



Azure AI on containers Global AI community

11 avril 2025



global AI community

Virtuelle Francophone

MS Azure AlknowEDGE

déployer et gérer des services Azure AI
« déconnectés » pour l'analyse de
documents



Serge RETKOWSKY

France AI Global Black Belt
@Microsoft



Alibek JAKUPOV

France AI Global Black Belt
@Microsoft



Paul PETON

AI Architect @Expertime
Microsoft MVP IA



Live vendredi 11/04/2025 à partir de 13h

Agenda

Azure AI Services

Enterprise Challenges

Azure AI containers

Demos

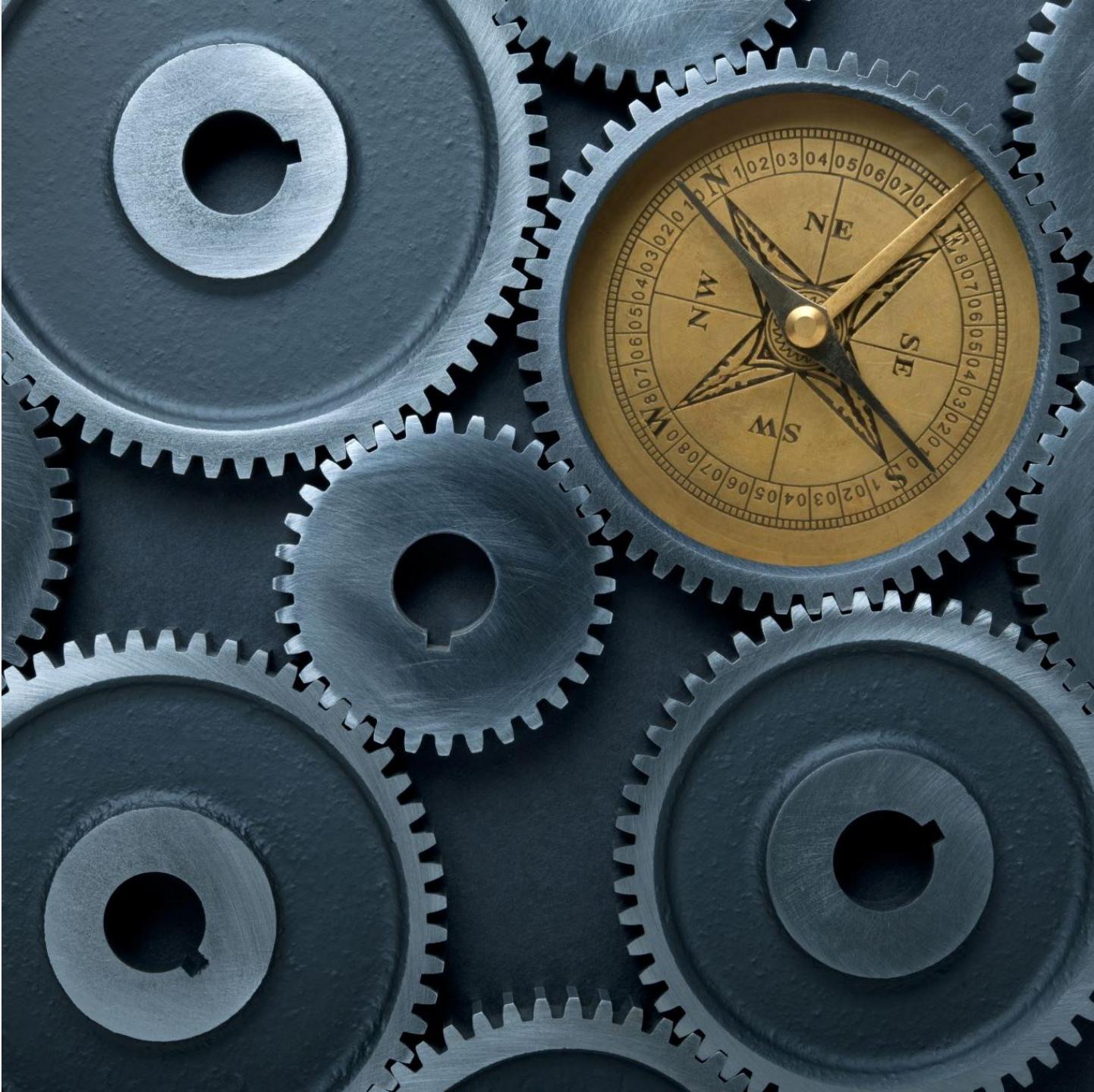
Accelerator

Customers

Pricing

Documentation

Enterprise challenges



Enterprise Challenges

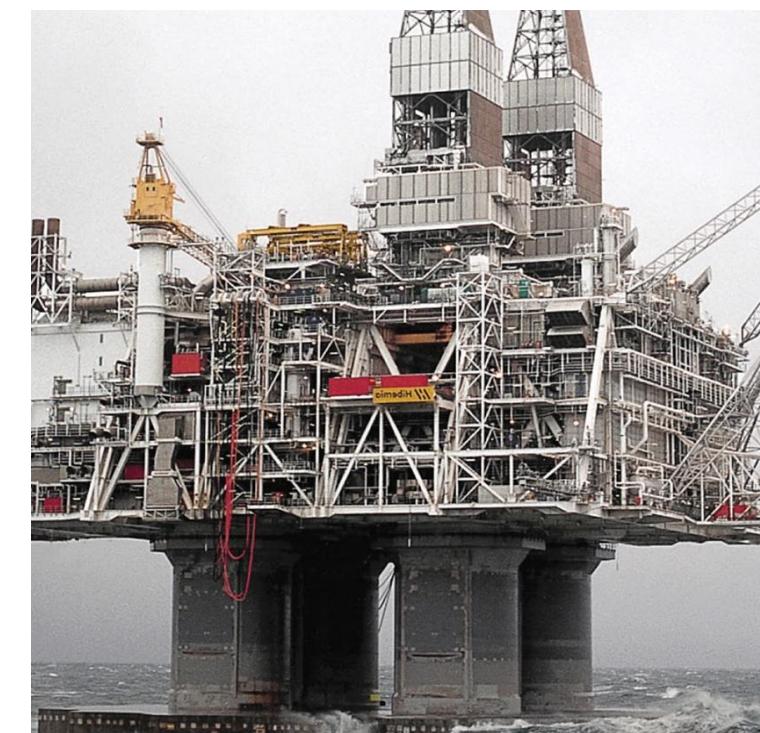
Data security & privacy concerns



Unable to upload all audio/text data to the cloud



Low bandwidth or intermittent connectivity



Overview of AI services containers





Azure AI Foundry



Copilot Studio



Visual Studio



GitHub



Azure AI
Foundry SDK

Model Catalog

Foundational models

Open-source models

Task models

Industry models

Azure
OpenAI Service

Azure
AI Search

Azure AI
Agent Service

Azure AI
Content Safety

Azure Machine
Learning

Evaluations

Customization

Governance

Monitoring

Observability



Azure AI Containers



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.



Azure AI containers allow developers to use the same intelligent APIs that are available in Azure, but with the **benefits of containerization**.



Container services offer **similar feature capabilities as the corresponding cloud service**. Customers can deploy the containers on-premise. The core AI technology, pricing tiers, API keys, and API signature are the same between the container and the corresponding cloud services.

Azure AI task-specific models and capabilities

Speech



Enable every device to hear and speak

- Speech to text
- Text to speech
- Custom Neural Voice
- Speech Translation
- Speaker Diarization
- Language Identification

Translator



Understand any language

- Text translation
- Document translation
- Microsoft Translator Pro (preview)
- Embedded translation

Language



Extract information & process intents

- Summarization
- Entity extraction (PII/PHI)
- Conversational AI
- Text analytics
- Sentiment mining

Content Understanding



Understand multimodal contents

- Document Intelligence
- Image analysis
- Video analysis
- Audio analysis
- Facial recognition

Responsible AI



Ensure AI use remains safe

- Content safety filters
- Safety evaluations

AI task-specific models available in



Microsoft
Public Cloud



Microsoft
Government Cloud



Connected
Containers



Disconnected
Containers

Regulatory Compliance



Azure AI Container Capabilities



[Immutable infrastructure](#) | Adapt to changes and avoid configuration drift



[Control over data](#) | Choose where data is processed with drive data, management, identity & security consistency



[Control over model updates](#) | Flexible versioning and control over when to deploy in your environments



[Portable architecture](#) | Create portable app architecture to deploy on Azure, on-prem, and edge endpoints



[High throughput, Low latency](#) | Run apps physically close to data and you control the throughput

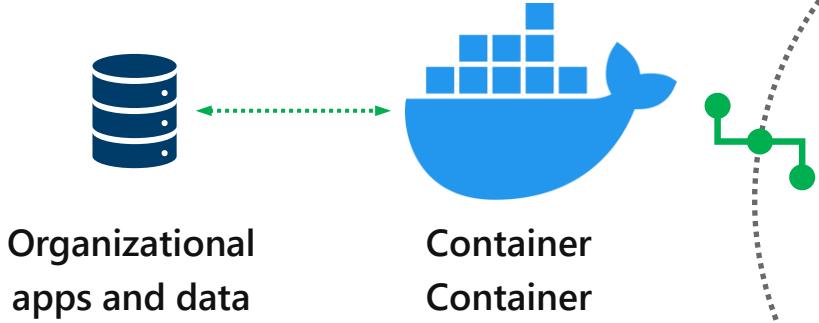


[Scalability](#) | Build your own apps on a scalable cluster foundation based on your business requirements

AI Container environment options

Connected containers

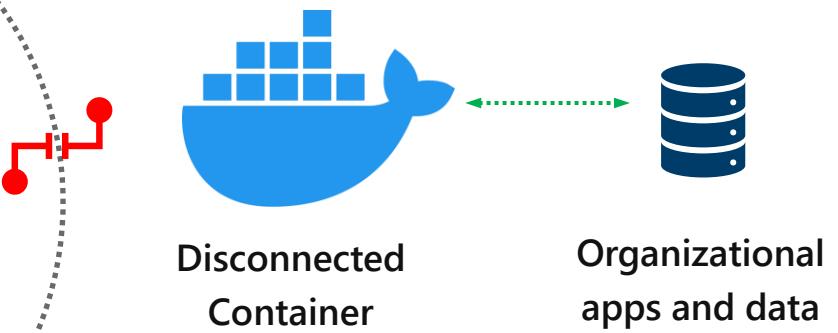
Pay-as-you-go consumption model where **only** billing data is sent to the cloud



Designed for specific security and data governance requirements where billing can be sent

Disconnected containers

Pay up front for a years' worth of consumption with **no** data sent to the cloud



Designed for environments with no cloud connectivity

AI task-specific models available in Containers



Speech

- Speech to Text
- Text to Speech
- Neural Speech to Text
- Custom Text to Speech*
- Speech Language Identification (Preview)*
- Speaker Diarization*



Translator

- Text Translation
- Document Translation
- Transliteration



Language

- Language Detection
- Key Phrase Extraction
- Sentiment Analysis
- PII Redaction
- Summarization
- CLU
- NER
- Custom NER*
- Text Analytics for Health*



Content Understanding

- Document Intelligence includes OCR - designed for documents
- Computer vision OCR optimized for images*



Responsible AI

Content Safety (limited access preview)

* Available only in connected containers

How to apply for a disconnected container?

A form is available:

[Azure AI Services Application for
Disconnected containers](#)

**When the request is accepted,
the customer will have access to
commitment tier disconnected
containers from your Azure portal**

Create Document Intelligence ...

Resource group * ⓘ rg-aiservices-demo-dev-eastus-001 Create new

Instance Details

Region ⓘ France Central

Name * ⓘ docintel-aiservicesdemo-dev

Pricing tier * ⓘ Commitment tier disconnected containers DC0

[View full pricing details](#)

Commitment tier pricing

Select a commitment tier pricing for the capabilities you wish you use. With commitment tier pricing, you will be charged a fixed amount every billing cycle for a given quota. If you need additional quota, you could either increase the number of units.

