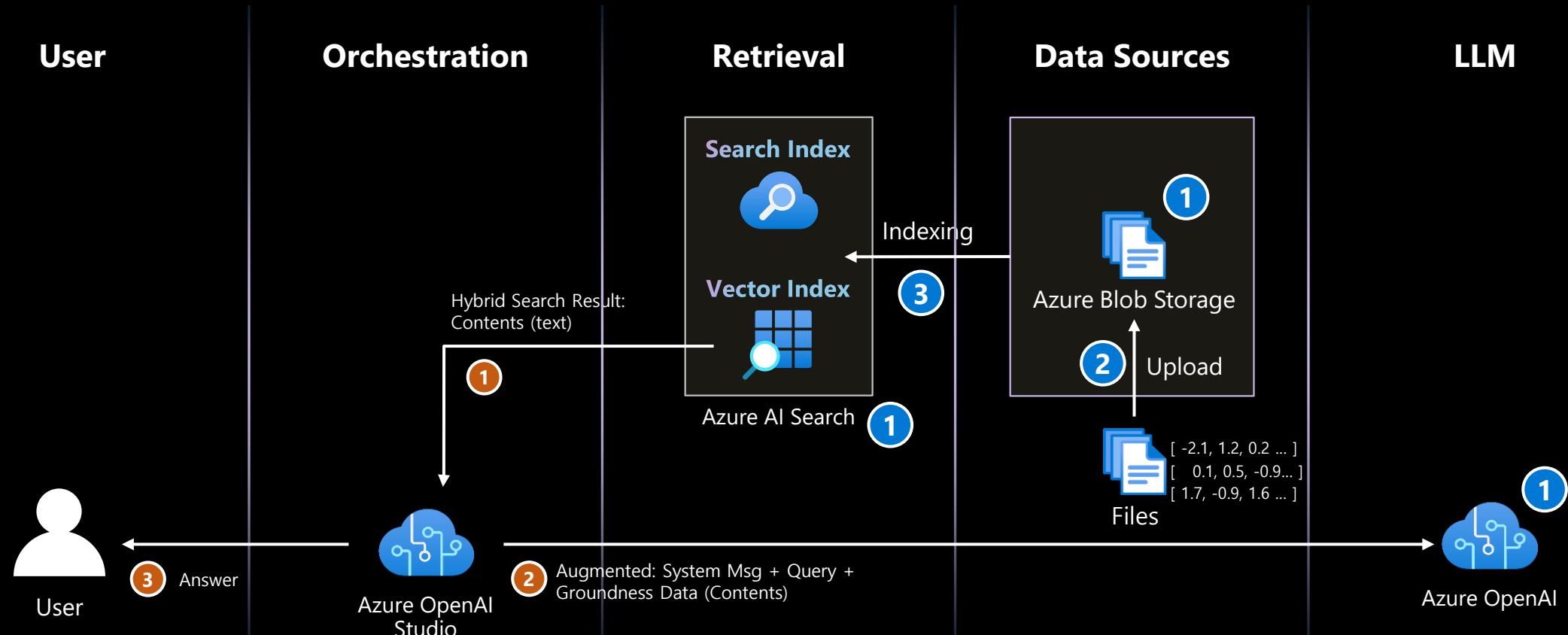
RAG 아키텍처 만들기 - on your data

Retrieval Augmented Generation (RAG) – Wikipedia 챗봇 서비스



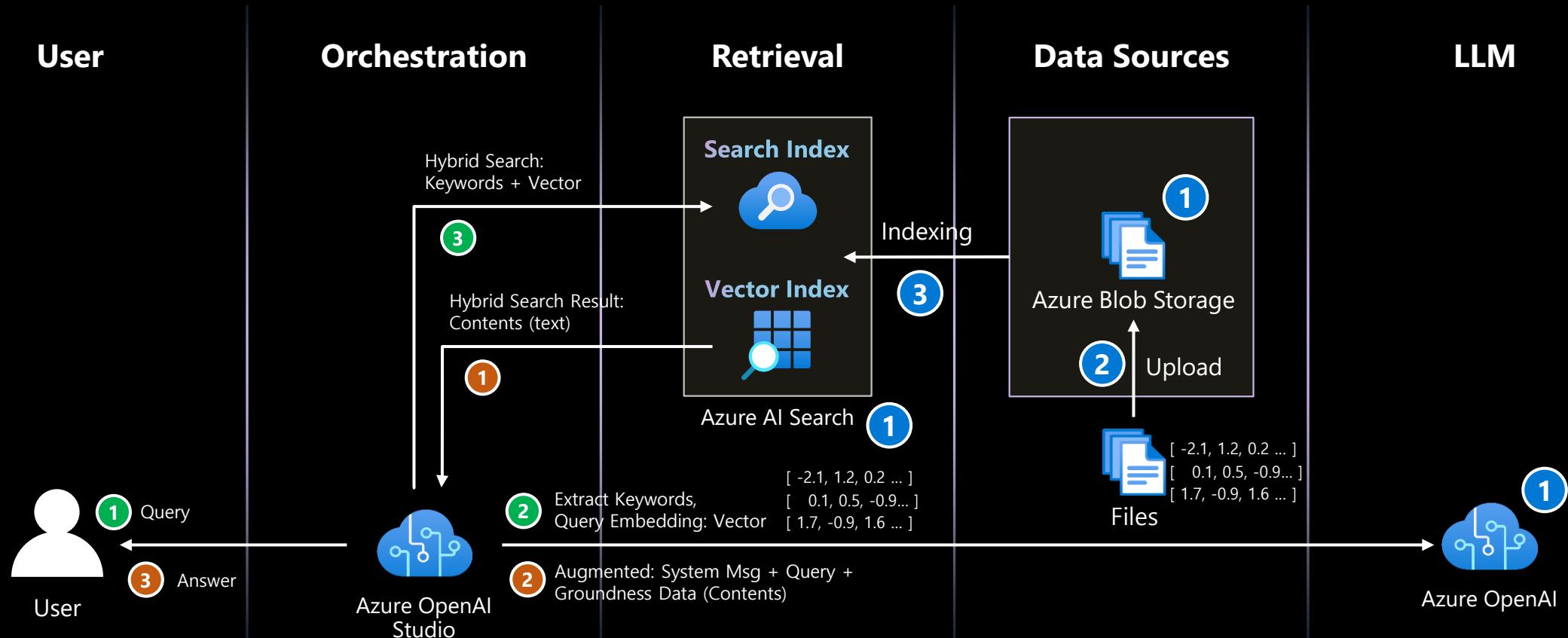
RAG 아키텍처 만들기 - on your data

Retrieval Augmented Generation (RAG) – Wikipedia 챗봇 서비스

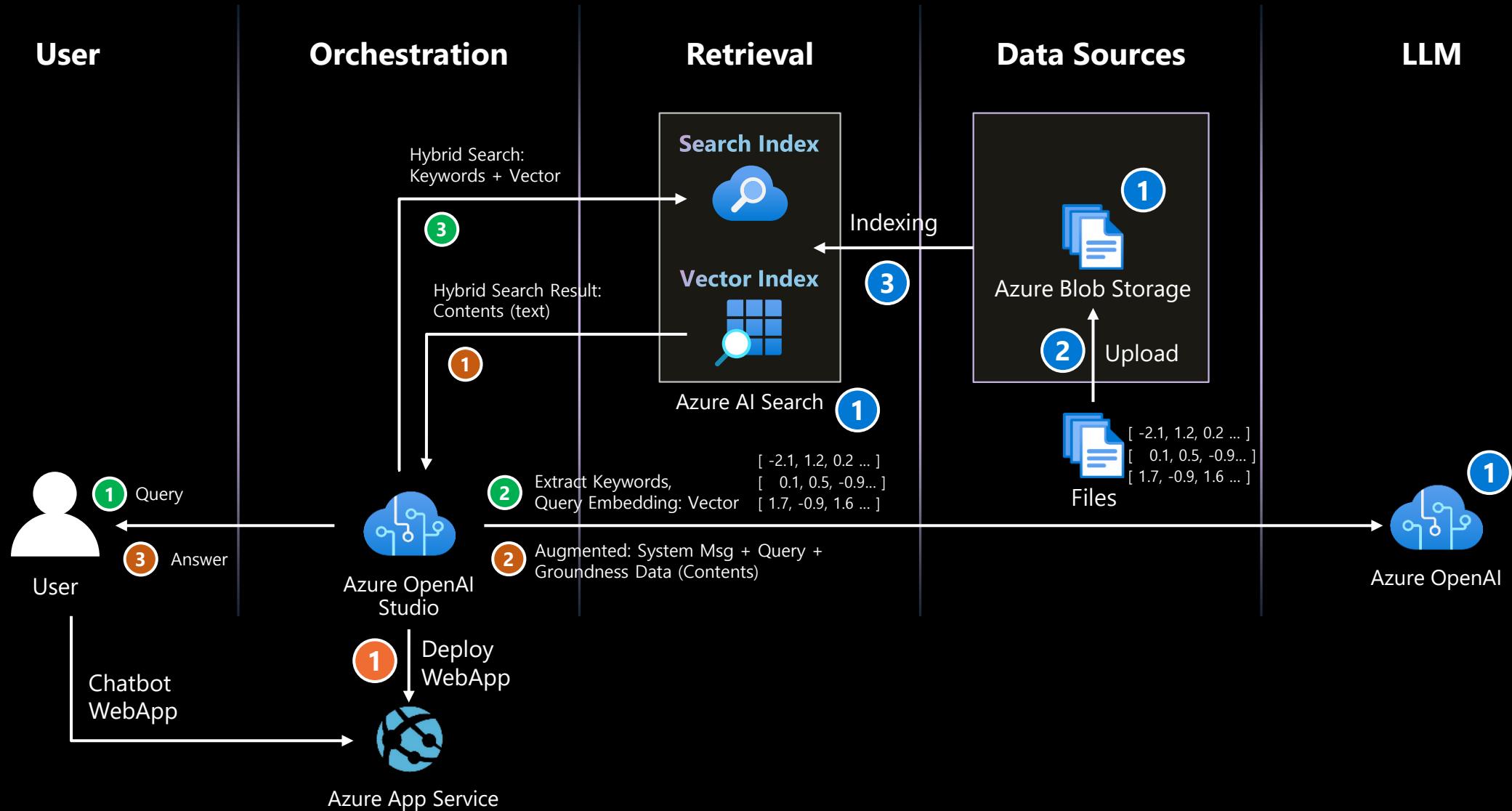


RAG 아키텍처 만들기 - on your data

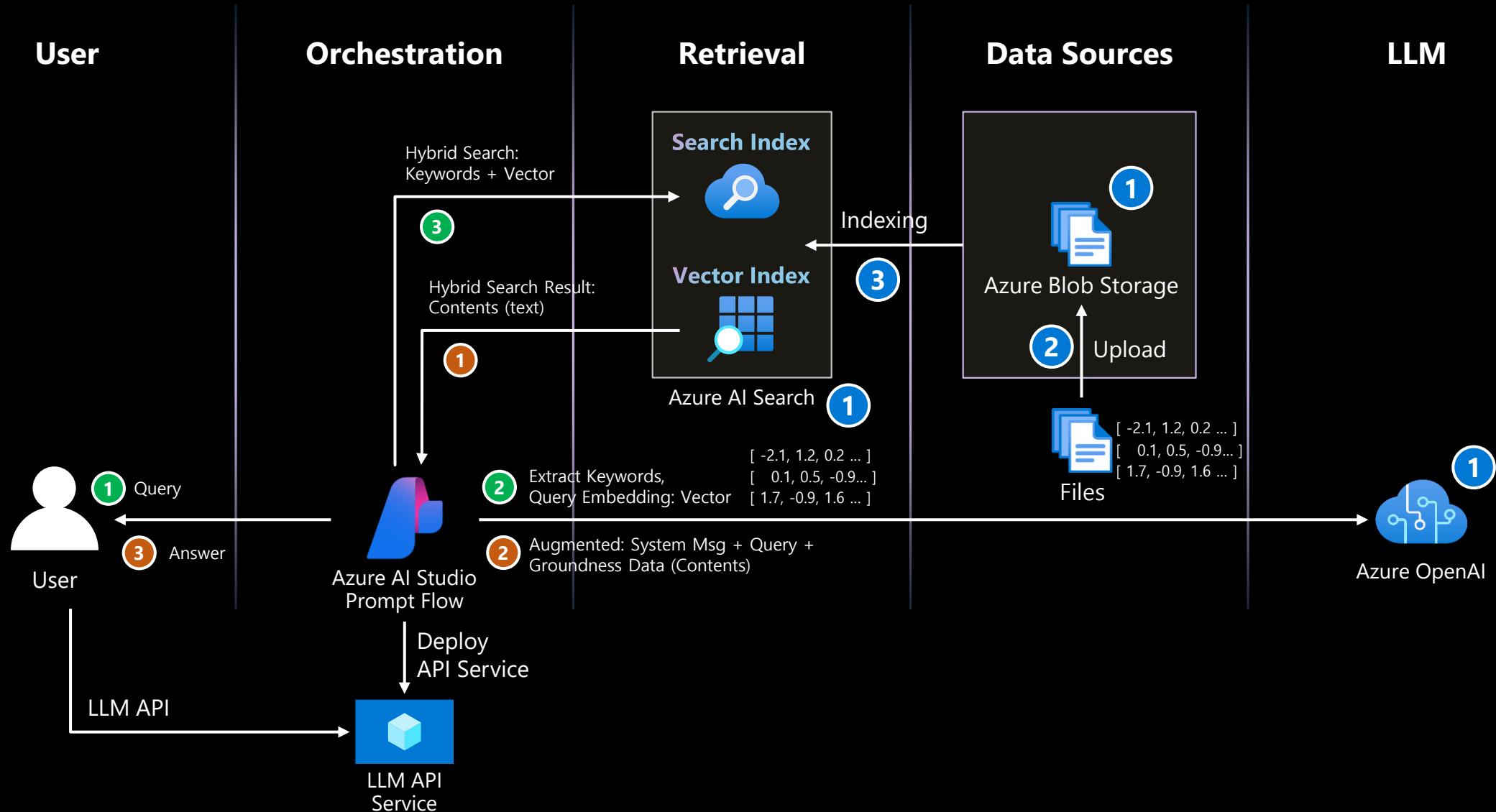
Retrieval Augmented Generation (RAG) – Wikipedia 챗봇 서비스



RAG 아키텍처 만들기 - on your data



LLM App 만들기 - Prompt Flow





프롬프트 엔지니어링

프롬프트란?

■ 프롬프트란?

프롬프트에는 원하는 작업을 수행하기 위해 언어 모델에 전달되는 지침 및 컨텍스트가 포함됩니다.



What is the Prompt in AI?



In AI, a prompt is a piece of text or code that is given to a language model or other machine learning model as input. The model then generates a response based on the prompt and its internal understanding of the data it has been trained on.

Prompts are often used in natural language processing (NLP) applications, such as chatbots, language translation, and text completion. For example, a chatbot might receive a user's message as a prompt, and generate a response based on its understanding of the user's intent and the context of the conversation.

Prompts can also be used to fine-tune language models and other machine learning models for specific tasks. For example, a language model might be trained on a large corpus of text and then fine-tuned on a smaller dataset using a specific prompt or set of prompts. This can improve the model's performance on the target task by giving it more specific training examples.

Overall, prompts are an important tool in AI for both generating responses and fine-tuning machine learning models.

프롬프트 엔지니어링이란?

■ 프롬프트란?

프롬프트에는 원하는 작업을 수행하기 위해 언어 모델에 전달되는 지침 및 컨텍스트가 포함됩니다.

■ 프롬프트 엔지니어링이란?

Prompt Engineering은 다양한 응용 프로그램에 언어 모델을 효율적으로 사용하기 위해 프롬프트를 개발하고 최적화하는 방법입니다.

