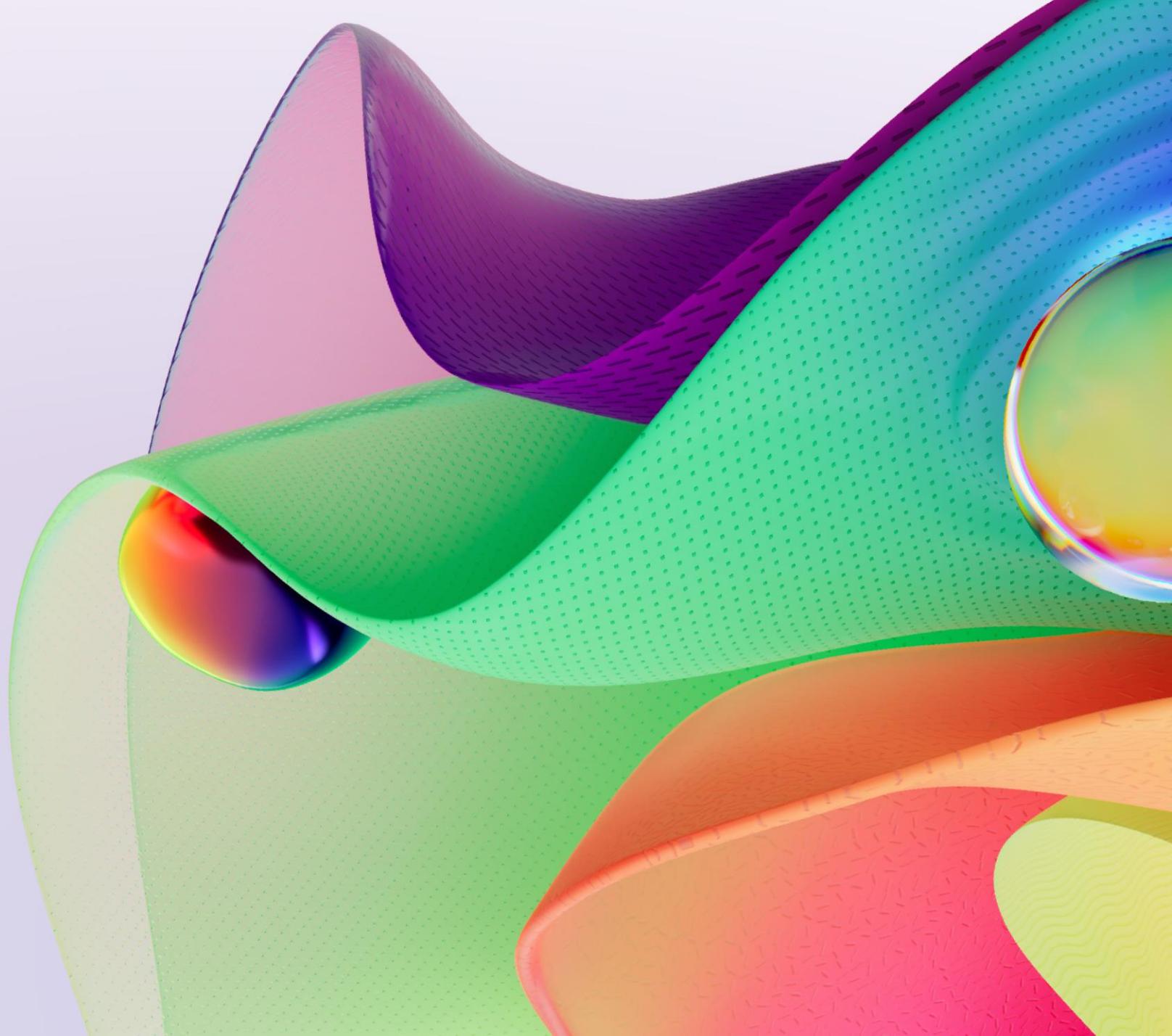




Microsoft AI Tour

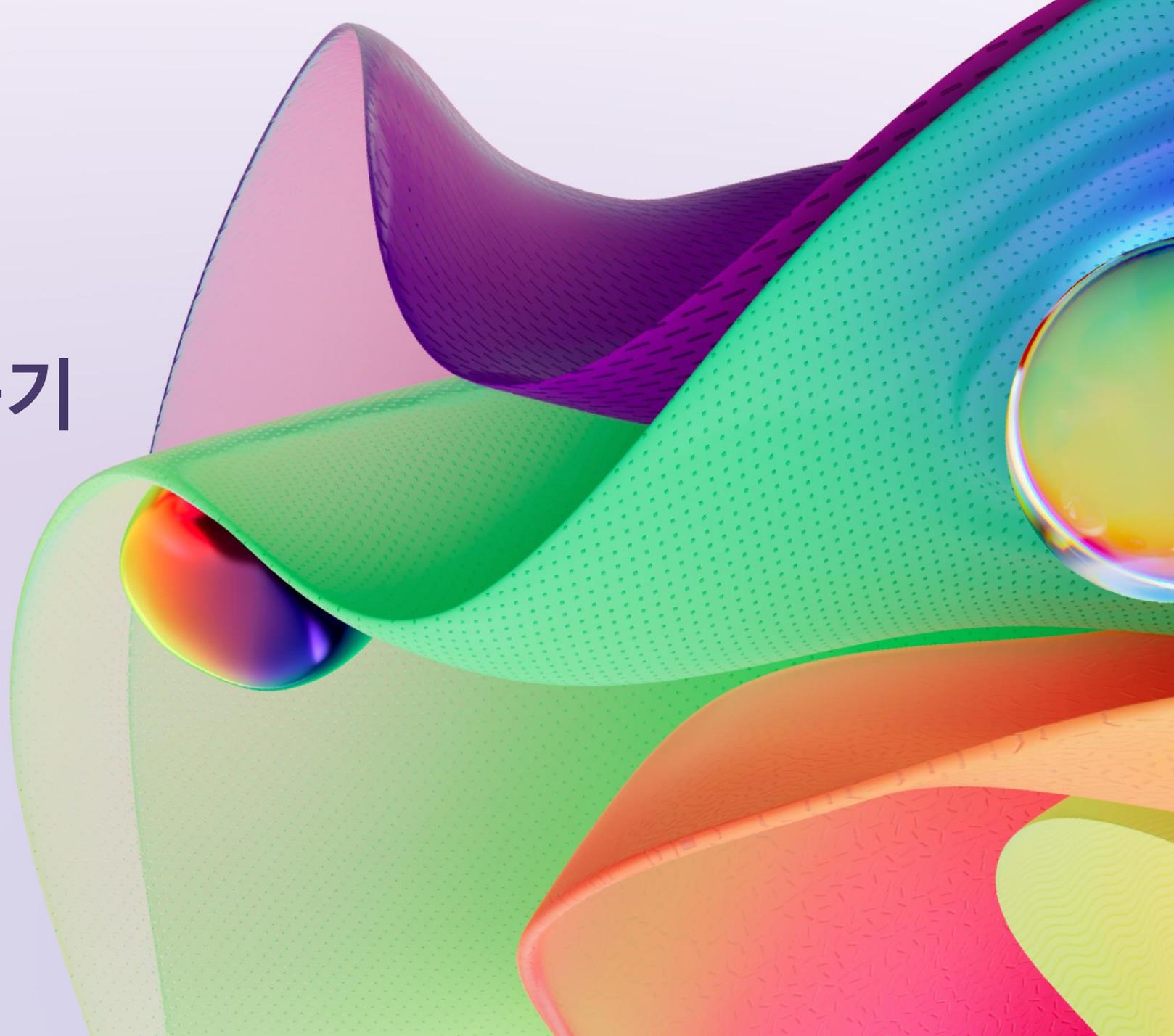
In partnership with  **NVIDIA.**





Workshop: 실전 RAG 정복하기

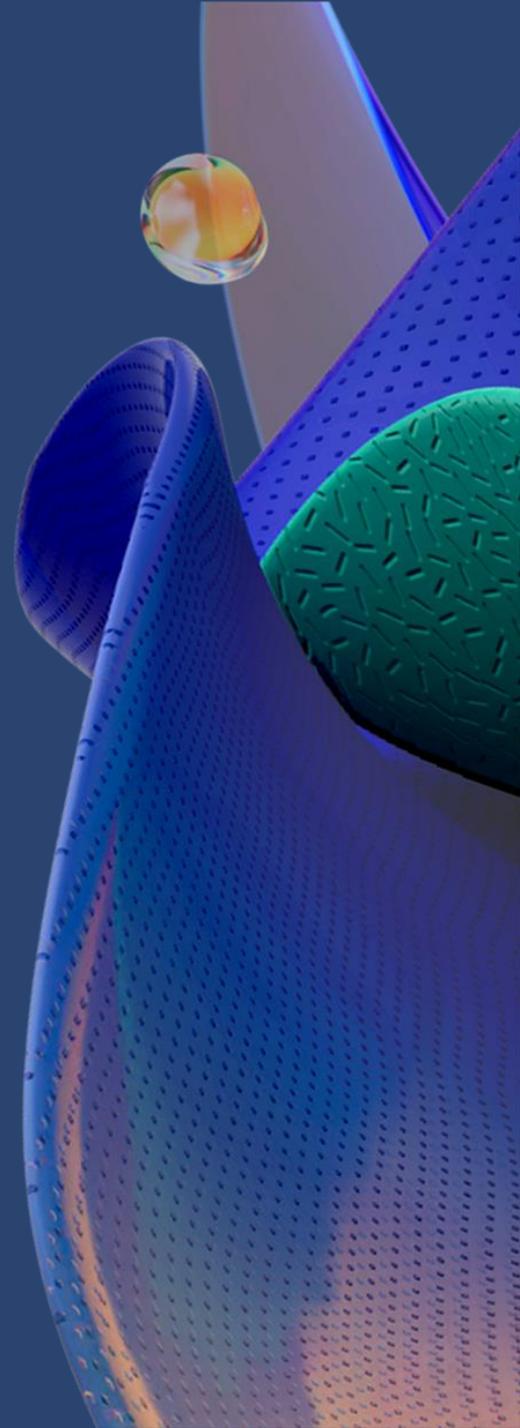
김현수
기술 전략 매니저, DN 팀



어젠다

- 워크샵 소개 (실전 RAG 정복하기)
- Retrieval Augmented Generation 소개
- Retrieval (검색)
- Augmented (증강)
- Generation (생성)
- 참고 자료

워크샵 소개



워크샵 소개 - Retrieval Augmented Generation (RAG)

(내가 가지고 있는 데이터 기반)
4월과 8월의 차이를 표로 그려줘.



사용자

User 3 minutes ago

4월과 8월의 차이를 위치별 계절, 시작 요일, 마지막 요일, 유래로 구분하여 테이블로 표현해줘.

Chatbot 2 minutes ago

April	August	
위치 별 계 절	북반구에서는 봄, 남반구에서는 가을	북반구에서는 여름, 남반구에서는 겨울
시작 요일	일년 중 1월과 윤년의 경우에는 일요일, 그 외의 경우는 수요일	일년 중 1월을 제외하고는 화요일
마지 막 요 일	일년 중 12월과 그 외의 경우는 수요일	수요일
유래	April은 라틴어 "aperire"에서 유래되었으며, 이는 봄에 꽃이 핀다는 의미입니다. 또 다른 이론으로는 그리스 신 여신 아프로디테에서 유래되었을 수도 있습니다.	August는 로마 황제 아우구스투스 카이사르를 의미합니다. Augustus는 8월이 그의 이름을 따서 명명된 것입니다.

Duration: 7.1818s

User 2 minutes ago

베리에는 어떤 종류들이 있어?

Chatbot 2 minutes ago

베리에는 여러 종류가 있습니다. 예를 들면 라즈베리, 딸기, 수박딸기, 리코델리베리, 블루베리 등이 있습니다. 그러나 영어에서 베리라고 불리는 과일 중에는 실제로는 "진정한 베리"가 아닌 것도 있습니다. 이러한 예로는 라즈베리, 딸기, 수박딸기, 블랙베리, 크랜베리, 보이센베리 등이 있습니다. 반대로 베리라고 불리지 않는 과일 중에는 토마토, 바나나, 가지, 구아바, 석류, 고추 등도 진정한 베리입니다. 그 외에도 호박, 오이, 멜론, 오렌지, 레몬과 같은 과일들도 약간 다른 구조를 가지고 있으며 다른 이름으로 불릴 수 있습니다.

Duration: 5.5688s

워크샵 소개 - 실전 RAG 정복하기 워크샵

세상에서 가장
지능적인 서비스를
Microsoft의 **Azure**로
구축할 수 있습니다.

