

Azure AI / ML Studio에서 소형 언어 모델 Phi-3를 학습하고 사용하는 방법

MLSA 송민석



MLSA 송민석

인하대학교 산업경영공학과, 인공지능공학과 전공



Microsoft Tech Community 에 블로그 작성

Fine-Tune and Integrate Custom Phi-3 Models with Prompt Flow in Azure AI Studio

Fine-Tune and Integrate Custom Phi-3 Models with Prompt Flow: Step-by-Step guide

Build a Chatbot Service to Ensure Safe Conversations Using Azure Content Safety & Azure OpenAI

Teach ChatGPT to Answer Questions: Azure AI Search & Azure OpenAI (Lang Chain)

Teach ChatGPT to Answer Questions: Azure AI Search & Azure OpenAI (Semantic Kernel)

Microsoft Phi-3 CookBook Contributor

- Microsoft Phi-3 공식 오픈소스 프로젝트에 기여
- 다양한 End-to-End Sample 제작 및 기여

목차

소개: MLSA 송민석



사전에 필요한 세팅

Azure ML Studio에서
Phi-3 파인튜닝하는
방법



Azure AI Studio의
Prompt flow 에서
배포한 모델 연결하는
방법

Azure AI / ML Studio



Properties

| | |
|--|-------------------------------------|
| Status ✔ Completed | Created by Song Minseok |
| Created on Jul 12, 2024 12:40 PM | Job type Pipeline |
| Start time Jul 12, 2024 12:40 PM | Experiment ChatCompletion |
| Duration 32m 18.27s | Arguments None |
| Compute duration 32m 18.27s | Registered models None |

Graph

```
graph TD; inputs([inputs]) --> integrate_with_promptflow[integrate_with_promptflow  
✔ Completed]; integrate_with_promptflow --> outputs([outputs])
```

Chat

Who founded Microsoft?

Microsoft was founded by Bill Gates and Paul Allen on April 4, 1975.
Time spent: 0.77 sec

Input anything to test...