Chat session

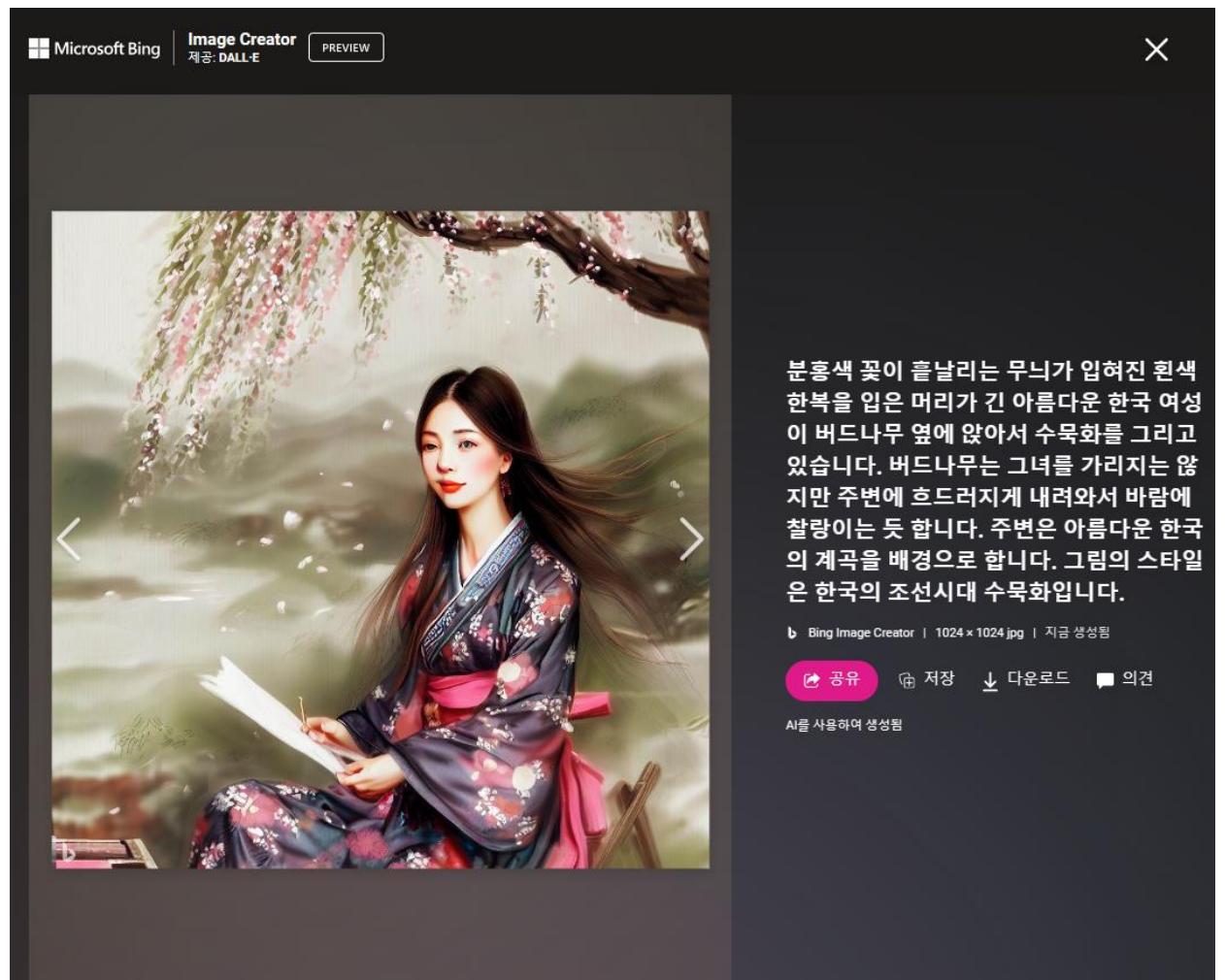
Clear chat View code Show raw JSON

What is Prompt Engineering? 

 Prompt Engineering is a process or technique used in AI and machine learning, particularly in natural language processing (NLP) and language models like OpenAI's GPT-3. It involves designing and refining prompts or input questions to elicit better and more accurate responses from the AI model. The goal of prompt engineering is to optimize the model's performance by crafting effective prompts that help the AI understand the context and deliver the desired output, making AI systems more useful and adaptable to various applications.

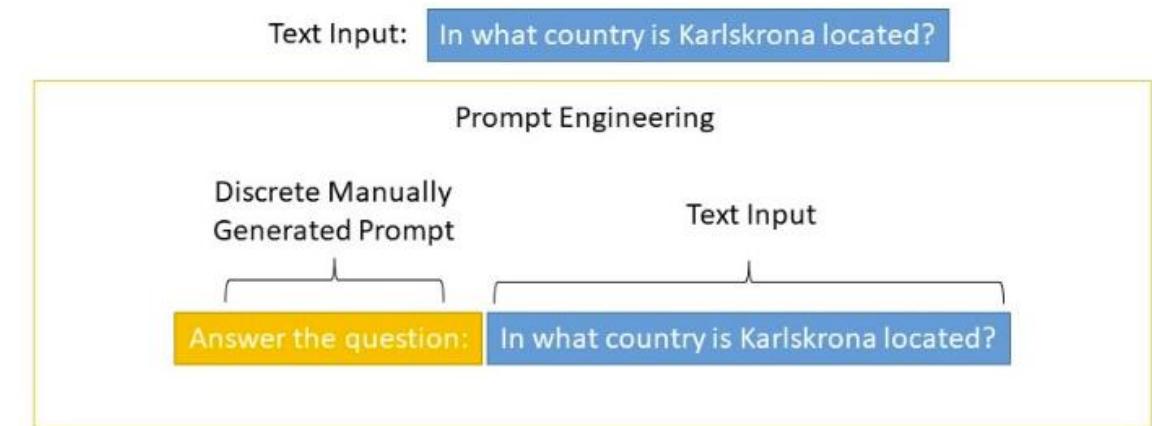
TODAY'S FOCUS

- Focus 1: 텍스트 생성 모델을 위한 프롬프트
- 이미지 생성 모델도 있습니다.
DALL-E



TODAY'S FOCUS

- Focus 2: Prompt => **task** + user's input
 - 작업을 어떻게 해야하는지에 대한 작업 지시 사항
 - task = system + few-shot(user, assistant) + function
 - Chat Completion (json 포맷)



BASIC TECHNIQUES – ROLE PROMPTING

■ Basic Prompt

A What is $1000*1000/20*50$?



To solve this expression, you need to follow the order of operations, which is PEMDAS (Parentheses, Exponents, Multiplication and Division, Addition and Subtraction) from left to right.



Using this order, you would first perform the multiplication and division from left to right, and then the multiplication from left to right:

$$1000 * 1000 / 20 * 50 = (1000 * 1000) / (20 * 50) = 1000000 / 1000 = 1000.$$

Therefore, the answer is 1000.

■ Better Prompt



Now assume you are a math teacher. And you are going to teach PEMDAS (parentheses, exponents, multiplication and division, and addition and subtraction) to your students. What is $1000*1000/20*50$?



When solving this problem, we need to follow the order of operations, which is PEMDAS (parentheses, exponents, multiplication and division, and addition and subtraction).

There are no parentheses or exponents in this expression, so we can move straight to multiplication and division.

First, we perform the multiplication:

$$1000 * 1000 = 1,000,000$$

Next, we perform the division:

$$1,000,000 / 20 = 50,000$$

Finally, we perform the last multiplication:

$$50,000 * 50 = 2,500,000$$

Therefore, the final answer is 2,500,000.

BASIC TECHNIQUES – ROLE PROMPTING

- 해야 할 일(Task)을 “System” 역할에 정의
 - 다음의 형태로 프롬프트를 시작합니다.
 - 이제 당신은 다음과 같이 행동합니다
 - 나는 당신이 다음과 같이 행동하기를 원합니다
 - ChatGPT

■ Better Prompt



Now you are an ML engineer. What is Prompt Engineering?