프롬프트 엔지니어링 워크샵 (2 days)

24년 7월부터 1,000명 이상 수강한 **강좌 요약본**



워크샵 소개 - 워크샵 참여 방법

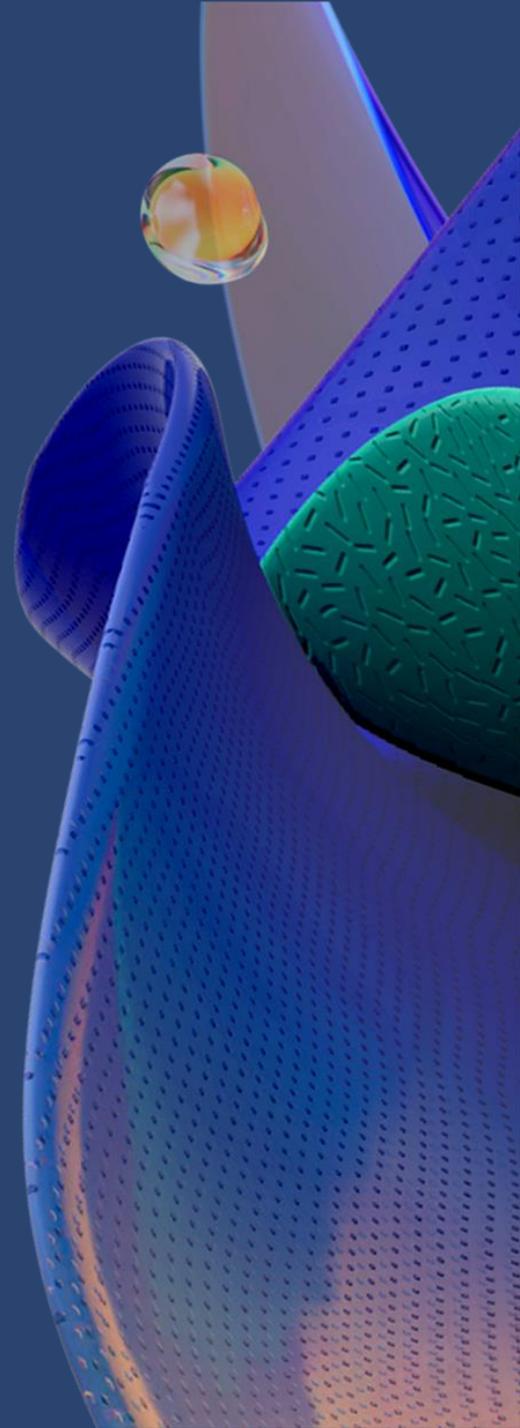
The screenshot shows the GitHub repository page for 'azure-openai-samples-kr'. The repository is public and has 115 stars and 51 forks. It is a fork of 'Azure/azure-openai-samples'. The repository contains various files and folders, including '.devcontainer', 'quick_start', '.env.sample', '.gitignore', 'AUTHORS.md', 'CODE_OF_CONDUCT.md', 'LICENSE', 'README.md', 'SECURITY.md', 'SUPPORT.md', and 'requirements.txt'. The 'README.md' file is selected, showing the title 'GPT 기초, 사용 사례 및 샘플 솔루션 - 한국어 버전'. A callout box with a dark background and white text points to the 'Open in GitHub Codespaces' button, which is highlighted with a red box. The text in the callout box reads: '한 번의 '클릭' 이면, 개발환경 설정 완료'.



워크샵 실습 링크:

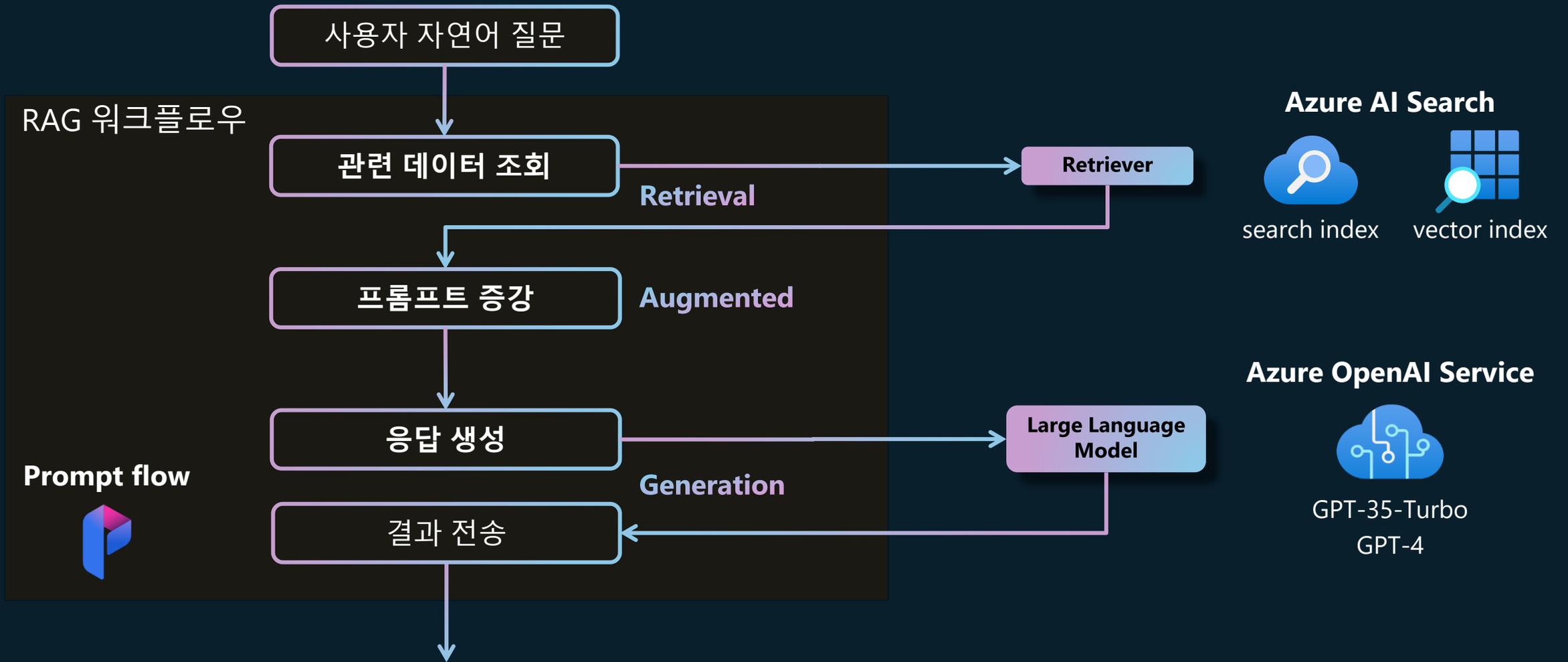
<https://github.com/HyounsooKim/azure-openai-samples-kr>

Retrieval Augmented Generation 소개



RAG – Retrieval Augmented Generation 란?

질문에 대한 답변을 생성할 때, 관련 정보들을 검색하여 그 결과를 생성



RAG 사례 - 고객경험 증대를 위한 고객센터



빠르고 정확하게 고객들의 고민을 해결



1 Intent Classifier: 사전 설정한 Intent 중 가장 관련성 높은 Intent를 결정

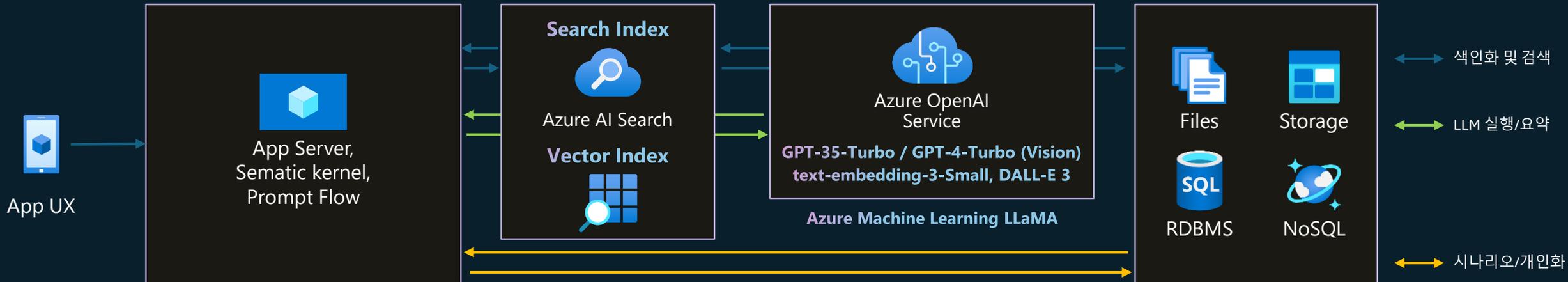
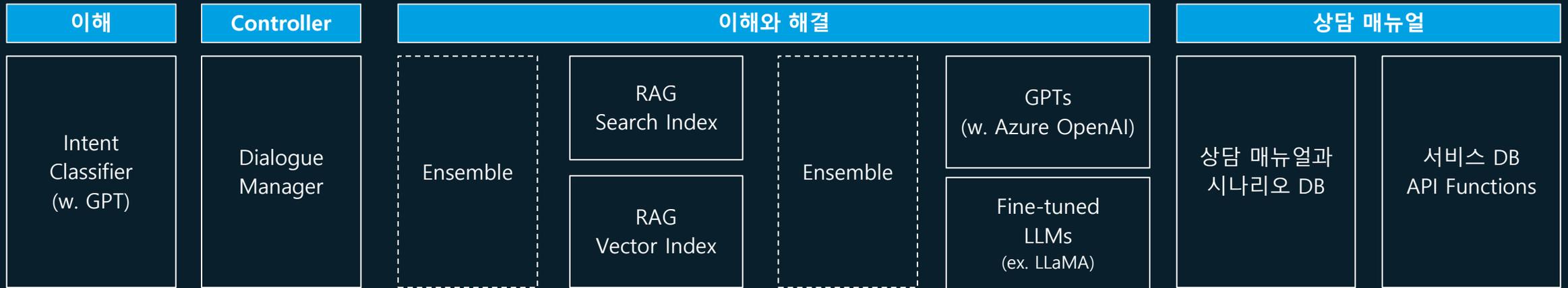
2 Dialogue Manager: 선택한 Intent에 맞게 이후 프로세스를 운영