[Learn more](#)

Custom API * ⓘ Disable Custom API capability

Unit count * ⓘ 1 units

Read * ⓘ 2M transactions/month for 12,960 USD per unit per year

Unit count * ⓘ 1 units

Annual cost (in USD): \$12960

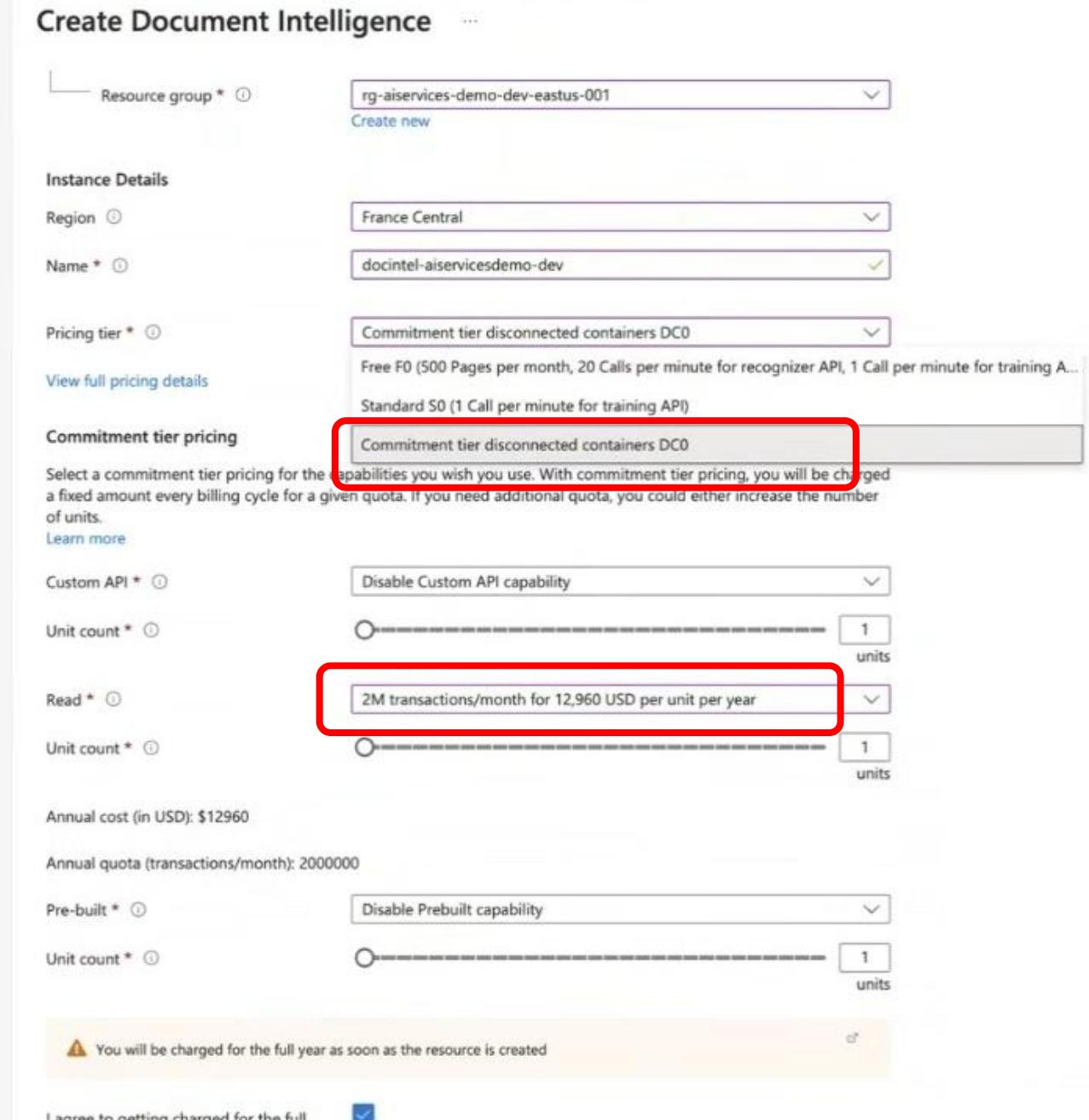
Annual quota (transactions/month): 2000000

Pre-built * ⓘ Disable Prebuilt capability

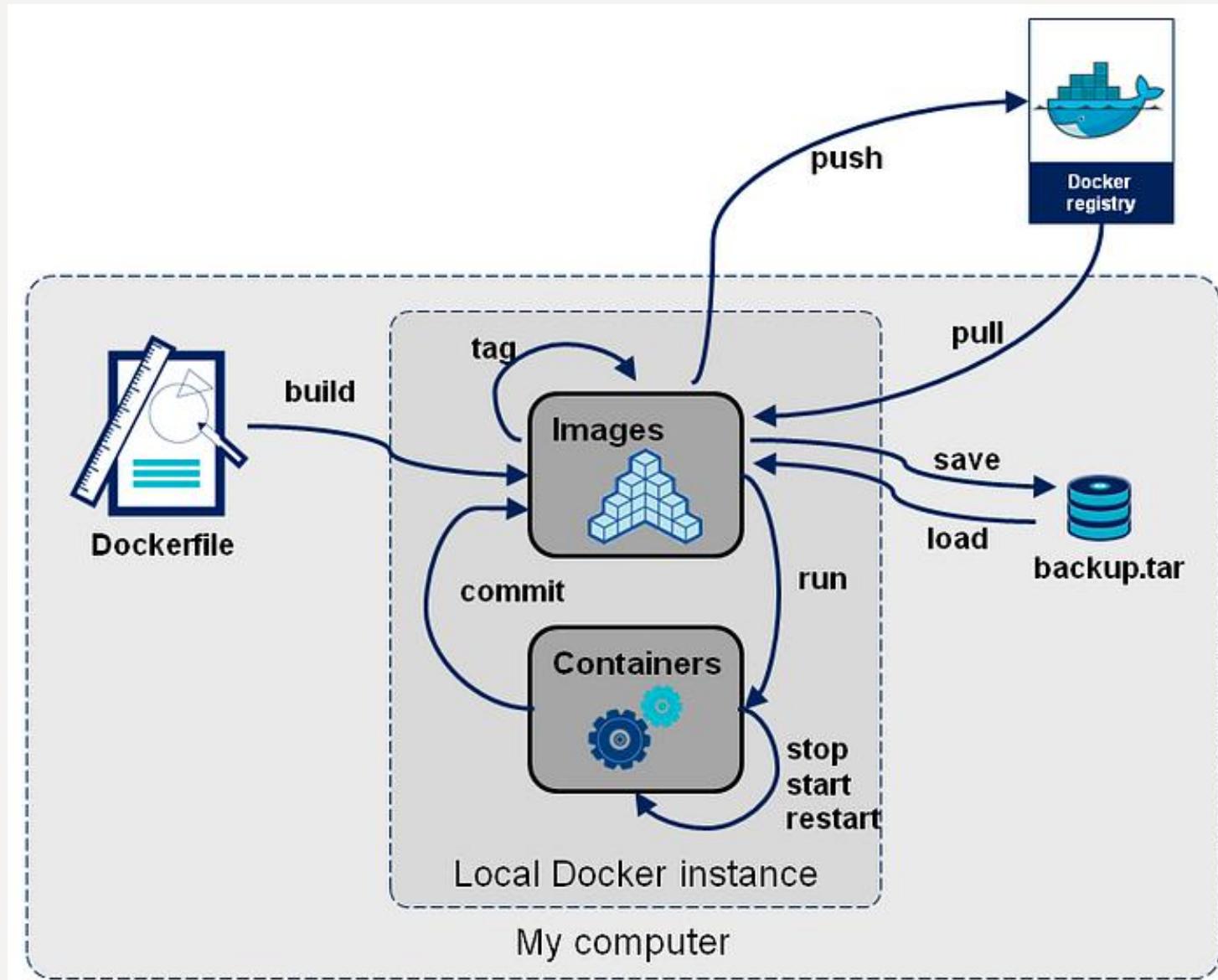
Unit count * ⓘ 1 units

⚠ You will be charged for the full year as soon as the resource is created

I agree to getting charged for the full



Docker



Docker

- The Azure AI services container images can be found in the [Microsoft Artifact Registry](#) catalog.
- Then Run the container (for example) :

```
docker run --rm -it -p 5000:5000 --memory 12g --cpus 4 \
-v /mnt/d/TranslatorContainer:/usr/local/models \
-e apikey={API_KEY} \
-e eula=accept \
-e billing={ENDPOINT_URI} \
-e Languages=en,fr,es,ar,ru \
mcr.microsoft.com/azure-cognitive-services/translator/text-translation:latest
```

Docker

docker desktop PERSONAL

Search Ctrl+K

Ask Gordon BETA

Containers

Images

Volumes

Builds

Docker Hub

Docker Scout

Extensions

Give feedback

View and manage your local and Docker Hub images. [Learn more](#)

Local Docker Hub repositories

7.5 GB / 16.55 GB in use 6 images Last refresh: 53 minutes ago

Search

Name	Tag	Image ID	Created	Size	Actions
swimmingpools	latest	cb83fd9848cd	19 hours ago	344.3 MB	View More Delete
damageclassification	latest	b50e4ee1faa9	19 hours ago	376.66 MB	View More Delete
mcr.microsoft.com/azure-cognitive-services/form-recognizer/layout-4.0	2024-11-30	bec17568a31b	14 days ago	6.05 GB	View More Delete
mcr.microsoft.com/azure-cognitive-services/textanalytics/summarization	cpu	7b62d8976c58	2 months ago	955.59 MB	View More Delete
mcr.microsoft.com/azure-cognitive-services/form-recognizer/read-3.1	latest	be3a0f26c787	5 months ago	5.21 GB	View More Delete
ollama/ollama	latest	96b7667cb536	2 days ago	5.23 GB	View More Delete

Docker

docker desktop PERSONAL

Search Ctrl+K

Ask Gordon BETA

Containers Give feedback

View all your running containers and applications. [Learn more](#)

Container CPU usage i
0.01% / 1600% (16 CPUs available)

Container memory usage i
167.7MB / 30.47GB

Show charts

Search

Only show running containers

<input type="checkbox"/>	Name	Container ID	Image	Port(s)	CPU (%)	Last started	Actions
<input type="checkbox"/>	ollama	257159622cff	ollama/ollama	11434:11434	0%	54 minutes ago	D ⋮ U
<input type="checkbox"/>	trusting_meninsky	a581d318ab04	swimmingpools	80:80	0%	1 hour ago	D ⋮ U
<input type="checkbox"/>	vibrant_mccarthy	615152411e3b	damageclassification	80:80 ↗	0.01%	21 minutes ago	D ⋮ U
<input type="checkbox"/>	ai-knowledge	-	-	-	0%	7 days ago	D ⋮ U

Where to find the Azure AI containers?

