

Azure AI services documentation

Build cutting-edge, market-ready, responsible applications for your organization with AI.



OVERVIEW

What are Azure AI services?



OVERVIEW

What is Azure AI Studio?



QUICKSTART

Chat with Azure OpenAI models using your own data



CONCEPT

Responsible use of AI



Azure OpenAI

Perform a wide variety of natural language tasks.



Azure AI Search

Bring AI-powered cloud search to your mobile and web applications.



Content Safety

An AI service that detects unwanted contents



Speech

Speech to text, text to speech, translation, and speaker recognition



Document Intelligence

Turn documents into intelligent data-driven solutions.



Vision

Analyze content in images and videos.



Custom Vision

Customize image recognition for your business.



Face

Detect and identify people and emotions in images.



Translator

Use AI-powered translation technology to translate more than 100 in-use, at-risk, and endangered languages and dialects.



Language

Build apps with industry-leading natural language understanding capabilities.



Video Indexer

Extract actionable insights from your videos.



Immersive Reader

Help users read and comprehend text.

Further resources

Demo & customization studios

- [Azure AI Studio ↗](#)
- [Azure OpenAI Studio ↗](#)
- [Content Safety ↗](#)
- [Speech ↗](#)
- [Document Intelligence ↗](#)
- [Vision ↗](#)
- [Custom Vision ↗](#)
- [Custom Translator ↗](#)
- [Language ↗](#)

Explore more AI resources

- [Azure AI Studio](#)
- [Azure Machine Learning](#)
- [Semantic Kernel](#)
- [AI Builder](#)
- [Windows AI](#)
- [GitHub Copilot ↗](#)

Training & certification

[AI learning and community hub](#)

Identify principles and practices for responsible AI

[AI Learning paths and modules](#)

[AI Engineer career path](#)

Deprecated Azure AI services

[Content Moderator](#)

[Language Understanding \(LUIS\)](#)

[QnA Maker](#)

[Metrics Advisor](#)

[Anomaly Detector](#)

[Personalizer](#)

What are Azure AI services?

Article • 08/28/2024

Azure AI services help developers and organizations rapidly create intelligent, cutting-edge, market-ready, and responsible applications with out-of-the-box and prebuilt and customizable APIs and models. Example applications include natural language processing for conversations, search, monitoring, translation, speech, vision, and decision-making.

💡 Tip

Try Azure AI services including Azure OpenAI, Content Safety, Speech, Vision, and more in [Azure AI Studio](#). For more information, see [What is Azure AI Studio?](#).

Most [Azure AI services](#) are available through REST APIs and client library SDKs in popular development languages. For more information, see each service's documentation.

Available Azure AI services

When building AI applications, use the following Azure AI services:

[+] Expand table

Service	Description
 Azure AI Search	Bring AI-powered cloud search to your mobile and web apps.
 Azure OpenAI	Perform a wide variety of natural language tasks.
 Bot Service	Create bots and connect them across channels.
 Content Safety	An AI service that detects unwanted contents.
 Custom Vision	Customize image recognition for your business.
 Document Intelligence	Turn documents into intelligent data-driven solutions.
 Face	Detect and identify people and emotions in images.
 Immersive Reader	Help users read and comprehend text.

Service	Description
 Language	Build apps with industry-leading natural language understanding capabilities.
 Speech	Speech to text, text to speech, translation, and speaker recognition.
 Translator	Use AI-powered translation technology to translate more than 100 in-use, at-risk, and endangered languages and dialects.
 Video Indexer	Extract actionable insights from your videos.
 Vision	Analyze content in images and videos.

The following Azure AI services are scheduled for retirement. These services are still available for existing applications but don't use them for new AI applications:

[\[+\] Expand table](#)

Service	Description
 Anomaly Detector (retired)	Identify potential problems early on.
 Content Moderator (retired)	Detect potentially offensive or unwanted content.
 Language understanding (retired)	Understand natural language in your apps.
 Metrics Advisor (retired)	An AI service that detects unwanted contents.
 Personalizer (retired)	Create rich, personalized experiences for each user.
 QnA maker (retired)	Distill information into easy-to-navigate questions and answers.

Pricing tiers and billing

Pricing tiers (and the amount you get billed) are based on the number of transactions you send using your authentication information. Each pricing tier specifies the:

- Maximum number of allowed transactions per second (TPS).
- Service features enabled within the pricing tier.
- Cost for a predefined number of transactions. Going above this number causes an extra charge as specified in the [pricing details](#) for your service.

 Note

Many of the Azure AI services have a free tier you can use to try the service. To use the free tier, use `F0` as the SKU for your resource.

Development options

The tools that you can use to customize and configure models are different from tools that you use to call the Azure AI services. Out of the box, most Azure AI services allow you to send data and receive insights without any customization. For example:

- You can send an image to the Azure AI Vision service to detect words and phrases or count the number of people in the frame
- You can send an audio file to the Speech service and get transcriptions and translate the speech to text at the same time

Azure offers a wide range of tools that are designed for different types of users, many of which can be used with Azure AI services. Designer-driven tools are the easiest to use, and are quick to set up and automate, but might have limitations when it comes to customization. Our REST APIs and client libraries provide users with more control and flexibility, but require more effort, time, and expertise to build a solution. If you use REST APIs and client libraries, there's an expectation that you're comfortable working with modern programming languages like C#, Java, Python, JavaScript, or another popular programming language.

Let's take a look at the different ways that you can work with the Azure AI services.

Client libraries and REST APIs

Azure AI services client libraries and REST APIs provide direct access to your service. These tools provide programmatic access to the Azure AI services, their baseline models, and in many cases allow you to programmatically customize your models and solutions.

- **Target user(s):** Developers and data scientists
- **Benefits:** Provides the greatest flexibility to call the services from any language and environment
- **UI:** N/A - Code only
- **Subscription(s):** Azure account + Azure AI services resources

If you want to learn more about available client libraries and REST APIs, use our [Azure AI services overview](#) to pick a service and get started with one of our quickstarts.

Continuous integration and deployment

You can use Azure DevOps and GitHub Actions to manage your deployments. In the [following section](#), we have two examples of CI/CD integrations to train and deploy custom models for Speech and the Language Understanding (LUIS) service.

- **Target user(s):** Developers, data scientists, and data engineers
- **Benefits:** Allows you to continuously adjust, update, and deploy applications and models programmatically. There's significant benefit when regularly using your data to improve and update models for Speech, Vision, Language, and Decision
- **UI tools:** N/A - Code only
- **Subscription(s):** Azure account + Azure AI services resource + GitHub account

Continuous integration and delivery with DevOps and GitHub Actions

Language Understanding and the Speech service offer continuous integration and continuous deployment solutions that are powered by Azure DevOps and GitHub Actions. These tools are used for automated training, testing, and release management of custom models.

- [CI/CD for Custom Speech](#)
- [CI/CD for LUIS](#)

On-premises containers

Many of the Azure AI services can be deployed in containers for on-premises access and use. Using these containers gives you the flexibility to bring Azure AI services closer to your data for compliance, security, or other operational reasons. For a complete list of Azure AI containers, see [On-premises containers for Azure AI services](#).

Training models

Some services allow you to bring your own data, then train a model. Trained custom models allow you to extend the model using the service's data and algorithm with your own data. The output matches your needs. When you bring your own data, you might need to tag the data in a way specific to the service. For example, if you're training a model to identify flowers, you can provide a catalog of flower images along with the location of the flower in each image to train the model.

Azure AI services in the ecosystem

With Azure and Azure AI services, you have access to a broad ecosystem, such as:

- Automation and integration tools like Logic Apps and Power Automate.
- Deployment options such as Azure Functions and the App Service.
- Azure AI services Docker containers for secure access.
- Tools like Apache Spark, Azure Databricks, Azure Synapse Analytics, and Azure Kubernetes Service for big data scenarios.

To learn more, see [Azure AI services ecosystem](#).

Regional availability

The APIs in Azure AI services are hosted on a growing network of Microsoft-managed data centers. You can find the regional availability for each API in [Azure region list](#).

Looking for a region we don't support yet? Let us know by filing a feature request on our [UserVoice forum](#).

Language support

Azure AI services support a wide range of cultural languages at the service level. You can find the language availability for each API in the [supported languages list](#).

Security

Azure AI services provide a layered security model, including [authentication](#) with Microsoft Entra credentials, a valid resource key, and [Azure Virtual Networks](#).

Certifications and compliance

Azure AI services awarded certifications include Cloud Security Alliance STAR Certification, FedRAMP Moderate, and HIPAA BAA.

To understand privacy and data management, go to the [Trust Center](#).

Help and support

Azure AI services provide several support options to help you move forward with creating intelligent applications. Azure AI services also have a strong community of

developers that can help answer your specific questions. For a full list of support options available to you, see [Azure AI services support and help options](#).

Next steps

- Learn how to [get started with Azure](#)
 - [Try Azure AI services and more in Azure AI Studio](#)
 - [Plan and manage costs for Azure AI services](#)
-

Feedback

Was this page helpful?



[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

Natural language support for Azure AI services

Article • 08/28/2024

Azure AI services enable you to build applications that see, hear, articulate, and understand users. Our language support capabilities enable users to communicate with your applications in natural ways and empower global outreach. Use the links in the tables to view language support and availability by service.

Language supported services

The following table provides links to language support reference articles by supported service.

[] Expand table

Azure AI Language support	Description
 Content Moderator (retired)	Detect potentially offensive or unwanted content.
 Document Intelligence	Turn documents into intelligent data-driven solutions.
 Immersive Reader	Help users read and comprehend text.
 Language service	Build apps with industry-leading natural language understanding capabilities.
 Language Understanding (LUIS) (retired)	Understand natural language in your apps.
 QnA Maker (retired)	Distill information into easy-to-navigate questions and answers.
 Speech Service	Configure speech-to-text, text-to-speech, translation, and speaker recognition applications.
 Translator	Translate more than 100 in-use, at-risk, and endangered languages and dialects.
 Video Insights	Extract actionable insights from your videos.

Azure AI Language support	Description
Video Indexer	
 Vision	Analyze content in images and videos.

Language independent services

These Azure AI services are language agnostic and don't have limitations based on human language.

[Expand table](#)

Azure AI service	Description
 Anomaly Detector	Identify potential problems early on.
 Custom Vision	Customize image recognition for your business.
 Face	Detect and identify people and emotions in images.
 Personalizer	Create rich, personalized experiences for users.

See also

- [What are Azure AI services?](#)
- [How to create an Azure AI services resource](#)

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

Quickstart: Create an Azure AI services resource

Article • 08/28/2024

Learn how to create and manage an Azure AI services resource. An Azure AI services resource allows you to access multiple Azure AI services with a single set of credentials.

You can access Azure AI services through two different resource kinds:

- Azure AI services multi-service resource:
 - Access multiple Azure AI services with a single set of credentials.
 - Consolidates billing from the services you use.
- Single-service resource such as Face and Vision:
 - Access a single Azure AI service with a unique set of credentials for each service created.
 - Most Azure AI services offer a free tier to try it out.

Azure AI services are Azure [resources](#) that you create under your Azure subscription.

After you create a resource, you can use the keys and endpoint generated to authenticate your applications.

Supported services with a multi-service resource

The multi-service resource enables access to the following Azure AI services with a single set of credentials. Some services are available via the multi-service resource and single-service resource.

Tip

We recommend whenever possible to use the **Azure AI services** resource (where the API kind is `AI Services`) to access multiple Azure AI services with a single set of credentials. For services not available via the multi-service resource (such as Face and Custom Vision), you can create a single-service resource.

Expand table

Service	Description	Kind (via API)
 Azure OpenAI	Perform a wide variety of natural language tasks.	AI Services OpenAI
 Content Safety	An AI service that detects unwanted contents.	AI Services ContentSafety
 Custom Vision	Customize image recognition for your business.	CustomVision.Prediction (Prediction only) CustomVision.Training (Training only)
 Document Intelligence	Turn documents into intelligent data-driven solutions.	AI Services FormRecognizer
 Face	Detect and identify people and emotions in images.	Face
 Language	Build apps with industry-leading natural language understanding capabilities.	AI Services TextAnalytics
 Speech	Speech to text, text to speech, translation, and speaker recognition.	AI Services Speech
 Translator	Use AI-powered translation technology to translate more than 100 in-use, at-risk, and endangered languages and dialects.	AI Services TextTranslation
 Vision	Analyze content in images and videos.	AI Services (Training and Prediction) ComputerVision

Prerequisites

- A valid Azure subscription - [Create one for free ↗](#).

Create a new Azure AI services resource

The Azure AI services multi-service resource is listed under **Azure AI services > Azure AI services** in the portal. Look for the logo as shown here:

The screenshot shows the Azure AI services portal. In the top left, there's a blue square icon with a white 'A' and a red border. To its right, the text 'Azure AI services | Azure AI services' is displayed, followed by a gear icon and three dots. Below this, a search bar contains the placeholder 'Search'. To the right of the search bar are buttons for '+ Create', 'Manage deleted resources', and 'Manage view'. A sidebar on the left lists several options: 'Overview', 'All Azure AI services', 'Azure AI services' (which is expanded to show 'Azure AI services' again, highlighted with a red box), 'Azure OpenAI', and 'AI Search'. On the right side of the main area, there's a 'Filter for any field...' input, an 'Add filter' button, a message 'Showing 0 to 0 of 0 records.', a 'List view' dropdown set to 'List view', and sorting options for 'Name ↑↓', 'Kind ↑↓', and 'Location ↑↓' with a magnifying glass icon. A 'No grouping' button is also present.

ⓘ Important

Azure provides more than one resource kinds named Azure AI services. Be sure to select the one that is listed under **Azure AI services > Azure AI services** with the logo as shown previously.

To create an Azure AI services resource follow these instructions:

1. Select this link to create an **Azure AI services** resource:

<https://portal.azure.com/#create/Microsoft.CognitiveServicesAI Services>

2. On the **Create** page, provide the following information:

[+] Expand table

Project details	Description
Subscription	Select one of your available Azure subscriptions.
Resource group	The Azure resource group that will contain your Azure AI services resource. You can create a new group or add it to a pre-existing group.
Region	The location of your Azure AI service instance. Different locations may introduce latency, but have no impact on the runtime availability of your resource.
Name	A descriptive name for your Azure AI services resource. For example, <i>MyAI Services Resource</i> .
Pricing tier	The cost of your Azure AI services account depends on the options you choose and your usage. For more information, see the API pricing

Project details	Description
	details  .

Home > Azure AI services >

Create Azure AI services

Basics Network Identity Tags Review + create

Get access to Vision, Language, Search, and Speech Azure AI services with a single API key. Quickly connect services together to achieve more insights into your content and easily integrate with other services like Azure Search.

[Learn more](#)

Project Details

Subscription * 

Resource group * 

[Create new](#)

Instance Details

Region 

Name * 

✓

 Location specifies the region only for included regional services. This does not specify a region for included non-regional services. Click here for more details.

Pricing tier * 

- Configure other settings for your resource as needed, read and accept the conditions (as applicable), and then select **Review + create**.

Tip

If your subscription doesn't allow you to create an Azure AI services resource, you may need to enable the privilege of that [Azure resource provider](#) using the [Azure portal](#), [PowerShell command](#) or an [Azure CLI command](#). If you are not the subscription owner, ask someone with the role of *Owner* or *Admin* to complete the registration for you or ask for the `/register/action` privileges to be granted to your account.

Clean up resources

If you want to clean up and remove an Azure AI services subscription, you can delete the resource or resource group. Deleting the resource group also deletes any other resources contained in the group.

1. In the Azure portal, expand the menu on the left side to open the menu of services, and choose **Resource Groups** to display the list of your resource groups.
2. Locate the resource group containing the resource to be deleted.
3. If you want to delete the entire resource group, select the resource group name. On the next page, Select **Delete resource group**, and confirm.
4. If you want to delete only the Azure AI services resource, select the resource group to see all the resources within it. On the next page, select the resource that you want to delete, select the ellipsis menu for that row, and select **Delete**.

Pricing

Pricing tiers (and the amount you're billed) are based on the number of transactions that you send by using your authentication information. Each pricing tier specifies the:

- Maximum number of allowed transactions per second (TPS).
- Service features enabled within the pricing tier.
- Cost for a predefined number of transactions. Going above this number will cause an extra charge, as specified in the [pricing details](#) for your service.

Related content

- Go to the [Azure AI services hub](#).
- Try AI services in the [Azure AI Studio](#).

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

Create an Azure AI services resource using Bicep

Article • 08/28/2024

Follow this quickstart to create Azure AI services resource using Bicep.

Azure AI services help developers and organizations rapidly create intelligent, cutting-edge, market-ready, and responsible applications with out-of-the-box and prebuilt and customizable APIs and models. Example applications include natural language processing for conversations, search, monitoring, translation, speech, vision, and decision-making.

💡 Tip

Try Azure AI services including Azure OpenAI, Content Safety, Speech, Vision, and more in [Azure AI Studio](#). For more information, see [What is Azure AI Studio?](#).

Most [Azure AI services](#) are available through REST APIs and client library SDKs in popular development languages. For more information, see each service's documentation.

[Bicep](#) is a domain-specific language (DSL) that uses declarative syntax to deploy Azure resources. It provides concise syntax, reliable type safety, and support for code reuse. Bicep offers the best authoring experience for your infrastructure-as-code solutions in Azure.

Things to consider

Using Bicep to create an Azure AI services resource lets you create a multi-service resource. This enables you to:

- Access multiple Azure AI services with a single key and endpoint.
- Consolidate billing from the services you use.

Prerequisites

- If you don't have an Azure subscription, [create one for free](#).

Review the Bicep file

The Bicep file used in this quickstart is from [Azure Quickstart Templates](#).

```
Bicep

@description('That name is the name of our application. It has to be unique. Type a name followed by your resource group name. (<name>-<resourceGroupName>)')
param aiServicesName string =
'aiServices-${uniqueString(resourceGroup().id)}'

@description('Location for all resources.')
param location string = resourceGroup().location

@allowed([
's0'
])
param sku string = 'S0'

resource account 'Microsoft.CognitiveServices/accounts@2023-05-01' = {
  name: aiServicesName
  location: location
  identity: {
    type: 'SystemAssigned'
  }
  sku: {
    name: sku
  }
  kind: 'AI Services'
  properties: {
    publicNetworkAccess: 'Disabled'
    networkAcls: {
      defaultAction: 'Deny'
    }
    disableLocalAuth: true
  }
}
```

One Azure resource is defined in the Bicep file. The `kind` field in the Bicep file defines the type of resource.

As needed, change the `sku` parameter value to the [pricing](#) instance you want. The `sku` depends on the resource `kind` that you use. For example, use `TextAnalytics` for the Azure AI Language service. The `TextAnalytics` kind uses `s` instead of `s0` for the `sku` value.

Deploy the Bicep file

1. Save the Bicep file as **main.bicep** to your local computer.
2. Deploy the Bicep file using either Azure CLI or Azure PowerShell.

CLI

```
Azure CLI  
az group create --name exampleRG --location eastus  
az deployment group create --resource-group exampleRG --template-file main.bicep
```

When the deployment finishes, you should see a message indicating the deployment succeeded.

Review deployed resources

Use the Azure portal, Azure CLI, or Azure PowerShell to list the deployed resources in the resource group.

CLI

```
Azure CLI  
az resource list --resource-group exampleRG
```

Clean up resources

When no longer needed, use the Azure portal, Azure CLI, or Azure PowerShell to delete the resource group and its resources.

CLI

```
Azure CLI  
az group delete --name exampleRG
```

Related content

- Authenticate requests to Azure AI services.
 - What are Azure AI services?
 - Natural language support
 - Use Azure AI services as containers.
 - Plan and manage costs for Azure AI services.
-

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

Quickstart: Create an Azure AI services resource by using an ARM template

Article • 08/28/2024

This quickstart shows you how to use an Azure Resource Manager template (ARM template) to create a resource in Azure AI services.

Azure AI services help developers and organizations rapidly create intelligent, cutting-edge, market-ready, and responsible applications with out-of-the-box and prebuilt and customizable APIs and models. Example applications include natural language processing for conversations, search, monitoring, translation, speech, vision, and decision-making.

💡 Tip

Try Azure AI services including Azure OpenAI, Content Safety, Speech, Vision, and more in [Azure AI Studio](#). For more information, see [What is Azure AI Studio?](#).

Most [Azure AI services](#) are available through REST APIs and client library SDKs in popular development languages. For more information, see each service's documentation.

An [Azure Resource Manager template](#) is a JavaScript Object Notation (JSON) file that defines the infrastructure and configuration for your project. The template uses declarative syntax. You describe your intended deployment without writing the sequence of programming commands to create the deployment.

Prerequisites

- If you don't have an Azure subscription, [create one for free](#).

Review the template

The template that you use in this quickstart is from [Azure Quickstart Templates](#).

JSON

```
{  
  "$schema": "https://schema.management.azure.com/schemas/2019-04-  
  01/deploymentTemplate.json#",  
  "resources": [  
    {  
      "type": "Microsoft.Web/sites",  
      "name": "my-new-site",  
      "location": "West US",  
      "apiVersion": "2019-04-01",  
      "dependsOn": [  
        "Microsoft.Storage/storageAccounts",  
        "Microsoft.Web/sites/config"  
      ],  
      "properties": {  
        "siteConfig": {  
          "appSettings": [  
            {  
              "name": "WEBSITE_CONTENT_TYPE",  
              "value": "text/html"  
            }  
          ]  
        },  
        "scmSiteName": "my-new-site",  
        "serverFarmId": "/subscriptions/.../resourceGroups/.../providers/Microsoft.Web/serverFarms/.../sites/.../scmSite"  
      }  
    }  
  ]  
}
```

```
"contentVersion": "1.0.0.0",
"metadata": {
    "_generator": {
        "name": "bicep",
        "version": "0.29.47.4906",
        "templateHash": "6912458133303539897"
    }
},
"parameters": {
    "aiServicesName": {
        "type": "string",
        "defaultValue": "[format('aiServices-{0}', uniqueString(resourceGroup().id))]",
        "metadata": {
            "description": "That name is the name of our application. It has to be unique.Type a name followed by your resource group name. (<name>-<resourceGroupName>)"
        }
    },
    "location": {
        "type": "string",
        "defaultValue": "[resourceGroup().location]",
        "metadata": {
            "description": "Location for all resources."
        }
    },
    "sku": {
        "type": "string",
        "defaultValue": "S0",
        "allowedValues": [
            "S0"
        ]
    }
},
"resources": [
{
    "type": "Microsoft.CognitiveServices/accounts",
    "apiVersion": "2023-05-01",
    "name": "[parameters('aiServicesName')]",
    "location": "[parameters('location')]",
    "identity": {
        "type": "SystemAssigned"
    },
    "sku": {
        "name": "[parameters('sku')]"
    },
    "kind": "AI Services",
    "properties": {
        "publicNetworkAccess": "Disabled",
        "networkAcls": {
            "defaultAction": "Deny"
        },
        "disableLocalAuth": true
    }
}
]
```

```
]  
}
```

One Azure resource is defined in the Bicep file. The `kind` field in the Bicep file defines the type of resource.

As needed, change the `sku` parameter value to the [pricing tier](#) instance you want. The `sku` depends on the resource `kind` that you use. For example, use `TextAnalytics` for the Azure AI Language service. The `TextAnalytics` kind uses `s` instead of `s0` for the `sku` value.

Deploy the template

Azure portal

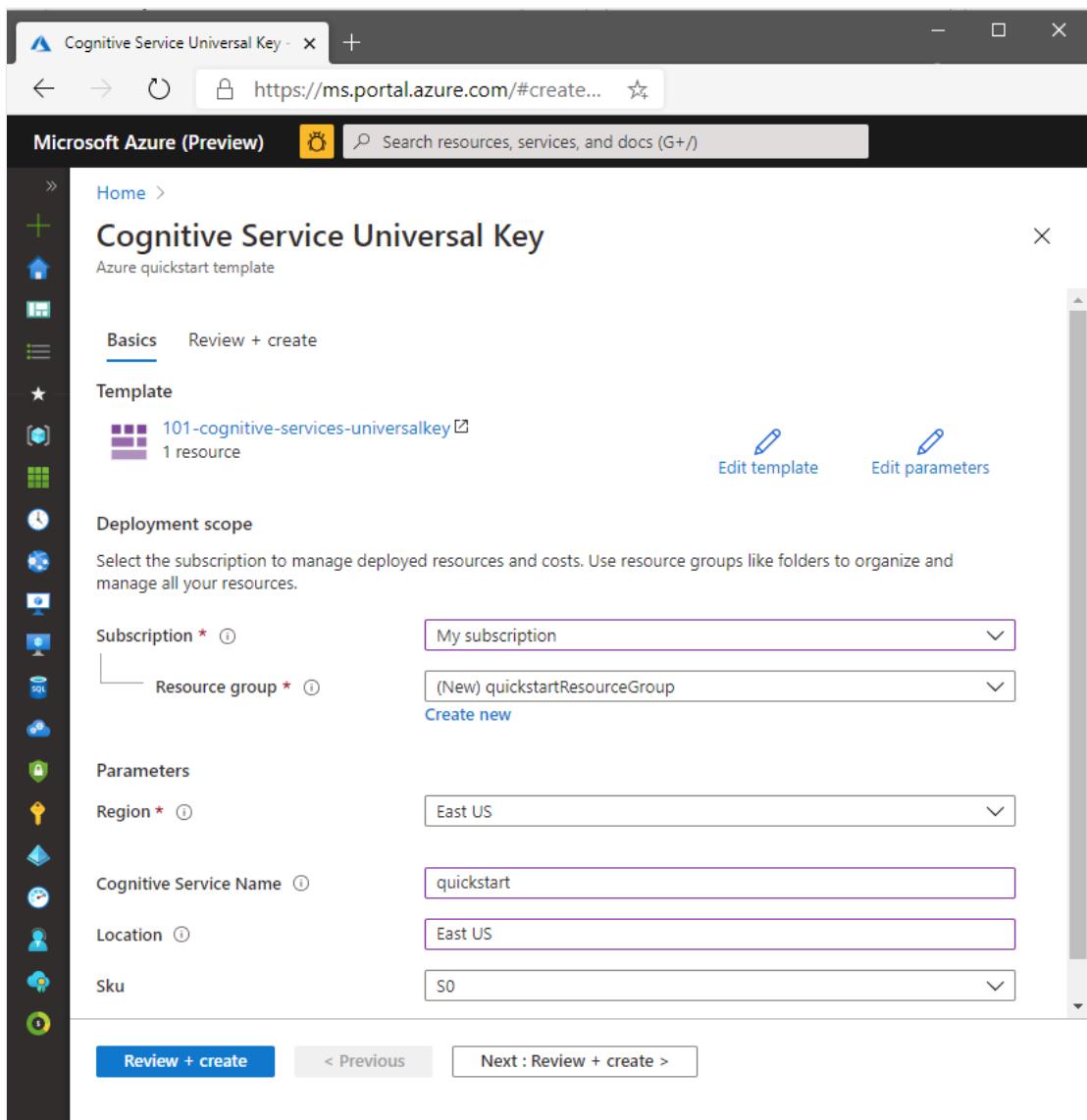
1. Select the Deploy to Azure button.



2. Enter the following values.

[Expand table](#)

Value	Description
Subscription	Select an Azure subscription.
Resource group	Select Create new , enter a unique name for the resource group, and then select OK .
Region	Select a region (for example, East US).
AI service Name	Replace the value with a unique name for your Azure AI services resource. You'll need the name in the next section when you validate the deployment.
Location	Replace with the region that you selected.
Sku	Select the pricing tier for your resource.



3. Select **Review + Create**, and then select **Create**. When deployment is successful, the **Go to resource** button is available.

💡 Tip

If your subscription doesn't allow you to create an Azure AI services resource, you might need to enable the privilege of that [Azure resource provider](#) by using the [Azure portal](#), a [PowerShell command](#) or an [Azure CLI command](#). If you're not the subscription owner, ask the subscription owner or someone with an admin role to complete the registration for you. Or ask for the `/register/action` privileges to be granted to your account.

Review deployed resources

Azure portal

When your deployment finishes, you can select the **Go to resource** button to see your new resource. You can also find the resource group by:

1. Selecting **Resource groups** from the left pane.
2. Selecting the resource group name.

Clean up resources

If you want to clean up and remove an Azure AI services subscription, you can delete the resource or the resource group. Deleting the resource group also deletes any other resources that the group contains.

Azure portal

1. On the left pane, select **Resource groups** to display the list of your resource groups.
2. Locate the resource group that contains the resource to be deleted.
3. Right-click the resource group, select **Delete resource group**, and then confirm.

Related content

- For more information on how to securely work with Azure AI services, see [Authenticate requests to Azure AI services](#).
- For a list of Azure AI services, see [What are Azure AI services?](#).
- For a list of natural languages that Azure AI services support, see [Natural language support in Azure AI services](#).
- To understand how to use Azure AI services on-premises, see [What are Azure AI containers?](#).
- To estimate the cost of using Azure AI services, see [Plan and manage costs for Azure AI Studio](#).

Feedback

Was this page helpful?

 Yes

 No

Quickstart: Create an Azure AI services resource using Terraform

Article • 08/28/2024

This article shows how to use Terraform to create an [Azure AI services multi-service resource](#) using [Terraform](#).

Azure AI services help developers and organizations rapidly create intelligent, cutting-edge, market-ready, and responsible applications with out-of-the-box and prebuilt and customizable APIs and models. Example applications include natural language processing for conversations, search, monitoring, translation, speech, vision, and decision-making.

💡 Tip

Try Azure AI services including Azure OpenAI, Content Safety, Speech, Vision, and more in [Azure AI Studio](#). For more information, see [What is Azure AI Studio?](#).

Most [Azure AI services](#) are available through REST APIs and client library SDKs in popular development languages. For more information, see each service's documentation.

[Terraform](#) enables the definition, preview, and deployment of cloud infrastructure. Using Terraform, you create configuration files using [HCL syntax](#). The HCL syntax allows you to specify the cloud provider - such as Azure - and the elements that make up your cloud infrastructure. After you create your configuration files, you create an *execution plan* that allows you to preview your infrastructure changes before they're deployed. Once you verify the changes, you apply the execution plan to deploy the infrastructure.

In this article, you learn how to:

- ✓ Create a random pet name for the Azure resource group name using [random_pet](#)
- ✓ Create an Azure resource group using [azurerm_resource_group](#)
- ✓ Create a random string using [random_string](#)
- ✓ Create an Azure AI services multi-service resource using [azurerm_cognitive_account](#)

Prerequisites

- [Install and configure Terraform](#)

Implement the Terraform code

ⓘ Note

The sample code for this article is located in the [Azure Terraform GitHub repo](#). You can view the log file containing the [test results from current and previous versions of Terraform](#).

See more [articles and sample code showing how to use Terraform to manage Azure resources](#)

1. Create a directory in which to test and run the sample Terraform code and make it the current directory.
2. Create a file named `main.tf` and insert the following code:

```
Terraform

resource "random_pet" "rg_name" {
    prefix = var.resource_group_name_prefix
}

resource "azurerm_resource_group" "rg" {
    name      = random_pet.rg_name.id
    location  = var.resource_group_location
}

resource "random_string" "azurerm_cognitive_account_name" {
    length   = 13
    lower    = true
    numeric  = false
    special  = false
    upper    = false
}

resource "azurerm_cognitive_account" "cognitive_service" {
    name          =
    "CognitiveService-${random_string.azurerm_cognitive_account_name.result}"
    location      = azurerm_resource_group.rg.location
    resource_group_name = azurerm_resource_group.rg.name
    sku_name      = var.sku
    kind          = "CognitiveServices"
}
```

3. Create a file named `outputs.tf` and insert the following code:

```
Terraform
```

```

output "resource_group_name" {
    value = azurerm_resource_group.rg.name
}

output "azurerm_cognitive_account_name" {
    value = azurerm_cognitive_account.cognitive_service.name
}

```

4. Create a file named `providers.tf` and insert the following code:

Terraform

```

terraform {
    required_version = ">=1.0"
    required_providers {
        azurerm = {
            source  = "hashicorp/azurerm"
            version = "~>3.0"
        }
        random = {
            source  = "hashicorp/random"
            version = "~>3.0"
        }
    }
    provider "azurerm" {
        features {}
    }
}

```

5. Create a file named `variables.tf` and insert the following code:

Terraform

```

variable "resource_group_location" {
    type      = string
    description = "Location for all resources."
    default    = "eastus"
}

variable "resource_group_name_prefix" {
    type      = string
    description = "Prefix of the resource group name that's combined with
a random ID so name is unique in your Azure subscription."
    default    = "rg"
}

variable "sku" {
    type      = string
    description = "The sku name of the Azure Analysis Services server to
create. Choose from: B1, B2, D1, S0, S1, S2, S3, S4, S8, S9. Some skus

```

```
are region specific. See https://docs.microsoft.com/en-us/azure/analysis-services/analysis-services-overview#availability-by-region
    default      = "S0"
}
```

Initialize Terraform

Run [terraform init](#) to initialize the Terraform deployment. This command downloads the Azure provider required to manage your Azure resources.

Console

```
terraform init -upgrade
```

Key points:

- The `-upgrade` parameter upgrades the necessary provider plugins to the newest version that complies with the configuration's version constraints.

Create a Terraform execution plan

Run [terraform plan](#) to create an execution plan.

Console

```
terraform plan -out main.tfplan
```

Key points:

- The `terraform plan` command creates an execution plan, but doesn't execute it. Instead, it determines what actions are necessary to create the configuration specified in your configuration files. This pattern allows you to verify whether the execution plan matches your expectations before making any changes to actual resources.
- The optional `-out` parameter allows you to specify an output file for the plan. Using the `-out` parameter ensures that the plan you reviewed is exactly what is applied.

Apply a Terraform execution plan

Run [terraform apply](#) to apply the execution plan to your cloud infrastructure.

Console

```
terraform apply main.tfplan
```

Key points:

- The example `terraform apply` command assumes you previously ran `terraform plan -out main.tfplan`.
- If you specified a different filename for the `-out` parameter, use that same filename in the call to `terraform apply`.
- If you didn't use the `-out` parameter, call `terraform apply` without any parameters.

Verify the results

Azure CLI

1. Get the Azure resource name in which the Azure AI services multi-service resource was created.

Console

```
resource_group_name=$(terraform output -raw resource_group_name)
```

2. Get the Azure AI services multi-service resource name.

Console

```
azurerm_aiservices_account_name=$(terraform output -raw  
azurerm_aiservices_account_name)
```

3. Run [az cognitiveservices account show](#) to show the Azure AI services account you created in this article.

Azure CLI

```
az cognitiveservices account show --name  
$azurerm_aiservices_account_name \
```

```
--resource-group
```

```
$resource_group_name
```

Clean up resources

When you no longer need the resources created via Terraform, do the following steps:

1. Run [terraform plan ↗](#) and specify the `destroy` flag.

```
Console
```

```
terraform plan -destroy -out main.destroy.tfplan
```

Key points:

- The `terraform plan` command creates an execution plan, but doesn't execute it. Instead, it determines what actions are necessary to create the configuration specified in your configuration files. This pattern allows you to verify whether the execution plan matches your expectations before making any changes to actual resources.
- The optional `-out` parameter allows you to specify an output file for the plan. Using the `-out` parameter ensures that the plan you reviewed is exactly what is applied.

2. Run [terraform apply ↗](#) to apply the execution plan.

```
Console
```

```
terraform apply main.destroy.tfplan
```

Troubleshoot Terraform on Azure

[Troubleshoot common problems when using Terraform on Azure](#)

Related content

- [Learn more about Azure AI services resources](#)

 **Note:** The author created this article with assistance from AI. [Learn more](#)

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

Azure AI services and the AI ecosystem

Article • 08/28/2024

[Azure AI services](#) provides capabilities to solve general problems such as analyzing text for emotional sentiment or analyzing images to recognize objects or faces. You don't need special machine learning or data science knowledge to use these services.

Azure Machine Learning

Azure AI services and Azure Machine Learning both have the end-goal of applying artificial intelligence (AI) to enhance business operations, though how each provides this in the respective offerings is different.

Generally, the audiences are different:

- Azure AI services are for developers without machine-learning experience.
- Azure Machine Learning is tailored for data scientists.

Azure AI services for big data

With Azure AI services for big data you can embed continuously improving, intelligent models directly into Apache Spark™ and SQL computations. These tools liberate developers from low-level networking details, so that they can focus on creating smart, distributed applications. Azure AI services for big data support the following platforms and connectors: Azure Databricks, Azure Synapse, Azure Kubernetes Service, and Data Connectors.

- **Target user(s):** Data scientists and data engineers
- **Benefits:** the Azure AI services for big data let users channel terabytes of data through Azure AI services using Apache Spark™. It's easy to create large-scale intelligent applications with any datastore.
- **UI:** N/A - Code only
- **Subscription(s):** Azure account + Azure AI services resources

To learn more about big data for Azure AI services, see [Azure AI services in Azure Synapse Analytics](#).

Azure Functions and Azure Service Web Jobs

[Azure Functions](#) and [Azure App Service Web Jobs](#) both provide code-first integration services designed for developers and are built on [Azure App Services](#). These products provide serverless infrastructure for writing code. Within that code you can make calls to our services using our client libraries and REST APIs.

- **Target user(s):** Developers and data scientists
- **Benefits:** Serverless compute service that lets you run event-triggered code.
- **UI:** Yes
- **Subscription(s):** Azure account + Azure AI services resource + Azure Functions subscription

Azure Logic Apps

[Azure Logic Apps](#) share the same workflow designer and connectors as Power Automate but provide more advanced control, including integrations with Visual Studio and DevOps. Power Automate makes it easy to integrate with your Azure AI services resources through service-specific connectors that provide a proxy or wrapper around the APIs. These are the same connectors as those available in Power Automate.

- **Target user(s):** Developers, integrators, IT pros, DevOps
- **Benefits:** Designer-first (declarative) development model providing advanced options and integration in a low-code solution
- **UI:** Yes
- **Subscription(s):** Azure account + Azure AI services resource + Logic Apps deployment

Power Automate

Power Automate is a service in the [Power Platform](#) that helps you create automated workflows between apps and services without writing code. We offer several connectors to make it easy to interact with your Azure AI services resource in a Power Automate solution. Power Automate is built on top of Logic Apps.

- **Target user(s):** Business users (analysts) and SharePoint administrators
- **Benefits:** Automate repetitive manual tasks simply by recording mouse clicks, keystrokes and copy paste steps from your desktop!
- **UI tools:** Yes - UI only
- **Subscription(s):** Azure account + Azure AI services resource + Power Automate Subscription + Office 365 Subscription

AI Builder

[AI Builder](#) is a Microsoft Power Platform capability you can use to improve business performance by automating processes and predicting outcomes. AI Builder brings the power of AI to your solutions through a point-and-click experience. Many Azure AI services such as the Language service, and Azure AI Vision have been directly integrated here and you don't need to create your own Azure AI services.

- **Target user(s):** Business users (analysts) and SharePoint administrators
- **Benefits:** A turnkey solution that brings the power of AI through a point-and-click experience. No coding or data science skills required.
- **UI tools:** Yes - UI only
- **Subscription(s):** AI Builder

Next steps

- Learn how you can build generative AI applications in the [Azure AI Studio](#).
- Get answers to frequently asked questions in the [Azure AI FAQ article](#)
- Create your Azure AI services resource in the [Azure portal](#) or with [Azure CLI](#).
- Keep up to date with [service updates](#).

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

Custom subdomain names for Azure AI services

Article • 08/28/2024

Starting in July 2019, Azure AI services use custom subdomain names for each resource created through the [Azure portal](#), [Azure Cloud Shell](#), or [Azure CLI](#). Unlike regional endpoints, which were common for all customers in a specific Azure region, custom subdomain names are unique to the resource. Custom subdomain names are required to enable features like Microsoft Entra ID for authentication.

How does this impact existing resources?

Azure AI services resources created before July 1, 2019 use the regional endpoints for the associated service. These endpoints work with existing and new resources.

If you'd like to migrate an existing resource to use custom subdomain names to enable features like Microsoft Entra ID, follow these instructions:

1. Sign in to the Azure portal and locate the Azure AI services resource that you'd like to add a custom subdomain name to.
2. In the **Overview** blade, locate and select **Generate Custom Domain Name**.
3. This opens a panel with instructions to create a unique custom subdomain for your resource.

Warning

After you've created a custom subdomain name it **cannot** be changed.

Do I need to update my existing resources?

No. The regional endpoint will continue to work for new and existing Azure AI services and the custom subdomain name is optional. Even if a custom subdomain name is added the regional endpoint will continue to work with the resource.

What if an SDK asks me for the region for a resource?

Warning

Speech Services use custom subdomains with **private endpoints only**. In all other cases, use **regional endpoints** with Speech Services and associated SDKs.

Regional endpoints and custom subdomain names are both supported and can be used interchangeably. However, the full endpoint is required.

Region information is available in the **Overview** blade for your resource in the [Azure portal](#). For the full list of regional endpoints, see [Is there a list of regional endpoints?](#)

Are custom subdomain names regional?

Yes. Using a custom subdomain name doesn't change any of the regional aspects of your Azure AI services resource.

What are the requirements for a custom subdomain name?

A custom subdomain name is unique to your resource. The name can only include alphanumeric characters and the `-` character; it must be between 2 and 64 characters in length and cannot end with a `-`.

Can I change a custom domain name?

No. After a custom subdomain name is created and associated with a resource it cannot be changed.

Can I reuse a custom domain name?

Each custom subdomain name is unique, so in order to reuse a custom subdomain name that you've assigned to an Azure AI services resource, you'll need to delete the existing resource. After the resource has been deleted, you can reuse the custom subdomain name.

Is there a list of regional endpoints?

Yes. This is a list of regional endpoints that you can use with Azure AI services resources.

 **Note**

The Translator service and Bing Search APIs use global endpoints.

[\[+\] Expand table](#)

Endpoint type	Region	Endpoint
Public	Global (Translator & Bing)	https://api.cognitive.microsoft.com
	Australia East	https://australiaeast.api.cognitive.microsoft.com
	Brazil South	https://brazilsouth.api.cognitive.microsoft.com
	Canada Central	https://canadacentral.api.cognitive.microsoft.com
	Central US	https://centralus.api.cognitive.microsoft.com
	East Asia	https://eastasia.api.cognitive.microsoft.com
	East US	https://eastus.api.cognitive.microsoft.com
	East US 2	https://eastus2.api.cognitive.microsoft.com
	France Central	https://francecentral.api.cognitive.microsoft.com
	India Central	https://centralindia.api.cognitive.microsoft.com
	Japan East	https://japaneast.api.cognitive.microsoft.com
	Korea Central	https://koreacentral.api.cognitive.microsoft.com
	North Central US	https://northcentralus.api.cognitive.microsoft.com
	North Europe	https://northeurope.api.cognitive.microsoft.com
	South Africa North	https://southafricanorth.api.cognitive.microsoft.com
	South Central US	https://southcentralus.api.cognitive.microsoft.com
	Southeast Asia	https://southeastasia.api.cognitive.microsoft.com
	UK South	https://uksouth.api.cognitive.microsoft.com
	West Central US	https://westcentralus.api.cognitive.microsoft.com
	West Europe	https://westeurope.api.cognitive.microsoft.com

Endpoint type	Region	Endpoint
	West US	https://westus.api.cognitive.microsoft.com
	West US 2	https://westus2.api.cognitive.microsoft.com
US Gov	US Gov Virginia	https://virginia.api.cognitive.microsoft.us
China	China East 2	https://chinaeast2.api.cognitive.azure.cn
	China North	https://chinanorth.api.cognitive.azure.cn

See also

- [What are Azure AI services?](#)
- [Authentication](#)

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

Plan and manage costs for Azure AI Studio

Article • 08/29/2024

ⓘ Important

Some of the features described in this article might only be available in preview. This preview is provided without a service-level agreement, and we don't recommend it for production workloads. Certain features might not be supported or might have constrained capabilities. For more information, see [Supplemental Terms of Use for Microsoft Azure Previews](#).

This article describes how you plan for and manage costs for Azure AI Studio. First, you use the Azure pricing calculator to help plan for Azure AI Studio costs before you add any resources for the service to estimate costs. Next, as you add Azure resources, review the estimated costs.

💡 Tip

Azure AI Studio does not have a specific page in the Azure pricing calculator. Azure AI Studio is composed of several other Azure services, some of which are optional. This article provides information on using the pricing calculator to estimate costs for these services.

You use Azure AI services in Azure AI Studio. Costs for Azure AI services are only a portion of the monthly costs in your Azure bill. You're billed for all Azure services and resources used in your Azure subscription, including the third-party services.

Prerequisites

Cost analysis in Microsoft Cost Management supports most Azure account types, but not all of them. To view the full list of supported account types, see [Understand Cost Management data](#). To view cost data, you need at least read access for an Azure account. For information about assigning access to Microsoft Cost Management data, see [Assign access to data](#).

Estimate costs before using Azure AI services

Use the [Azure pricing calculator](#) to estimate costs before you add Azure AI services.

1. Select a product such as Azure OpenAI in the Azure pricing calculator.

The screenshot shows the Azure Pricing Calculator interface. At the top, there are tabs for 'Products', 'Example scenarios', 'Saved estimates', and 'FAQs'. Below the tabs, a blue header bar says 'Select a product to include it in your estimate.' A search bar contains the text 'openai'. Underneath, a card for 'Azure OpenAI Service' is shown, with a brief description: 'Apply advanced coding and language models to a variety of use cases'. To the right of the card is a magnifying glass icon inside a circle. There are also 'Edit', 'Delete', and 'Copy' icons at the top right of the calculator interface.

2. Enter the number of units you plan to use. For example, enter the number of tokens for prompts and completions.

The screenshot shows the 'Your Estimate' page for the Azure OpenAI Service. At the top, it lists 'Azure OpenAI Service' with a note about Language Models, GPT-3.5-Turbo-4K, 100000 x 100... and 'Upfront: \$0.00' and 'Monthly: \$150.00'. Below this, the 'Azure OpenAI Service' section has dropdowns for 'Region: East US', 'Model type: Language Models', and 'Model: GPT-3.5-Turbo-4K'. The 'Prompt' section shows '100000 x1,000 tokens' at '\$0.0015 Per 1,000 tokens' with a total of '\$150.00'. The 'Completion' section shows '0 x1,000 tokens' at '\$0.0020 Per 1,000 tokens' with a total of '\$0.00'. At the bottom right, there are buttons for 'Upfront cost' and 'Monthly cost'.

3. You can select more than one product to estimate costs for multiple products. For example, select Virtual Machines to add potential costs for compute resources.

The screenshot shows the 'Your Estimate' page with two selected products: 'Azure OpenAI Service' and 'Virtual Machines'. The 'Azure OpenAI Service' row shows 'Monthly: \$150.00' with a red box around it. The 'Virtual Machines' row shows 'Monthly: \$137.24' with a red box around it. Below these, the 'Support' section shows 'Included' with a support icon. The 'Select your program/offer' section includes a note about 'LICENSING PROGRAM: Microsoft Customer Agreement (MCA)' and a link to 'Log in to see your Azure agreement pricing.' There are also checkboxes for 'Show Dev/Test Pricing' and 'Estimated upfront cost' (with '\$0.00') and 'Estimated monthly cost' (with '\$287.24' in a red box).

As you add new resources to your project, return to this calculator and add the same resource here to update your cost estimates.

Costs that typically accrue with Azure AI Studio

When you create resources for a hub, resources for other Azure services are also created. They are:

[+] Expand table

Service pricing page	Description with example use cases
Azure AI services	You pay to use services such as Azure OpenAI, Speech, Content Safety, Vision, Document Intelligence, and Language. Costs vary for each service and for some features within each service. For more information about provisioning of Azure AI services, see Azure AI Studio hubs .
Azure AI Search	An example use case is to store data in a vector search index .
Azure Machine Learning	Compute instances are needed to run Visual Studio Code (Web or Desktop) and prompt flow via Azure AI Studio. When you create a compute instance, the virtual machine (VM) stays on so it's available for your work. Enable idle shutdown to save on cost when the VM is idle for a specified time period. Or set up a schedule to automatically start and stop the compute instance to save cost when you aren't planning to use it.
Azure Virtual Machine	Azure Virtual Machines gives you the flexibility of virtualization for a wide range of computing solutions with support for Linux, Windows Server, SQL Server, Oracle, IBM, SAP, and more.
Azure Container Registry Basic account	Provides storage of private Docker container images, enabling fast, scalable retrieval, and network-close deployment of container workloads on Azure.
Azure Blob Storage	Can be used to store Azure AI Studio project files .
Key Vault	A key vault for storing secrets.
Azure Private Link	Azure Private Link enables you to access Azure PaaS Services (for example, Azure Storage and SQL Database) over a private endpoint in your virtual network.

Costs might accrue before resource deletion

Before you delete a hub in the Azure portal or with Azure CLI, the following sub resources are common costs that accumulate even when you aren't actively working in the workspace. If you're planning on returning to your hub at a later time, these resources might continue to accrue costs:

- Azure AI Search (for the data)
- Virtual machines
- Load Balancer
- Azure Virtual Network
- Bandwidth

Each VM is billed per hour it's running. Cost depends on VM specifications. VMs that are running but not actively working on a dataset are still charged via the load balancer. For each compute instance, one load balancer is billed per day. Every 50 nodes of a compute cluster have one standard load balancer billed. Each load balancer is billed around \$0.33/day. To avoid load balancer costs on stopped compute instances and compute clusters, delete the compute resource.

Compute instances also incur P10 disk costs even in stopped state. This cost is because any user content saved to disk is persisted across the stopped state similar to Azure VMs. We're working on making the OS disk size/ type configurable to better control costs. For Azure Virtual Networks, one virtual network is billed per subscription and per region. Virtual networks can't span regions or subscriptions. Setting up private endpoints in virtual network setups might also incur charges. If your virtual network uses an Azure Firewall, the firewall might also incur charges. Bandwidth usage is charged; the more data transferred, the more you're charged.

💡 Tip

Using a managed virtual network is free. However some features of the managed network rely on Azure Private Link (for private endpoints) and Azure Firewall (for FQDN rules) and will incur charges. For more information, see [Managed virtual network isolation](#).

Costs might accrue after resource deletion

After you delete a hub in the Azure portal or with Azure CLI, the following resources continue to exist. They continue to accrue costs until you delete them.

- Azure Container Registry
- Azure Blob Storage
- Key Vault
- Application Insights (if you enabled it for your hub)

Monitor costs

As you use Azure AI Studio with hubs, you incur costs. Azure resource usage unit costs vary by time intervals (seconds, minutes, hours, and days) or by unit usage (bytes, megabytes, and so on). You can see the incurred costs in [cost analysis](#).

When you use cost analysis, you view hub costs in graphs and tables for different time intervals. Some examples are by day, current and prior month, and year. You also view costs against budgets and forecasted costs. Switching to longer views over time can help you identify spending trends. And you see where overspending might occur. If you create budgets, you can also easily see where they're exceeded.

Monitor Azure AI Studio project costs

You can get to cost analysis from the [Azure portal](#). You can also get to cost analysis from the [Azure AI Studio](#).

Important

Your AI Studio project costs are only a subset of your overall application or solution costs. You need to monitor costs for all Azure resources used in your application or solution. For more information, see [Azure AI Studio hubs](#).

