Azure OpenAI

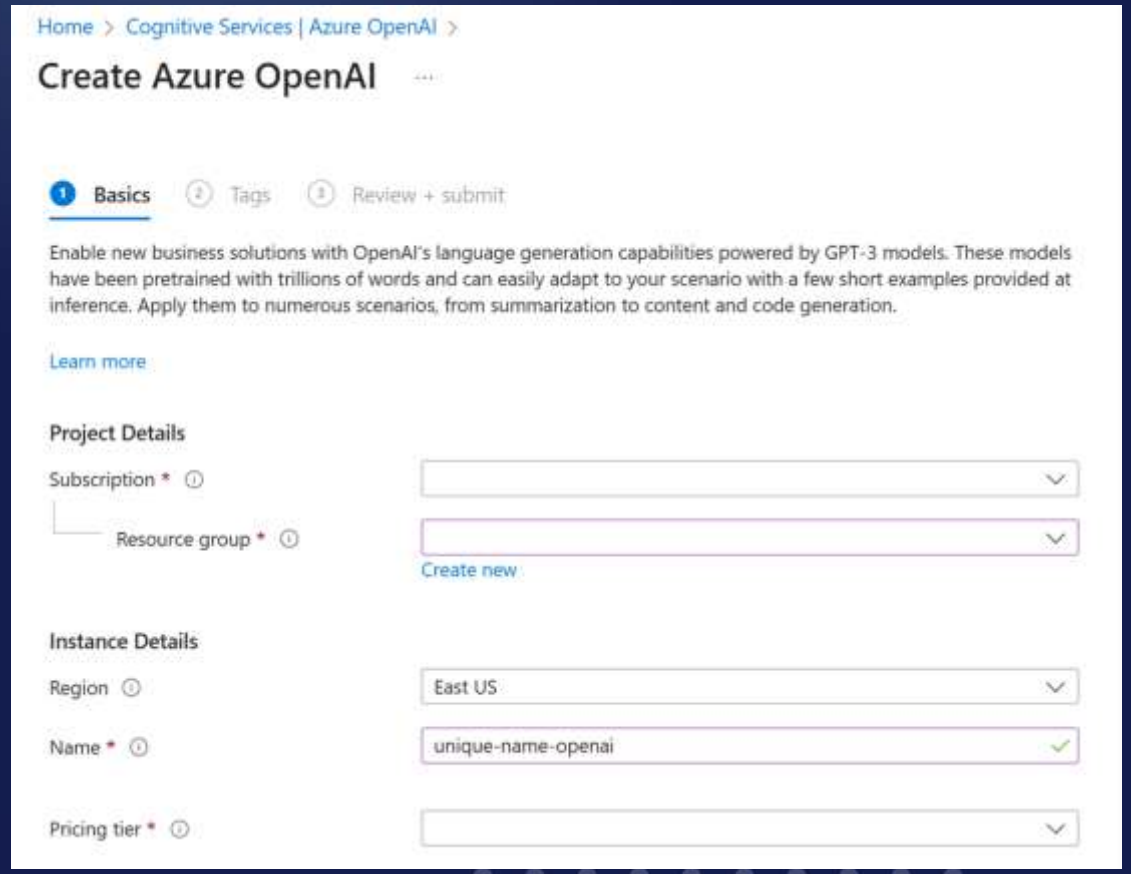
인공지능 기술과 빅테크 전쟁

김영욱
Hello AI
Microsoft AI MVP
Microsoft Regional Director
youngwook@outlook.com

Hello AI

Azure OpenAI Service에 액세스

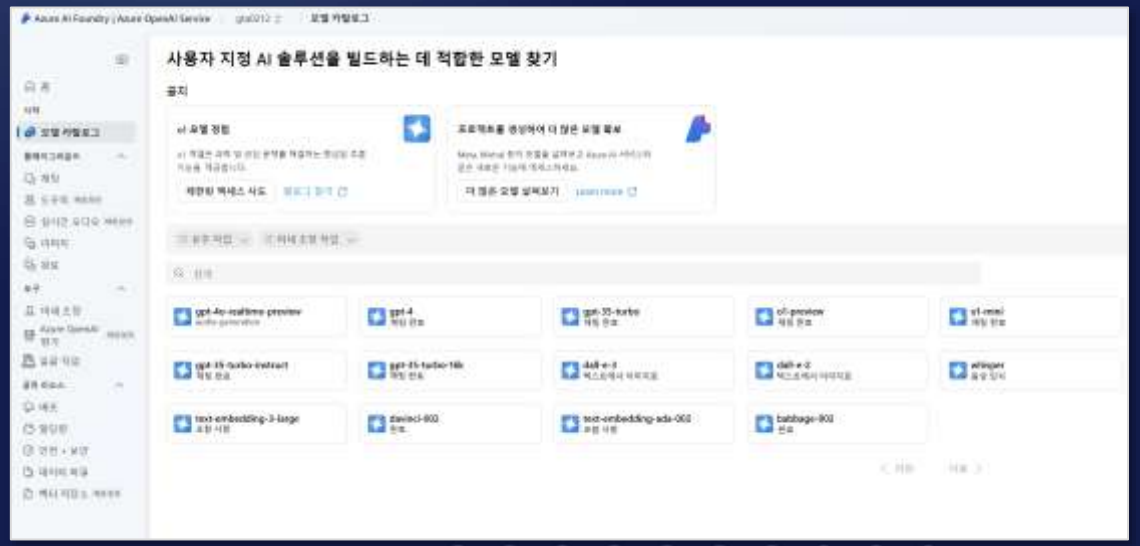
Azure OpenAI Service 리소스를 만들 때 구독 이름, 리소스 그룹 이름, 지역, 고유한 instance 이름을 제공하고 가격 책정 계층을 선택해야 합니다.