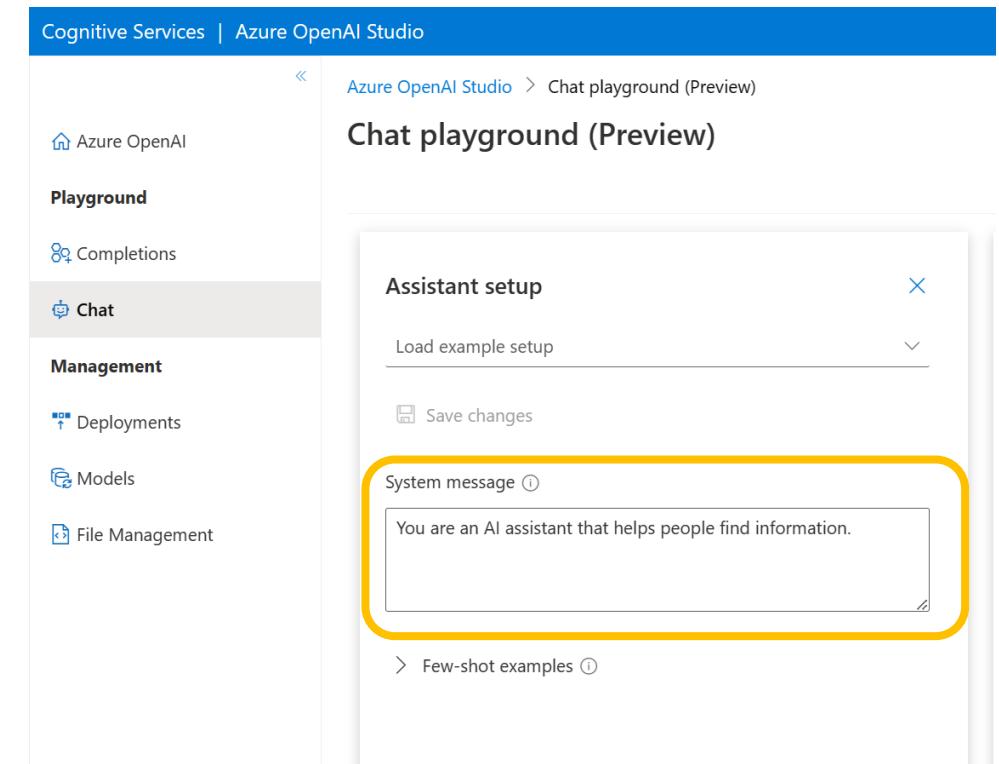


Prompt engineering is a technique used in natural language processing (NLP) to generate text that is more accurate and relevant to a particular task. It involves the creation of a set of prompts, which are inputs given to a language model to elicit a specific response. These prompts can take many forms, including questions, keywords, or sentence fragments, and they are designed to provide the model with the necessary context to generate a coherent and relevant response.



BASIC TECHNIQUES – ROLE PROMPTING

- 해야 할 일(Task)을 “System” 역할에 정의
 - 다음의 형태로 프롬프트를 시작합니다.
 - 이제 당신은 다음과 같이 행동합니다
 - 나는 당신이 다음과 같이 행동하기를 원합니다
 - ChatGPT
 - AOA Chat Playground (채팅)

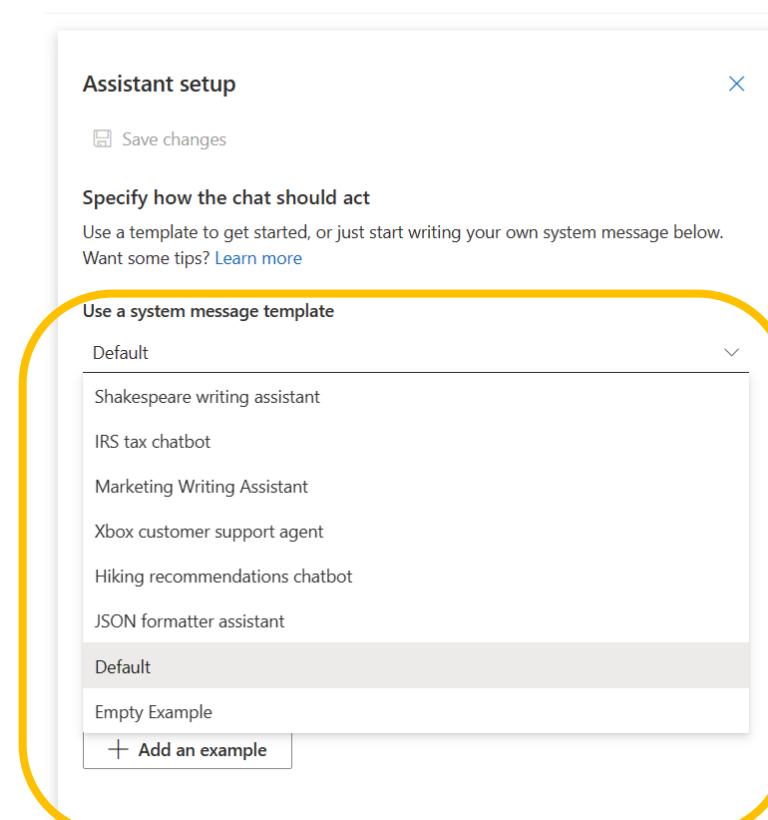


BASIC TECHNIQUES – ROLE PROMPTING

- 해야 할 일(Task)을 “System” 역할에 정의
 - 다음의 형태로 프롬프트를 시작합니다.
 - 이제 당신은 다음과 같이 행동합니다
 - 나는 당신이 다음과 같이 행동하기를 원합니다
 - ChatGPT
 - AOAII Chat Playground (채팅)
 - 템플릿을 활용해 봅시다.

Azure OpenAI Studio > Chat playground (Preview)

Chat playground (Preview)



BASIC TECHNIQUES – ROLE PROMPTING

- 해야 할 일(Task)을 “System” 역할에 정의
 - 다음의 형태로 프롬프트를 시작합니다.
 - 이제 당신은 다음과 같이 행동합니다
 - 나는 당신이 다음과 같이 행동하기를 원합니다
 - ChatGPT
 - AOAII Chat Playground (채팅)
 - 템플릿을 활용해 봅시다.
 - Code로 보기

```
response = openai.ChatCompletion.create(  
    engine="gpt-4", # engine = "deployment_name".  
    messages=[  
        {"role": "system", "content": "You are a helpful assistant."},  
        {"role": "user", "content": "Does Azure OpenAI support customer managed keys?"},  
        {"role": "assistant", "content": "Yes, customer managed keys are supported by Azure OpenAI."},  
        {"role": "user", "content": "Do other Azure Cognitive Services support this too?"}  
    ]  
)
```

BASIC TECHNIQUES

Add more context to the model

- 해야 할 일(Task)을 “System” 역할에 정의
 - 다음의 형태로 프롬프트를 시작합니다.
 - 이제 당신은 다음과 같이 행동합니다
 - Then
 - 정확히(Exactly) 어떤 정보가 제공되나요?
 - 작업을 완료하기 위해 어떤 특정(Specific) 요청이 제공됩니까?

Basic Prompt



How to house train a dog?



House training a dog involves teaching them to eliminate outside and not indoors. Here are some steps you can follow to house train your dog:

1. Establish a Routine: Establish a consistent routine for your dog's meals, playtime, and potty breaks. This will help them learn when they should go outside.
2. Choose a Spot: Choose a specific spot outside where you want your dog to go potty. Take your dog to this spot every time they need to eliminate.
3. Use a Command: Use a consistent command, such as "go potty," every time you take your dog outside to eliminate.
4. Reward Good Behavior: When your dog successfully eliminates outside, praise and reward them with treats or verbal praise.
5. Supervise and Restrict Access: Until your dog is fully house trained, supervise them closely and restrict their access to the rest of the house. Use a crate or baby gates to keep them in a small area.
6. Consistency is Key: Consistency is crucial in house training a dog. Stick to the routine and be patient with your dog.
7. Clean Up Accidents: If your dog has an accident indoors, clean it up immediately with an enzymatic cleaner. Do not punish or scold your dog as this can lead to anxiety and make house training more difficult.

With patience and consistency, you can successfully house train your dog. Remember to praise and reward good behavior, and be patient with your furry friend.

BASIC TECHNIQUES

Add more context to the model

- 해야 할 일(Task)을 “System” 역할에 정의합니다.
 - 다음의 형태로 프롬프트를 시작합니다.
 - 이제 당신은 다음과 같이 행동합니다
 - Then
 - 정확히(Exactly) 어떤 정보가 제공되나요?
 - 작업을 완료하기 위해 어떤 특정(Specific) 요청이 제공됩니까?

Better Prompt

Now assume that you are a puppy trainer. You have one agent who has a new born 3-month-old golden retriever. Please send your agent an email to provide some suggestions on how to house train this puppy.

- throughout the day.
2. Choose a Spot: Choose a specific spot outside where you want your puppy to go potty. Take your puppy to this spot every time they need to eliminate.
3. Use a Command: Use a consistent command, such as "go potty," every time you take your puppy outside to eliminate.
4. Reward Good Behavior: When your puppy successfully eliminates outside, praise and reward them with treats or verbal praise.
5. Supervise and Restrict Access: Until your puppy is fully house trained, supervise them closely and restrict their access to the rest of the house. Use a crate or baby gates to keep them in a small area.

Additionally, it is important to remember that puppies have a limited bladder capacity and will need to go potty frequently. At three months old, your puppy may need to go outside every 2-3 hours, especially after waking up from a nap, after meals, and after playtime.