For the examples in this section, assume that all Azure AI Studio resources are in the same resource group. But you can have resources in different resource groups. For example, your Azure AI Search resource might be in a different resource group than your project.

Here's an example of how to monitor costs for a project. The costs are used as an example only. Your costs vary depending on the services that you use and the amount of usage.

1. Sign in to [Azure AI Studio](#).
2. Select your project and select **Settings** from the left navigation section. Select **View cost for resources** from the **Total cost** section. The [Azure portal](#) opens to the

resource group for your project.

The screenshot shows the 'Settings' page in the Azure AI Studio Preview. The left sidebar includes sections for Project overview, Settings (which is selected and highlighted with a red box), Get started, Model catalog, Model benchmarks, Prompt catalog, Project playground (Chat, Completion, Assistants, Images), Tools (Code, Prompt flow, Tracing, Evaluation, Fine-tuning), Components (Data, Indexes, Deployments, Content filters), and a 'Project properties' section. The main content area is divided into several sections: 'Connected resources' (listing four items: contosoai0596127155, contosoai0596127155_aoai, contoso-outdoor-proj/workspaceblobstore, and contoso-outdoor-proj/workspaceartifactstore, each with a 'Type' column), 'Computes' (listing one item: mycompute, showing State as Stopped and VM size as Standard_E4s_v3, with a 'Start' button and a '+ Create compute' link), 'Prompt flow runtimes' (empty), and 'Project members' (listing 'Me (You)' and 'Groups and application permissions'). A 'Total cost' section at the bottom right shows a total of \$212.06, with a 'View cost for resources' button and a magnifying glass icon.

3. Expand the **Resource** column to see the costs for each service that's underlying your **project**. But this view doesn't include costs for all resources that you use in a project.

The screenshot shows the 'Cost analysis' page in the Azure portal. The top navigation bar includes Save, Save as, Delete view, Share, Subscribe, Refresh, Download, Cost by resource, Configure resource group, Try preview, Help, and a search bar. The main area displays a summary of \$212.06 with filters for Scope (rg-contosoairesource), View (CustomView), Date (Dec 2023), and Resource (/subscriptions/a). Below this is a table titled 'Cost by resource' with columns for Resource, Resource type, Location, Resource group name, Tags, and Cost. The first row shows 'contoso-outdoor-proj' as the resource, 'Azure Machine Learning workspace' as the type, 'us east 2' as the location, 'rg-contosoairesource' as the group name, and 'amiresourcetype:provisioner.b...' as the tags. The cost is listed as \$212.06. A detailed breakdown table below shows individual service costs: Virtual Machines (\$204.92), Load Balancer (\$2.93), Storage (\$2.93), Virtual Network (\$1.25), Load Balancer (\$0.03), and Bandwidth (<\$0.01).

4. Select Costs by resource > Resources.

The screenshot shows the Azure Cost analysis interface. At the top, there's a toolbar with various icons like Save, Save as, Delete view, Share, Refresh, Download, and a dropdown menu labeled "Cost by resource". Below the toolbar, the scope is set to "rg-contosoairesource". The main area displays the total actual cost as \$212.06. A dropdown menu for "Cost by resource" is open, showing options for Resources, Services, Reservations, and Cost by resource. The "Cost by resource" option is highlighted with a red box. The table below lists one row of data: "contoso-outdoor-proj" (Azure Machine Learning workspace) located in "us east 2" under the "rg-contosoairesource" resource group. The total cost for this project is \$212.06.

- On the **Cost analysis** page where you're taken to, make sure the scope is set to your resource group.

The screenshot shows the Azure Cost Management page for the resource group "rg-contosoairesource". The scope is set to "Applied AI Docs Team / rg-contosoairesource (change)". The main area displays the total cost as \$222.97. A table lists five resources: "contoso-outdoor-proj" (Machine learning), "contoso_ai_resource" (Machine learning), "crcntosoairesource" (Container registry), "stcontosoai" (Storage account), and "kv-contoso" (Key vault). The total cost for "contoso-outdoor-proj" is \$212.06. A red box highlights the "contoso-outdoor-proj" row. A circular magnifying glass icon is visible on the right side of the table.

In this example:

- The resource group name is **rg-contosoairesource**.
- The total cost for all resources and services in the resource group is **\$222.97**. In this example, \$222.97 is the total cost for your application or solution that you're building with Azure AI Studio. Again, this example assumes that all Azure AI Studio resources are in the same resource group. But you can have resources in different resource groups.
- The project name is **contoso-outdoor-proj**.
- The costs that are limited to resources and services in the **project** total **\$212.06**.

- Expand **contoso-outdoor-proj** to see the costs for services underlying the **project** resource.

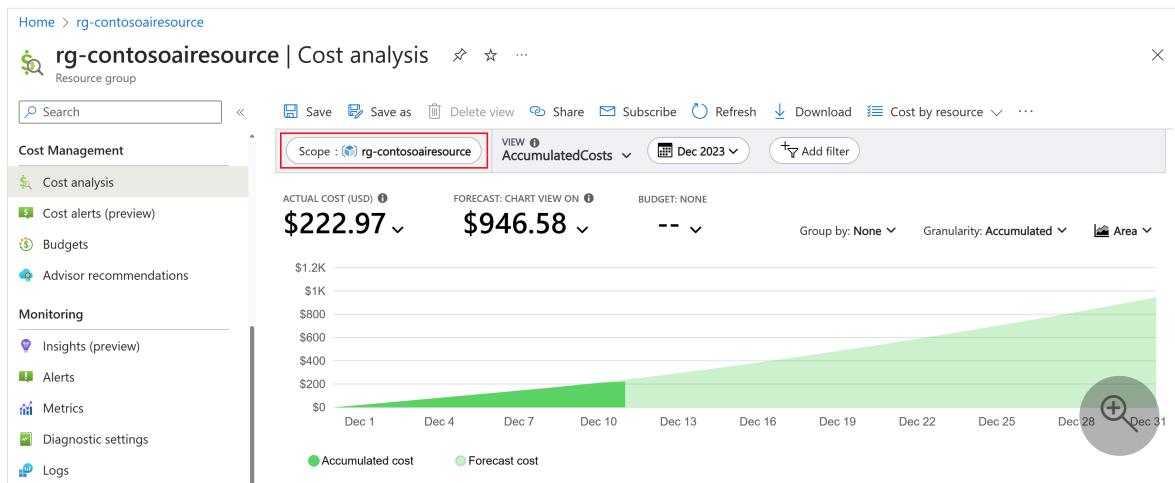
The screenshot shows the Azure Cost Management interface for a resource group named 'rg-contosoairesource'. The main view displays a table of resources with columns for Name, Type, Resource group, Location, and Total cost. A specific row for 'contoso-outdoor-proj' is expanded, showing detailed service-level costs. The 'Storage' service has the highest cost at \$204.92.

Name	Type	Resource group	Location	Total
contoso-outdoor-proj	Machine learning	rg-contosoairesource	us east 2	\$212.06
Virtual Machines	Virtual Machines Dv2 Series	Virtual Machines Dv2 Series...	D3 v2/DS3 v2	\$204.92
Load Balancer	Load Balancer	Load Balancer	Standard Included LB Rules ...	\$2.93
Storage	Premium SSD Managed Disks	Premium SSD Managed Dis...	P6 LRS Disk	\$2.08
Virtual Network	IP Addresses	IP Addresses - Standard IPv4	Standard IPv4 Static Public IP	\$1.25
Load Balancer	Load Balancer	Load Balancer	Standard Data Processed	\$0.03

7. Expand **contoso_ai_resource** to see the costs for services underlying the **hub** resource. You can also apply a filter to focus on other costs in your resource group.

You can also view resource group costs directly from the Azure portal. To do so:

1. Sign in to [Azure portal](#).
2. Select **Resource groups**.
3. Find and select the resource group that contains your Azure AI Studio resources.
4. From the left navigation menu, select **Cost analysis**.



For more information, see the [Azure pricing calculator](#).

Monitor costs for models offered through the Azure Marketplace

Models deployed as a service using pay-as-you-go are offered through the Azure Marketplace. The model publishers might apply different costs depending on the offering. Each project in Azure AI Studio has its own subscription with the offering,

which allows you to monitor the costs and the consumption happening on that project. Use Microsoft Cost Management² to monitor the costs:

1. Sign in to [Azure portal](#).
2. On the left navigation area, select **Cost Management + Billing** and then, on the same menu, select **Cost Management**.
3. On the left navigation area, under the section **Cost Management**, select now **Cost Analysis**.
4. Select a view such as **Resources**. The cost associated with each resource is displayed.

The screenshot shows the Microsoft Cost Management interface for an Azure subscription. The left sidebar has sections for Cost Management (with 'Cost analysis' selected), Billing, Products + services, and Settings. The main content area shows a 'Recent' section with a 'Resources' card (also highlighted with a red box). Below it are 'Recommended' cards for 'Accumulated costs', 'Resource groups', 'Daily costs', and 'Visualize your costs'. A search bar and a 'Browse all views' link are at the bottom of the main content area.

5. On the Type column, select the filter icon to filter all the resources of type **microsoft.saas/resources**. This type corresponds to resources created from offers from the Azure Marketplace. For convenience, you can filter by resource types containing the string **SaaS**.

The screenshot shows a filtered list of resources. A modal dialog is open over the table, showing a dropdown menu with 'Contains' selected and 'SaaS' typed in. The table below lists resources with columns for Name, Type, Subscription, Tags, and Total cost. The 'Type' column shows values like 'SaaS' and 'microsoft.saas/resources'. The 'Subscription' column shows 'Azure subscription 1' repeated. The 'Total' column shows various dollar amounts. A magnifying glass icon is in the bottom right corner of the table area.

Name	Type	Subscription	Tags	Total
> Ilama-2... ...	SaaS	Azure subscription 1	--	\$217.59
> Ilama-2... ...	SaaS	Azure subscription 1	--	\$151.38
> Ilama-2... ...	SaaS	Azure subscription 1	--	\$116.99
> Ilama-2... ...	SaaS	Azure subscription 1	--	\$88.00
> Ilama-2... ...	SaaS	Azure subscription 1	--	\$87.97
> Ilama-2... ...	SaaS	Azure subscription 1	--	\$44.81
> Ilama-2... ...	SaaS	Azure subscription 1	--	\$0.02
> Ilama-2... ...	SaaS	Azure subscription 1	--	<\$0.01
< Ilama-2... ...	SaaS	Azure subscription 1	--	<\$0.01

6. One resource is displayed for each model offer per project. Naming of those resources is **[Model offer name]-[GUID]**.

7. Select to expand the resource details to get access to each of the costs meters associated with the resource.

- **Tier** represents the offering.
- **Product** is the specific product inside the offering.

Some model providers might use the same name for both.

Scope: 📈 Azure subscription 1 (change)						
Resources		+ Add new				
Back		Customize	Download	...		
Filter rows		< Dec 2023 >				
Name	Type	Resource group	Location	Subscription	Tags	Total ↓
llama-2-70b...	SaaS	paygo-finetuned-model	global	Azure subscription 1	--	\$217.59
Service	Tier	Product	Meter			
SaaS	...	Meta Llama-2-70B	Meta Llama-2-70B	Meta Llama-2-70B - Meta Llama-2-70B - paygo-finetuned-model-inference-ho...		
SaaS	...	Meta Llama-2-70B	Meta Llama-2-70B	Meta Llama-2-70B - Meta Llama-2-70B - paygo-finetuned-model-inference-ou...		
SaaS	...	Meta Llama-2-70B	Meta Llama-2-70B	Meta Llama-2-70B - Meta Llama-2-70B - paygo-finetuned-model-inference-in...		
llama-2-7b...	SaaS	paygo-finetuned-model	global	Azure subscription 1	--	\$151.38
Service	Tier	Product	Meter			
SaaS	...	Meta Llama 2 7B	Meta Llama-2-7B	Meta Llama 2 7B - Meta Llama-2-7B - paygo-finetun...		
SaaS	...	Meta Llama 2 7B	Meta Llama-2-7B	Meta Llama 2 7B - Meta Llama-2-7B - paygo-finetun...		
SaaS	...	Meta Llama 2 7B	Meta Llama-2-7B	Meta Llama 2 7B - Meta Llama-2-7B - paygo-finetun...		

💡 Tip

Remember that one resource is created per each project, per each plan your project subscribes to.

8. When you expand the details, costs are reported per each of the meters associated with the offering. Each meter might track different sources of costs like inferencing, or fine tuning. The following meters are displayed (when some cost is associated with them):

[+] Expand table

Meter	Group	Description
paygo-inference-input-tokens	Base model	Costs associated with the tokens used as input for inference of a base model.
paygo-inference-output-tokens	Base model	Costs associated with the tokens generated as output for the inference of base model.
paygo-finetuned-model-inference-hosting	Fine-tuned model	Costs associated with the hosting of an inference endpoint for a fine-tuned model. This value isn't the cost of hosting the model, but the cost of having an endpoint serving it.

Meter	Group	Description
paygo-finetuned-model-inference-input-tokens	Fine-tuned model	Costs associated with the tokens used as input for inference of a fine tuned model.
paygo-finetuned-model-inference-output-tokens	Fine-tuned model	Costs associated with the tokens generated as output for the inference of a fine tuned model.

Create budgets

You can create [budgets](#) to manage costs and create [alerts](#) that automatically notify stakeholders of spending anomalies and overspending risks. Alerts are based on spending compared to budget and cost thresholds. Budgets and alerts are created for Azure subscriptions and resource groups, so they're useful as part of an overall cost monitoring strategy.

Budgets can be created with filters for specific resources or services in Azure if you want more granularity present in your monitoring. Filters help ensure that you don't accidentally create new resources that cost you more money. For more about the filter options when you create a budget, see [Group and filter options](#).

Export cost data

You can also [export your cost data](#) to a storage account. Exporting data is helpful when you or others need to do more data analysis for costs. For example, finance teams can analyze the data using Excel or Power BI. You can export your costs on a daily, weekly, or monthly schedule and set a custom date range. Exporting cost data is the recommended way to retrieve cost datasets.

Understand the full billing model for Azure AI services

Azure AI services run on Azure infrastructure that accrues costs along with Azure AI when you deploy the new resource. It's important to understand that extra infrastructure might accrue cost. You need to manage that cost when you make changes to deployed resources.

When you create or use Azure AI services resources, you might get charged based on the services that you use. There are two billing models available for Azure AI services:

- Pay-as-you-go: Pay-as-you-go pricing, you're billed according to the Azure AI services offering that you use, based on its billing information.
- Commitment tiers: With commitment tier pricing, you commit to using several service features for a fixed fee, enabling you to have a predictable total cost based on the needs of your workload. You're billed according to the plan you choose. See [Quickstart: purchase commitment tier pricing](#) for information on available services, how to sign up, and considerations when purchasing a plan.

Note

If you use the resource above the quota provided by the commitment plan, you will be charged for the additional usage as per the overage amount mentioned in the Azure portal when you purchase a commitment plan.

You can pay for Azure AI services charges with your Azure Prepayment (previously called monetary commitment) credit. However, you can't use Azure Prepayment credit to pay for charges for third-party products and services including ones from the Azure Marketplace.

For more information, see the [Azure pricing calculator](#).

Next steps

- Learn [how to optimize your cloud investment with Microsoft Cost Management](#).
- Learn more about managing costs with [cost analysis](#).
- Learn about how to [prevent unexpected costs](#).
- Take the [Cost Management](#) guided learning course.

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

Autoscale AI services limits

Article • 08/28/2024

This article provides guidance for how customers can access higher rate limits on their Azure AI services resources.

Overview

Each Azure AI services resource has a pre-configured static call rate (transactions per second) which limits the number of concurrent calls that customers can make to the backend service in a given time frame. The autoscale feature will automatically increase/decrease a customer's resource's rate limits based on near-real-time resource usage metrics and backend service capacity metrics.

Get started with the autoscale feature

This feature is disabled by default for every new resource. Follow these instructions to enable it.

Azure portal

Go to your resource's page in the Azure portal, and select the **Overview** tab on the left pane. Under the **Essentials** section, find the **Autoscale** line and select the link to view the **Autoscale Settings** pane and enable the feature.

The screenshot shows the Azure portal interface with the 'Face' resource selected. The left sidebar shows navigation options like Overview, Activity log, Access control (IAM), Tags, and more. The main pane displays the 'Overview' tab under the 'Essentials' section. It shows details such as Resource group (move), Status: Active, Location: centralus, Subscription (move), Subscription ID, and Tags. A red box highlights the 'Autoscale' link under the 'Tags' section. To the right, a modal window titled 'Autoscale Settings' is open. It contains a descriptive text about autoscale and two buttons: 'Enabled' (highlighted with a red box) and 'Disabled'. Below the buttons is a 'Save' button. Another red box highlights the 'Disabled' status of the Autoscale setting. The bottom of the modal shows a note: 'Autoscale automatically increases/decreases your resource rate limit based on near real-time usage and backend service capacity. Learn more about Autoscale.'

Frequently asked questions

Does enabling the autoscale feature mean my resource will never be throttled again?

No, you may still get 429 errors for rate limit excess. If your application triggers a spike, and your resource reports a 429 response, autoscale checks the available capacity projection section to see whether the current capacity can accommodate a rate limit increase and respond within five minutes.

If the available capacity is enough for an increase, autoscale gradually increases the rate limit cap of your resource. If you continue to call your resource at a high rate that results in more 429 throttling, your TPS rate will continue to increase over time. If this action continues for one hour or more, you should reach the maximum rate (up to 1000 TPS) currently available at that time for that resource.

If the available capacity isn't enough for an increase, the autoscale feature waits five minutes and checks again.

What if I need a higher default rate limit?

By default, Azure AI services resources have a default rate limit of 10 TPS. If you need a higher default TPS, submit a ticket by following the [New Support Request](#) link on your resource's page in the Azure portal. Remember to include a business justification in the request.

Will this feature increase my Azure spend?

Azure AI services pricing hasn't changed and can be accessed [here](#). We'll only bill for successful calls made to Azure AI services APIs. However, increased call rate limits mean more transactions are completed, and you may receive a higher bill.

Be aware of potential errors and their consequences. If a bug in your client application causes it to call the service hundreds of times per second, that would likely lead to a much higher bill, whereas the cost would be much more limited under a fixed rate limit. Errors of this kind are your responsibility. We highly recommend that you perform development and client update tests against a resource with a fixed rate limit prior to using the autoscale feature.

Can I disable this feature if I'd rather limit the rate than have unpredictable spending?

Yes, you can disable the autoscale feature through Azure portal or CLI and return to your default call rate limit setting. If your resource was previously approved for a higher default TPS, it goes back to that rate. It can take up to five minutes for the changes to go into effect.

Which services support the autoscale feature?

Autoscale feature is available for the following services:

- [Azure AI Vision](#)
- [Language](#) (only available for sentiment analysis, key phrase extraction, named entity recognition, and text analytics for health)
- [Anomaly Detector](#)
- [Content Moderator](#)
- [Custom Vision \(Prediction\)](#)
- [Immersive Reader](#)
- [LUIS](#)
- [Metrics Advisor](#)
- [Personalizer](#)
- [QnAMaker](#)
- [Document Intelligence](#)

Can I test this feature using a free subscription?

No, the autoscale feature isn't available to free tier subscriptions.

Next steps

- [Plan and Manage costs for Azure AI services.](#)
- [Optimize your cloud investment with Azure Cost Management.](#)
- Learn about how to [prevent unexpected costs](#).
- Take the [Cost Management](#) guided learning course.

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

Purchase commitment tier pricing

Article • 08/28/2024

Azure AI offers commitment tier pricing, allowing discounted rates compared to the pay-as-you-go pricing model. With commitment tier pricing, you can commit to using the following Azure AI services features for a fixed fee, enabling you to have a predictable total cost based on the needs of your workload:

- Speech to text (Standard)
- Text to speech (Neural)
- Text Translation (Standard)
- Language Understanding standard (Text Requests)
- Azure AI Language
 - Sentiment Analysis
 - Key Phrase Extraction
 - Language Detection
 - Named Entity Recognition (NER)
- Azure AI Vision - OCR
- Document Intelligence – Custom/Invoice

For more information, see [Azure AI services pricing](#).

Create a new resource

1. Sign in to the [Azure portal](#) and select **Create a new resource** for one of the applicable Azure AI services or Azure AI services listed.
2. Enter the applicable information to create your resource. Be sure to select the standard pricing tier.

Note

If you intend to purchase a commitment tier for disconnected container usage, you will need to request separate access and select the **Commitment tier disconnected containers** pricing tier. For more information, see [disconnected containers](#).

Microsoft Azure

Search resources, services, and docs (G+)

Create a resource

Dashboard

All services

FAVORITES

All resources

Tags

Resource groups

SQL databases

Dedicated SQL pools (formerly SQL DW)

Subscriptions

Azure Cosmos DB

Azure Active Directory

Recent

Automation Accounts

Event Hubs

Cognitive Search

Cognitive Services

API Management services

Management groups

Key vaults

Templates

Cost Management + Billing

Dashboard > Create a resource > Form Recognizer >

Create Form Recognizer

Basics Network Identity Tags Review + create

Accelerate your business processes by automating information extraction. Form Recognizer applies advanced machine learning to accurately extract text, key/value pairs, and tables from documents. With just a few samples, Form Recognizer tailors its understanding to your documents, both on-premises and in the cloud. Turn forms into usable data at a fraction of the time and cost, so you can focus more time acting on the information rather than compiling it. [Learn more.](#)

Project Details

Subscription * ⓘ my-subscription

Resource group * ⓘ my-test-group [Create new](#)

Instance Details

Region ⓘ West Europe

Name * ⓘ my-resource

The free tier (F0) for this resource type is already being used by your subscription, therefore it will not appear in the dropdown below.

Pricing tier * ⓘ Standard S0 (1 Call per minute for training API)

[View full pricing details](#)

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

- Once your resource is created, you can change your pricing from pay-as-you-go, to a commitment plan.

Purchase a commitment plan by updating your Azure resource

- Sign in to the [Azure portal](#) with your Azure subscription.
- In your Azure resource for one of the applicable features listed, select **Commitment tier pricing**.
- Select **Change** to view the available commitments for hosted API and container usage. Choose a commitment plan for one or more of the following offerings:
 - Web:** web-based APIs, where you send data to Azure for processing.
 - Connected container:** Docker containers that enable you to [deploy Azure AI services on premises](#), and maintain an internet connection for billing and metering.

Choose a monthly commitment plan based on the needs of your workload. You can customize the 'Standard' pricing tier by selecting one of several commitment tier options. Each commitment tier offers a discounted rate compared to the pay-as-you-go pricing model. Commitment plans are charged monthly, except the first month upon purchase which is pro-rated(cost and quota) based on the number of days remaining in that month. For the subsequent months, the charge is incurred on the first day of the month. Any overage will be charged as and when incurred after you have consumed the allowed quota for your chosen tier. [Learn more about commitment tier pricing](#)

Commitment plans for: Web

Custom, Invoice ([Change](#))

Tier : No commitment (Pay as you go)

Commitment plans for: Connected container

Custom, Invoice ([Change](#))

Tier : No commitment (Pay as you go)

4. In the window that appears, select both a **Tier** and **Auto-renewal** option.

- **Commitment tier** - The commitment tier for the feature. The commitment tier is enabled immediately when you select **Purchase** and you're charged the commitment amount on a pro-rated basis.
- **Auto-renewal** - Choose how you want to renew, change, or cancel the current commitment plan starting with the next billing cycle. If you decide to autorenew, the **Auto-renewal date** is the date (in your local timezone) when you'll be charged for the next billing cycle. This date coincides with the start of the calendar month.

⊗ Caution

Once you select **Purchase** you will be charged for the tier you select. Once purchased, the commitment plan is non-refundable.

Commitment plans are charged monthly, except the first month upon purchase which is pro-rated (cost and quota) based on the number of days remaining in that month. For the subsequent months, the charge is incurred on the first day of the month.



Purchase commitment plan



Purchase a commitment plan here. Commitment plans are charged monthly, except the first month upon purchase which is pro-rated (cost and quota) based on the number of days remaining in that month. For the subsequent months, the charge is incurred on the first day of the month. You can also configure or cancel your commitment settings for the next month. If you want to cancel your commitment plan after the current month, choose "Do not auto-renew" in the 'Manage your plan' dropdown below. [Learn more about commitment plan management.](#)

Hosting model ⓘ

Web

Capabilities ⓘ

Custom, Invoice

Commitment tier * ⓘ



Manage your plan ⓘ



Auto-renewal date ⓘ

2021-11-30

Purchase

Cancel



Overage pricing

If you use the resource above the quota provided, you're charged for the additional usage as per the overage amount mentioned in the commitment tier.

Purchase a different commitment plan

The commitment plans have a calendar month commitment period. You can purchase a commitment plan at any time from the default pay-as-you-go pricing model. When you purchase a plan, you're charged a pro-rated price for the remaining month. During the commitment period, you can't change the commitment plan for the current month. However, you can choose a different commitment plan for the next calendar month. The billing for the next month would happen on the first day of the next month.

If you need a larger commitment plan than any of the ones offered, contact

csgate@microsoft.com.

End a commitment plan

If you decide that you don't want to continue purchasing a commitment plan, you can set your resource's autorenewal to **Do not auto-renew**. Your commitment plan expires on the displayed commitment end date. After this date, you won't be charged for the commitment plan. You're able to continue using the Azure resource to make API calls, charged at pay-as-you-go pricing. You have until midnight (UTC) on the last day of each month to end a commitment plan, and not be charged for the following month.

Purchase a commitment tier pricing plan for disconnected containers

Commitment plans for disconnected containers have a calendar year commitment period. These are different plans than web and connected container commitment plans. When you purchase a commitment plan, you'll be charged the full price immediately. During the commitment period, you can't change your commitment plan, however you can purchase additional unit(s) at a pro-rated price for the remaining days in the year. You have until midnight (UTC) on the last day of your commitment, to end a commitment plan.

You can choose a different commitment plan in the **Commitment Tier pricing** settings of your resource.

Overage pricing for disconnected containers

To use a disconnected container beyond the quota initially purchased with your disconnected container commitment plan, you can purchase additional quota by updating your commitment plan at any time.

To purchase additional quota, go to your resource in Azure portal and adjust the "unit count" of your disconnected container commitment plan using the slider. This will add additional monthly quota and you will be charged a pro-rated price based on the remaining days left in the current billing cycle.

See also

- [Azure AI services pricing ↗](#).

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

Enable diagnostic logging for Azure AI services

Article • 08/28/2024

This guide provides step-by-step instructions to enable diagnostic logging for an Azure AI service. These logs provide rich, frequent data about the operation of a resource that are used for issue identification and debugging. Before you continue, you must have an Azure account with a subscription to at least one Azure AI service, such as [Speech Services](#).

Prerequisites

To enable diagnostic logging, you'll need somewhere to store your log data. This tutorial uses Azure Storage and Log Analytics.

- [Azure storage](#) - Retains diagnostic logs for policy audit, static analysis, or backup. The storage account does not have to be in the same subscription as the resource emitting logs as long as the user who configures the setting has appropriate Azure RBAC access to both subscriptions.
- [Log Analytics](#) - A flexible log search and analytics tool that allows for analysis of raw logs generated by an Azure resource.

ⓘ Note

- Additional configuration options are available. To learn more, see [Collect and consume log data from your Azure resources](#).
- "Trace" in diagnostic logging is only available for [Custom question answering](#).

Enable diagnostic log collection

Let's start by enabling diagnostic logging using the Azure portal.

ⓘ Note

To enable this feature using PowerShell or the Azure CLI, use the instructions provided in [Collect and consume log data from your Azure resources](#).

1. Navigate to the Azure portal. Then locate and select an Azure AI services resource. For example, your subscription to Speech Services.
2. Next, from the left-hand navigation menu, locate **Monitoring** and select **Diagnostic settings**. This screen contains all previously created diagnostic settings for this resource.
3. If there is a previously created resource that you'd like to use, you can select it now. Otherwise, select **+ Add diagnostic setting**.
4. Enter a name for the setting. Then select **Archive to a storage account** and **Send to log Analytics**.
5. When prompted to configure, select the storage account and OMS workspace that you'd like to use to store your diagnostic logs. **Note:** If you don't have a storage account or OMS workspace, follow the prompts to create one.
6. Select **Audit**, **RequestResponse**, and **AllMetrics**. Then set the retention period for your diagnostic log data. If a retention policy is set to zero, events for that log category are stored indefinitely.
7. Select **Save**.

It can take up to two hours before logging data is available to query and analyze. So don't worry if you don't see anything right away.

View and export diagnostic data from Azure Storage

Azure Storage is a robust object storage solution that is optimized for storing large amounts of unstructured data. In this section, you'll learn to query your storage account for total transactions over a 30-day timeframe and export the data to excel.

1. From the Azure portal, locate the Azure Storage resource that you created in the last section.
2. From the left-hand navigation menu, locate **Monitoring** and select **Metrics**.
3. Use the available drop-downs to configure your query. For this example, let's set the time range to **Last 30 days** and the metric to **Transaction**.
4. When the query is complete, you'll see a visualization of transaction over the last 30 days. To export this data, use the **Export to Excel** button located at the top of the page.

Learn more about what you can do with diagnostic data in [Azure Storage](#).

View logs in Log Analytics

Follow these instructions to explore log analytics data for your resource.

1. From the Azure portal, locate and select **Log Analytics** from the left-hand navigation menu.
2. Locate and select the resource you created when enabling diagnostics.
3. Under **General**, locate and select **Logs**. From this page, you can run queries against your logs.

Sample queries

Here are a few basic Kusto queries you can use to explore your log data.

Run this query for all diagnostic logs from Azure AI services for a specified time period:

```
Kusto

AzureDiagnostics
| where ResourceProvider == "MICROSOFT.COGNITIVESERVICES"
```

Run this query to see the 10 most recent logs:

```
Kusto

AzureDiagnostics
| where ResourceProvider == "MICROSOFT.COGNITIVESERVICES"
| take 10
```

Run this query to group operations by **Resource**:

```
Kusto

AzureDiagnostics
| where ResourceProvider == "MICROSOFT.COGNITIVESERVICES" |
summarize count() by Resource
```

Run this query to find the average time it takes to perform an operation:

```
Kusto

AzureDiagnostics
| where ResourceProvider == "MICROSOFT.COGNITIVESERVICES"
| summarize avg(DurationMs)
by OperationName
```

Run this query to view the volume of operations over time split by **OperationName** with counts binned for every 10s.

```
AzureDiagnostics
| where ResourceProvider == "MICROSOFT.COGNITIVESERVICES"
| summarize count()
by bin(TimeGenerated, 10s), OperationName
| render areachart kind=unstacked
```

Next steps

- To understand how to enable logging, as well as the metrics and log categories that are supported by the various Azure services, read the [Overview of metrics](#) in Microsoft Azure and the [Overview of Azure Diagnostic Logs](#).
- Read these articles to learn about event hubs:
 - [What is Azure Event Hubs?](#)
 - [Get started with Event Hubs](#)
- Read [Understand log searches in Azure Monitor logs](#).

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

Recover or purge deleted Azure AI services resources

Article • 09/12/2024

This article provides instructions on how to recover or purge an Azure AI services resource that is already deleted.

Once you delete a resource, you won't be able to create another one with the same name for 48 hours. To create a resource with the same name, you need to purge the deleted resource.

ⓘ Note

- The instructions in this article are applicable to both a multi-service resource and a single-service resource. A multi-service resource enables access to multiple Azure AI services using a single key and endpoint. On the other hand, a single-service resource enables access to just that specific Azure AI service for which the resource was created.
- Charges for provisioned deployments on a deleted resource will continue until the resource is purged. To prevent this, delete a resource's deployment before deleting the resource.

Recover a deleted resource

The following prerequisites must be met before you can recover a deleted resource:

- The resource to be recovered must have been deleted within the past 48 hours.
- The resource to be recovered must not have been purged already. A purged resource can't be recovered.
- Before you attempt to recover a deleted resource, make sure that the resource group for that account exists. If the resource group was deleted, you must recreate it. Recovering a resource group isn't possible. For more information, see [Manage resource groups](#).
- If the deleted resource used customer-managed keys with Azure Key Vault and the key vault have also been deleted, then you must restore the key vault before you restore the Azure AI services resource. For more information, see [Azure Key Vault recovery management](#).

- If the deleted resource used a customer-managed storage and storage account has also been deleted, you must restore the storage account before you restore the Azure AI services resource. For instructions, see [Recover a deleted storage account](#).

To recover a deleted Azure AI services resource, use the following commands. Where applicable, replace:

- {subscriptionID} with your Azure subscription ID
- {resourceGroup} with your resource group
- {resourceName} with your resource name
- {location} with the location of your resource



If you need to recover a deleted resource, navigate to the hub of the Azure AI services API type and select "Manage deleted resources" from the menu. For example, if you would like to recover an "Anomaly detector" resource, search for "Anomaly detector" in the search bar and select the service. Then select **Manage deleted resources**.

Select the subscription in the dropdown list to locate the deleted resource you would like to recover. Select one or more of the deleted resources and select **Recover**.

Name	Deletion date	Scheduled purge date
MyAnomalyDetectorResou...	7/7/2023, 10:31:53 AM	7/7/2023, 10:31:53 AM

Note

It can take a couple of minutes for your deleted resource(s) to recover and show up in the list of the resources. Select the **Refresh** button in the menu to

update the list of resources.

Purge a deleted resource

Your subscription must have

`Microsoft.CognitiveServices/locations/resourceGroups/deletedAccounts/delete`

permissions to purge resources, such as [Cognitive Services Contributor](#) or [Contributor](#).

When using `Contributor` to purge a resource the role must be assigned at the subscription level. If the role assignment is only present at the resource or resource group level, you can't access the purge functionality.

To purge a deleted Azure AI services resource, use the following commands. Where applicable, replace:

- `{subscriptionID}` with your Azure subscription ID
- `{resourceGroup}` with your resource group
- `{resourceName}` with your resource name
- `{location}` with the location of your resource

ⓘ Note

Once a resource is purged, it is permanently deleted and cannot be restored. You will lose all data and keys associated with the resource.

Azure portal

If you need to purge a deleted resource, the steps are similar to recovering a deleted resource.

1. Navigate to the hub of the Azure AI services API type of your deleted resource. For example, if you would like to purge an "Anomaly detector" resource, search for "Anomaly detector" in the search bar and select the service. Then select **Manage deleted resources** from the menu.
2. Select the subscription in the dropdown list to locate the deleted resource you would like to purge.
3. Select one or more deleted resources and select **Purge**. Purging permanently deletes an Azure AI services resource.

The screenshot shows the Microsoft Azure portal interface. On the left, there's a sidebar with various service categories like Home, Dashboard, All services, Resource groups, App Services, etc. The main area is titled 'Cognitive Services | Anomaly detector'. A sub-menu on the left lists services under 'Cognitive Services': Azure OpenAI, Speech, Language, Vision, and Decision. Under 'Decision', 'Anomaly detector' is selected. To the right, a modal window titled 'Manage deleted resources' is open. It displays a table with one record: 'Name' (MyAnomalyDetectorResource), 'Deletion date' (7/5/2023, 10:31:53 AM), and 'Scheduled purge date' (7/7/2023, 10:31:53 AM). There are checkboxes next to each column header. At the bottom of the modal are buttons for 'Recover', 'Purge', 'Close', and 'Refresh'.

Related content

- [Create an Azure AI services resource](#)
- [Create an Azure AI services resource using an ARM template](#)

Feedback

Was this page helpful?

Yes

No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

What are Azure AI containers?

Article • 09/25/2024

Azure AI services provide several [Docker containers](#) that let you use the same APIs that are available in Azure, on-premises. Using these containers gives you the flexibility to bring Azure AI services closer to your data for compliance, security or other operational reasons. Container support is currently available for a subset of Azure AI services.

<https://www.youtube-nocookie.com/embed/hdfbn4Q8jbo>

Containerization is an approach to software distribution in which an application or service, including its dependencies & configuration, is packaged together as a container image. With little or no modification, a container image can be deployed on a container host. Containers are isolated from each other and the underlying operating system, with a smaller footprint than a virtual machine. Containers can be instantiated from container images for short-term tasks, and removed when no longer needed.

Features and benefits

- **Immutable infrastructure:** Enable DevOps teams to leverage a consistent and reliable set of known system parameters, while being able to adapt to change. Containers provide the flexibility to pivot within a predictable ecosystem and avoid configuration drift.
- **Control over data:** Choose where your data gets processed by Azure AI services. This can be essential if you can't send data to the cloud but need access to Azure AI services APIs. Support consistency in hybrid environments – across data, management, identity, and security.
- **Control over model updates:** Flexibility in versioning and updating of models deployed in their solutions.
- **Portable architecture:** Enables the creation of a portable application architecture that can be deployed on Azure, on-premises and the edge. Containers can be deployed directly to [Azure Kubernetes Service](#), [Azure Container Instances](#), or to a [Kubernetes](#) cluster deployed to [Azure Stack](#). For more information, see [Deploy Kubernetes to Azure Stack](#).
- **High throughput / low latency:** Provide customers the ability to scale for high throughput and low latency requirements by enabling Azure AI services to run physically close to their application logic and data. Containers don't cap transactions per second (TPS) and can be made to scale both up and out to handle demand if you provide the necessary hardware resources.

- **Scalability:** With the ever growing popularity of containerization and container orchestration software, such as Kubernetes; scalability is at the forefront of technological advancements. Building on a scalable cluster foundation, application development caters to high availability.

Containers in Azure AI services

Azure AI containers provide the following set of Docker containers, each of which contains a subset of functionality from services in Azure AI services. You can find instructions and image locations in the tables below.

 **Note**

See [Install and run Document Intelligence containers](#) for Azure AI Document Intelligence container instructions and image locations.

Decision containers

 Expand table

Service	Container	Description	Availability
Anomaly detector	Anomaly Detector (image 	The Anomaly Detector API enables you to monitor and detect abnormalities in your time series data with machine learning.	Generally available

Language containers

 Expand table

Service	Container	Description	Availability
LUIS	LUIS (image 	Loads a trained or published Language Understanding model, also known as a LUIS app, into a docker container and provides access to the query predictions from the container's API endpoints. You can collect query logs from the container and upload these back to the LUIS portal  to improve the app's prediction accuracy.	Generally available. This container can also run in disconnected environments .

Service	Container	Description	Availability
Language service	Key Phrase Extraction (image )	Extracts key phrases to identify the main points. For example, for the input text "The food was delicious and there were wonderful staff," the API returns the main talking points: "food" and "wonderful staff".	Generally available. This container can also run in disconnected environments.
Language service	Text Language Detection (image )	For up to 120 languages, detects which language the input text is written in and report a single language code for every document submitted on the request. The language code is paired with a score indicating the strength of the score.	Generally available. This container can also run in disconnected environments.
Language service	Sentiment Analysis (image )	Analyzes raw text for clues about positive or negative sentiment. This version of sentiment analysis returns sentiment labels (for example <i>positive</i> or <i>negative</i>) for each document and sentence within it.	Generally available. This container can also run in disconnected environments.
Language service	Text Analytics for health (image )	Extract and label medical information from unstructured clinical text.	Generally available
Language service	Named Entity Recognition (image )	Extract named entities from text.	Generally available. This container can also run in disconnected environments.
Language service	Custom Named Entity Recognition (image )	Extract named entities from text, using a custom model you create using your data.	Generally available
Language service	Summarization (image )	Summarize text from various sources.	Public preview. This container can also run in disconnected environments.
Translator	Translator (image )	Translate text in several languages and dialects.	Generally available. Gated - request access  . This container can also run in

Service	Container	Description	Availability
			disconnected environments.

Speech containers

[\[+\] Expand table](#)

Service	Container	Description	Availability
Speech Service API	Speech to text (image ↗)	Transcribes continuous real-time speech into text.	Generally available. This container can also run in disconnected environments.
Speech Service API	Custom Speech to text (image ↗)	Transcribes continuous real-time speech into text using a custom model.	Generally available. This container can also run in disconnected environments.
Speech Service API	Neural Text to speech (image ↗)	Converts text to natural-sounding speech using deep neural network technology, allowing for more natural synthesized speech.	Generally available. This container can also run in disconnected environments.
Speech Service API	Speech language identification (image ↗)	Determines the language of spoken audio.	Preview

Vision containers

[\[+\] Expand table](#)

Service	Container	Description	Availability
Azure AI Vision	Read OCR (image ↗)	The Read OCR container allows you to extract printed and handwritten text from images and documents with support for JPEG, PNG, BMP, PDF, and TIFF file formats. For more information, see the Read API documentation .	Generally Available. This container can also run in disconnected environments.
Spatial Analysis	Spatial analysis (image ↗)	Analyzes real-time streaming video to understand spatial relationships between people, their movement, and interactions with objects in physical environments.	Preview

Additionally, some containers are supported in the [Azure AI services multi-service resource](#) offering. You can create one single Azure AI services resource and use the same billing key across supported services for the following services:

- Azure AI Vision
- LUIS
- Language service

Prerequisites

You must satisfy the following prerequisites before using Azure AI containers:

Docker Engine: You must have Docker Engine installed locally. Docker provides packages that configure the Docker environment on [macOS](#), [Linux](#), and [Windows](#). On Windows, Docker must be configured to support Linux containers. Docker containers can also be deployed directly to [Azure Kubernetes Service](#) or [Azure Container Instances](#).

Docker must be configured to allow the containers to connect with and send billing data to Azure.

Familiarity with Microsoft Container Registry and Docker: You should have a basic understanding of both Microsoft Container Registry and Docker concepts, like registries, repositories, containers, and container images, as well as knowledge of basic `docker` commands.

For a primer on Docker and container basics, see the [Docker overview](#).

Individual containers can have their own requirements, as well, including server and memory allocation requirements.

Azure AI services container security

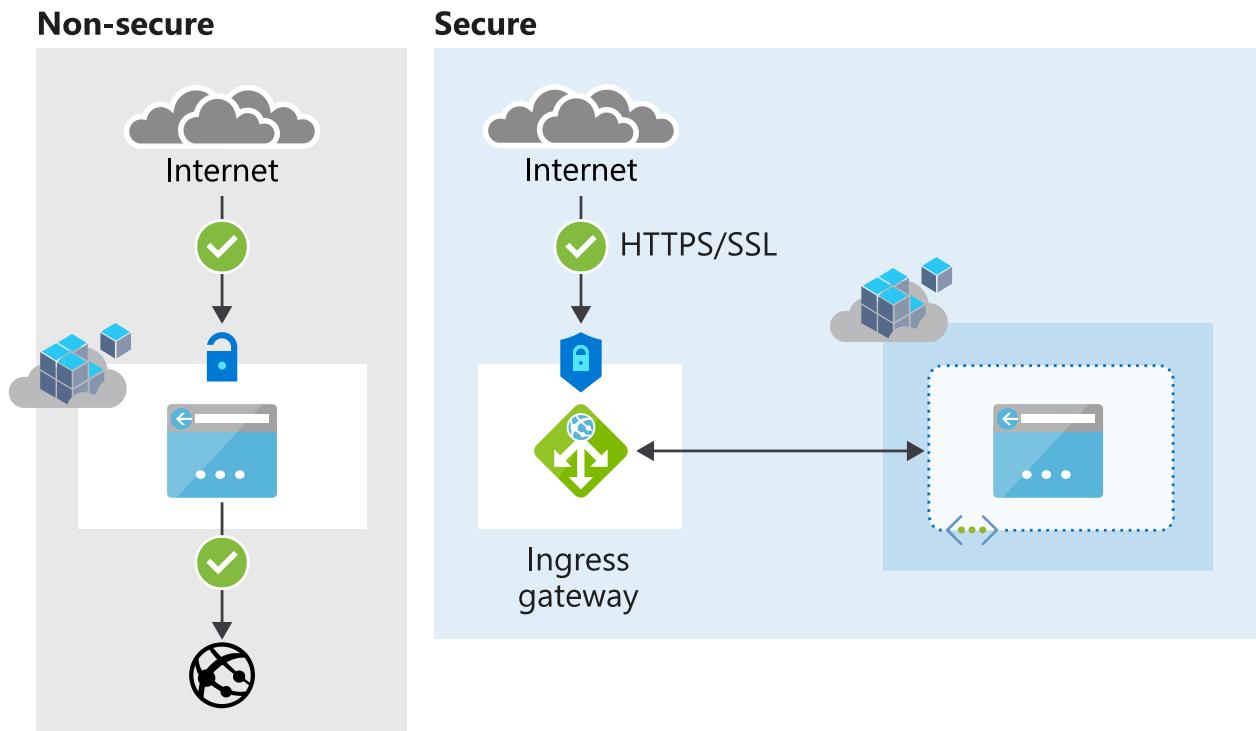
Security should be a primary focus whenever you're developing applications. The importance of security is a metric for success. When you're architecting a software solution that includes Azure AI containers, it's vital to understand the limitations and capabilities available to you. For more information about network security, see [Configure Azure AI services virtual networks](#).

Important

By default there is *no security* on the Azure AI services container API. The reason for this is that most often the container will run as part of a pod which is protected

from the outside by a network bridge. However, it is possible for users to construct their own authentication infrastructure to approximate the authentication methods used when accessing the [cloud-based Azure AI services](#).

The following diagram illustrates the default and **non-secure** approach:



As an example of an alternative and *secure* approach, consumers of Azure AI containers could augment a container with a front-facing component, keeping the container endpoint private. Let's consider a scenario where we use [Istio](#) as an ingress gateway. Istio supports HTTPS/TLS and client-certificate authentication. In this scenario, the Istio frontend exposes the container access, presenting the client certificate that is approved beforehand with Istio.

[Nginx](#) is another popular choice in the same category. Both Istio and Nginx act as a service mesh and offer additional features including things like load-balancing, routing, and rate-control.

Container networking

The Azure AI containers are required to submit metering information for billing purposes. Failure to allowlist various network channels that the Azure AI containers rely on will prevent the container from working.

Allowlist Azure AI services domains and ports

The host should allowlist **port 443** and the following domains:

- *.cognitive.microsoft.com
- *.cognitiveservices.azure.com

Disable deep packet inspection

Deep packet inspection (DPI) [↗](#) is a type of data processing that inspects in detail the data sent over a computer network, and usually takes action by blocking, rerouting, or logging it accordingly.

Disable DPI on the secure channels that the Azure AI containers create to Microsoft servers. Failure to do so will prevent the container from functioning correctly.

Developer samples

Developer samples are available at our [GitHub repository ↗](#).

Next steps

Learn about [container recipes](#) you can use with the Azure AI services.

Install and explore the functionality provided by containers in Azure AI services:

- [Anomaly Detector containers](#)
- [Azure AI Vision containers](#)
- [Language Understanding \(LUIS\) containers](#)
- [Speech Service API containers](#)
- [Language service containers](#)
- [Translator containers](#)

Feedback

Was this page helpful?



[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

FAQ for Azure AI containers

FAQ

General questions

What is available?

Azure AI containers allow developers to use the same intelligent APIs that are available in Azure, but with the [benefits](#) of containerization. Some containers are available as a gated preview that might require an application to access. Other containers are publicly available as an un gated preview, or are generally available. You can find a full list of containers and their availability in the [Container support in Azure AI services](#) article. You can also view the containers in the [Microsoft Container Registry \(MCR\)](#).

Is there any difference between the Azure AI services cloud and the containers?

Azure AI containers are an alternative to the Azure AI services cloud. Containers offer the same capabilities as the corresponding cloud services. Customers can deploy the containers on-premises or in Azure. The core AI technology, pricing tiers, API keys, and API signature are the same between the container and the corresponding cloud services. There are [features and benefits](#) for choosing containers over their cloud service equivalent.

How do I access and use a gated preview container?

Previously, gated preview containers were hosted on the [containerpreview.azurecr.io](#) repository. As of September 22, 2020, these containers are hosted on the Microsoft Container Registry. Downloading them doesn't require you to use the `docker login` command. You can run a gated preview container if your Azure resource was created with the approved Azure subscription ID. You won't be able to run the container if your Azure subscription hasn't been approved after completion of the [request form](#).

Will containers be available for all Azure AI services offerings, and what's the next set of

containers I should expect?

We'd like to make more Azure AI services offerings available as containers. Contact your local Microsoft account manager to get updates on new container releases and other Azure AI services announcements.

What is the service-level agreement (SLA) for Azure AI containers?

ⓘ Important

To learn more about service-level agreements for Azure AI services, [visit our SLA page ↗](#).

Azure AI provides SLAs for cloud hosted services that can be viewed on our SLA page.

However, Azure AI services in containers don't provide an SLA as it is on-premises software. Customers control Azure AI services container configurations of resources, so Microsoft is unable to offer an SLA for general availability (GA). Customers are free to deploy containers on-premises and define the host environments.

Are these containers available in Sovereign clouds?

Standard Azure AI containers can be used in the Sovereign clouds. The containers can run and connect to the billing endpoint in these clouds once authorized, but container images must be pulled from the Public cloud container registry. Containers that rely on language models downloaded at runtime, such as Translator or Custom Speech to text, also are configured to pull models from Public endpoints.

Purchasing disconnected containers isn't currently supported in Sovereign clouds. Disconnected containers purchased in Public cloud along with all dependencies can be transferred to Sovereign clouds and run on Sovereign cloud infrastructure.

Versioning

How are containers updated to the latest version?

Customers can choose when to update the containers that they've deployed. Containers are marked with standard [Docker tags](#) such as `latest` to indicate the most recent version. We encourage customers to pull the latest versions of containers as they're released. For details on how to get notified when an image is updated, see [Azure Container Registry webhooks](#).

Container license files are used as keys to decrypt certain files within each container image. If these encrypted files happen to be updated within a new container image, the license file you have may fail to start the container even if it worked with the previous version of the container image. To avoid this issue, we recommend that you download a new license file from the resource endpoint for your container provided in Azure portal after you pull new image versions from mcr.microsoft.com.

To download a new license file, you can add `DownloadLicense=True` to your docker run command along with a license mount, your API Key, and your billing endpoint. Refer to your [container's documentation](#) for detailed instructions.

What versions are supported?

Generally, only the current version of the container is supported. We encourage customers to stay current to get the latest patches and technology.

How are updates versioned?

Major version changes indicate that there's a breaking change to the API signature. We anticipate that this indication will generally coincide with major version changes to the corresponding Azure AI services cloud offering. Minor version changes indicate bug fixes, model updates, or new features that don't make a breaking change to the API signature.

Technical questions

How can I diagnose potential errors in my deployment environment?

If you're having trouble running an Azure AI services container, you can try using the Microsoft diagnostics container. Use this container to diagnose common errors in your deployment environment that might prevent Azure AI containers from functioning as expected.

To get the container, use the following `docker pull` command:

Bash

```
docker pull mcr.microsoft.com/azure-cognitive-services/diagnostic
```

Then run the container. Replace `{ENDPOINT_URI}` with your endpoint, and replace `{API_KEY}` with your key to your resource:

Bash

```
docker run --rm mcr.microsoft.com/azure-cognitive-services/diagnostic \
eula=accept \
Billing={ENDPOINT_URI} \
ApiKey={API_KEY}
```

The container will test for network connectivity to the billing endpoint.

How should I run the Azure AI containers on IoT devices?