지역별 가용성
Azure OpenAI Service는 다양한 유형의 모델에 대한 액세스를 제공합니다. 특정 모델은 일부 지역에서만 사용할 수 있습니다. 지역 가용성에 대한 Azure OpenAI 모델 가용성 가이드를 참조하세요. 지역당 두 개의 Azure OpenAI 리소스를 만들 수 있습니다.

Azure AI Foundry

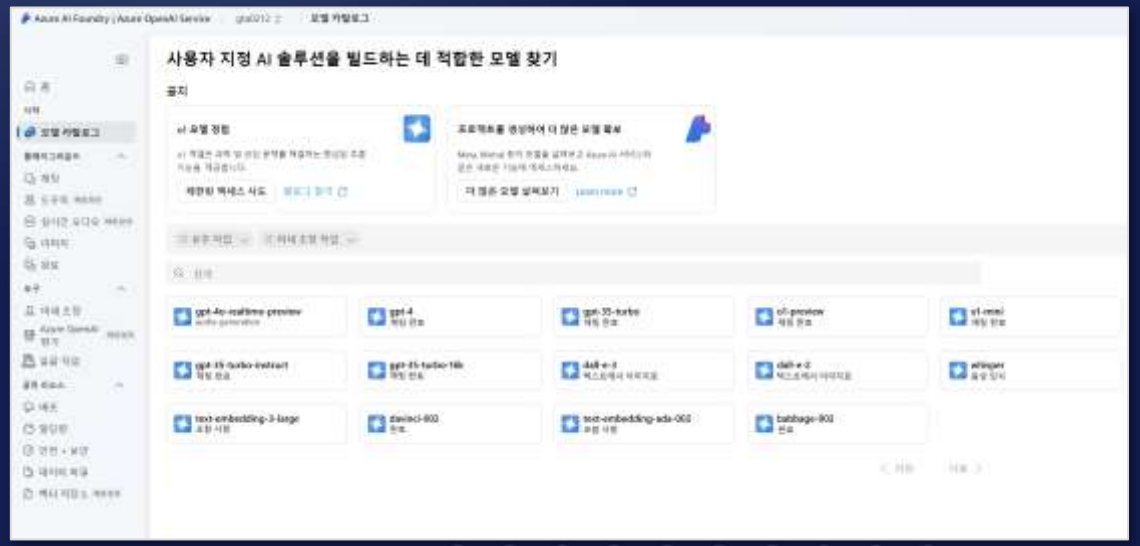
Azure OpenAI Service 리소스를 만들 때 구독 이름, 리소스 그룹 이름, 지역, 고유한 instance 이름을 제공하고 가격 책정 계층을 선택해야 합니다.