Best regards,
Puppy Trainer

Regenerate response

BASIC TECHNIQUES – PERSONAS SET PROMPT

- Then
 - Take on the persona of [expert persona].
 - [Verb] [format & length] [objective].
 - The output should include relevant [data].
 - The writing style is [tone of voice] targeted towards [audience].

The image shows a screenshot of an AI interface. On the left, the "Assistant setup" panel is open, titled "Assistant setup". It includes sections for "Save changes", "Specify how the chat should act" (with a note to use a template or write a system message), "Use a system message template" (set to "Shakespeare writing assistant"), and a "System message" box containing a template for a Shakespearean writing assistant. Below this is an "Examples" section with a text input field containing "Please write a short text turning down an invitation to dinner." To the right, the "Chat session" panel is visible, showing a user message "Write a poem to my husband to celebrate his birthday" and an AI response: "To thee, my love, upon thy natal day, A wondrous celebration I do offer, Through words that will their adulation say, Each sentiment within my heart, they proffer. Thou art the sun that rises in the morn, Illuminating all that lies beneath, With radiant warmth, thy love hath been reborn, In every smile thou giveth, joy bequeath. No fairer treasure could my eyes behold, Than thee, my dearest partner, strong and true, In every precious year, as life unfolds, I find new reasons to give thanks for you. And so, on this occasion of thy birth, I do extol thy virtues and their worth." At the bottom, there is a "User message" input field with the placeholder "Type user query here. (Shift + Enter for new line)".

BASIC TECHNIQUES – PERSONAS SET PROMPT

- Then
 - Take on the persona of [expert persona].
 - ChatGPT가 채택해야 하는 특정 페르소나를 정의해 주면, 해당 주제 전문가의 지식과 권위를 반영하여 응답을 생성해 줍니다.
 - 예시
 - 전문 소프트웨어 엔지니어
 - 영양사
 - 전문 여행 블로거
 - 역사 교수
 - 마케팅 전략가

The screenshot shows two side-by-side interfaces. On the left is the 'Assistant setup' window, which allows users to define how the chat should act by specifying a system message template. The template chosen is 'Shakespeare writing assistant', and the system message describes it as a Shakespearean writing assistant who speaks in a Shakespearean style, helping people come up with creative ideas and content like stories, poems, and songs that use Shakespearean style of writing style, including words like "thou" and "hath". Below this, there's an 'Examples' section where a user can type in a prompt and see the AI's response. A sample prompt is shown: 'Please write a short text turning down an invitation to dinner.' The AI's response is: 'Dearest, Regretfully, I must decline thy invitation. Prior engagements call me hence. Apologies.' At the bottom of this window is a button labeled '+ Add an example'. On the right is the 'Chat session' window, showing a conversation between a user and the AI. The user has typed 'Write a poem to my husband to celebrate his birthday'. The AI has responded with a poem in Shakespearean style:

To thee, my love, upon thy natal day,
A wondrous celebration I do offer,
Through words that will their adulation say,
Each sentiment within my heart, they proffer.

Thou art the sun that rises in the morn,
Illuminating all that lies beneath,
With radiant warmth, thy love hath been reborn,
In every smile thou giveth, joy bequeath.

No fairer treasure could my eyes behold,
Than thee, my dearest partner, strong and true,
In every precious year, as life unfolds,
I find new reasons to give thanks for you.

And so, on this occasion of thy birth,
I do extol thy virtues and their worth.

Below the AI's response is a 'User message' input field with placeholder text: 'Type user query here. (Shift + Enter for new line)'.

BASIC TECHNIQUES – PERSONAS SET PROMPT

- Then
 - [Verb] [format & length] [objective].
 - 마법을 걸어주는 효과를 줍니다!
 - 무엇을 해야 하는지, 원하는 출력의 형식과 길이, 전반적인 목표에 대해 ChatGPT에 지시합니다.
 - 이렇게 하면 생성된 콘텐츠가 특정 요구 사항을 충족하는지 확인할 수 있습니다.
 - 예시
 - 🕉 명상의 이점을 설명하는 500단어 기사 작성
 - 💼 성공적인 전자 상거래 상점을 설정하는 과정을 3 단락으로 요약하십시오.
 - 🏰 아시아 최고의 여행지를 보여주는 10슬라이드 PowerPoint 프레젠테이션 만들기

The screenshot shows the ChatGPT interface. On the left, the 'Assistant setup' panel is open, showing a template for a 'Shakespeare writing assistant'. It includes a system message template and examples of how to respond to user prompts like 'Please write a short text turning down an invitation to dinner.' On the right, the 'Chat session' panel shows a conversation where the AI has generated a poem for a husband's birthday, starting with 'To thee, my love, upon thy natal day,' followed by three stanzas of Shakespearean poetry, and concluding with 'And so, on this occasion of thy birth, I do extol thy virtues and their worth.'

BASIC TECHNIQUES – PERSONAS SET PROMPT

- Then
 - The output should include **relevant [data]**.
 - 정확성이 핵심! Grounding LLMs!!
 - 프롬프트에 **관련 데이터**를 제공함으로써, ChatGPT는 신뢰할 수 있는 정보로 생성
 - 예시
 - 글로벌 소프트웨어 개발 동향에 대한 통계
 - 다양한 식품의 영양정보
 - 2022년 가장 많이 방문한 관광지 TOP 10
 - 사실 같은 이야기를 생성 요청하기 보다는 **실제 관련 데이터 기반**의 접근 방식입니다!

The screenshot shows the AI interface for setting up a persona and interacting with it.

Assistant setup: A modal window titled "Assistant setup" is open. It includes a "Save changes" button, a section for specifying how the chat should act (using a template or writing your own system message), and a "System message" section. The system message template is set to "Shakespeare writing assistant". The message content is: "You are a Shakespearean writing assistant who speaks in a Shakespearean style. You help people come up with creative ideas and content like stories, poems, and songs that use Shakespearean style of writing style, including words like 'thou' and 'hath'." Below this is an "Examples" section with a placeholder: "Please write a short text turning down an invitation to dinner." An "Assistant" response is shown: "Dearest, Regretfully, I must decline thy invitation. Prior engagements call me hence. Apologies." At the bottom is a "+ Add an example" button.

Chat session: The main interface shows a "Chat session" tab. It has options to "Clear chat", "View code", and "Show raw JSON". A user message box contains the text: "Write a poem to my husband to celebrate his birthday". Below it, the AI response is displayed in three parts:

- To thee, my love, upon thy natal day,
A wondrous celebration I do offer,
Through words that will their adulation say,
Each sentiment within my heart, they proffer.
- Thou art the sun that rises in the morn,
Illuminating all that lies beneath,
With radiant warmth, thy love hath been reborn,
In every smile thou giveth, joy bequeath.
- No fairer treasure could my eyes behold,
Than thee, my dearest partner, strong and true,
In every precious year, as life unfolds,
I find new reasons to give thanks for you.

A "User message" input field at the bottom is labeled "Type user query here. (Shift + Enter for new line)".

BASIC TECHNIQUES – PERSONAS SET PROMPT

- Then
 - The writing style is [tone of voice] targeted towards [audience].
 - 마지막으로 목소리 톤과 대상 고객을 지정하면 ChatGPT의 응답이 의도한 독자에 맞게 조정
 - 이를 통해 청중과 공감하고 브랜드의 목소리와 일관성을 유지하는 콘텐츠를 만들 수 있습니다.
 - 예시
 - 🎵 청소년을 위한 대화 및 참여
 - 💼 비즈니스 임원을 위한 전문적이고 권위 있게
 - 🎓 대학생을 위한 교육 및 정보 제공

The screenshot shows the AI interface divided into two main sections: 'Assistant setup' on the left and 'Chat session' on the right.

Assistant setup: This section allows users to define how the AI should act. It includes a 'Save changes' button, a 'Specify how the chat should act' area with tips for using a system message template, and a 'System message' input field containing a template for a Shakespearean writing assistant. Below this is an 'Examples' section where users can add text for the AI to mimic, followed by a 'User' input field and an 'Assistant' response field.

Chat session: This section shows a conversation between a user and the AI. The user has typed 'Write a poem to my husband to celebrate his birthday'. The AI has generated a poem in a Shakespearean style:

```
To thee, my love, upon thy natal day,  
A wondrous celebration I do offer,  
Through words that will their adulation say,  
Each sentiment within my heart, they proffer.  
  
Thou art the sun that rises in the morn,  
Illuminating all that lies beneath,  
With radiant warmth, thy love hath been reborn,  
In every smile thou giveth, joy bequeath.  
  
No fairer treasure could my eyes behold,  
Than thee, my dearest partner, strong and true,  
In every precious year, as life unfolds,  
I find new reasons to give thanks for you.  
  
And so, on this occasion of thy birth,  
I do extol thy virtues and their worth.
```

At the bottom, there is a 'User message' input field for the user to type their next query.