Microsoft | Microsoft Artifact Registry

Search Employee Sign In

Microsoft Artifact Registry

Discover official Microsoft artifacts such as container images

Showing 58 results Sort by

Filters

Search Partial Exact

Last Published Date

After Before

Category

- Application Frameworks
- Application Infrastructure
- Application Services
- Base Images
- Databases
- DevOps Tools
- IoT Edge Modules
- Messaging Services
- Monitoring
- Programming Languages

 Microsoft Azure AI services Text Analytics Summarization Last Published: 03/31/2025 Azure AI services Summarization Images Learn more	 Microsoft Azure AI services Text Analytics Sentiment Last Published: 03/31/2025 IoT Edge Modules Learn more	 Microsoft Azure AI services Text Analytics Healthcare Last Published: 03/31/2025 Azure AI services Healthcare Images Learn more	 Microsoft Azure AI services Text Analytics Custom NER Last Published: 03/30/2025 Azure AI services Custom NER Images Learn more	 Microsoft Azure AI services Conversational Language Understanding Last Published: 03/27/2025 Azure AI services Conversational Language Understanding Images Learn more
 Microsoft Azure AI services Document Intelligence Layout 4.0 Last Published: 03/27/2025 Azure AI services Document Intelligence Layout Images Learn more	 Microsoft Azure AI services Neural Text To Speech Last Published: 03/26/2025 Azure AI services Neural Text To Speech Images Learn more	 Microsoft Azure AI services Text Analytics PII Entity Recognition Last Published: 03/25/2025 IoT Edge Modules Learn more	 Microsoft Azure AI services Text Analytics Language Last Published: 03/25/2025 Azure AI services Language Detection Images Learn more	 Microsoft Azure AI services Text Analytics Named Entity Recognition Last Published: 03/25/2025 Azure AI services NER Images Learn more

<https://mcr.microsoft.com/en-us/catalog>

Microsoft Artifact Registry

Microsoft | Microsoft Artifact Registry Search Employee Sign In

Back to catalog

Azure AI services Document Intelligence Layout 4.0 Microsoft

Repository: mcr.microsoft.com/azure-cognitive-services/form-recognizer/layout-4.0

Last Published 03/27/2025

Pulls 34

Project License

Help Documentation Support

About Tags

Filter: By Name, Artifact Type, Pkg, OS, Arch, or Digest

Artifact	Type	Digest	Last Published
2024-11-30	Docker Image	sha256:bec17568a31b01a7e45af56d150bee91a6da28e37d51907d31330e92555f4f6e	03/27/2025
2024-11-30.20250326.1-4778ec1d	Docker Image	sha256:bec17568a31b01a7e45af56d150bee91a6da28e37d51907d31330e92555f4f6e	03/27/2025
latest	Docker Image	sha256:bec17568a31b01a7e45af56d150bee91a6da28e37d51907d31330e92555f4f6e	03/27/2025
2024-11-30.20250319.1-8406c6d7	Docker Image	sha256:adae1d7c2cab7dc162cb5d05231a1a6c1c1dbe29c666fc238cb54f06abdf3076	03/19/2025

< Previous 1 Next >

<https://mcr.microsoft.com/en-us/catalog>

Containers are also available on Docker Hub

The screenshot shows the Docker Hub interface. On the left, there's a sidebar titled "Repositories" with a search bar and a list of Microsoft repositories:

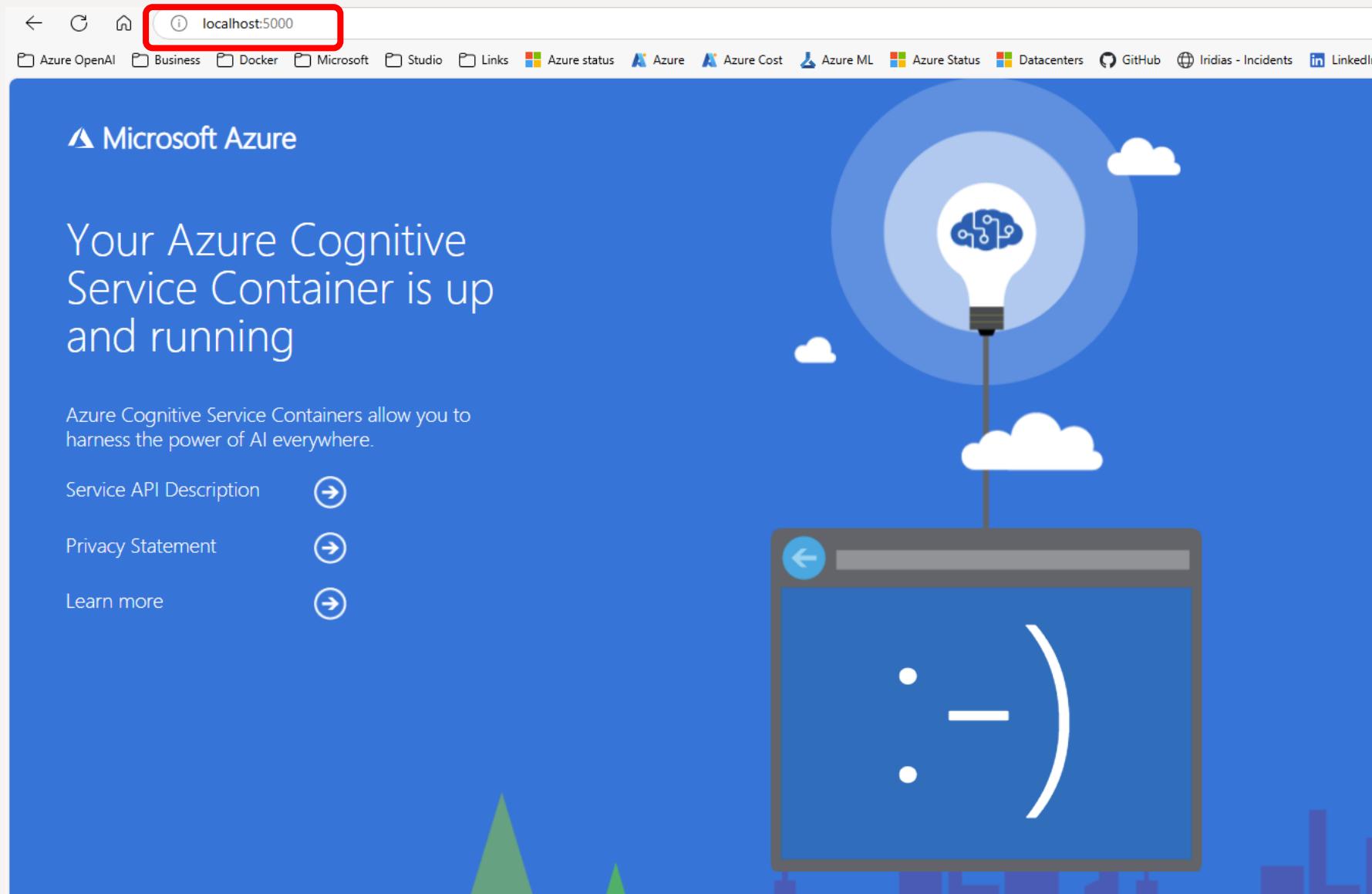
- microsoft/azure-cognitive-services-form-recognizer-layout-4-0** (Verified Publisher) - Updated 23 days ago. Description: Azure AI services Document Intelligence Layout Images.
- microsoft/dotnet-nightly-yarp** (Verified Publisher) - Updated 2 months ago. Description: Official images for YARP (Yet Another Reverse Proxy).
- microsoft/dotnet-nightly-reverse-proxy** (Verified Publisher) - Updated 2 months ago. Description: Official images for the .NET Reverse Proxy (YARP).
- microsoft/azurerm-polymer-ubuntu20-04-py39-cpu-inference** (Verified Publisher)

The main content area shows the details for the **Azure AI services Document Intelligence Layout 4.0** image, also marked as a Verified Publisher. It was created 23 days ago. The image is described as "Azure AI services Document Intelligence Layout Images".

On the right side of the image page, there's a "Docker Pull Command" section with the command `docker pull mcr.microsoft.com/azure-cognitive-services/form-recognizer/layout-4.0`. This section is highlighted with a red box.

At the bottom of the page, the URL <https://hub.docker.com/u/microsoft> is displayed.

How to use Azure Cognitive Services in containers?



How to use Azure Cognitive Services in containers?

The screenshot shows a web browser window with the URL `localhost:5000/api-docs/index.html` highlighted by a red box. The page displays the API documentation for the Form Recognizer service. On the left, a sidebar lists various endpoints: Analyze, Authentication, HealthCheck, Ready, DisconnectedUsage, Status, and InternalWebHook. The `Analyze` endpoint is expanded, showing its description: "Form Recognizer extracts information from forms and images into structured data." Below this, the `Analyze` endpoint is further expanded to show its detailed description: "Analyzes document with model." It also provides a link to the detailed information at <https://westus2.dev.cognitive.microsoft.com/docs/services/form-recognizer-api-2023-07-31>. The `PATH PARAMETERS` section for the `/formrecognizer/documentModels/{modelId}:analyze` endpoint includes a required parameter `modelId` with a default value of "prebuilt-read". The `QUERY PARAMETERS` section is currently empty. On the right side of the page, there is a large dark panel containing a `POST` request sample for the `/formrecognizer/documentModels/{modelId}:analyze` endpoint. The `Payload` section shows a JSON object with a `Content type` of `application/json` and a `urlSource` field set to `"http://example.com/images/test.jpg"`. A "Copy" button is located next to the payload field.

localhost:5000/api-docs/index.html

Analyze > Form Recognizer: prebuilt-read (2023-07-31)

Authentication > Download OpenAPI specification: [Download](#)

HealthCheck > Form Recognizer extracts information from forms and images into structured data.

Ready >

DisconnectedUsage >

Status > Analyze

InternalWebHook >

Analyzes document with model.

See <https://westus2.dev.cognitive.microsoft.com/docs/services/form-recognizer-api-2023-07-31> for detailed information.

PATH PARAMETERS

modelId **required** string
Default: "prebuilt-read"
Value: "prebuilt-read"
Unique model name.

QUERY PARAMETERS

POST /formrecognizer/documentModels/{modelId}:analyze

Request samples

Payload

Content type application/json

Copy

{
 "urlSource": "http://example.com/images/test.jpg"

How to use Azure Cognitive Services in containers?

The screenshot shows a browser window with the URL `localhost:5000/api-docs/index.html` highlighted by a red box. The page displays the API documentation for the `Form Recognizer: prebuilt-read` service. It includes sections for `Analyze`, `Authentication`, `HealthCheck`, `Ready`, `DisconnectedUsage`, `Status`, and `InternalWebHook`. The `Authentication` section has a link to download the OpenAPI specification, which is also highlighted by a red box. Below the sections, there is a detailed description of the `Analyze` endpoint, which analyzes documents with a model. It includes a table for `PATH PARAMETERS` and `QUERY PARAMETERS`.

Analyze > Form Recognizer: prebuilt-read (2023-07-31)

Authentication > Download OpenAPI specification: [Download](#)

HealthCheck > Form Recognizer extracts information from forms and images into structured data.

Ready >

DisconnectedUsage >

Status > Analyze

InternalWebHook >

Analyze document with model.

See <https://westus2.dev.cognitive.microsoft.com/docs/services/form-recognizer-api-2023-0>

PATH PARAMETERS

modelId	string Default: "prebuilt-read" Value: "prebuilt-read" Unique model name.
---------	--