- 모델 및 미세 조정 기능의 전체 카탈로그를 사용한 프로 코드 개발
- 클라우드 인프라를 완전히 제어하는 PaaS 서비스
- 프롬프트 및 모델 오케스트레이션
- 성능, 안정성, 확장성, 책임 있는 AI 안전을 테스트하는 평가 엔진
- 사용자 지정 앱 및 서비스에서 사용하기 위해 Azure에 엔드포인트로 배포

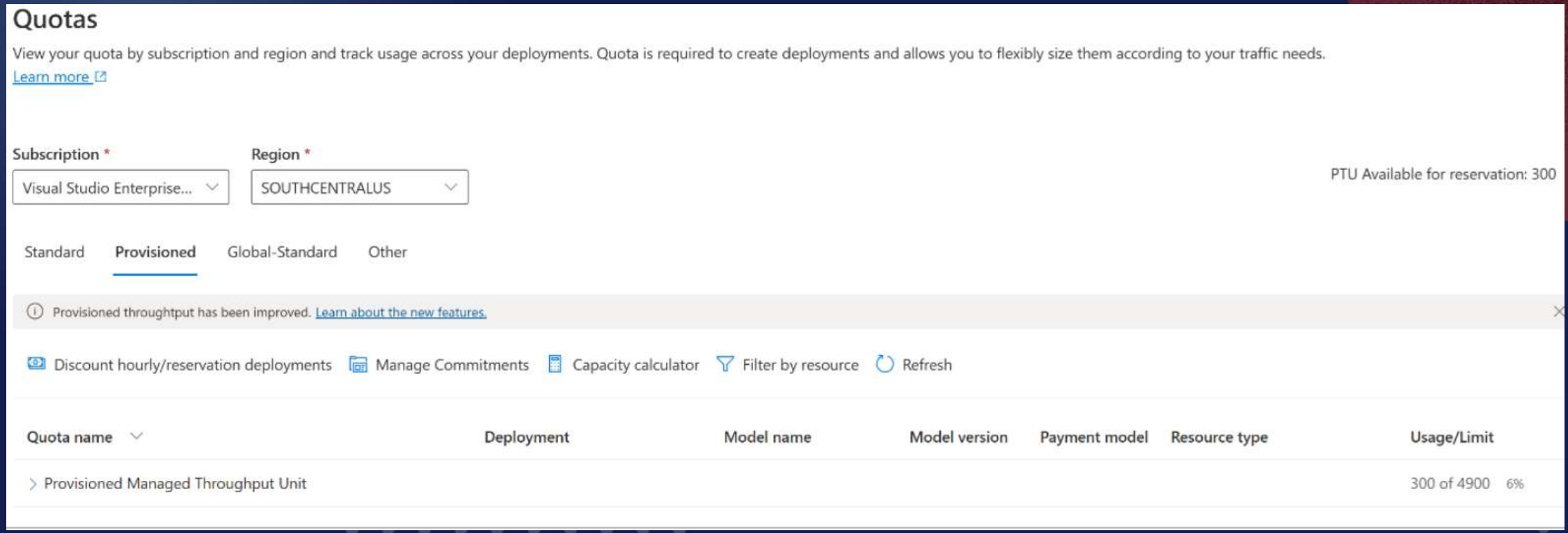
Azure AI Foundry

Azure OpenAI Service 리소스를 만들 때 구독 이름, 리소스 그룹 이름, 지역, 고유한 instance 이름을 제공하고 가격 책정 계층을 선택해야 합니다.