3 RAG (Retrieval-Augmented Generation)와 Vector Search: 사전 구축 매뉴얼과 시나리오 DB 조회

4 LLMs: Azure OpenAI와 자체 sLLM을 조합하여 사용

RAG – Azure 기반 아키텍처

용도에 맞는 완전 관리형 서비스로 안전하고 빠르게 구축



RAG – Retrieval Augmented Generation 아키텍처

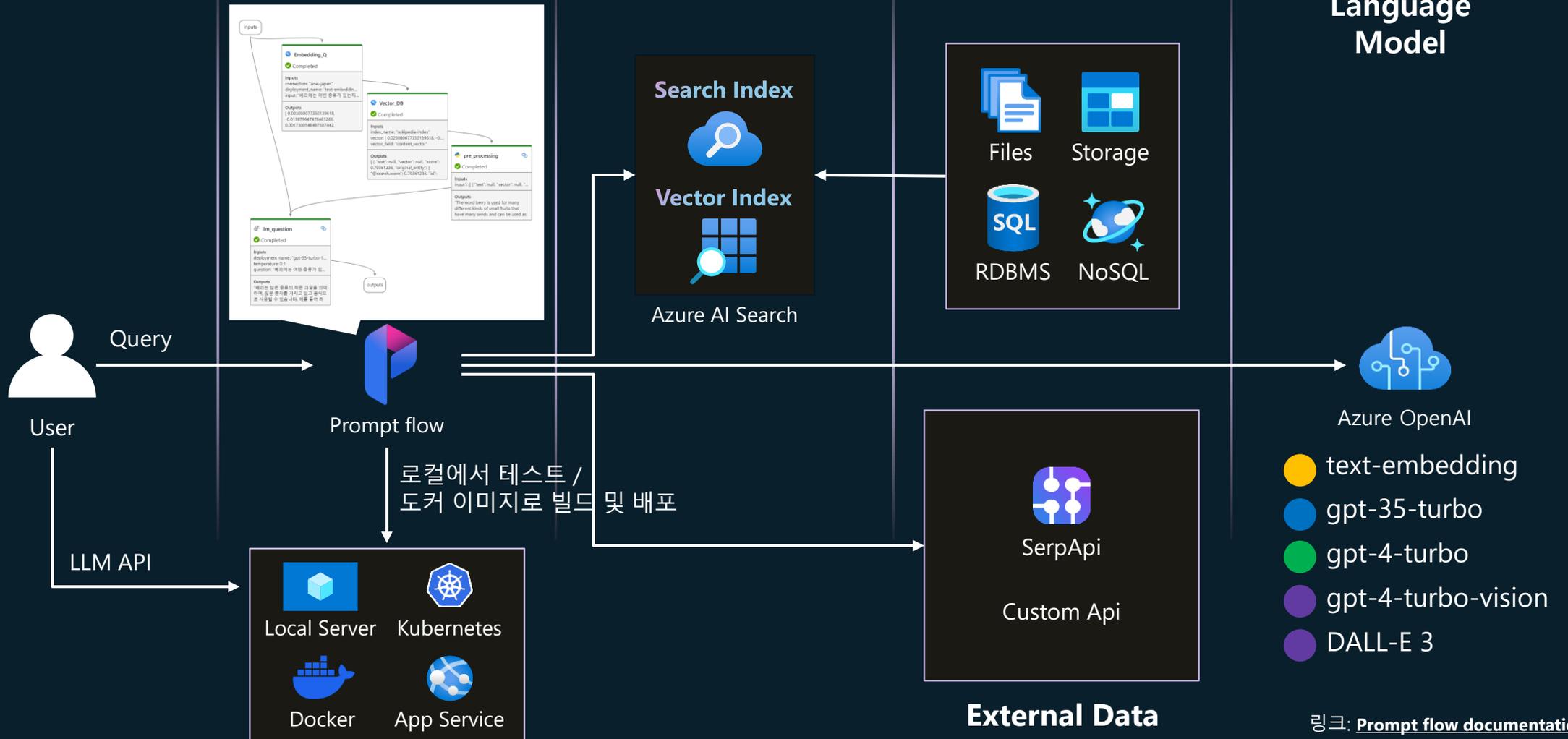
사용자

오케스트레이션

검색

원본 데이터

Large Language Model



RAG – 워크샵을 위한 사전 준비 리소스

Azure Portal과 CLI를 통한 리소스 생성



**Azure 계정
생성**



**구독 ID
생성**

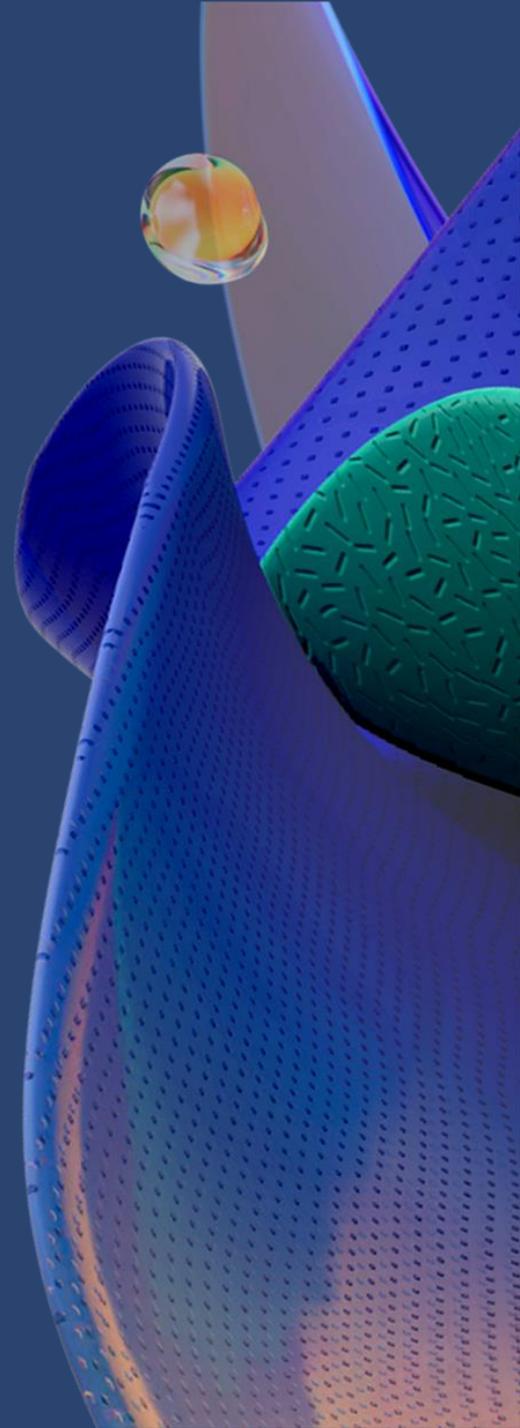


**Azure OpenAI Service
(구독별 액세스 승인 신청)**



Azure AI Search

Retrieval (검색)



Retrieval – 검색을 위한 리소스 사전 확인

사용자

오케스트레이션

검색

원본 데이터

Large Language Model

하이브리드 검색:
키워드 + 벡터

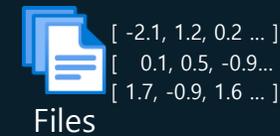
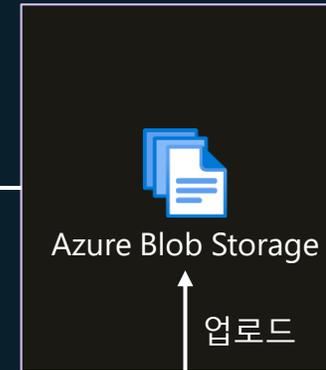
색인화
(인덱싱)

업로드

3

2

1



Azure OpenAI Studio

● text-embedding

● gpt-35-turbo



Azure OpenAI

Retrieval – 1. 문서를 벡터로 변환 (Embedding)

```
wiki_data.csv ×
quick_start > data > wiki_data.csv > data
1 id,url,title,text
2582 93,https://simple.wikipedia.org/wiki/Boil,Boil,"Boil might mean:
2583
2584 Boiling, heating a liquid to the point where it turns into gas
2585 Boil, a type of Staphylococcal
2586
2587 Basic English 850 words"
2588 94,https://simple.wikipedia.o
lower part of a man's face.
2589
2590 The hair that grows on the upp
only below the lower lip and a
a lot of hair and a big beard,
men shave part or all of their
long. Women with hirsutism will
2591
2592 Some animals also have hair like
beard.
2593
2594 Facial hair"
2595 98,https://simple.wikipedia.org/wiki/Black,Black,"In light, black is lack of all
color. In painting, however, the black pigment is the combination of all colors. In
heraldry, black is called ""sable"". It is the opposite of white.
2596
2597 Origin of black
2598 The word ""black"" comes from Old English blac ("black, dark", also, ""ink""), from
Proto-Germanic *blakkaz ("burned"), and from Proto-Indo-European *bhleg-. Black is
the darkest color/tone on a scale.
2599
2600 Black in science
2601 In science, an object that is black absorbs the light that hits it. Because these
objects do not reflect any light, the human eye can't see any color coming from that
object. The brain then sees these objects as black.
2602
2603 A way to create black objects is to mix pigments. A pigment works by reflecting only
```

```
wiki_data_embeddings.csv ×
quick_start > data > wiki_data_embeddings.csv > data
1 id,url,title,text,n_tokens,content_vector
010489335283637047, 0.003925760742276907]"
53 98,https://simple.wikipedia.org/wiki/Black,Black,"In light, black is lack of all color.
In painting, however, the black pigment is the combination of all colors. In heraldry,
black is called ""sable"". It is the opposite of white. Origin of black The word
""black"" comes from Old English blac ("black, dark", also, ""ink""), from
Proto-Germanic *blakkaz ("burned"), and from Proto-Indo-European *bhleg-. Black is
the darkest color/tone on a scale. Black in science In science, an object that is black
absorbs the light that hits it. Because these objects do not reflect any light, the
human eye can't see any color coming from that object. The brain then sees these
objects as black. A way to create black objects is to mix pigments. A pigment works by
reflecting only the color of the pigment. For example, a blue pigment absorbs all
other colors except blue. By mixing pigments in the right quantities, black can be made. In
heraldry, black objects become quickly warm because they absorb much light. Meaning of
black Black is associated with power, elegance, formality, safety, birth, male, and
mystery. Black is a dark color, the darkest color there is. Black, along with gray and
white, is a neutral color. This means that it is not a hot color or a cool color. Black
is a color seen with fear and the unknown (black holes). It can have a bad meaning
(blackbird, black bunny) or a good meaning ('in the black', 'black is beautiful').
Black can stand for strength and power. It can be a formal, elegant, and high-class
color (black tie, black Mercedes, black man). Black clothing is dark in emo and goth
subculture. Related pages List of colors Blackbody radiation Black people Basic English
850 words",376,["0.021400097757577896, -0.007918923161923885, 0.0051758079789578915, -0.
019930345937609673, -0.026506219059228897, -0.009331658482551575, 0.010567010380327702,
-0.022743146866559982, -0.011092827655375004, -0.023908812552690506, 0.
004038650542497635, -0.0003430873330216855, -0.004513786174356937, -0.
006423830054700375, -0.027823926880955696, 0.02345268242061138, 0.044472672045230865, 0.
01766236498951912, 0.0023202444426715374, 0.012473887763917446, -0.017117543146014214,
0.03223318234086037, -0.015394385904073715, 0.011815032921731472, -0.04105169698596001,
0.00752614438533783, 0.02245173044502735, -0.020285112783312798, -0.005369029939174652,
-0.02005704864859581, -0.009198620915412903, 0.015026948414742947, -0.
00987648032605648, -0.009268307127058506, -0.01904342509806156, 0.013861282728612423,
-0.018815360963344574, -0.023503363132476807, -0.01549574825912714, -0.
012290168553590775, -0.0007689273916184902, 0.022046281024813652, -0.0126956170424819,
0.002294903853908181, -0.006094403099268675, 0.006829279009252787, -0.
012315508000523187, -0.0220585412889710, -0.010091026640475008, 0.022388278158211708, 0
```

Azure
OpenAI
Service



text-embedding-ada-002
text-embedding-3-small
text-embedding-3-large

98,https://simple.wikipedia.org/wiki/Black,Black,"In light, black is lack of all color. In painting, however, the black pigment is the combination of all colors. In heraldry, black is called ""sable"". It is the opposite of white.

Origin of black

The word ""black"" comes from Old English blac ("black, dark", also, ""ink""), from Proto-Germanic *blakkaz ("burned"), and from Proto-Indo-European *bhleg-. Black is the darkest color/tone on a scale.

Black in science

In science, an object that is black absorbs the light that hits it. Because these objects do not reflect any light, the human eye can't see any color coming from that object. The brain then sees these objects as black.

A way to create black objects is to mix pigments. A pigment works by reflecting only

850 words",376,["0.021400097757577896, -0.007918923161923885, 0.0051758079789578915, -0.019930345937609673, -0.026506219059228897, -0.009331658482551575, 0.010567010380327702, -0.022743146866559982, -0.011092827655375004, -0.023908812552690506, 0.004038650542497635, -0.0003430873330216855, -0.004513786174356937, -0.006423830054700375, -0.027823926880955696, 0.02345268242061138, 0.044472672045230865, 0.01766236498951912, 0.0023202444426715374, 0.012473887763917446, -0.017117543146014214, 0.03223318234086037, -0.015394385904073715, 0.011815032921731472, -0.04105169698596001, 0.00752614438533783, 0.02245173044502735, -0.020285112783312798, -0.005369029939174652, -0.02005704864859581, -0.009198620915412903, 0.015026948414742947, -0.00987648032605648, -0.009268307127058506, -0.01904342509806156, 0.013861282728612423, -0.018815360963344574, -0.023503363132476807, -0.01549574825912714, -0.012290168553590775, -0.0007689273916184902, 0.022046281024813652, -0.0126956170424819, 0.002294903853908181, -0.006094403099268675, 0.006829279009252787, -0.012315508000523187, -0.0220585412889710, -0.010091026640475008, 0.022388278158211708, 0

Retrieval – 1. 문서를 벡터로 변환 (Embedding)

임베딩 데모 시연



GitHub Codespace
(Repository 에서 즉시
통합 개발 환경 제공)



Azure OpenAI Service
text-embedding-ada-002
(텍스트를 벡터로 변환)

Retrieval – 1. 문서를 벡터로 변환 (Embedding)

임베딩 데모 시연

Retrieval – 2. 문서를 벡터로 변환하여 업로드 (Upload)

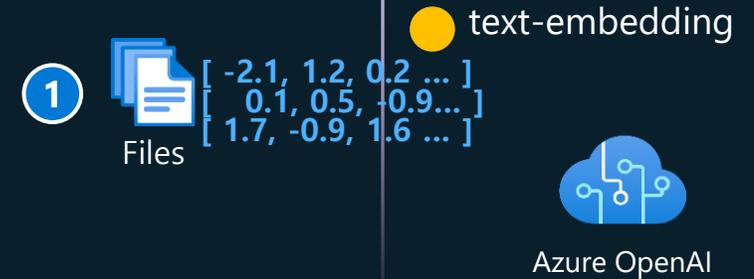
사용자

오케스트레이션

검색

원본 데이터

Large
Language
Model



Retrieval – 2. 문서를 벡터로 변환하여 업로드 (Upload)

검색용 CSV 파일 업로드

Retrieval – 2. 문서를 벡터로 변환하여 업로드 (Upload)

벡터 데이터가 추가된 CSV 문서를 Azure Storage Account의 컨테이너에 업로드

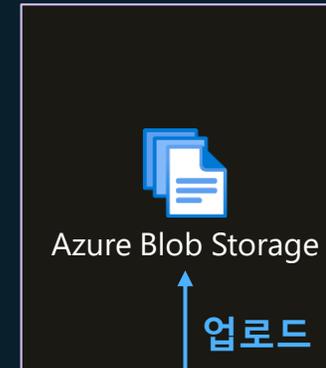
사용자

오케스트레이션

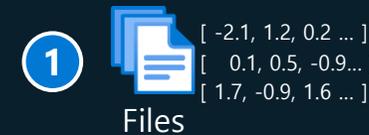
검색

원본 데이터

LLM



업로드



● text-embedding

● gpt-35-turbo



Azure OpenAI

Retrieval – 3. 검색을 위한 색인화 (Indexing)

색인화 과정 데모 시연



Azure Storage accounts
(벡터 CSV 파일 업로드)



Azure AI Search

Retrieval – 3. 검색을 위한 색인화 (Indexing)

색인화 과정 데모 시연

Retrieval – 3. 검색을 위한 색인화 (Indexing)