Whether you don't have a reliable internet connection, you want to save on bandwidth cost, you have low-latency requirements, or you're dealing with sensitive data that needs to be analyzed on-site, [Azure IoT Edge with Azure AI containers](#) gives you consistency with the cloud.

Are these containers compatible with OpenShift?

We don't test containers with OpenShift, but generally, Azure AI containers should run on any platform that supports Docker images. If you're using OpenShift, we recommend running the containers as `root-user`.

How do I provide product feedback and feature recommendations?

We encourage customers to [voice their concerns](#) publicly and to up-vote others who have done the same where potential issues overlap. You can use the feedback tool for both product feedback and feature recommendations.

What status messages and errors do Azure AI containers return?

Here are the status messages and errors:

[+] Expand table

Status	Description
Valid	Your API key is valid. No action is needed.
Invalid	Your API key is invalid. You must provide a valid API key to run the container. Find your API key and service region in the Azure portal, in the Keys and Endpoint section for your Azure AI services resource.
Mismatch	You've provided an API key or endpoint for a different kind of Azure AI services resource. Find your API key and service region in the Azure portal, in the Keys and Endpoint section for your Azure AI services resource.
CouldNotConnect	The container couldn't connect to the billing endpoint. Check the <code>Retry-After</code> value and wait for this period to end before you make more requests.
OutOfQuota	The API key has exceeded the quota. You can either upgrade your pricing tier or wait for more quota to become available. Find your tier in the Azure portal, in the Pricing Tier section of your Azure AI service resource.
BillingEndpointBusy	The billing endpoint is currently busy. Check the <code>Retry-After</code> value and wait for this period to end before you make more requests.
ContainerUseUnauthorized	The provided API key isn't authorized for use with this container. You're likely using a gated container, so make sure your Azure subscription

Status	Description
	ID is approved by submitting an online request .
<pre>[ERROR] Failed to download: context deadline exceeded</pre>	The model file download request to our servers timed out. Make sure you have a strong internet connection to download all required files within one hour. Model downloads apply to Text Translation and some Speech service containers.
<pre>The provided license path was not found. Please ensure a volume is mounted and a directory exists at the location specified by Mounts:License</pre>	Disconnected containers only. There is no license file stored locally at the location specified in your docker run command. You may have mounted the license volume incorrectly. Be sure to check your local file system and provide a valid filepath to the desired local storage location for the license file using the -v argument in your docker run command.
<pre>Unknown</pre>	The server is currently unable to process billing requests.

How do I get support?

Customer support channels are the same as for the Azure AI services cloud-based APIs. All Azure AI containers include logging features that help us and the community support customers. Here are options for more support:

- Find contact information in your [Azure support plan](#).
- Try these areas of [Stack Overflow](#), a question-and-answer site for professional and enthusiast programmers:
 - [Azure AI services](#)
 - [Microsoft Cognitive](#)

How does billing work?

Customers are charged based on consumption, similar to the Azure AI services cloud. The containers need to be configured to send metering data to Azure, and transactions are billed accordingly. Resources used across the hosted and on-premises services add to the single quota with tiered pricing, counting against both usages. For more information, see the billing section of the [container](#) you're using.

Important

Azure AI containers are *not licensed* to run without being connected to Azure for metering. Customers need to enable the containers to always communicate billing information with the metering service. Azure AI containers don't send customer data to Microsoft.

Here's an example of the information that a container communicates for metering:

JSON

```
usageRequestBody": {  
    "id": "1234abcd-1234-56ab-ab12-1234abcd",  
    "containerType": "<container-type>",  
    "containerVersion": "<container-version>",  
    "containerId": "<contianer-id>",  
    "meter": {  
        "name": "<meter-name>",  
        "quantity": 0.0  
    },  
    "requestTime": 12345687890,  
    "apiType": "<api-type>"  
},
```

What is the current support warranty for containers?

Microsoft's standard warranty for enterprise software applies for all containers formally announced as generally available (GA). There's no warranty for previews.

What happens to Azure AI containers when internet connectivity is lost?

Azure AI containers are *not licensed* to run without being connected to Azure for metering. Customers need to enable the containers to always communicate with the metering service.

How long can the container operate without being connected to Azure?

Azure AI containers are *not licensed* to run without being connected to Azure for metering. Customers need to enable the containers to always communicate with the

metering service.

What hardware is required to run these containers?

Azure AI containers are x64-based containers that can run any compatible Linux node, VM, and edge device that supports x64 Linux Docker containers. They all require CPU processors. For more information, see the container requirements and recommendations section of the [container](#) you're using.

Are these containers currently supported on Windows?

Azure AI containers are Linux containers, but there's some support for Linux containers on Windows. For more information about Linux containers on Windows, see the [Docker documentation](#).

How do Azure AI containers compare to AWS and Google offerings?

Microsoft is the first cloud provider to move its pretrained AI models in containers with simple billing per transaction as though customers are using a cloud service. Microsoft believes a hybrid cloud gives customers more choice.

What compliance certifications do containers have?

Azure AI containers don't have any compliance certifications.

What regions are Azure AI containers available in?

Containers can be run anywhere in any region, but they need a key and need to call back to Azure for metering. All supported regions for the cloud service are supported for the containers' metering call.

- How to use Azure AI containers in disconnected environments

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

Use Docker containers in disconnected environments

Article • 08/28/2024

Containers enable you to run Azure AI services APIs in your own environment, and are great for your specific security and data governance requirements. Disconnected containers enable you to use several of these APIs disconnected from the internet.

Currently, the following containers can be run in this manner:

- [Speech to text](#)
- [Custom Speech to text](#)
- [Neural Text to speech](#)
- [Text Translation \(Standard\)](#)
- Azure AI Language
 - [Sentiment Analysis](#)
 - [Key Phrase Extraction](#)
 - [Language Detection](#)
 - [Summarization](#)
 - [Named Entity Recognition](#)
- [Azure AI Vision - Read](#)
- [Document Intelligence](#)

Before attempting to run a Docker container in an offline environment, make sure you know the steps to successfully download and use the container. For example:

- Host computer requirements and recommendations.
- The Docker `pull` command you'll use to download the container.
- How to validate that a container is running.
- How to send queries to the container's endpoint, once it's running.

Request access to use containers in disconnected environments

Fill out and submit the [request form](#) to request access to the containers disconnected from the internet.

The form requests information about you, your company, and the user scenario for which you'll use the container. After you submit the form, the Azure AI services team reviews it and emails you with a decision within 10 business days.

Important

- On the form, you must use an email address associated with an Azure subscription ID.
- The Azure resource you use to run the container must have been created with the approved Azure subscription ID.
- Check your email (both inbox and junk folders) for updates on the status of your application from Microsoft.

After you're approved, you'll be able to run the container after you download it from the Microsoft Container Registry (MCR), described later in the article.

You won't be able to run the container if your Azure subscription hasn't been approved.

Access is limited to customers that meet the following requirements:

- Your organization should be identified as strategic customer or partner with Microsoft.
- Disconnected containers are expected to run fully offline, hence your use cases must meet one of below or similar requirements:
 - Environment or device(s) with zero connectivity to internet.
 - Remote location that occasionally has internet access.
 - Organization under strict regulation of not sending any kind of data back to cloud.
- Application completed as instructed - Please pay close attention to guidance provided throughout the application to ensure you provide all the necessary information required for approval.

Purchase a commitment tier pricing plan for disconnected containers

Create a new resource

1. Sign in to the [Azure portal](#) and select **Create a new resource** for one of the applicable Azure AI services listed above.
2. Enter the applicable information to create your resource. Be sure to select **Commitment tier disconnected containers** as your pricing tier.

Note

- You will only see the option to purchase a commitment tier if you have been approved by Microsoft.
- Pricing details are for example only.

3. Select **Review + Create** at the bottom of the page. Review the information, and select **Create**.

Configure container for disconnected usage

See the following documentation for steps on downloading and configuring the container for disconnected usage:

- [Vision - Read](#)
- [Language Understanding \(LUIS\)](#)
- [Text Translation \(Standard\)](#)
- [Document Intelligence](#)

Speech service

- [Speech to text](#)
- [Custom Speech to text](#)
- [Neural Text to speech](#)

Language service

- [Sentiment Analysis](#)
- [Key Phrase Extraction](#)
- [Language Detection](#)

Environment variable names in Kubernetes deployments

Some Azure AI Containers, for example Translator, require users to pass environmental variable names that include colons (:) when running the container. This will work fine when using Docker, but Kubernetes does not accept colons in environmental variable names. To resolve this, you can replace colons with double underscore characters (__) when deploying to Kubernetes. See the following example of an acceptable format for environment variable names:

Kubernetes

```
env:  
- name: Mounts__License  
  value: "/license"  
- name: Mounts__Output  
  value: "/output"
```

This example replaces the default format for the `Mounts:License` and `Mounts:Output` environment variable names in the docker run command.

Container image and license updates

Container license files are used as keys to decrypt certain files within each container image. If these encrypted files happen to be updated within a new container image, the license file you have may fail to start the container even if it worked with the previous version of the container image. To avoid this issue, we recommend that you download a new license file from the resource endpoint for your container provided in Azure portal after you pull new image versions from mcr.microsoft.com.

To download a new license file, you can add `DownloadLicense=True` to your docker run command along with a license mount, your API Key, and your billing endpoint. Refer to your [container's documentation](#) for detailed instructions.

Usage records

When operating Docker containers in a disconnected environment, the container will write usage records to a volume where they're collected over time. You can also call a REST endpoint to generate a report about service usage.

Arguments for storing logs

When run in a disconnected environment, an output mount must be available to the container to store usage logs. For example, you would include `-v /host/output:{OUTPUT_PATH}` and `Mounts:Output={OUTPUT_PATH}` in the example below, replacing `{OUTPUT_PATH}` with the path where the logs will be stored:

Docker

```
docker run -v /host/output:{OUTPUT_PATH} ... <image> ... Mounts:Output={OUTPUT_PATH}
```

Get records using the container endpoints

The container provides two endpoints for returning records about its usage.

Get all records

The following endpoint will provide a report summarizing all of the usage collected in the mounted billing record directory.

HTTP

```
https://<service>/records/usage-logs/
```

It will return JSON similar to the example below.

JSON

```
{
  "apiType": "noop",
  "serviceName": "noop",
  "meters": [
    {
      "name": "Sample.Meter",
      "quantity": 253
    }
  ]
}
```

Get records for a specific month

The following endpoint will provide a report summarizing usage over a specific month and year.

HTTP

```
https://<service>/records/usage-logs/{MONTH}/{YEAR}
```

it will return a JSON response similar to the example below:

JSON

```
{
  "apiType": "string",
  "serviceName": "string",
  "meters": [
```

```
{  
    "name": "string",  
    "quantity": 253  
}  
]  
}
```

Purchase a commitment plan to use containers in disconnected environments

Commitment plans for disconnected containers have a calendar year commitment period. When you purchase a plan, you'll be charged the full price immediately. During the commitment period, you can't change your commitment plan, however you can purchase additional unit(s) at a pro-rated price for the remaining days in the year. You have until midnight (UTC) on the last day of your commitment, to end a commitment plan.

You can choose a different commitment plan in the **Commitment Tier pricing** settings of your resource.

End a commitment plan

If you decide that you don't want to continue purchasing a commitment plan, you can set your resource's auto-renewal to **Do not auto-renew**. Your commitment plan will expire on the displayed commitment end date. After this date, you won't be charged for the commitment plan. You'll be able to continue using the Azure resource to make API calls, charged at pay-as-you-go pricing. You have until midnight (UTC) on the last day of the year to end a commitment plan for disconnected containers, and not be charged for the following year.

Troubleshooting

If you run the container with an output mount and logging enabled, the container generates log files that are helpful to troubleshoot issues that happen while starting or running the container.

💡 Tip

For more troubleshooting information and guidance, see [Disconnected containers Frequently asked questions \(FAQ\)](#).

Next steps

[Azure AI containers overview](#)

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

FAQ for Azure AI services disconnected containers

FAQ

General questions

Which services are available as disconnected containers?

Azure AI containers allow developers to use the same intelligent APIs that are available in Azure, but with the [benefits](#) of containerization. Disconnected containers are offered through commitment tier pricing, each offering a discounted rate compared to the pay-as-you-go pricing model. With commitment tier pricing, you can commit to using the following Azure AI services features for a fixed fee, enabling you to have a predictable total cost based on the needs of your workload.

- Azure AI Speech
 - Speech to text
 - Custom Speech to text
 - Neural Text to speech
- Text Translation (Standard)
- Azure AI Language
 - Sentiment Analysis
 - Key Phrase Extraction
 - Language Detection
 - Named Entity Recognition (NER)
- Azure AI Vision - Read
- Document Intelligence

For more information about commitment tier pricing, reach out to your Microsoft account team or contact at Microsoft. You can also email azureaicontainergating@service.microsoft.com.

Will containers be available for other Azure AI services offerings, and what's the next set of containers I should expect?

We'd like to make more Azure AI services offerings available as containers. Contact your local Microsoft account manager to get updates on new container releases and other Azure AI services announcements.

How do I get access to disconnected container?

Fill out and submit the [request form](#) to request access to the container. Access is limited to customers that meet the following requirements:

- Your organization must have an enterprise agreement or an equivalent agreement and should be identified as strategic customer or partner with Microsoft.
- Valid business scenario/use case - Disconnected containers are expected to run fully offline hence, your use cases must meet one of below or similar requirements:
 - Environment or devices with zero connectivity to internet.
 - Remote location that occasionally has internet access.
 - Organization under strict regulation of not sending any kind of data back to cloud.
- Application completed as instructed - Please pay close attention to guidance provided throughout the application to ensure you provide all the necessary information required for approval.

What if my use case can't satisfy the requirements listed above?

If your use case can't satisfy above requirements but you're interested in running containers on premises, you might be able to use [connected containers](#).

What are some reasons my application might be denied?

Possible causes for a denied application are as follows:

- Not being an existing Microsoft partner or enterprise agreement customer
- Invalid Azure subscription ID. Your ID must belong to the applicant organization, and you can't apply on behalf of other organizations.
- The application was submitted with a personal email address (For example: @hotmail.com, @gmail.com, @yahoo.com, etc.)
- Invalid justification or business scenario provided
- Other missing or inaccurate data

How long does it take to approve my application to access disconnected containers?

After you submit the form, the Azure AI services team will review it and email you a decision within 10 business days.

Once your application is approved, the Azure AI services gating team will communicate details for purchasing the commitment tier pricing, and instructions to download and run the containers.

If you have any questions on the application, gating process or other information needed, email azureaicontainergating@service.microsoft.com

How do I download the disconnected containers?

These containers are hosted on the Microsoft Container Registry and available for download on [Microsoft Artifact Registry](#) and [Microsoft Container Registry \(MCR\)](#). You won't be able to run the container if your Azure subscription has not been approved after completion of the request form. Once you're approved, you can follow the steps in [Use Docker containers in disconnected environments](#)

What is the service-level agreement (SLA) for Azure AI containers?

Azure AI services disconnected containers don't have an SLA, because customers control configuration of the resources used. Customers are free to deploy containers on-premises and define the host environments.

Are these containers available in sovereign clouds?

Not everyone is familiar with the term sovereign cloud, so let's begin with definition: The sovereign cloud consists of the [Azure Government](#), [Azure Germany](#), and [Microsoft Azure operated by 21Vianet](#) clouds. Currently, customers are able to create disconnected containers resource from public cloud.

Versioning

How are containers updated to the latest version?

Customers can choose when to update the containers that they've deployed. Containers will be marked with standard Docker tags such as latest to indicate the most recent version. We encourage customers to pull the latest versions of containers as they're released. For details on how to get notified when an image is updated, see Azure Container Registry webhooks.

What versions will be supported?

The current and last major version of the container will be supported. However, we encourage customers to stay current to get the latest technology to be compliant.

How are updates versioned?

Major version changes indicate that there's a breaking change to the API signature. We anticipate that this will generally coincide with major version changes to the corresponding Azure AI services cloud offering. Minor version changes indicate bug fixes, model updates, or new features that don't make a breaking change to the API signature.

Technical questions

How can I diagnose potential errors in my deployment environment?

If you're having trouble running an Azure AI services container, you can try using the Microsoft diagnostics container. Use this container to diagnose common errors in your deployment environment that might prevent Azure AI containers from functioning as expected.

To get the container, use the following `docker pull` command:

```
Bash
```

```
docker pull mcr.microsoft.com/azure-cognitive-services/diagnostic
```

Then run the container. Replace `{ENDPOINT_URI}` with your endpoint, and replace `{API_KEY}` with your key to your resource:

Bash

```
docker run --rm mcr.microsoft.com/azure-cognitive-services/diagnostic \
eula=accept \
Billing={ENDPOINT_URI} \
ApiKey={API_KEY}
```

The container will test for network connectivity to the billing endpoint.

Are these containers compatible with OpenShift?

We don't test containers with OpenShift, but generally, Azure AI containers should run on any platform that supports Docker images. If you're using OpenShift, we recommend running the containers as root-user.

How do I provide product feedback and feature recommendations?

We encourage customers to [voice their concerns](#) publicly and to up-vote others who have done the same where potential issues overlap. You can use the feedback tool for both product feedback and feature recommendations.

What status messages and errors do Azure AI containers return?

Here are the status messages and errors:

[] Expand table

Status	Description
Valid	Your API key is valid. No action is needed.
Invalid	Your API key is invalid. You must provide a valid API key to configure the container for use in disconnected environments. Find your API key and service region in the Azure portal, in the Keys and Endpoint section for your Azure AI services resource.

Status	Description
Mismatch	You've provided an API key or endpoint for a different kind of Azure AI services resource. Find your API key and service region in the Azure portal, in the Keys and Endpoint section for your Azure AI services resource.
CouldNotConnect	The container couldn't connect to the endpoint.
OutOfQuota	The API key has exceeded the quota. You can either upgrade your pricing tier or wait for additional quota to become available. Find your tier in the Azure portal, in the Pricing Tier section of your Azure AI service resource.
ContainerUseUnauthorized	The provided API key isn't authorized for use with this container. You're likely using a gated container, so make sure your Azure subscription ID is approved by submitting an online request ↗ .
Unknown	The server is currently unable to process requests.

How do I get support?

Customer support channels are the same as for the Azure AI services cloud-based APIs. All Azure AI containers include logging features that will help us and the community support customers. Here are options for more support:

- Find contact information in your [Azure support plan ↗](#).
- Try these areas of Stack Overflow, a question-and-answer site for professional and enthusiast programmers:
- [Azure AI services ↗](#)
- [Microsoft Cognitive ↗](#)

How does billing work?

Customers are charged based on the commitment tier they purchase. Choose an annual commitment plan based on the needs of your workload. You can customize the "Commitment tier disconnected containers DC0" pricing tier by selecting one of several commitment tier options.

Each commitment tier offers a discounted rate as compared to the pay-as-you-go pricing model. Commitment plans are charged annually starting from the day the resource was created. Additional units can be added at any time in the year and are charged pro-rated based on the number of days remaining in the billing year. For more information, see the following pricing pages:

💡 Tip

To find the pricing information for disconnected containers on these pricing pages, scroll down to **Commitment Tiers** section.

- [Speech Services ↗](#) for:
 - Speech to text
 - Custom Speech to text
 - Neural Text to speech
- [Text Translation \(Standard\) ↗](#)
- [Azure AI Language pricing ↗](#) for
- Sentiment Analysis
- Key Phrase Extraction
- Language Detection
- [Azure AI Vision - Read ↗](#)
- [Document Intelligence - Custom/Invoice ↗](#)

How do I monitor the container's usage?

Docker containers running in a disconnected environment will write usage records to a volume where they're collected over time. These containers also provide a REST endpoint to generate a report about usage. See [Use containers in disconnected environments](#) for more information.

What is the current support warranty for containers?

Microsoft's standard warranty for enterprise software will apply when a container image is formally announced as General Available (GA). There's no warranty for container images in preview.

What hardware is required to run these containers?

Azure AI containers are x64-based containers that can run any compatible Linux node, Virtual Machine (VM), and edge device that supports x64 Linux Docker containers. They all require CPU processors. You can find the minimum and recommended configurations for each container offering on these pages:

- Speech Service for:
 - [Speech to text](#)
 - [Custom Speech to text](#)
 - [Neural Text to speech](#)
- [Text Translation \(Standard\)](#)
- Azure AI Language documentation for
- [Sentiment Analysis](#)
- [Key Phrase Extraction](#)
- [Language Detection](#)
- [Named Entity Recognition](#)
- [Azure AI Vision - Read](#)
- [Document Intelligence - Custom/Invoice](#)

Are these containers currently supported on Windows?

Azure AI containers are Linux containers, but there's some support for Linux containers on Windows. For more information about Linux containers on Windows, see the [Docker documentation](#).

What compliance certifications do containers have?

Azure AI containers don't have any compliance certifications.

Feedback

Was this page helpful?



[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

Create containers for reuse

Article • 08/28/2024

Use these container recipes to create Azure AI containers that can be reused. Containers can be built with some or all configuration settings so that they are *not* needed when the container is started.

Once you have this new layer of container (with settings), and you have tested it locally, you can store the container in a container registry. When the container starts, it will only need those settings that are not currently stored in the container. The private registry container provides configuration space for you to pass those settings in.

Docker run syntax

Any `docker run` examples in this document assume a Windows console with a `\` line continuation character. Consider the following for your own use:

- Do not change the order of the arguments unless you are very familiar with docker containers.
- If you are using an operating system other than Windows, or a console other than Windows console, use the correct console/terminal, folder syntax for mounts, and line continuation character for your console and system. Because the Azure AI services container is a Linux operating system, the target mount uses a Linux-style folder syntax.
- `docker run` examples use the directory off the `c:\` drive to avoid any permission conflicts on Windows. If you need to use a specific directory as the input directory, you may need to grant the docker service permission.

Store no configuration settings in image

The example `docker run` commands for each service do not store any configuration settings in the container. When you start the container from a console or registry service, those configuration settings need to pass in. The private registry container provides configuration space for you to pass those settings in.

Reuse recipe: store all configuration settings with container

In order to store all configuration settings, create a `Dockerfile` with those settings.

Issues with this approach:

- The new container has a separate name and tag from the original container.
- In order to change these settings, you will have to change the values of the Dockerfile, rebuild the image, and republish to your registry.
- If someone gets access to your container registry or your local host, they can run the container and use the Azure AI services endpoints.
- If the Azure AI service that you're using doesn't require input mounts, don't add the `COPY` lines to your Dockerfile.

Create Dockerfile, pulling from the existing Azure AI services container you want to use, then use docker commands in the Dockerfile to set or pull in information the container needs.

This example:

- Sets the billing endpoint, `{BILLING_ENDPOINT}` from the host's environment key using `ENV`.
- Sets the billing API-key, `{ENDPOINT_KEY}` from the host's environment key using ``ENV``.

Reuse recipe: store billing settings with container

This example shows how to build the Language service's sentiment container from a Dockerfile.

Dockerfile

```
FROM mcr.microsoft.com/azure-cognitive-services/sentiment:latest
ENV billing={BILLING_ENDPOINT}
ENV apikey={ENDPOINT_KEY}
ENV EULA=accept
```

Build and run the container [locally](#) or from your [private registry container](#) as needed.

Reuse recipe: store billing and mount settings with container

This example shows how to use Language Understanding, saving billing and models from the Dockerfile.

- Copies the Language Understanding (LUIS) model file from the host's file system using `COPY`.
- The LUIS container supports more than one model. If all models are stored in the same folder, you all need one `COPY` statement.
- Run the docker file from the relative parent of the model input directory. For the following example, run the `docker build` and `docker run` commands from the relative parent of `/input`. The first `/input` on the `COPY` command is the host computer's directory. The second `/input` is the container's directory.

Dockerfile

```
FROM <container-registry>/<cognitive-service-container-name>:<tag>
ENV billing={BILLING_ENDPOINT}
ENV apikey={ENDPOINT_KEY}
ENV EULA=accept
COPY /input /input
```

Build and run the container [locally](#) or from your [private registry container](#) as needed.

How to use container on your local host

To build the Docker file, replace `<your-image-name>` with the new name of the image, then use:

Console

```
docker build -t <your-image-name> .
```

To run the image, and remove it when the container stops (`--rm`):

Console

```
docker run --rm <your-image-name>
```

How to add container to private registry

Follow these steps to use the Dockerfile and place the new image in your private container registry.

1. Create a `Dockerfile` with the text from reuse recipe. A `Dockerfile` doesn't have an extension.

2. Replace any values in the angle brackets with your own values.
3. Build the file into an image at the command line or terminal, using the following command. Replace the values in the angle brackets, <>, with your own container name and tag.

The tag option, `-t`, is a way to add information about what you have changed for the container. For example, a container name of `modified-LUIS` indicates the original container has been layered. A tag name of `with-billing-and-model` indicates how the Language Understanding (LUIS) container has been modified.

Bash

```
docker build -t <your-new-container-name>:<your-new-tag-name> .
```

4. Sign in to Azure CLI from a console. This command opens a browser and requires authentication. Once authenticated, you can close the browser and continue working in the console.

Azure CLI

```
az login
```

5. Sign in to your private registry with Azure CLI from a console.

Replace the values in the angle brackets, `<my-registry>`, with your own registry name.

Azure CLI

```
az acr login --name <my-registry>
```

You can also sign in with docker login if you are assigned a service principal.

Bash

```
docker login <my-registry>.azurecr.io
```

6. Tag the container with the private registry location. Replace the values in the angle brackets, `<my-registry>`, with your own registry name.

Bash

```
docker tag <your-new-container-name>:<your-new-tag-name> <my-registry>.azurecr.io/<your-new-container-name-in-registry>:<your-new-tag-name>
```

If you don't use a tag name, `latest` is implied.

7. Push the new image to your private container registry. When you view your private container registry, the container name used in the following CLI command will be the name of the repository.

Bash

```
docker push <my-registry>.azurecr.io/<your-new-container-name-in-registry>:<your-new-tag-name>
```

Next steps

[Create and use Azure Container Instance](#)

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

Deploy and run container on Azure Container Instance

Article • 08/28/2024

With the following steps, scale Azure AI services applications in the cloud easily with Azure [Container Instances](#). Containerization helps you focus on building your applications instead of managing the infrastructure. For more information on using containers, see [features and benefits](#).

Prerequisites

The recipe works with any Azure AI services container. The Azure AI services resource must be created before using the recipe. Each Azure AI service that supports containers has a "How to install" article for installing and configuring the service for a container. Some services require a file or set of files as input for the container, it is important that you understand and have used the container successfully before using this solution.

- An Azure resource for the Azure AI service you're using.
- Azure AI service resource **endpoint URL** - review your specific service's "How to install" for the container, to find where the endpoint URL is from within the Azure portal, and what a correct example of the URL looks like. The exact format can change from service to service.
- Azure AI service resource **key** - the keys are on the **Keys** page for the Azure resource. You only need one of the two keys. The key is a string of 32 alpha-numeric characters.
- A single Azure AI services container on your local host (your computer). Make sure you can:
 - Pull down the image with a `docker pull` command.
 - Run the local container successfully with all required configuration settings with a `docker run` command.
 - Call the container's endpoint, getting a response of HTTP 2xx and a JSON response back.

All variables in angle brackets, `<>`, need to be replaced with your own values. This replacement includes the angle brackets.

 **Important**

The LUIS container requires a `.gz` model file that is pulled in at runtime. The container must be able to access this model file via a volume mount from the container instance. To upload a model file, follow these steps:

1. [Create an Azure file share](#). Take note of the Azure Storage account name, key, and file share name as you'll need them later.
2. [export your LUIS model \(packaged app\) from the LUIS portal](#).
3. In the Azure portal, navigate to the **Overview** page of your storage account resource, and select **File shares**.
4. Select the file share name that you recently created, then select **Upload**. Then upload your packaged app.

Azure portal

Create an Azure Container Instance resource using the Azure portal

1. Go to the [Create](#) page for Container Instances.
2. On the **Basics** tab, enter the following details:

[Expand table](#)

Setting	Value
Subscription	Select your subscription.
Resource group	Select the available resource group or create a new one such as <code>cognitive-services</code> .
Container name	Enter a name such as <code>cognitive-container-instance</code> . The name must be in lower caps.
Location	Select a region for deployment.
Image type	If your container image is stored in a container registry that doesn't require credentials, choose <code>Public</code> . If accessing your container image requires credentials, choose <code>Private</code> . Refer to container repositories and images for details on whether or not the container image is <code>Public</code> or <code>Private</code> ("Public Preview").

Setting	Value
Image name	<p>Enter the Azure AI services container location. The location is what's used as an argument to the <code>docker pull</code> command. Refer to the container repositories and images for the available image names and their corresponding repository.</p> <p>The image name must be fully qualified specifying three parts. First, the container registry, then the repository, finally the image name: <code><container-registry>/<repository>/<image-name></code>.</p> <p>Here is an example, <code>mcr.microsoft.com/azure-cognitive-services/keyphrase</code> would represent the Key Phrase Extraction image in the Microsoft Container Registry under the Azure AI services repository. Another example is, <code>containerpreview.azurecr.io/microsoft/cognitive-services-speech-to-text</code> which would represent the Speech to text image in the Microsoft repository of the Container Preview container registry.</p>
OS type	Linux
Size	<p>Change size to the suggested recommendations for your specific Azure AI container:</p> <ul style="list-style-type: none"> 2 CPU cores 4 GB

3. On the **Networking** tab, enter the following details:

[] [Expand table](#)

Setting	Value
Ports	Set the TCP port to <code>5000</code> . Exposes the container on port 5000.

4. On the **Advanced** tab, enter the required **Environment Variables** for the container billing settings of the Azure Container Instance resource:

[] [Expand table](#)

Key	Value
ApiKey	Copied from the Keys and endpoint page of the resource. It is a 32 alphanumeric-character string with no spaces or dashes, <code>xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx</code> .
Billing	Your endpoint URL copied from the Keys and endpoint page of the resource.

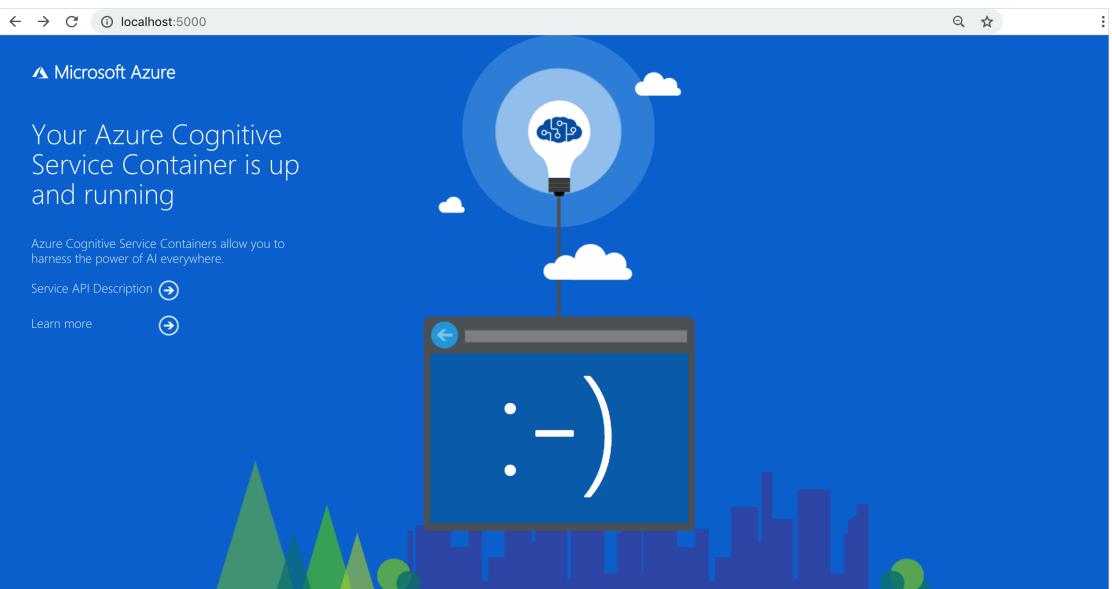
Key	Value
Eula	accept

5. Select **Review and Create**
6. After validation passes, click **Create** to finish the creation process
7. When the resource is successfully deployed, it's ready

Use the Container Instance

Azure portal

1. Select the **Overview** and copy the IP address. It will be a numeric IP address such as `55.55.55.55`.
2. Open a new browser tab and use the IP address, for example, `http://<IP-address>:5000` (`http://55.55.55.55:5000`). You will see the container's home page, letting you know the container is running.



3. Select **Service API Description** to view the swagger page for the container.
4. Select any of the **POST APIs** and select **Try it out**. The parameters are displayed including the input. Fill in the parameters.
5. Select **Execute** to send the request to your Container Instance.

You have successfully created and used Azure AI containers in Azure Container Instance.

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

Deploy a language detection container to Azure Kubernetes Service

Article • 08/28/2024

Learn how to deploy the language detection container. This procedure shows you how to create the local Docker containers, push the containers to your own private container registry, run the container in a Kubernetes cluster, and test it in a web browser.

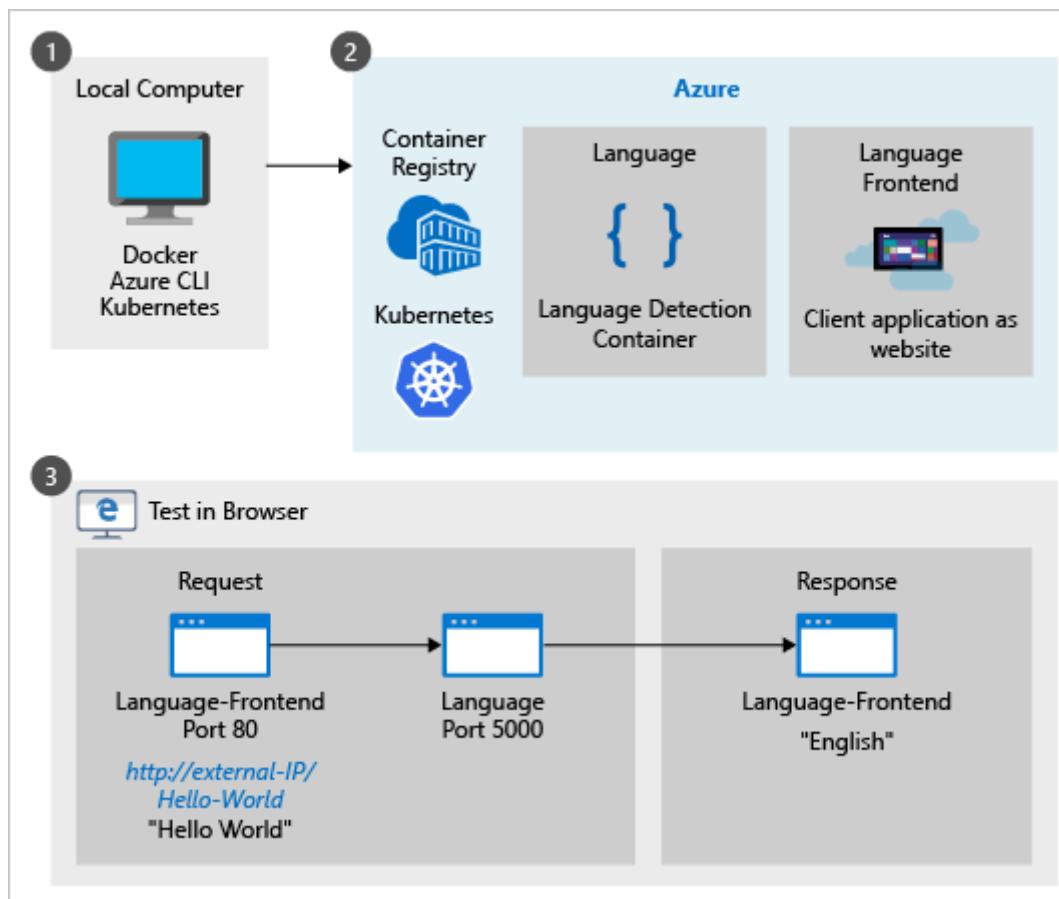
Prerequisites

This procedure requires several tools that must be installed and run locally. Do not use Azure Cloud Shell.

- Use an Azure subscription. If you don't have an Azure subscription, create a [free account](#) before you begin.
- [Git](#) for your operating system so you can clone the [sample](#) used in this procedure.
- [Azure CLI](#).
- [Docker engine](#) and validate that the Docker CLI works in a console window.
- [kubectl](#).
- An Azure resource with the correct pricing tier. Not all pricing tiers work with this container:
 - [Language](#) resource with F0 or Standard pricing tiers only.
 - [Azure AI services](#) resource with the S0 pricing tier.

Running the sample

This procedure loads and runs the Azure AI services container sample for language detection. The sample has two containers, one for the client application and one for the Azure AI services container. We'll push both of these images to the Azure Container Registry. Once they are on your own registry, create an Azure Kubernetes Service to access these images and run the containers. When the containers are running, use the `kubectl` CLI to watch the containers performance. Access the client application with an HTTP request and see the results.



The sample containers

The sample has two container images, one for the frontend website. The second image is the language detection container returning the detected language (culture) of text. Both containers are accessible from an external IP when you are done.

The language-frontend container

This website is equivalent to your own client-side application that makes requests of the language detection endpoint. When the procedure is finished, you get the detected language of a string of characters by accessing the website container in a browser with `http://<external-IP>/<text-to-analyze>`. An example of this URL is `http://132.12.23.255/helloworld!`. The result in the browser is English.

The language container

The language detection container, in this specific procedure, is accessible to any external request. The container hasn't been changed in any way so the standard Azure AI services container-specific language detection API is available.

For this container, that API is a POST request for language detection. As with all Azure AI containers, you can learn more about the container from its hosted Swagger information, <http://<external-IP>:5000/swagger/index.html>.

Port 5000 is the default port used with the Azure AI containers.

Create Azure Container Registry service

To deploy the container to the Azure Kubernetes Service, the container images need to be accessible. Create your own Azure Container Registry service to host the images.

1. Sign in to the Azure CLI

```
Azure CLI
```

```
az login
```

2. Create a resource group named `cogserv-container-rg` to hold every resource created in this procedure.

```
Azure CLI
```

```
az group create --name cogserv-container-rg --location westus
```

3. Create your own Azure Container Registry with the format of your name then `registry`, such as `pattyregistry`. Do not use dashes or underline characters in the name.

```
Azure CLI
```

```
az acr create --resource-group cogserv-container-rg --name pattyregistry --sku Basic
```

Save the results to get the `loginServer` property. This will be part of the hosted container's address, used later in the `language.yml` file.

```
Azure CLI
```

```
az acr create --resource-group cogserv-container-rg --name pattyregistry --sku Basic
```

```
Output
```

```
{  
    "adminUserEnabled": false,  
    "creationDate": "2019-01-02T23:49:53.783549+00:00",  
    "id": "/subscriptions/xxxxxxxx-xxxx-xxxx-xxxx-  
xxxxxxxxxx/resourceGroups/cogserv-container-  
rg/providers/Microsoft.ContainerRegistry/registries/pattyregistry",  
    "location": "westus",  
    "loginServer": "pattyregistry.azurecr.io",  
    "name": "pattyregistry",  
    "provisioningState": "Succeeded",  
    "resourceGroup": "cogserv-container-rg",  
    "sku": {  
        "name": "Basic",  
        "tier": "Basic"  
    },  
    "status": null,  
    "storageAccount": null,  
    "tags": {},  
    "type": "Microsoft.ContainerRegistry/registries"  
}
```

4. Sign in to your container registry. You need to login before you can push images to your registry.

Azure CLI

```
az acr login --name pattyregistry
```

Get website Docker image

1. The sample code used in this procedure is in the Azure AI containers samples repository. Clone the repository to have a local copy of the sample.

Console

```
git clone https://github.com/Azure-Samples/cognitive-services-  
containers-samples
```

Once the repository is on your local computer, find the website in the [\dotnet\Language\FrontendService](#) directory. This website acts as the client application calling the language detection API hosted in the language detection container.

2. Build the Docker image for this website. Make sure the console is in the [\FrontendService](#) directory where the Dockerfile is located when you run the

following command:

Console

```
docker build -t language-frontend -t  
pattyregistry.azurecr.io/language-frontend:v1 .
```

To track the version on your container registry, add the tag with a version format, such as v1.

3. Push the image to your container registry. This may take a few minutes.

Console

```
docker push pattyregistry.azurecr.io/language-frontend:v1
```

If you get an unauthorized: authentication required error, login with the az acr login --name <your-container-registry-name> command.

When the process is done, the results should be similar to:

Output

```
The push refers to repository [pattyregistry.azurecr.io/language-  
frontend]  
82ff52ee6c73: Pushed  
07599c047227: Pushed  
816caf41a9a1: Pushed  
2924be3aed17: Pushed  
45b83a23806f: Pushed  
ef68f6734aa4: Pushed  
v1: digest:  
sha256:31930445deee181605c0cde53dab5a104528dc1ff57e5b3b34324f0d8a0eb286  
size: 1580
```

Get language detection Docker image

1. Pull the latest version of the Docker image to the local machine. This may take a few minutes. If there is a newer version of this container, change the value from 1.1.006770001-amd64-preview to the newer version.

Console

```
docker pull mcr.microsoft.com/azure-cognitive-  
services/language:1.1.006770001-amd64-preview
```

2. Tag image with your container registry. Find the latest version and replace the version `1.1.006770001-amd64-preview` if you have a more recent version.

```
Console
```

```
docker tag mcr.microsoft.com/azure-cognitive-services/language  
pattyregistry.azurecr.io/language:1.1.006770001-amd64-preview
```

3. Push the image to your container registry. This may take a few minutes.

```
Console
```

```
docker push pattyregistry.azurecr.io/language:1.1.006770001-amd64-  
preview
```

Get Container Registry credentials

The following steps are needed to get the required information to connect your container registry with the Azure Kubernetes Service you create later in this procedure.

1. Create service principal.

```
Azure CLI
```

```
az ad sp create-for-rbac
```

Save the results `appId` value for the assignee parameter in step 3, `<appId>`. Save the `password` for the next section's client-secret parameter `<client-secret>`.

```
Output
```

```
{  
  "appId": "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",  
  "displayName": "azure-cli-2018-12-31-18-39-32",  
  "name": "http://azure-cli-2018-12-31-18-39-32",  
  "password": "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",  
  "tenant": "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx"  
}
```

2. Get your container registry ID.

```
Azure CLI
```

```
az acr show --resource-group cogserv-container-rg --name pattyregistry
```

```
--query "id" --output table
```

Save the output for the scope parameter value, <acrId>, in the next step. It looks like:

Output

```
/subscriptions/xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx/resourceGroups/cogserv-container-rg/providers/Microsoft.ContainerRegistry/registries/pattyregistry
```

Save the full value for step 3 in this section.

3. To grant the correct access for the AKS cluster to use images stored in your container registry, create a role assignment. Replace <appId> and <acrId> with the values gathered in the previous two steps.

Azure CLI

```
az role assignment create --assignee <appId> --scope <acrId> --role Reader
```

Create Azure Kubernetes Service

1. Create the Kubernetes cluster. All the parameter values are from previous sections except the name parameter. Choose a name that indicates who created it and its purpose, such as patty-kube.

Azure CLI

```
az aks create --resource-group cogserv-container-rg --name patty-kube --node-count 2 --service-principal <appId> --client-secret <client-secret> --generate-ssh-keys
```

This step may take a few minutes. The result is:

Output

```
{
  "aadProfile": null,
  "addonProfiles": null,
  "agentPoolProfiles": [
    {
      "count": 2,
      "dnsPrefix": null,
```

```

        "fqdn": null,
        "maxPods": 110,
        "name": "nodepool1",
        "osDiskSizeGb": 30,
        "osType": "Linux",
        "ports": null,
        "storageProfile": "ManagedDisks",
        "vmSize": "Standard_DS1_v2",
        "vnetSubnetId": null
    }
],
"dnsPrefix": "patty-kube--65a101",
"enableRbac": true,
"fqdn": "patty-kube--65a101-341f1f54.hcp.westus.azmk8s.io",
"id": "/subscriptions/xxxxxxxx-xxxx-xxxx-xxxx-
xxxxxxxx/resourcegroups/cogserv-container-
rg/providers/Microsoft.ContainerService/managedClusters/patty-kube",
"kubernetesVersion": "1.9.11",
"linuxProfile": {
    "adminUsername": "azureuser",
    "ssh": {
        "publicKeys": [
            {
                "keyData": "ssh-rsa AAAAB3NzaC...ohR2d81mFC"
            }
        ]
    }
},
"location": "westus",
"name": "patty-kube",
"networkProfile": {
    "dnsServiceIp": "10.0.0.10",
    "dockerBridgeCidr": "172.17.0.1/16",
    "networkPlugin": "kubenet",
    "networkPolicy": null,
    "podCidr": "10.244.0.0/16",
    "serviceCidr": "10.0.0.0/16"
},
"nodeResourceGroup": "MC_patty_westus",
"provisioningState": "Succeeded",
"resourceGroup": "cogserv-container-rg",
"servicePrincipalProfile": {
    "clientId": "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
    "keyVaultSecretRef": null,
    "secret": null
},
"tags": null,
"type": "Microsoft.ContainerService/ManagedClusters"
}

```

The service is created but it doesn't have the website container or language detection container yet.

2. Get credentials of the Kubernetes cluster.

Azure CLI

```
az aks get-credentials --resource-group cogserv-container-rg --name patty-kube
```

Load the orchestration definition into your Kubernetes service

This section uses the **kubectl** CLI to talk with the Azure Kubernetes Service.

1. Before loading the orchestration definition, check **kubectl** has access to the nodes.

Console

```
kubectl get nodes
```

The response looks like:

Output

NAME	STATUS	ROLES	AGE	VERSION
aks-nodepool1-13756812-0	Ready	agent	6m	v1.9.11
aks-nodepool1-13756812-1	Ready	agent	6m	v1.9.11

2. Copy the following file and name it `language.yml`. The file has a `service` section and a `deployment` section each for the two container types, the `language-frontend` website container and the `language` detection container.

yml

```
# A service which exposes the .net frontend app container through a dependable hostname: http://language-frontend:5000
apiVersion: v1
kind: Service
metadata:
  name: language-frontend
  labels:
    run: language-frontend
spec:
  selector:
    app: language-frontend
  type: LoadBalancer
  ports:
    - name: front
```

```

    port: 80
    targetPort: 80
    protocol: TCP
---
# A deployment declaratively indicating how many instances of the .net
frontend app container we want up
apiVersion: apps/v1beta1
kind: Deployment
metadata:
  name: language-frontend
spec:
  replicas: 1
  template:
    metadata:
      labels:
        app: language-frontend
    spec:
      containers:
        - name: language-frontend
          image: # < URI of the Frontend App image >
          ports:
            - name: public-port
              containerPort: 80
          livenessProbe:
            httpGet:
              path: /status
              port: public-port
            initialDelaySeconds: 30
            timeoutSeconds: 1
            periodSeconds: 10
          imagePullSecrets:
            - name: # < Name of the registry secret providing access to the
              frontend image >
          automountServiceAccountToken: false
---
# A service which exposes the cognitive-service containers through a
dependable hostname: http://language:5000
apiVersion: v1
kind: Service
metadata:
  name: language
  labels:
    run: language
spec:
  selector:
    app: language
  type: LoadBalancer
  ports:
    - name: language
      port: 5000
      targetPort: 5000
      protocol: TCP
---
# A deployment declaratively indicating how many instances of the
cognitive-service container we want up

```

```

apiVersion: apps/v1beta1
kind: Deployment
metadata:
  name: language
spec:
  replicas: 1
  template:
    metadata:
      labels:
        app: language
    spec:
      containers:
        - name: language
          image: # < URI of the Language Image >
          ports:
            - name: public-port
              containerPort: 5000
          livenessProbe:
            httpGet:
              path: /status
              port: public-port
          initialDelaySeconds: 30
          timeoutSeconds: 1
          periodSeconds: 10
      args:
        - "eula=accept"
        - "apikey=" # < API Key for the Language Service >
        - "billing=" # < Language billing endpoint URI >
      imagePullSecrets:
        - name: # < Name of the registry secret providing access to the
          Language image >
      automountServiceAccountToken: false

```

3. Change the language-frontend deployment lines of `language.yaml` based on the following table to add your own container registry image names, client secret, and Language service settings.

[] Expand table

Language-frontend deployment settings	Purpose
Line 32 <code>image</code> property	Image location for the frontend image in your Container Registry <code><container-registry-name>.azurecr.io/language-frontend:v1</code>
Line 44 <code>name</code> property	Container Registry secret for the image, referred to as <code><client-secret></code> in a previous section.

4. Change the language deployment lines of `language.yml` based on the following table to add your own container registry image names, client secret, and Language service settings.

[Expand table](#)

Language deployment settings	Purpose
Line 78 <code>image</code> property	Image location for the language image in your Container Registry <code><container-registry-name>.azurecr.io/language:1.1.006770001-amd64-preview</code>
Line 95 <code>name</code> property	Container Registry secret for the image, referred to as <code><client-secret></code> in a previous section.
Line 91 <code>apiKey</code> property	Your Language service resource key
Line 92 <code>billing</code> property	The billing endpoint for your Language service resource. <code>https://westus.api.cognitive.microsoft.com/text/analytics/v2.1</code>

Because the `apiKey` and `billing endpoint` are set as part of the Kubernetes orchestration definition, the website container doesn't need to know about these or pass them as part of the request. The website container refers to the language detection container by its orchestrator name `language`.

5. Load the orchestration definition file for this sample from the folder where you created and saved the `language.yml`.

Console

```
kubectl apply -f language.yml
```

The response is:

Output

```
service "language-frontend" created
deployment.apps "language-frontend" created
service "language" created
deployment.apps "language" created
```

Get external IPs of containers

For the two containers, verify the `language-frontend` and `language` services are running and get the external IP address.

Console

```
kubectl get all
```

Output

NAME	READY	STATUS	RESTARTS	AGE
pod/language-586849d8dc-7zvz5	1/1	Running	0	13h
pod/language-frontend-68b9969969-bz9bg	1/1	Running	1	13h
NAME	TYPE	CLUSTER-IP	EXTERNAL-IP	
PORT(S)	AGE			
service/kubernetes	ClusterIP	10.0.0.1	<none>	
443/TCP	14h			
service/language	LoadBalancer	10.0.39.169	104.42.172.68	
5000:30161/TCP	13h			
service/language-frontend	LoadBalancer	10.0.42.136	104.42.37.219	
80:30943/TCP	13h			
NAME	DESIRED	CURRENT	UP-TO-DATE	
AVAILABLE AGE				
deployment.extensions/language	1	1	1	1
13h				
deployment.extensions/language-frontend	1	1	1	1
13h				
NAME	DESIRED	CURRENT		
READY AGE				
replicaset.extensions/language-586849d8dc	1	1	1	1
13h				
replicaset.extensions/language-frontend-68b9969969	1	1	1	1
13h				
NAME	DESIRED	CURRENT	UP-TO-DATE	
AVAILABLE AGE				
deployment.apps/language	1	1	1	1
13h				
deployment.apps/language-frontend	1	1	1	1
13h				
NAME	DESIRED	CURRENT	READY	
AGE				
replicaset.apps/language-586849d8dc	1	1	1	
13h				

```
replicaset.apps/language-frontend-68b9969969    1        1        1  
13h
```

If the `EXTERNAL-IP` for the service is shown as pending, rerun the command until the IP address is shown before moving to the next step.