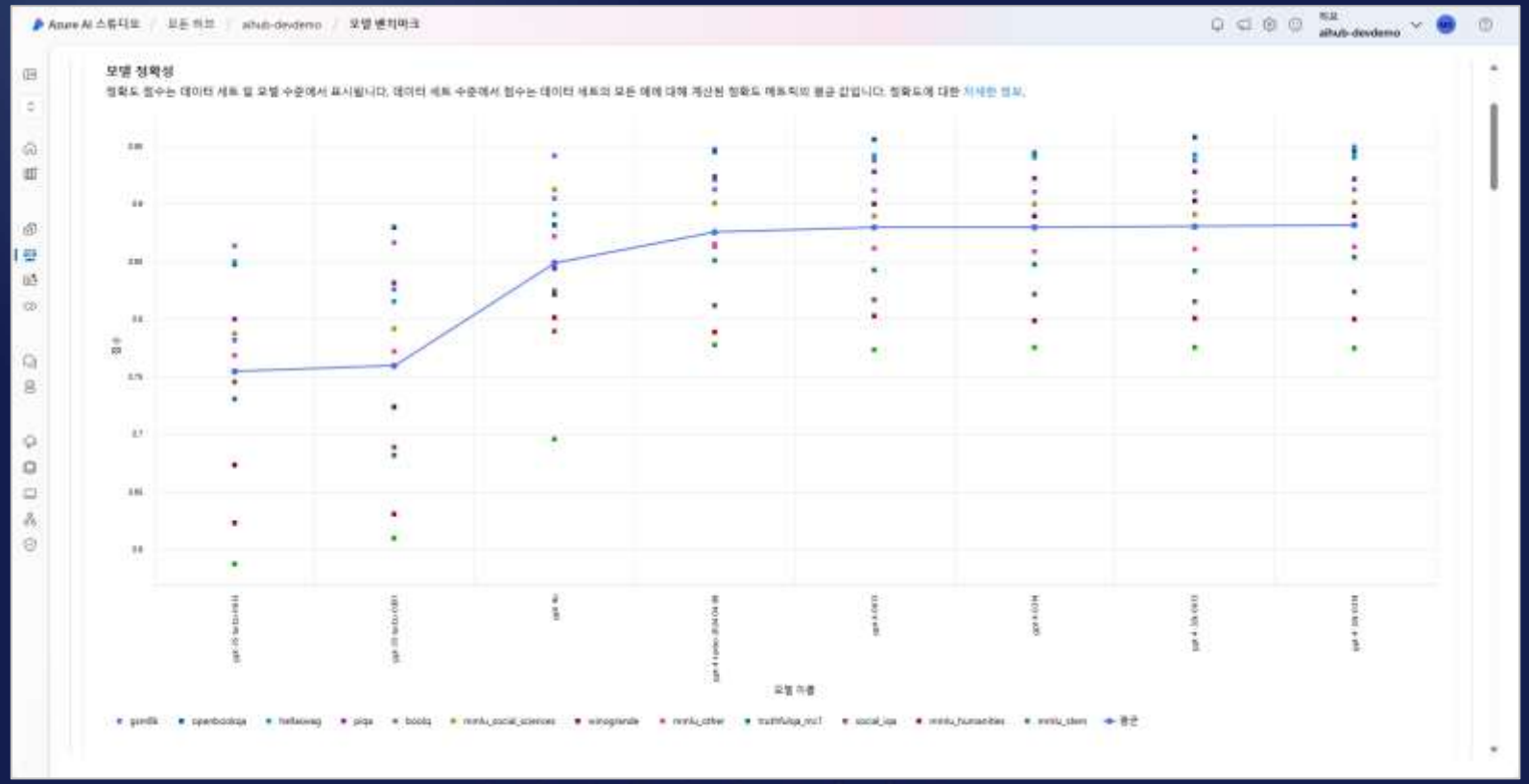
- 모델 및 미세 조정 기능의 전체 카탈로그를 사용한 프로 코드 개발
- 클라우드 인프라를 완전히 제어하는 PaaS 서비스
- 프롬프트 및 모델 오케스트레이션
- 성능, 안정성, 확장성, 책임 있는 AI 안전을 테스트하는 평가 엔진
- 사용자 지정 앱 및 서비스에서 사용하기 위해 Azure에 엔드포인트로 배포

모델 할당량(Quotas) 확인



모델 벤치마크

간 모델 별로 품질을 평가할 수 있는 지표들을 제공한다.