사용자

오케스트레이션

검색

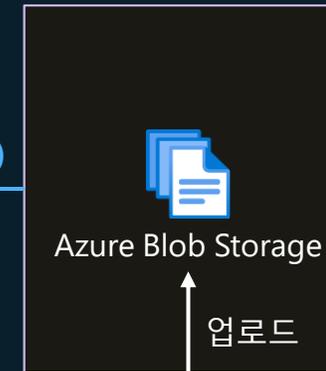
원본 데이터

Large Language Model

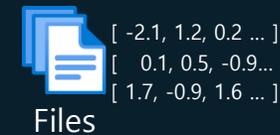


색인화 (인덱싱)

2



1



● text-embedding

● gpt-35-turbo



Azure OpenAI

Retrieval – 4. 테스트 (Azure OpenAI On Your Data)

벡터 DB로 연결 후 테스트



**Azure OpenAI Service
On Your Data**

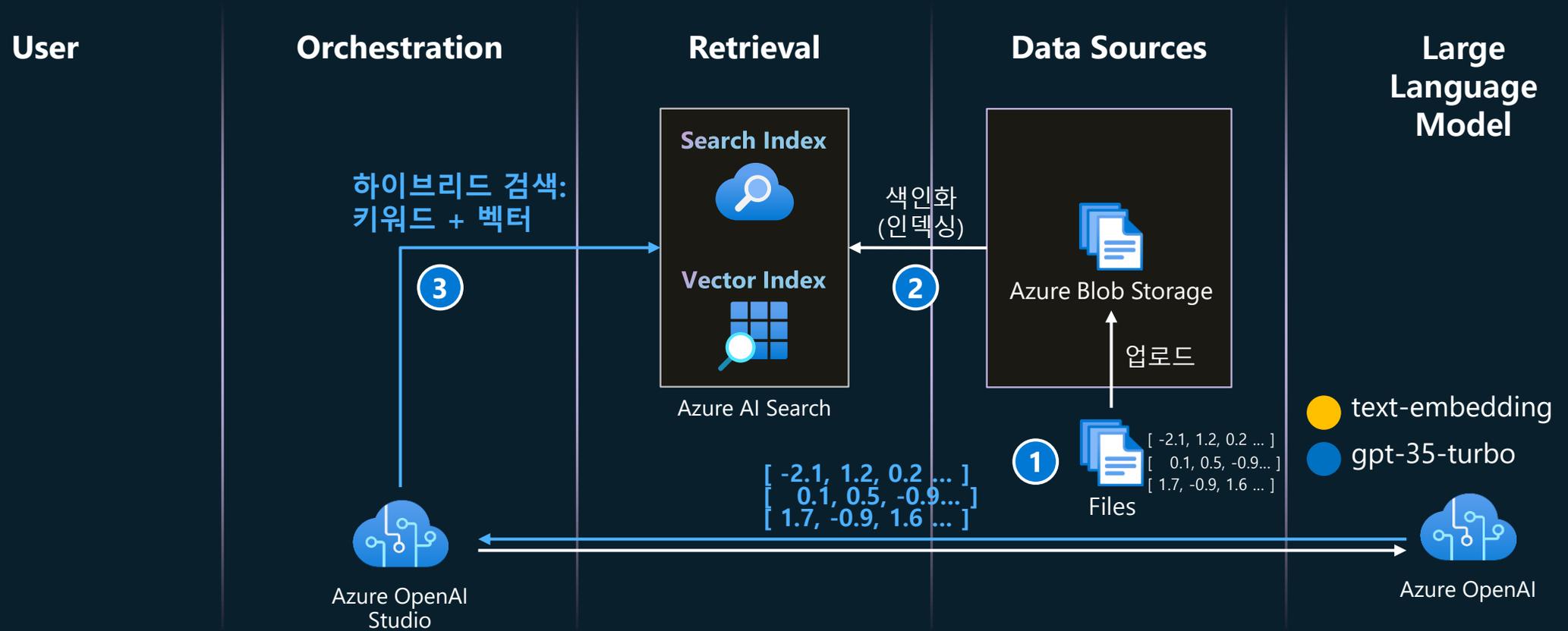


Azure AI Search

Retrieval – 4. 테스트 (Azure OpenAI On Your Data)

벡터 DB로 연결 후 테스트

Retrieval – 4. 테스트 (Azure OpenAI On Your Data)



Retrieval – 4. 테스트 (Azure OpenAI On Your Data)

채팅 플레이그라운드

배포 대상

설치 가져오기 | 설정 내보내기 | 패널 표시

설정

확인 데이터 추가

자신의 데이터 원본에 대한 인사이트를 얻으세요. 귀하의 데이터는 Azure 구독에 안전하게 저장됩니다. 데이터를 보호하는 방법에 대해 자세히 알아보세요.

데이터 원본: 리소스 검색:
Azure AI Search genaithon-2024
인덱스:
wikipedia-index

고급 설정

데이터 콘텐츠에 대한 응답 제한

엄격성(1~5)

3

검색된 문서(3~20개)

5

데이터 원본 제거

채팅 재생 | 채팅 지우기 | 플레이그라운드 설정 | 코드 보기 | JSON 표시

4월과 8월의 차이를 특징별로 요약해서 표로 그려줘.

4월과 8월의 차이점을 특징별로 요약하여 표로 정리하겠습니다.

특징	4월 (1)	8월 (2)
순서	1년 중 네 번째 달	1년 중 여덟 번째 달
일수	30일	31일
시작하는 요일	7월과 같고, 윤년에는 1월과 같음	윤년에는 2월과 같고, 보통년에는 다른 어떤 달과도 같지 않음
끝나는 요일	매년 12월과 같음	매년 11월과 같음
계절	북반구에서는 봄, 남반구에서는 가을	북반구에서는 여름, 남반구에서는 겨울

축제 및 기념일 - 4월 1일: 만우절

- 부활절(3월 22일부터 4월 25일 사이의 일요일)
- 4월 22일: 지구의 날 | 8월 1일: 스위스 국경일
- 8월 15일: 성모 승천일(서방 기독교)
- 8월 15일: 인도 독립기념일
- 8월 17일: 인도네시아 독립기념일
- 8월 31일: 말레이시아 독립기념일 | 역사적 사건 | 4월 1일: 영국 왕립 공군 창설
- 4월 4일: 마틴 루터 킹 주니어 암살
- 4월 12일: 유리 가가린의 우주 비행 | 8월 1일: 스위스 전통적 장립일
- 8월 6일: 히로시마에 원자폭탄 투하
- 8월 15일: 일본의 항복 발표 | 꽃 | 스위트피와 데이지 | 글라디올러스 | 보석 | 다이아몬드 | 페리도트 | 별자리 | 양자리(3월 21일 - 4월 19일), 황소자리(4월 20일 - 5월 20일) | 사자자리(7월 23일 - 8월 22일), 처녀자리(8월 23일 - 9월 22일) |

2 참조

1 Citation 1

2 Citation 2

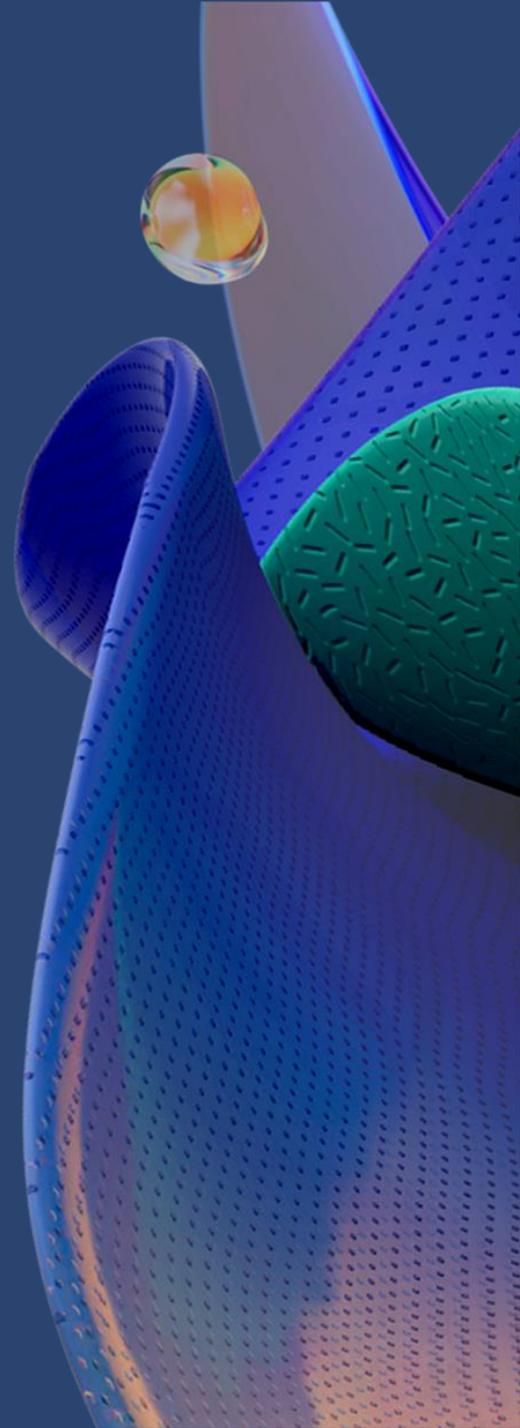
여기에 사용자 쿼리를 입력하세요.(줄 바꿈 입력: Shift + Enter)

인용

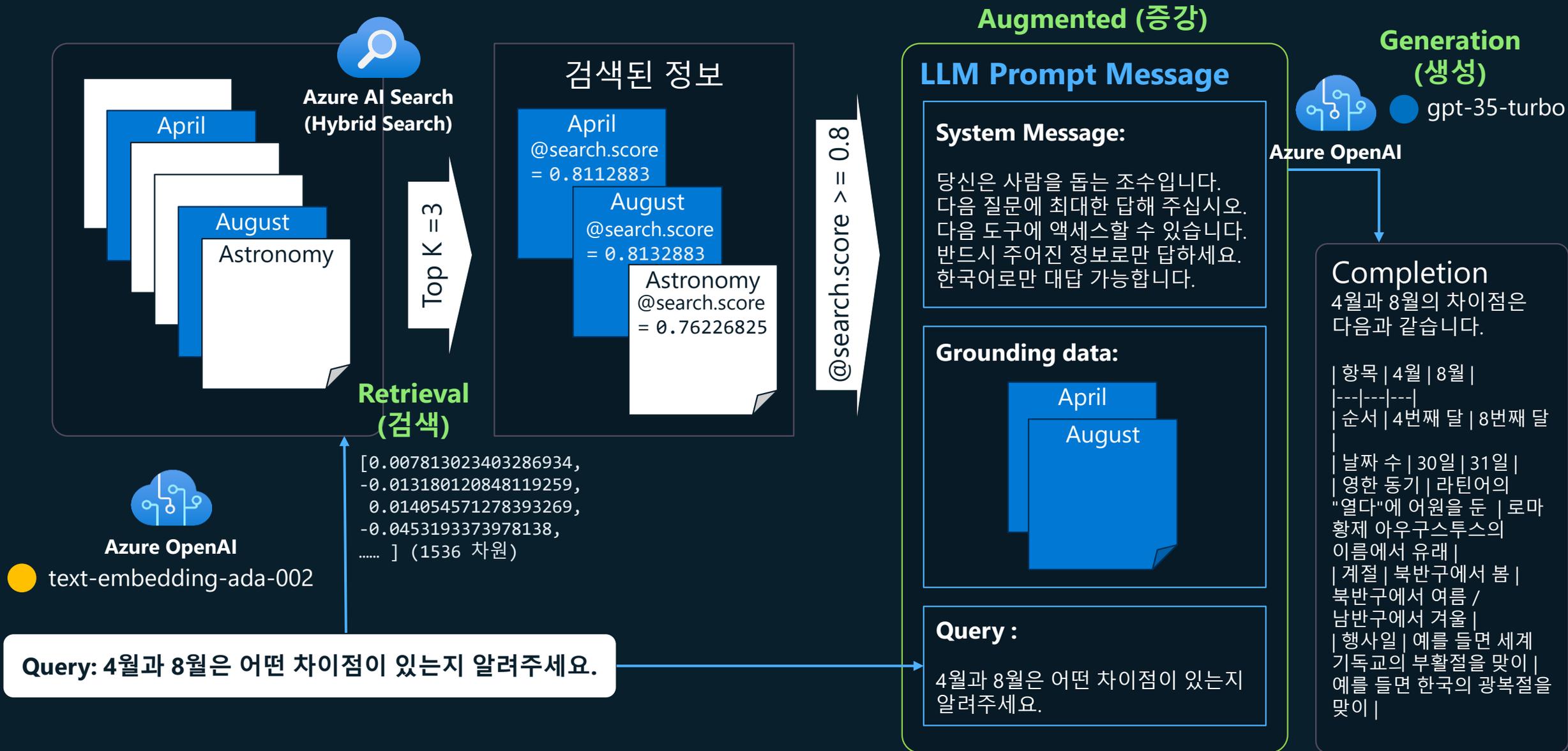
April

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. April in poetry Poets use April to mean the end of winter. For example: April showers bring May flowers. Events in April Fixed Events April 1 - April Fools' Day April 1 - Islamic Republic Day (Iran) April 2 - International Children's Book Day April 2 - Thai Heritage and Conservation Day April 2 - World Autism Awareness Day April 2 - Malvinas Day (Argentina) April 4 - Independence Day (Senegal) April 4 - International Day for Landmine Awareness and Assistance April 4 - Peace Day (Angola) April 5 - End of Tax Year (United Kingdom) April 6 - Tartan Day (Canada and United States) April 6 - Chakri Day (Thailand) April 7 - Day of Maternity and Beauty (Armenia) April 7 - Genocide Memorial Day (Rwanda) April 7 - World Health Day April 7 - Women's Day (Mozambique) April 8 - Buddha's Birthday (Buddhism) April 9 - Martyrs' Day (Tunisia) April 9 - Day of National Unity (Georgia) April 9 - Day of the Finnish language April 12 - Cosmonauts' Day (Russia),