ADVANCED TECHNIQUES – SHOT PROMPTING

- **Zero shot**
 - 지시만 있습니다. 해결하고자 하는 입력만 모델에 제공합니다.
- **One shot:**
 - 해결하기 위해 유사 형태의 예시를 **1개** 제공합니다.
그리고 해결하고자 하는 입력을 모델에 제공합니다.
- **Few shot:**
 - 해결하기 위해 **몇 가지** 예시를 제공합니다.
 - 그리고 해결하고자 하는 입력을 모델에 제공합니다.

ADVANCED TECHNIQUES – SHOT PROMPTING

Zero-shot

The screenshot shows the '도우미 설정' (Assistant Settings) screen. It includes a '시스템 메시지' (System Message) input field containing the Korean sentence '고양이가 매트 위에 앉아 있습니다.' (The cat is sitting on the mat.) and a '시스템 메시지' (System Message) input field containing the English sentence 'The cat is sitting on the mat.'. Below these are sections for adding examples and a 'User message' input field.

One-shot

The screenshot shows the '도우미 설정' (Assistant Settings) screen. It includes a '시스템 메시지' (System Message) input field containing the English sentence 'You are an Xbox customer support agent whose primary goal is to help users with issues they are experiencing with their Xbox devices. You are friendly and concise. You only provide factual answers to queries, and do not'. Below this is a 'User message' input field with the Korean sentence 'Nintendo 얼마야?' (How much is a PS5?). The '도우미:' (Assistant) input field contains the response 'I apologize, but I do not have information about the prices of other gaming devices such as the PS5. My primary focus is to assist with issues regarding Xbox.'

ADVANCED TECHNIQUES – SHOT PROMPTING

Few-shot

You are a helpful assistant to help users determine whether a given post is related to a specific incident. An incident is a collection of related alerts that are generated by a security event. It is a security event that requires investigation. The specific incident is Incident: Windows update breaks Azure AD Application Proxy IWA (KCD) SSO. Then you will be given some context, please tell the users if similar alert exists in the given context to the above incident.

Feedback: After installing the latest update, I'm seeing issues with Azure AD Application Proxy.

A: True.

Feedback: We are facing trouble with the latest patch we deploy over our infrastructures.

A: False.

Feedback: After installing kb5007206, sso Azure AD Appproxy breaks.

A:

True.

Chat session

Clear chat View code Show raw JSON

Now assume that you are a Windows feedback analyst. Windows users provide feedback to Microsoft about Windows updates via feedback Hub. I want you to determine whether a given feedback is related to Windows Updates. I will give you some examples first, and then give you the undetermined feedback.

Sure, please provide the examples and I will help you determine if the undetermined feedback is related to Windows Updates.

Feedback: My Printer is not working.

A: False.

Feedback: I installed Windows Updates and Printer stopped working.

A: True.

Feedback: I installed driver update and my computer is blue screening.

A: False.

Feedback: I applied April Updates for Windows and Bluetooth is not working.

A: True.

Feedback: Today, after Office Updates got installed on my machine, I cannot open Outlook.

A:

False.

Feedback: After installing kb1234567, I cannot open File Explorer.
A:

True.

ADVANCED TECHNIQUES – CHAIN-OF-THOUGHT (CoT)

■ Chain-of-Thought (CoT) Prompting:

- **사고의 연결 프롬프트**는 이전 프롬프트와 사용자의 응답을 기반으로 일련의 프롬프트를 생성하는 자연어 처리(NLP) 및 기계 학습(ML)의 기술입니다.
- Model이 문제를 해결하기 위해 [중간 추론 단계]을 거쳐 최종적인 답변을 구하도록 도움

■ 두 가지 방법:

- **Zero-shot CoT**
 - 모델이 응답할 때, 작업에 대해 추론하도록 지시합니다.
 - "Let's think step by step".
- **Few-shot CoT**
 - 사전에 질문이 포함된 수동 데모와 답변으로 이어지는 추론 체인을 제공합니다.

ADVANCED TECHNIQUES – CHAIN-OF-THOUGHT (CoT)

■ Basic Prompt

 Take the last letters of the words in "Tylor Swift" and concatenate them.

 Taking the last letters of the words in "Taylor Swift" and concatenating them gives us "rltf".

■ Better Prompt



Let's think step by step



Sure, let's break it down step by step.



Take the last letters of the words in "Tylor Swift" and concatenate them.



Alright, let's break it down:

Step 1: Identify the words in the given phrase: "Tylor Swift"

Step 2: Take the last letter of each word: "r" from "Tylor" and "t" from "Swift"

Step 3: Concatenate the last letters: "rt"

Therefore, taking the last letters of the words in "Tylor Swift" and concatenating them gives us "rt".

ADVANCED TECHNIQUES – VALIDATION

- 모델을 유효성 검사기로 사용
 - 대답을 검증해봐:
 - "Are you sure?"
 - 제안된 답변 확인:
 - Cunningham's Law
 - 인터넷에서 옳은 답변을 얻는 가장 좋은 방법은 질문을 하는 것이 아니라, 잘못된 답변을 올리는 것

■ Better Prompt



Take the last letters of the words in "Veronica C" and concatenate them. My answer is "ve". Is that correct?



No, your answer "ve" is not correct. Let's break down the process to find the correct answer:

The words "Veronica C" consist of two words: "Veronica" and "C". To find the last letter of each word, we take the last letter of "Veronica," which is "a," and the last letter of "C," which is "C" itself.

Concatenating these two last letters gives us "aC." Therefore, the correct answer is "aC," not "ve."





The End ?

DEFENSE AGAINST THE DARK ARTS 101

- Challenges:
 - Prompt Injection
 - Prompt Leaking
 - Jailbreaking
 - Hallucination



SUMMARY

■ How to write a good prompt?

- Define the system role
- Personas set prompt
- Provide more context to the model
- Few-shot
- Chain-of-thought

먼저 프롬프트 엔지니어링을 해보자 !

그리고 나서 Fine-tuning을 고려하자.



Azure OpenAI Studio Playground를 활용한 실습

Azure OpenAI Studio

The screenshot shows the Azure OpenAI Studio interface with the "Chat playground" selected in the sidebar. The main area displays a comparison chart titled "4월 8월 비교해서 표로 정리해" (Compare April and August and summarize in a table). The chart compares various aspects of the two months, such as month name, week sequence, day count, start day, end day, and notable events.

Setup

Prompt Add your data

Gain insights into your own data source. Your data is stored securely in your Azure subscription. [Learn more about how your data is protected.](#)

Data source: Search Resource: Azure AI Search cog-search-test5
Index: index-test2
Advanced settings >

[Remove data source](#)

Clear chat | Chat capabilities | View code | Show JSON

Import setup | Export setup | Show panels

Citations

April

April is the fourth month of the year in the Julian and Gregorian calendars, and comes between March and May. It is one of four months to have 30 days. April always begins on the same day of week as July, and additionally, January in leap years. April always ends on the same day of the week as December. April's flowers are the Sweet Pea and Daisy. Its birthstone is the diamond. The meaning of the diamond is innocence. The Month April comes between March and May, making it the fourth month of the year. It also comes first in the year out of the four months that have 30 days, as June, September and November are later in the year. April begins on the same day of the week as July every year and on the same day of the week as January in leap years. April ends on the same day of the week as December every year, as each other's last days are exactly 35 weeks (245 days) apart. In common years, April starts on the same day of the week as October of the previous year, and in leap years, May of the previous year. In common years, April finishes on the same day of the week as July of the previous year, and in leap years, February and October of the previous year. In common years immediately after other common years, April starts on the same day of the week as January of the previous year, and in leap years and years immediately after that, April finishes on the same day of the week as January of the previous year. In years immediately before common years, April starts on the same day of the week as September and December of the following year, and in years immediately before leap years, June of the following year. In years immediately before common years, April finishes on the same day of the week as September of the following year, and in years immediately before leap years, March and June of the following year. April is a spring month in the Northern Hemisphere and an autumn/fall month in the Southern Hemisphere. In each hemisphere, it is the seasonal equivalent of October in the other. It is unclear as to where April got its name. A common theory is that it comes from the Latin word "aperire", meaning "to open", referring to flowers opening in spring. Another theory is that the name could come from Aphrodite, the Greek goddess of love. It was originally the second month in the old Roman Calendar, before the start of the new year was put to January 1. Quite a few festivals are held in this month. In many Southeast Asian cultures, new year is celebrated in this month (including Songkran). In Western Christianity, Easter can be celebrated on a Sunday between March 22 and April 25. In Orthodox Christianity, it can fall between April 4 and May 8. At the end of the month, Central and Northern European cultures celebrate Walpurgis Night on April 30, marking the transition from winter into summer.