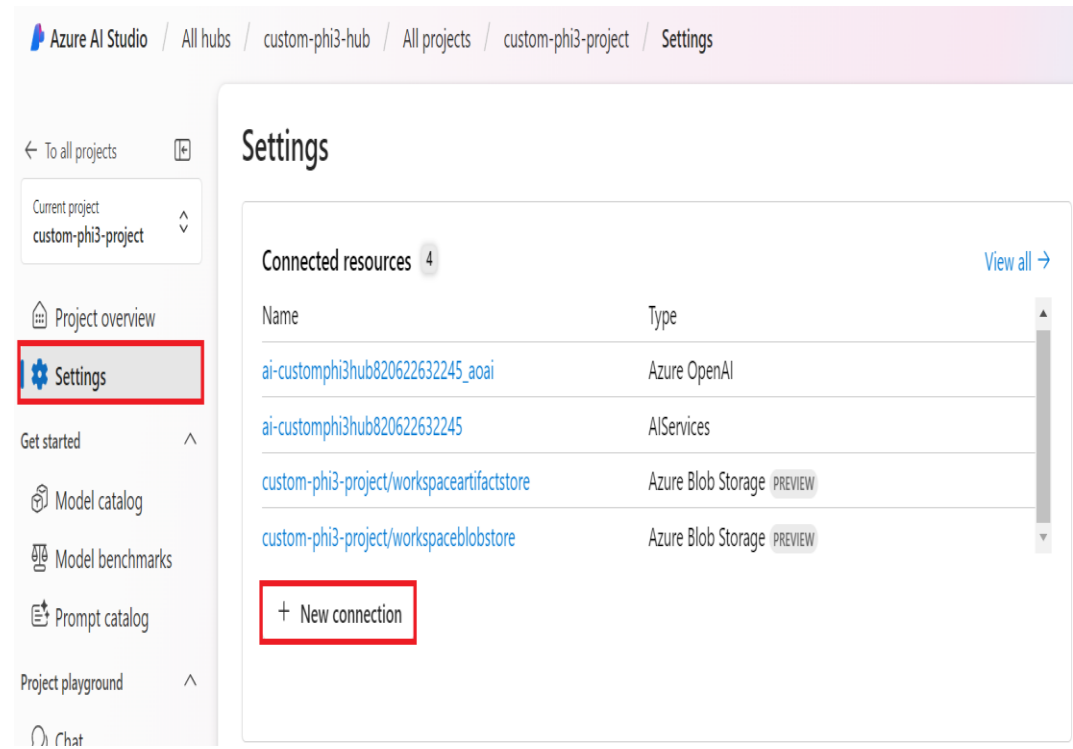
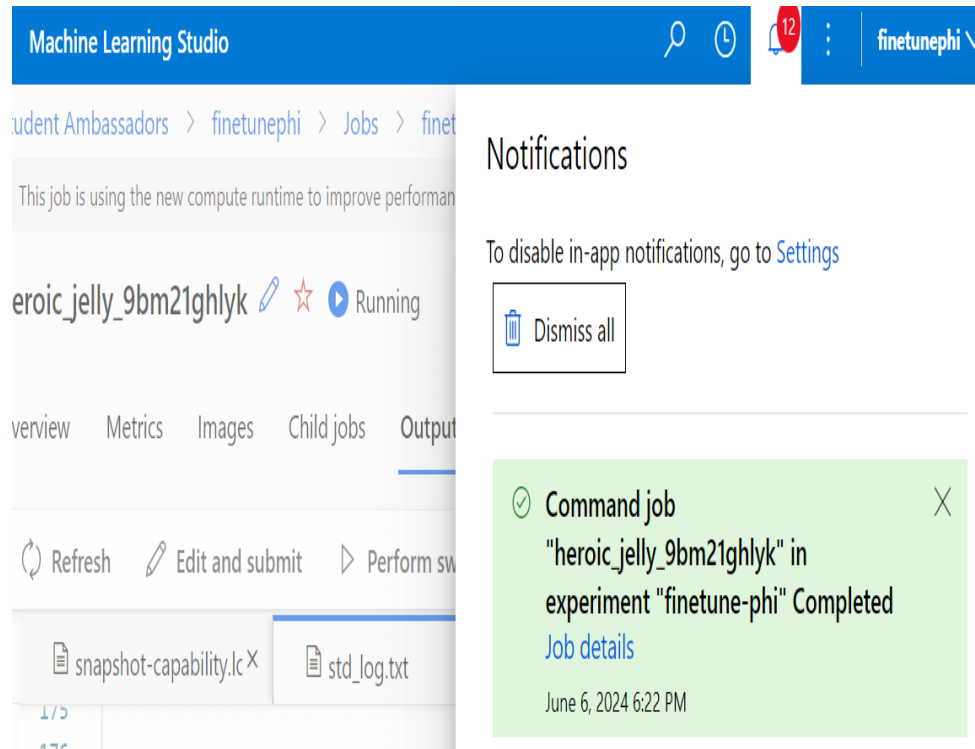
실습 자료 링크

<https://github.com/skytin1004/Fine-tune-phi-aistudio-kr>



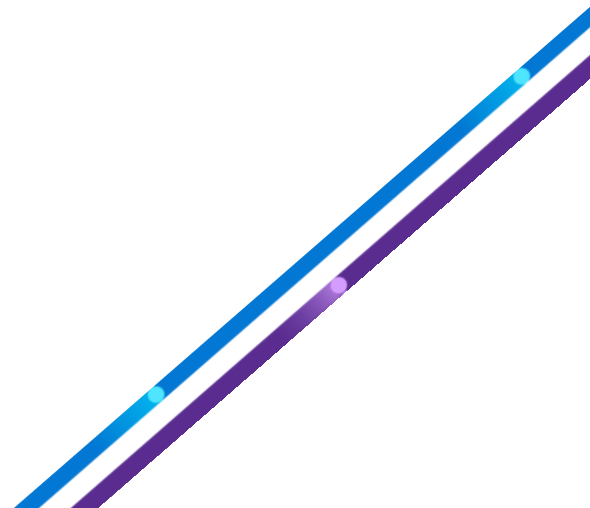
Azure ML Studio / Azure AI Studio 란?