Augmented (증강)



Augmented - 검색 데이터 활용 과정



Augmented – 1. 프롬프트 엔지니어링의 중요성

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

Augmented – 1. 프롬프트 엔지니어링의 중요성

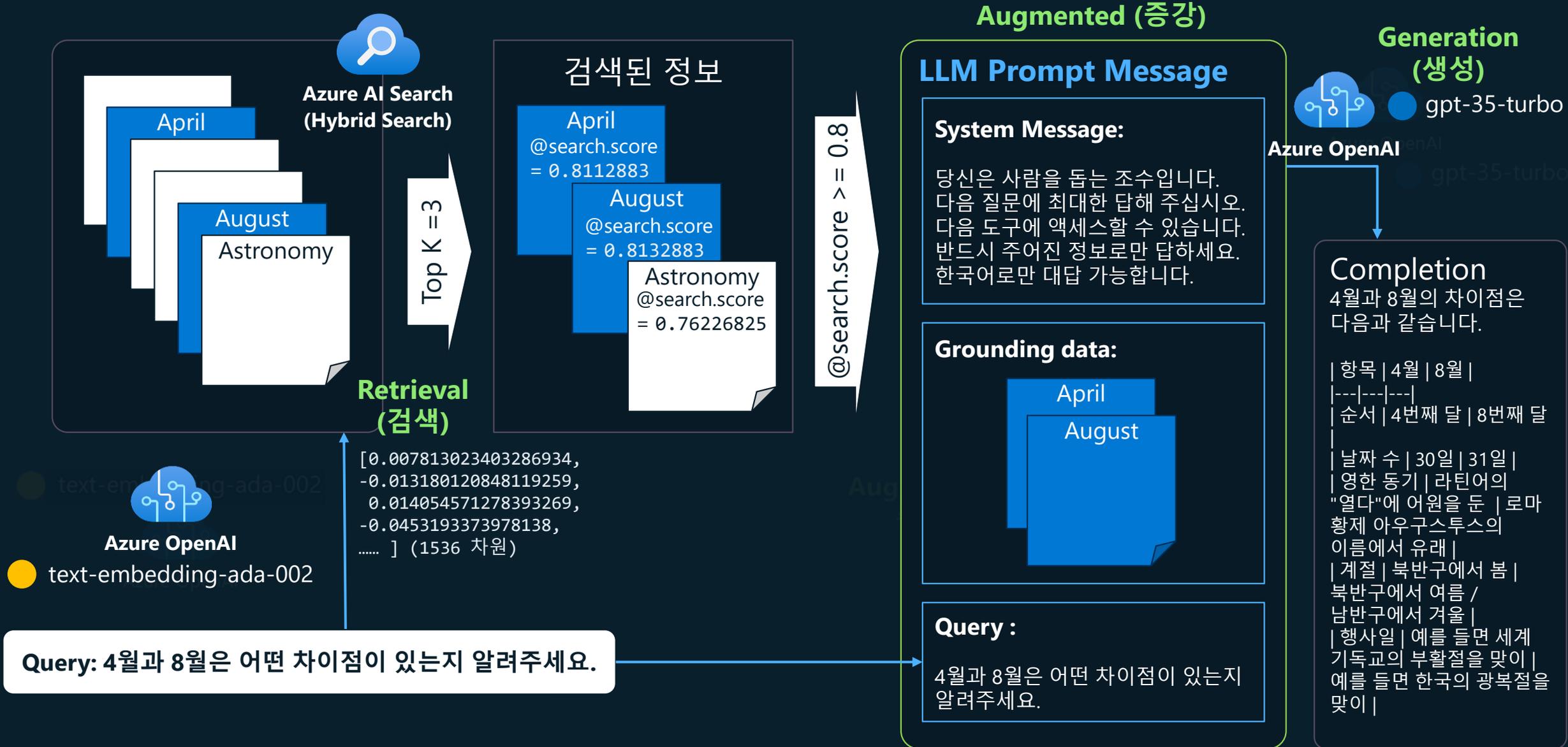
프롬프트 엔지니어링 데모 시연



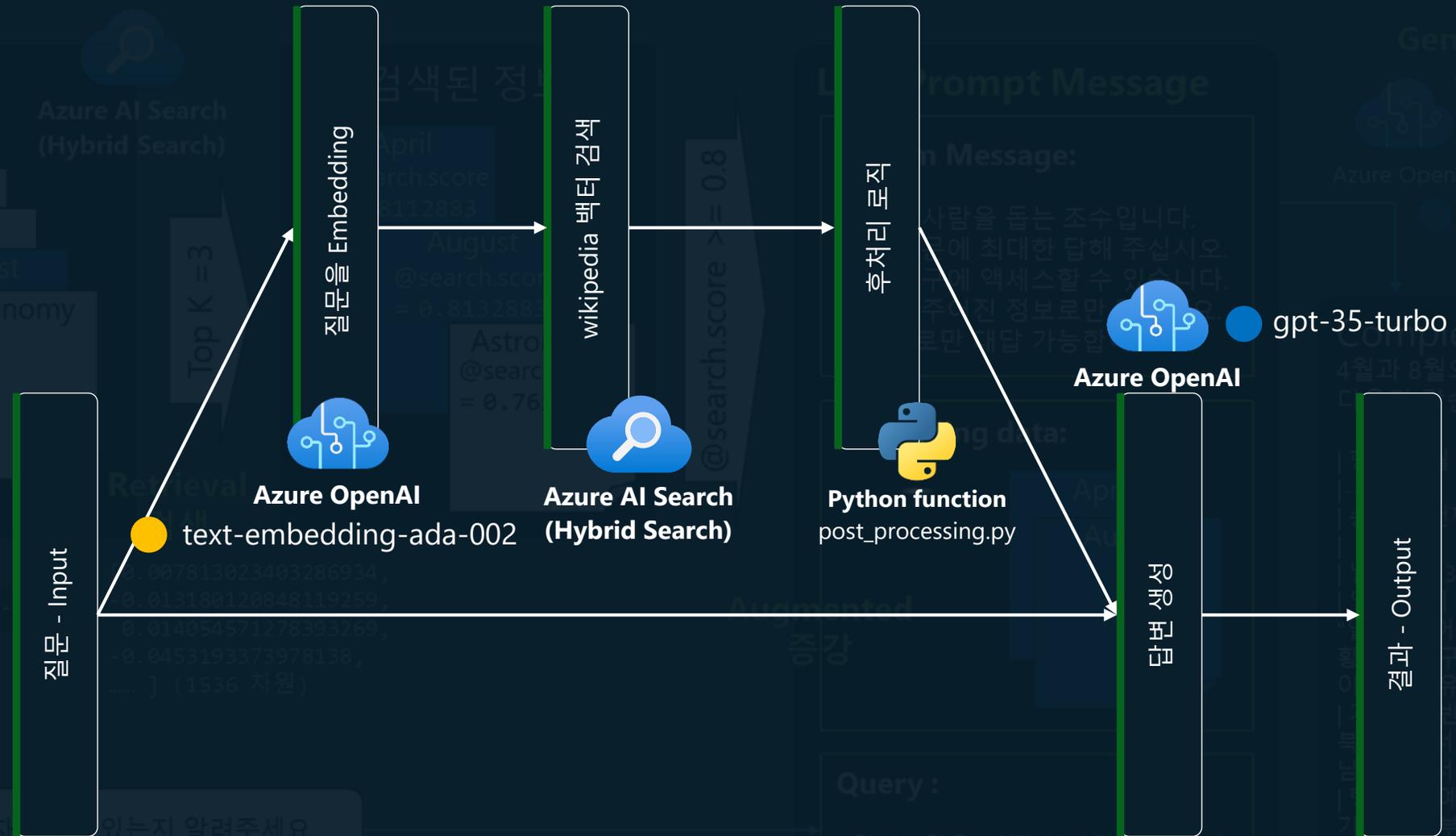
GitHub Codespace

관련 링크: https://github.com/HyounsooKim/azure-openai-samples-kr/blob/main/quick_start/02_OpenAI_getting_started.ipynb

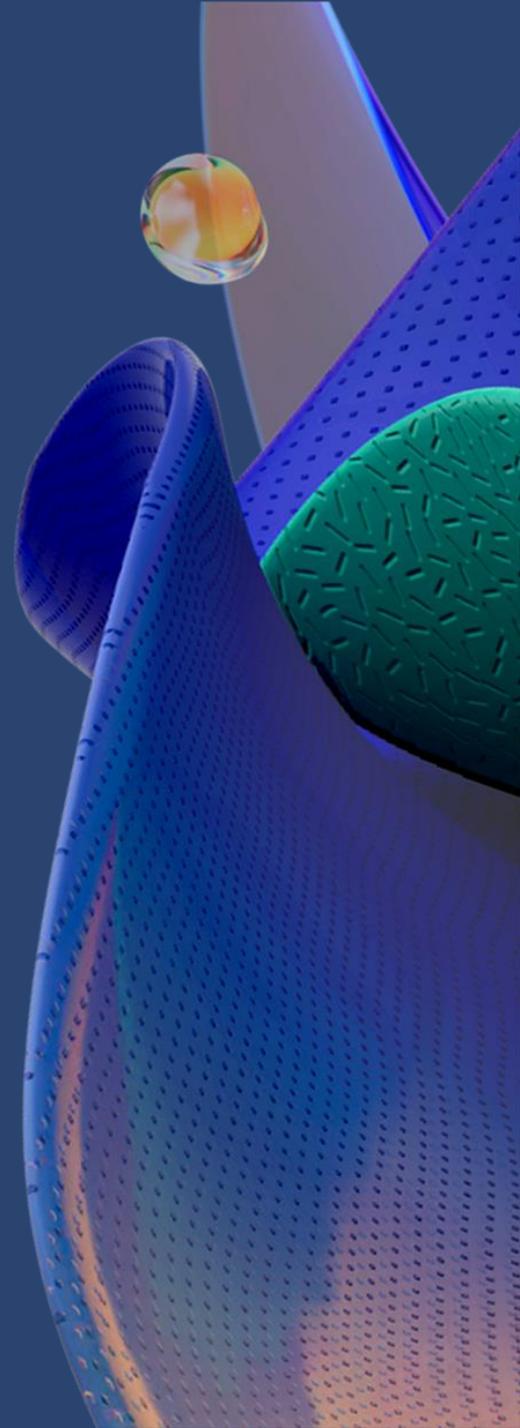
Augmented – 2. 검색 데이터 활용 과정 간략화



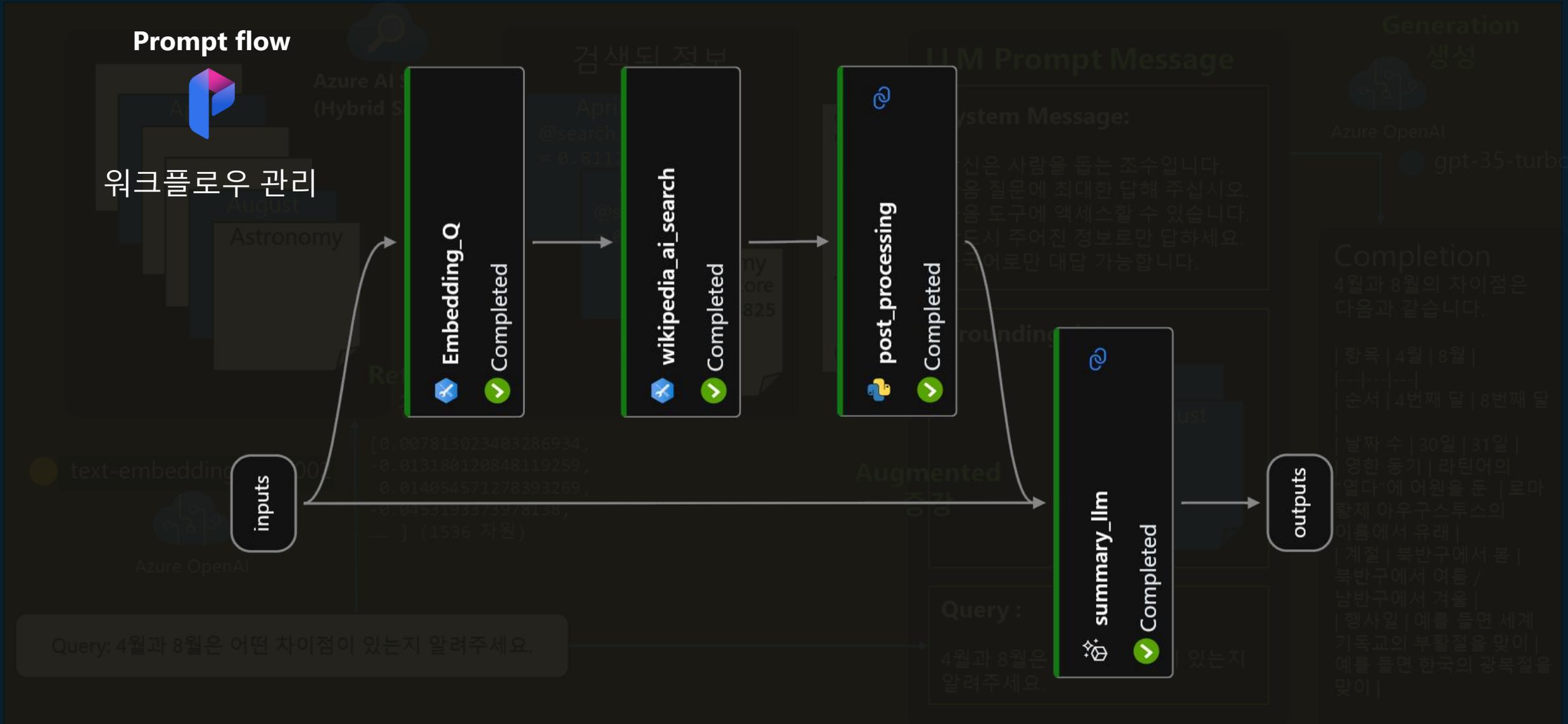
Augmented - 2. 검색 데이터 활용 과정 간략화



Generation (생성)