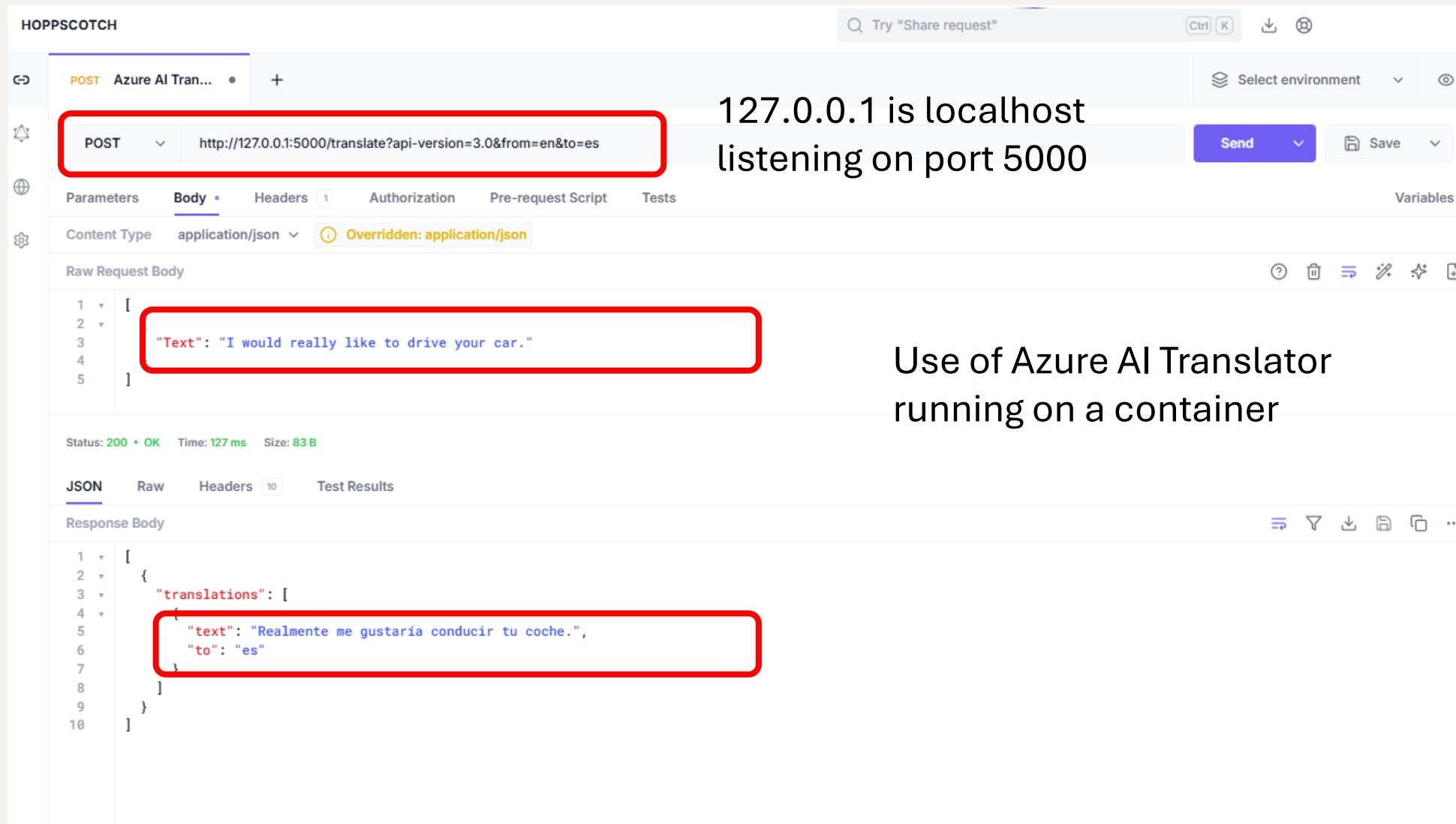
QUERY PARAMETERS

The screenshot shows a code editor displaying the `swagger.json` file. This file defines the OpenAPI specification for the Form Recognizer service. It includes definitions for paths such as `/authentication/renew` (POST method), `/ContainerLiveness` (GET method), and `/ContainerReadiness` (GET method). Each path specifies its tags, parameters, and responses. The `responses` section for each path includes descriptions for both `200` and `503` status codes.

```
swagger.json
{
  "openapi": "3.0.1",
  "info": {
    "title": "Form Recognizer: prebuilt-read",
    "description": "Form Recognizer extracts information from forms and images into structured data.",
    "version": "2023-07-31"
  },
  "paths": {
    "/authentication/renew": {
      "post": {
        "tags": [
          "Authentication"
        ],
        "parameters": [
          {
            "name": "token",
            "in": "query",
            "schema": {
              "type": "string"
            }
          }
        ],
        "responses": {
          "200": {
            "description": "Success"
          }
        }
      }
    },
    "/ContainerLiveness": {
      "get": {
        "tags": [
          "HealthCheck"
        ],
        "responses": {
          "200": {
            "description": "Service is healthy."
          },
          "503": {
            "description": "Service is unhealthy."
          }
        }
      }
    },
    "/ContainerReadiness": {
      "get": {
        "tags": [
          "HealthCheck"
        ],
        "responses": {
          "200": {
            "description": "Service is ready."
          },
          "503": {
            "description": "Service is not ready to serve request."
          }
        }
      }
    }
  }
}
```

Swagger file

An example of an Azure AI service on a container



The screenshot shows the Hopscotch API testing tool interface. A red box highlights the request URL: `http://127.0.0.1:5000/translate?api-version=3.0&from=en&to=es`. Another red box highlights the JSON request body:

```
1 [  
2   "3 ]
```

The response status is 200 OK. A red box highlights the JSON response body:

```
1 [  
2   {  
3     "translations": [  
4       {  
5         "text": "Realmente me gustaría conducir tu coche.",  
6         "to": "es"  
7       }  
8     ]  
9   }  
10 ]
```

Text on the right side of the screen:

127.0.0.1 is localhost
listening on port 5000

Use of Azure AI Translator
running on a container

An example of an Azure AI service on a container



Use of an image classification model built with Azure Custom Vision and deployed as a container to classify an image as a damage or not (two class classification in this example)

HOPSCOTCH

POST Azure AI cont... +

Body

Content Type application/octet-stream

car5.jpg Choisir un fichier car5.jpg

Status: 200 • OK Time: 177 ms Size: 247 B

JSON Raw Headers Test Results

Response Body

```
1 {  
2   "created": "2024-12-19T14:17:57.703092",  
3   "id": "",  
4   "iteration": "",  
5   "predictions": [  
6     {  
7       "boundingBox": null,  
8       "probability": 0.9876402,  
9       "tagId": "",  
10      "tagName": "damage"  
11    },  
12    {  
13      "boundingBox": null,  
14      "probability": 0.01235984,  
15      "tagId": "",  
16      "tagName": "no_damage"  
17    }  
18  ],  
19  "project": ""  
20}
```

An example of an Azure AI service on a container



HOPPSOTCH

POST Azure AI cont... ● +

POST http://127.0.0.1:5000/image

Parameters Body Headers Authorization Pre-request

Content Type application/octet-stream Override

test1.jpg Choisir un fichier test1.jpg

Status: 200 • OK Time: 124 ms Size: 9.25 KB

JSON Raw Headers Test Results

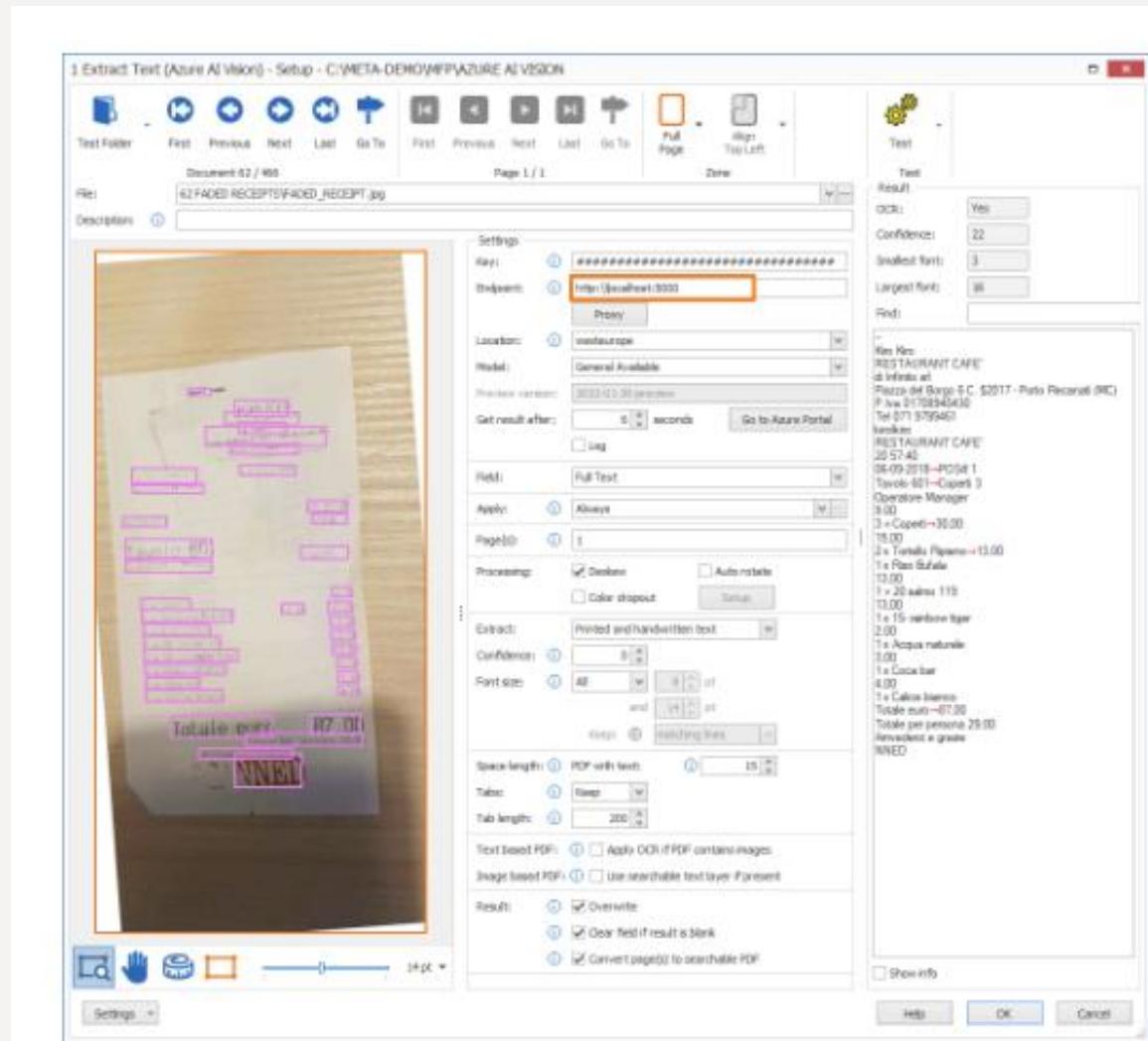
Response Body

```
1  {
2   "created": "2024-12-19T14:27:22.246462",
3   "id": "",
4   "iteration": "",
5   "predictions": [
6     {
7       "boundingBox": {
8         "height": 0.12696499,
9         "left": 0.29223132,
10        "top": 0.0675808,
11        "width": 0.07697695
12      },
13      "probability": 0.99999982,
14      "tagId": 0,
15      "tagName": "pool"
16    },
17    {
18      "boundingBox": {
19        "height": 0.04564708,
20        "left": 0.96127725,
21        "top": 0.05788725,
22        "width": 0.03868902
23      },
24      "probability": 0.99912864,
25      "tagId": 0,
26      "tagName": "pool"
27    },
28    {
29    }
```

A screenshot of the Hoppsotch API testing tool. It shows a POST request to 'http://127.0.0.1:5000/image'. The 'Body' tab is selected, showing a file named 'test1.jpg' being uploaded. The response status is 200 OK with a time of 124 ms and size of 9.25 KB. The 'JSON' tab is selected, displaying a JSON response with two predictions. Each prediction includes a bounding box (height, width, top-left coordinates), probability (e.g., 0.99999982), tag ID (0), and tag name ('pool'). A red box highlights the first prediction's bounding box and probability values.

Use of an object classification model built with Azure Custom Vision and deployed as a container to detect objects like swimming pools from an aerial image.

An example of an Azure AI service application on a container



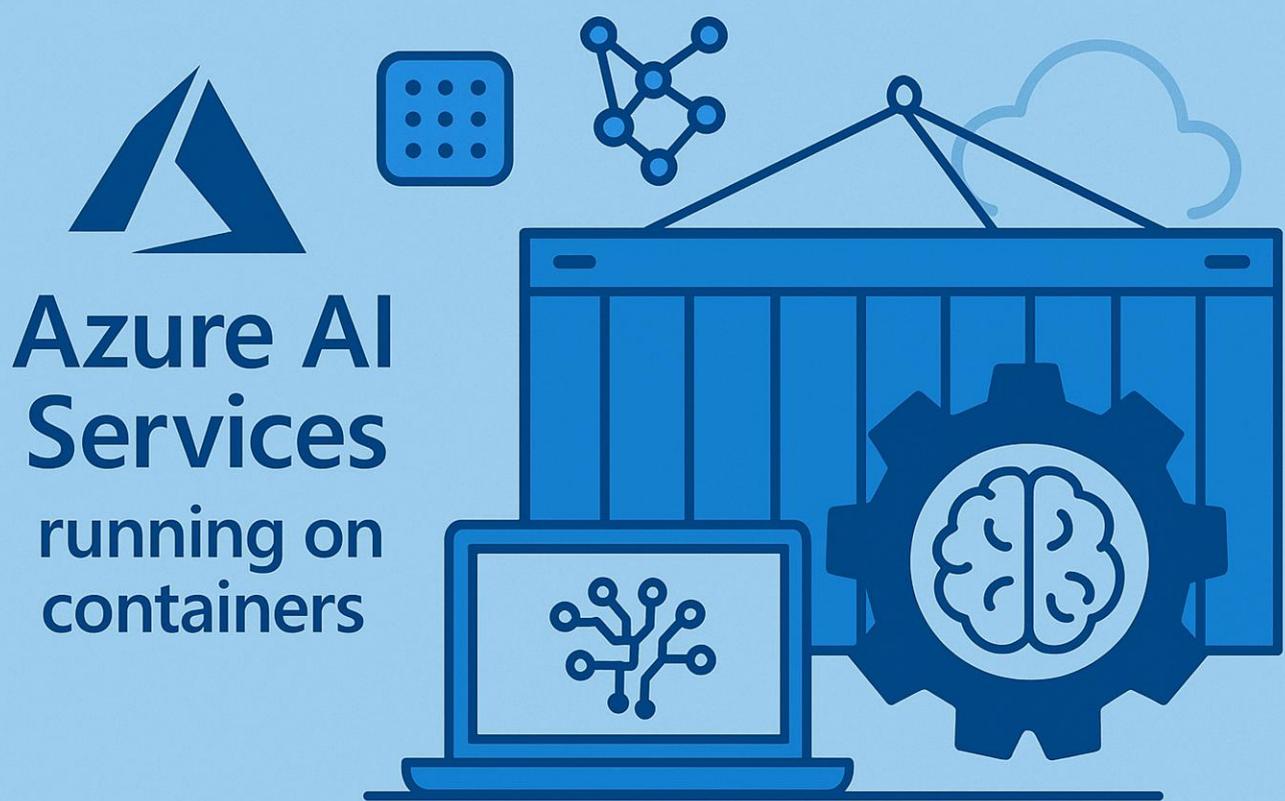
The screenshot shows the 'Extract Text (Azure AI Vision)' application interface. On the left, a receipt image is displayed with several lines of text redacted in red. The main window contains various configuration options:

- Test Folder:** Document 12 / 400
- Document:** 62_FADED RECEIPTS\FADED_RECEIPT.jpg
- Page:** Page 1 / 1
- Result:** OCR: Yes, Confidence: 22, Smallest font: 3, Largest font: 30.
- Endpoint:** https://westeurope.azuredl.ai/ocr/v3.1/analyze (highlighted with an orange box).
- Location:** westeurope
- Model:** General Available
- Process timeout:** 20000 ms (30 seconds)
- Get result after:** 5 seconds
- Ref:** Full Text
- Apply:** Always
- Page(s):** 1
- Processing:** Deskew (checked), Auto rotate, Color dropout, Setup.
- Extract:** Printed and handwritten text.
- Confidence:** 0.0
- Font size:** All (radio button selected), and 8pt, 10pt, 12pt, 14pt, 16pt, 18pt, 20pt, 22pt, 24pt.
- Space length:** POP with lines, 15pt.
- Tab:** Keep, Tab length: 200pt.
- Text based PDF:** Apply OCR if PDF contains images.
- Image based PDF:** Use searchable text layer if present.
- Result:** Overwrite, Gear Red if result is blank, Convert pages to searchable PDF.

On the right, the 'Extract Text (Azure AI Document Intelligence)' configuration window is shown, also with the 'Endpoint' field highlighted in orange:

- Key:** https://westeurope.azuredl.ai/ocr/v3.1/analyze
- Endpoint:** http://localhost:5000 (highlighted with an orange box)
- Location:** westeurope
- Api version:** 2023-07-31 (3.1 GA)
- Prebuilt model:** Invoice
- Get result after:** 6 seconds, Log checkbox.
- Apply:** Always
- Page(s):** 1
- Test page(s):** (empty)
- Processing:** Deskew (checked), Auto rotate (checked), Color dropout, Setup.
- Confidence:** 0
- Tabs:** Keep
- Result:** Overwrite (checked), Clear field if result is blank (checked), Convert page(s) to searchable PDF (checked), Searchable barcodes (unchecked).

Demos



Azure Document Intelligence demo on a container



Demo steps

1. We will utilize Azure Document Intelligence within a container.
2. This allows us to perform OCR on documents locally, with no access to the cloud.
3. Subsequently, we can query the documents using SLM, such as phi-3, operating in a Docker container.



Document Intelligence Layout API & GenAI on a container – French documents

The screenshot shows the Docker Desktop interface with the 'Images' tab selected. The left sidebar includes 'Containers', 'Images' (selected), 'Volumes', 'Builds', 'Docker Scout', and 'Extensions'. The main area displays a table of images under the 'Local' tab, with 'Hub repositories' listed below it. A search bar and filter icons are at the top of the table. The table columns are: Name, Tag, Image ID, Created, Size, and Actions. The table lists ten images, with the last one, 'mcr.microsoft.com/azure-cognitive-services/speechservices/neural-text-to-speech', highlighted by a mouse cursor. The status bar at the bottom shows 'Engine running', system resources (RAM 4.05 GB, CPU 0.00%, Disk 950.04 GB avail. of 1081.10 GB), and a terminal window.

Name	Tag	Image ID	Created	Size	Actions
poolsdetection	latest	4fdb8e47908d	1 day ago	343.3 MB	View Edit Delete
damageclassification	latest	3574d040b60e	7 days ago	375.66 MB	View Edit Delete
ollama/ollama	latest	bb26c2494105	22 days ago	6.54 GB	View Edit Delete
mcr.microsoft.com/azure-cognitive-services/form-recognizer/businesscard-3.0	latest	27e5e3045a63	1 month ago	6.46 GB	View Edit Delete
mcr.microsoft.com/azure-cognitive-services/form-recognizer/layout-3.1	latest	6d835d79cebd	1 month ago	5.21 GB	View Edit Delete
mcr.microsoft.com/azure-cognitive-services/form-recognizer/read-3.1	latest	be3a0f26c787	1 month ago	5.21 GB	View Edit Delete
mcr.microsoft.com/azure-cognitive-services/translator/text-translation	latest	7a2a2f4e0b4c	1 month ago	4.39 GB	View Edit Delete
mcr.microsoft.com/azure-cognitive-services/speechservices/speech-to-text	latest	3eb8b2c7a866	2 months ago	28.07 GB	View Edit Delete
mcr.microsoft.com/azure-cognitive-services/speechservices/neural-text-to-speech	latest	bc3bb58c53d4	2 months ago	6.26 GB	View Edit Delete
heinous/phi-3-on-cpu	latest	b73841cff97c	6 months ago	5.98 GB	View Edit Delete

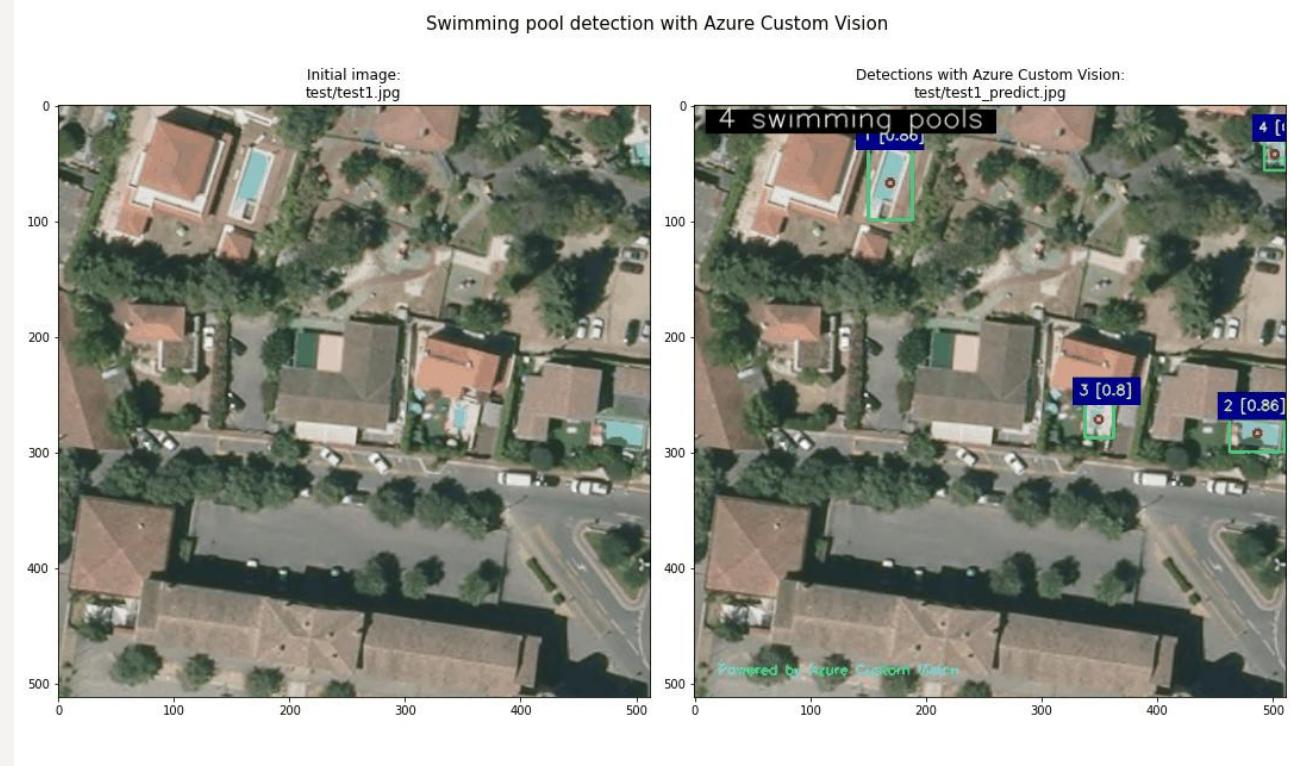
Azure AI Custom Vision demo with model Edge inferencing on a container



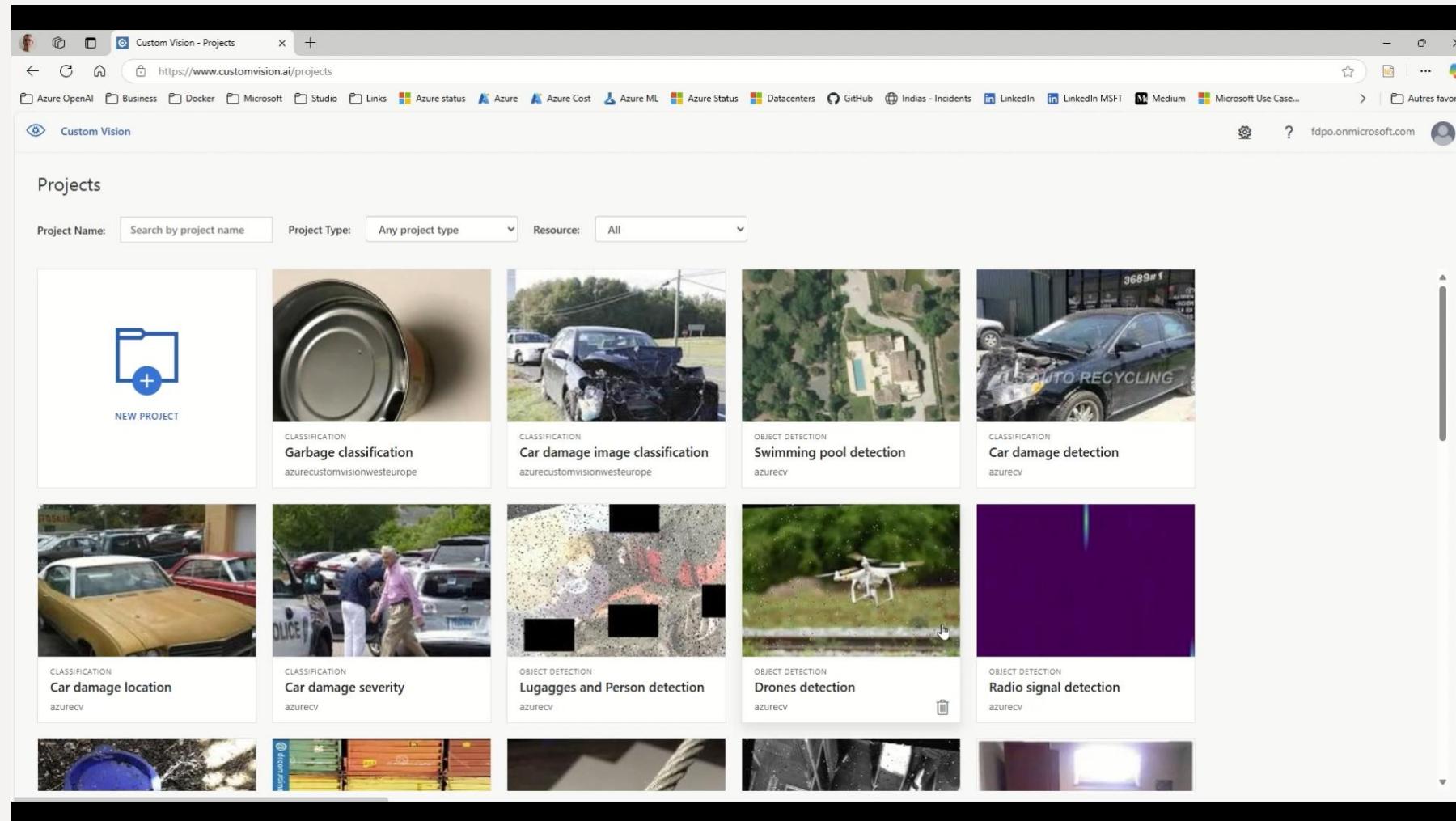
Object Detection model

Our goal is to identify swimming pools in aerial photographs.