Test the language detection container

Open a browser and navigate to the external IP of the `language` container from the previous section: `http://<external-ip>:5000/swagger/index.html`. You can use the `Try it` feature of the API to test the language detection endpoint.

The screenshot shows the Swagger UI interface for the Language Detection Cognitive Service API. At the top, there's a green header bar with the word "swagger" and a dropdown menu "Select a spec" set to "Language Detection Cognitive Service API V1". Below the header, the main title "Language Detection Cognitive Service API" is followed by a small "v2" badge. Underneath the title, a brief description explains that the API returns detected language and a score between 0 and 1, with higher scores indicating 100% certainty. It also mentions support for 120 languages. The interface is divided into sections: "Detect Language" and "Status". The "Detect Language" section has a green "POST" button next to the URL "/text/analytics/v2.0/languages Detect Language". The "Status" section has a blue "GET" button next to the URL "/status". Both sections have a small downward arrow icon at the end of their respective headers.

Test the client application container

Change the URL in the browser to the external IP of the `language-frontend` container using the following format: `http://<external-ip>/helloworld`. The English culture text of `helloworld` is predicted as `English`.

Clean up resources

When you are done with the cluster, delete the Azure resource group.

```
Azure CLI
```

```
az group delete --name cogserv-container-rg
```

Related information

- [kubectl for Docker Users ↗](#)

Next steps

[Azure AI containers](#)

Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

Use Docker Compose to deploy multiple containers

Article • 08/28/2024

This article shows you how to deploy multiple Azure AI containers. Specifically, you'll learn how to use Docker Compose to orchestrate multiple Docker container images.

Docker Compose [↗](#) is a tool for defining and running multi-container Docker applications. In Compose, you use a YAML file to configure your application's services. Then, you create and start all the services from your configuration by running a single command.

It can be useful to orchestrate multiple container images on a single host computer. In this article, we'll pull together the Read and Document Intelligence containers.

Prerequisites

This procedure requires several tools that must be installed and run locally:

- An Azure subscription. If you don't have one, create a [free account ↗](#) before you begin.
- [Docker Engine ↗](#). Confirm that the Docker CLI works in a console window.
- An Azure resource with the correct pricing tier. Only the following pricing tiers work with this container:
 - **Azure AI Vision** resource with F0 or Standard pricing tier only.
 - **Document Intelligence** resource with F0 or Standard pricing tier only.
 - **Azure AI services** resource with the S0 pricing tier.
- If you're using a gated preview container, You will need to complete the [online request form ↗](#) to use it.

Docker Compose file

The YAML file defines all the services to be deployed. These services rely on either a [DockerFile](#) or an existing container image. In this case, we'll use two preview images. Copy and paste the following YAML file, and save it as *docker-compose.yaml*. Provide the appropriate **apikey**, **billing**, and **EndpointUri** values in the file.

YAML

```

version: '3.7'
services:
  forms:
    image: "mcr.microsoft.com/azure-cognitive-services/form-
recognizer/layout"
    environment:
      eula: accept
      billing: # < Your Document Intelligence billing URL >
      apikey: # < Your Document Intelligence API key >
      FormRecognizer__ComputerVisionApiKey: # < Your Document Intelligence
API key >
      FormRecognizer__ComputerVisionEndpointUri: # < Your Document
Intelligence URI >
    volumes:
      - type: bind
        source: E:\publicpreview\output
        target: /output
      - type: bind
        source: E:\publicpreview\input
        target: /input
    ports:
      - "5010:5000"

  ocr:
    image: "mcr.microsoft.com/azure-cognitive-services/vision/read:3.1-
preview"
    environment:
      eula: accept
      apikey: # < Your Azure AI Vision API key >
      billing: # < Your Azure AI Vision billing URL >
    ports:
      - "5021:5000"

```

Important

Create the directories on the host machine that are specified under the **volumes** node. This approach is required because the directories must exist before you try to mount an image by using volume bindings.

Start the configured Docker Compose services

A Docker Compose file enables the management of all the stages in a defined service's life cycle: starting, stopping, and rebuilding services; viewing the service status; and log streaming. Open a command-line interface from the project directory (where the docker-compose.yaml file is located).

(!) Note

To avoid errors, make sure that the host machine correctly shares drives with Docker Engine. For example, if *E:\publicpreview* is used as a directory in the *docker-compose.yaml* file, share drive **E** with Docker.

From the command-line interface, execute the following command to start (or restart) all the services defined in the *docker-compose.yaml* file:

Console

```
docker-compose up
```

The first time Docker executes the **docker-compose up** command by using this configuration, it pulls the images configured under the **services** node and then downloads and mounts them:

Console

```
Pulling forms (mcr.microsoft.com/azure-cognitive-services/form-
recognizer/layout:)...  
latest: Pulling from azure-cognitive-services/form-recognizer/layout  
743f2d6c1f65: Pull complete  
72befba99561: Pull complete  
2a40b9192d02: Pull complete  
c7715c9d5c33: Pull complete  
f0b33959f1c4: Pull complete  
b8ab86c6ab26: Pull complete  
41940c21ed3c: Pull complete  
e3d37dd258d4: Pull complete  
cdb5eb761109: Pull complete  
fd93b5f95865: Pull complete  
ef41dcbe5857: Pull complete  
4d05c86a4178: Pull complete  
34e811d37201: Pull complete  
Pulling ocr (mcr.microsoft.com/azure-cognitive-services/vision/read:3.1-
preview:)...  
latest: Pulling from /azure-cognitive-services/vision/read:3.1-preview  
f476d66f5408: Already exists  
8882c27f669e: Already exists  
d9af21273955: Already exists  
f5029279ec12: Already exists  
1a578849dcd1: Pull complete  
45064b1ab0bf: Download complete  
4bb846705268: Downloading [=====>  
] 187.1MB/222.8MB  
c56511552241: Waiting
```

```
e91d2aa0f1ad: Downloading [=====]>
] 162.2MB/176.1MB
```

After the images are downloaded, the image services are started:

Console

```
Starting docker_ocr_1 ... done
Starting docker_forms_1 ... done
Attaching to docker_ocr_1,
docker_forms_1
forms_1 | forms_1 | forms_1 | Notice: This Preview is made
available to you on the condition that you agree to the Supplemental Terms
of Use for Microsoft Azure Previews [https://go.microsoft.com/fwlink/?linkid=2018815], which supplement your agreement
[https://go.microsoft.com/fwlink/?linkid=2018657] governing your use of
Azure. If you do not have an existing agreement governing your use of Azure,
you agree that your agreement governing use of Azure is the Microsoft Online
Subscription Agreement [https://go.microsoft.com/fwlink/?linkid=2018755]
(which incorporates the Online Services Terms
[https://go.microsoft.com/fwlink/?linkid=2018760]). By using the Preview you
agree to these terms.

forms_1 |
forms_1 |
forms_1 | Using '/input' for reading models and other read-only data.
forms_1 | Using '/output/forms/812d811d1bcc' for writing logs and other
output data.
forms_1 | Logging to console.
forms_1 | Submitting metering to
'https://westus2.api.cognitive.microsoft.com/'.
forms_1 | WARNING: No access control enabled!
forms_1 | warn: Microsoft.AspNetCore.Server.Kestrel[0]
forms_1 |         Overriding address(es) 'http://+:80'. Binding to endpoints
defined in UseKestrel() instead.
forms_1 | Hosting environment: Production
forms_1 | Content root path: /app/forms
forms_1 | Now listening on: http://0.0.0.0:5000
forms_1 | Application started. Press Ctrl+C to shut down.

ocr_1 |
ocr_1 |
ocr_1 | Notice: This Preview is made available to you on the condition
that you agree to the Supplemental Terms of Use for Microsoft Azure Previews
[https://go.microsoft.com/fwlink/?linkid=2018815], which supplement your
agreement [https://go.microsoft.com/fwlink/?linkid=2018657] governing your
use of Azure. If you do not have an existing agreement governing your use of Azure,
you agree that your agreement governing use of Azure is the Microsoft
Online Subscription Agreement [https://go.microsoft.com/fwlink/?linkid=2018755]
(which incorporates the Online Services Terms
[https://go.microsoft.com/fwlink/?linkid=2018760]). By using the Preview you
agree to these terms.

ocr_1 |
ocr_1 |
ocr_1 | Logging to console.
ocr_1 | Submitting metering to
'https://westcentralus.api.cognitive.microsoft.com/'.
```

```
ocr_1    | WARNING: No access control enabled!
ocr_1    | Hosting environment: Production
ocr_1    | Content root path: /
ocr_1    | Now listening on: http://0.0.0.0:5000
ocr_1    | Application started. Press Ctrl+C to shut down.
```

Verify the service availability

💡 Tip

You can use the [docker images](#) command to list your downloaded container images. For example, the following command lists the ID, repository, and tag of each downloaded container image, formatted as a table:

```
docker images --format "table {{.ID}}\t{{.Repository}}\t{{.Tag}}"
```

IMAGE ID	REPOSITORY	TAG
<image-id>	<repository-path/name>	<tag-name>

Here's some example output:

IMAGE ID	REPOSITORY
TAG	
2ce533f88e80	mcr.microsoft.com/azure-cognitive-services/form-recognizer/layout
	latest
4be104c126c5	mcr.microsoft.com/azure-cognitive-services/vision/read:3.1-preview
	latest

Test containers

Open a browser on the host machine and go to **localhost** by using the specified port from the *docker-compose.yaml* file, such as <http://localhost:5021/swagger/index.html>. For example, you could use the **Try It** feature in the API to test the Document Intelligence endpoint. Both containers swagger pages should be available and testable.

The screenshot shows the Swagger UI for the Cognitive Service Form Recognizer API V1 (Preview). At the top, there's a header bar with the URL 'localhost:5010/swagger/index.html' and the 'swagger' logo. To the right, there are dropdown menus for 'Select a spec' and 'Cognitive Service Form Recognizer API V1 (Preview)'.

The main content area has a title 'Cognitive Service Form Recognizer API V1 (Preview) ↗' and a sub-section 'Extracts information from forms and images into structured data based on a model created by a set of representative training forms.' Below this, there are two sections: 'Form' and 'Status'. The 'Form' section contains several API endpoints:

- POST /formrecognizer/v1.0-preview/custom/train** Train a model using a set of documents.
- GET /formrecognizer/v1.0-preview/custom/models/{id}/keys** Get keys of a model.
- GET /formrecognizer/v1.0-preview/custom/models/{id}** Get information about a model.
- DELETE /formrecognizer/v1.0-preview/custom/models/{id}** Delete a model and all associated pre-processing data.
- GET /formrecognizer/v1.0-preview/custom/models** Get all models.
- POST /formrecognizer/v1.0-preview/custom/models/{id}/analyze** Analyze a document to extract key-value pairs and table information.

The 'Status' section contains one endpoint:

- GET /status**

Next steps

[Azure AI containers](#)

Feedback

Was this page helpful?

Yes

No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

Tutorial: Create a container image for deployment to Azure Container Instances

Article • 08/29/2024

Azure Container Instances enables deployment of Docker containers onto Azure infrastructure without provisioning any virtual machines or adopting a higher-level service. In this tutorial, you package a small Node.js web application into a container image that can be run using Azure Container Instances.

In this article, part one of the series, you:

- ✓ Clone application source code from GitHub
- ✓ Create a container image from application source
- ✓ Test the image in a local Docker environment

In tutorial parts two and three, you upload your image to Azure Container Registry, and then deploy it to Azure Container Instances.

Before you begin

You must satisfy the following requirements to complete this tutorial:

Azure CLI: You must have Azure CLI version 2.0.29 or later installed on your local computer. To find the version, run `az --version`. If you need to install or upgrade, see [Install the Azure CLI](#).

Docker: This tutorial assumes a basic understanding of core Docker concepts like containers, container images, and basic `docker` commands. For a primer on Docker and container basics, see the [Docker overview](#).

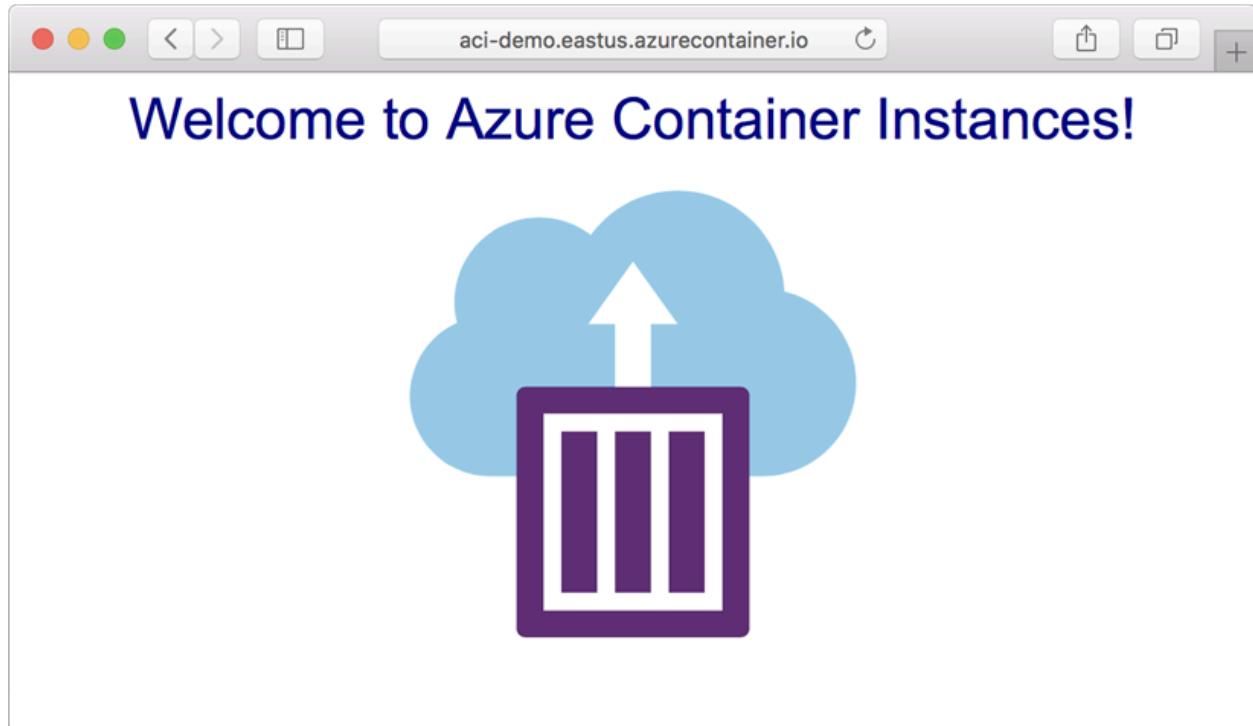
Docker: To complete this tutorial, you need Docker installed locally. Docker provides packages that configure the Docker environment on [macOS](#), [Windows](#), and [Linux](#).

ⓘ Important

Because the Azure Cloud shell does not include the Docker daemon, you *must* install both the Azure CLI and Docker Engine on your *local computer* to complete this tutorial. You cannot use the Azure Cloud Shell for this tutorial.

Get application code

The sample application in this tutorial is a simple web app built in [Node.js](#). The application serves a static HTML page, and looks similar to the following screenshot:



Use Git to clone the sample application's repository:

```
git
git clone https://github.com/Azure-Samples/aci-helloworld.git
```

You can also [download the ZIP archive](#) from GitHub directly.

Build the container image

The Dockerfile in the sample application shows how the container is built. It starts from an [official Node.js image](#) based on [Alpine Linux](#), a small distribution that is well suited for use with containers. It then copies the application files into the container, installs dependencies using the Node Package Manager, and finally, starts the application.

```
Dockerfile
FROM node:8.9.3-alpine
RUN mkdir -p /usr/src/app
COPY ./app/* /usr/src/app/
WORKDIR /usr/src/app
```

```
RUN npm install  
CMD node /usr/src/app/index.js
```

Use the [docker build](#) command to create the container image and tag it as *aci-tutorial-app*:

Bash

```
docker build ./aci-helloworld -t aci-tutorial-app
```

Output from the [docker build](#) command is similar to the following (truncated for readability):

Bash

```
docker build ./aci-helloworld -t aci-tutorial-app
```

Output

```
Sending build context to Docker daemon 119.3kB  
Step 1/6 : FROM node:8.9.3-alpine  
8.9.3-alpine: Pulling from library/node  
88286f41530e: Pull complete  
84f3a4bf8410: Pull complete  
d0d9b2214720: Pull complete  
Digest:  
sha256:c73277ccc763752b42bb2400d1aaecb4e3d32e3a9dbedd0e49885c71bea07354  
Status: Downloaded newer image for node:8.9.3-alpine  
--> 90f5ee24bee2  
...  
Step 6/6 : CMD node /usr/src/app/index.js  
--> Running in f4a1ea099eec  
--> 6edad76d09e9  
Removing intermediate container f4a1ea099eec  
Successfully built 6edad76d09e9  
Successfully tagged aci-tutorial-app:latest
```

Use the [docker images](#) command to see the built image:

Bash

```
docker images
```

Your newly built image should appear in the list:

Bash

```
docker images
```

Output

REPOSITORY	TAG	IMAGE ID	CREATED	SIZE
aci-tutorial-app	latest	5c745774dfa9	39 seconds ago	68.1 MB

Run the container locally

Before you deploy the container to Azure Container Instances, use [docker run](#) to run it locally and confirm that it works. The `-d` switch lets the container run in the background, while `-p` allows you to map an arbitrary port on your computer to port 80 in the container.

Bash

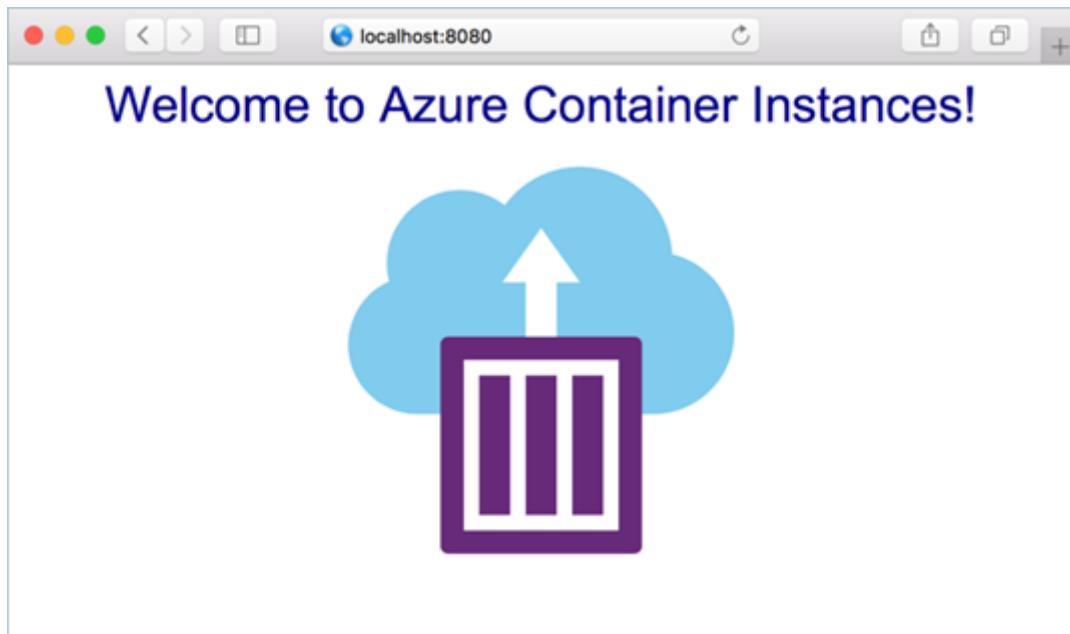
```
docker run -d -p 8080:80 aci-tutorial-app
```

Output from the `docker run` command displays the running container's ID if the command was successful:

Bash

```
docker run -d -p 8080:80 aci-tutorial-app
```
a2e3e4435db58ab0c664ce521854c2e1a1bda88c9cf2fcff46aedf48df86cccf
```

Now, navigate to `http://localhost:8080` in your browser to confirm that the container is running. You should see a web page similar to the following screenshot:



## Next steps

In this tutorial, you created a container image that can be deployed in Azure Container Instances, and verified that it runs locally. So far, you completed the following steps:

- ✓ Cloned the application source from GitHub
- ✓ Created a container image from the application source
- ✓ Tested the container locally

Advance to the next tutorial in the series to learn about storing your container image in Azure Container Registry:

[Push image to Azure Container Registry](#)

---

## Feedback

Was this page helpful?

Yes

No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

# Quickstart: Create a private container registry using the Azure CLI

Article • 09/19/2024

Azure Container Registry is a private registry service for building, storing, and managing container images and related artifacts. In this quickstart, you create an Azure container registry instance with the Azure CLI. Then, use Docker commands to push a container image into the registry, and finally pull and run the image from your registry.

This quickstart requires that you are running the Azure CLI (version 2.0.55 or later recommended). Run `az --version` to find the version. If you need to install or upgrade, see [Install Azure CLI](#).

You must also have Docker installed locally. Docker provides packages that easily configure Docker on any [macOS](#), [Windows](#), or [Linux](#) system.

Because the Azure Cloud Shell doesn't include all required Docker components (the `dockerd` daemon), you can't use the Cloud Shell for this quickstart.

## Create a resource group

Create a resource group with the `az group create` command. An Azure resource group is a logical container into which Azure resources are deployed and managed.

The following example creates a resource group named *myResourceGroup* in the *eastus* location.

Azure CLI

```
az group create --name myResourceGroup --location eastus
```

## Create a container registry

In this quickstart you create a *Basic* registry, which is a cost-optimized option for developers learning about Azure Container Registry. For details on available service tiers, see [Container registry service tiers](#).

Create an ACR instance using the `az acr create` command. The registry name must be unique within Azure, and contain 5-50 lowercase alphanumeric characters. In the following example, *mycontainerregistry* is used. Update this to a unique value.

## Azure CLI

```
az acr create --resource-group myResourceGroup \
--name mycontainerregistry --sku Basic
```

When the registry is created, the output is similar to the following:

### JSON

```
{
 "adminUserEnabled": false,
 "creationDate": "2019-01-08T22:32:13.175925+00:00",
 "id": "/subscriptions/00000000-0000-0000-0000-
0000000000/resourceGroups/myResourceGroup/providers/Microsoft.ContainerReg
istry/registries/mycontainerregistry",
 "location": "eastus",
 "loginServer": "mycontainerregistry.azurecr.io",
 "name": "mycontainerregistry",
 "provisioningState": "Succeeded",
 "resourceGroup": "myResourceGroup",
 "sku": {
 "name": "Basic",
 "tier": "Basic"
 },
 "status": null,
 "storageAccount": null,
 "tags": {},
 "type": "Microsoft.ContainerRegistry/registries"
}
```

Take note of `loginServer` in the output, which is the fully qualified registry name (all lowercase). Throughout the rest of this quickstart `<registry-name>` is a placeholder for the container registry name, and `<login-server>` is a placeholder for the registry's login server name.

### Tip

In this quickstart, you create a *Basic* registry, which is a cost-optimized option for developers learning about Azure Container Registry. Choose other tiers for increased storage and image throughput, and capabilities such as connection using a [private endpoint](#). For details on available service tiers (SKUs), see [Container registry service tiers](#).

## Log in to registry

Before pushing and pulling container images, you must log in to the registry. To do so, use the [az acr login](#) command. Specify only the registry resource name when logging in with the Azure CLI. Don't use the fully qualified login server name.

```
Azure CLI
```

```
az acr login --name <registry-name>
```

Example:

```
Azure CLI
```

```
az acr login --name mycontainerregistry
```

The command returns a `Login Succeeded` message once completed.

## Push image to registry

To push an image to an Azure Container registry, you must first have an image. If you don't yet have any local container images, run the following [docker pull](#) command to pull an existing public image. For this example, pull the `hello-world` image from Microsoft Container Registry.

```
docker pull mcr.microsoft.com/hello-world
```

Before you can push an image to your registry, you must tag it with the fully qualified name of your registry login server. The login server name is in the format `<registry-name>.azurecr.io` (must be all lowercase), for example, `mycontainerregistry.azurecr.io`.

Tag the image using the [docker tag](#) command. Replace `<login-server>` with the login server name of your ACR instance.

```
docker tag mcr.microsoft.com/hello-world <login-server>/hello-world:v1
```

Example:

```
docker tag mcr.microsoft.com/hello-world
mycontainerregistry.azurecr.io/hello-world:v1
```

Finally, use [docker push](#) to push the image to the registry instance. Replace <login-server> with the login server name of your registry instance. This example creates the **hello-world** repository, containing the `hello-world:v1` image.

```
docker push <login-server>/hello-world:v1
```

After pushing the image to your container registry, remove the `hello-world:v1` image from your local Docker environment. (Note that this [docker rmi](#) command does not remove the image from the **hello-world** repository in your Azure container registry.)

```
docker rmi <login-server>/hello-world:v1
```

## List container images

The following example lists the repositories in your registry:

Azure CLI

```
az acr repository list --name <registry-name> --output table
```

Output:

Result

```

```

```
hello-world
```

The following example lists the tags on the **hello-world** repository.

Azure CLI

```
az acr repository show-tags --name <registry-name> --repository hello-world
--output table
```

Output:

```
Result
```

```

```

```
v1
```

## Run image from registry

Now, you can pull and run the `hello-world:v1` container image from your container registry by using [docker run](#):

```
docker run <login-server>/hello-world:v1
```

Example output:

```
Unable to find image 'mycontainerregistry.azurecr.io/hello-world:v1' locally
```

```
v1: Pulling from hello-world
```

```
Digest:
```

```
sha256:662dd8e65ef7ccf13f417962c2f77567d3b132f12c95909de6c85ac3c326a345
```

```
Status: Downloaded newer image for mycontainerregistry.azurecr.io/hello-world:v1
```

```
Hello from Docker!
```

```
This message shows that your installation appears to be working correctly.
```

```
[...]
```

## Clean up resources

When no longer needed, you can use the [az group delete](#) command to remove the resource group, the container registry, and the container images stored there.

```
Azure CLI
```

```
az group delete --name myResourceGroup
```

## Next steps

In this quickstart, you created an Azure Container Registry with the Azure CLI, pushed a container image to the registry, and pulled and ran the image from the registry. Continue to the Azure Container Registry tutorials for a deeper look at ACR.

[Azure Container Registry tutorials](#)

[Azure Container Registry Tasks tutorials](#)

---

## Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

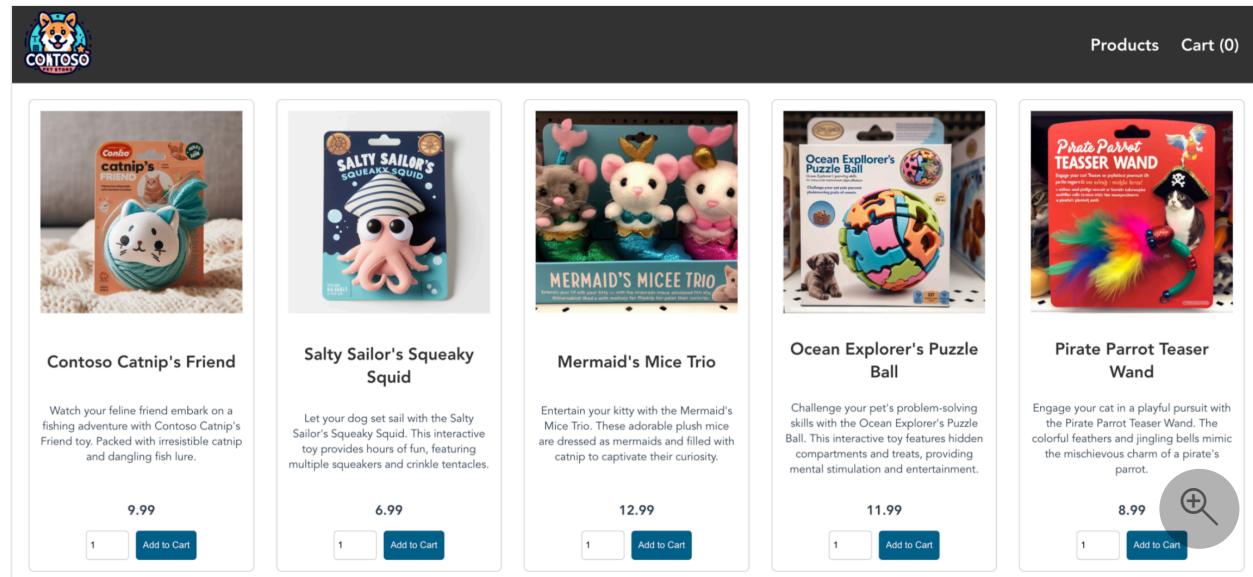
# Tutorial - Prepare an application for Azure Kubernetes Service (AKS)

Article • 09/06/2024

In this tutorial, part one of seven, you prepare a multi-container application to use in Kubernetes. You use existing development tools like Docker Compose to locally build and test the application. You learn how to:

- ✓ Clone a sample application source from GitHub.
- ✓ Create a container image from the sample application source.
- ✓ Test the multi-container application in a local Docker environment.

Once completed, the following application runs in your local development environment:



In later tutorials, you upload the container image to an Azure Container Registry (ACR), and then deploy it into an AKS cluster.

## Before you begin

This tutorial assumes a basic understanding of core Docker concepts such as containers, container images, and `docker` commands. For a primer on container basics, see [Get started with Docker ↗](#).

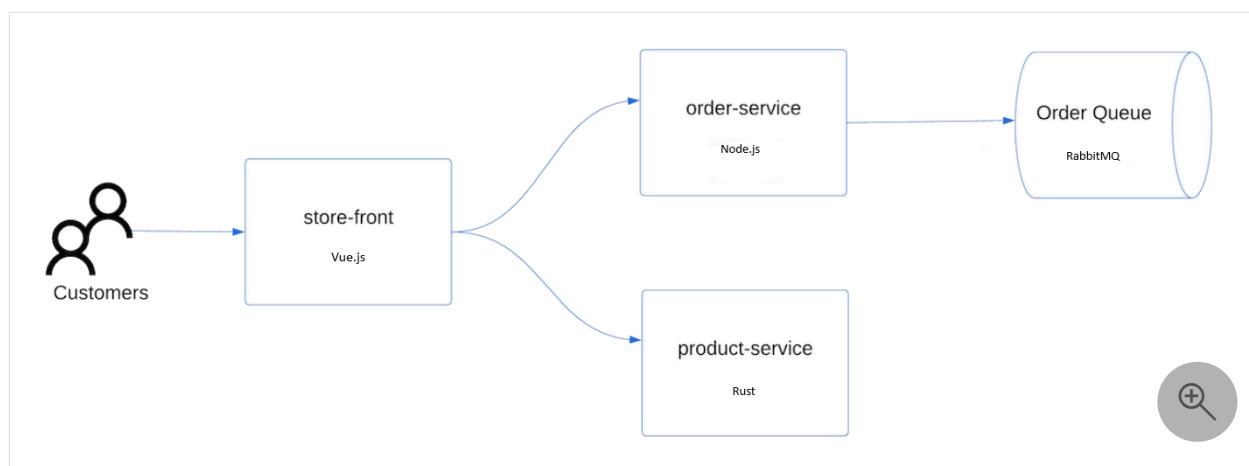
To complete this tutorial, you need a local Docker development environment running Linux containers. Docker provides packages that configure Docker on a [Mac ↗](#), [Windows ↗](#), or [Linux ↗](#) system.

## (!) Note

Azure Cloud Shell doesn't include the Docker components required to complete every step in these tutorials. Therefore, we recommend using a full Docker development environment.

# Get application code

The [sample application](#) used in this tutorial is a basic store front app including the following Kubernetes deployments and services:



- **Store front:** Web application for customers to view products and place orders.
- **Product service:** Shows product information.
- **Order service:** Places orders.
- **Rabbit MQ:** Message queue for an order queue.

## Git

1. Use [git](#) to clone the sample application to your development environment.

### Console

```
git clone https://github.com/Azure-Samples/aks-store-demo.git
```

2. Change into the cloned directory.

### Console

```
cd aks-store-demo
```

## Review Docker Compose file

The sample application you create in this tutorial uses the [docker-compose-quickstart](#) YAML file ↗ from the [repository](#) ↗ you cloned.

YAML

```
services:
 rabbitmq:
 image: rabbitmq:3.13.2-management-alpine
 container_name: 'rabbitmq'
 restart: always
 environment:
 - "RABBITMQ_DEFAULT_USER=username"
 - "RABBITMQ_DEFAULT_PASS=password"
 ports:
 - 15672:15672
 - 5672:5672
 healthcheck:
 test: ["CMD", "rabbitmqctl", "status"]
 interval: 30s
 timeout: 10s
 retries: 5
 volumes:
 - ./rabbitmq_enabled_plugins:/etc/rabbitmq/enabled_plugins
 networks:
 - backend_services
order-service:
 build: src/order-service
 container_name: 'order-service'
 restart: always
 ports:
 - 3000:3000
 healthcheck:
 test: ["CMD", "wget", "-O", "/dev/null", "-q", "http://order-service:3000/health"]
 interval: 30s
 timeout: 10s
 retries: 5
 environment:
 - ORDER_QUEUE_HOSTNAME=rabbitmq
 - ORDER_QUEUE_PORT=5672
 - ORDER_QUEUE_USERNAME=username
 - ORDER_QUEUE_PASSWORD=password
 - ORDER_QUEUE_NAME=orders
 - ORDER_QUEUE_RECONNECT_LIMIT=3
 networks:
```

```

 - backend_services
depends_on:
 rabbitmq:
 condition: service_healthy
product-service:
 build: src/product-service
 container_name: 'product-service'
 restart: always
 ports:
 - 3002:3002
 healthcheck:
 test: ["CMD", "wget", "-O", "/dev/null", "-q", "http://product-
service:3002/health"]
 interval: 30s
 timeout: 10s
 retries: 5
 environment:
 - AI_SERVICE_URL=http://ai-service:5001/
networks:
 - backend_services
store-front:
 build: src/store-front
 container_name: 'store-front'
 restart: always
 ports:
 - 8080:8080
 healthcheck:
 test: ["CMD", "wget", "-O", "/dev/null", "-q", "http://store-
front:80/health"]
 interval: 30s
 timeout: 10s
 retries: 5
 environment:
 - VUE_APP_PRODUCT_SERVICE_URL=http://product-service:3002/
 - VUE_APP_ORDER_SERVICE_URL=http://order-service:3000/
networks:
 - backend_services
depends_on:
 - product-service
 - order-service
networks:
 backend_services:
 driver: bridge

```

## Create container images and run application

Docker

You can use [Docker Compose](#) to automate building container images and the deployment of multi-container applications.

## Docker

1. Create the container image, download the RabbitMQ image, and start the application using the `docker compose` command:

Console

```
docker compose -f docker-compose-quickstart.yml up -d
```

2. View the created images using the `docker images` command.

Console

```
docker images
```

The following condensed example output shows the created images:

Output

| REPOSITORY                     | TAG                      |
|--------------------------------|--------------------------|
| IMAGE ID                       |                          |
| aks-store-demo-product-service | latest                   |
| 72f5cd7e6b84                   |                          |
| aks-store-demo-order-service   | latest                   |
| 54ad5de546f9                   |                          |
| aks-store-demo-store-front     | latest                   |
| 1125f85632ae                   |                          |
| rabbitmq                       | 3.13.2-management-alpine |
| b1dafc50c098                   |                          |
| ...                            |                          |

3. View the running containers using the `docker ps` command.

Console

```
docker ps
```

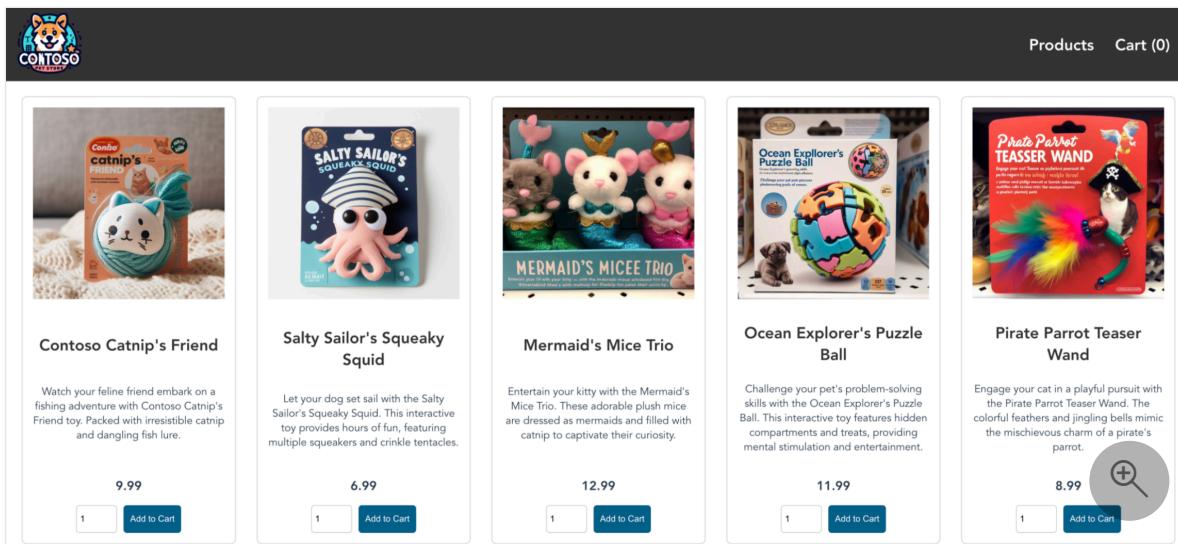
The following condensed example output shows four running containers:

Output

| CONTAINER ID | IMAGE                             |
|--------------|-----------------------------------|
| f27fe74cf0a  | aks-store-demo-product-service    |
| df1eaa137885 | aks-store-demo-order-service      |
| b3ce9e496e96 | aks-store-demo-store-front        |
| 31df28627ffa | rabbitmq:3.13.2-management-alpine |

## Test application locally

To see your running application, navigate to <http://localhost:8080> in a local web browser. The sample application loads, as shown in the following example:



On this page, you can view products, add them to your cart, and then place an order.

## Clean up resources

Since you validated the application's functionality, you can stop and remove the running containers. ***Do not delete the container images*** - you use them in the next tutorial.

- Stop and remove the container instances and resources using the [docker-compose down ↗](#) command.

Console

```
docker compose down
```

# Next steps

Azure CLI

In this tutorial, you created a sample application, created container images for the application, and then tested the application. You learned how to:

- ✓ Clone a sample application source from GitHub.
- ✓ Create a container image from the sample application source.
- ✓ Test the multi-container application in a local Docker environment.

In the next tutorial, you learn how to store container images in an ACR.

[Push images to Azure Container Registry](#)

# Azure AI services security

Article • 08/28/2024

Security should be considered a top priority in the development of all applications, and with the growth of artificial intelligence enabled applications, security is even more important. This article outlines various security features available for Azure AI services. Each feature addresses a specific liability, so multiple features can be used in the same workflow.

For a comprehensive list of Azure service security recommendations, see the [Azure AI services security baseline](#) article.

## Security features

[+] Expand table

| Feature                                        | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">Transport Layer Security (TLS)</a> | All of the Azure AI services endpoints exposed over HTTP enforce the TLS 1.2 protocol. With an enforced security protocol, consumers attempting to call an Azure AI services endpoint should follow these guidelines: <ul style="list-style-type: none"><li>• The client operating system (OS) needs to support TLS 1.2.</li><li>• The language (and platform) used to make the HTTP call need to specify TLS 1.2 as part of the request. Depending on the language and platform, specifying TLS is done either implicitly or explicitly.</li><li>• For .NET users, consider the <a href="#">Transport Layer Security best practices</a></li></ul>                                                         |
| <a href="#">Authentication options</a>         | Authentication is the act of verifying a user's identity. Authorization, by contrast, is the specification of access rights and privileges to resources for a given identity. An identity is a collection of information about a <a href="#">principal</a> , and a principal can be either an individual user or a service.<br><br>By default, you authenticate your own calls to Azure AI services using the subscription keys provided; this is the simplest method but not the most secure. The most secure authentication method is to use managed roles in Microsoft Entra ID. To learn about this and other authentication options, see <a href="#">Authenticate requests to Azure AI services</a> . |
| <a href="#">Key rotation</a>                   | Each Azure AI services resource has two API keys to enable secret rotation. This is a security precaution that lets you regularly change the keys that can access your service, protecting the privacy of your service if a key gets leaked. To learn about this and other authentication options, see <a href="#">Rotate keys</a> .                                                                                                                                                                                                                                                                                                                                                                       |

| Feature                     | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Environment variables       | <p>Environment variables are name-value pairs that are stored within a specific development environment. Environment variables are more secure than using hardcoded values in your code. For instructions on how to use environment variables in your code, see the <a href="#">Environment variables guide</a>.</p> <p>However, if your environment is compromised, the environment variables are compromised as well, so this isn't the most secure approach. The most secure authentication method is to use managed roles in Microsoft Entra ID. To learn about this and other authentication options, see <a href="#">Authenticate requests to Azure AI services</a>.</p>                  |
| Customer-managed keys (CMK) | <p>This feature is for services that store customer data at rest (longer than 48 hours). While this data is already double-encrypted on Azure servers, users can get extra security by adding another layer of encryption, with keys they manage themselves. You can link your service to Azure Key Vault and manage your data encryption keys there.</p> <p>Check to see if CMK is supported by the service that you want to use in the <a href="#">Customer-managed keys</a> documentation.</p>                                                                                                                                                                                               |
| Virtual networks            | <p>Virtual networks allow you to specify which endpoints can make API calls to your resource. The Azure service will reject API calls from devices outside of your network. You can set a formula-based definition of the allowed network, or you can define an exhaustive list of endpoints to allow. This is another layer of security that can be used in combination with others.</p>                                                                                                                                                                                                                                                                                                       |
| Data loss prevention        | <p>The data loss prevention feature lets an administrator decide what types of URLs their Azure resource can take as inputs (for those API calls that take URLs as input). This can be done to prevent the possible exfiltration of sensitive company data: If a company stores sensitive information (such as a customer's private data) in URL parameters, a bad actor inside that company could submit the sensitive URLs to an Azure service, which surfaces that data outside the company. Data loss prevention lets you configure the service to reject certain URI forms on arrival.</p>                                                                                                 |
| Customer Lockbox            | <p>The Customer Lockbox feature provides an interface for customers to review and approve or reject data access requests. It's used in cases where a Microsoft engineer needs to access customer data during a support request. For information on how Customer Lockbox requests are initiated, tracked, and stored for later reviews and audits, see the <a href="#">Customer Lockbox guide</a>.</p> <p>Customer Lockbox is available for the following services:</p> <ul style="list-style-type: none"> <li>• Azure OpenAI</li> <li>• Translator</li> <li>• Conversational language understanding</li> <li>• Custom text classification</li> <li>• Custom named entity recognition</li> </ul> |

| Feature                       | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                               | <ul style="list-style-type: none"><li>Orchestration workflow</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Bring your own storage (BYOS) | <p>The Speech service doesn't currently support Customer Lockbox. However, you can arrange for your service-specific data to be stored in your own storage resource using bring-your-own-storage (BYOS). BYOS allows you to achieve similar data controls to Customer Lockbox. Keep in mind that Speech service data stays and is processed in the Azure region where the Speech resource was created. This applies to any data at rest and data in transit. For customization features like Custom Speech and Custom Voice, all customer data is transferred, stored, and processed in the same region where the Speech service resource and BYOS resource (if used) reside.</p> <p>To use BYOS with Speech, follow the <a href="#">Speech encryption of data at rest</a> guide.</p> <p>Microsoft doesn't use customer data to improve its Speech models. Additionally, if endpoint logging is disabled and no customizations are used, then no customer data is stored by Speech.</p> |

## Next steps

- Explore [Azure AI services](#) and choose a service to get started.

## Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

# Authenticate requests to Azure AI services

Article • 08/28/2024

Each request to an Azure AI service must include an authentication header. This header passes along a resource key or authentication token, which is used to validate your subscription for a service or group of services. In this article, you'll learn about three ways to authenticate a request and the requirements for each.

- Authenticate with a [single-service](#) or [multi-service](#) resource key
- Authenticate with a [token](#)
- Authenticate with [Microsoft Entra ID](#)

## Prerequisites

Before you make a request, you need an Azure account and an Azure AI services subscription. If you already have an account, go ahead and skip to the next section. If you don't have an account, we have a guide to get you set up in minutes: [Create an Azure AI services resource](#).

Go to your resource in the Azure portal. The **Keys & Endpoint** section can be found in the **Resource Management** section. Copy your endpoint and access key as you'll need both for authenticating your API calls. You can use either `KEY1` or `KEY2`. Always having two keys allows you to securely rotate and regenerate keys without causing a service disruption.

## Authentication headers

Let's quickly review the authentication headers available for use with Azure AI services.

[+] [Expand table](#)

| Header                       | Description                                                                                                                                                                 |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ocp-Apim-Subscription-Key    | Use this header to authenticate with a resource key for a specific service or a multi-service resource key.                                                                 |
| Ocp-Apim-Subscription-Region | This header is only required when using a multi-service resource key with the <a href="#">Azure AI Translator service</a> . Use this header to specify the resource region. |

| Header        | Description                                                                                                                                                                                                 |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Authorization | Use this header if you are using an access token. The steps to perform a token exchange are detailed in the following sections. The value provided follows this format: <code>Bearer &lt;TOKEN&gt;</code> . |

## Authenticate with a single-service resource key

The first option is to authenticate a request with a resource key for a specific service, like Azure AI Translator. The keys are available in the Azure portal for each resource that you've created. Go to your resource in the Azure portal. The **Keys & Endpoint** section can be found in the **Resource Management** section. Copy your endpoint and access key as you'll need both for authenticating your API calls. You can use either `KEY1` or `KEY2`. Always having two keys allows you to securely rotate and regenerate keys without causing a service disruption.

To use a resource key to authenticate a request, it must be passed along as the `Ocp-Apim-Subscription-Key` header. This is a sample call to the Azure AI Translator service:

This is a sample call to the Translator service:

### cURL

```
curl -X POST 'https://api.cognitive.microsofttranslator.com/translate?api-version=3.0&from=en&to=de' \
-H 'Ocp-Apim-Subscription-Key: YOUR_SUBSCRIPTION_KEY' \
-H 'Content-Type: application/json' \
--data-binary '[{"text": "How much for the cup of coffee?"}]' | json_pp
```

## Authenticate with a multi-service resource key

You can use a [multi-service](#) resource key to authenticate requests. The main difference is that the multi-service resource key isn't tied to a specific service, rather, a single key can be used to authenticate requests for multiple Azure AI services. See [Azure AI services pricing](#) for information about regional availability, supported features, and pricing.

The resource key is provided in each request as the `Ocp-Apim-Subscription-Key` header.

## Supported regions

When using the [Azure AI services multi-service](#) resource key to make a request to `api.cognitive.microsoft.com`, you must include the region in the URL. For example:

`westus.api.cognitive.microsoft.com`.

When using a multi-service resource key with [Azure AI Translator](#), you must specify the resource region with the `Ocp-Apim-Subscription-Region` header.

Multi-service resource authentication is supported in these regions:

- `australiaeast`
- `brazilsouth`
- `canadacentral`
- `centralindia`
- `eastasia`
- `eastus`
- `japaneast`
- `northeurope`
- `southcentralus`
- `southeastasia`
- `uksouth`
- `westcentralus`
- `westeurope`
- `westus`
- `westus2`
- `francecentral`
- `koreacentral`
- `northcentralus`
- `southafricanorth`
- `uaenorth`
- `switzerlandnorth`

## Sample requests

This is a sample call to the Azure AI Translator service:

cURL

```
curl -X POST 'https://api.cognitive.microsofttranslator.com/translate?api-version=3.0&from=en&to=de' \
-H 'Ocp-Apim-Subscription-Key: YOUR_SUBSCRIPTION_KEY' \
-H 'Ocp-Apim-Subscription-Region: YOUR_SUBSCRIPTION_REGION' \
-H 'Content-Type: application/json' \
--data-binary '[{"text": "How much for the cup of coffee?"}]' | json_pp
```

# Authenticate with an access token

Some Azure AI services accept, and in some cases require, an access token. Currently, these services support access tokens:

- Text Translation API
- Speech Services: Speech to text API
- Speech Services: Text to speech API

## ⚠️ Warning

The services that support access tokens may change over time, please check the API reference for a service before using this authentication method.

Both single service and multi-service resource keys can be exchanged for authentication tokens. Authentication tokens are valid for 10 minutes. They're stored in JSON Web Token (JWT) format and can be queried programmatically using the [JWT libraries](#).

Access tokens are included in a request as the `Authorization` header. The token value provided must be preceded by `Bearer`, for example: `Bearer YOUR_AUTH_TOKEN`.

## Sample requests

Use this URL to exchange a resource key for an access token: `https://YOUR-REGION.api.cognitive.microsoft.com/sts/v1.0/issueToken`.

### cURL

```
curl -v -X POST \
"https://YOUR-REGION.api.cognitive.microsoft.com/sts/v1.0/issueToken" \
-H "Content-type: application/x-www-form-urlencoded" \
-H "Content-length: 0" \
-H "Ocp-Apim-Subscription-Key: YOUR_SUBSCRIPTION_KEY"
```

These multi-service regions support token exchange:

- `australiaeast`
- `brazilsouth`
- `canadacentral`
- `centralindia`
- `eastasia`
- `eastus`

- `japaneast`
- `northeurope`
- `southcentralus`
- `southeastasia`
- `uksouth`
- `westcentralus`
- `westeurope`
- `westus`
- `westus2`

After you get an access token, you'll need to pass it in each request as the `Authorization` header. This is a sample call to the Azure AI Translator service:

#### CURL

```
curl -X POST 'https://api.cognitive.microsofttranslator.com/translate?api-version=3.0&from=en&to=de' \
-H 'Authorization: Bearer YOUR_AUTH_TOKEN' \
-H 'Content-Type: application/json' \
--data-binary '[{"text": "How much for the cup of coffee?"}]' | json_pp
```

## Authenticate with Microsoft Entra ID

### ⓘ Important

Microsoft Entra authentication always needs to be used together with custom subdomain name of your Azure resource. [Regional endpoints](#) do not support Microsoft Entra authentication.

In the previous sections, we showed you how to authenticate against Azure AI services using a single-service or multi-service subscription key. While these keys provide a quick and easy path to start development, they fall short in more complex scenarios that require Azure [role-based access control \(Azure RBAC\)](#). Let's take a look at what's required to authenticate using Microsoft Entra ID.

In the following sections, you'll use either the Azure Cloud Shell environment or the Azure CLI to create a subdomain, assign roles, and obtain a bearer token to call the Azure AI services. If you get stuck, links are provided in each section with all available options for each command in Azure Cloud Shell/Azure CLI.

## ⓘ Important

If your organization is doing authentication through Microsoft Entra ID, you should [disable local authentication](#) (authentication with keys) so that users in the organization must always use Microsoft Entra ID.

## Create a resource with a custom subdomain

The first step is to create a custom subdomain. If you want to use an existing Azure AI services resource which does not have custom subdomain name, follow the instructions in [Azure AI services custom subdomains](#) to enable custom subdomain for your resource.

1. Start by opening the Azure Cloud Shell. Then [select a subscription](#):

```
PowerShell
```

```
Set-AzContext -SubscriptionName <SubscriptionName>
```

2. Next, [create an Azure AI services resource](#) with a custom subdomain. The subdomain name needs to be globally unique and cannot include special characters, such as: ".", "!", ",", ".":

```
PowerShell
```

```
$account = New-AzCognitiveServicesAccount -ResourceGroupName
<RESOURCE_GROUP_NAME> -name <ACCOUNT_NAME> -Type <ACCOUNT_TYPE> -
SkuName <SUBSCRIPTION_TYPE> -Location <REGION> -CustomSubdomainName
<UNIQUE_SUBDOMAIN>
```

3. If successful, the **Endpoint** should show the subdomain name unique to your resource.

## Assign a role to a service principal

Now that you have a custom subdomain associated with your resource, you're going to need to assign a role to a service principal.

## ⓘ Note

Keep in mind that Azure role assignments may take up to five minutes to propagate.

1. First, let's register an [Microsoft Entra application](#).