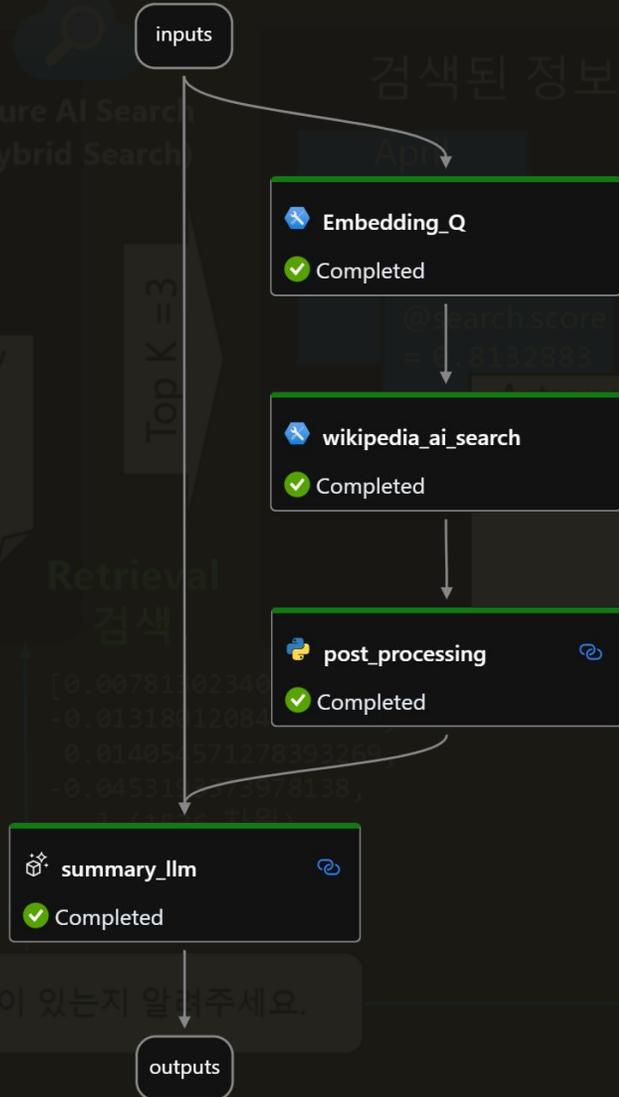
Generation – 1. Prompt Flow 등장배경



Generation – 1. Prompt Flow 등장배경

Prompt flow

워크플로우 관리



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

Generation – 2. 사용자 질의와 동작 구조

사용자

오케스트레이션

검색

원본 데이터

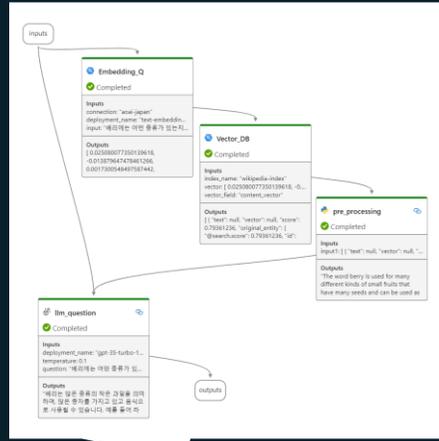
Large Language Model



Query

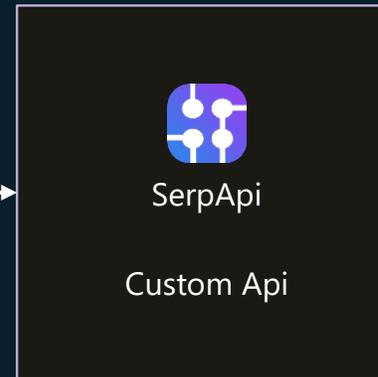
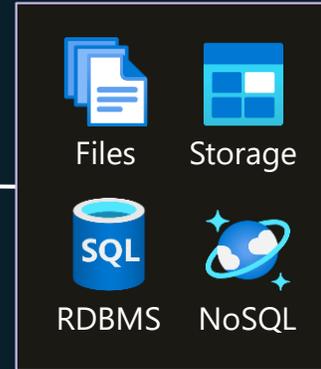
User

LLM API



Prompt flow

로컬에서 테스트 / 도커 이미지로 빌드 및 배포



External Data



Azure OpenAI

- text-embedding
- gpt-35-turbo
- gpt-4-turbo
- gpt-4-turbo-vision
- DALL-E 3

링크: [Prompt flow documentation](#)

Generation – 3. LLMOps를 위한 Prompt flow 예시



(아래 URL을 분류해줘.)

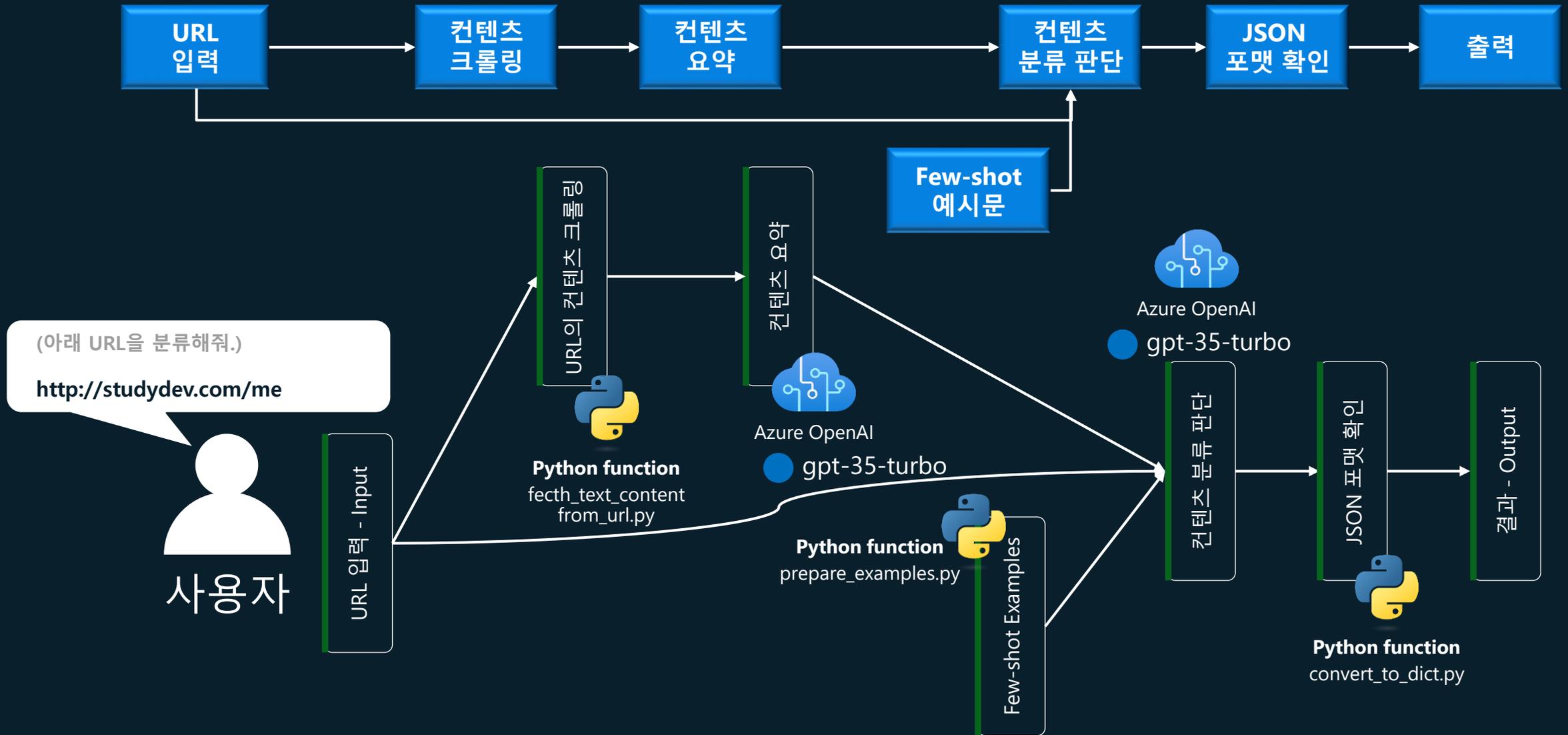
<http://studydev.com/me>



사용자

Few-shot
예시문

Generation – 3. LLMOps를 위한 Prompt flow 예시



Generation – 3. LLMOps를 위한 Prompt flow 예시

Web-Classification 데모



GitHub Codespace
(IDE + Source)



Prompt flow
(Workflow)



Azure OpenAI Service
(LLM/Embedding)

Generation – 3. LLMOps를 위한 Prompt flow 예시

Web-Classification 데모

관련 링크: <https://github.com/HyounsooKim/promptflow-kr/tree/main/examples/flows/standard/web-classification>

Generation – 4. Prompt flow의 API 기능



The screenshot shows a web browser window titled 'Local Server Test App' with the URL `https://fuzzy-meme-wrr5w67x47w3vgvq-8080.app.github.dev`. The page displays a chat interface for 'web-classification(1.0.0)'. The user has entered the URL `http://studydev.com/me`. The chatbot has responded with a JSON object:

```
1 {  
2   "category": "Academic",  
3   "evidence": "Both"  
4 }
```

The duration of the response is 2.5066s. Below this, the user has entered another URL, `http://studydev.com/me`, and the chatbot has responded with a similar JSON object:

```
1 {  
2   "category": "Profile",  
3   "evidence": "Both"  
4 }
```

The duration of this response is 3.4631s. At the bottom of the page, there is an input field labeled 'url *' and a submit button.