모델 배포

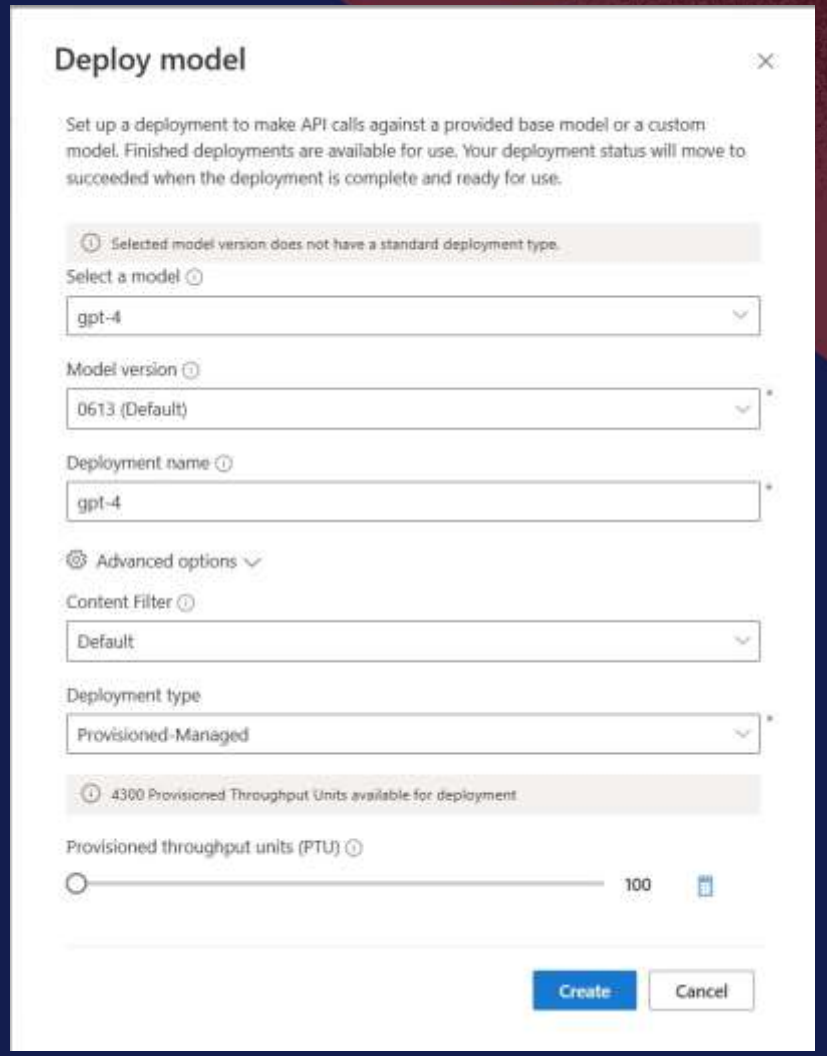
1.Azure AI Foundry 포털에 로그인합니다.

2.프로비전된 배포에 사용하도록 설정된 구독을 선택하고 할당량이 있는 지역에서 원하는 리소스를 선택합니다.

3.왼쪽 탐색 메뉴의 관리에서 배포를 선택합니다.

4.새 배포 만들기를 선택하고 다음 필드를 구성합니다. 고급 옵션 드롭다운 메뉴를 확장합니다.

5.각 필드에 값을 작성합니다. 예를 들면 다음과 같습니다.