- Step 1: We will create and train an object detection model using Azure AI Custom Vision, which will be conducted in the cloud.
- Step 2: After that, we can implement the trained model in a Docker environment for edge inferencing. There are also extra edge functionalities available (such as ONNX, TensorFlow, iOS, etc.).



Step 1: Build and train an object detection model using Azure Custom Vision



Step 2: Object detection edge inferencing on Docker

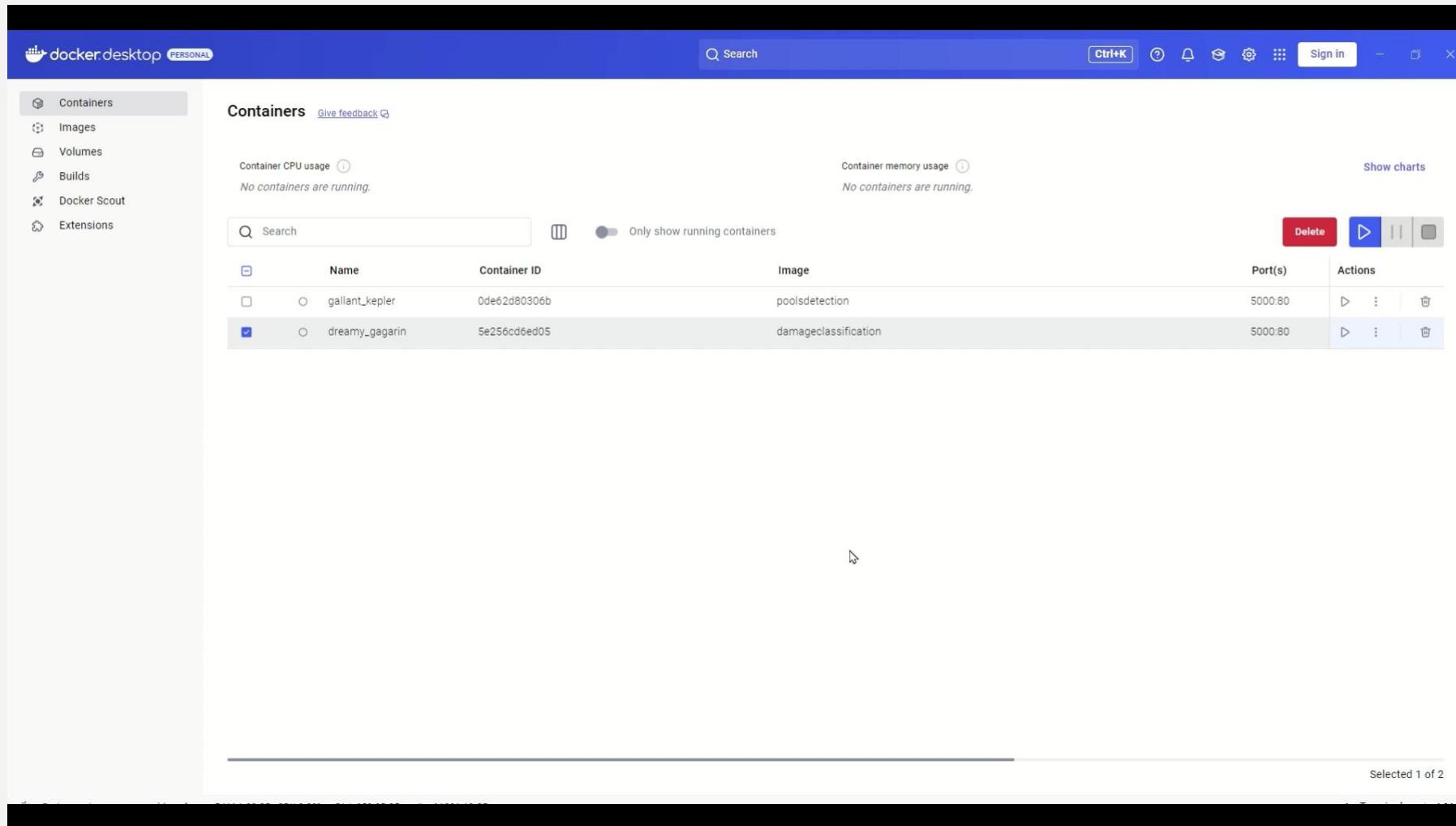


Image classification model for edge inferencing

Image classification Docker deployed model using Azure AI Custom Vision

Upload an image to classify

Drop Image Here
- or -
Click to Upload

ClearSubmit

Examples

Classification results

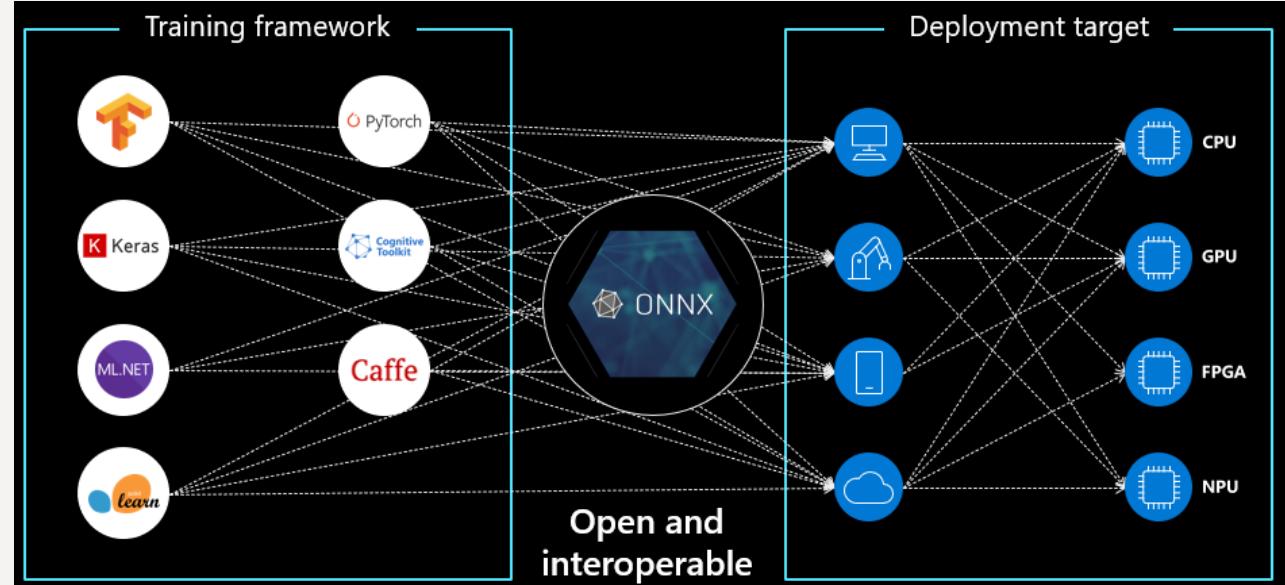
Message

Elapsed time

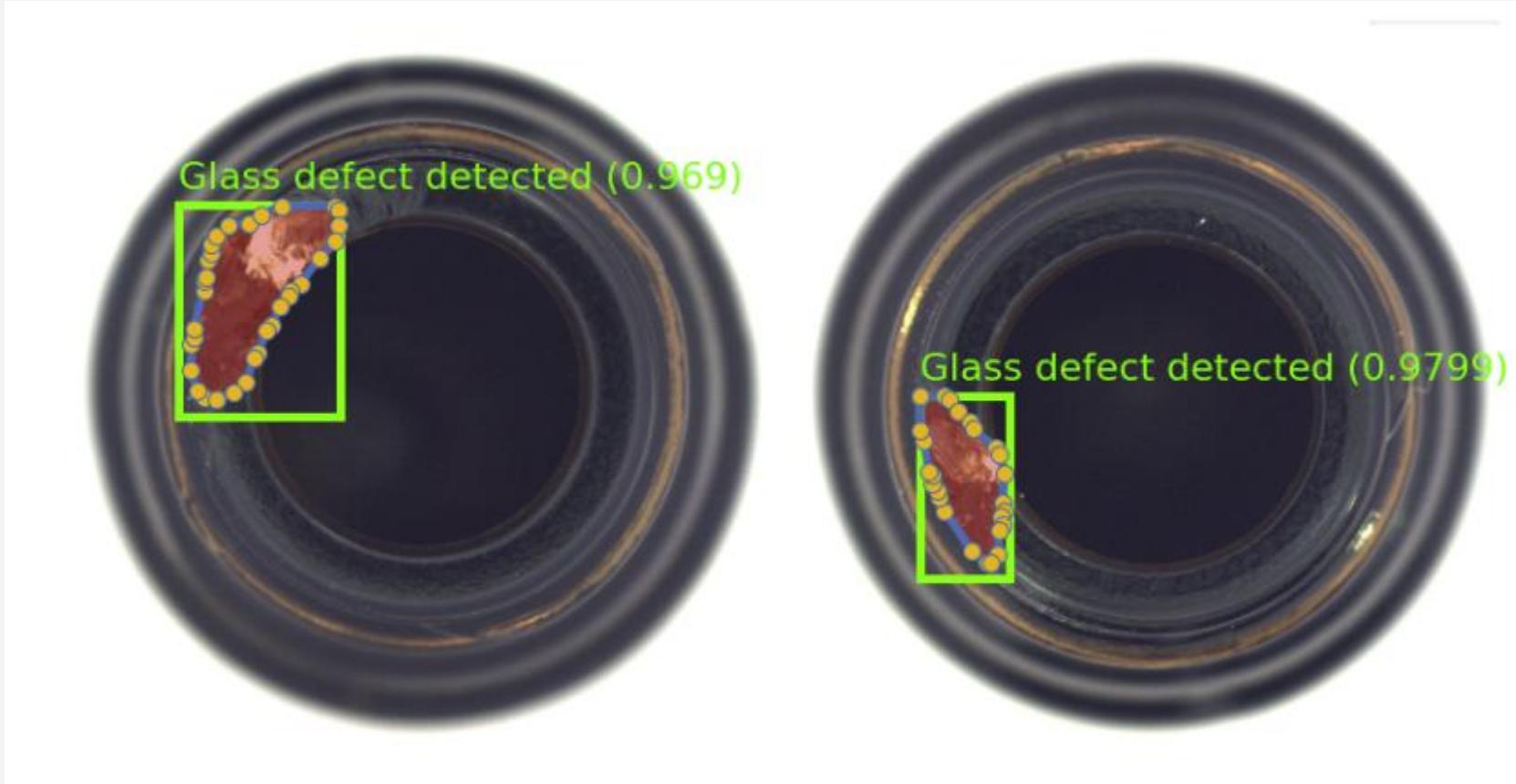
Flag

Instance segmentation model with ONNX deployment

- **ONNX is an open format to represent both deep learning and traditional models.** With ONNX, AI developers can more easily move models between state-of-the-art tools and choose the combination that is best for them. ONNX is developed and supported by a community of partners such as **Microsoft, Facebook and AWS**.
- ONNX is widely supported and can be found in many frameworks, tools, and hardware. Enabling interoperability between different frameworks and streamlining the path from research to production helps increase the speed of innovation in the AI community. ONNX helps to solve the challenge of hardware dependency related to AI models and enables deploying same AI models to several HW accelerated targets.