```
PowerShell
```

```
$SecureStringPassword = ConvertTo-SecureString -String <YOUR_PASSWORD>
-AsPlainText -Force

$app = New-AzureADApplication -DisplayName <APP_DISPLAY_NAME> -
IdentifierUris <APP_URIS> -PasswordCredentials $SecureStringPassword
```

You're going to need the **ApplicationId** in the next step.

2. Next, you need to [create a service principal](#) for the Microsoft Entra application.

```
PowerShell
```

```
New-AzADServicePrincipal -ApplicationId <APPLICATION_ID>
```

ⓘ Note

If you register an application in the Azure portal, this step is completed for you.

3. The last step is to [assign the "Cognitive Services User" role](#) to the service principal (scoped to the resource). By assigning a role, you're granting service principal access to this resource. You can grant the same service principal access to multiple resources in your subscription.

ⓘ Note

The ObjectId of the service principal is used, not the ObjectId for the application. The ACCOUNT\_ID will be the Azure resource Id of the Azure AI services account you created. You can find Azure resource Id from "properties" of the resource in Azure portal.

```
Azure CLI
```

```
New-AzRoleAssignment -ObjectId <SERVICE_PRINCIPAL_OBJECTID> -Scope
<ACCOUNT_ID> -RoleDefinitionName "Cognitive Services User"
```

## Sample request

In this sample, a password is used to authenticate the service principal. The token provided is then used to call the Computer Vision API.

### 1. Get your TenantId:

```
PowerShell

$context=Get-AzContext
$context.Tenant.Id
```

### 2. Get a token:

```
PowerShell

$tenantId = $context.Tenant.Id
$clientId = $app.ApplicationId
$clientSecret = "<YOUR_PASSWORD>"
$resourceUrl = "https://cognitiveservices.azure.com/"

$tokenEndpoint =
"https://login.microsoftonline.com/$tenantId/oauth2/token"
$body = @{
 grant_type = "client_credentials"
 client_id = $clientId
 client_secret = $clientSecret
 resource = $resourceUrl
}

$responseToken = Invoke-RestMethod -Uri $tokenEndpoint -Method Post -
Body $body
$accessToken = $responseToken.access_token
```

#### ⚠ Note

Anytime you use passwords in a script, the most secure option is to use the PowerShell Secrets Management module and integrate with a solution such as Azure Key Vault.

### 3. Call the Computer Vision API:

```
PowerShell

$url = $account.Endpoint+"vision/v1.0/models"
$result = Invoke-RestMethod -Uri $url -Method Get -Headers
@{"Authorization"="Bearer $accessToken"} -Verbose
$result | ConvertTo-Json
```

Alternatively, the service principal can be authenticated with a certificate. Besides service principal, user principal is also supported by having permissions delegated through another Microsoft Entra application. In this case, instead of passwords or certificates, users would be prompted for two-factor authentication when acquiring token.

## Authorize access to managed identities

Azure AI services support Microsoft Entra authentication with [managed identities for Azure resources](#). Managed identities for Azure resources can authorize access to Azure AI services resources using Microsoft Entra credentials from applications running in Azure virtual machines (VMs), function apps, virtual machine scale sets, and other services. By using managed identities for Azure resources together with Microsoft Entra authentication, you can avoid storing credentials with your applications that run in the cloud.

### Enable managed identities on a VM

Before you can use managed identities for Azure resources to authorize access to Azure AI services resources from your VM, you must enable managed identities for Azure resources on the VM. To learn how to enable managed identities for Azure Resources, see:

- [Azure portal](#)
- [Azure PowerShell](#)
- [Azure CLI](#)
- [Azure Resource Manager template](#)
- [Azure Resource Manager client libraries](#)

For more information about managed identities, see [Managed identities for Azure resources](#).

## Use Azure key vault to securely access credentials

You can [use Azure Key Vault](#) to securely develop Azure AI services applications. Key Vault enables you to store your authentication credentials in the cloud, and reduces the chances that secrets may be accidentally leaked, because you won't store security information in your application.

Authentication is done via Microsoft Entra ID. Authorization may be done via Azure role-based access control (Azure RBAC) or Key Vault access policy. Azure RBAC can be used for both management of the vaults and access data stored in a vault, while key vault access policy can only be used when attempting to access data stored in a vault.

## See also

- [What are Azure AI services?](#)
  - [Azure AI services pricing ↗](#)
  - [Custom subdomains](#)
- 

## Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

# Disable local authentication in Azure AI Services

Article • 09/25/2024

Azure AI Services provides Microsoft Entra authentication support for all resources. This feature provides you with seamless integration when you require centralized control and management of identities and resource credentials. Organizations can disable local authentication methods and enforce Microsoft Entra authentication instead.

## How to disable local authentication

You can disable local authentication using the Azure policy **Azure AI Services resources should have key access disabled (disable local authentication)**. Set it at the subscription level or resource group level to enforce the policy for a group of services.

If you're creating an account using Bicep / ARM template, you can set the property `disableLocalAuth` to `true` to disable local authentication. For more information, see [Microsoft.CognitiveServices accounts - Bicep, ARM template, & Terraform](#)

You can also use PowerShell with the Azure CLI to disable local authentication for an individual resource. First sign in with the `Connect-AzAccount` command. Then use the `Set-AzCognitiveServicesAccount` cmdlet with the parameter `-DisableLocalAuth $true`, like the following example:

PowerShell

```
Set-AzCognitiveServicesAccount -ResourceGroupName "my-resource-group" -Name "my-resource-name" -DisableLocalAuth $false
```

## Verify local authentication status

Disabling local authentication doesn't take effect immediately. Allow a few minutes for the service to block future authentication requests.

You can use PowerShell to determine whether the local authentication policy is currently enabled. First sign in with the `Connect-AzAccount` command. Then use the cmdlet [`Get-AzCognitiveServicesAccount`](#) to retrieve your resource, and check the property `DisableLocalAuth`. A value of `true` means local authentication is disabled.

# Re-enable local authentication

To enable local authentication, execute the PowerShell cmdlet `Set-AzCognitiveServicesAccount` with the parameter `-DisableLocalAuth $false`. Allow a few minutes for the service to accept the change to allow local authentication requests.

## Next steps

- [Authenticate requests to Azure AI services](#)
- 

## Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

# Develop Azure AI services applications with Azure Key Vault

Article • 08/28/2024

Learn how to develop Azure AI services applications securely by using [Azure Key Vault](#).

Key Vault reduces the risk that secrets may be accidentally leaked, because you avoid storing security information in your application.

## Prerequisites

- A valid Azure subscription - [Create one for free ↗](#)
- [Visual Studio IDE ↗](#)
- An [Azure Key Vault](#)
- An [Azure AI services resource](#)

### Note

Review the documentation and quickstart articles for the Azure AI service you're using to get an understanding of:

- The credentials and other information you will need to send API calls.
- The packages and code you will need to run your application.

## Get your credentials from your Azure AI services resource

Before you add your credential information to your Azure key vault, you need to retrieve them from your Azure AI services resource. For example, if your service needs a key and endpoint you would find them using the following steps:

1. Navigate to your Azure resource in the [Azure portal ↗](#).
2. From the collapsible menu on the left, select **Keys and Endpoint**.

The screenshot shows the 'MyDemoResource | Keys and Endpoint' page in the Azure portal. On the left, a sidebar lists various service management options like Overview, Activity log, Access control (IAM), Tags, and Diagnose and solve problems. Under Resource Management, the 'Keys and Endpoint' option is highlighted with a red box. The main content area displays two keys: 'KEY 1' and 'KEY 2', each with a copy icon. Below them is the 'Endpoint' field containing the URL 'https://mydemoresource.cognitiveservices.azure.com/' (also highlighted with a red box). The 'Location' field shows 'westus'. A large search icon is located in the bottom right corner of the main content area.

Some Azure AI services require different information to authenticate API calls, such as a key and region. Make sure to retrieve this information before continuing on.

## Add your credentials to your key vault

For your application to retrieve and use your credentials to authenticate API calls, you will need to add them to your [key vault secrets](#).

Repeat these steps to generate a secret for each required resource credential. For example, a key and endpoint. These secret names will be used later to authenticate your application.

1. Open a new browser tab or window. Navigate to your key vault in the [Azure portal](#).
2. From the collapsible menu on the left, select **Objects > Secrets**.
3. Select **Generate/Import**.

The screenshot shows the Azure Key Vault interface. At the top, there's a navigation bar with 'Home > my-key-vault'. Below it is a search bar and a 'Generate/Import' button, which is highlighted with a red box. To the right of the search bar are 'Refresh', 'Restore Backup', 'View sample code', and 'Manage deleted secrets' buttons. On the left, there's a sidebar with links like 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', 'Access policies', 'Events', and 'Objects' (with 'Keys', 'Secrets', and 'Certificates' listed). The main area shows a table with two rows: 'endpoint' (Type: key, Status: Enabled) and 'key' (Type: key, Status: Enabled). A large circular 'Add' button with a plus sign is located on the right side of the table.

4. On the **Create a secret** screen, enter the following values:

[Expand table](#)

| Name           | Value                                                                                                      |
|----------------|------------------------------------------------------------------------------------------------------------|
| Upload options | Manual                                                                                                     |
| Name           | A secret name for your key or endpoint. For example: "CognitiveServicesKey" or "CognitiveServicesEndpoint" |
| Value          | Your Azure AI services resource key or endpoint.                                                           |

Later your application will use the secret "Name" to securely access the "Value".

5. Leave the other values as their defaults. Select **Create**.

**Tip**

Make sure to remember the names that you set for your secrets, as you'll use them later in your application.

You should now have named secrets for your resource information.

## Create an environment variable for your key vault's name

We recommend creating an environment variable for your Azure key vault's name. Your application will read this environment variable at runtime to retrieve your key and

endpoint information.

To set environment variables, use one the following commands. `KEY_VAULT_NAME` with the name of the environment variable, and replace `Your-Key-Vault-Name` with the name of your key vault, which will be stored in the environment variable.

#### Azure CLI

Create and assign persisted environment variable, given the value.

#### CMD

```
setx KEY_VAULT_NAME "Your-Key-Vault-Name"
```

In a new instance of the **Command Prompt**, read the environment variable.

#### CMD

```
echo %KEY_VAULT_NAME%
```

## Authenticate to Azure using Visual Studio

Developers using Visual Studio 2017 or later can authenticate a Microsoft Entra account through Visual Studio. This enables you to access secrets in your key vault by signing into your Azure subscription from within the IDE.

To authenticate in Visual Studio, select **Tools** from the top navigation menu, and select **Options**. Navigate to the **Azure Service Authentication** option to sign in with your user name and password.

## Authenticate using the command line

Before you can grant access to your key vault, you must authenticate with your Microsoft Entra user name and password.

#### Azure CLI

To authenticate with the [Azure CLI](#), run the `az login` command.

#### Azure CLI

```
az login
```

On systems with a default web browser, the Azure CLI will launch the browser to authenticate. For systems without a default web browser, the `az login` command will use the device code authentication flow. You can also force the Azure CLI to use the device code flow rather than launching a browser by specifying the `--use-device-code` argument.

If you have multiple subscriptions, make sure to [select the Azure subscription](#) that contains your key vault.

## Grant access to your key vault

Create an access policy for your key vault that grants secret permissions to your user account.

Azure CLI

To set the access policy, run the `az keyvault set-policy` command. Replace `Your-Key-Vault-Name` with the name of your key vault. Replace `user@domain.com` with your Microsoft Entra user name.

Azure CLI

```
az keyvault set-policy --name Your-Key-Vault-Name --upn user@domain.com
--secret-permissions delete get list set purge
```

## Create a new C# application

Using the Visual Studio IDE, create a new .NET Core console app. This will create a "Hello World" project with a single C# source file: `program.cs`.

Install the following client libraries by right-clicking on the solution in the **Solution Explorer** and selecting **Manage NuGet Packages**. In the package manager that opens select **Browse** and search for the following libraries, and select **Install** for each:

- `Azure.Security.KeyVault.Secrets`
- `Azure.Identity`

# Import the example code

Copy the following example code into your `program.cs` file. Replace `Your-Key-Secret-Name` and `Your-Endpoint-Secret-Name` with the secret names that you set in your key vault.

C#

```
using System;
using System.Threading.Tasks;
using Azure;
using Azure.Identity;
using Azure.Security.KeyVault.Secrets;
using System.Net;

namespace key_vault_console_app
{
 class Program
 {
 static async Task Main(string[] args)
 {
 //Name of your key vault
 var keyVaultName =
Environment.GetEnvironmentVariable("KEY_VAULT_NAME");

 //variables for retrieving the key and endpoint from your key
 //vault.

 //Set these variables to the names you created for your secrets
 const string keySecretName = "Your-Key-Secret-Name";
 const string endpointSecretName = "Your-Endpoint-Secret-Name";

 //Endpoint for accessing your key vault
 var kvUri = $"https://'{keyVaultName}'.vault.azure.net";

 var keyVaultClient = new SecretClient(new Uri(kvUri), new
DefaultAzureCredential());

 Console.WriteLine($"Retrieving your secrets from
{keyVaultName}.");

 //Key and endpoint secrets retrieved from your key vault
 var keySecret = await
keyVaultClient.GetSecretAsync(keySecretName);
 var endpointSecret = await
keyVaultClient.GetSecretAsync(endpointSecretName);
 Console.WriteLine($"Your key secret value is:
{keySecret.Value.Value}");
 Console.WriteLine($"Your endpoint secret value is:
{endpointSecret.Value.Value}");
 Console.WriteLine("Secrets retrieved successfully");

 }
 }
}
```

```
 }
}
```

## Run the application

Run the application by selecting the **Debug** button at the top of Visual Studio. Your key and endpoint secrets will be retrieved from your key vault.

## Send a test Language service call (optional)

If you're using a multi-service resource or Language resource, you can update [your application](#) by following these steps to send an example Named Entity Recognition call by retrieving a key and endpoint from your key vault.

1. Install the `Azure.AI.TextAnalytics` library by right-clicking on the solution in the **Solution Explorer** and selecting **Manage NuGet Packages**. In the package manager that opens select **Browse** and search for the following libraries, and select **Install** for each:
  - Microsoft.Azure.CognitiveServices.Language.TextAnalytics
  - Microsoft.Azure.CognitiveServices.Core
2. Add the following directive to the top of your `program.cs` file.

```
C#

using Azure.AI.TextAnalytics;
```

3. Add the following code sample to your application.

```
C#

// Example method for extracting named entities from text
private static void EntityRecognitionExample(string keySecret, string endpointSecret)
{
 //String to be sent for Named Entity Recognition
 var exampleString = "I had a wonderful trip to Seattle last week.";

 AzureKeyCredential azureKeyCredential = new
 AzureKeyCredential(keySecret);
 Uri endpoint = new Uri(endpointSecret);
 var languageServiceClient = new TextAnalyticsClient(endpoint,
 azureKeyCredential);

 Console.WriteLine($"Sending a Named Entity Recognition (NER)
request");
 var response =
```

```
languageServiceClient.RecognizeEntities(exampleString);
 Console.WriteLine("Named Entities:");
 foreach (var entity in response.Value)
 {
 Console.WriteLine($"\\tText: {entity.Text}, \\tCategory:
{entity.Category}, \\tSub-Category: {entity.SubCategory}");
 Console.WriteLine($"\\t\\tScore:
{entity.ConfidenceScore:F2}, \\tLength: {entity.Length}, \\tOffset:
{entity.Offset}\\n");
 }
}
```

4. Add the following code to call `EntityRecognitionExample()` from your main method, with your key and endpoint values.

C#

```
EntityRecognitionExample(keySecret.Value.Value,
endpointSecret.Value.Value);
```

5. Run the application.

## Next steps

- See [What are Azure AI services](#) for available features you can develop along with [Azure Key Vault](#).
- For additional information on secure application development, see:
  - [Best practices for using Azure Key Vault](#)
  - [Azure AI services security](#)
  - [Azure security baseline for Azure AI services](#)

---

## Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

# Rotate keys in Azure AI services

Article • 08/28/2024

Each Azure AI services resource has two API keys to enable secret rotation. This is a security precaution that lets you regularly change the keys that can access your service, protecting the privacy of your resource if a key gets leaked.

## How to rotate keys

You can rotate keys using the following procedure:

1. If you're using both keys in production, change your code so that only one key is in use. In this guide, assume it's key 1.

This is a necessary step because once a key is regenerated, the older version of that key stops working immediately. This would cause clients using the older key to get `401 access denied` errors.

2. Once you have only key 1 in use, you can regenerate key 2. Go to your resource's page on the Azure portal, select the **Keys and Endpoint** tab, and select the **Regenerate Key 2** button at the top of the page.
3. Next, update your code to use the newly generated key 2.

It helps to have logs or availability to check that users of the key have successfully swapped from using key 1 to key 2 before you proceed.

4. Now you can regenerate key 1 using the same process.
5. Finally, update your code to use the new key 1.

## See also

- [What are Azure AI services?](#)
- [Azure AI services security features](#)

## Feedback

Was this page helpful?

 Yes

 No



# Use environment variables with Azure AI services

Article • 08/28/2024

This guide shows you how to set and retrieve environment variables for your Azure AI services credentials when you test applications.

## Important

If you use an API key, store it securely somewhere else, such as in [Azure Key Vault](#). Don't include the API key directly in your code, and never post it publicly.

For more information about AI services security, see [Authenticate requests to Azure AI services](#).

## Set an environment variable

To set environment variables, use one the following commands, where the `ENVIRONMENT_VARIABLE_KEY` is the named key and `value` is the value stored in the environment variable.

### Command Line

Use the following command to create and assign a persisted environment variable, given the input value.

#### CMD

```
:: Assigns the env var to the value
setx ENVIRONMENT_VARIABLE_KEY "value"
```

In a new instance of the Command Prompt, use the following command to read the environment variable.

#### CMD

```
:: Prints the env var value
echo %ENVIRONMENT_VARIABLE_KEY%
```

## 💡 Tip

After you set an environment variable, restart your integrated development environment (IDE) to ensure that the newly added environment variables are available.

# Retrieve an environment variable

To use an environment variable in your code, it must be read into memory. Use one of the following code snippets, depending on which language you're using. These code snippets demonstrate how to get an environment variable given the `ENVIRONMENT_VARIABLE_KEY` and assign the value to a program variable named `value`.

C#

For more information, see [Environment.GetEnvironmentVariable](#) .

C#

```
using static System.Environment;

class Program
{
 static void Main()
 {
 // Get the named env var, and assign it to the value variable
 var value =
 GetEnvironmentVariable(
 "ENVIRONMENT_VARIABLE_KEY");
 }
}
```

## Next steps

- Explore [Azure AI services](#) and choose a service to get started.

## Feedback

Was this page helpful?

Yes

No



# Customer-managed keys for encryption

Article • 08/28/2024

Azure AI is built on top of multiple Azure services. While the data is stored securely using encryption keys that Microsoft provides, you can enhance security by providing your own (customer-managed) keys. The keys you provide are stored securely using Azure Key Vault.

## Prerequisites

- An Azure subscription.
- An Azure Key Vault instance. The key vault contains the key(s) used to encrypt your services.
  - The key vault instance must enable soft delete and purge protection.
  - The managed identity for the services secured by a customer-managed key must have the following permissions in key vault:
    - wrap key
    - unwrap key
    - get

For example, the managed identity for Azure Cosmos DB would need to have those permissions to the key vault.

## How metadata is stored

The following services are used by Azure AI to store metadata for your Azure AI resource and projects:

 Expand table

| Service         | What it's used for                                                 | Example                                                       |
|-----------------|--------------------------------------------------------------------|---------------------------------------------------------------|
| Azure Cosmos DB | Stores metadata for your Azure AI projects and tools               | Flow creation timestamps, deployment tags, evaluation metrics |
| Azure AI Search | Stores indices that are used to help query your AI studio content. | An index based off your model deployment names                |
| Azure Storage   | Stores artifacts created by Azure AI                               | Fine-tuned models                                             |

| Service | What it's used for | Example |
|---------|--------------------|---------|
| Account | projects and tools |         |

All of the above services are encrypted using the same key at the time that you create your Azure AI resource for the first time, and are set up in a managed resource group in your subscription once for every Azure AI resource and set of projects associated with it. Your Azure AI resource and projects read and write data using managed identity. Managed identities are granted access to the resources using a role assignment (Azure role-based access control) on the data resources. The encryption key you provide is used to encrypt data that is stored on Microsoft-managed resources. It's also used to create indices for Azure AI Search, which are created at runtime.

## Customer-managed keys

When you don't use a customer-managed key, Microsoft creates and manages these resources in a Microsoft owned Azure subscription and uses a Microsoft-managed key to encrypt the data.

When you use a customer-managed key, these resources are *in your Azure subscription* and encrypted with your key. While they exist in your subscription, these resources are managed by Microsoft. They're automatically created and configured when you create your Azure AI resource.

### ⓘ Important

When using a customer-managed key, the costs for your subscription will be higher because these resources are in your subscription. To estimate the cost, use the [Azure pricing calculator](#).

These Microsoft-managed resources are located in a new Azure resource group is created in your subscription. This group is in addition to the resource group for your project. This resource group contains the Microsoft-managed resources that your key is used with. The resource group is named using the formula of <Azure AI resource group name><GUID>. It isn't possible to change the naming of the resources in this managed resource group.

### 💡 Tip

- The [Request Units](#) for the Azure Cosmos DB automatically scale as needed.

- If your AI resource uses a private endpoint, this resource group will also contain a Microsoft-managed Azure Virtual Network. This VNet is used to secure communications between the managed services and the project. You cannot provide your own VNet for use with the Microsoft-managed resources. You also cannot modify the virtual network. For example, you cannot change the IP address range that it uses.

### Important

If your subscription does not have enough quota for these services, a failure will occur.

### Warning

Don't delete the managed resource group that contains this Azure Cosmos DB instance, or any of the resources automatically created in this group. If you need to delete the resource group or Microsoft-managed services in it, you must delete the Azure AI resources that uses it. The resource group resources are deleted when the associated AI resource is deleted.

The process to enable Customer-Managed Keys with Azure Key Vault for Azure AI services varies by product. Use these links for service-specific instructions:

- [Azure OpenAI encryption of data at rest](#)
- [Custom Vision encryption of data at rest](#)
- [Face Services encryption of data at rest](#)
- [Document Intelligence encryption of data at rest](#)
- [Translator encryption of data at rest](#)
- [Language service encryption of data at rest](#)
- [Speech encryption of data at rest](#)
- [Content Moderator encryption of data at rest](#)
- [Personalizer encryption of data at rest](#)

## How compute data is stored

Azure AI uses compute resources for compute instance and serverless compute when you fine-tune models or build flows. The following table describes the compute options and how data is encrypted by each one:

| Compute            | Encryption                                                                                   |
|--------------------|----------------------------------------------------------------------------------------------|
| Compute instance   | Local scratch disk is encrypted.                                                             |
| Serverless compute | OS disk encrypted in Azure Storage with Microsoft-managed keys. Temporary disk is encrypted. |

**Compute instance** The OS disk for compute instance is encrypted with Microsoft-managed keys in Microsoft-managed storage accounts. If the project was created with the `hbi_workspace` parameter set to `TRUE`, the local temporary disk on compute instance is encrypted with Microsoft managed keys. Customer managed key encryption isn't supported for OS and temp disk.

**Serverless compute** The OS disk for each compute node stored in Azure Storage is encrypted with Microsoft-managed keys. This compute target is ephemeral, and clusters are typically scaled down when no jobs are queued. The underlying virtual machine is de-provisioned, and the OS disk is deleted. Azure Disk Encryption isn't supported for the OS disk.

Each virtual machine also has a local temporary disk for OS operations. If you want, you can use the disk to stage training data. This environment is short-lived (only during your job) and encryption support is limited to system-managed keys only.

## Limitations

- Encryption keys don't pass down from the Azure AI resource to dependent resources including Azure AI Services and Azure Storage when configured on the Azure AI resource. You must set encryption specifically on each resource.
- The customer-managed key for encryption can only be updated to keys in the same Azure Key Vault instance.
- After deployment, you can't switch from Microsoft-managed keys to Customer-managed keys or vice versa.
- Resources that are created in the Microsoft-managed Azure resource group in your subscription can't be modified by you or be provided by you at the time of creation as existing resources.
- You can't delete Microsoft-managed resources used for customer-managed keys without also deleting your project.

## Next steps

- Azure AI services Customer-Managed Key Request Form [↗](#) is still required for Speech and Content Moderator.
  - [What is Azure Key Vault?](#)
- 

## Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

# Configure Azure AI services virtual networks

Article • 08/28/2024

Azure AI services provide a layered security model. This model enables you to secure your Azure AI services accounts to a specific subset of networks. When network rules are configured, only applications that request data over the specified set of networks can access the account. You can limit access to your resources with *request filtering*, which allows requests that originate only from specified IP addresses, IP ranges, or from a list of subnets in [Azure Virtual Networks](#).

An application that accesses an Azure AI services resource when network rules are in effect requires authorization. Authorization is supported with [Microsoft Entra ID](#) credentials or with a valid API key.

## Important

Turning on firewall rules for your Azure AI services account blocks incoming requests for data by default. To allow requests through, one of the following conditions needs to be met:

- The request originates from a service that operates within an Azure Virtual Network on the allowed subnet list of the target Azure AI services account. The endpoint request that originated from the virtual network needs to be set as the [custom subdomain](#) of your Azure AI services account.
- The request originates from an allowed list of IP addresses.

Requests that are blocked include those from other Azure services, from the Azure portal, and from logging and metrics services.

## Note

We recommend that you use the Azure Az PowerShell module to interact with Azure. To get started, see [Install Azure PowerShell](#). To learn how to migrate to the Az PowerShell module, see [Migrate Azure PowerShell from AzureRM to Az](#).

## Scenarios

To secure your Azure AI services resource, you should first configure a rule to deny access to traffic from all networks, including internet traffic, by default. Then, configure rules that grant access to traffic from specific virtual networks. This configuration enables you to build a secure network boundary for your applications. You can also configure rules to grant access to traffic from select public internet IP address ranges and enable connections from specific internet or on-premises clients.

Network rules are enforced on all network protocols to Azure AI services, including REST and WebSocket. To access data by using tools such as the Azure test consoles, explicit network rules must be configured. You can apply network rules to existing Azure AI services resources, or when you create new Azure AI services resources. After network rules are applied, they're enforced for all requests.

## Supported regions and service offerings

Virtual networks are supported in [regions where Azure AI services are available](#). Azure AI services support service tags for network rules configuration. The services listed here are included in the `CognitiveServicesManagement` service tag.

- ✓ Anomaly Detector
- ✓ Azure OpenAI
- ✓ Content Moderator
- ✓ Custom Vision
- ✓ Face
- ✓ Language Understanding (LUIS)
- ✓ Personalizer
- ✓ Speech service
- ✓ Language
- ✓ QnA Maker
- ✓ Translator

### Note

If you use Azure OpenAI, LUIS, Speech Services, or Language services, the `CognitiveServicesManagement` tag only enables you to use the service by using the SDK or REST API. To access and use Azure OpenAI Studio, LUIS portal, Speech Studio, or Language Studio from a virtual network, you need to use the following tags:

- `AzureActiveDirectory`
- `AzureFrontDoor.Frontend`

- `AzureResourceManager`
- `CognitiveServicesManagement`
- `CognitiveServicesFrontEnd`
- `Storage` (Speech Studio only)

For information on configuring Azure AI Studio, see the [Azure AI Studio documentation](#).

## Change the default network access rule

By default, Azure AI services resources accept connections from clients on any network. To limit access to selected networks, you must first change the default action.

### Warning

Making changes to network rules can impact your applications' ability to connect to Azure AI services. Setting the default network rule to *deny* blocks all access to the data unless specific network rules that *grant* access are also applied.

Before you change the default rule to deny access, be sure to grant access to any allowed networks by using network rules. If you allow listing for the IP addresses for your on-premises network, be sure to add all possible outgoing public IP addresses from your on-premises network.

## Manage default network access rules

You can manage default network access rules for Azure AI services resources through the Azure portal, PowerShell, or the Azure CLI.

Azure portal

1. Go to the Azure AI services resource you want to secure.
2. Select **Resource Management** to expand it, then select **Networking**.

Home > contoso-rg > contoso-custom-vision

**contoso-custom-vision** | Networking

Custom vision | Directory: Microsoft

Firewalls and virtual networks Private endpoint connections

Save Discard Refresh

Access control settings allowing access to Azure AI services account will remain in effect for up to three minutes after saving updated settings restricting access.

Allow access from

All networks  Selected Networks and Private Endpoints  Disabled

Configure network security for your Azure AI services account. [Learn more](#).

Virtual networks

Secure your Azure AI services account with virtual networks. + Add existing virtual network + Add new virtual network

| Virtual Network      | Subnet | Address range | Endpoint Status | Resource group | Subscription |
|----------------------|--------|---------------|-----------------|----------------|--------------|
| No network selected. |        |               |                 |                |              |

Firewall

Add IP ranges to allow access from the internet or your on-premises networks. [Learn more](#).

Add your client IP address

Address range

IP address or CIDR

3. To deny access by default, under **Firewalls and virtual networks**, select **Selected Networks and Private Endpoints**.

With this setting alone, unaccompanied by configured virtual networks or address ranges, all access is effectively denied. When all access is denied, requests that attempt to consume the Azure AI services resource aren't permitted. The Azure portal, Azure PowerShell, or the Azure CLI can still be used to configure the Azure AI services resource.

4. To allow traffic from all networks, select **All networks**.

Home > contoso-rg > contoso-custom-vision

**contoso-custom-vision** | Networking

Custom vision | Directory: Microsoft

Firewalls and virtual networks Private endpoint connections

Save Discard Refresh

Allow access from

All networks  Selected Networks and Private Endpoints  Disabled

All networks, including the internet, can access this resource. [Learn more](#).

Resource Management

Keys and Endpoint

Encryption

Pricing tier

Networking

Identity

5. Select **Save** to apply your changes.

## Grant access from a virtual network

You can configure Azure AI services resources to allow access from specific subnets only. The allowed subnets might belong to a virtual network in the same subscription or in a different subscription. The other subscription can belong to a different Microsoft Entra tenant. When the subnet belongs to a different subscription, the Microsoft.CognitiveServices resource provider needs to be also registered for that subscription.

Enable a *service endpoint* for Azure AI services within the virtual network. The service endpoint routes traffic from the virtual network through an optimal path to the Azure AI service. For more information, see [Virtual Network service endpoints](#).

The identities of the subnet and the virtual network are also transmitted with each request. Administrators can then configure network rules for the Azure AI services resource to allow requests from specific subnets in a virtual network. Clients granted access by these network rules must continue to meet the authorization requirements of the Azure AI services resource to access the data.

Each Azure AI services resource supports up to 100 virtual network rules, which can be combined with IP network rules. For more information, see [Grant access from an internet IP range](#) later in this article.

## Set required permissions

To apply a virtual network rule to an Azure AI services resource, you need the appropriate permissions for the subnets to add. The required permission is the default *Contributor* role or the *Cognitive Services Contributor* role. Required permissions can also be added to custom role definitions.

The Azure AI services resource and the virtual networks that are granted access might be in different subscriptions, including subscriptions that are part of a different Microsoft Entra tenant.

### Note

Configuration of rules that grant access to subnets in virtual networks that are a part of a different Microsoft Entra tenant are currently supported only through PowerShell, the Azure CLI, and the REST APIs. You can view these rules in the Azure portal, but you can't configure them.

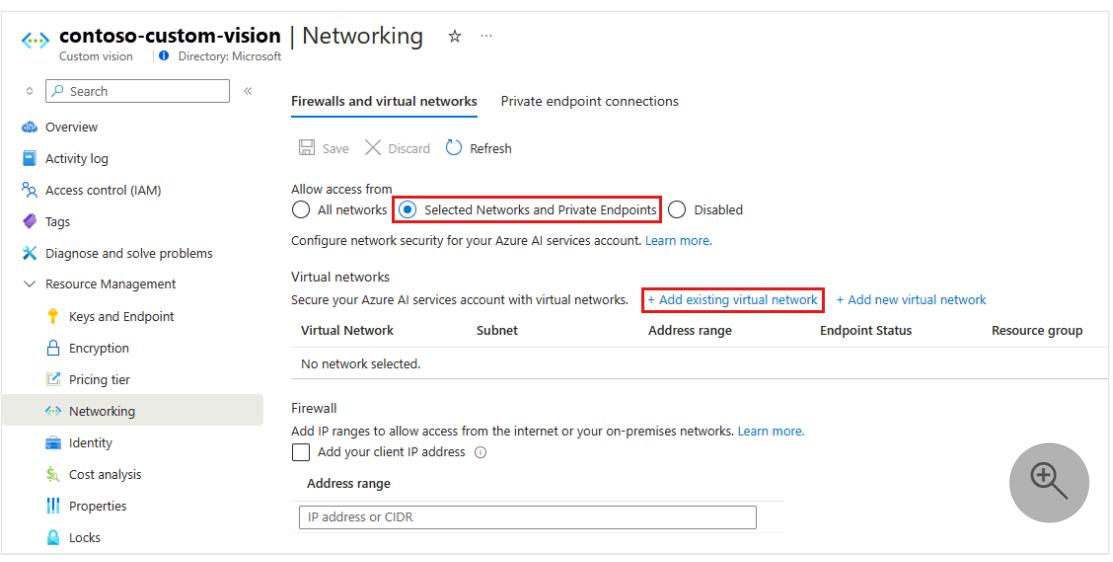
## Configure virtual network rules

You can manage virtual network rules for Azure AI services resources through the Azure portal, PowerShell, or the Azure CLI.

Azure portal

To grant access to a virtual network with an existing network rule:

1. Go to the Azure AI services resource you want to secure.
2. Select **Resource Management** to expand it, then select **Networking**.
3. Confirm that you selected **Selected Networks and Private Endpoints**.
4. Under **Allow access from**, select **Add existing virtual network**.



5. Select the **Virtual networks** and **Subnets** options, and then select **Enable**.

## Add networks

X

Subscription \*

Contoso Subscription

Virtual networks \*

contoso-rg

Subnets \*

default (Service endpoint required)

**i** The following networks don't have service endpoints enabled for 'Microsoft.CognitiveServices'. Enabling access will take up to 15 minutes to complete. After starting this operation, it is safe to leave and return later if you do not wish to wait.

| Virtual network | Service endpoint status |     |
|-----------------|-------------------------|-----|
| contoso-rg      | Not enabled             | ... |
| default         | Not enabled             | ... |

Enable

### ⚠ Note

If a service endpoint for Azure AI services wasn't previously configured for the selected virtual network and subnets, you can configure it as part of this operation.

Currently, only virtual networks that belong to the same Microsoft Entra tenant are available for selection during rule creation. To grant access to a subnet in a virtual network that belongs to another tenant, use PowerShell, the Azure CLI, or the REST APIs.

6. Select **Save** to apply your changes.

To create a new virtual network and grant it access:

1. On the same page as the previous procedure, select **Add new virtual network**.

The screenshot shows the Azure portal interface for managing a Custom Vision service named "contoso-custom-vision". The left sidebar lists various management options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Resource Management (Keys and Endpoint, Encryption, Pricing tier), Networking (Identity, Cost analysis, Properties, Locks), and Firewall. The "Networking" section is currently selected. The main content area displays settings for firewalls and virtual networks. Under "Allow access from", the radio button for "Selected Networks and Private Endpoints" is selected. Below this, there's a note about securing the Azure AI services account with virtual networks and buttons for "+ Add existing virtual network" and "+ Add new virtual network", both of which are highlighted with a red box. A table below shows columns for Virtual Network, Subnet, Address range, Endpoint Status, and Resource group, with a message stating "No network selected.". The bottom right of the blade features a circular icon with a plus sign and a magnifying glass.

2. Provide the information necessary to create the new virtual network, and then select **Create**.

**Create virtual network**

\* Name  
widgets-vnet

\* Address space ⓘ  
10.1.0.0/16  
10.1.0.0 - 10.1.255.255 (65536 addresses)

\* Subscription  
widgets-subscription

\* Resource group  
widgets-resource-group  
[Create new](#)

\* Location  
(US) West US 2

Subnet

\* Name  
default

\* Address range ⓘ  
10.1.0.0/24  
10.1.0.0 - 10.1.0.255 (256 addresses)

DDoS protection ⓘ  
 Basic  Standard

Service endpoint ⓘ  
Microsoft.CognitiveServices

Firewall ⓘ  
 Disabled  Enabled

**Create**

3. Select **Save** to apply your changes.

To remove a virtual network or subnet rule:

1. On the same page as the previous procedures, select ...(**More options**) to open the context menu for the virtual network or subnet, and select **Remove**.

The screenshot shows the 'Firewalls and virtual networks' blade in the Azure portal. At the top, there are buttons for 'Save', 'Discard', and 'Refresh'. Below that, there's a section for 'Allow access from' with three options: 'All networks' (radio button), 'Selected Networks and Private Endpoints' (selected radio button), and 'Disabled'. A note says 'Configure network security for your Azure AI services account. [Learn more](#)'. Under 'Virtual networks', there's a table with columns 'Virtual Network', 'Subnet', 'Address range', 'Endpoint Status', 'Resource group', and 'Subscription'. One row is shown for 'contoso-01-vnet' with subnet '1'. The 'Subscription' column shows 'contoso-rg'. To the right of the table are two buttons: a red-bordered 'Remove' button and a '...' button. Below the table is a 'Firewall' section with a note about adding IP ranges. It includes a checkbox for 'Add your client IP address' and a 'Remove' button. There's also a search icon with a magnifying glass.

2. Select **Save** to apply your changes.

### ⓘ Important

Be sure to [set the default rule](#) to *deny*, or network rules have no effect.

## Grant access from an internet IP range

You can configure Azure AI services resources to allow access from specific public internet IP address ranges. This configuration grants access to specific services and on-premises networks, which effectively block general internet traffic.

You can specify the allowed internet address ranges by using [CIDR format \(RFC 4632\)](#) in the form `192.168.0.0/16` or as individual IP addresses like `192.168.0.1`.

### 💡 Tip

Small address ranges that use `/31` or `/32` prefix sizes aren't supported. Configure these ranges by using individual IP address rules.

IP network rules are only allowed for *public internet* IP addresses. IP address ranges reserved for private networks aren't allowed in IP rules. Private networks include addresses that start with `10.*`, `172.16.*` - `172.31.*`, and `192.168.*`. For more information, see [Private Address Space \(RFC 1918\)](#).

Currently, only IPv4 addresses are supported. Each Azure AI services resource supports up to 100 IP network rules, which can be combined with [virtual network rules](#).

# Configure access from on-premises networks

To grant access from your on-premises networks to your Azure AI services resource with an IP network rule, identify the internet-facing IP addresses used by your network. Contact your network administrator for help.

If you use Azure ExpressRoute on-premises for Microsoft peering, you need to identify the NAT IP addresses. For more information, see [What is Azure ExpressRoute](#).

For Microsoft peering, the NAT IP addresses that are used are either customer provided or supplied by the service provider. To allow access to your service resources, you must allow these public IP addresses in the resource IP firewall setting.

## Managing IP network rules

You can manage IP network rules for Azure AI services resources through the Azure portal, PowerShell, or the Azure CLI.

Azure portal

1. Go to the Azure AI services resource you want to secure.
2. Select **Resource Management** to expand it, then select **Networking**.
3. Confirm that you selected **Selected Networks and Private Endpoints**.
4. Under **Firewalls and virtual networks**, locate the **Address range** option. To grant access to an internet IP range, enter the IP address or address range (in [CIDR format](#)). Only valid public IP (nonreserved) addresses are accepted.

[Firewalls and virtual networks](#)   [Private endpoint connections](#)

[Save](#)  [Discard](#)  [Refresh](#)

Allow access from:

All networks  Selected Networks and Private Endpoints  Disabled

Configure network security for your Azure AI services account. [Learn more](#).

Virtual networks

Secure your Azure AI services account with virtual networks. [+ Add existing virtual network](#) [+ Add new virtual network](#)

| Virtual Network      | Subnet | Address range | Endpoint Status | Resource group |
|----------------------|--------|---------------|-----------------|----------------|
| No network selected. |        |               |                 |                |

Firewall

Add IP ranges to allow access from the internet or your on-premises networks. [Learn more](#).

Add your client IP address [\(i\)](#)

Address range

To remove an IP network rule, select the trash can icon next to the address range.

5. Select **Save** to apply your changes.

**Important**

Be sure to [set the default rule](#) to *deny*, or network rules have no effect.

## Use private endpoints

You can use [private endpoints](#) for your Azure AI services resources to allow clients on a virtual network to securely access data over [Azure Private Link](#). The private endpoint uses an IP address from the virtual network address space for your Azure AI services resource. Network traffic between the clients on the virtual network and the resource traverses the virtual network and a private link on the Microsoft Azure backbone network, which eliminates exposure from the public internet.

Private endpoints for Azure AI services resources let you:

- Secure your Azure AI services resource by configuring the firewall to block all connections on the public endpoint for the Azure AI service.
- Increase security for the virtual network, by enabling you to block exfiltration of data from the virtual network.

- Securely connect to Azure AI services resources from on-premises networks that connect to the virtual network by using [Azure VPN Gateway](#) or [ExpressRoutes](#) with private-peering.

## Understand private endpoints

A private endpoint is a special network interface for an Azure resource in your [virtual network](#). Creating a private endpoint for your Azure AI services resource provides secure connectivity between clients in your virtual network and your resource. The private endpoint is assigned an IP address from the IP address range of your virtual network. The connection between the private endpoint and the Azure AI service uses a secure private link.

Applications in the virtual network can connect to the service over the private endpoint seamlessly. Connections use the same connection strings and authorization mechanisms that they would use otherwise. The exception is Speech Services, which require a separate endpoint. For more information, see [Private endpoints with the Speech Services](#) in this article. Private endpoints can be used with all protocols supported by the Azure AI services resource, including REST.

Private endpoints can be created in subnets that use service endpoints. Clients in a subnet can connect to one Azure AI services resource using private endpoint, while using service endpoints to access others. For more information, see [Virtual Network service endpoints](#).

When you create a private endpoint for an Azure AI services resource in your virtual network, Azure sends a consent request for approval to the Azure AI services resource owner. If the user who requests the creation of the private endpoint is also an owner of the resource, this consent request is automatically approved.

Azure AI services resource owners can manage consent requests and the private endpoints through the **Private endpoint connection** tab for the Azure AI services resource in the [Azure portal](#) ↗.

## Specify private endpoints

When you create a private endpoint, specify the Azure AI services resource that it connects to. For more information on creating a private endpoint, see:

- [Create a private endpoint by using the Azure portal](#)
- [Create a private endpoint by using Azure PowerShell](#)
- [Create a private endpoint by using the Azure CLI](#)

# Connect to private endpoints

## ⓘ Note

Azure OpenAI Service uses a different private DNS zone and public DNS zone forwarder than other Azure AI services. For the correct zone and forwarder names, see [Azure services DNS zone configuration](#).

Clients on a virtual network that use the private endpoint use the same connection string for the Azure AI services resource as clients connecting to the public endpoint. The exception is the Speech service, which requires a separate endpoint. For more information, see [Use private endpoints with the Speech service](#) in this article. DNS resolution automatically routes the connections from the virtual network to the Azure AI services resource over a private link.

By default, Azure creates a [private DNS zone](#) attached to the virtual network with the necessary updates for the private endpoints. If you use your own DNS server, you might need to make more changes to your DNS configuration. For updates that might be required for private endpoints, see [Apply DNS changes for private endpoints](#) in this article.

## Use private endpoints with the Speech service

See [Use Speech service through a private endpoint](#).

## Apply DNS changes for private endpoints

When you create a private endpoint, the DNS `CNAME` resource record for the Azure AI services resource is updated to an alias in a subdomain with the prefix `privatelink`. By default, Azure also creates a private DNS zone that corresponds to the `privatelink` subdomain, with the DNS A resource records for the private endpoints. For more information, see [What is Azure Private DNS](#).

When you resolve the endpoint URL from outside the virtual network with the private endpoint, it resolves to the public endpoint of the Azure AI services resource. When it's resolved from the virtual network hosting the private endpoint, the endpoint URL resolves to the private endpoint's IP address.

This approach enables access to the Azure AI services resource using the same connection string for clients in the virtual network that hosts the private endpoints and clients outside the virtual network.

If you use a custom DNS server on your network, clients must be able to resolve the fully qualified domain name (FQDN) for the Azure AI services resource endpoint to the private endpoint IP address. Configure your DNS server to delegate your private link subdomain to the private DNS zone for the virtual network.

### 💡 Tip

When you use a custom or on-premises DNS server, you should configure your DNS server to resolve the Azure AI services resource name in the `privatelink` subdomain to the private endpoint IP address. Delegate the `privatelink` subdomain to the private DNS zone of the virtual network. Alternatively, configure the DNS zone on your DNS server and add the DNS A records.

For more information on configuring your own DNS server to support private endpoints, see the following resources:

- [Name resolution that uses your own DNS server](#)
- [DNS configuration](#)

## Grant access to trusted Azure services for Azure OpenAI

You can grant a subset of trusted Azure services access to Azure OpenAI, while maintaining network rules for other apps. These trusted services will then use managed identity to authenticate your Azure OpenAI service. The following table lists the services that can access Azure OpenAI if the managed identity of those services have the appropriate role assignment.

 [Expand table](#)

| Service                | Resource provider name                         |
|------------------------|------------------------------------------------|
| Azure AI Services      | <code>Microsoft.CognitiveServices</code>       |
| Azure Machine Learning | <code>Microsoft.MachineLearningServices</code> |
| Azure AI Search        | <code>Microsoft.Search</code>                  |

You can grant networking access to trusted Azure services by creating a network rule exception using the REST API or Azure portal:

# Using the Azure CLI

```
Bash

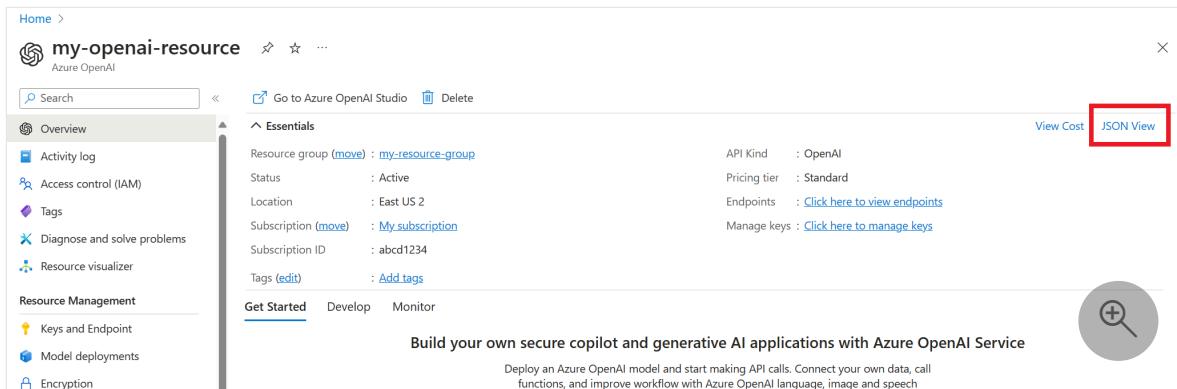
accessToken=$(az account get-access-token --resource
https://management.azure.com --query "accessToken" --output tsv)
rid="/subscriptions/<your subscription id>/resourceGroups/<your resource
group>/providers/Microsoft.CognitiveServices/accounts/<your Azure AI
resource name>"

curl -i -X PATCH https://management.azure.com$rid?api-version=2023-10-01-
preview \
-H "Content-Type: application/json" \
-H "Authorization: Bearer $accessToken" \
-d \
'
{
 "properties": {
 "networkAcls": {
 "bypass": "AzureServices"
 }
 }
}
'
```

To revoke the exception, set `networkAcls.bypass` to `None`.

To verify if the trusted service has been enabled from the Azure portal,

## 1. Use the JSON View from the Azure OpenAI resource overview page



The screenshot shows the Azure portal interface for an Azure OpenAI resource named "my-openai-resource". The left sidebar has sections for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Resource visualizer, Resource Management (Keys and Endpoint, Model deployments, Encryption), and Get Started. The main content area displays the "Essentials" section with details like Resource group (my-resource-group), Status (Active), Location (East US 2), Subscription (My subscription), Subscription ID (abcd1234), and Tags (Add tags). To the right of the essentials, there are fields for API Kind (OpenAI), Pricing tier (Standard), Endpoints (Click here to view endpoints), and Manage keys (Click here to manage keys). At the top right, there are "View Cost" and "JSON View" buttons, with "JSON View" being highlighted by a red box. Below the essentials, there is a call-to-action: "Build your own secure copilot and generative AI applications with Azure OpenAI Service" and a note: "Deploy an Azure OpenAI model and start making API calls. Connect your own data, call functions, and improve workflow with Azure OpenAI language, image and speech".

## 2. Choose your latest API version under API versions. Only the latest API version is supported, `2023-10-01-preview`.

## Resource JSON

Resource ID

|                 |                  |                      |
|-----------------|------------------|----------------------|
| /subscriptions/ | /resourceGroups/ | /providers/Microsoft |
|-----------------|------------------|----------------------|

API Versions

2023-10-01-preview