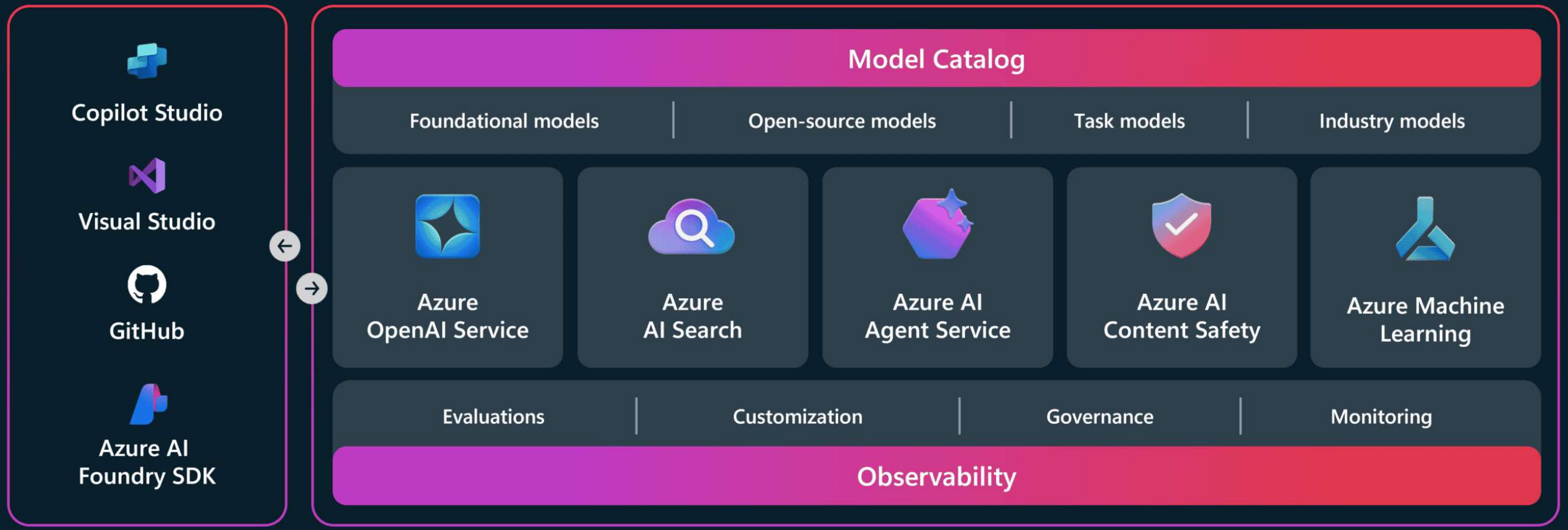
모델 배포

항목	설명	예시
모델 선택	배포하려는 특정 모델을 선택합니다.	**GPT-4**
모델 버전	배포할 모델 버전을 선택합니다.	**0613**
배포 이름	배포 이름은 코드에서 클라이언트 라이브러리 및 **REST API**를 사용하여 모델을 호출하는 데 사용됩니다.	**gpt-4**
콘텐츠 필터	배포에 적용할 필터링 정책을 지정합니다. 콘텐츠 필터링 방법에 대해 자세히 알아봅니다.	기본값
배포 유형	이는 처리량과 성능에 영향을 미칩니다. 배포에 대한 배포 대화 상자 드롭다운에서 전역 프로비전된 관리형, **DataZone** 프로비전 관리 또는 프로비저닝된 관리형 선택	프로비전-관리
프로비전된 처리량 단위	배포에 포함할 처리량을 선택합니다.	**100**

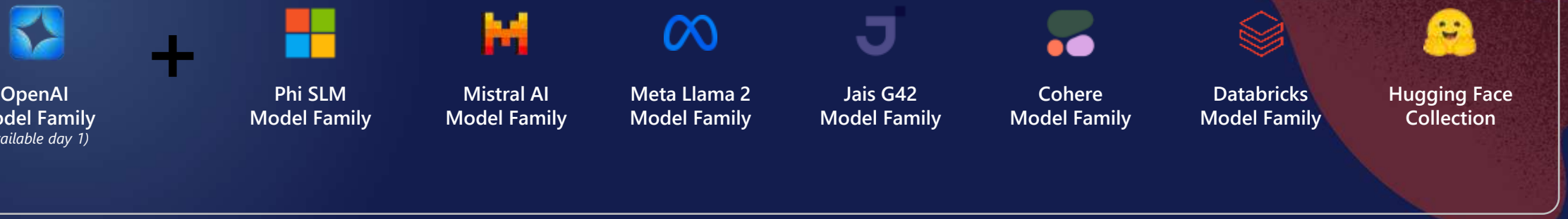
실습: Azure OpenAI 호출

```
import os from openai import AzureOpenAI client =
AzureOpenAI( azure_endpoint =
os.getenv("AZURE_OPENAI_ENDPOINT"),
api_key=os.getenv("AZURE_OPENAI_API_KEY"), api_version="2024-10-
21" ) response = client.chat.completions.create( model="gpt-4", #
model = "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 services support
this too?"} ] ) print(response.choices[0].message.content)
```