- 기존에 코드작업을 통해 언어 모델을 파인튜닝 및 배포 혹은 평가하는 과정을 웹 서비스에서 쉽게 작업할 수 있게 하는 플랫폼



사전에 필요한 세팅

1. Azure 가입하기
2. Azure Machine Learning resource 만들기
3. Azure Managed Identity 만들기
4. Azure Management Identity에 필수적인 권한 부여하기
5. Azure GPU 신청하기



사전에 필요한 세팅

1. Azure Machine Learning resource 만들기

- Azure 접속
- Azure 검색창에 “Azure Machine Learning” 검색
- Azure Machine Learning 리소스에서 + Create 선택
- 페이지 양식 채우고 Review + create 선택 (관련된 모든 리소스 전부 생성해야함.)

Azure Machine Learning ...

Create a machine learning workspace

Subscription * ⓘ

Resource group * ⓘ
[Create new](#)

Workspace details

Configure your basic workspace settings like its storage connection, authentication, container, and more. [Learn more](#) ⓘ

Name * ⓘ ✓

Region * ⓘ

Storage account * ⓘ
[Create new](#)

Key vault * ⓘ
[Create new](#)

Application insights * ⓘ
[Create new](#)

Container registry ⓘ
[Create new](#)

[Review + create](#) [< Previous](#) [Next : Networking](#)

사전에 필요한 세팅

2. Azure Managed Identity 만들기

- Azure 접속
- Azure 검색창에 “Azure Managed Identity ” 검색
- Azure Managed Identity 리소스에서 + Create 선택
- 페이지 양식 채우고 Review + create 선택

Create User Assigned Managed Identity ...

Basics Tags Review + create

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ

Subscription 1 ▼

Resource group * ⓘ

TestGroup ▼

[Create new](#)

Instance details

Region * ⓘ

East US 2 ▼

Name * ⓘ

finetunephi-managedidentity ✓

Previous

Next

Review + create

사전에 필요한 세팅

2-1. Azure Managed Identity 권한 부여하기 (Contributor)

- Azure Managed Identity 리소스에서 Azure role assignment 선택
- + Add role assignment 선택 후 양식 채우기
- Scope -> ResourceGroup
- Role은 Contributor으로 선택

Home > finetunephi-managedidentity

finetunephi-managedidentity | Azure Managed Identity

Search

Overview

Activity log

Access control (IAM)

Tags

Azure role assignments

Associated resources (preview)

Settings

Federated credentials

Properties

Locks

+ Add role assignment

If this identity has role assignments, you can select one to edit.

Subscription *

Subscription 1

Role

Storage Blob Data Reader

AcrPull

Contributor

Add role assignment (Preview)

Scope ⓘ

Resource group

Subscription

Subscription 1

Resource group ⓘ

TestGroup

Role ⓘ

Contributor ⓘ

[Learn more about RBAC](#)

Save Discard

사전에 필요한 세팅

2-2. Azure Managed Identity 권한 부여하기 (Storage Blob Data Reader)

- Azure 검색창에 “storage account” 검색
- Storage account 리소스에서 이전에 생성했던 storage account 선택
- 왼쪽 메뉴에서 Access Control(IAM) 선택
- + Add -> Add role assignment 선택
- Storage Blob Data Reader 권한을 이전에 생성한 Managed Identity에 부여

[Role](#) [Members](#) [Conditions](#) [Review + assign](#)

A role definition is a collection of permissions. You can use the built-in roles or you can create your own custom roles. [Learn more](#)

[Job function roles](#) Privileged administrator roles

Grant access to Azure resources based on job function, such as the ability to create virtual machines.

| <input type="text" value="Storage Blob Data Reader"/> | | | | | Type : All | Category : All |
|---|---|-------------|-------------|----------------------|------------|----------------|
| Name ↑↓ | Description ↑↓ | Type ↑↓ | Category ↑↓ | Details | | |
| Storage Blob Data Rea... | Allows for read access to Azure Storage ... | BuiltInRole | Storage | View | | |

[Review + assign](#)