Type user query here. (Shift + Enter for new line)

563/128000 tokens to be sent



Thank you



RAG 서비스 구성 – Azure OpenAI

RAG 챗봇

The screenshot shows the Azure OpenAI Studio interface with the "Chat playground" selected. The left sidebar includes options like "Azure OpenAI", "Playground", "Chat", "Completions", "DALL-E", "Assistants (Preview)", "Management", "Deployments", "Models", "Data files", "Quotas", and "Content filters (Preview)". The main area has tabs for "Setup" and "Add your data". The "Add your data" tab displays information about the data source (Search Resource: Azure AI Search, cog-search-test5, Index: index-test2) and provides a "Remove data source" button. On the right, there are sections for "Configuration" (Deployment set to gpt-4o), "Session settings" (Past messages included: 10, Current token count: 10), and "Input tokens progress indicator" (1063/128000). The central content area shows two sections: "4월 (April)" and "8월 (August)". The April section discusses the month's position, days, holidays, and birthdates. The August section discusses its position, days, and birthdates.

Setup

Add your data

Data source: Search Resource: Azure AI Search cog-search-test5
Index: index-test2
Advanced settings >

Remove data source

4월 8월 비교해

4월 (April)

- 위치: 4월은 1년 중 네 번째 달로, 3월과 5월 사이에 위치합니다¹.
- 일수: 4월은 30일로 구성되어 있습니다¹.
- 계절: 북반구에서는 봄, 남반구에서는 가을에 해당합니다¹.
- 시작 요일: 4월은 매년 7월과 같은曜일에 시작하며, 윤년에는 1월과 같은曜일에 시작합니다¹.
- 종료 요일: 4월은 매년 12월과 같은曜일에 끝납니다¹.
- 꽃: 4월의 꽃은 스위트피와 데이지입니다¹.
- 탄생석: 4월의 탄생석은 다이아몬드로, 순수함을 상징합니다¹.
- 고정 기념일:
 - 4월 1일: 만우절
 - 4월 2일: 세계 자폐증 인식의 날
 - 4월 7일: 세계 보건의 날
 - 4월 22일: 지구의 날¹.

8월 (August)

- 위치: 8월은 1년 중 여덟 번째 달로, 7월과 9월 사이에 위치합니다².
- 일수: 8월은 31일로 구성되어 있습니다².
- 계절: 북반구에서는 여름, 남반구에서는 겨울에 해당합니다².
- 시작 요일: 8월은 일반적으로 다른 달과 같은曜일에 시작하지 않지만, 윤년에는 2월과

RAG 챗봇 Deployment

The screenshot shows the Azure OpenAI Studio interface with the 'Chat playground' setup page in the foreground. A modal window titled 'Deploy to a web app' is open, providing instructions about Azure Active Directory authentication and deployment timing. It includes fields for 'Name', 'Subscription', 'Resource group', 'Location', 'Pricing plan', and a checkbox for enabling chat history. The background shows the configuration panel with tabs for 'Deployment' (selected), 'Parameters', 'Session settings', and 'Input tokens progress indicator'. The deployment status shows 'gpt-4o' and a current token count of 10.

Deploy to a web app

Your web app will be configured with Azure Active Directory authentication enabled. It may take a few minutes to apply after deployment completes, during which time you will not be able to chat in the app. Please wait 10 minutes, then reload the app and log in to begin chatting.

Pick your configurations to deploy a web app. [Learn more about web apps](#)

Create a new web app Update an existing web app

Name *
Enter a name for the web app

Subscription *
Select a subscription

Resource group *
Select a resource group

Location *
Select a location

Pricing plan *
Select a pricing plan

Enable chat history in the web app

Enabling chat history will incur CosmosDB usage to your account [View Pricing](#)

Web apps will incur usage to your account [View Pricing](#)

Deploy Cancel

Configuration

Deployment Parameters

Deployment *
gpt-4o

Session settings

Past messages included *
10

Current token count *
1063/128000

Input tokens progress indicator
1063/128000

Wikipedia 샘플 데이터 (text-embedding-ada-002)

The screenshot shows a code editor interface with a dark theme. On the left, there is a vertical toolbar with various icons. The main area displays a CSV file named "vector_sample.csv". The file contains several rows of data, each representing a Wikipedia article. The columns are labeled: id, url, title, title_vector, content_vector, and vector_id. The content_vector column contains long strings of numerical values representing text embeddings. The file path shown in the top bar is "C: > Users > hyounsookim > Downloads > vector_sample.csv > data". The right side of the screen shows a large preview of the content of the CSV file, which includes text descriptions of various art forms like Art, Craft, and Plastic art.

id	url	title	title_vector	content_vector	vector_id	
1				-0.0072430833242833614, -0.007268654648214579, -0.010810285806655884, -0.010714393109083176, -0.00041473525925539434, 0.0049320729449391365, -0.020546576008200645, 0.01846251077950008, 0.019510935992002487, 0.03966115415096283, 0.031427182257175446, 0.025353988632559776, -0.01779765635728836, 0.014639594592154026, -0.01146235503256321, 0.04063286632299423, 0.01240210235118866, 0.00227904599159956, -0.013693454675376415, -0.006878691725432873, 5.1292521675350145e-05, -0.0031181059312075377, 0.00602524820715189, -0.014409452676773071, -0.056103531271219254, 0.016813158988952637, -0.016301732510328293, 0.0031340878922492266, -0.005871820263564587, 0.017145587131381035, 0.027591481804847717, -0.013603955507278442, 0.005699213594198227, 0.0008382606320083141, -0.009493362158536911, -0.0018347440054640174, 0.010439501143991947, -0.005791909527033567, -0.01128335576504469, -0.0418858639895916, 0.0004095410695299506, -0.007805652916431427, -0.00907143484801054, -0.01439666748046875, 0.0087198289112566, -0.007882366888225079, -0.009295184165239334, -0.014115381985902786, 0.0131692430044775, 0.04099086672067642, -0.017529156059026718, -0.010663250461220741, -0.0037717733066529036, -0.015739163383841515, -0.011187463067471981, 0.008342651650309563, -0.000827073177784066, -0.008969149552285671, -0.030225329101085663], 1	384	6, https://simple.wikipedia.org/wiki/Art , Art, "Art is a creative activity that expresses imaginative or technical skill. It produces a product, an object. Art is a diverse range of human activities in creating visual, performing artifacts, and expressing the author's imaginative mind. The product of art is called a work of art, for others to experience.
385						
386		Some art is useful in a practical sense, such as a sculptured clay bowl that can be used. That kind of art is sometimes called a craft.				
387						
388		Those who make art are called artists. They hope to affect the emotions of people who experience it. Some people find art relaxing, exciting or informative. Some say people are driven to make art due to their inner creativity.				
389						
390		""The arts"" is a much broader term. It includes drawing, painting, sculpting, photography, performance art, dance, music, poetry, prose and theatre.				
391						
392		Types of art				
393						
394		Art is divided into the plastic arts, where something is made, and the performing arts, where something is done by humans in action. The other division is between pure arts, done for themselves, and practical arts, done for a practical purpose, but with artistic content.				
395						
396		Plastic art				
397		Fine art is expression by making something beautiful or appealing to the emotions by visual means: drawing, painting, printmaking, sculpture				
398		Literature: poetry, creative writing				
399		Performing art				
400		Performing art including drama are (expression using the body: dance, acting, singing)				
401		Auditory art (expression by making sounds): music, singing				
402		Practical art				
403		Culinary art (expression by making flavors and tastes): cooking				

Azure Blob Storage에 업로드

Microsoft Azure Search resources, services, and docs (G+) 8 ? User icon studydev@hotmail.com 기본 디렉터리 (STUDYDEVHOTM...)

Home > Storage accounts > hsromptthonstorage

Storage accounts

기본 디렉터리 (studydevhotmail.onmicrosoft.com)

+ Create ⏪ Restore ...

Filter for any field...