Azure AI Foundry



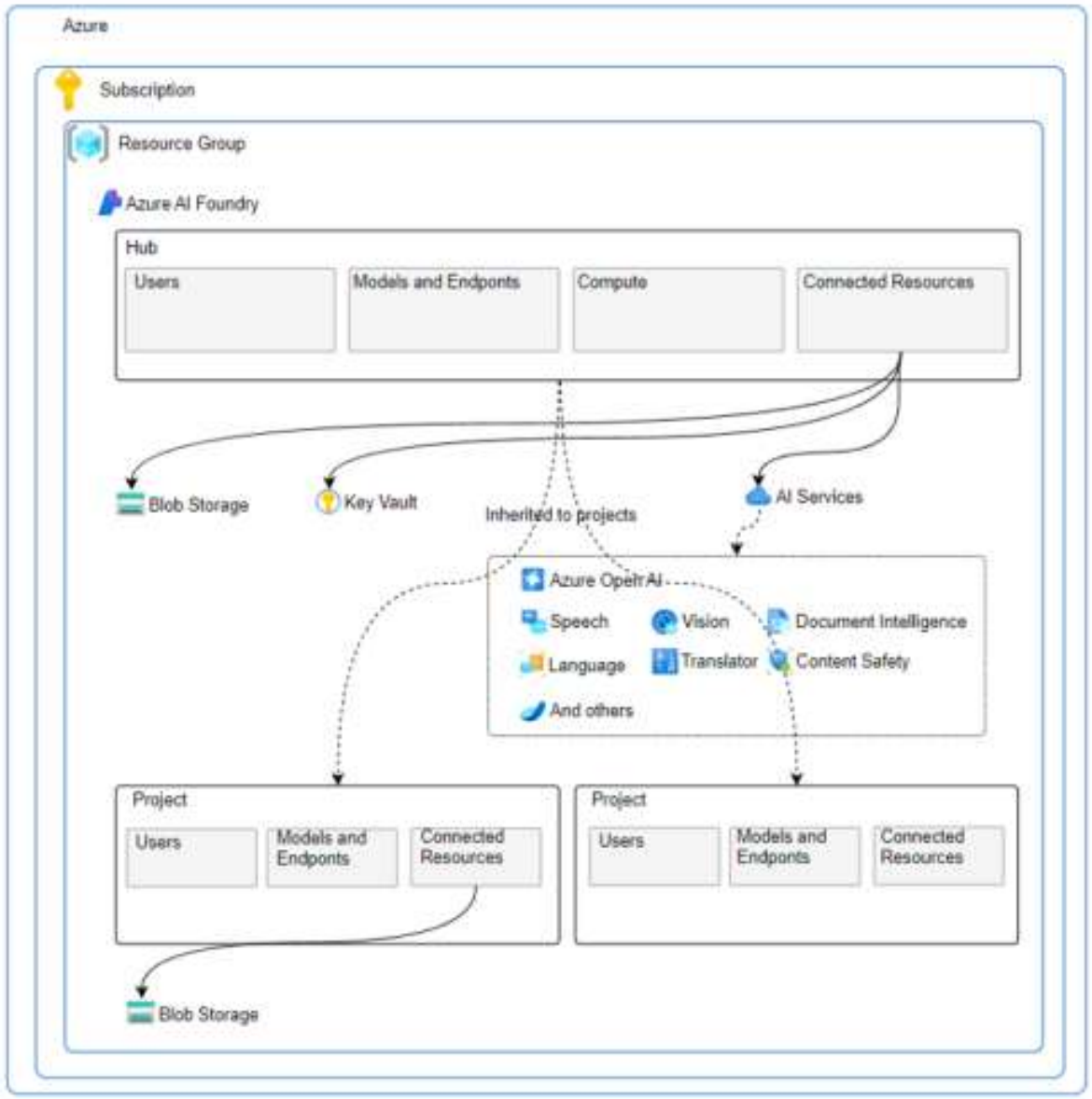
Now 1800 + Frontier, task, and Open Models



- 통합 개발 환경: Azure AI Foundry는 웹 포털, SDK, CLI를 통해 AI 모델을 빌드, 평가, 배포할 수 있는 통합 환경을 제공합니다.
- 다양한 AI 모델 지원: Microsoft, OpenAI, Hugging Face, Meta, Mistral 등 다양한 공급업체의 프런티어 및 오픈 소스 모델을 포함하는 광범위한 모델 카탈로그를 제공합니다.
- 에이전트 관리 및 오케스트레이션: 다양한 AI 에이전트를 통합 관리할 수 있는 도구를 제공하여 반복적인 작업을 자동화하고 사용자 생산성을 향상시킵니다.
- 개발자 친화적인 SDK: AI 애플리케이션과 에이전트를 맞춤화, 테스트, 배포할 수 있는 강력한 기능을 제공하는 SDK를 제공합니다.