```
75 "networkAcls": {
76 "bypass": "AzureServices",
77 "defaultAction": "Deny",
78 "virtualNetworkRules": [],
79 "ipRules": []
80 },
```

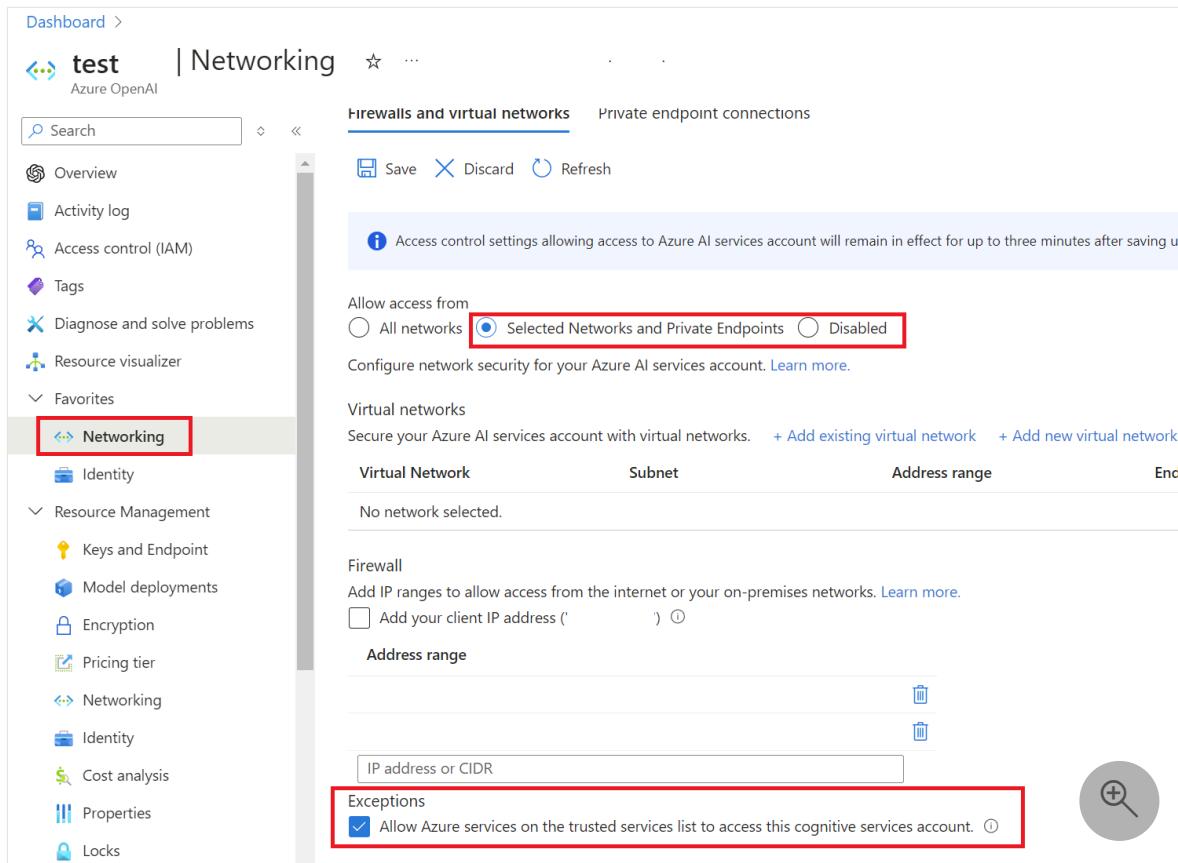


## Using the Azure portal

1. Navigate to your Azure OpenAI resource, and select **Networking** from the navigation menu.
2. Under **Exceptions**, select **Allow Azure services on the trusted services list to access this cognitive services account**.

### Tip

You can view the **Exceptions** option by selecting either **Selected networks and private endpoints** or **Disabled** under **Allow access from**.



The screenshot shows the Azure portal interface for an Azure OpenAI resource named 'test'. The left sidebar has 'Networking' selected. The main content area is titled 'Networking' and shows the 'Firewalls and virtual networks' tab is active. A note says 'Access control settings allowing access to Azure AI services account will remain in effect for up to three minutes after saving'. Below it, 'Allow access from' has 'Selected Networks and Private Endpoints' selected. The 'Exceptions' section at the bottom is highlighted with a red box, showing the checkbox 'Allow Azure services on the trusted services list to access this cognitive services account' is checked.

# Pricing

For pricing details, see [Azure Private Link pricing](#).

## Next steps

- Explore the various [Azure AI services](#)
- Learn more about [Virtual Network service endpoints](#)

---

## Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

# Configure data loss prevention for Azure AI services

Article • 08/28/2024

Azure AI services data loss prevention capabilities allow customers to configure the list of outbound URLs their Azure AI services resources are allowed to access. This creates another level of control for customers to prevent data loss. In this article, we'll cover the steps required to enable the data loss prevention feature for Azure AI services resources.

## Prerequisites

Before you make a request, you need an Azure account and an Azure AI services subscription. If you already have an account, go ahead and skip to the next section. If you don't have an account, we have a guide to get you set up in minutes: [Create an Azure AI services multi-service resource](#).

## Enabling data loss prevention

There are two parts to enable data loss prevention. First, the resource property `restrictOutboundNetworkAccess` must be set to `true`. When this is set to true, you also need to provide the list of approved URLs. The list of URLs is added to the `allowedFqdnList` property. The `allowedFqdnList` property contains an array of comma-separated URLs.

### ⓘ Note

- The `allowedFqdnList` property value supports a maximum of 1000 URLs.
- The property supports both IP addresses and fully qualified domain names i.e., `www.microsoft.com`, values.
- It can take up to 15 minutes for the updated list to take effect.

### Azure CLI

1. Install the [Azure CLI](#) and [sign in](#), or select [Try it](#).
2. View the details of the Azure AI services resource.

#### Azure CLI

```
az cognitiveservices account show \
-g "myresourcegroup" -n "myaccount" \
```

3. View the current properties of the Azure AI services resource.

#### Azure CLI

```
az rest -m get \
-u /subscriptions/{subscription ID}/resourceGroups/{resource
group}/providers/Microsoft.CognitiveServices/accounts/{account
name}?api-version=2021-04-30 \
```

4. Configure the restrictOutboundNetworkAccess property and update the allowed FqdnList with the approved URLs

#### Azure CLI

```
az rest -m patch \
-u /subscriptions/{subscription ID}/resourceGroups/{resource
group}/providers/Microsoft.CognitiveServices/accounts/{account
name}?api-version=2021-04-30 \
-b '{"properties": { "restrictOutboundNetworkAccess": true,
"allowedFqdnList": ["microsoft.com"] }}'
```

## Supported services

The following services support data loss prevention configuration:

- Azure OpenAI
- Azure AI Vision
- Content Moderator
- Custom Vision
- Face
- Document Intelligence
- Speech Service
- QnA Maker

## Next steps

- [Configure Virtual Networks](#)

---

# Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

# Azure security baseline for Cognitive Services

Article • 12/12/2023

This security baseline applies guidance from the [Microsoft cloud security benchmark version 1.0](#) to Cognitive Services. The Microsoft cloud security benchmark provides recommendations on how you can secure your cloud solutions on Azure. The content is grouped by the security controls defined by the Microsoft cloud security benchmark and the related guidance applicable to Cognitive Services.

You can monitor this security baseline and its recommendations using Microsoft Defender for Cloud. Azure Policy definitions will be listed in the Regulatory Compliance section of the Microsoft Defender for Cloud portal page.

When a feature has relevant Azure Policy Definitions, they are listed in this baseline to help you measure compliance with the Microsoft cloud security benchmark controls and recommendations. Some recommendations may require a paid Microsoft Defender plan to enable certain security scenarios.

## ⓘ Note

Features not applicable to Cognitive Services have been excluded. To see how Cognitive Services completely maps to the Microsoft cloud security benchmark, see the [full Cognitive Services security baseline mapping file ↗](#).

## Security profile

The security profile summarizes high-impact behaviors of Cognitive Services, which may result in increased security considerations.

ⓘ [Expand table](#)

| Service Behavior Attribute                              | Value     |
|---------------------------------------------------------|-----------|
| Product Category                                        | AI+ML     |
| Customer can access HOST / OS                           | No Access |
| Service can be deployed into customer's virtual network | False     |
| Stores customer content at rest                         | True      |

# Network security

For more information, see the [Microsoft cloud security benchmark: Network security](#).

## NS-1: Establish network segmentation boundaries

### Features

#### Virtual Network Integration

**Description:** Service supports deployment into customer's private Virtual Network (VNet). [Learn more](#).

[+] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| False     | Not Applicable     | Not Applicable               |

**Configuration Guidance:** This feature is not supported to secure this service.

#### Network Security Group Support

**Description:** Service network traffic respects Network Security Groups rule assignment on its subnets. [Learn more](#).

[+] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| False     | Not Applicable     | Not Applicable               |

**Feature notes:** Though Network Security Groups for this service is not supported, a service-level firewall may be configured. For more information, please visit: [Managing IP network rules](#)

**Configuration Guidance:** This feature is not supported to secure this service.

## NS-2: Secure cloud services with network controls

### Features

## Azure Private Link

**Description:** Service native IP filtering capability for filtering network traffic (not to be confused with NSG or Azure Firewall). [Learn more](#).

[+] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** Deploy private endpoints for all Azure resources that support the Private Link feature, to establish a private access point for the resources.

**Reference:** [Use private endpoints](#)

## Disable Public Network Access

**Description:** Service supports disabling public network access either through using service-level IP ACL filtering rule (not NSG or Azure Firewall) or using a 'Disable Public Network Access' toggle switch. [Learn more](#).

[+] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** Disable public network access either using the service-level IP ACL filtering rule or a toggling switch for public network access.

**Reference:** [Change the default network access rule](#)

## Microsoft Defender for Cloud monitoring

Azure Policy built-in definitions - Microsoft.CognitiveServices:

[+] [Expand table](#)

| Name<br>(Azure portal)      | Description                                                                                                                                                               | Effect(s)                   | Version<br>(GitHub)     |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|-------------------------|
| Cognitive Services accounts | To improve the security of Cognitive Services accounts, ensure that it isn't exposed to the public internet and can only be accessed from a private endpoint. Disable the | Audit,<br>Deny,<br>Disabled | <a href="#">3.0.1 ↗</a> |

| Name<br>(Azure portal)                                                         | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Effect(s)                   | Version<br>(GitHub)     |
|--------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|-------------------------|
| should<br>Name<br>disable<br>(Azure portal)<br>public<br>network<br>access ↗   | public network access property as described in <a href="https://go.microsoft.com/fwlink/?linkid=2129800">https://go.microsoft.com/fwlink/?linkid=2129800</a> . This option disables access from any public address space outside the Azure IP range, and denies all logins that match IP or virtual network-based firewall rules. This reduces data leakage risks.                                                                                                                                               |                             |                         |
| Cognitive<br>Services<br>accounts<br>should<br>restrict<br>network<br>access ↗ | Network access to Cognitive Services accounts should be restricted. Configure network rules so only applications from allowed networks can access the Cognitive Services account. To allow connections from specific internet or on-premises clients, access can be granted to traffic from specific Azure virtual networks or to public internet IP address ranges.                                                                                                                                             | Audit,<br>Deny,<br>Disabled | <a href="#">3.0.0 ↗</a> |
| Cognitive<br>Services<br>should use<br>private link ↗                          | Azure Private Link lets you connect your virtual networks to Azure services without a public IP address at the source or destination. The Private Link platform handles the connectivity between the consumer and services over the Azure backbone network. By mapping private endpoints to Cognitive Services, you'll reduce the potential for data leakage. Learn more about private links at: <a href="https://go.microsoft.com/fwlink/?linkid=2129800">https://go.microsoft.com/fwlink/?linkid=2129800</a> . | Audit,<br>Disabled          | <a href="#">3.0.0 ↗</a> |

## Identity management

For more information, see the [Microsoft cloud security benchmark: Identity management](#).

### IM-1: Use centralized identity and authentication system

#### Features

##### Azure AD Authentication Required for Data Plane Access

**Description:** Service supports using Azure AD authentication for data plane access.

[Learn more.](#)

Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** Use Azure Active Directory (Azure AD) as the default authentication method to control your data plane access.

**Reference:** [Authenticate with Azure Active Directory](#)

## Local Authentication Methods for Data Plane Access

**Description:** Local authentications methods supported for data plane access, such as a local username and password. [Learn more](#).

[+] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Feature notes:** While you can authenticate against Azure Cognitive Services using a single-service or multi-service subscription key, or use those keys to authenticate with access tokens, these authentication methods fall short in more complex scenarios that require Azure role-based access control (Azure RBAC). Avoid the usage of local authentication methods or accounts, these should be disabled wherever possible. Instead use Azure AD to authenticate where possible.

**Configuration Guidance:** Restrict the use of local authentication methods for data plane access. Instead, use Azure Active Directory (Azure AD) as the default authentication method to control your data plane access.

**Reference:** [Authenticate with an access token](#)

## Microsoft Defender for Cloud monitoring

Azure Policy built-in definitions - Microsoft.CognitiveServices:

[+] Expand table

| Name<br>(Azure portal)                                                                        | Description                                                                                                                                                                                                                                                  | Effect(s)                   | Version<br>(GitHub) |
|-----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|---------------------|
| <a href="#">Cognitive Services accounts should have local authentication methods disabled</a> | Disabling local authentication methods improves security by ensuring that Cognitive Services accounts require Azure Active Directory identities exclusively for authentication. Learn more at: <a href="https://aka.ms/cs/auth">https://aka.ms/cs/auth</a> . | Audit,<br>Deny,<br>Disabled | 1.0.0 ↗             |

## IM-3: Manage application identities securely and automatically

### Features

#### Managed Identities

**Description:** Data plane actions support authentication using managed identities. [Learn more.](#)

[+] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** Use Azure managed identities instead of service principals when possible, which can authenticate to Azure services and resources that support Azure Active Directory (Azure AD) authentication. Managed identity credentials are fully managed, rotated, and protected by the platform, avoiding hard-coded credentials in source code or configuration files.

**Reference:** [Authorize access to managed identities](#)

#### Service Principals

**Description:** Data plane supports authentication using service principals. [Learn more.](#)

[+] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** There is no current Microsoft guidance for this feature configuration. Please review and determine if your organization wants to configure this security feature.

**Reference:** [Authenticate requests to Azure Cognitive Services](#)

## IM-7: Restrict resource access based on conditions

## Features

### Conditional Access for Data Plane

**Description:** Data plane access can be controlled using Azure AD Conditional Access Policies. [Learn more.](#)

[+] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** Define the applicable conditions and criteria for Azure Active Directory (Azure AD) conditional access in the workload. Consider common use cases such as blocking or granting access from specific locations, blocking risky sign-in behavior, or requiring organization-managed devices for specific applications.

### IM-8: Restrict the exposure of credential and secrets

## Features

### Service Credential and Secrets Support Integration and Storage in Azure Key Vault

**Description:** Data plane supports native use of Azure Key Vault for credential and secrets store. [Learn more.](#)

[+] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** Ensure that secrets and credentials are stored in secure locations such as Azure Key Vault, instead of embedding them into code or configuration files.

**Reference:** [Develop Azure Cognitive Services applications with Key Vault](#)

## Privileged access

For more information, see the [Microsoft cloud security benchmark: Privileged access](#).

## PA-1: Separate and limit highly privileged/administrative users

### Features

#### Local Admin Accounts

Description: Service has the concept of a local administrative account. [Learn more](#).

[ ] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| False     | Not Applicable     | Not Applicable               |

Configuration Guidance: This feature is not supported to secure this service.

## PA-7: Follow just enough administration (least privilege) principle

### Features

#### Azure RBAC for Data Plane

Description: Azure Role-Based Access Control (Azure RBAC) can be used to manage access to service's data plane actions. [Learn more](#).

[ ] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

Configuration Guidance: Use Azure role-based access control (Azure RBAC) to manage Azure resource access through built-in role assignments. Azure RBAC roles can be assigned to users, groups, service principals, and managed identities.

Reference: [Authenticate with Azure Active Directory](#)

## PA-8: Determine access process for cloud provider support

### Features

#### Customer Lockbox

Description: Customer Lockbox can be used for Microsoft support access. [Learn more.](#)

 Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

Configuration Guidance: In support scenarios where Microsoft needs to access your data, use Customer Lockbox to review, then approve or reject each of Microsoft's data access requests.

## Data protection

For more information, see the [Microsoft cloud security benchmark: Data protection](#).

### DP-1: Discover, classify, and label sensitive data

#### Features

#### Sensitive Data Discovery and Classification

Description: Tools (such as Azure Purview or Azure Information Protection) can be used for data discovery and classification in the service. [Learn more.](#)

 Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| False     | Not Applicable     | Not Applicable               |

Configuration Guidance: This feature is not supported to secure this service.

## DP-2: Monitor anomalies and threats targeting sensitive data

### Features

#### Data Leakage/Loss Prevention

**Description:** Service supports DLP solution to monitor sensitive data movement (in customer's content). [Learn more.](#)

  [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** Cognitive Services data loss prevention capabilities allow customers to configure the list of outbound URLs their Cognitive Services resources are allowed to access. This creates another level of control for customers to prevent data loss.

**Reference:** [Configure data loss prevention for Azure Cognitive Services](#)

## DP-3: Encrypt sensitive data in transit

### Features

#### Data in Transit Encryption

**Description:** Service supports data in-transit encryption for data plane. [Learn more.](#)

  [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | True               | Microsoft                    |

**Configuration Guidance:** No additional configurations are required as this is enabled on a default deployment.

**Reference:** [Azure Cognitive Services security](#)

## DP-4: Enable data at rest encryption by default

### Features

#### Data at Rest Encryption Using Platform Keys

**Description:** Data at-rest encryption using platform keys is supported, any customer content at rest is encrypted with these Microsoft managed keys. [Learn more.](#)

[+] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | True               | Microsoft                    |

**Configuration Guidance:** No additional configurations are required as this is enabled on a default deployment.

**Reference:** [Configure customer-managed keys with Azure Key Vault for Cognitive Services](#)

## DP-5: Use customer-managed key option in data at rest encryption when required

### Features

#### Data at Rest Encryption Using CMK

**Description:** Data at-rest encryption using customer-managed keys is supported for customer content stored by the service. [Learn more.](#)

[+] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** If required for regulatory compliance, define the use case and service scope where encryption using customer-managed keys are needed. Enable and implement data at rest encryption using customer-managed key for those services.

Reference: [Configure customer-managed keys with Azure Key Vault for Cognitive Services](#)

## Microsoft Defender for Cloud monitoring

Azure Policy built-in definitions - Microsoft.CognitiveServices:

[+] Expand table

| Name<br>(Azure portal)                                                                                | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Effect(s)             | Version<br>(GitHub)   |
|-------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|-----------------------|
| <a href="#">Cognitive Services accounts should enable data encryption with a customer-managed key</a> | Customer-managed keys are commonly required to meet regulatory compliance standards. Customer-managed keys enable the data stored in Cognitive Services to be encrypted with an Azure Key Vault key created and owned by you. You have full control and responsibility for the key lifecycle, including rotation and management. Learn more about customer-managed keys at <a href="https://go.microsoft.com/fwlink/?linkid=2121321">https://go.microsoft.com/fwlink/?linkid=2121321</a> . | Audit, Deny, Disabled | <a href="#">2.1.0</a> |

## DP-6: Use a secure key management process

### Features

#### Key Management in Azure Key Vault

Description: The service supports Azure Key Vault integration for any customer keys, secrets, or certificates. [Learn more](#).

[+] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

Configuration Guidance: Use Azure Key Vault to create and control the life cycle of your encryption keys, including key generation, distribution, and storage. Rotate and revoke your keys in Azure Key Vault and your service based on a defined schedule or when there is a key retirement or compromise. When there is a need to use customer-managed key (CMK) in the workload, service, or application level, ensure you follow the best practices for key management: Use a key hierarchy to generate a separate data

encryption key (DEK) with your key encryption key (KEK) in your key vault. Ensure keys are registered with Azure Key Vault and referenced via key IDs from the service or application. If you need to bring your own key (BYOK) to the service (such as importing HSM-protected keys from your on-premises HSMs into Azure Key Vault), follow recommended guidelines to perform initial key generation and key transfer.

**Reference:** [Configure customer-managed keys with Azure Key Vault for Cognitive Services](#)

## DP-7: Use a secure certificate management process

### Features

#### Certificate Management in Azure Key Vault

**Description:** The service supports Azure Key Vault integration for any customer certificates. [Learn more.](#)

[ ] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| False     | Not Applicable     | Not Applicable               |

**Configuration Guidance:** This feature is not supported to secure this service.

## Asset management

*For more information, see the [Microsoft cloud security benchmark: Asset management](#).*

## AM-2: Use only approved services

### Features

#### Azure Policy Support

**Description:** Service configurations can be monitored and enforced via Azure Policy. [Learn more.](#)

[ ] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** Use Microsoft Defender for Cloud to configure Azure Policy to audit and enforce configurations of your Azure resources. Use Azure Monitor to create alerts when there is a configuration deviation detected on the resources. Use Azure Policy [deny] and [deploy if not exists] effects to enforce secure configuration across Azure resources.

**Reference:** [Azure Policy built-in policy definitions for Azure Cognitive Services](#)

## Logging and threat detection

*For more information, see the [Microsoft cloud security benchmark: Logging and threat detection](#).*

### LT-1: Enable threat detection capabilities

#### Features

##### Microsoft Defender for Service / Product Offering

**Description:** Service has an offering-specific Microsoft Defender solution to monitor and alert on security issues. [Learn more](#).

[ ] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| False     | Not Applicable     | Not Applicable               |

**Configuration Guidance:** This feature is not supported to secure this service.

### LT-4: Enable logging for security investigation

#### Features

## Azure Resource Logs

**Description:** Service produces resource logs that can provide enhanced service-specific metrics and logging. The customer can configure these resource logs and send them to their own data sink like a storage account or log analytics workspace. [Learn more.](#)

[+] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| True      | False              | Customer                     |

**Configuration Guidance:** Enable resource logs for the service. For example, Key Vault supports additional resource logs for actions that get a secret from a key vault or and Azure SQL has resource logs that track requests to a database. The content of resource logs varies by the Azure service and resource type.

**Reference:** [Enable diagnostic logging for Azure Cognitive Services](#)

## Backup and recovery

For more information, see the [Microsoft cloud security benchmark: Backup and recovery](#).

### BR-1: Ensure regular automated backups

#### Features

##### Azure Backup

**Description:** The service can be backed up by the Azure Backup service. [Learn more.](#)

[+] Expand table

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| False     | Not Applicable     | Not Applicable               |

**Configuration Guidance:** This feature is not supported to secure this service.

#### Service Native Backup Capability

**Description:** Service supports its own native backup capability (if not using Azure Backup). [Learn more](#).

[+] [Expand table](#)

| Supported | Enabled By Default | Configuration Responsibility |
|-----------|--------------------|------------------------------|
| False     | Not Applicable     | Not Applicable               |

**Configuration Guidance:** This feature is not supported to secure this service.

## Next steps

- See the [Microsoft cloud security benchmark overview](#)
- Learn more about [Azure security baselines](#)

# Azure Policy Regulatory Compliance controls for Azure AI services

Article • 08/28/2024

[Regulatory Compliance in Azure Policy](#) provides Microsoft created and managed initiative definitions, known as *built-ins*, for the **compliance domains** and **security controls** related to different compliance standards. This page lists the **compliance domains** and **security controls** for Azure AI services. You can assign the built-ins for a **security control** individually to help make your Azure resources compliant with the specific standard.

The title of each built-in policy definition links to the policy definition in the Azure portal. Use the link in the **Policy Version** column to view the source on the [Azure Policy GitHub repo](#).

## Important

Each control is associated with one or more [Azure Policy](#) definitions. These policies might help you [assess compliance](#) with the control. However, there often isn't a one-to-one or complete match between a control and one or more policies. As such, **Compliant** in Azure Policy refers only to the policies themselves. This doesn't ensure that you're fully compliant with all requirements of a control. In addition, the compliance standard includes controls that aren't addressed by any Azure Policy definitions at this time. Therefore, compliance in Azure Policy is only a partial view of your overall compliance status. The associations between controls and Azure Policy Regulatory Compliance definitions for these compliance standards can change over time.

## CMMC Level 3

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance - CMMC Level 3](#). For more information about this compliance standard, see [Cybersecurity Maturity Model Certification \(CMMC\)](#).

 Expand table

| <b>Domain</b>                        | <b>Control ID</b> | <b>Control title</b>                                                                                                                                                                                 | <b>Policy (Azure portal)</b>                                                                                | <b>Policy version (GitHub)</b> |
|--------------------------------------|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|--------------------------------|
| Access Control                       | AC.1.001          | Limit information system access to authorized users, processes acting on behalf of authorized users, and devices (including other information systems).                                              | <a href="#">Azure AI Services resources should restrict network access ↗</a>                                | 3.2.0 ↗                        |
| Access Control                       | AC.1.002          | Limit information system access to the types of transactions and functions that authorized users are permitted to execute.                                                                           | <a href="#">Azure AI Services resources should restrict network access ↗</a>                                | 3.2.0 ↗                        |
| Access Control                       | AC.2.016          | Control the flow of CUI in accordance with approved authorizations.                                                                                                                                  | <a href="#">Azure AI Services resources should restrict network access ↗</a>                                | 3.2.0 ↗                        |
| Configuration Management             | CM.3.068          | Restrict, disable, or prevent the use of nonessential programs, functions, ports, protocols, and services.                                                                                           | <a href="#">Azure AI Services resources should restrict network access ↗</a>                                | 3.2.0 ↗                        |
| System and Communications Protection | SC.1.175          | Monitor, control, and protect communications (i.e., information transmitted or received by organizational systems) at the external boundaries and key internal boundaries of organizational systems. | <a href="#">Azure AI Services resources should restrict network access ↗</a>                                | 3.2.0 ↗                        |
| System and Communications Protection | SC.3.177          | Employ FIPS-validated cryptography when used to protect the confidentiality of CUI.                                                                                                                  | <a href="#">Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) ↗</a> | 2.2.0 ↗                        |
| System and Communications Protection | SC.3.183          | Deny network communications traffic by default and allow network communications traffic by exception (i.e., deny all, permit by exception).                                                          | <a href="#">Azure AI Services resources should restrict network access ↗</a>                                | 3.2.0 ↗                        |

# FedRAMP High

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance - FedRAMP High](#). For more information about this compliance standard, see [FedRAMP High](#).

[Expand table](#)

| Domain         | Control ID | Control title                       | Policy (Azure portal)                                                                        | Policy version (GitHub) |
|----------------|------------|-------------------------------------|----------------------------------------------------------------------------------------------|-------------------------|
| Access Control | AC-2       | Account Management                  | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| Access Control | AC-2 (1)   | Automated System Account Management | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| Access Control | AC-2 (7)   | Role-Based Schemes                  | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| Access Control | AC-3       | Access Enforcement                  | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| Access Control | AC-4       | Information Flow Enforcement        | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗      |
| Access Control | AC-4       | Information Flow Enforcement        | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                 |
| Access Control | AC-17      | Remote Access                       | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗      |
| Access Control | AC-17 (1)  | Automated Monitoring / Control      | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗      |

| <b>Domain</b>                        | <b>Control ID</b> | <b>Control title</b>                                     | <b>Policy (Azure portal)</b>                                                                 | <b>Policy version (GitHub)</b> |
|--------------------------------------|-------------------|----------------------------------------------------------|----------------------------------------------------------------------------------------------|--------------------------------|
| Identification And Authentication    | IA-2              | Identification And Authentication (Organizational Users) | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Identification And Authentication    | IA-4              | Identifier Management                                    | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| System And Communications Protection | SC-7              | Boundary Protection                                      | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| System And Communications Protection | SC-7              | Boundary Protection                                      | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                        |
| System And Communications Protection | SC-7 (3)          | Access Points                                            | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| System And Communications Protection | SC-7 (3)          | Access Points                                            | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                        |
| System And Communications Protection | SC-12             | Cryptographic Key Establishment And Management           | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) ↗  | 2.2.0 ↗                        |

## FedRAMP Moderate

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance - FedRAMP Moderate](#). For more information about this compliance standard, see [FedRAMP Moderate](#) ↗.

Expand table

| <b>Domain</b>                     | <b>Control ID</b> | <b>Control title</b>                                     | <b>Policy (Azure portal)</b>                                                                 | <b>Policy version (GitHub)</b> |
|-----------------------------------|-------------------|----------------------------------------------------------|----------------------------------------------------------------------------------------------|--------------------------------|
| Access Control                    | AC-2              | Account Management                                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Access Control                    | AC-2 (1)          | Automated System Account Management                      | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Access Control                    | AC-2 (7)          | Role-Based Schemes                                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Access Control                    | AC-3              | Access Enforcement                                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Access Control                    | AC-4              | Information Flow Enforcement                             | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Access Control                    | AC-4              | Information Flow Enforcement                             | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                        |
| Access Control                    | AC-17             | Remote Access                                            | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Access Control                    | AC-17 (1)         | Automated Monitoring / Control                           | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Identification And Authentication | IA-2              | Identification And Authentication (Organizational Users) | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |

| Domain                               | Control ID | Control title                                  | Policy (Azure portal)                                                                        | Policy version (GitHub) |
|--------------------------------------|------------|------------------------------------------------|----------------------------------------------------------------------------------------------|-------------------------|
| Identification And Authentication    | IA-4       | Identifier Management                          | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| System And Communications Protection | SC-7       | Boundary Protection                            | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗      |
| System And Communications Protection | SC-7       | Boundary Protection                            | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                 |
| System And Communications Protection | SC-7 (3)   | Access Points                                  | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗      |
| System And Communications Protection | SC-7 (3)   | Access Points                                  | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                 |
| System And Communications Protection | SC-12      | Cryptographic Key Establishment And Management | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) ↗  | 2.2.0 ↗                 |

## Microsoft cloud security benchmark

The [Microsoft cloud security benchmark](#) provides recommendations on how you can secure your cloud solutions on Azure. To see how this service completely maps to the Microsoft cloud security benchmark, see the [Azure Security Benchmark mapping files](#) ↗.

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance - Microsoft cloud security benchmark](#).

Expand table

| <b>Domain</b>                | <b>Control ID</b> | <b>Control title</b>                                                     | <b>Policy (Azure portal)</b>                                                                 | <b>Policy version (GitHub)</b> |
|------------------------------|-------------------|--------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|--------------------------------|
| Network Security             | NS-2              | Secure cloud services with network controls                              | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Network Security             | NS-2              | Secure cloud services with network controls                              | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                        |
| Network Security             | NS-2              | Secure cloud services with network controls                              | Azure AI Services resources should use Azure Private Link ↗                                  | 1.0.0 ↗                        |
| Identity Management          | IM-1              | Use centralized identity and authentication system                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Data Protection              | DP-5              | Use customer-managed key option in data at rest encryption when required | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) ↗  | 2.2.0 ↗                        |
| Logging and Threat Detection | LT-3              | Enable logging for security investigation                                | Diagnostic logs in Azure AI services resources should be enabled ↗                           | 1.0.0 ↗                        |

## NIST SP 800-171 R2

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance - NIST SP 800-171 R2](#). For more information about this compliance standard, see [NIST SP 800-171 R2 ↗](#).

Expand table

| <b>Domain</b>  | <b>Control ID</b> | <b>Control title</b>                                                              | <b>Policy (Azure portal)</b>                               | <b>Policy version (GitHub)</b> |
|----------------|-------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------|--------------------------------|
| Access Control | 3.1.1             | Limit system access to authorized users, processes acting on behalf of authorized | [Deprecated]: Cognitive Services should use private link ↗ | 3.0.1-deprecated ↗             |

| <b>Domain</b>                                 | <b>Control ID</b> | <b>Control title</b>                                                                                                               | <b>Policy (Azure portal)</b>                                                                 | <b>Policy version (GitHub)</b> |
|-----------------------------------------------|-------------------|------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|--------------------------------|
| users, and devices (including other systems). |                   |                                                                                                                                    |                                                                                              |                                |
| Access Control                                | 3.1.1             | Limit system access to authorized users, processes acting on behalf of authorized users, and devices (including other systems).    | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Access Control                                | 3.1.12            | Monitor and control remote access sessions.                                                                                        | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Access Control                                | 3.1.13            | Employ cryptographic mechanisms to protect the confidentiality of remote access sessions.                                          | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Access Control                                | 3.1.14            | Route remote access via managed access control points.                                                                             | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Access Control                                | 3.1.2             | Limit system access to the types of transactions and functions that authorized users are permitted to execute.                     | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Access Control                                | 3.1.3             | Control the flow of CUI in accordance with approved authorizations.                                                                | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Access Control                                | 3.1.3             | Control the flow of CUI in accordance with approved authorizations.                                                                | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                        |
| System and Communications Protection          | 3.13.1            | Monitor, control, and protect communications (i.e., information transmitted or received by organizational systems) at the external | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |

| <b>Domain</b>                        | <b>Control ID</b> | <b>Control title</b>                                                                                                                                                                                 | <b>Policy (Azure portal)</b>                                                                | <b>Policy version (GitHub)</b> |
|--------------------------------------|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|--------------------------------|
|                                      |                   | boundaries and key internal boundaries of organizational systems.                                                                                                                                    |                                                                                             |                                |
| System and Communications Protection | 3.13.1            | Monitor, control, and protect communications (i.e., information transmitted or received by organizational systems) at the external boundaries and key internal boundaries of organizational systems. | Azure AI Services resources should restrict network access ↗                                | 3.2.0 ↗                        |
| System and Communications Protection | 3.13.10           | Establish and manage cryptographic keys for cryptography employed in organizational systems.                                                                                                         | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) ↗ | 2.2.0 ↗                        |
| System and Communications Protection | 3.13.2            | Employ architectural designs, software development techniques, and systems engineering principles that promote effective information security within organizational systems.                         | [Deprecated]: Cognitive Services should use private link ↗                                  | 3.0.1-deprecated ↗             |
| System and Communications Protection | 3.13.2            | Employ architectural designs, software development techniques, and systems engineering principles that promote effective information security within organizational systems.                         | Azure AI Services resources should restrict network access ↗                                | 3.2.0 ↗                        |
| System and Communications Protection | 3.13.5            | Implement subnetworks for publicly accessible system components that are physically or logically separated from internal networks.                                                                   | [Deprecated]: Cognitive Services should use private link ↗                                  | 3.0.1-deprecated ↗             |
| System and Communications            | 3.13.5            | Implement subnetworks for publicly accessible                                                                                                                                                        | Azure AI Services resources should                                                          | 3.2.0 ↗                        |

| <b>Domain</b>                        | <b>Control ID</b> | <b>Control title</b>                                                                                                                        | <b>Policy (Azure portal)</b>                                                                 | <b>Policy version (GitHub)</b> |
|--------------------------------------|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|--------------------------------|
| Protection                           |                   | system components that are physically or logically separated from internal networks.                                                        | restrict network access ↗                                                                    |                                |
| System and Communications Protection | 3.13.6            | Deny network communications traffic by default and allow network communications traffic by exception (i.e., deny all, permit by exception). | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                        |
| Identification and Authentication    | 3.5.1             | Identify system users, processes acting on behalf of users, and devices.                                                                    | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Identification and Authentication    | 3.5.2             | Authenticate (or verify) the identities of users, processes, or devices, as a prerequisite to allowing access to organizational systems.    | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Identification and Authentication    | 3.5.5             | Prevent reuse of identifiers for a defined period.                                                                                          | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Identification and Authentication    | 3.5.6             | Disable identifiers after a defined period of inactivity.                                                                                   | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |

## NIST SP 800-53 Rev. 4

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance - NIST SP 800-53 Rev. 4](#). For more information about this compliance standard, see [NIST SP 800-53 Rev. 4](#).

[Expand table](#)

| Domain                            | Control ID | Control title                                            | Policy (Azure portal)                                                                        | Policy version (GitHub) |
|-----------------------------------|------------|----------------------------------------------------------|----------------------------------------------------------------------------------------------|-------------------------|
| Access Control                    | AC-2       | Account Management                                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| Access Control                    | AC-2 (1)   | Automated System Account Management                      | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| Access Control                    | AC-2 (7)   | Role-Based Schemes                                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| Access Control                    | AC-3       | Access Enforcement                                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| Access Control                    | AC-4       | Information Flow Enforcement                             | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗      |
| Access Control                    | AC-4       | Information Flow Enforcement                             | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                 |
| Access Control                    | AC-17      | Remote Access                                            | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗      |
| Access Control                    | AC-17 (1)  | Automated Monitoring / Control                           | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗      |
| Identification And Authentication | IA-2       | Identification And Authentication (Organizational Users) | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |

| <b>Domain</b>                        | <b>Control ID</b> | <b>Control title</b>                           | <b>Policy (Azure portal)</b>                                                                 | <b>Policy version (GitHub)</b> |
|--------------------------------------|-------------------|------------------------------------------------|----------------------------------------------------------------------------------------------|--------------------------------|
| Identification And Authentication    | IA-4              | Identifier Management                          | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| System And Communications Protection | SC-7              | Boundary Protection                            | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| System And Communications Protection | SC-7              | Boundary Protection                            | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                        |
| System And Communications Protection | SC-7 (3)          | Access Points                                  | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| System And Communications Protection | SC-7 (3)          | Access Points                                  | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                        |
| System And Communications Protection | SC-12             | Cryptographic Key Establishment And Management | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) ↗  | 2.2.0 ↗                        |

## NIST SP 800-53 Rev. 5

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance - NIST SP 800-53 Rev. 5](#). For more information about this compliance standard, see [NIST SP 800-53 Rev. 5](#) ↗.

[] Expand table

| <b>Domain</b>  | <b>Control ID</b> | <b>Control title</b> | <b>Policy (Azure portal)</b>                                | <b>Policy version (GitHub)</b> |
|----------------|-------------------|----------------------|-------------------------------------------------------------|--------------------------------|
| Access Control | AC-2              | Account Management   | Azure AI Services resources should have key access disabled | 1.1.0 ↗                        |

| <b>Domain</b>                     | <b>Control ID</b> | <b>Control title</b>                                     | <b>Policy (Azure portal)</b>                                                                 | <b>Policy version (GitHub)</b> |
|-----------------------------------|-------------------|----------------------------------------------------------|----------------------------------------------------------------------------------------------|--------------------------------|
|                                   |                   |                                                          | (disable local authentication) ↗                                                             |                                |
| Access Control                    | AC-2 (1)          | Automated System Account Management                      | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Access Control                    | AC-2 (7)          | Privileged User Accounts                                 | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Access Control                    | AC-3              | Access Enforcement                                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Access Control                    | AC-4              | Information Flow Enforcement                             | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Access Control                    | AC-4              | Information Flow Enforcement                             | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                        |
| Access Control                    | AC-17             | Remote Access                                            | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Access Control                    | AC-17 (1)         | Monitoring and Control                                   | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗             |
| Identification and Authentication | IA-2              | Identification and Authentication (organizational Users) | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                        |
| Identification and Authentication | IA-4              | Identifier Management                                    | Azure AI Services resources should have key access disabled                                  | 1.1.0 ↗                        |

| <b>Domain</b>                        | <b>Control ID</b> | <b>Control title</b>                           | <b>Policy (Azure portal)</b>                                                                | <b>Policy version (GitHub)</b> |
|--------------------------------------|-------------------|------------------------------------------------|---------------------------------------------------------------------------------------------|--------------------------------|
| (disable local authentication) ↗     |                   |                                                |                                                                                             |                                |
| System and Communications Protection | SC-7              | Boundary Protection                            | [Deprecated]: Cognitive Services should use private link ↗                                  | 3.0.1-deprecated ↗             |
| System and Communications Protection | SC-7              | Boundary Protection                            | Azure AI Services resources should restrict network access ↗                                | 3.2.0 ↗                        |
| System and Communications Protection | SC-7 (3)          | Access Points                                  | [Deprecated]: Cognitive Services should use private link ↗                                  | 3.0.1-deprecated ↗             |
| System and Communications Protection | SC-7 (3)          | Access Points                                  | Azure AI Services resources should restrict network access ↗                                | 3.2.0 ↗                        |
| System and Communications Protection | SC-12             | Cryptographic Key Establishment and Management | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) ↗ | 2.2.0 ↗                        |

## NL BIO Cloud Theme

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance details for NL BIO Cloud Theme](#). For more information about this compliance standard, see [Baseline Information Security Government Cybersecurity - Digital Government \(digitaleoverheid.nl\)](#) ↗.

[ ] Expand table

| <b>Domain</b>                                   | <b>Control ID</b> | <b>Control title</b>                                                                | <b>Policy (Azure portal)</b>                                                                | <b>Policy version (GitHub)</b> |
|-------------------------------------------------|-------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|--------------------------------|
| U.05.2 Data protection - Cryptographic measures | U.05.2            | Data stored in the cloud service shall be protected to the latest state of the art. | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) ↗ | 2.2.0 ↗                        |

| Domain                                            | Control ID | Control title                                                                                                                   | Policy (Azure portal)                                                                        | Policy version (GitHub) |
|---------------------------------------------------|------------|---------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|-------------------------|
| U.07.1 Data separation - Isolated                 | U.07.1     | Permanent isolation of data is a multi-tenant architecture. Patches are realized in a controlled manner.                        | [Deprecated]: Cognitive Services should use private link ↗                                   | 3.0.1-deprecated ↗      |
| U.07.1 Data separation - Isolated                 | U.07.1     | Permanent isolation of data is a multi-tenant architecture. Patches are realized in a controlled manner.                        | Azure AI Services resources should restrict network access ↗                                 | 3.2.0 ↗                 |
| U.07.3 Data separation - Management features      | U.07.3     | U.07.3 - The privileges to view or modify CSC data and/or encryption keys are granted in a controlled manner and use is logged. | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| U.10.2 Access to IT services and data - Users     | U.10.2     | Under the responsibility of the CSP, access is granted to administrators.                                                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| U.10.3 Access to IT services and data - Users     | U.10.3     | Only users with authenticated equipment can access IT services and data.                                                        | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| U.10.5 Access to IT services and data - Competent | U.10.5     | Access to IT services and data is limited by technical measures and has been implemented.                                       | Azure AI Services resources should have key access disabled (disable local authentication) ↗ | 1.1.0 ↗                 |
| U.11.3 Cryptoservices - Encrypted                 | U.11.3     | Sensitive data is always encrypted, with private keys managed by the CSC.                                                       | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) ↗  | 2.2.0 ↗                 |

# Reserve Bank of India IT Framework for Banks v2016

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance - RBI ITF Banks v2016](#). For more information about this compliance standard, see [RBI ITF Banks v2016 \(PDF\)](#).

[+] Expand table

| Domain        | Control ID | Control title      | Policy<br>(Azure portal)                                                                  | Policy version<br>(GitHub) |
|---------------|------------|--------------------|-------------------------------------------------------------------------------------------|----------------------------|
| Metrics       |            | Metrics-21.1       | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) | 2.2.0                      |
| Anti-Phishing |            | Anti-Phishing-14.1 | Azure AI Services resources should restrict network access                                | 3.2.0                      |

## System and Organization Controls (SOC) 2

To review how the available Azure Policy built-ins for all Azure services map to this compliance standard, see [Azure Policy Regulatory Compliance details for System and Organization Controls \(SOC\) 2](#). For more information about this compliance standard, see [System and Organization Controls \(SOC\) 2](#).

[+] Expand table

| Domain                               | Control ID | Control title                                                       | Policy<br>(Azure portal)                                                                  | Policy version<br>(GitHub) |
|--------------------------------------|------------|---------------------------------------------------------------------|-------------------------------------------------------------------------------------------|----------------------------|
| Logical and Physical Access Controls | CC6.1      | Logical access security software, infrastructure, and architectures | Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK) | 2.2.0                      |

## Next steps

- Learn more about [Azure Policy Regulatory Compliance](#).
- See the built-ins on the [Azure Policy GitHub repo](#).

---

# Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

# Responsible use of AI with Azure AI services

Article • 08/28/2024

Azure AI services provides information and guidelines on how to responsibly use artificial intelligence in applications. Below are the links to articles that provide this guidance for the different services within the Azure AI services suite.

## Vision

- [Azure AI Vision - Image Analysis](#)
- [Azure AI Vision - OCR](#)
- [Azure AI Vision - Face](#)
- [Azure AI Vision - Spatial Analysis](#)
- [Azure Custom Vision](#)
- [Azure Video Indexer](#)

## Language

- [Azure AI Language](#)
- [Azure AI Language - Custom text classification](#)
- [Azure AI Language - Named entity recognition](#)
- [Azure AI Language - Custom named entity recognition](#)
- [Azure AI Language - Entity linking](#)
- [Azure AI Language - Language detection](#)
- [Azure AI Language - Key phrase extraction](#)
- [Azure AI Language - Personally identifiable information detection](#)
- [Azure AI Language - Question Answering](#)
- [Azure AI Language - Sentiment Analysis and opinion mining](#)
- [Azure AI Language - Text Analytics for health](#)
- [Azure AI Language - Summarization](#)
- [Language Understanding](#)

## Speech

- [Azure AI Speech - Pronunciation Assessment](#)
- [Azure AI Speech - Speaker Recognition](#)
- [Azure AI Speech - Text to speech](#)

- Azure AI Speech - Speech to text

## Search

- Azure AI Search

## Other

- Azure OpenAI
  - Azure AI Content Safety
  - Azure AI Document Intelligence
  - Anomaly Detector
  - Personalizer
  - QnA Maker
- 

## Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

# Limited Access features for Azure AI services

Article • 08/28/2024

Our vision is to empower developers and organizations to use AI to transform society in positive ways. We encourage responsible AI practices to protect the rights and safety of individuals. To achieve this, Microsoft has implemented a Limited Access policy grounded in our [AI Principles](#) to support responsible deployment of Azure services.

## What is Limited Access?

Limited Access services require registration, and only customers managed by Microsoft—meaning those who are working directly with Microsoft account teams—are eligible for access. The use of these services is limited to the use case selected at the time of registration. Customers must acknowledge that they've reviewed and agree to the terms of service. Microsoft may require customers to reverify this information.

Limited Access services are made available to customers under the terms governing their subscription to Microsoft Azure Services (including the [Service Specific Terms](#)). Review these terms carefully as they contain important conditions and obligations governing your use of Limited Access services.

## List of Limited Access services

The following services are Limited Access:

- [Custom Neural Voice](#): Pro features
- [Custom text to speech avatar](#): All features
- [Speaker Recognition](#): All features
- [Face API](#): Identify and Verify features, face ID property
- [Azure AI Vision](#): Celebrity Recognition feature
- [Azure AI Video Indexer](#): Celebrity Recognition and Face Identify features
- [Azure OpenAI](#): Azure OpenAI Service, modified abuse monitoring, and modified content filters

Features of these services that aren't listed above are available without registration.

## FAQ about Limited Access

## How do I register for access?

Submit a registration form for each Limited Access service you would like to use:

- [Custom Neural Voice](#) : Pro features
- [Custom text to speech avatar](#) : All features
- [Speaker Recognition](#) : All features
- [Face API](#) : Identify and Verify features
- [Azure AI Vision](#) : Celebrity Recognition feature
- [Azure AI Video Indexer](#) : Celebrity Recognition and Face Identify features
- [Azure OpenAI](#): Azure OpenAI Service, modified abuse monitoring, and modified content filters

## How long will the registration process take?

You'll receive communication from us about your application within 5-10 business days. In some cases, reviews can take longer. You'll receive an email as soon as your application is reviewed.

## Who is eligible to use Limited Access services?

Limited Access services are available only to customers managed by Microsoft. Additionally, Limited Access services are only available for certain use cases, and customers must select their intended use case in their registration form.

Please use an email address affiliated with your organization in your registration form. Registration forms submitted with personal email addresses will be denied.

If you're not a managed customer, we invite you to submit an application using the same forms and we will reach out to you about any opportunities to join an eligibility program.

## What is a managed customer? What if I don't know whether I'm a managed customer?

Managed customers work with Microsoft account teams. We invite you to submit a registration form for the features you'd like to use, and we'll verify your eligibility for access. We are not able to accept requests to become a managed customer at this time.

## What happens if I'm an existing customer and I don't register?

Existing customers have until June 30, 2023 to submit a registration form and be approved to continue using Limited Access services after June 30, 2023. We recommend allowing 10 business days for review. Without an approved application, you will be denied access after June 30, 2023.

## I'm an existing customer who applied for access to Custom Neural Voice or Speaker Recognition, do I have to register to keep using these services?

We're always looking for opportunities to improve our Responsible AI program, and Limited Access is an update to our service gating processes. If you've previously applied for and been granted access to Custom Neural Voice or Speaker Recognition, we request that you submit a new registration form to continue using these services beyond June 30, 2023.

If you're an existing customer using Custom Neural Voice or Speaker Recognition on June 21, 2022, you have until June 30, 2023 to submit a registration form with your selected use case and receive approval to continue using these services after June 30, 2023. We recommend allowing 10 days for application processing. Existing customers can continue using the service until June 30, 2023, after which they must be approved for access. The registration forms can be found here:

- [Custom Neural Voice](#) : Pro features
- [Speaker Recognition](#) : All features

## What if my use case isn't on the registration form?

Limited Access features are only available for the use cases listed on the registration forms. If your desired use case isn't listed, let us know in this [feedback form](#) so we can improve our service offerings.

## Where can I use Limited Access services?

Search [here](#) for a Limited Access service to view its regional availability. In the Brazil South and UAE North datacenter regions, we are prioritizing access for commercial customers managed by Microsoft.

Detailed information about supported regions for Custom Neural Voice and Speaker Recognition operations can be found [here](#).

## What happens to my data if my application is denied?

If you're an existing customer and your application for access is denied, you will no longer be able to use Limited Access features after June 30, 2023. Your data is subject to Microsoft's data retention [policies](#).

## Help and support

Report abuse of Limited Access services [here](#).

---

## Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback](#) | [Get help at Microsoft Q&A](#)

# Azure AI services SDK reference

Article • 08/28/2024

This article provides an overview of available Azure AI client libraries and packages with links to service and feature level reference documentation.

## Available Azure AI services

Select a service from the table and learn how Azure AI can help you meet your development goals.

## Supported services

[+] Expand table

| Service                                                                                             | Description                                                   | Reference documentation                                                                                                                                  |
|-----------------------------------------------------------------------------------------------------|---------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
|  Azure AI Search | Bring AI-powered cloud search to your mobile and web apps.    | <ul style="list-style-type: none"><li>• <a href="#">Azure AI Search SDK for .NET</a></li><li>• <a href="#">Azure AI Search NuGet package ↗</a></li></ul> |
|  Azure OpenAI    | Perform a wide variety of natural language tasks.             | <ul style="list-style-type: none"><li>• <a href="#">Azure OpenAI SDK for .NET</a></li><li>• <a href="#">Azure OpenAI NuGet package ↗</a></li></ul>       |
|  Bot Service     | Create bots and connect them across channels.                 | <ul style="list-style-type: none"><li>• <a href="#">Bot service SDK for .NET ↗</a></li><li>• <a href="#">Bot Builder (NuGet package) ↗</a></li></ul>     |
|  Content Safety  | Detect harmful content in applications and services.          | <ul style="list-style-type: none"><li>• <a href="#">Content Safety SDK for .NET</a></li><li>• <a href="#">Content Safety NuGet package ↗</a></li></ul>   |
|  Custom Vision   | Customize image recognition for your applications and models. | <ul style="list-style-type: none"><li>• <a href="#">Custom Vision SDK for .NET</a></li></ul>                                                             |

| Service                                                                                                 | Description                                                                                                              | Reference documentation                                                                                                                                                                                                                                                                                                      |
|---------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                         |                                                                                                                          | <ul style="list-style-type: none"> <li>• <a href="#">Custom Vision NuGet package (prediction)</a> ↗</li> <li>• <a href="#">Custom Vision NuGet package (training)</a> ↗</li> </ul>                                                                                                                                           |
|  Document Intelligence | Turn documents into intelligent data-driven solutions.                                                                   | <ul style="list-style-type: none"> <li>• <a href="#">Document Intelligence SDK for .NET</a></li> <li>• <a href="#">Document Intelligence NuGet package</a> ↗</li> </ul>                                                                                                                                                      |
|  Face                  | Detect, recognize, and identify human faces in images.                                                                   | <ul style="list-style-type: none"> <li>• <a href="#">Face SDK for .NET</a></li> <li>• <a href="#">Face NuGet package</a> ↗</li> </ul>                                                                                                                                                                                        |
|  Immersive Reader      | Help users with text readability and comprehension.                                                                      | <ul style="list-style-type: none"> <li>• <a href="#">Immersive Reader C# quickstart</a></li> </ul>                                                                                                                                                                                                                           |
|  Language             | Build applications with natural language understanding capabilities.                                                     | <ul style="list-style-type: none"> <li>• <a href="#">Language SDK for .NET (text analysis)</a></li> <li>• <a href="#">Language NuGet package (text analysis)</a> ↗</li> <li>• <a href="#">Language SDK for .NET (Question Answering)</a></li> <li>• <a href="#">Language NuGet package (question answering)</a> ↗</li> </ul> |
|  Speech              | Add speech to text, text to speech, translation, and speaker recognition capabilities to applications.                   | <ul style="list-style-type: none"> <li>• <a href="#">Speech SDK for .NET</a></li> <li>• <a href="#">Speech NuGet package (Speech CLI)</a> ↗</li> </ul>                                                                                                                                                                       |
|  Translator          | Use AI-powered translation technology to translate more than 100 in-use, at-risk, and endangered languages and dialects. | <ul style="list-style-type: none"> <li>• <a href="#">Translator SDK for .NET (text)</a></li> <li>• <a href="#">Translator NuGet package (text)</a> ↗</li> <li>• <a href="#">Translator SDK for .NET (batch)</a></li> </ul>                                                                                                   |

| <b>Service</b>                                                                                  | <b>Description</b>                                       | <b>Reference documentation</b>                                                                                                            |
|-------------------------------------------------------------------------------------------------|----------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                 |                                                          | <ul style="list-style-type: none"> <li>• <a href="#">Translator NuGet package (batch) ↗</a></li> </ul>                                    |
|  Video Indexer | Extract actionable insights from your videos.            | <a href="#">Video Indexer C# code samples ↗</a>                                                                                           |
|  Vision        | Analyze content in digital images and rich media assets. | <ul style="list-style-type: none"> <li>• <a href="#">Vision SDK for .NET</a></li> <li>• <a href="#">Vision NuGet package ↗</a></li> </ul> |

## Deprecated services

[\[+\] Expand table](#)

| <b>Service</b>                                                                                                                         | <b>Description</b>                                             | <b>Reference documentation</b>                                                                                                                                                                                                                                                           |
|----------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  Anomaly Detector<br>(deprecated 2023)               | Monitor and detect anomalies in time series data.              | <ul style="list-style-type: none"> <li>• <a href="#">Anomaly Detector SDK for .NET</a></li> <li>• <a href="#">Anomaly Detector NuGet package ↗</a></li> </ul>                                                                                                                            |
|  Content Moderator<br>(deprecated 2024)             | Monitor and detect potentially offensive and unwanted content. | <ul style="list-style-type: none"> <li>• <a href="#">Content Moderator SDK for .NET</a></li> <li>• <a href="#">Content Moderator NuGet package ↗</a></li> </ul>                                                                                                                          |
|  Language understanding (LUIS)<br>(deprecated 2023) | Use natural language understanding in your applications.       | <ul style="list-style-type: none"> <li>• <a href="#">LUIS SDK for .NET (authoring) ↗</a></li> <li>• <a href="#">LUIS NuGet package (authoring) ↗</a></li> <li>• <a href="#">LUIS SDK for .NET (prediction) ↗</a></li> <li>• <a href="#">LUIS NuGet package (prediction) ↗</a></li> </ul> |
|  Metrics Advisor<br>(deprecated 2023)               | Detect harmful content in applications and services.           | <ul style="list-style-type: none"> <li>• <a href="#">Metrics Advisor SDK for .NET</a></li> </ul>                                                                                                                                                                                         |

| Service                                                                                                             | Description                                                                 | Reference documentation                                                                                                        |
|---------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                     |                                                                             | <ul style="list-style-type: none"><li>Metrics Advisor NuGet package <a href="#">↗</a></li></ul>                                |
|  Personalizer<br>(deprecated 2023) | Present rich, personalized content and experiences to users.                | <ul style="list-style-type: none"><li>Personalizer SDK for .NET</li><li>Personalizer NuGet package <a href="#">↗</a></li></ul> |
|  QnA maker<br>(deprecated 2022)    | Distill information into easy-to-navigate question and answer applications. | <ul style="list-style-type: none"><li>QnA maker SDK for .NET</li><li>QnA maker NuGet package <a href="#">↗</a></li></ul>       |

## Next steps

- View Azure AI REST API reference
- 

## Feedback

Was this page helpful?

 Yes

 No

Provide product feedback [↗](#) | Get help at Microsoft Q&A

# Azure AI services REST API reference

Article • 08/28/2024

This article provides an overview of available Azure AI services REST APIs with links to service and feature level reference documentation.

## Available Azure AI services

Select a service from the table to learn how it can help you meet your development goals.

## Supported services

[\[...\] Expand table](#)

| Service documentation                                                                                                     | Description                                                                  | Reference documentation                                                                                                                                                                                                             |
|---------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  <a href="#">Azure AI Search</a>       | Bring AI-powered cloud search to your mobile and web apps                    | <a href="#">Azure AI Search API</a>                                                                                                                                                                                                 |
|  <a href="#">Azure OpenAI</a>          | Perform a wide variety of natural language tasks                             | <a href="#">Azure OpenAI APIs</a> <ul style="list-style-type: none"><li>• <a href="#">resource creation &amp; deployment</a></li><li>• <a href="#">completions &amp; embeddings</a></li><li>• <a href="#">fine-tuning</a></li></ul> |
|  <a href="#">Bot Service</a>           | Create bots and connect them across channels                                 | <a href="#">Bot Service API</a>                                                                                                                                                                                                     |
|  <a href="#">Content Safety</a>        | An AI service that detects unwanted contents                                 | <a href="#">Content Safety API</a>                                                                                                                                                                                                  |
|  <a href="#">Custom Vision</a>         | Customize image recognition for your business applications.                  | <a href="#">Custom Vision APIs</a> <ul style="list-style-type: none"><li>• <a href="#">prediction</a></li><li>• <a href="#">training</a></li></ul>                                                                                  |
|  <a href="#">Document Intelligence</a> | Turn documents into intelligent data-driven solutions                        | <a href="#">Document Intelligence API</a>                                                                                                                                                                                           |
|  <a href="#">Face</a>                  | Detect and identify people and emotions in images                            | <a href="#">Face API</a>                                                                                                                                                                                                            |
|  <a href="#">Language</a>              | Build apps with industry-leading natural language understanding capabilities | <a href="#">REST API</a>                                                                                                                                                                                                            |

| <b>Service documentation</b>                                                                                    | <b>Description</b>                                                             | <b>Reference documentation</b>                                                                                  |
|-----------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
|  <a href="#">Speech</a>        | Speech to text, text to speech, translation, and speaker recognition           | <a href="#">Speech APIs</a><br>• <a href="#">speech to text</a><br>• <a href="#">text to speech</a>             |
|  <a href="#">Translator</a>    | Translate more than 100 in-use, at-risk, and endangered languages and dialects | <a href="#">Translator APIs</a><br>• <a href="#">text translation</a><br>• <a href="#">document translation</a> |
|  <a href="#">Video Indexer</a> | Extract actionable insights from your videos                                   | <a href="#">Video Indexer API</a>                                                                               |
|  <a href="#">Vision</a>        | Analyze content in images and videos                                           | <a href="#">Vision API</a>                                                                                      |

## Deprecated services

[\[ \] Expand table](#)

| <b>Service documentation</b>                                                                                                                           | <b>Description</b>                                              | <b>Reference documentation</b>        |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|---------------------------------------|
|  <a href="#">Anomaly Detector</a><br>(deprecated 2023)              | Identify potential problems early on                            | <a href="#">Anomaly Detector API</a>  |
|  <a href="#">Content Moderator</a><br>(deprecated 2024)             | Detect potentially offensive or unwanted content                | <a href="#">Content Moderator API</a> |
|  <a href="#">Language understanding (LUIS)</a><br>(deprecated 2023) | Understand natural language in your apps                        | <a href="#">LUIS API</a>              |
|  <a href="#">Metrics Advisor</a><br>(deprecated 2023)               | An AI service that detects unwanted contents                    | <a href="#">Metrics Advisor API</a>   |
|  <a href="#">Personalizer</a><br>(deprecated 2023)                  | Create rich, personalized experiences for each user             | <a href="#">Personalizer API</a>      |
|  <a href="#">QnA maker</a><br>(deprecated 2022)                     | Distill information into easy-to-navigate questions and answers | <a href="#">QnA Maker API</a>         |

## Next steps

- [View Azure AI SDK reference](#)

# Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

# Accounts

Reference

Service: Azure AI Services

API Version: 2023-05-01

## Operations

[\[+\] Expand table](#)

|                        |                                                                                                                                                                     |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Create                 | Create Cognitive Services Account. Accounts is a resource group wide resource type. It holds the keys for developer to access intelligent APIs. It's also the re... |
| Delete                 | Deletes a Cognitive Services account from the resource group.                                                                                                       |
| Get                    | Returns a Cognitive Services account specified by the parameters.                                                                                                   |
| List                   | Returns all the resources of a particular type belonging to a subscription.                                                                                         |
| List By Resource Group | Returns all the resources of a particular type belonging to a resource group                                                                                        |
| List Keys              | Lists the account keys for the specified Cognitive Services account.                                                                                                |
| List Models            | List available Models for the requested Cognitive Services account                                                                                                  |
| List Skus              | List available SKUs for the requested Cognitive Services account                                                                                                    |
| List Usages            | Get usages for the requested Cognitive Services account                                                                                                             |
| Regenerate Key         | Regenerates the specified account key for the specified Cognitive Services account.                                                                                 |
| Update                 | Updates a Cognitive Services account                                                                                                                                |

# Microsoft Azure Cognitive Services management client library for .NET

Article • 05/07/2024

Microsoft Azure Cognitive Services brings AI within reach of every developer and data scientist. With leading models, a variety of use cases can be unlocked. All it takes is an API call to embed the ability to see, hear, speak, search, understand, and accelerate advanced decision-making into your apps. Enable developers and data scientists of all skill levels to easily add AI capabilities to their apps.

This library supports managing Microsoft Azure Cognitive Services resources.

This library follows the [new Azure SDK guidelines](#), and provides many core capabilities:

- Support MSAL.NET, Azure.Identity is out of box for supporting MSAL.NET.
- Support [OpenTelemetry](<https://opentelemetry.io/>) for distributed tracing.
- HTTP pipeline with custom policies.
- Better error-handling.
- Support uniform telemetry across all languages.

## Getting started

### Install the package

Install the Microsoft Azure Cognitive Services management library for .NET with [NuGet](#):

.NET CLI