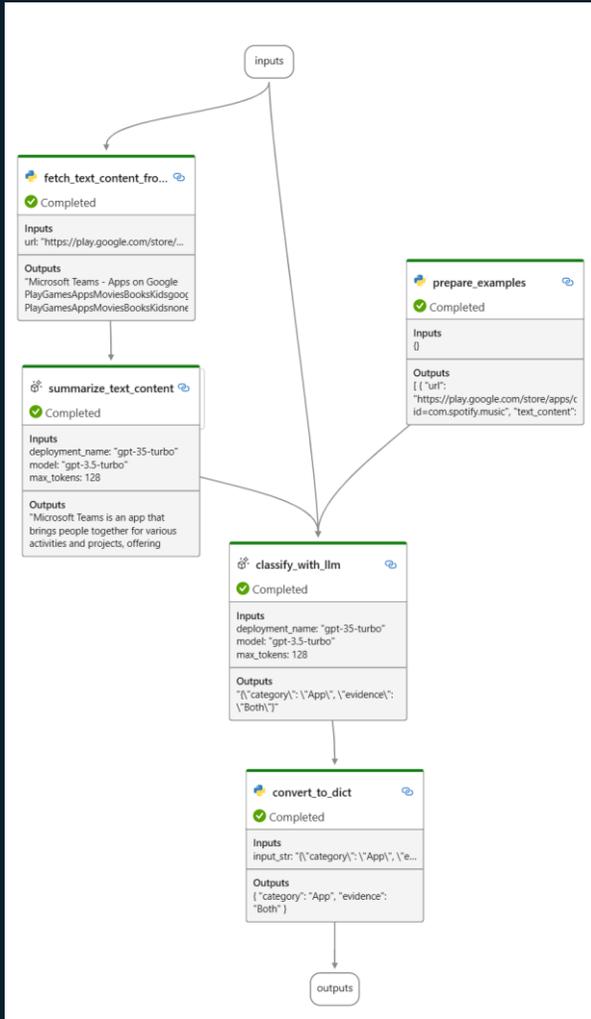
Generation – 4. Prompt flow의 테스트 기능

Streaming Dark Theme

filter condition (UI ONLY! NOT implement yet) Columns Filter

Kind	Name	Input	Output	Start time	End time	Latency	Total tokens	Status
Flow	wikipedia_rag	{ question: "4월과 8월을 비교하고 표를 그려줘" }	{ output: "다음은 4월과 8월을 비교한 표입니다" }	2024. 04. 29. 오전 10:53:10	2024. 04. 29. 오전 10:53:19	⌚ 8.72s	⓪ 7,336	✔
Flow	wikipedia_rag	{ question: "4월과 8월을 비교하고 표를 그려줘" }	null	2024. 04. 29. 오전 10:52:21	2024. 04. 29. 오전 10:52:24	⌚ 2.67s	⓪ 0	✖
Flow	web_classification	{ url: "https://play.google.com/store/apps/de" }	{ category: "App" evidence: "Both" }	2024. 04. 29. 오전 10:47:57	2024. 04. 29. 오전 10:47:59	⌚ 2.63s	⓪ 1,252	✔
Flow	web_classification	{ url: "http://azure.studydev.com/openai/aoai" }	{ category: "PDF" evidence: "Both" }	2024. 04. 29. 오전 10:45:26	2024. 04. 29. 오전 10:46:01	⌚ 35.81s	⓪ 2,195	✔
Flow	web_classification	{ url: "http://www.studydev.com/development/m" }	{ category: "Academic" evidence: "Both" }	2024. 04. 29. 오전 10:43:35	2024. 04. 29. 오전 10:43:37	⌚ 2.90s	⓪ 2,078	✔
Flow	web_classification	{ url: "http://studydev.com/me" }	{ category: "None" evidence: "None" }	2024. 04. 29. 오전 10:40:26	2024. 04. 29. 오전 10:40:31	⌚ 5.39s	⓪ 2,213	✔
Flow	web_classification	{ url: "http://studydev.com/me" }	{ category: "Profile" evidence: "Both" }	2024. 04. 29. 오전 10:40:25	2024. 04. 29. 오전 10:40:29	⌚ 3.41s	⓪ 1,900	✔
Flow	web_classification	{ url: "http://studydev.com/me" }	{ category: "Profile" evidence: "Text content" }	2024. 04. 29. 오전 10:39:05	2024. 04. 29. 오전 10:39:09	⌚ 3.86s	⓪ 1,917	✔

Generation – 4. Prompt flow의 로깅 기능



The screenshot shows a web application interface for 'web_classification'. At the top, it displays 'web_classification' with a green status indicator, a clock icon, '2.63s', and a refresh icon. On the right, it says 'Created on: 2024. 04. 29. 오전 10:47:57' and 'Next trace'. Below this is a sidebar with 'Kind' and 'Flow' buttons. The main area shows a flow log for 'web_classification' with a total time of 2.63s. The log includes the following steps:

- Function: fetch_text_content_from_url (0.24s)
- Function: prepare_examples (1.00ms)
- Function: summarize_text_content (1.66s)
 - LLM: openai_chat (556 tokens, 1.66s)
- Function: classify_with_llm (0.72s)
 - LLM: openai_chat (696 tokens, 0.71s)
- Function: convert_to_dict (0 ms)

On the right side of the interface, there is a chat window titled 'openai_chat' using 'gpt-35-turbo' with 556 tokens and 1.66s. The chat shows a 'User' message and an 'Assistant' response.

User: Text: Microsoft Teams - Apps on Google PlayGamesAppsMoviesBooksKidsgoogle_logo PlayGamesAppsMoviesBooksKidsonesearchhelp_outline Sign in with Googleplay_appsLibrary & devicespaymentPayments & subscriptionsreviewsMy Play activityredeemOffersPlay PasssettingsSettingsPrivacy Policy • Terms of ServiceGamesAppsMoviesBooksKidsMicrosoft TeamsMicrosoft Corporation4.6star7.36M reviews100M+ DownloadsRated for 3+infoInstallShareAdd to wishlistAbout this apparrow_forwardWhether you're connecting with your community for an upcoming activity or working with teammates on a project, Microsoft Teams helps bring people together so that they can get things done. It's the only app that has communities, events, chats, channels, meetings, storage, tasks, and calendars in one place—so you can easily connect and manage access to information. Get your community, family, friends, or work mates together to accomplish tasks, share ideas, and make plans. Join audio and video calls in a secure setting, collaborate in documents, and store files and photos with built-in cloud storage. You can do it all in Microsoft Teams. Easily connect with anyone: • Meet securely with communities, teammates, family, or friends. • Set up a meeting within seconds and invite anyone by sharing a link or calendar invite. • Chat 1-1 or to your entire community, @mention people in chats to get their attention. • Create a dedicated community to discuss specific topics and make plans*. • Work closely and collaborate by keeping conversations organized by specific topics and projects with teams and channels. • Video or audio call anyone directly in Teams or instantly convert a group chat to a call. • Use GIFs, emojis, and message animations to express yourself when words aren't enough. Accomplish plans and projects together: • Send photos and videos in chats to quickly and easily share important moments. • Use cloud storage to access shared documents and files on the go. • Organize shared content in a community — events, photos, lin

Assistant: Microsoft Teams is an app that brings people together for various activities and projects, offering features such as communities, events, chats, channels, meetings, storage, tasks, and calendars in one place. Users can connect with others through audio and video calls, collaborate on documents, and store files and photos with built-in cloud storage. The app allows for secure meetings, 1-1 chats, and group discussions, as well as sharing of photos and videos. It also enables easy access to shared documents and files on the go, making it convenient for organizing and collaborating on plans and projects.

Generation – 4. Prompt flow의 예외 처리 확인 기능

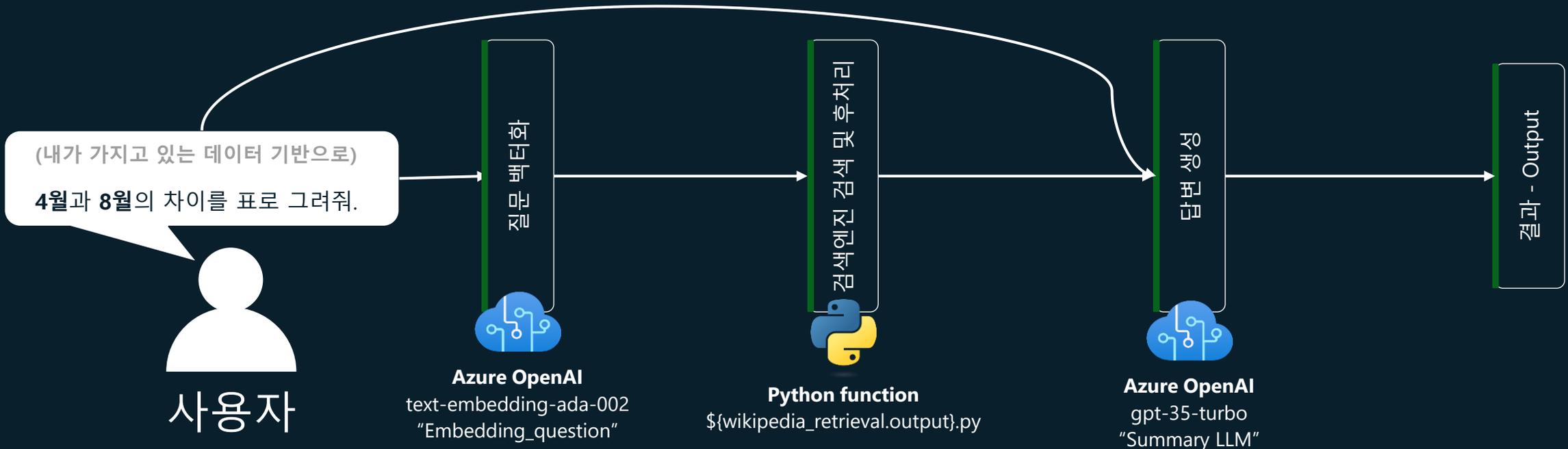
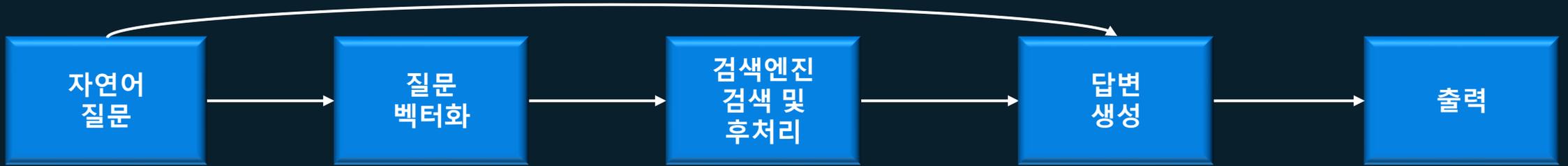
The screenshot displays the Azure AI Studio interface for a workflow named 'wikipedia_rag'. The workflow is shown as a tree structure on the left, with the following components:

- Flow: wikipedia_rag (2.68s)
 - Function: Embedding_Q (2.35s)
 - Embedding: openai_embeddings (2.15s)
 - Function: wikipedia_retrieval (0.15s)

The 'wikipedia_retrieval' function step is expanded on the right, showing an error message. The error details are as follows:

```
{
  name: "exception"
  timestamp: "2024-04-29T01:52:24.154822"
  attributes: {
    exception.type: "TypeError"
    exception.message:
      key must be a string.
    exception.stacktrace:
      Traceback (most recent call last):
        File "/usr/local/lib/python3.9/site-packages/opentelemetry/trace/__init__.py", line 570, in use_span
          yield span
        File "/usr/local/lib/python3.9/site-packages/opentelemetry/sdk/trace/__init__.py", line 1071, in start_as_current_span
          yield span
        File "/usr/local/lib/python3.9/site-packages/opentelemetry/trace/__init__.py", line 437, in start_as_current_span
          yield span
        File "/root/.local/lib/python3.9/site-packages/promptflow/tracing/_trace.py", line 415, in wrapped
          output = func(*args, **kwargs)
        File "/workspaces/promptflow-kr/wikipedia-rag/wikipedia_retrieval.py", line 16, in my_python_tool
          credential = AzureKeyCredential(os.getenv("AZURE_AI_SEARCH_API_KEY"))
        File "/usr/local/lib/python3.9/site-packages/azure/core/credentials.py", line 75, in __init__
          raise TypeError("key must be a string.")
      TypeError: key must be a string.
    exception.escaped: "False"
  }
}
```

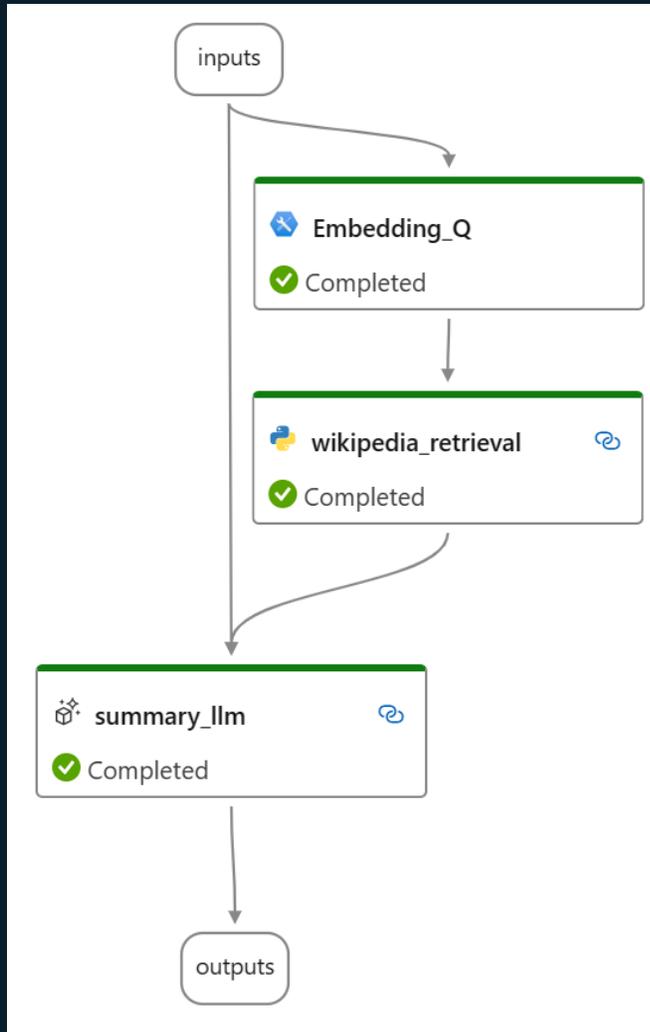
Generation – 5. RAG 기반 Prompt flow 예시



Generation – 5. RAG 기반 Prompt flow 예시

Wikipedia-rag

Generation – 5. RAG 기반 Prompt flow 예시



wikipedia-rag(1.0.0)

User a minute ago

예술의 종류에는 어떤 것들이 있는지 수집한 콘텐츠 기준으로만 답변을 생성해서 알려줘.