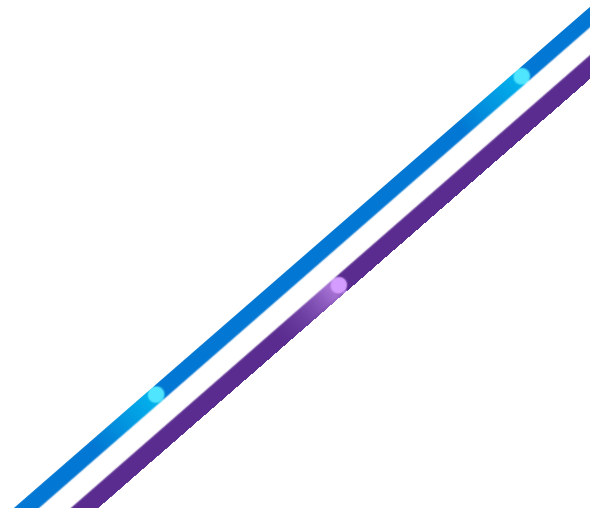
[Previous](#)

[Next](#)

사전에 필요한 세팅

2-3. Azure Managed Identity 권한 부여하기 (ACR Pull)

- Azure 검색창에 “container registry” 검색
- container registry 리소스에서 이전에 생성했던 container registry 선택
- 왼쪽 메뉴에서 Access Control(IAM) 선택
- + Add -> Add role assignment 선택
- Acr Pull 권한을 이전에 생성한 Managed Identity에 부여



사전에 필요한 세팅

3. Azure GPU 신청하기

- [Azure ML Studio](#)에 접속
- 왼쪽 탭에서 **Quota**를 선택
- **Request quota**를 선택
- **Standard NCADSA100v4, Standard NCSv3 Family** 요청

Azure AI | Machine Learning Studio

Search All workspaces

Student Ambassadors > Quota > Subscription 1 > East US 2

Subscription 1 (East US 2)

Request quota Configure workspace quota

Subscription
Dedicated quota ⓘ

0 cores used | 68 cores available

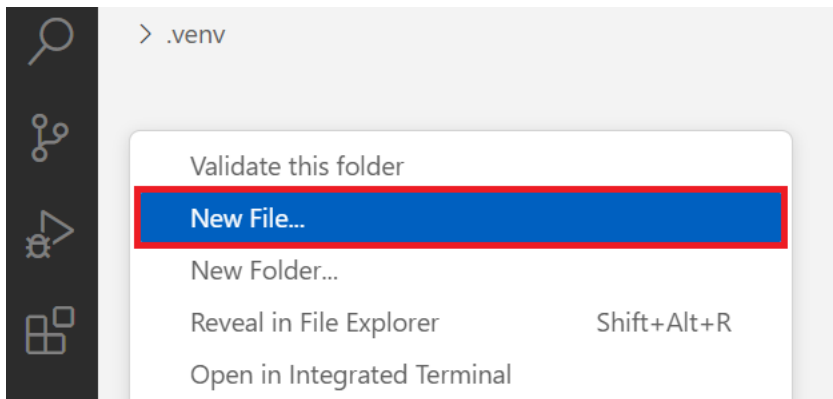
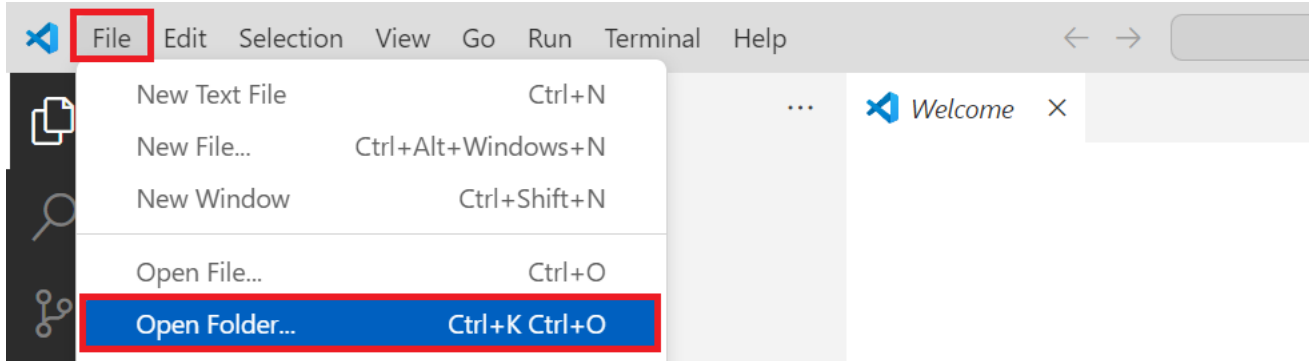
Number of dedicated cores you can use for your subscription