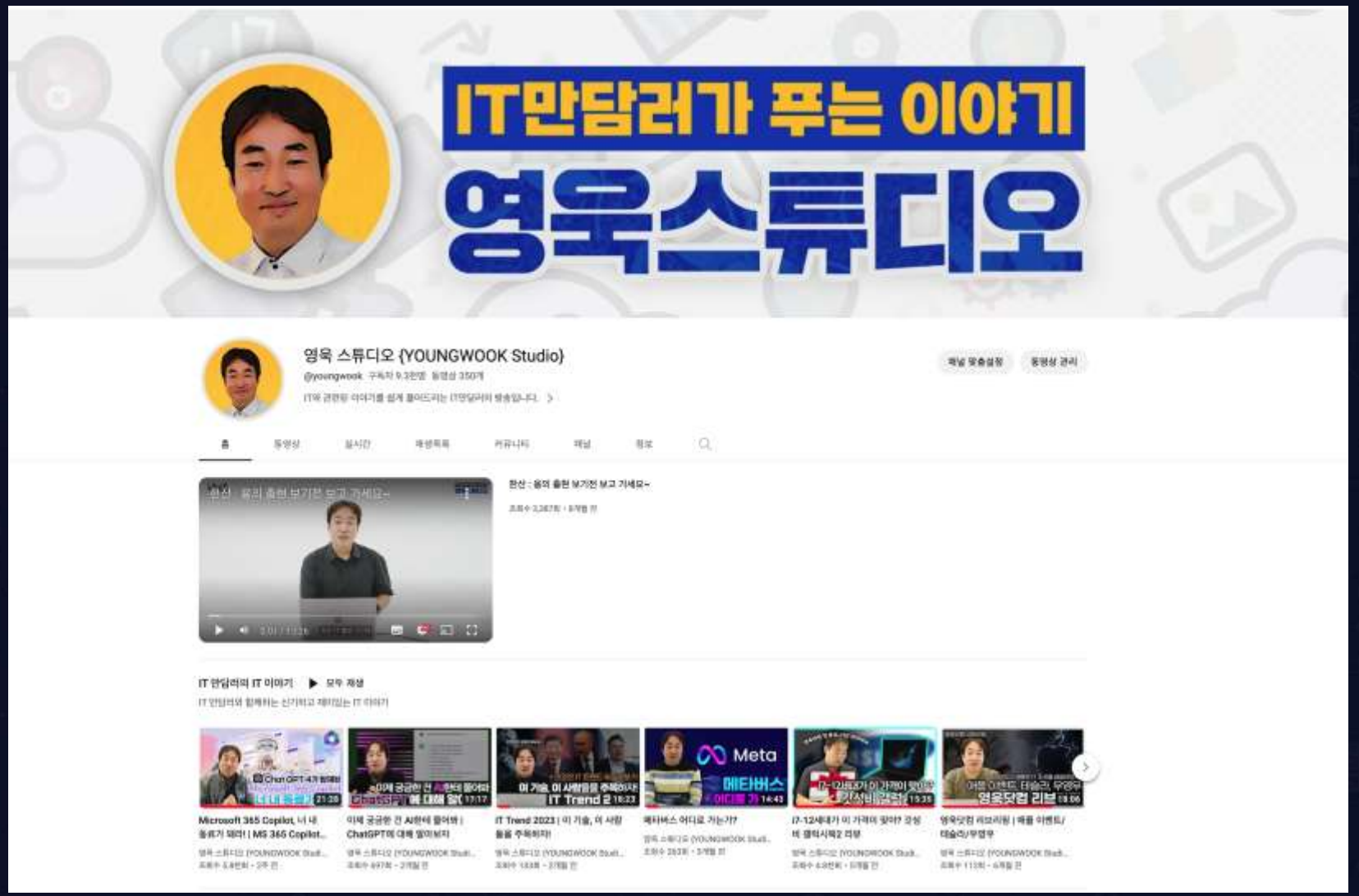
Azure AI Foundry Hubs and Projects

Walkthrough of my implementation of a hub and projects



실습: Azure Foundry 실습

```
import os from openai import AzureOpenAI client =
AzureOpenAI( azure_endpoint =
os.getenv("AZURE_OPENAI_ENDPOINT"),
api_key=os.getenv("AZURE_OPENAI_API_KEY"), api_version="2024-10-
21" ) response = client.chat.completions.create( model="gpt-4", #
model = "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 services support
this too?"} ] ) print(response.choices[0].message.content)
```


IT만담러가 푸는 이야기
영욱스튜디오
영욱 스튜디오 (YOUNGWOOK Studio)
@youngwook
홈
동영상
재생목록
커뮤니티
채널
정보
완산 : 옴의 출현 보기전 보고 가세요~
IT 만담러의 IT 이야기
모두 재생
Microsoft 365 Copilot, 너 내 동료가 돼라! | MS 365 Copilot
이제 궁금한 건 AI한테 물어봐 | ChatGPT에 대해 알아보자
IT Trend 2023 | 이 기술, 이 사람 들을 주목하자!
메타버스 어디로 가는가?
I7-12세대가 이 가격이 맞아? 고성비 갤럭시북2 리뷰
영욱닷컴 리브리뷰 | 배틀 이벤트/테슬라/무엄무