Instance segmentation with ONNX Edge deployment



[retkowsky/InstanceSegmentation: Instance Segmentation CV model with AutoML for Images](#)

Azure AI Speech services on a container

1. Text to speech
2. Speech to Text



AZURE SPEECH

Azure Speech services containers

Microsoft | Microsoft Artifact Registry Search Employee Sign In

Microsoft Artifact Registry

Discover official Microsoft artifacts such as container images

Showing 7 results Sort by

Filters

Search Partial Exact

Last Published Date
After Before

Category

- Application Frameworks
- Application Infrastructure
- Application Services
- Base Images
- Databases
- DevOps Tools
- IoT Edge Modules
- Messaging Services
- Monitoring

<p>Microsoft Azure AI services Neural Text To Speech</p> <p>Last Published: 12/04/2024</p> <p>Azure AI services Neural Text To Speech Images</p> <p>Learn more</p>	<p>Microsoft Azure AI services Speech Language Detection</p> <p>Last Published: 10/21/2024</p> <p>Azure AI services Speech Language Detection Images</p> <p>Learn more</p>	<p>Microsoft Azure AI services Custom Speech To Text</p> <p>Last Published: 10/21/2024</p> <p>Azure AI services Custom Speech To Text Images</p> <p>Learn more</p>	<p>Microsoft Azure AI services Speech To Text</p> <p>Last Published: 10/21/2024</p> <p>Azure AI services Speech To Text Images</p> <p>Learn more</p>	<p>Microsoft Azure AI services Custom Text To Speech</p> <p>Last Published: 12/16/2024</p> <p>Azure AI services Custom Text To Speech Images</p> <p>Learn more</p>
<p>Microsoft Azure AI services Text To Speech</p> <p>Last Published: 09/28/2021</p> <p>Azure AI services Text To Speech Images</p> <p>Learn more</p>				

Speech to text on container

The screenshot shows the Docker Desktop application window. The title bar reads "docker desktop PERSONAL". The left sidebar has a "Personal" section with icons for Containers, Images (which is selected), Volumes, Builds, Docker Scout, and Extensions. The main area is titled "Images" with a "Local" tab selected. It shows a list of 10 images. A progress bar at the top indicates "18.88 GB / 41.34 GB in use". The table lists the following images:

Name	Tag	Image ID	Created	Size	Actions
poolsdetection	latest	4fdb8e47908d	1 day ago	343.3 MB	More Options
damageclassification	latest	3574d040b60e	7 days ago	375.66 MB	More Options
ollama/ollama	latest	bb26c2494105	21 days ago	6.54 GB	More Options
mcr.microsoft.com/azure-cognitive-services/form-recognizer/businesscard-3.0	latest	27e5e3045a63	1 month ago	6.46 GB	More Options
mcr.microsoft.com/azure-cognitive-services/form-recognizer/layout-3.1	latest	6d835d79cebd	1 month ago	5.21 GB	More Options
mcr.microsoft.com/azure-cognitive-services/form-recognizer/read-3.1	latest	be3a0f26c787	1 month ago	5.21 GB	More Options
mcr.microsoft.com/azure-cognitive-services/translator/text-translation	latest	7a2a2f4e0b4c	1 month ago	4.39 GB	More Options
mcr.microsoft.com/azure-cognitive-services/speechservices/speech-to-text	latest	3eb8b2c7a866	2 months ago	28.07 GB	More Options
mcr.microsoft.com/azure-cognitive-services/speechservices/neural-text-to-speech	latest	bc3bb58c53d4	2 months ago	6.26 GB	More Options
heinous/phi-3-on-cpu	latest	b73841cff97c	6 months ago	5.98 GB	More Options

At the bottom right, it says "Showing 10 items". The status bar at the bottom of the window shows "Last refresh: 10 minutes ago".

Sentiment Analysis with Azure AI Language on a container



Sentiment analysis with Azure AI Language on a container

The image shows a comparison between two tools: HOPPSOTCH on the left and Postman on the right. A large blue arrow points from the HOPPSOTCH interface to the Postman interface.

HOPPSOTCH (Left):

- REST:** Selected.
- Method:** POST to Azure AI cont...
- URL:** http://localhost:5000/text/analytics/v3.0/sentiment
- Body:** application/json
- Raw Request Body:**

```
1  {
2    "documents": [
3      {
4        "language": "en",
5        "id": "1",
6        "text": "I love using Azure for sentiment analysis!"
7      }
8    ]
9 }
```

Postman (Right):

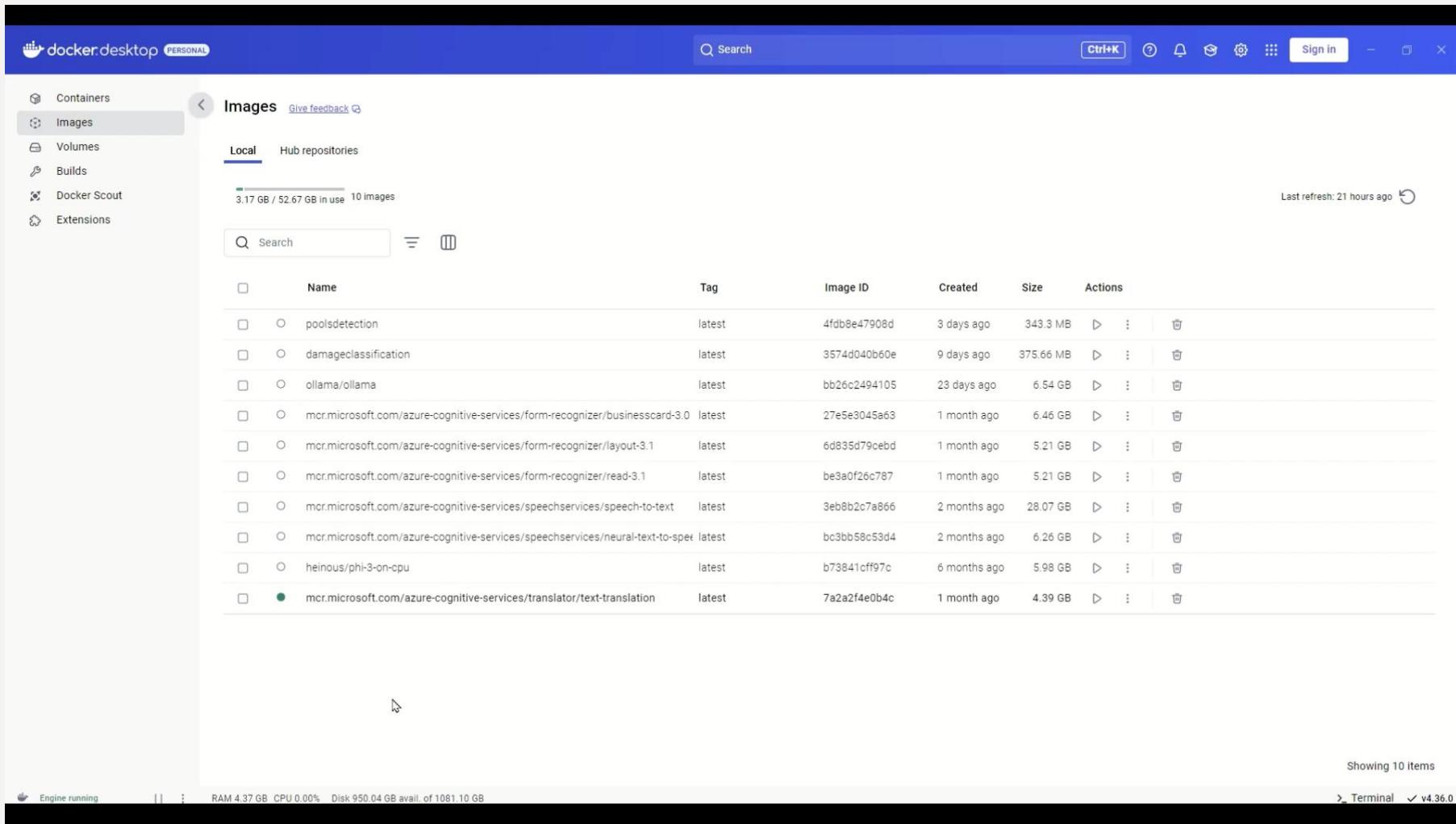
- Status:** 200 • OK | Time: 8939 ms | Size: 351 B
- JSON:** Selected.
- Headers:** 4
- Test Results:**
- Response Body:**

```
1  {
2    "documents": [
3      {
4        "id": "1",
5        "sentiment": "positive",
6        "confidenceScores": {
7          "positive": 0.99,
8          "neutral": 0.01,
9          "negative": 0.0
10       },
11       "sentences": [
12         {
13           "sentiment": "positive",
14           "confidenceScores": {
15             "positive": 0.99,
16             "neutral": 0.01,
17             "negative": 0.0
18           },
19           "offset": 0,
20           "length": 42,
21           "text": "I love using Azure for sentiment analysis!"
22         }
23       ]
24     }
25   }
```

Azure AI Translator on a container



Azure AI Translator on a container



The screenshot shows the Docker Desktop interface with the "Images" tab selected. The left sidebar includes options for Containers, Images (which is selected), Volumes, Builds, Docker Scout, and Extensions. The main area displays a table of 10 local images, with a "Hub repositories" section below it.

Name	Tag	Image ID	Created	Size	Actions
poolsdetection	latest	4fdb8e47908d	3 days ago	343.3 MB	⋮
damageclassification	latest	3574d040b60e	9 days ago	375.66 MB	⋮
ollama/ollama	latest	bb26c2494105	23 days ago	6.54 GB	⋮
mcr.microsoft.com/azure-cognitive-services/form-recognizer/businesscard-3.0	latest	27e5e3045a63	1 month ago	6.46 GB	⋮
mcr.microsoft.com/azure-cognitive-services/form-recognizer/layout-3.1	latest	6d835d79cebd	1 month ago	5.21 GB	⋮
mcr.microsoft.com/azure-cognitive-services/form-recognizer/read-3.1	latest	be3a0f26c787	1 month ago	5.21 GB	⋮
mcr.microsoft.com/azure-cognitive-services/speechservices/speech-to-text	latest	3eb8b2c7a866	2 months ago	28.07 GB	⋮
mcr.microsoft.com/azure-cognitive-services/speechservices/neural-text-to-speech	latest	bc3bb58c53d4	2 months ago	6.26 GB	⋮
heinous/phi-3-on-cpu	latest	b73841cff97c	6 months ago	5.98 GB	⋮
mcr.microsoft.com/azure-cognitive-services/translator/text-translation	latest	7a2a2f4e0b4c	1 month ago	4.39 GB	⋮

Showing 10 items

Engine running | RAM 4.37 GB CPU 0.00% Disk 950.04 GB avail. of 1081.10 GB | Terminal ✓ v4.36.0



Demos videos

Azure AI on the Edge
de Serge Analytics
Playlist · Publique · 7 vidéos · Aucune vue
Azure AI demos services on the edge

1 Azure AI Speech - Speech to Text on container
Serge Analytics · Aucune vue · il y a 29 secondes · 0:40

2 Azure AI Custom Vision - Object detection model training
Serge Analytics · Aucune vue · il y a 1 minute · 1:14

3 Azure AI Document Intelligence and Phi3 Layout API - English demo
Serge Analytics · Aucune vue · il y a 1 minute · 1:56

4 Azure AI Custom Vision - Image classification Docker deployed model
Serge Analytics · Aucune vue · il y a 2 minutes · 1:54

5 Azure AI Custom Vision - Object detection Docker deployed model
Serge Analytics · Aucune vue · il y a 3 minutes · 1:06

6 Azure AI Document Intelligence and Phi3 Layout API - French demo
Serge Analytics · Aucune vue · il y a 5 minutes · 2:06

<https://aka.ms/azureai-edge-demosvideos>

Quick assessment

1. You plan to use an Azure AI services container in a local Docker host. Which of the following statements is true?

- Client applications must pass a subscription key to the Azure resource endpoint before using the container.
- The container must be able to connect to the Azure resource endpoint to send usage data for billing.
- All data passed from the client application to the container is forwarded to the Azure resource endpoint.

2. Which of the following parameters must you specify when deploying an Azure AI services container image?

- EULA
- ResourceGroup
- SubscriptionName

3. You plan to use the language detection functionality of Azure AI Language in a container. Which container image should you deploy?