```
dotnet add package Azure.ResourceManager.CognitiveServices
```

### Prerequisites

- You must have an [Microsoft Azure subscription](#).

### Authenticate the Client

To create an authenticated client and start interacting with Microsoft Azure resources, see the [quickstart guide here](#).

## Key concepts

Key concepts of the Microsoft Azure SDK for .NET can be found [here](#).

## Documentation

Documentation is available to help you learn how to use this package:

- [Quickstart](#).
- [API References](#).
- [Authentication](#).

## Examples

Code samples for using the management library for .NET can be found in the following locations

- [.NET Management Library Code Samples](#)

## Troubleshooting

- File an issue via [GitHub Issues](#).
- Check [previous questions](#) or ask new ones on Stack Overflow using Azure and .NET tags.

## Next steps

For more information about Microsoft Azure SDK, see [this website](#).

## Contributing

For details on contributing to this repository, see the [contributing guide](#).

This project welcomes contributions and suggestions. Most contributions require you to agree to a Contributor License Agreement (CLA) declaring that you have the right to, and actually do, grant us the rights to use your contribution. For details, visit <https://cla.microsoft.com>.

When you submit a pull request, a CLA-bot will automatically determine whether you need to provide a CLA and decorate the PR appropriately (for example, label, comment). Follow the instructions provided by the bot. You'll only need to do this action once across all repositories using our CLA.

This project has adopted the [Microsoft Open Source Code of Conduct](#). For more information, see the [Code of Conduct FAQ](#) or contact [opencode@microsoft.com](mailto:opencode@microsoft.com) with any other questions or comments.

 **Collaborate with us on GitHub**

The source for this content can be found on GitHub, where you can also create and review issues and pull requests. For more information, see [our contributor guide](#).



**Azure SDK for .NET feedback**

Azure SDK for .NET is an open source project. Select a link to provide feedback:

-  [Open a documentation issue](#)
-  [Provide product feedback](#)

# @azure/arm-cognitiveservices package

Reference

## Classes

[+] Expand table

[CognitiveServicesManagementClient](#)

## Interfaces

[+] Expand table

|                                                               |                                                                                                                    |
|---------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| <a href="#">AbusePenalty</a>                                  | The abuse penalty.                                                                                                 |
| <a href="#">Account</a>                                       | Cognitive Services account is an Azure resource representing the provisioned account, it's type, location and SKU. |
| <a href="#">AccountListResult</a>                             | The list of cognitive services accounts operation response.                                                        |
| <a href="#">AccountModel</a>                                  | Cognitive Services account Model.                                                                                  |
| <a href="#">AccountModelListResult</a>                        | The list of cognitive services accounts operation response.                                                        |
| <a href="#">AccountProperties</a>                             | Properties of Cognitive Services account.                                                                          |
| <a href="#">AccountSku</a>                                    | Cognitive Services resource type and SKU.                                                                          |
| <a href="#">AccountSkuListResult</a>                          | The list of cognitive services accounts operation response.                                                        |
| <a href="#">Accounts</a>                                      | Interface representing a Accounts.                                                                                 |
| <a href="#">AccountsCreateOptionalParams</a>                  | Optional parameters.                                                                                               |
| <a href="#">AccountsDeleteOptionalParams</a>                  | Optional parameters.                                                                                               |
| <a href="#">AccountsGetOptionalParams</a>                     | Optional parameters.                                                                                               |
| <a href="#">AccountsListByResourceGroupNextOptionalParams</a> | Optional parameters.                                                                                               |
| <a href="#">AccountsListByResourceGroupOptionalParams</a>     | Optional parameters.                                                                                               |

|                                                           |                                                                                    |
|-----------------------------------------------------------|------------------------------------------------------------------------------------|
| <a href="#">AccountsListKeysOptional</a>                  | Optional parameters.                                                               |
| <a href="#">Params</a>                                    |                                                                                    |
| <a href="#">AccountsListModelsNextOptional</a>            | Optional parameters.                                                               |
| <a href="#">OptionalParams</a>                            |                                                                                    |
| <a href="#">AccountsListModelsOptional</a>                | Optional parameters.                                                               |
| <a href="#">OptionalParams</a>                            |                                                                                    |
| <a href="#">AccountsListNextOptional</a>                  | Optional parameters.                                                               |
| <a href="#">Params</a>                                    |                                                                                    |
| <a href="#">AccountsListOptional</a>                      | Optional parameters.                                                               |
| <a href="#">Params</a>                                    |                                                                                    |
| <a href="#">AccountsListSkusOptional</a>                  | Optional parameters.                                                               |
| <a href="#">Params</a>                                    |                                                                                    |
| <a href="#">AccountsListUsagesOptional</a>                | Optional parameters.                                                               |
| <a href="#">OptionalParams</a>                            |                                                                                    |
| <a href="#">AccountsRegenerateKeyOptional</a>             | Optional parameters.                                                               |
| <a href="#">OptionalParams</a>                            |                                                                                    |
| <a href="#">AccountsUpdateOptional</a>                    | Optional parameters.                                                               |
| <a href="#">Params</a>                                    |                                                                                    |
| <a href="#">ApiKeys</a>                                   | The access keys for the cognitive services account.                                |
| <a href="#">ApiProperties</a>                             | The api properties for special APIs.                                               |
| <a href="#">AzureEntityResource</a>                       | The resource model definition for an Azure Resource Manager resource with an etag. |
| <a href="#">CallRateLimit</a>                             | The call rate limit Cognitive Services account.                                    |
| <a href="#">CapacityConfig</a>                            | The capacity configuration.                                                        |
| <a href="#">CheckDomainAvailabilityOptional</a>           | Optional parameters.                                                               |
| <a href="#">OptionalParams</a>                            |                                                                                    |
| <a href="#">CheckDomainAvailabilityParameter</a>          | Check Domain availability parameter.                                               |
| <a href="#">CheckSkuAvailabilityOptional</a>              | Optional parameters.                                                               |
| <a href="#">OptionalParams</a>                            |                                                                                    |
| <a href="#">CheckSkuAvailabilityParameter</a>             | Check SKU availability parameter.                                                  |
| <a href="#">CognitiveServicesManagementClientOptional</a> | Optional parameters.                                                               |

| Params                                                   |                                                                                        |
|----------------------------------------------------------|----------------------------------------------------------------------------------------|
| CommitmentCost                                           | Cognitive Services account commitment cost.                                            |
| CommitmentPeriod                                         | Cognitive Services account commitment period.                                          |
| CommitmentPlan                                           | Cognitive Services account commitment plan.                                            |
| CommitmentPlanAccount Association                        | The commitment plan association.                                                       |
| CommitmentPlanAccount AssociationListResult              | The list of cognitive services Commitment Plan Account Association operation response. |
| CommitmentPlan Association                               | The commitment plan association.                                                       |
| CommitmentPlanListResult                                 | The list of cognitive services accounts operation response.                            |
| CommitmentPlanProperties                                 | Properties of Cognitive Services account commitment plan.                              |
| CommitmentPlans                                          | Interface representing a CommitmentPlans.                                              |
| CommitmentPlansCreate OrUpdateAssociation OptionalParams | Optional parameters.                                                                   |
| CommitmentPlansCreate OrUpdateOptionalParams             | Optional parameters.                                                                   |
| CommitmentPlansCreate OrUpdatePlanOptional Params        | Optional parameters.                                                                   |
| CommitmentPlansDelete AssociationHeaders                 | Defines headers for CommitmentPlans_deleteAssociation operation.                       |
| CommitmentPlansDelete AssociationOptionalParams          | Optional parameters.                                                                   |
| CommitmentPlansDelete OptionalParams                     | Optional parameters.                                                                   |
| CommitmentPlansDelete PlanHeaders                        | Defines headers for CommitmentPlans_deletePlan operation.                              |
| CommitmentPlansDelete PlanOptionalParams                 | Optional parameters.                                                                   |
| CommitmentPlansGet AssociationOptionalParams             | Optional parameters.                                                                   |

|                                                |                                                             |
|------------------------------------------------|-------------------------------------------------------------|
| <a href="#">CommitmentPlansGet</a>             | Optional parameters.                                        |
| <a href="#">OptionalParams</a>                 |                                                             |
| <a href="#">CommitmentPlansGetPlan</a>         | Optional parameters.                                        |
| <a href="#">OptionalParams</a>                 |                                                             |
| <a href="#">CommitmentPlansList</a>            | Optional parameters.                                        |
| <a href="#">AssociationsNextOptionalParams</a> |                                                             |
| <a href="#">CommitmentPlansList</a>            | Optional parameters.                                        |
| <a href="#">AssociationsOptionalParams</a>     |                                                             |
| <a href="#">CommitmentPlansListNext</a>        | Optional parameters.                                        |
| <a href="#">OptionalParams</a>                 |                                                             |
| <a href="#">CommitmentPlansList</a>            | Optional parameters.                                        |
| <a href="#">OptionalParams</a>                 |                                                             |
| <a href="#">CommitmentPlansListPlans</a>       | Optional parameters.                                        |
| <a href="#">ByResourceGroupNext</a>            |                                                             |
| <a href="#">OptionalParams</a>                 |                                                             |
| <a href="#">CommitmentPlansListPlans</a>       | Optional parameters.                                        |
| <a href="#">ByResourceGroupOptionalParams</a>  |                                                             |
| <a href="#">CommitmentPlansListPlans</a>       | Optional parameters.                                        |
| <a href="#">BySubscriptionNext</a>             |                                                             |
| <a href="#">OptionalParams</a>                 |                                                             |
| <a href="#">CommitmentPlansListPlans</a>       | Optional parameters.                                        |
| <a href="#">BySubscriptionOptionalParams</a>   |                                                             |
| <a href="#">CommitmentPlansUpdate</a>          | Defines headers for CommitmentPlans_updatePlan operation.   |
| <a href="#">PlanHeaders</a>                    |                                                             |
| <a href="#">CommitmentPlansUpdate</a>          | Optional parameters.                                        |
| <a href="#">PlanOptionalParams</a>             |                                                             |
| <a href="#">CommitmentQuota</a>                | Cognitive Services account commitment quota.                |
| <a href="#">CommitmentTier</a>                 | Cognitive Services account commitment tier.                 |
| <a href="#">CommitmentTierListResult</a>       | The list of cognitive services accounts operation response. |
| <a href="#">CommitmentTiers</a>                | Interface representing a CommitmentTiers.                   |
| <a href="#">CommitmentTiersListNext</a>        | Optional parameters.                                        |
| <a href="#">OptionalParams</a>                 |                                                             |

|                                         |                                                             |
|-----------------------------------------|-------------------------------------------------------------|
| <a href="#">CommitmentTiersList</a>     | Optional parameters.                                        |
| <a href="#">OptionalParams</a>          |                                                             |
| <a href="#">DeletedAccounts</a>         | Interface representing a DeletedAccounts.                   |
| <a href="#">DeletedAccountsGet</a>      | Optional parameters.                                        |
| <a href="#">OptionalParams</a>          |                                                             |
| <a href="#">DeletedAccountsListNext</a> | Optional parameters.                                        |
| <a href="#">OptionalParams</a>          |                                                             |
| <a href="#">DeletedAccountsList</a>     | Optional parameters.                                        |
| <a href="#">OptionalParams</a>          |                                                             |
| <a href="#">DeletedAccountsPurge</a>    | Optional parameters.                                        |
| <a href="#">OptionalParams</a>          |                                                             |
| <a href="#">Deployment</a>              | Cognitive Services account deployment.                      |
| <a href="#">DeploymentListResult</a>    | The list of cognitive services accounts operation response. |
| <a href="#">DeploymentModel</a>         | Properties of Cognitive Services account deployment model.  |
| <a href="#">DeploymentProperties</a>    | Properties of Cognitive Services account deployment.        |
| <a href="#">DeploymentScaleSettings</a> | Properties of Cognitive Services account deployment model.  |
| <a href="#">Deployments</a>             | Interface representing a Deployments.                       |
| <a href="#">DeploymentsCreate</a>       | Optional parameters.                                        |
| <a href="#">OrUpdateOptionalParams</a>  |                                                             |
| <a href="#">DeploymentsDelete</a>       | Optional parameters.                                        |
| <a href="#">OptionalParams</a>          |                                                             |
| <a href="#">DeploymentsGetOptional</a>  | Optional parameters.                                        |
| <a href="#">Params</a>                  |                                                             |
| <a href="#">DeploymentsListNext</a>     | Optional parameters.                                        |
| <a href="#">OptionalParams</a>          |                                                             |
| <a href="#">DeploymentsListOptional</a> | Optional parameters.                                        |
| <a href="#">Params</a>                  |                                                             |
| <a href="#">DomainAvailability</a>      | Domain availability.                                        |
| <a href="#">Encryption</a>              | Properties to configure Encryption                          |
| <a href="#">ErrorAdditionalInfo</a>     | The resource management error additional info.              |
| <a href="#">ErrorDetail</a>             | The error detail.                                           |

|                                                  |                                                                                                                                                                |
|--------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">ErrorResponse</a>                    | Common error response for all Azure Resource Manager APIs to return error details for failed operations. (This also follows the OData error response format.). |
| <a href="#">Identity</a>                         | Identity for the resource.                                                                                                                                     |
| <a href="#">IpRule</a>                           | A rule governing the accessibility from a specific ip address or ip range.                                                                                     |
| <a href="#">KeyVaultProperties</a>               | Properties to configure keyVault Properties                                                                                                                    |
| <a href="#">MetricName</a>                       | A metric name.                                                                                                                                                 |
| <a href="#">Model</a>                            | Cognitive Services Model.                                                                                                                                      |
| <a href="#">ModelDeprecationInfo</a>             | Cognitive Services account ModelDeprecationInfo.                                                                                                               |
| <a href="#">ModelListResult</a>                  | The list of cognitive services models.                                                                                                                         |
| <a href="#">ModelSku</a>                         | Describes an available Cognitive Services Model SKU.                                                                                                           |
| <a href="#">Models</a>                           | Interface representing a Models.                                                                                                                               |
| <a href="#">ModelsListNextOptionalParams</a>     | Optional parameters.                                                                                                                                           |
| <a href="#">ModelsListOptionalParams</a>         | Optional parameters.                                                                                                                                           |
| <a href="#">MultiRegionSettings</a>              | The multiregion settings Cognitive Services account.                                                                                                           |
| <a href="#">NetworkRuleSet</a>                   | A set of rules governing the network accessibility.                                                                                                            |
| <a href="#">Operation</a>                        | Details of a REST API operation, returned from the Resource Provider Operations API                                                                            |
| <a href="#">OperationDisplay</a>                 | Localized display information for this particular operation.                                                                                                   |
| <a href="#">OperationListResult</a>              | A list of REST API operations supported by an Azure Resource Provider. It contains an URL link to get the next set of results.                                 |
| <a href="#">Operations</a>                       | Interface representing a Operations.                                                                                                                           |
| <a href="#">OperationsListNextOptionalParams</a> | Optional parameters.                                                                                                                                           |
| <a href="#">OperationsListOptionalParams</a>     | Optional parameters.                                                                                                                                           |
| <a href="#">PatchResourceTags</a>                | The object being used to update tags of a resource, in general used for PATCH operations.                                                                      |
| <a href="#">PatchResourceTagsAndSku</a>          | The object being used to update tags and sku of a resource, in general used for PATCH operations.                                                              |

|                                                                        |                                                                                                                 |
|------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| <a href="#">PrivateEndpoint</a>                                        | The Private Endpoint resource.                                                                                  |
| <a href="#">PrivateEndpointConnection</a>                              | The Private Endpoint Connection resource.                                                                       |
| <a href="#">PrivateEndpointConnectionListResult</a>                    | A list of private endpoint connections                                                                          |
| <a href="#">PrivateEndpointConnectionProperties</a>                    | Properties of the PrivateEndpointConnectProperties.                                                             |
| <a href="#">PrivateEndpointConnections</a>                             | Interface representing a PrivateEndpointConnections.                                                            |
| <a href="#">PrivateEndpointConnectionsCreateOrUpdateOptionalParams</a> | Optional parameters.                                                                                            |
| <a href="#">PrivateEndpointConnectionsDeleteOptionalParams</a>         | Optional parameters.                                                                                            |
| <a href="#">PrivateEndpointConnectionsGetOptionalParams</a>            | Optional parameters.                                                                                            |
| <a href="#">PrivateEndpointConnectionsListOptionalParams</a>           | Optional parameters.                                                                                            |
| <a href="#">PrivateLinkResource</a>                                    | A private link resource                                                                                         |
| <a href="#">PrivateLinkResourceListResult</a>                          | A list of private link resources                                                                                |
| <a href="#">PrivateLinkResourceProperties</a>                          | Properties of a private link resource.                                                                          |
| <a href="#">PrivateLinkResources</a>                                   | Interface representing a PrivateLinkResources.                                                                  |
| <a href="#">PrivateLinkResourcesListOptionalParams</a>                 | Optional parameters.                                                                                            |
| <a href="#">PrivateLinkService ConnectionState</a>                     | A collection of information about the state of the connection between service consumer and provider.            |
| <a href="#">ProxyResource</a>                                          | The resource model definition for a Azure Resource Manager proxy resource. It will not have tags and a location |
| <a href="#">QuotaLimit</a>                                             |                                                                                                                 |
| <a href="#">RegenerateKeyParameters</a>                                | Regenerate key parameters.                                                                                      |

|                                        |                                                                                          |
|----------------------------------------|------------------------------------------------------------------------------------------|
| RegionSetting                          | The call rate limit Cognitive Services account.                                          |
| RequestMatchPattern                    |                                                                                          |
| Resource                               | Common fields that are returned in the response for all Azure Resource Manager resources |
| ResourceSku                            | Describes an available Cognitive Services SKU.                                           |
| ResourceSkuListResult                  | The Get Skus operation response.                                                         |
| ResourceSkuRestrictionInfo             |                                                                                          |
| ResourceSkuRestrictions                | Describes restrictions of a SKU.                                                         |
| ResourceSkus                           | Interface representing a ResourceSkus.                                                   |
| ResourceSkusListNext<br>OptionalParams | Optional parameters.                                                                     |
| ResourceSkusListOptional<br>Params     | Optional parameters.                                                                     |
| Sku                                    | The resource model definition representing SKU                                           |
| SkuAvailability                        | SKU availability.                                                                        |
| SkuAvailabilityListResult              | Check SKU availability result list.                                                      |
| SkuCapability                          | SkuCapability indicates the capability of a certain feature.                             |
| SkuChangeInfo                          | Sku change info of account.                                                              |
| SystemData                             | Metadata pertaining to creation and last modification of the resource.                   |
| ThrottlingRule                         |                                                                                          |
| Usage                                  | The usage data for a usage request.                                                      |
| UsageListResult                        | The response to a list usage request.                                                    |
| Usages                                 | Interface representing a Usages.                                                         |
| UsagesListNextOptional<br>Params       | Optional parameters.                                                                     |
| UsagesListOptionalParams               | Optional parameters.                                                                     |
| UserAssignedIdentity                   | User-assigned managed identity.                                                          |
| UserOwnedStorage                       | The user owned storage for Cognitive Services account.                                   |

## VirtualNetworkRule

A rule governing the accessibility from a specific virtual network.

# Type Aliases

[ ] [Expand table](#)

|                                                         |                                                                                                                                                                                                       |
|---------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">AbusePenaltyAction</a>                      | Defines values for AbusePenaltyAction.<br><a href="#">KnownAbusePenaltyAction</a> can be used interchangeably with AbusePenaltyAction, this enum contains the known values that the service supports. |
| <h2>Known values supported by the service</h2>          |                                                                                                                                                                                                       |
| <a href="#">AccountsCreateResponse</a>                  | Contains response data for the create operation.                                                                                                                                                      |
| <a href="#">AccountsGetResponse</a>                     | Contains response data for the get operation.                                                                                                                                                         |
| <a href="#">AccountsListByResourceGroupNextResponse</a> | Contains response data for the listByResourceGroupNext operation.                                                                                                                                     |
| <a href="#">AccountsListByResourceGroupResponse</a>     | Contains response data for the listByResourceGroup operation.                                                                                                                                         |
| <a href="#">AccountsListKeysResponse</a>                | Contains response data for the listKeys operation.                                                                                                                                                    |
| <a href="#">AccountsListModelsNextResponse</a>          | Contains response data for the listModelsNext operation.                                                                                                                                              |
| <a href="#">AccountsListModelsResponse</a>              | Contains response data for the listModels operation.                                                                                                                                                  |
| <a href="#">AccountsListNextResponse</a>                | Contains response data for the listNext operation.                                                                                                                                                    |
| <a href="#">AccountsListResponse</a>                    | Contains response data for the list operation.                                                                                                                                                        |
| <a href="#">AccountsListSkusResponse</a>                | Contains response data for the listSkus operation.                                                                                                                                                    |

|                                                                               |                                                                                                                                                                                          |
|-------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">AccountsListUsages</a>                                            | Contains response data for the listUsages operation.                                                                                                                                     |
| <a href="#">Response</a>                                                      |                                                                                                                                                                                          |
| <a href="#">AccountsRegenerate</a>                                            | Contains response data for the regenerateKey operation.                                                                                                                                  |
| <a href="#">KeyResponse</a>                                                   |                                                                                                                                                                                          |
| <a href="#">AccountsUpdate</a>                                                | Contains response data for the update operation.                                                                                                                                         |
| <a href="#">Response</a>                                                      |                                                                                                                                                                                          |
| <a href="#">ActionType</a>                                                    | <p>Defines values for ActionType.</p> <p><a href="#">KnownActionType</a> can be used interchangeably with ActionType, this enum contains the known values that the service supports.</p> |
| <h2>Known values supported by the service</h2>                                |                                                                                                                                                                                          |
| <b>Internal</b>                                                               |                                                                                                                                                                                          |
| <a href="#">CheckDomain</a>                                                   | Contains response data for the checkDomainAvailability operation.                                                                                                                        |
| <a href="#">AvailabilityResponse</a>                                          |                                                                                                                                                                                          |
| <a href="#">CheckSkuAvailability</a>                                          | Contains response data for the checkSkuAvailability operation.                                                                                                                           |
| <a href="#">Response</a>                                                      |                                                                                                                                                                                          |
| <a href="#">CommitmentPlan</a>                                                | Defines values for CommitmentPlanProvisioningState.                                                                                                                                      |
| <a href="#">ProvisioningState</a>                                             | <a href="#">KnownCommitmentPlanProvisioningState</a> can be used interchangeably with CommitmentPlanProvisioningState, this enum contains the known values that the service supports.    |
| <h2>Known values supported by the service</h2>                                |                                                                                                                                                                                          |
| Accepted<br>Creating<br>Deleting<br>Moving<br>Failed<br>Succeeded<br>Canceled |                                                                                                                                                                                          |
| <a href="#">CommitmentPlans</a>                                               | Contains response data for the createOrUpdateAssociation operation.                                                                                                                      |
| <a href="#">CreateOrUpdate</a>                                                |                                                                                                                                                                                          |
| <a href="#">AssociationResponse</a>                                           |                                                                                                                                                                                          |
| <a href="#">CommitmentPlans</a>                                               | Contains response data for the createOrUpdatePlan operation.                                                                                                                             |
| <a href="#">CreateOrUpdatePlan</a>                                            |                                                                                                                                                                                          |
| <a href="#">Response</a>                                                      |                                                                                                                                                                                          |
| <a href="#">CommitmentPlans</a>                                               | Contains response data for the createOrUpdate operation.                                                                                                                                 |
| <a href="#">CreateOrUpdate</a>                                                |                                                                                                                                                                                          |
| <a href="#">Response</a>                                                      |                                                                                                                                                                                          |

|                                     |                                                                                                                                                                                        |
|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">CommitmentPlans</a>     | Contains response data for the getAssociation operation.                                                                                                                               |
| <a href="#">GetAssociation</a>      |                                                                                                                                                                                        |
| <a href="#">Response</a>            |                                                                                                                                                                                        |
| <a href="#">CommitmentPlans</a>     | Contains response data for the getPlan operation.                                                                                                                                      |
| <a href="#">GetPlanResponse</a>     |                                                                                                                                                                                        |
| <a href="#">CommitmentPlans</a>     | Contains response data for the get operation.                                                                                                                                          |
| <a href="#">GetResponse</a>         |                                                                                                                                                                                        |
| <a href="#">CommitmentPlansList</a> | Contains response data for the listAssociationsNext operation.                                                                                                                         |
| <a href="#">AssociationsNext</a>    |                                                                                                                                                                                        |
| <a href="#">Response</a>            |                                                                                                                                                                                        |
| <a href="#">CommitmentPlansList</a> | Contains response data for the listAssociations operation.                                                                                                                             |
| <a href="#">Associations</a>        |                                                                                                                                                                                        |
| <a href="#">Response</a>            |                                                                                                                                                                                        |
| <a href="#">CommitmentPlansList</a> | Contains response data for the listNext operation.                                                                                                                                     |
| <a href="#">NextResponse</a>        |                                                                                                                                                                                        |
| <a href="#">CommitmentPlansList</a> | Contains response data for the listPlansByResourceGroupNext operation.                                                                                                                 |
| <a href="#">PlansByResource</a>     |                                                                                                                                                                                        |
| <a href="#">GroupNextResponse</a>   |                                                                                                                                                                                        |
| <a href="#">CommitmentPlansList</a> | Contains response data for the listPlansByResourceGroup operation.                                                                                                                     |
| <a href="#">PlansByResource</a>     |                                                                                                                                                                                        |
| <a href="#">GroupResponse</a>       |                                                                                                                                                                                        |
| <a href="#">CommitmentPlansList</a> | Contains response data for the listPlansBySubscriptionNext operation.                                                                                                                  |
| <a href="#">PlansBySubscription</a> |                                                                                                                                                                                        |
| <a href="#">NextResponse</a>        |                                                                                                                                                                                        |
| <a href="#">CommitmentPlansList</a> | Contains response data for the listPlansBySubscription operation.                                                                                                                      |
| <a href="#">PlansBySubscription</a> |                                                                                                                                                                                        |
| <a href="#">Response</a>            |                                                                                                                                                                                        |
| <a href="#">CommitmentPlansList</a> | Contains response data for the list operation.                                                                                                                                         |
| <a href="#">Response</a>            |                                                                                                                                                                                        |
| <a href="#">CommitmentPlans</a>     | Contains response data for the updatePlan operation.                                                                                                                                   |
| <a href="#">UpdatePlanResponse</a>  |                                                                                                                                                                                        |
| <a href="#">CommitmentTiersList</a> | Contains response data for the listNext operation.                                                                                                                                     |
| <a href="#">NextResponse</a>        |                                                                                                                                                                                        |
| <a href="#">CommitmentTiersList</a> | Contains response data for the list operation.                                                                                                                                         |
| <a href="#">Response</a>            |                                                                                                                                                                                        |
| <a href="#">CreatedByType</a>       | Defines values for CreatedByType.<br><a href="#">KnownCreatedByType</a> can be used interchangeably with CreatedByType, this enum contains the known values that the service supports. |

## Known values supported by the service

User  
Application  
ManagedIdentity  
Key

|                                     |                                                                                                                                                                                                                                                          |
|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DeletedAccountsGetResponse          | Contains response data for the get operation.                                                                                                                                                                                                            |
| DeletedAccountsListNextResponse     | Contains response data for the listNext operation.                                                                                                                                                                                                       |
| DeletedAccountsListResponse         | Contains response data for the list operation.                                                                                                                                                                                                           |
| DeploymentModelVersionUpgradeOption | Defines values for DeploymentModelVersionUpgradeOption.<br><a href="#">KnownDeploymentModelVersionUpgradeOption</a> can be used interchangeably with DeploymentModelVersionUpgradeOption, this enum contains the known values that the service supports. |

## Known values supported by the service

OnceNewDefaultVersionAvailable  
OnceCurrentVersionExpired  
NoAutoUpgrade

|                             |                                                                                                                                                                                                                                  |
|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DeploymentProvisioningState | Defines values for DeploymentProvisioningState.<br><a href="#">KnownDeploymentProvisioningState</a> can be used interchangeably with DeploymentProvisioningState, this enum contains the known values that the service supports. |
|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## Known values supported by the service

Accepted  
Creating  
Deleting  
Moving  
Failed  
Succeeded  
Disabled  
Canceled

|                     |                                                                                                                                                                                                          |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DeploymentScaleType | Defines values for DeploymentScaleType.<br><a href="#">KnownDeploymentScaleType</a> can be used interchangeably with DeploymentScaleType, this enum contains the known values that the service supports. |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## Known values supported by the service

|                                   |                                                                                                                                                                                     |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                   | Standard<br>Manual                                                                                                                                                                  |
| DeploymentsCreateOrUpdateResponse | Contains response data for the createOrUpdate operation.                                                                                                                            |
| DeploymentsGetResponse            | Contains response data for the get operation.                                                                                                                                       |
| DeploymentsListNextResponse       | Contains response data for the listNext operation.                                                                                                                                  |
| DeploymentsListResponse           | Contains response data for the list operation.                                                                                                                                      |
| HostingModel                      | Defines values for HostingModel.<br><a href="#">KnownHostingModel</a> can be used interchangeably with HostingModel, this enum contains the known values that the service supports. |

## Known values supported by the service

|           |                                                                                                                                                                            |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           | Web<br>ConnectedContainer<br>DisconnectedContainer<br>ProvisionedWeb                                                                                                       |
| KeyName   | Defines values for KeyName.                                                                                                                                                |
| KeySource | Defines values for KeySource.<br><a href="#">KnownKeySource</a> can be used interchangeably with KeySource, this enum contains the known values that the service supports. |

## Known values supported by the service

|                      |                                                                                                                                                                                                             |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                      | Microsoft.CognitiveServices<br>Microsoft.KeyVault                                                                                                                                                           |
| ModelLifecycleStatus | Defines values for ModelLifecycleStatus.<br><a href="#">KnownModelLifecycleStatus</a> can be used interchangeably with ModelLifecycleStatus, this enum contains the known values that the service supports. |

## Known values supported by the service

|  |                               |
|--|-------------------------------|
|  | GenerallyAvailable<br>Preview |
|--|-------------------------------|

|                                                                  |                                                                                                                                                                                                                                                                               |
|------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">ModelsListNextResponse</a>                           | Contains response data for the listNext operation.                                                                                                                                                                                                                            |
| <a href="#">ModelsListResponse</a>                               | Contains response data for the list operation.                                                                                                                                                                                                                                |
| <a href="#">NetworkRuleAction</a>                                | Defines values for NetworkRuleAction.<br><a href="#">KnownNetworkRuleAction</a> can be used interchangeably with NetworkRuleAction, this enum contains the known values that the service supports.                                                                            |
| <h2>Known values supported by the service</h2>                   |                                                                                                                                                                                                                                                                               |
|                                                                  | Allow<br>Deny                                                                                                                                                                                                                                                                 |
| <a href="#">OperationsListNextResponse</a>                       | Contains response data for the listNext operation.                                                                                                                                                                                                                            |
| <a href="#">OperationsListResponse</a>                           | Contains response data for the list operation.                                                                                                                                                                                                                                |
| <a href="#">Origin</a>                                           | Defines values for Origin.<br><a href="#">KnownOrigin</a> can be used interchangeably with Origin, this enum contains the known values that the service supports.                                                                                                             |
| <h2>Known values supported by the service</h2>                   |                                                                                                                                                                                                                                                                               |
|                                                                  | user<br>system<br>user,system                                                                                                                                                                                                                                                 |
| <a href="#">PrivateEndpointConnectionProvisioningState</a>       | Defines values for PrivateEndpointConnectionProvisioningState.<br><a href="#">KnownPrivateEndpointConnectionProvisioningState</a> can be used interchangeably with PrivateEndpointConnectionProvisioningState, this enum contains the known values that the service supports. |
| <h2>Known values supported by the service</h2>                   |                                                                                                                                                                                                                                                                               |
|                                                                  | Succeeded<br>Creating<br>Deleting<br>Failed                                                                                                                                                                                                                                   |
| <a href="#">PrivateEndpointConnectionsCreateOrUpdateResponse</a> | Contains response data for the createOrUpdate operation.                                                                                                                                                                                                                      |

|                                                        |                                                                                                                                                                                                                                                                                                                                  |
|--------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">PrivateEndpointConnectionsGetResponse</a>  | Contains response data for the get operation.                                                                                                                                                                                                                                                                                    |
| <a href="#">PrivateEndpointConnectionsListResponse</a> | Contains response data for the list operation.                                                                                                                                                                                                                                                                                   |
| <a href="#">PrivateEndpointServiceConnectionStatus</a> | Defines values for PrivateEndpointServiceConnectionStatus. <a href="#">KnownPrivateEndpointServiceConnectionStatus</a> can be used interchangeably with PrivateEndpointServiceConnectionStatus, this enum contains the known values that the service supports.<br><br><b>Possible values:</b><br>Pending<br>Approved<br>Rejected |
| <a href="#">PrivateLinkResourcesListResponse</a>       | Contains response data for the list operation.                                                                                                                                                                                                                                                                                   |
| <a href="#">ProvisioningState</a>                      | Defines values for ProvisioningState. <a href="#">KnownProvisioningState</a> can be used interchangeably with ProvisioningState, this enum contains the known values that the service supports.<br><br><b>Possible values:</b><br>Accepted<br>Creating<br>Deleting<br>Moving<br>Failed<br>Succeeded<br>ResolvingDNS              |
| <a href="#">PublicNetworkAccess</a>                    | Defines values for PublicNetworkAccess. <a href="#">KnownPublicNetworkAccess</a> can be used interchangeably with PublicNetworkAccess, this enum contains the known values that the service supports.<br><br><b>Possible values:</b><br>Enabled<br>Disabled                                                                      |

|                                  |                                                                                                                                                                                                 |
|----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">QuotaUsageStatus</a> | Defines values for QuotaUsageStatus.<br><a href="#">KnownQuotaUsageStatus</a> can be used interchangeably with QuotaUsageStatus, this enum contains the known values that the service supports. |
|                                  | <b>Included</b><br><b>Blocked</b><br><b>InOverage</b><br><b>Unknown</b>                                                                                                                         |

## Known values supported by the service

|                                              |                                                                                                                                                                                           |
|----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">ResourceSku</a>                  | Defines values for ResourceSkuRestrictionsReasonCode.                                                                                                                                     |
| <a href="#">RestrictionsReasonCode</a>       | <a href="#">KnownResourceSkuRestrictionsReasonCode</a> can be used interchangeably with ResourceSkuRestrictionsReasonCode, this enum contains the known values that the service supports. |
|                                              | <b>QuotaId</b><br><b>NotAvailableForSubscription</b>                                                                                                                                      |
| <a href="#">ResourceSku</a>                  | Defines values for ResourceSkuRestrictionsType.                                                                                                                                           |
| <a href="#">RestrictionsType</a>             |                                                                                                                                                                                           |
| <a href="#">ResourceSkusListNextResponse</a> | Contains response data for the listNext operation.                                                                                                                                        |
| <a href="#">ResourceSkusListResponse</a>     | Contains response data for the list operation.                                                                                                                                            |
| <a href="#">RoutingMethods</a>               | Defines values for RoutingMethods.<br><a href="#">KnownRoutingMethods</a> can be used interchangeably with RoutingMethods, this enum contains the known values that the service supports. |
|                                              | <b>Priority</b><br><b>Weighted</b><br><b>Performance</b>                                                                                                                                  |
| <a href="#">SkuTier</a>                      | Defines values for SkuTier.<br><a href="#">KnownSkuTier</a> can be used interchangeably with SkuTier, this enum contains the known values that the service supports.                      |

## Known values supported by the service

Free  
Basic  
Standard  
Premium  
Enterprise

|          |                                                                                                                                                                         |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| UnitType | Defines values for UnitType.<br><a href="#">KnownUnitType</a> can be used interchangeably with UnitType, this enum contains the known values that the service supports. |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## Known values supported by the service

Count  
Bytes  
Seconds  
Percent  
CountPerSecond  
BytesPerSecond  
Milliseconds

|                        |                                                    |
|------------------------|----------------------------------------------------|
| UsagesListNextResponse | Contains response data for the listNext operation. |
| UsagesListResponse     | Contains response data for the list operation.     |

## Enums

[+] [Expand table](#)

|                                                          |                                                                                               |
|----------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| <a href="#">KnownAbusePenaltyAction</a>                  | Known values of <a href="#">AbusePenaltyAction</a> that the service accepts.                  |
| <a href="#">KnownActionType</a>                          | Known values of <a href="#">ActionType</a> that the service accepts.                          |
| <a href="#">KnownCommitmentPlanProvisioningState</a>     | Known values of <a href="#">CommitmentPlanProvisioningState</a> that the service accepts.     |
| <a href="#">KnownCreatedByType</a>                       | Known values of <a href="#">CreatedByType</a> that the service accepts.                       |
| <a href="#">KnownDeploymentModelVersionUpgradeOption</a> | Known values of <a href="#">DeploymentModelVersionUpgradeOption</a> that the service accepts. |
| <a href="#">KnownDeploymentProvisioningState</a>         | Known values of <a href="#">DeploymentProvisioningState</a> that the service accepts.         |
| <a href="#">KnownDeploymentScaleType</a>                 | Known values of <a href="#">DeploymentScaleType</a> that the service accepts.                 |

|                                                 |                                                                                                      |
|-------------------------------------------------|------------------------------------------------------------------------------------------------------|
| KnownHostingModel                               | Known values of <a href="#">HostingModel</a> that the service accepts.                               |
| KnownKeySource                                  | Known values of <a href="#">KeySource</a> that the service accepts.                                  |
| KnownModelLifecycleStatus                       | Known values of <a href="#">ModelLifecycleStatus</a> that the service accepts.                       |
| KnownNetworkRuleAction                          | Known values of <a href="#">NetworkRuleAction</a> that the service accepts.                          |
| KnownOrigin                                     | Known values of <a href="#">Origin</a> that the service accepts.                                     |
| KnownPrivateEndpointConnectionProvisioningState | Known values of <a href="#">PrivateEndpointConnectionProvisioningState</a> that the service accepts. |
| KnownPrivateEndpointServiceConnectionStatus     | Known values of <a href="#">PrivateEndpointServiceConnectionStatus</a> that the service accepts.     |
| KnownProvisioningState                          | Known values of <a href="#">ProvisioningState</a> that the service accepts.                          |
| KnownPublicNetworkAccess                        | Known values of <a href="#">PublicNetworkAccess</a> that the service accepts.                        |
| KnownQuotaUsageStatus                           | Known values of <a href="#">QuotaUsageStatus</a> that the service accepts.                           |
| KnownResourceSkuRestrictionsReasonCode          | Known values of <a href="#">ResourceSkuRestrictionsReasonCode</a> that the service accepts.          |
| KnownRoutingMethods                             | Known values of <a href="#">RoutingMethods</a> that the service accepts.                             |
| KnownSkuTier                                    | Known values of <a href="#">SkuTier</a> that the service accepts.                                    |
| KnownUnitType                                   | Known values of <a href="#">UnitType</a> that the service accepts.                                   |

## Functions

[+] [Expand table](#)

|                               |                                                                                                                                                                       |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| getContinuationToken(unknown) | Given the last <code>.value</code> produced by the <code>byPage</code> iterator, returns a continuation token that can be used to begin paging from that point later. |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## Function Details

### getContinuationToken(unknown)

Given the last `.value` produced by the `byPage` iterator, returns a continuation token that can be used to begin paging from that point later.