Chatbot a minute ago

예술의 종류에는 다양한 것들이 있습니다. 예술은 시각적 미술, 연극, 음악, 춤, 문학, 사진, 조각, 건축 등 다양한 활동을 포괄하는 개념입니다. 이 중에서도 주요 예술의 종류는 다음과 같습니다:

1. 시각미술: 그림, 회화, 조각, 인쇄물 등을 통해 아름다움이나 감정을 표현하는 미술입니다.
2. 연극: 배우들이 몸 사용하며 대사를 말하고 움직임으로 이야기를 전달하는 예술입니다.
3. 음악: 소리를 통해 표현되는 예술로, 악기를 연주하거나 노래를 부르는 등의 형태로 이루어집니다.
4. 춤: 몸을 사용하여 리듬에 맞추어 움직이는 예술로, 다양한 춤의 종류가 있습니다.
5. 문학: 시, 소설, 수필 등의 글을 통해 이야기나 감정을 표현하는 예술입니다.
6. 사진: 사진기를 사용하여 현실을 사진으로 담아내는 예술입니다.
7. 조각: 나무, 돌, 금속 등의 재료를 사용하여 형태를 만들어내는 예술입니다.
8. 건축: 건물을 디자인하고 건설하는 예술로, 인간의 활동 공간을 창조하는 역할을 합니다.

이외에도 예술의 종류는 다양하며, 각각의 분야에서도 다양한 스타일과 장르가 존재합니다. 예술은 창조적인 활동으로써, 사람들의 상상력과 감정을 표현하는 매체로 사용됩니다.

Duration: 14.8381s

User a few seconds ago

4월과 8월의 차이점을 표로 만들어서 비교해줘.

Chatbot a few seconds ago

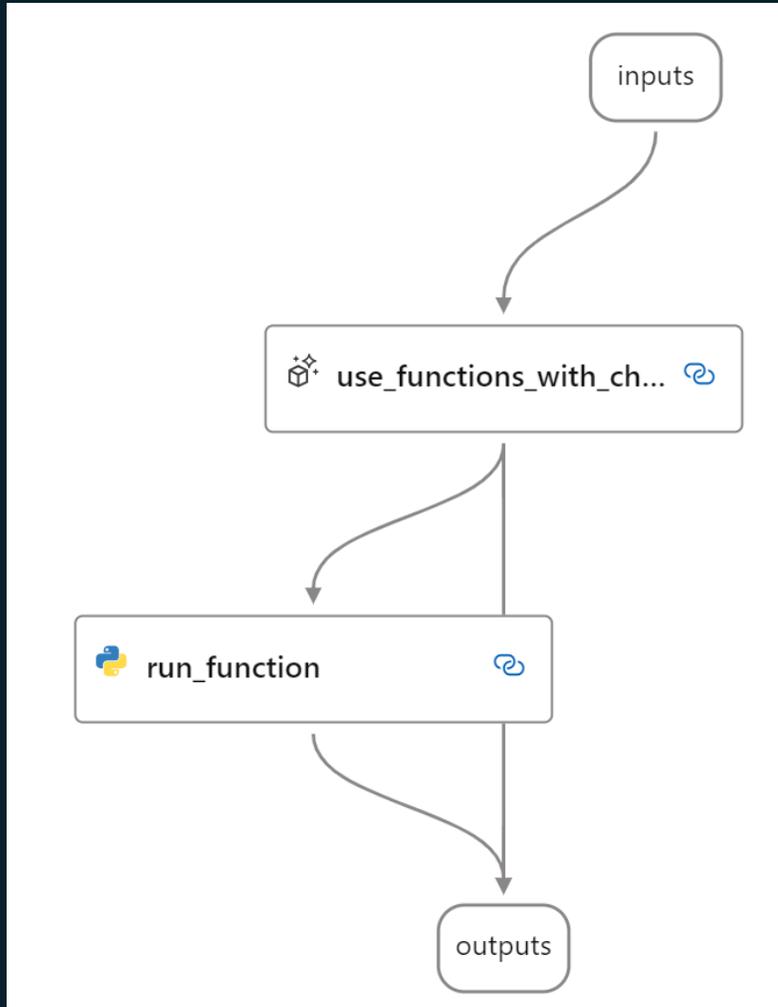
	4월	8월
일수	30일	31일
원래 이름	없음	Sextilis
이름 유래	라틴어에서 "열다"에서 유래 로마 황제 Augustus Caesar에 따름	
시작 요일	모든 해의 7월과 같은 요일	평년의 2월과 같은 요일
끝 날짜 요일	모든 해의 12월과 같은 요일	모든 해의 11월과 같은 요일
이른바 변동되는 행사들	없음	뜨레인 2월의 행사들을 가짐
북반구 계절	봄	여름
남반구 계절	가을	겨울

위 표는 4월과 8월의 주요 차이점을 비교하기 위해 작성되었습니다. 4월은 30일이며, 이름은 특별한 유래가 없습니다. 반면에 8월은 31일로 이루어져 있으며, 로마 황제 Augustus Caesar에 따라 이름이 지어졌습니다. 또한, 4월은 모든 해의 7월과 같은 요일로 시작하지만, 8월은 평년의 2월과 같은 요일로 시작합니다. 끝으로, 4월은 모든 해의 12월과 같은 요일로 끝나지만, 8월은 모든 해의 11월과 같은 요일로 끝납니다. 계절적으로는 4월은 북반구에서 봄에 해당하며, 8월은 여름에 해당합니다. 남반구에서는 4월은 가을에 해당하며, 8월은 겨울에 해당합니다. 특별한 행사나 이벤트는 8월에만 갖고 있으며, 4월에는 그런 행사는 없습니다.

Duration: 14.4042s

question *

Generation – 6. 자연어 기반 자동화 (Function Calling)



The screenshot shows a chat interface with a user asking a question and a chatbot responding with a structured JSON output. The user's question is "한국 마이크로소프트에서 강남역까지 얼마나 걸려?". The chatbot's response is a JSON object containing details about the distance and duration between the Microsoft office in Korea and Gangnam Station.

```
16     "tool_calls": null
17   }
18 }

Duration: 1.5654s

User a few seconds ago

한국 마이크로소프트에서 강남역까지 얼마나 걸려?

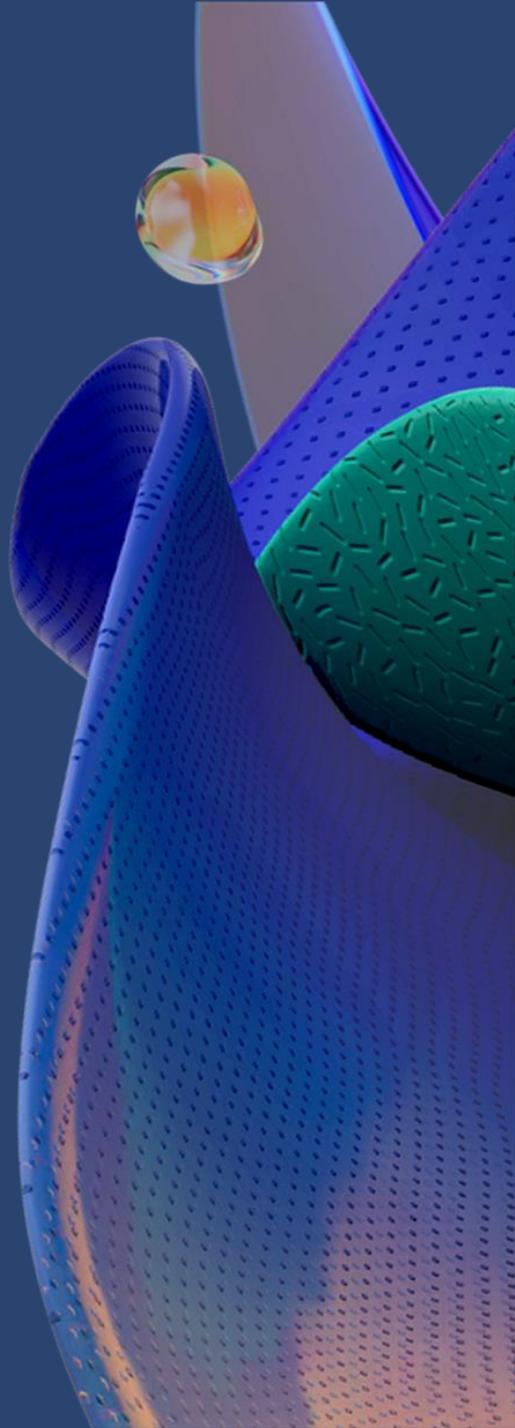
Chatbot a few seconds ago

1 {
2   "answer": {
3     "destination_name": "강남역 2호선",
4     "distance": "10.49km",
5     "duration": "0시간 48분 26초",
6     "origin_name": "한국마이크로소프트",
7     "taxi_fare": 18100,
8     "toll_fare": 0
9   },
10  "llm_output": {
11    "content": null,
12    "function_call": {
13      "arguments": "{\n\"origin\": \"한국 마이크로소프트\", \n\"destination\": \"강남역\"}",
14      "name": "get_directions"
15    },
16    "role": "assistant",
17    "tool_calls": null
18  }
19 }
```

Duration: 2.6976s

question *

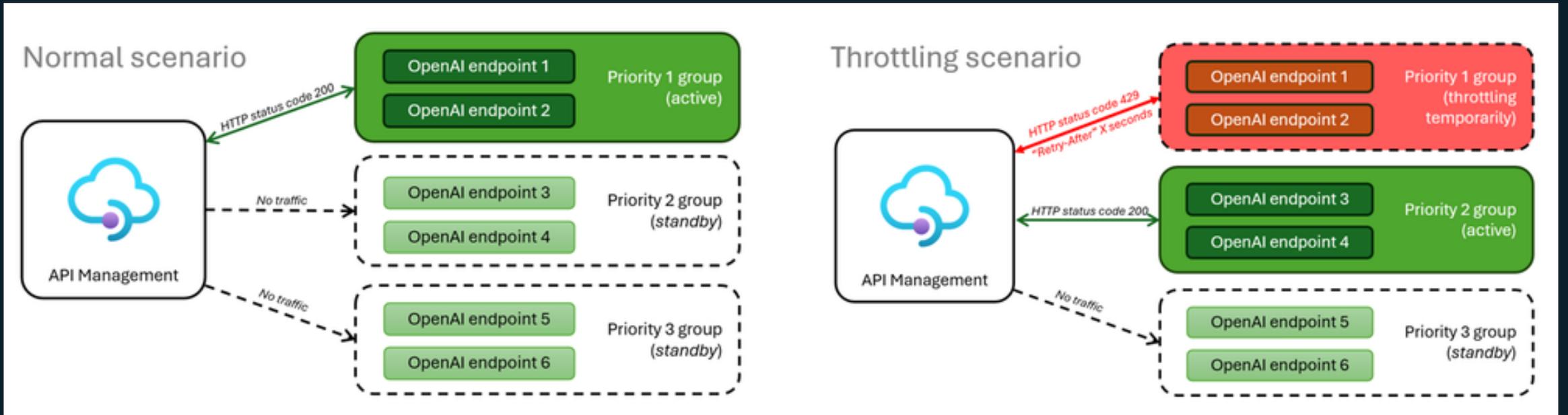
참고 자료



LLMApp에서 고려해야 할 요소 Tokens per minute

LLM은 리전별 모델을 분당 호출할 수 있는 토큰 수가 제약이 있습니다. 따라서 여러 리전에 배포하여 사용하는 것이 TPM을 확보하는 가장 빠른 방법입니다. 다만, 시간대에 따라서 성능이 다를 수 있으므로 테스트가 필요합니다.

- Azure OpenAI는 토큰 제한(TPM)이 발생하면 **429 또는 TooManyRequests HTTP 상태 코드**와 다음 요청을 시도할 때까지 기다려야 하는 시간을 나타내는 Retry-After 응답 헤더를 발행합니다.
- 따라서, 이런 상황에서 **우선순위에 근거하여 다른 Endpoint를 활용하여 API를 호출**하도록 구성합니다. (API management가 이 기능을 제공하며, 일반적인 로드밸런싱 보다 더 스마트하게 구성 가능합니다.)



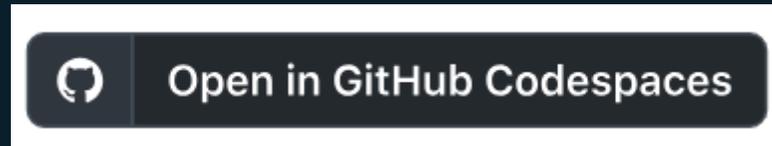
관련 자료

[워크샵 소개 자료](#), [AOAI 소개](#), [프롬프트 엔지니어링](#), [실습 소개](#)

[GtiHub: Prompt Engineering Workshop](#)

[Github: Prompt flow](#)

Call to action!



위 워크샵의 **GitHub Codespace**에서 **Azure**를 활용해 보세요.

Microsoft Learn에서 AI를 통해 기술 시작하기

AI 기술을 구축하고, 커뮤니티와 연결하며, Microsoft 개인
인증 정보를 획득하고, 전문가로부터 배우며, Cloud Skills
Challenge에 참여하세요.

aka.ms/LearnAtAITour

Discord에서 Azure AI 커뮤니티 가입

오늘 열정적인 세션을 진행해 주신
Microsoft 펠로우, Experts 및 MVP 분들과
행사 이후에도 AI에 관해 토론하고 싶으신가요?

Discord의 Azure AI 커뮤니티에서
여러분이 AI에 관해 궁금해하던 질문을
전문가들과 자유롭게 공유할 수 있습니다.

aka.ms/AzureAI/Discord



감사합니다!