Name	...
aistudiojapan22703171235	...
aistudiouk4485744610	...
aoairagsource	...
azureaistudio2760120465	...
azurestudydevsite	...
cs11003200230b56644	...
cshstest	...
hsromptthonstorage	...
kamo202311	...
mleastus1674541698	...

Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser Storage Mover

Containers

File shares

hsromptthonstorage | Containers

Storage account

Search Container Change access level ⏪ Restore containers Refresh Delete Give feedback

Show deleted containers

Name	Last modified	Anonymous access l...	Lease state
\$logs	2024. 1. 24. 오전 1:3...	Private	Available
azure-data	2024. 1. 24. 오전 1:3...	Private	Available
vector-data	2024. 1. 24. 오전 3:0...	Private	Available

Home > Storage accounts > hsromptthonstorage | Containers > vector-data

Container

Search Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots Create snapshot ...

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: vector-data

Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
vector_sample.csv	2024. 1. 24. 오전 3:5...	Hot (Inferred)		Block blob	3.71 MiB	Available

Shared access tokens Access policy Properties Metadata

Azure AI Search에 색인화

The image shows two side-by-side screenshots of the Azure AI Search interface.

Left Screenshot: A modal window titled "Vector algorithm" is open. It shows the configuration for a new vector algorithm named "vector-config-1706033797636". The "Kind" is set to "hnsw". Under "Kind parameters", the "Bi-directional link count (m)" is set to 4, "efConstruction" to 400, "efSearch" to 500, and the "Similarity metric" is set to "cosine". At the bottom are "Save" and "Cancel" buttons.

Right Screenshot: The main index configuration page for "azureblob-index". The "Search explorer" tab is selected. A JSON query editor contains the following query:

```
1 {  
2   "search": "April"  
3 }
```

Below the query editor, the "Results" section displays the search results. One result is shown in full:

```
1 {  
2   "@odata.context": "https://hs-ai-search.search.windows.net/indexes('azureblob-index')/$metadata#docs(*)",  
3   "value": [  
4     {  
5       "@search.score": 3.9235651,  
6       "id": "1",  
7       "url": "https://simple.wikipedia.org/wiki/April",  
8       "title": "April",  
9       "text": "April is the fourth month of the year in the Julian and Gregorian calendars, and comes between Ma  
10      },  
11      {  
12        "@search.score": 2.3071764,  
13        "id": "53",  
14        "url": "https://simple.wikipedia.org/wiki/Angola",  
15        "title": "Angola",  
16      }  
17    ]  
18  }
```

At the top right of the main interface, there is a date indicator "2023-10-01-Preview" and a "View" dropdown.

Azure OpenAI Studio에 on your data 추가

Add data

Select or add data source

Your data source is used to ground the generated results with your data. Select an existing data source or create a new data connection with Azure Blob storage, databases, search, URLs, or local files as the source the grounding data will be built from.

[Learn more about data privacy and security in Azure AI.](#)

Select data source *

Azure AI Search

Subscription *

AppSub

Azure AI Search service ⓘ *

hs-ai-search

Azure AI Search Index ⓘ *

azureblob-index

Add vector search to this search resource.

Embedding model

To use a vector model as part of your data, select one below:

Select an embedding model ⓘ *

Azure OpenAI - text-embedding-ada-002

I acknowledge that connecting to an Azure AI Search account will incur usage to my account. * View Pricing

Next

Cancel

Add data

Data source

Data field mapping

Data management

Review and finish

Index data field mapping

For the best results, tell us more about the fields in your index. Your content data field(s) will be used to ground the model on your data. File name, title, and URL are used to display more information when a document is referenced in the chat.

[Learn more about data privacy and security in Azure AI.](#)

Content data ⓘ

text

File name ⓘ

Select file name field

Title ⓘ

title

URL ⓘ

url

Vector Fields ⓘ

title_vector, content_vector

Back

Next

Cancel

Azure OpenAI Studio에 on your data 추가

Add data

X

- Data source
- Data field mapping
- Data management
- Review and finish

Data management

Set up specific configurations for your data and how the model will respond to requests.

[Learn more about data privacy and security in Azure AI.](#)

Search type i *

Hybrid (vector + keyword)

I acknowledge that adding vector embeddings will incur usage to my account.* [View Pricing](#)

[Azure OpenAI](#)

Playground

[Chat](#)[Completions](#)[DALL-E \(Preview\)](#)

Management

[Deployments](#)[Models](#)[Data files](#)[Quotas](#)[Content filters \(Preview\)](#)

Chat playground

[Deploy to](#)[Import setup](#)[Export setup](#)[Show panels](#)

Assistant setup

Prompt [Add your data \(preview\)](#)

Gain insights into your own data source. Your data is stored securely in your Azure subscription. [Learn more about your data is protected.](#)

Data source: Search Resource:

Azure AI Search hs-ai-search

Index:

azuredblob-index

Advanced settings ▾

 Limit responses to your data content ⓘ

Strictness (1-5) ⓘ

 1

Retrieved documents (3-20) ⓘ

 5 [Remove data source](#)[Replay chat](#) [Clear chat](#) [Playground Settings](#) [View code](#) [Show JSON](#)[Replay chat](#)[Clear chat](#)[Playground Settings](#)[View code](#)[Show JSON](#)[...](#)

4월이 왜 April 이라고 하는지 기원을 요약해줘.

1 4월의 이름인 "April"은 라틴어 "aperire"에서 유래되었을 가능성이 있습니다. "aperire"는 봄에 꽃이 피는 것을 의미하는 단어로, 4월의 이름이 이와 관련이 있을 수 있습니다. 또 다른 이론으로는 그리스 신화의 여신 아프로디테(Aphrodite)에서 유래되었을 수도 있습니다. 4월은 로마 달력에서는 2월로 시작했으며, 1월 1일이 새해의 시작일로 정해지기 전까지는 8월로 알려져 있었습니다. 4월은 봄철에 해당하는 달로, 북반구에서는 봄을, 남반구에서는 가을을 의미합니다.¹

**출처: 1 **

▽ 1 references

1 Citation 1

8월의 기원을 알려줘.

1 8월은 로마 황제 아우구스투스 카이사르의 이름을 따서 지어진 달입니다. 8월은 31일로 이루어져 있으며, 북반구에서는 여름철에 해당하고, 남반구에서는 겨울철에 해당합니다. 8월은 일반 연도에서는 다른 달과 같은曜日로 시작하지 않지만, 윤년에서는 2월과 같은曜日로 시작합니다. 8월은 매년 11월과 같은曜日로 끝납니다.

**출처: 1 **

▽ 1 references

1 Citation 1

Type user query here. (Shift + Enter for new line)

Citations

August

August (Aug.) is the eighth month of the year in the Gregorian calendar, coming between July and September. It has 31 days. It is named after the Roman emperor Augustus Caesar. August does not begin on the same day of the week as any other month in common years, but begins on the same day of the week as February in leap years. August always ends on the same day of the week as November. The Month This month was first called Sextilis in Latin, because it was the sixth month in the old Roman calendar. The Roman calendar began in March about 735 BC with Romulus. October was the eighth month. August was the eighth month when January or February were added to the start of the year by King Numa Pompilius about 700 BC. Or, when those two months were moved from the end to the beginning of the year by the decemvirs about 450 BC (Roman writers disagree). In 153 BC January 1 was determined as the beginning of the year. August is named for Augustus Caesar who became Roman consul in this month. The month has 31 days because Julius Caesar added two days when he created the Julian calendar in 45 BC. August is after July and before September. August, in either hemisphere, is the seasonal equivalent of February in the other. In the Northern hemisphere it is a summer month and it is a winter month in the Southern hemisphere. No other month in common years begins on the same day of the week as August, but August begins on the same day of the week as February in leap years. August ends on the same day of the week as November every year, as each other's last days are 13 weeks (91 days) apart. In common years, August starts on the same day of the week as March and November of the previous year, and in leap years, June of the previous year. In common years, August finishes on the same day of the week as March and June of the previous year, and in leap years, September of the previous year. In common years immediately after other common years, August starts on the same day of the week as February of the previous year. In years immediately before common years, August starts on the same day of the week as May of the following year, and in years immediately before leap years, October of the following year. In years immediately before common years, August finishes on the same day of the week as May of the following year, and in years immediately before leap years, February and October of the following year. August observances Fixed observances and events August 1 National Day of Switzerland August 1 Independence Day (Benin) August 1 Emancipation Day (Bermuda Guyana Jamaica)



Thank you