Each virtual machine family has a limit on the number of dedicated cores which can be used at any given moment, regardless of how many total cores are be

| Virtual machine family | Usage | Quota ↓ |
|--|--------------|----------------------|
| Standard NCSv3 Family Cluster Dedicated vCPUs | 0 cores used | 24 cores available ⓘ |
| Standard NCADSA100v4 Family Cluster Dedicated vCPUs | 0 cores used | 24 cores available ⓘ |

Quota

예시: 데이터셋 다운로드




Visual
Editing e



예시: 파인튜닝 하기




※GPU 할당이 승인되면 파인튜닝을 진행할 수 있습니다.

phi3-workspace > Model catalog > azureml > Phi-3-mini

instruct  PREVIEW

Artifacts Security

 Fine-tuning task: chat-completion  Languages: en

luate  Fine-tune  Deploy  View license

it is a 3.8B parameters, lightweight, state-of-the-art open model assets that includes both synthetic data and the filtered publicly th a focus on high-quality and reasoning dense properties. The -3 family with the Mini version in two variants 4K and 128K which (tokens) that it can support.


: a post-training process that incorporates both supervised fine- nce optimization for the instruction following and safety

Fine-tune Phi-3-mini-4k-instruct


Quickly fine-tune this model with your (or our) data or sample data

Select task type *

Chat completion


Training data * 

+ Select data

Validation data * 

Provide different validation data

+ Select data

Select Azure ML compute cluster * 

StandardNC24adsA100v4

+ New

Finish

Cancel

Advanced settings

예시: 파인튜닝한 모델 배포

※ 현재 배포과정에 시스템 오류가 생기고 있음,
MS에서 문제 해결하면 해보기 추천

The screenshot displays the Azure AI Machine Learning Studio interface. The left sidebar contains a navigation menu with the following items: Jobs, Components, Pipelines, Environments, Models, Endpoints (highlighted with a red box), Manage, Compute, Monitoring, Data Labeling, Linked Services (PREVIEW), and Connections (PREVIEW). The main content area shows the breadcrumb path: Student Ambassadors > phi3-workspace > Endpoints > fine-tuned-phi3-endpoint. Below the breadcrumb, the title 'fine-tuned-phi3-endpoint' is displayed. A tab bar at the top of the main content area includes 'Details', 'Test', 'Consume' (highlighted with a red box), 'Monitoring', 'PREVIEW', and 'Logs'. The 'Basic consumption info' section contains the following information:

- REST endpoint:** A text box containing the URL `https://fine-tuned-phi3-endpoint.eastus2.inference.ml.azure.com/score`, which is highlighted with a red box. A copy icon is visible to the right of the text box.
- Authentication:**
 - Primary key:** A text box containing a masked key (dots), highlighted with a red box. To its right is a copy icon and a 'Regenerate' button.
 - Secondary key:** A text box containing a masked key (dots). To its right is a copy icon and a 'Regenerate' button.

예시: 프롬프트 플로우와 결합해서 채팅하기

※ 현재 배포과정에 시스템 오류가 생기고 있음,
MS에서 문제 해결하면 해보기 추천

The screenshot displays a web-based development environment. The top navigation bar includes buttons for 'Clone', 'Save', 'Deploy', 'Evaluate', and 'Chat'. The 'Compute session running' status is indicated by a green checkmark. The left pane shows a code editor with the following Python code:

```
nt_url, json=data, headers=headers)

se: {response.json()}")

t"]
ved response from Azure ML Endpoint.")

xception as e:
zure ML Endpoint: {e}")
```

The right pane features a chat interface. The chat header shows 'Chat' with an information icon and window controls. A blue message box contains the question: 'Who founded Microsoft?'. Below it, a grey response box contains the answer: 'Microsoft was founded by Bill Gates and Paul Allen on April 4, 1975.' and the response time: 'Time spent: 0.77 sec'. At the bottom, there is an input field with the placeholder text 'Input anything to test...' and a send button.

감사합니다