```
function getContinuationToken(page: unknown): string | undefined
```

## Parameters

**page** `unknown`

An object from accessing `value` on the `IteratorResult` from a `byPage` iterator.

## Returns

`string | undefined`

The continuation token that can be passed into `byPage()` during future calls.

# cognitiveservices Package

Reference

## Packages

[\[\] Expand table](#)

aio

models

operations

## Classes

[\[\] Expand table](#)

[CognitiveServicesManagementClient](#)

Cognitive Services Management Client.

 Collaborate with us on GitHub

The source for this content can be found on GitHub, where you can also create and review issues and pull requests. For more information, see [our contributor guide](#).



Azure SDK for Python feedback

Azure SDK for Python is an open source project. Select a link to provide feedback:

 [Open a documentation issue](#)

 [Provide product feedback](#)

# com.microsoft.azure.management.cognitiveservices

Reference

Package: com.microsoft.azure.management.cognitiveservices

Maven Artifact: [com.microsoft.azure:azure-mgmt-cognitiveservices:1.10.0-beta](#)

This package contains the classes for CognitiveServicesManagementClient. Cognitive Services Management Client.

## Classes

[ ] [Expand table](#)

|                                                          |                                                                  |
|----------------------------------------------------------|------------------------------------------------------------------|
| <a href="#">CheckSkuAvailabilityParameter</a>            | Check SKU availability parameter.                                |
| <a href="#">CheckSkuAvailabilityResult</a>               | Check SKU availability result.                                   |
| <a href="#">CognitiveServicesAccountUpdateParameters</a> | The parameters to provide for the account.                       |
| <a href="#">CognitiveServicesResourceAndSku</a>          | Cognitive Services resource type and SKU.                        |
| <a href="#">Error</a>                                    | Cognitive Services error object.                                 |
| <a href="#">ErrorBody</a>                                | Cognitive Services error body.                                   |
| <a href="#">ErrorException</a>                           | Exception thrown for an invalid response with Error information. |
| <a href="#">Kind</a>                                     | Defines values for Kind.                                         |
| <a href="#">MetricName</a>                               | A metric name.                                                   |
| <a href="#">OperationDisplayInfo</a>                     | The operation supported by Cognitive Services.                   |
| <a href="#">ProvisioningState</a>                        | Defines values for ProvisioningState.                            |
| <a href="#">QuotaUsageStatus</a>                         | Defines values for QuotaUsageStatus.                             |
| <a href="#">RegenerateKeyParameters</a>                  | Regenerate key parameters.                                       |
| <a href="#">SkuName</a>                                  | Defines values for SkuName.                                      |
| <a href="#">UnitType</a>                                 | Defines values for UnitType.                                     |
| <a href="#">Usage</a>                                    | The usage data for a usage request.                              |

# Enums

 Expand table

|         |                             |
|---------|-----------------------------|
| KeyName | Defines values for KeyName. |
| SkuTier | Defines values for SkuTier. |

## Collaborate with us on GitHub

The source for this content can be found on GitHub, where you can also create and review issues and pull requests. For more information, see [our contributor guide](#).



## Azure SDK for Java feedback

Azure SDK for Java is an open source project. Select a link to provide feedback:

 [Open a documentation issue](#)

 [Provide product feedback](#)

# az cognitiveservices

Reference

Manage Azure Cognitive Services accounts.

This article lists the Azure CLI commands for Azure Cognitive Services account and subscription management only. Refer to the documentation at <https://docs.microsoft.com/azure/cognitive-services/> for individual services to learn how to use the APIs and supported SDKs.

## Commands

[Expand table](#)

| Name                                                                | Description                                                      | Type | Status |
|---------------------------------------------------------------------|------------------------------------------------------------------|------|--------|
| <a href="#">az cognitiveservices account</a>                        | Manage Azure Cognitive Services accounts.                        | Core | GA     |
| <a href="#">az cognitiveservices account commitment-plan</a>        | Manage commitment plans for Azure Cognitive Services accounts.   | Core | GA     |
| <a href="#">az cognitiveservices account commitment-plan create</a> | Create a commitment plan for Azure Cognitive Services account.   | Core | GA     |
| <a href="#">az cognitiveservices account commitment-plan delete</a> | Delete a commitment plan from Azure Cognitive Services account.  | Core | GA     |
| <a href="#">az cognitiveservices account commitment-plan list</a>   | Show all commitment plans from Azure Cognitive Services account. | Core | GA     |
| <a href="#">az cognitiveservices account commitment-plan show</a>   | Show a commitment plan from Azure Cognitive Services account.    | Core | GA     |
| <a href="#">az cognitiveservices account create</a>                 | Manage Azure Cognitive Services accounts.                        | Core | GA     |
| <a href="#">az cognitiveservices account delete</a>                 | Manage Azure Cognitive Services accounts.                        | Core | GA     |
| <a href="#">az cognitiveservices account deployment</a>             | Manage deployments for Azure Cognitive Services accounts.        | Core | GA     |
| <a href="#">az cognitiveservices account deployment create</a>      | Create a deployment for Azure Cognitive Services account.        | Core | GA     |

| Name                                                           | Description                                                | Type | Status |
|----------------------------------------------------------------|------------------------------------------------------------|------|--------|
| <a href="#">az cognitiveservices account deployment delete</a> | Delete a deployment from Azure Cognitive Services account. | Core | GA     |
| <a href="#">az cognitiveservices account deployment list</a>   | Show all deployments for Azure Cognitive Services account. | Core | GA     |
| <a href="#">az cognitiveservices account deployment show</a>   | Show a deployment for Azure Cognitive Services account.    | Core | GA     |
| <a href="#">az cognitiveservices account identity</a>          | Manage identity of Cognitive Services accounts.            | Core | GA     |
| <a href="#">az cognitiveservices account identity assign</a>   | Assign an identity of a Cognitive Services account.        | Core | GA     |
| <a href="#">az cognitiveservices account identity remove</a>   | Remove the identity from a Cognitive Services account.     | Core | GA     |
| <a href="#">az cognitiveservices account identity show</a>     | Show the identity of a Cognitive Services account.         | Core | GA     |
| <a href="#">az cognitiveservices account keys</a>              | Manage Azure Cognitive Services accounts.                  | Core | GA     |
| <a href="#">az cognitiveservices account keys list</a>         | Manage Azure Cognitive Services accounts.                  | Core | GA     |
| <a href="#">az cognitiveservices account keys regenerate</a>   | Manage Azure Cognitive Services accounts.                  | Core | GA     |
| <a href="#">az cognitiveservices account list</a>              | Manage Azure Cognitive Services accounts.                  | Core | GA     |
| <a href="#">az cognitiveservices account list-deleted</a>      | List soft-deleted Azure Cognitive Services accounts.       | Core | GA     |
| <a href="#">az cognitiveservices account list-kinds</a>        | List all valid kinds for Azure Cognitive Services account. | Core | GA     |
| <a href="#">az cognitiveservices account list-models</a>       | Manage Azure Cognitive Services accounts.                  | Core | GA     |
| <a href="#">az cognitiveservices account list-skus</a>         | Manage Azure Cognitive Services accounts.                  | Core | GA     |
| <a href="#">az cognitiveservices account list-usage</a>        | List usages for Azure Cognitive Services account.          | Core | GA     |
| <a href="#">az cognitiveservices account network-rule</a>      | Manage network rules.                                      | Core | GA     |

| Name                                                             | Description                                              | Type | Status     |
|------------------------------------------------------------------|----------------------------------------------------------|------|------------|
| <a href="#">az cognitiveservices account network-rule add</a>    | Add a network rule.                                      | Core | GA         |
| <a href="#">az cognitiveservices account network-rule list</a>   | List network rules.                                      | Core | GA         |
| <a href="#">az cognitiveservices account network-rule remove</a> | Remove a network rule.                                   | Core | GA         |
| <a href="#">az cognitiveservices account purge</a>               | Purge a soft-deleted Azure Cognitive Services account.   | Core | GA         |
| <a href="#">az cognitiveservices account recover</a>             | Recover a soft-deleted Azure Cognitive Services account. | Core | GA         |
| <a href="#">az cognitiveservices account show</a>                | Manage Azure Cognitive Services accounts.                | Core | GA         |
| <a href="#">az cognitiveservices account show-deleted</a>        | Show a soft-deleted Azure Cognitive Services account.    | Core | GA         |
| <a href="#">az cognitiveservices account update</a>              | Manage Azure Cognitive Services accounts.                | Core | GA         |
| <a href="#">az cognitiveservices commitment-tier</a>             | Manage commitment tiers for Azure Cognitive Services.    | Core | GA         |
| <a href="#">az cognitiveservices commitment-tier list</a>        | Show all commitment tiers for Azure Cognitive Services.  | Core | GA         |
| <a href="#">az cognitiveservices list</a>                        | Manage Azure Cognitive Services accounts.                | Core | Deprecated |
| <a href="#">az cognitiveservices model</a>                       | Manage model for Azure Cognitive Services.               | Core | GA         |
| <a href="#">az cognitiveservices model list</a>                  | Show all models for Azure Cognitive Services.            | Core | GA         |
| <a href="#">az cognitiveservices usage</a>                       | Manage usage for Azure Cognitive Services.               | Core | GA         |
| <a href="#">az cognitiveservices usage list</a>                  | Show all usages for Azure Cognitive Services.            | Core | GA         |

## az cognitiveservices list

 Edit

Deprecated

This command has been deprecated and will be removed in a future release. Use 'az cognitiveservices account list' instead.

Manage Azure Cognitive Services accounts.

This article lists the Azure CLI commands for Azure Cognitive Services account and subscription management only. Refer to the documentation at <https://docs.microsoft.com/azure/cognitive-services/> for individual services to learn how to use the APIs and supported SDKs.

Azure CLI

```
az cognitiveservices list [--resource-group]
```

## Examples

List all the Cognitive Services accounts in a resource group.

Azure CLI

```
az cognitiveservices list -g MyResourceGroup
```

## Optional Parameters

### --resource-group -g

Name of resource group. You can configure the default group using `az configure --defaults group=<name>`.

### ▼ Global Parameters

#### --debug

Increase logging verbosity to show all debug logs.

#### --help -h

Show this help message and exit.

## --only-show-errors

Only show errors, suppressing warnings.

## --output -o

Output format.

Accepted values: json, jsonc, none, table, tsv, yaml, yamlc

Default value: json

## --query

JMESPath query string. See <http://jmespath.org/> for more information and examples.

## --subscription

Name or ID of subscription. You can configure the default subscription using `az account set -s NAME_OR_ID`.

## --verbose

Increase logging verbosity. Use --debug for full debug logs.

### Collaborate with us on GitHub

The source for this content can be found on GitHub, where you can also create and review issues and pull requests. For more information, see [our contributor guide](#).



### Azure CLI feedback

Azure CLI is an open source project. Select a link to provide feedback:

 [Open a documentation issue](#)

 [Provide product feedback](#)

# AzureRM.CognitiveServices

Reference

This topic displays help topics for the Azure Cognitive Services cmdlets.

## Cognitive Services

 Expand table

[Get-AzureRmCognitiveServicesAccount](#)

Gets an account.

### Warning

The AzureRM PowerShell module has been officially deprecated as of February 29, 2024. Users are advised to migrate from AzureRM to the Az PowerShell module to ensure continued support and updates.

Although the AzureRM module may still function, it's no longer maintained or supported, placing any continued use at the user's discretion and risk. Please refer to our [migration resources](#) ↗ for guidance on transitioning to the Az module.

[Get-AzureRmCognitiveServicesAccountKey](#)

Gets the API keys for an account.

### Warning

The AzureRM PowerShell module has been officially deprecated as of February 29, 2024. Users are advised to migrate from AzureRM to the Az PowerShell module to ensure continued support and updates.

Although the AzureRM module may still function, it's no longer maintained or

supported, placing any continued use at the user's discretion and risk. Please refer to our [migration resources](#) for guidance on transitioning to the Az module.

`Get-AzureRmCognitiveServicesAccountSkus`

Gets the available SKUs for an account.

**⚠️ Warning**

The AzureRM PowerShell module has been officially deprecated as of February 29, 2024. Users are advised to migrate from AzureRM to the Az PowerShell module to ensure continued support and updates.

Although the AzureRM module may still function, it's no longer maintained or supported, placing any continued use at the user's discretion and risk. Please refer to our [migration resources](#) for guidance on transitioning to the Az module.

`Get-AzureRmCognitiveServicesAccountType`

Gets the available Cognitive Services Account Types.

**⚠️ Warning**

The AzureRM PowerShell module has been officially deprecated as of February 29, 2024. Users are advised to migrate from AzureRM to the Az PowerShell module to ensure continued support and updates.

Although the AzureRM module may still function, it's no longer maintained or supported, placing any continued use at the user's discretion and risk. Please refer to our [migration resources](#) for

guidance on transitioning to the Az module.

## Get-AzureRmCognitiveServicesAccountUsage

Get current usages for a Cognitive Services account.

### ⚠ Warning

The AzureRM PowerShell module has been officially deprecated as of February 29, 2024. Users are advised to migrate from AzureRM to the Az PowerShell module to ensure continued support and updates.

Although the AzureRM module may still function, it's no longer maintained or supported, placing any continued use at the user's discretion and risk. Please refer to our [migration resources](#) ↗ for guidance on transitioning to the Az module.

## New-AzureRmCognitiveServicesAccount

Creates a Cognitive Services account.

### ⚠ Warning

The AzureRM PowerShell module has been officially deprecated as of February 29, 2024. Users are advised to migrate from AzureRM to the Az PowerShell module to ensure continued support and updates.

Although the AzureRM module may still function, it's no longer maintained or supported, placing any continued use at the user's discretion and risk. Please refer to our [migration resources](#) ↗ for guidance on transitioning to the Az module.

## New-AzureRmCognitiveServicesAccountKey

Regenerates an account key.

 **Warning**

The AzureRM PowerShell module has been officially deprecated as of February 29, 2024. Users are advised to migrate from AzureRM to the Az PowerShell module to ensure continued support and updates.

Although the AzureRM module may still function, it's no longer maintained or supported, placing any continued use at the user's discretion and risk. Please refer to our [migration resources](#) for guidance on transitioning to the Az module.

[Remove-AzureRmCognitiveServicesAccount](#)

Deletes a Cognitive Services account.

 **Warning**

The AzureRM PowerShell module has been officially deprecated as of February 29, 2024. Users are advised to migrate from AzureRM to the Az PowerShell module to ensure continued support and updates.

Although the AzureRM module may still function, it's no longer maintained or supported, placing any continued use at the user's discretion and risk. Please refer to our [migration resources](#) for guidance on transitioning to the Az module.

[Set-AzureRmCognitiveServicesAccount](#)

Modifies an account.

 **Warning**

The AzureRM PowerShell module has been officially deprecated as of February 29, 2024. Users are advised to migrate

from AzureRM to the Az PowerShell module to ensure continued support and updates.

Although the AzureRM module may still function, it's no longer maintained or supported, placing any continued use at the user's discretion and risk. Please refer to our [migration resources](#) for guidance on transitioning to the Az module.

# Azure Policy built-in policy definitions for Azure AI services

Article • 08/28/2024

This page is an index of [Azure Policy](#) built-in policy definitions for Azure AI services. For additional Azure Policy built-ins for other services, see [Azure Policy built-in definitions](#).

The name of each built-in policy definition links to the policy definition in the Azure portal. Use the link in the **Version** column to view the source on the [Azure Policy GitHub repo](#).

## Azure AI services

[\[ \] Expand table](#)

| Name<br>(Azure portal)                                                                                     | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Effect(s)             | Version<br>(GitHub)   |
|------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|-----------------------|
| <a href="#">Azure AI Services resources should encrypt data at rest with a customer-managed key (CMK)</a>  | Using customer-managed keys to encrypt data at rest provides more control over the key lifecycle, including rotation and management. This is particularly relevant for organizations with related compliance requirements. This is not assessed by default and should only be applied when required by compliance or restrictive policy requirements. If not enabled, the data will be encrypted using platform-managed keys. To implement this, update the 'Effect' parameter in the Security Policy for the applicable scope. | Audit, Deny, Disabled | <a href="#">2.2.0</a> |
| <a href="#">Azure AI Services resources should have key access disabled (disable local authentication)</a> | Key access (local authentication) is recommended to be disabled for security. Azure OpenAI Studio, typically used in development/testing, requires key access and will not function if key access is disabled. After disabling, Microsoft Entra ID becomes the only access method, which allows maintaining minimum privilege principle and granular control. Learn more at: <a href="https://aka.ms/AI/auth">https://aka.ms/AI/auth</a>                                                                                        | Audit, Deny, Disabled | <a href="#">1.1.0</a> |
| <a href="#">Azure AI Services resources should restrict network access</a>                                 | By restricting network access, you can ensure that only allowed networks can access the service. This can be achieved by configuring network rules so that only applications from allowed networks can access the Azure AI service.                                                                                                                                                                                                                                                                                             | Audit, Deny, Disabled | <a href="#">3.2.0</a> |

| Name<br>(Azure portal)                                                                                             | Description                                                                                                                                                                                                                                                                                                                                                                                                                                | Effect(s)                   | Version<br>(GitHub)     |
|--------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|-------------------------|
| <a href="#">Azure AI Services resources should use Azure Private Link ↗</a>                                        | Azure Private Link lets you connect your virtual network to Azure services without a public IP address at the source or destination. The Private Link platform reduces data leakage risks by handling the connectivity between the consumer and services over the Azure backbone network. Learn more about private links at: <a href="https://aka.ms/AzurePrivateLink/Overview">https://aka.ms/AzurePrivateLink/Overview</a> ↗             | Audit, Disabled             | <a href="#">1.0.0 ↗</a> |
| <a href="#">Cognitive Services accounts should use a managed identity ↗</a>                                        | Assigning a managed identity to your Cognitive Service account helps ensure secure authentication. This identity is used by this Cognitive service account to communicate with other Azure services, like Azure Key Vault, in a secure way without you having to manage any credentials.                                                                                                                                                   | Audit, Deny, Disabled       | <a href="#">1.0.0 ↗</a> |
| <a href="#">Cognitive Services accounts should use customer owned storage ↗</a>                                    | Use customer owned storage to control the data stored at rest in Cognitive Services. To learn more about customer owned storage, visit <a href="https://aka.ms/cogsvc-cmk">https://aka.ms/cogsvc-cmk</a> ↗ .                                                                                                                                                                                                                               | Audit, Deny, Disabled       | <a href="#">2.0.0 ↗</a> |
| <a href="#">Configure Azure AI Services resources to disable local key access (disable local authentication) ↗</a> | Key access (local authentication) is recommended to be disabled for security. Azure OpenAI Studio, typically used in development/testing, requires key access and will not function if key access is disabled. After disabling, Microsoft Entra ID becomes the only access method, which allows maintaining minimum privilege principle and granular control. Learn more at: <a href="https://aka.ms/AI/auth">https://aka.ms/AI/auth</a> ↗ | DeployIfNotExists, Disabled | <a href="#">1.0.0 ↗</a> |
| <a href="#">Configure Cognitive Services accounts to disable local authentication methods ↗</a>                    | Disable local authentication methods so that your Cognitive Services accounts require Azure Active Directory identities exclusively for authentication. Learn more at: <a href="https://aka.ms/cs/auth">https://aka.ms/cs/auth</a> ↗ .                                                                                                                                                                                                     | Modify, Disabled            | <a href="#">1.0.0 ↗</a> |
| <a href="#">Configure Cognitive Services accounts to disable public network access ↗</a>                           | Disable public network access for your Cognitive Services resource so that it's not accessible over the public internet. This can reduce data leakage risks. Learn more at: <a href="https://go.microsoft.com/fwlink/?linkid=2129800">https://go.microsoft.com/fwlink/?linkid=2129800</a> ↗ .                                                                                                                                              | Disabled, Modify            | <a href="#">3.0.0 ↗</a> |
| <a href="#">Configure Cognitive Services accounts with private endpoints ↗</a>                                     | Private endpoints connect your virtual networks to Azure services without a public IP address at the source or destination. By mapping private endpoints to Cognitive Services, you'll reduce the potential for data                                                                                                                                                                                                                       | DeployIfNotExists, Disabled | <a href="#">3.0.0 ↗</a> |

| Name                                                                                                              | Description                                                                                                                                                                                                                                                                                                                                  | Effect(s)                                     | Version  |
|-------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|----------|
| (Azure portal)                                                                                                    |                                                                                                                                                                                                                                                                                                                                              |                                               | (GitHub) |
|                                                                                                                   | leakage. Learn more about private links at: <a href="https://go.microsoft.com/fwlink/?linkid=2129800">https://go.microsoft.com/fwlink/?linkid=2129800</a> .                                                                                                                                                                                  |                                               |          |
| Diagnostic logs in Azure AI services resources should be enabled ↗                                                | Enable logs for Azure AI services resources. This enables you to recreate activity trails for investigation purposes, when a security incident occurs or your network is compromised                                                                                                                                                         | AuditIfNotExists, Disabled                    | 1.0.0 ↗  |
| Enable logging by category group for Cognitive Services (microsoft.cognitiveservices/accounts) to Event Hub ↗     | Resource logs should be enabled to track activities and events that take place on your resources and give you visibility and insights into any changes that occur. This policy deploys a diagnostic setting using a category group to route logs to an Event Hub for Cognitive Services (microsoft.cognitiveservices/accounts).              | DeployIfNotExists, AuditIfNotExists, Disabled | 1.2.0 ↗  |
| Enable logging by category group for Cognitive Services (microsoft.cognitiveservices/accounts) to Log Analytics ↗ | Resource logs should be enabled to track activities and events that take place on your resources and give you visibility and insights into any changes that occur. This policy deploys a diagnostic setting using a category group to route logs to a Log Analytics workspace for Cognitive Services (microsoft.cognitiveservices/accounts). | DeployIfNotExists, AuditIfNotExists, Disabled | 1.1.0 ↗  |
| Enable logging by category group for Cognitive Services (microsoft.cognitiveservices/accounts) to Storage ↗       | Resource logs should be enabled to track activities and events that take place on your resources and give you visibility and insights into any changes that occur. This policy deploys a diagnostic setting using a category group to route logs to a Storage Account for Cognitive Services (microsoft.cognitiveservices/accounts).         | DeployIfNotExists, AuditIfNotExists, Disabled | 1.1.0 ↗  |

## Next steps

- See the built-ins on the [Azure Policy GitHub repo](#).
- Review the [Azure Policy definition structure](#).
- Review [Understanding policy effects](#).

---

## Feedback

Was this page helpful?

 Yes

 No



# Azure AI services support and help options

Article • 08/28/2024

Are you just starting to explore the functionality of Azure AI services? Perhaps you are implementing a new feature in your application. Or after using the service, do you have suggestions on how to improve it? Here are options for where you can get support, stay up-to-date, give feedback, and report bugs for Azure AI services.

## Get solutions to common issues

In the Azure portal, you can find answers to common AI service issues.

1. Go to your Azure AI services resource in the Azure portal. You can find it on the list on this page: [Azure AI services](#). If you're a United States government customer, use the [Azure portal for the United States government](#).
2. In the left pane, under **Help**, select **Support + Troubleshooting**.
3. Describe your issue in the text box, and answer the remaining questions in the form.
4. You'll find Learn articles and other resources that might help you resolve your issue.

## Create an Azure support request

### A

Explore the range of Azure support options and [choose the plan](#) that best fits, whether you're a developer just starting your cloud journey or a large organization deploying business-critical, strategic applications. Azure customers can create and manage support requests in the Azure portal.

To submit a support request for Azure AI services, follow the instructions on the [New support request](#) page in the Azure portal. Select **Cognitive Services** in the **Service type** dropdown field.

## Post a question on Microsoft Q&A

For quick and reliable answers on your technical product questions from Microsoft Engineers, Azure Most Valuable Professionals (MVPs), or our expert community, engage

with us on [Microsoft Q&A](#), Azure's preferred destination for community support.

If you can't find an answer to your problem using search, submit a new question to Microsoft Q&A. Use one of the following tags when you ask your question:

- [Azure AI services](#)

## Vision

- [Azure AI Vision](#)
- [Custom Vision](#)
- [Face](#)
- [Document Intelligence](#)
- [Video Indexer](#)

## Language

- [Immersive Reader](#)
- [Language Understanding \(LUIS\)](#)
- [QnA Maker](#)
- [Language service](#)
- [Translator](#)

## Speech

- [Speech service](#)

## Decision

- [Anomaly Detector](#)
- [Content Moderator](#)
- [Metrics Advisor](#)
- [Personalizer](#)

## Azure OpenAI

- [Azure OpenAI](#)

# Post a question to Stack Overflow



For answers on your developer questions from the largest community developer ecosystem, ask your question on Stack Overflow.

If you do submit a new question to Stack Overflow, please use one or more of the following tags when you create the question:

- [Azure AI services ↗](#)

## Vision

- [Azure AI Vision ↗](#)
- [Custom Vision ↗](#)
- [Face ↗](#)
- [Document Intelligence ↗](#)
- [Video Indexer ↗](#)

## Language

- [Immersive Reader ↗](#)
- [Language Understanding \(LUIS\) ↗](#)
- [QnA Maker ↗](#)
- [Language service ↗](#)
- [Translator ↗](#)

## Speech

- [Speech service ↗](#)

## Decision

- [Anomaly Detector ↗](#)
- [Content Moderator ↗](#)
- [Metrics Advisor ↗](#)
- [Personalizer ↗](#)

## Azure OpenAI

- [Azure OpenAI ↗](#)

# Submit feedback

To request new features, post them on <https://feedback.azure.com> ↗ . Share your ideas for making Azure AI services and its APIs work better for the applications you develop.

- [Azure AI services ↗](#)

## Vision

- [Azure AI Vision](#)
- [Custom Vision](#)
- [Face](#)
- [Document Intelligence](#)
- [Video Indexer](#)

## Language

- [Immersive Reader](#)
- [Language Understanding \(LUIS\)](#)
- [QnA Maker](#)
- [Language service](#)
- [Translator](#)

## Speech

- [Speech service](#)

## Decision

- [Anomaly Detector](#)
- [Content Moderator](#)
- [Metrics Advisor](#)
- [Personalizer](#)

# Stay informed

Staying informed about features in a new release or news on the Azure blog can help you find the difference between a programming error, a service bug, or a feature not yet available in Azure AI services.

- Learn more about product updates, roadmap, and announcements in [Azure Updates](#).
- News about Azure AI services is shared in the [Azure blog](#).
- Join the conversation on [Reddit](#) about Azure AI services.

# Next steps

[What are Azure AI services?](#)

# Feedback

Was this page helpful?

 Yes

 No

[Provide product feedback ↗](#) | [Get help at Microsoft Q&A](#)

# Compare Azure Government and global Azure

Article • 09/03/2024

Microsoft Azure Government uses same underlying technologies as global Azure, which includes the core components of [Infrastructure-as-a-Service \(IaaS\)](#), [Platform-as-a-Service \(PaaS\)](#), and [Software-as-a-Service \(SaaS\)](#). Both Azure and Azure Government have the same comprehensive security controls in place and the same Microsoft commitment on the safeguarding of customer data. Whereas both cloud environments are assessed and authorized at the FedRAMP High impact level, Azure Government provides an extra layer of protection to customers through contractual commitments regarding storage of customer data in the United States and limiting potential access to systems processing customer data to [screened US persons](#). These commitments may be of interest to customers using the cloud to store or process data subject to US export control regulations.

## ⓘ Note

These lists and tables do not include feature or bundle availability in the Azure Government Secret or Azure Government Top Secret clouds. For more information about specific availability for air-gapped clouds, please contact your account team.

## Export control implications

You're responsible for designing and deploying your applications to meet [US export control requirements](#) such as the requirements prescribed in the EAR, ITAR, and DoE 10 CFR Part 810. In doing so, you shouldn't include sensitive or restricted information in Azure resource names, as explained in [Considerations for naming Azure resources](#).

## Guidance for developers

Most of the currently available technical content assumes that applications are being developed on global Azure rather than on Azure Government. For this reason, it's important to be aware of two key differences in applications that you develop for hosting in Azure Government.

- Certain services and features that are in specific regions of global Azure might not be available in Azure Government.
- Feature configurations in Azure Government might differ from those in global Azure.

Therefore, it's important to review your sample code and configurations to ensure that you are building within the Azure Government cloud services environment.

For more information, see [Azure Government developer guide](#).

## ⓘ Note

This article has been updated to use the new Azure PowerShell Az module. You can still use the AzureRM module, which will continue to receive bug fixes until at least December 2020. To learn more about the new Az module and AzureRM compatibility, see [Introducing the new Azure PowerShell Az module](#). For Az module installation instructions, see [Install the Azure Az PowerShell module](#).

You can use AzureCLI or PowerShell to obtain Azure Government endpoints for services you provisioned:

- Use **Azure CLI** to run the `az cloud show` command and provide `AzureUSGovernment` as the name of the target cloud environment. For example,

#### Azure CLI

```
az cloud show --name AzureUSGovernment
```

should get you different endpoints for Azure Government.

- Use a **PowerShell** cmdlet such as [Get-AzEnvironment](#) to get endpoints and metadata for an instance of Azure service. For example,

#### PowerShell

```
Get-AzEnvironment -Name AzureUSGovernment
```

should get you properties for Azure Government. This cmdlet gets environments from your subscription data file.

Table below lists API endpoints in Azure vs. Azure Government for accessing and managing some of the more common services. If you provisioned a service that isn't listed in the table below, see the Azure CLI and PowerShell examples above for suggestions on how to obtain the corresponding Azure Government endpoint.

 [Expand table](#)

| Service category      | Service name                   | Azure Public                     | Azure Government                                     | Notes                                     |
|-----------------------|--------------------------------|----------------------------------|------------------------------------------------------|-------------------------------------------|
| AI + machine learning | Azure Bot Service              | botframework.com                 | botframework.azure.us                                |                                           |
|                       | Azure AI Document Intelligence | cognitiveservices.azure.com      | cognitiveservices.azure.us                           |                                           |
|                       | Azure OpenAI Service           | openai.azure.com                 | openai.azure.us                                      |                                           |
|                       | Computer Vision                | cognitiveservices.azure.com      | cognitiveservices.azure.us                           |                                           |
|                       | Custom Vision                  | cognitiveservices.azure.com      | cognitiveservices.azure.us<br><a href="#">Portal</a> |                                           |
|                       | Content Moderator              | cognitiveservices.azure.com      | cognitiveservices.azure.us                           |                                           |
|                       | Face API                       | cognitiveservices.azure.com      | cognitiveservices.azure.us                           |                                           |
|                       | Language Understanding         | cognitiveservices.azure.com      | cognitiveservices.azure.us<br><a href="#">Portal</a> | Part of <a href="#">Azure AI Language</a> |
|                       | Personalizer                   | cognitiveservices.azure.com      | cognitiveservices.azure.us                           |                                           |
|                       | QnA Maker                      | cognitiveservices.azure.com      | cognitiveservices.azure.us                           | Part of <a href="#">Azure AI Language</a> |
|                       | Speech service                 | See <a href="#">STT API docs</a> | <a href="#">Speech Studio</a>                        |                                           |
|                       |                                |                                  | See <a href="#">Speech service endpoints</a>         |                                           |
|                       |                                |                                  | <a href="#">Speech translation endpoints</a>         |                                           |

| Service category   | Service name                  | Azure Public                            | Azure Government                                                                                                                                                                                                              | Notes                                              |
|--------------------|-------------------------------|-----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|
|                    |                               |                                         | Virginia:<br><a href="https://usgovvirginia.s2s.speech.azure.us">https://usgovvirginia.s2s.speech.azure.us</a><br>Arizona:<br><a href="https://usgovarizona.s2s.speech.azure.us">https://usgovarizona.s2s.speech.azure.us</a> |                                                    |
|                    | Text Analytics                | cognitiveservices.azure.com             | cognitiveservices.azure.us                                                                                                                                                                                                    | Part of Azure AI Language                          |
|                    | Translator                    | See <a href="#">Translator API docs</a> | cognitiveservices.azure.us                                                                                                                                                                                                    |                                                    |
| Analytics          | Azure HDInsight               | azurehdinsight.net                      | azurehdinsight.us                                                                                                                                                                                                             |                                                    |
|                    | Event Hubs                    | servicebus.windows.net                  | servicebus.usgovcloudapi.net                                                                                                                                                                                                  |                                                    |
|                    | Power BI                      | app.powerbi.com                         | app.powerbigov.us                                                                                                                                                                                                             | Power BI US Gov ↗                                  |
| Compute            | Batch                         | batch.azure.com                         | batch.usgovcloudapi.net                                                                                                                                                                                                       |                                                    |
|                    | Cloud Services                | cloudapp.net                            | usgovcloudapp.net                                                                                                                                                                                                             |                                                    |
| Containers         | Azure Service Fabric          | clouddapp.azure.com                     | clouddapp.usgovcloudapi.net                                                                                                                                                                                                   |                                                    |
|                    | Container Registry            | azurecr.io                              | azurecr.us                                                                                                                                                                                                                    |                                                    |
| Databases          | Azure Cache for Redis         | redis.cache.windows.net                 | redis.cache.usgovcloudapi.net                                                                                                                                                                                                 | See <a href="#">How to connect to other clouds</a> |
|                    | Azure Cosmos DB               | documents.azure.com                     | documents.azure.us                                                                                                                                                                                                            |                                                    |
|                    | Azure Database for MariaDB    | mariadb.database.azure.com              | mariadb.database.usgovcloudapi.net                                                                                                                                                                                            |                                                    |
|                    | Azure Database for MySQL      | mysql.database.azure.com                | mysql.database.usgovcloudapi.net                                                                                                                                                                                              |                                                    |
|                    | Azure Database for PostgreSQL | postgres.database.azure.com             | postgres.database.usgovcloudapi.net                                                                                                                                                                                           |                                                    |
|                    | Azure SQL Database            | database.windows.net                    | database.usgovcloudapi.net                                                                                                                                                                                                    |                                                    |
| Identity           | Microsoft Entra ID            | login.microsoftonline.com               | login.microsoftonline.us                                                                                                                                                                                                      |                                                    |
|                    |                               | certauth.login.microsoftonline.com      | certauth.login.microsoftonline.us                                                                                                                                                                                             |                                                    |
|                    |                               | passwordreset.microsoftonline.com       | passwordreset.microsoftonline.us                                                                                                                                                                                              |                                                    |
| Integration        | Service Bus                   | servicebus.windows.net                  | servicebus.usgovcloudapi.net                                                                                                                                                                                                  |                                                    |
| Internet of Things | Azure IoT Hub                 | azure-devices.net                       | azure-devices.us                                                                                                                                                                                                              |                                                    |

| Service category          | Service name           | Azure Public                           | Azure Government                          | Notes                          |
|---------------------------|------------------------|----------------------------------------|-------------------------------------------|--------------------------------|
| Management and governance | Azure Maps             | atlas.microsoft.com                    | atlas.azure.us                            |                                |
|                           | Notification Hubs      | servicebus.windows.net                 | servicebus.usgovcloudapi.net              |                                |
|                           | Azure Automation       | azure-automation.net                   | azure-automation.us                       |                                |
|                           | Azure Monitor          | mms.microsoft.com                      | oms.microsoft.us                          | Log Analytics workspace portal |
|                           |                        | ods.opinsights.azure.com               | ods.opinsights.azure.us                   | Data collector API             |
|                           |                        | oms.opinsights.azure.com               | oms.opinsights.azure.us                   |                                |
|                           |                        | portal.loganalytics.io                 | portal.loganalytics.us                    |                                |
| Migration                 |                        | api.loganalytics.io                    | api.loganalytics.us                       |                                |
|                           |                        | docs.loganalytics.io                   | docs.loganalytics.us                      |                                |
|                           |                        | adx.monitor.azure.com                  | adx.monitor.azure.us                      | Data Explorer queries          |
|                           | Azure Resource Manager | management.azure.com                   | management.usgovcloudapi.net              |                                |
|                           | Gallery URL            | gallery.azure.com                      | gallery.azure.us                          |                                |
|                           | Microsoft Azure portal | portal.azure.com                       | portal.azure.us                           |                                |
|                           | Microsoft Intune       | enterpriseregistration.windows.net     | enterpriseregistration.microsoftonline.us | Enterprise registration        |
| Networking                |                        | manage.microsoft.com                   | manage.microsoft.us                       | Enterprise enrollment          |
|                           | Azure Site Recovery    | hypervrecoverymanager.windowsazure.com | hypervrecoverymanager.windowsazure.us     | Site Recovery service          |
|                           |                        | backup.windowsazure.com                | backup.windowsazure.us                    | Protection service             |
| Security                  |                        | blob.core.windows.net                  | blob.core.usgovcloudapi.net               | Storing VM snapshots           |
|                           | Traffic Manager        | trafficmanager.net                     | usgovtrafficmanager.net                   |                                |
| Storage                   | Key Vault              | vault.azure.net                        | vault.usgovcloudapi.net                   |                                |
|                           | Managed HSM            | managedhsm.azure.net                   | managedhsm.usgovcloudapi.net              |                                |
| Storage                   | Azure Backup           | backup.windowsazure.com                | backup.windowsazure.us                    |                                |

| Service category               | Service name              | Azure Public                 | Azure Government             | Notes |
|--------------------------------|---------------------------|------------------------------|------------------------------|-------|
| Virtual desktop infrastructure | Blob                      | blob.core.windows.net        | blob.core.usgovcloudapi.net  |       |
|                                | Queue                     | queue.core.windows.net       | queue.core.usgovcloudapi.net |       |
|                                | Table                     | table.core.windows.net       | table.core.usgovcloudapi.net |       |
|                                | File                      | file.core.windows.net        | file.core.usgovcloudapi.net  |       |
| Azure Virtual Desktop          | Azure Virtual Desktop     | See <a href="#">AVD docs</a> | See <a href="#">AVD docs</a> |       |
| Web                            | API Management            | management.azure.com         | management.usgovcloudapi.net |       |
|                                | API Management Gateway    | azure-api.net                | azure-api.us                 |       |
|                                | API Management management | management.azure-api.net     | management.azure-api.us      |       |
|                                | API Management Portal     | portal.azure-api.net         | portal.azure-api.us          |       |
|                                | App Configuration         | azconfig.io                  | azconfig.azure.us            |       |
|                                | App Service               | azurewebsites.net            | azurewebsites.us             |       |
|                                | Azure AI Search           | search.windows.net           | search.azure.us              |       |
|                                | Azure Functions           | azurewebsites.net            | azurewebsites.us             |       |

## Service availability

Microsoft's goal for Azure Government is to match service availability in Azure. For service availability in Azure Government, see [Products available by region](#). Services available in Azure Government are listed by category and whether they're Generally Available or available through Preview. If a service is available in Azure Government, that fact isn't reiterated in the rest of this article. Instead, you're encouraged to review [Products available by region](#) for the latest, up-to-date information on service availability.

In general, service availability in Azure Government implies that all corresponding service features are available to you. Variations to this approach and other applicable limitations are tracked and explained in this article based on the main service categories outlined in the [online directory of Azure services](#). Other considerations for service deployment and usage in Azure Government are also provided.

## AI + machine learning

This section outlines variations and considerations when using [Azure Bot Service](#), [Azure Machine Learning](#), and [Cognitive Services](#) in the Azure Government environment. For service availability, see [Products available by region](#).

## Azure Bot Service

The following Azure Bot Service **features aren't currently available** in Azure Government:

- Bot Framework Composer integration
- Channels (due to availability of dependent services)
  - Direct Line Speech Channel
  - Telephony Channel (Preview)
  - Microsoft Search Channel (Preview)
  - Kik Channel (deprecated)

For information on how to deploy Bot Framework and Azure Bot Service bots to Azure Government, see [Configure Bot Framework bots for US Government customers](#).

## Azure Machine Learning

For feature variations and limitations, see [Azure Machine Learning feature availability across cloud regions](#).

## Azure AI services: Content Moderator

The following Content Moderator **features aren't currently available** in Azure Government:

- Review UI and Review APIs.

## Azure AI Language Understanding (LUIS)

The following Language Understanding **features aren't currently available** in Azure Government:

- Speech Requests
- Prebuilt Domains

Azure AI Language Understanding (LUIS) is part of [Azure AI Language](#).

## Azure AI Speech

For feature variations and limitations, including API endpoints, see [Speech service in sovereign clouds](#).

## Azure AI services: OpenAI Service

For feature variations and limitations see [Azure OpenAI in Azure Gov](#).

## Azure AI services: Translator

For feature variations and limitations, including API endpoints, see [Translator in sovereign clouds](#).

## Analytics

This section outlines variations and considerations when using Analytics services in the Azure Government environment. For service availability, see [Products available by region](#).

## Azure HDInsight

For secured virtual networks, you'll want to allow network security groups (NSGs) access to certain IP addresses and ports. For Azure Government, you should allow the following IP addresses (all with an Allowed port of 443):

[Expand table](#)

| Region          | Allowed IP addresses                                          | Allowed port |
|-----------------|---------------------------------------------------------------|--------------|
| US DoD Central  | 52.180.249.174<br>52.180.250.239                              | 443          |
| US DoD East     | 52.181.164.168<br>52.181.164.151                              | 443          |
| US Gov Texas    | 52.238.116.212<br>52.238.112.86                               | 443          |
| US Gov Virginia | 13.72.49.126<br>13.72.55.55<br>13.72.184.124<br>13.72.190.110 | 443          |
| US Gov Arizona  | 52.127.3.176<br>52.127.3.178                                  | 443          |

For a demo on how to build data-centric solutions on Azure Government using HDInsight, see Azure AI services, HDInsight, and Power BI on Azure Government.

## Power BI

For usage guidance, feature variations, and limitations, see [Power BI for US government customers](#). For a demo on how to build data-centric solutions on Azure Government using Power BI, see Azure AI services, HDInsight, and Power BI on Azure Government.

## Power BI Embedded

To learn how to embed analytical content within your business process application, see [Tutorial: Embed a Power BI content into your application for national clouds](#).

## Databases

This section outlines variations and considerations when using Databases services in the Azure Government environment. For service availability, see [Products available by region](#).

## Azure Database for MySQL

The following Azure Database for MySQL features aren't currently available in Azure Government:

- Advanced Threat Protection

## Azure Database for PostgreSQL

For Flexible Server availability in Azure Government regions, see [Azure Database for PostgreSQL – Flexible Server](#).

The following Azure Database for PostgreSQL features aren't currently available in Azure Government:

- Azure Cosmos DB for PostgreSQL, formerly Azure Database for PostgreSQL – Hyperscale (Citus). For more information about supported regions, see [Regional availability for Azure Cosmos DB for PostgreSQL](#).
- The following features of the Single Server deployment option
  - Advanced Threat Protection
  - Backup with long-term retention

## Developer tools

This section outlines variations and considerations when using Developer tools in the Azure Government environment. For service availability, see [Products available by region](#).

## Enterprise Dev/Test subscription offer

- Enterprise Dev/Test subscription offer in existing or separate tenant is currently available only in Azure public as documented in [Azure EA portal administration](#).

## Identity

This section outlines variations and considerations when using Identity services in the Azure Government environment. For service availability, see [Products available by region](#).

## Microsoft Entra ID P1 and P2

For feature variations and limitations, see [Cloud feature availability](#).

For information on how to use Power BI capabilities for collaboration between Azure and Azure Government, see [Cross-cloud B2B](#).

The following features have known limitations in Azure Government:

- Limitations with B2B Collaboration in supported Azure US Government tenants:
  - For more information about B2B collaboration limitations in Azure Government and to find out if B2B collaboration is available in your Azure Government tenant, see [Microsoft Entra B2B in government and national clouds](#).
- Limitations with multi-factor authentication:
  - Trusted IPs isn't supported in Azure Government. Instead, use Conditional Access policies with named locations to establish when multi-factor authentication should and shouldn't be required based off the user's current IP address.

## Azure Active Directory B2C

Azure Active Directory B2C is **not available** in Azure Government.

## Microsoft Authentication Library (MSAL)

The Microsoft Authentication Library (MSAL) enables developers to acquire security tokens from the Microsoft identity platform to authenticate users and access secured web APIs. For feature variations and limitations, see [National clouds and MSAL](#).

## Management and governance

This section outlines variations and considerations when using Management and Governance services in the Azure Government environment. For service availability, see [Products available by region](#).

## Automation

The following Automation features aren't currently available in Azure Government:

- Automation analytics solution

## Azure Advisor

For feature variations and limitations, see [Azure Advisor in sovereign clouds](#).

## Azure Lighthouse

The following Azure Lighthouse features aren't currently available in Azure Government:

- Managed Service offers published to Azure Marketplace
- Delegation of subscriptions across a national cloud and the Azure public cloud, or across two separate national clouds, isn't supported
- Privileged Identity Management (PIM) feature isn't enabled, for example, just-in-time (JIT) / eligible authorization capability

## Azure Managed Grafana

The following document contains information about Azure Managed Grafana feature availability in Azure Government: [Azure Managed Grafana: Feature availability in sovereign clouds](#).

## Azure Monitor

Azure Monitor enables the same features in both Azure and Azure Government.

- System Center Operations Manager 2019 is supported equally well in both Azure and Azure Government.

The following options are available for previous versions of System Center Operations Manager:

- Integrating System Center Operations Manager 2016 with Azure Government requires an updated Advisor management pack that is included with Update Rollup 2 or later.
- System Center Operations Manager 2012 R2 requires an updated Advisor management pack included with Update Rollup 3 or later.

For more information, see [Connect Operations Manager to Azure Monitor](#).

### Frequently asked questions

- Can I migrate data from Azure Monitor logs in Azure to Azure Government?
  - No. It isn't possible to move data or your workspace from Azure to Azure Government.
- Can I switch between Azure and Azure Government workspaces from the Operations Management Suite portal?
  - No. The portals for Azure and Azure Government are separate and don't share information.

## Application Insights

Application Insights (part of Azure Monitor) enables the same features in both Azure and Azure Government. This section describes the supplemental configuration that is required to use Application Insights in Azure Government.

**Visual Studio** – In Azure Government, you can enable monitoring on your ASP.NET, ASP.NET Core, Java, and Node.js based applications running on Azure App Service. For more information, see [Application monitoring for Azure App Service overview](#). In Visual Studio, go to Tools|Options|Accounts|Registered Azure Clouds|Add New Azure Cloud and select Azure US Government as the Discovery endpoint. After that, adding an account in File|Account Settings will prompt you for which cloud you want to add from.

**SDK endpoint modifications** – In order to send data from Application Insights to an Azure Government region, you'll need to modify the default endpoint addresses that are used by the Application Insights SDKs. Each SDK requires slightly different modifications, as described in [Application Insights overriding default endpoints](#).

**Firewall exceptions** – Application Insights uses several IP addresses. You might need to know these addresses if the app that you're monitoring is hosted behind a firewall. For more information, see [IP addresses used by Azure Monitor](#) from where you can download Azure Government IP addresses.

! Note

Although these addresses are static, it's possible that we'll need to change them from time to time. All Application Insights traffic represents outbound traffic except for availability monitoring and webhooks, which require inbound firewall rules.

You need to open some **outgoing ports** in your server's firewall to allow the Application Insights SDK and/or Status Monitor to send data to the portal:

 Expand table

| Purpose   | URL                       | IP address  | Ports |
|-----------|---------------------------|-------------|-------|
| Telemetry | dc.applicationinsights.us | 23.97.4.113 | 443   |

## Cost Management and Billing

The following Azure Cost Management + Billing features aren't currently available in Azure Government:

- Cost Management + Billing for cloud solution providers (CSPs)

## Media

This section outlines variations and considerations when using Media services in the Azure Government environment. For service availability, see [Products available by region](#).

## Media Services

For Azure Media Services v3 feature variations in Azure Government, see [Azure Media Services v3 clouds and regions availability](#).

## Migration

This section outlines variations and considerations when using Migration services in the Azure Government environment. For service availability, see [Products available by region](#).

## Azure Migrate

The following Azure Migrate features aren't currently available in Azure Government:

- Containerizing Java Web Apps on Apache Tomcat (on Linux servers) and deploying them on Linux containers on App Service.
- Containerizing Java Web Apps on Apache Tomcat (on Linux servers) and deploying them on Linux containers on Azure Kubernetes Service (AKS).
- Containerizing ASP.NET apps and deploying them on Windows containers on AKS.
- Containerizing ASP.NET apps and deploying them on Windows containers on App Service.
- You can only create assessments for Azure Government as target regions and using Azure Government offers.

For more information, see [Azure Migrate support matrix](#). For a list of Azure Government URLs needed by the Azure Migrate appliance when connecting to the internet, see [Azure Migrate appliance URL access](#).

## Networking

This section outlines variations and considerations when using Networking services in the Azure Government environment. For service availability, see [Products available by region](#).

### Azure ExpressRoute

For an overview of ExpressRoute, see [What is Azure ExpressRoute?](#). For an overview of how **BGP communities** are used with ExpressRoute in Azure Government, see [BGP community support in National Clouds](#).

### Azure Front Door

Azure Front Door (AFD) Standard and Premium tiers are available in general availability in Azure Government regions US Gov Arizona and US Gov Texas. The following Azure Front Door feature isn't supported in Azure Government:

- Managed certificate for enabling HTTPS; instead use your own certificate.

### Private Link

- For Private Link services availability, see [Azure Private Link availability](#).
- For Private DNS zone names, see [Azure Private Endpoint DNS configuration](#).

### Traffic Manager

Traffic Manager health checks can originate from certain IP addresses for Azure Government. Review the [IP addresses in the JSON file](#) to ensure that incoming connections from these IP addresses are allowed at the endpoints to check its health status.

## Security

This section outlines variations and considerations when using Security services in the Azure Government environment. For service availability, see [Products available by region](#).

### Microsoft Defender for Endpoint

For feature variations and limitations, see [Microsoft Defender for Endpoint for US Government customers](#).

### Microsoft Defender for IoT

For feature variations and limitations, see [Cloud feature availability for US Government customers](#).

## Azure Information Protection

Azure Information Protection Premium is part of the [Enterprise Mobility + Security](#) suite. For details on this service and how to use it, see [Azure Information Protection Premium Government Service Description](#).

## Microsoft Defender for Cloud

For feature variations and limitations, see [Cloud feature availability for US Government customers](#).

## Microsoft Sentinel

For feature variations and limitations, see [Cloud feature availability for US Government customers](#).

## Storage

This section outlines variations and considerations when using Storage services in the Azure Government environment. For service availability, see [Products available by region](#).

## Azure NetApp Files

For Azure NetApp Files feature availability in Azure Government and how to access the Azure NetApp Files service within Azure Government, see [Azure NetApp Files for Azure Government](#).

## Azure Import/Export

With Import/Export jobs for US Gov Arizona or US Gov Texas, the mailing address is for US Gov Virginia. The data is loaded into selected storage accounts from the US Gov Virginia region. For all jobs, we recommend that you rotate your storage account keys after the job is complete to remove any access granted during the process. For more information, see [Manage storage account access keys](#).

## Web

This section outlines variations and considerations when using Web services in the Azure Government environment. For service availability, see [Products available by region](#).

## API Management

The following API Management **features aren't currently available** in Azure Government:

- Azure AD B2C integration

## App Service

The following App Service **resources aren't currently available** in Azure Government:

- App Service Certificate
- App Service Managed Certificate
- App Service Domain

The following App Service features aren't currently available in Azure Government:

- Deployment
  - Deployment options: only Local Git Repository and External Repository are available

## Azure Functions

When connecting your Functions app to Application Insights in Azure Government, make sure you use [APPLICATIONINSIGHTS\\_CONNECTION\\_STRING](#), which lets you customize the Application Insights endpoint.

## Next steps

Learn more about Azure Government:

- [Acquiring and accessing Azure Government ↗](#)
- [Azure Government overview](#)
- [Azure support for export controls](#)
- [Azure Government compliance](#)
- [Azure Government security](#)
- [Azure guidance for secure isolation](#)

Start using Azure Government:

- [Guidance for developers](#)
- [Connect with the Azure Government portal](#)

---

## Feedback

Was this page helpful?



[Get help at Microsoft Q&A](#)