- mcr.microsoft.com/azure-ai-services/textanalytics
- mcr.microsoft.com/azure-ai-services
- mcr.microsoft.com/azure-ai-services/textanalytics/language

Quick assessment

1. You plan to use an Azure AI services container in a local Docker host. Which of the following statements is true?

- Client applications must pass a subscription key to the Azure resource endpoint before using the container.
- The container must be able to connect to the Azure resource endpoint to send usage data for billing. ✓ Correct
- All data passed from the client application to the container is forwarded to the Azure resource endpoint.

2. Which of the following parameters must you specify when deploying an Azure AI services container image?

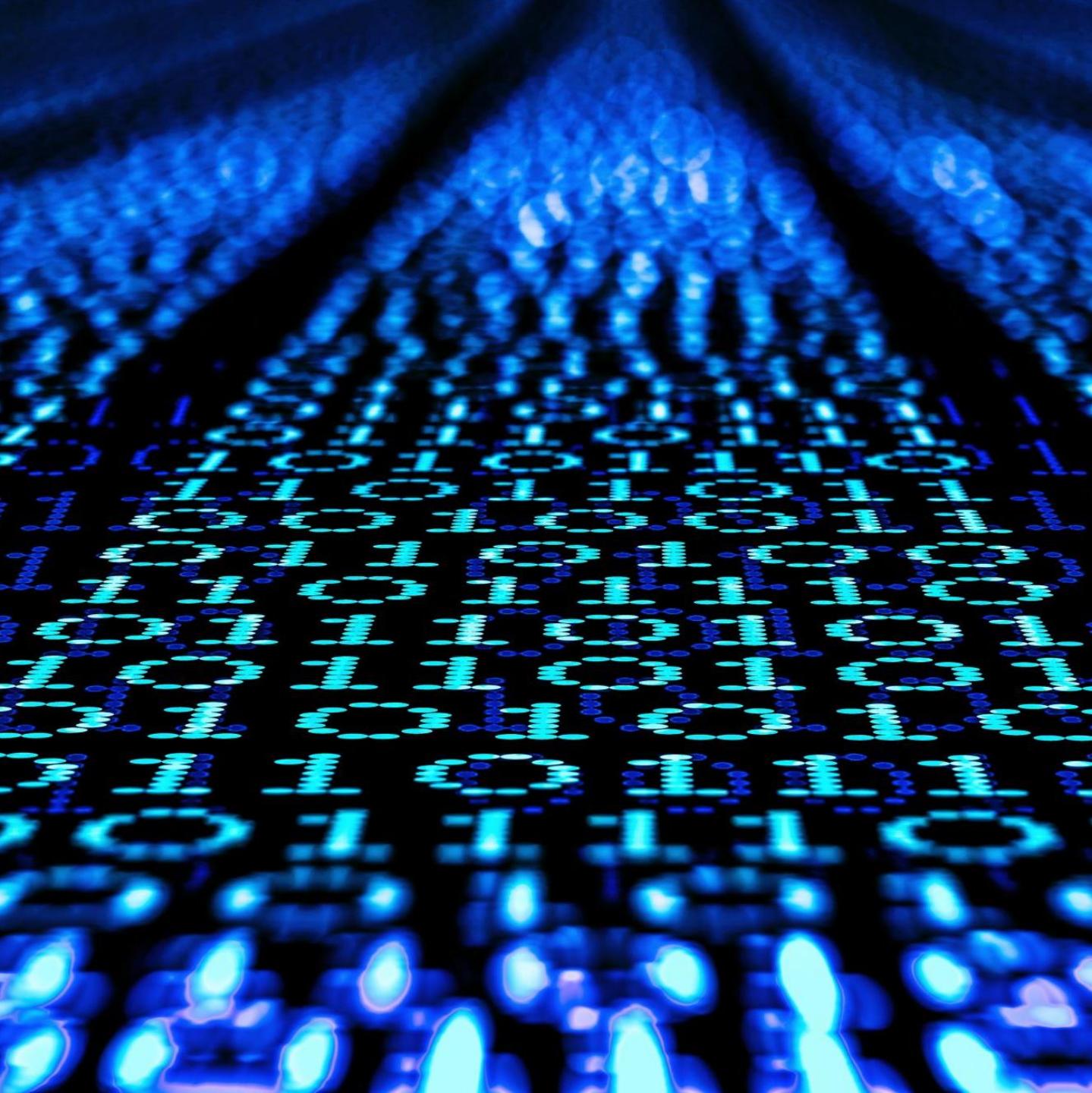
- EULA ✓ Correct
- ResourceGroup
- SubscriptionName

3. You plan to use the language detection functionality of Azure AI Language in a container. Which container image should you deploy?

- mcr.microsoft.com/azure-ai-services/textanalytics
- mcr.microsoft.com/azure-ai-services
- mcr.microsoft.com/azure-ai-services/textanalytics/language ✓ Correct

New!

AI-KnowlEDGE accelerator



New!

AI-knowlEDGE

AI-knowlEDGE is a desktop application to showcase Disconnected Containers.

[Azure-Samples/AI-KnowlEDGE](#)

AI-KnowlEDGE Public

Unwatch 14

main Branches Tags Go to file Add file Code

ajakupov-microsoft doc: update Docker instruction + Cognitive Services 3bab9f · 19 hours ago 76 Commits

.vscode feat: update launch json, add frontend 5 days ago

backend refactor: remove unused docker file 5 days ago

frontend feat: update logo 2 days ago

images doc: update logo 2 days ago

tests refactor: migrate ollama service 3 weeks ago

.gitignore refactor: remove ollama summary 3 weeks ago

LICENSE fix: cleaning up secrets 3 months ago

README.md doc: update Docker instruction + Cognitive Services 19 hours ago

requirements.txt feat: Add docker compose for summary last month

README Code of conduct MIT license

CONTRIBUTORS 1 FORKS 0 STARS 0 ISSUES 0 OPEN LICENSE MIT LINKEDIN

The screenshot shows the GitHub repository page for 'AI-KnowlEDGE'. At the top, there's a header with the repository name, a 'Public' button, and a 'Unwatch 14' link. Below the header are navigation links for 'main', 'Branches', 'Tags', 'Go to file', 'Add file', and 'Code'. The main content area displays a list of 76 commits from user 'ajakupov-microsoft'. Each commit includes a file icon, the file path, a brief description, and a timestamp. Below the commit list are links for 'README', 'Code of conduct', and 'MIT license'. At the bottom, there are buttons for 'CONTRIBUTORS' (1), 'FORKS' (0), 'STARS' (0), 'ISSUES' (0 OPEN), 'LICENSE' (MIT), and 'LINKEDIN'. A large watermark at the bottom features the text 'AI knowlEDGE' with star icons and 'Your Disconnected Documents Analyzer'.

AI-KnowEDGE Demo video: AIKnowEDGE demo video - YouTube

The screenshot displays the AI-KnowEDGE software interface. At the top, there's a navigation bar with icons for back, forward, search, and other system functions. Below the bar, the title "Document Analysis" is visible. A sub-header indicates the document is from "Investigation 2000". The main content area contains two large text blocks. The left block discusses the CIA's understanding of the threat presented by a communist-backed government agent (Bin Laden) and the aggressive level of that threat. It also mentions that the report fails to document such an assessment. The right block discusses the CIA's failure to understand the gravity and magnitude of the threat, and that they failed to take action in response. In fact, the memorandum of these officials before the 9/11 Commission and other evidence shows that they correctly understood the threat, and also noted previous efforts to negotiate of the Americans with their then-Security Advisor, then-Secretary of Defense and then-Secretary of State, and their deputies, and Attorney General, the intelligence community who did not act on findings, and senior officials behind George W. Bush and Vice President Cheney and his cabinet members. The right block also discusses the CIA's failure to brief the National Security Advisor on the threat posed by al-Qaeda and discussed anti- and post-9/11 intelligence to make informed policy decisions against the threat. It also discusses the memorandum of the 9/11 Commission, which detailed covert action findings, and Mr. Allen's testimony before Congress, and his meeting with the Director of Central Intelligence. The right block concludes with a statement that the CIA did better through their own analysis or their own research regarding the quality of our collection efforts, or the magnitude of the collection interests and possible threats to our national security. A summary section at the bottom right includes "Augmentation Summary" and "Download Summary".

Suggested Questions

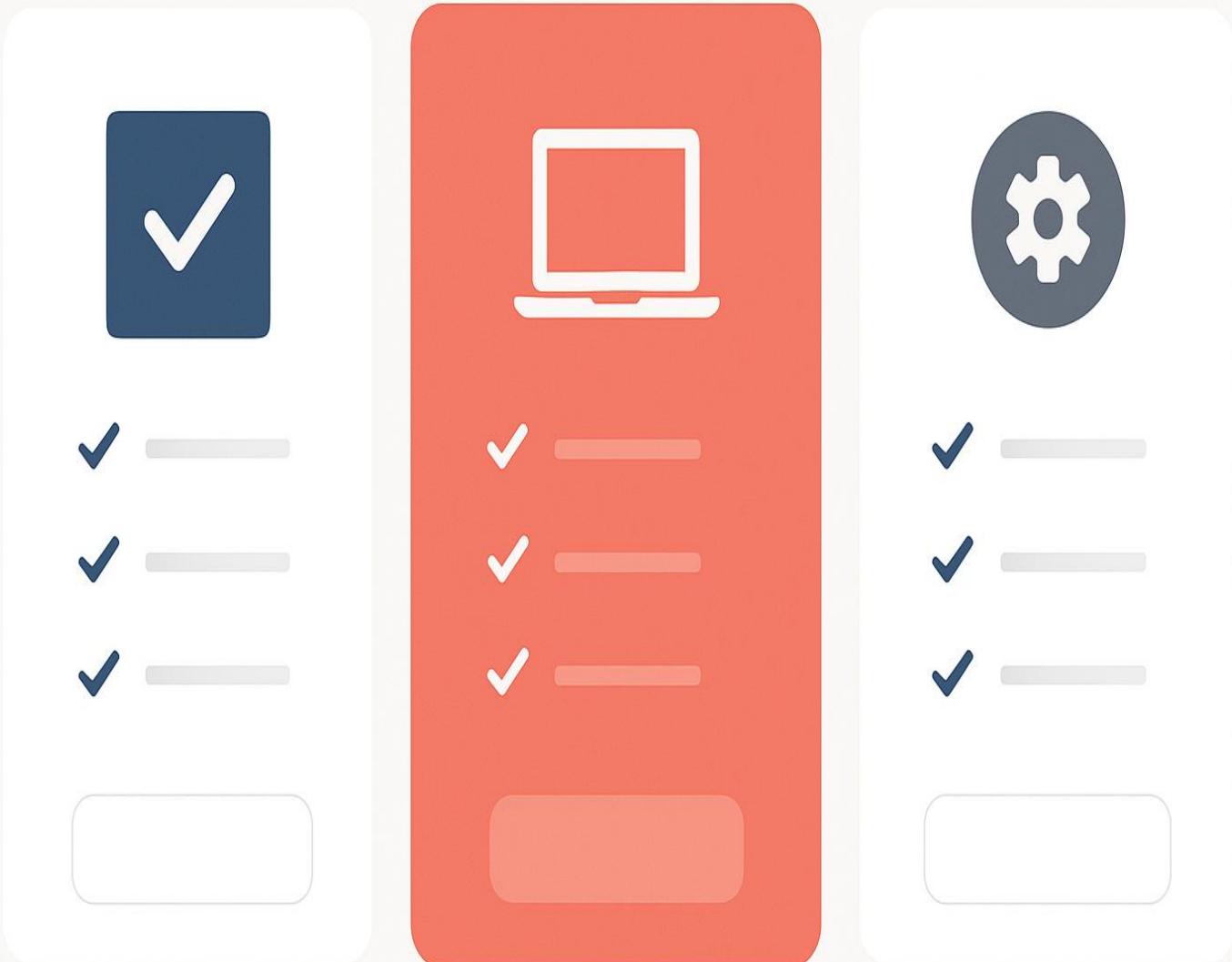
Generating suggested questions...

SUGGESTED QUESTIONS

```
Last login: Wed Apr  9 09:01:53 on ttys000
alibekjakupov@mac-1 AI-KnowLEDGE %
```



Pricing



Document Intelligence Pricing

Category	Features	Price (per year)	Max usage (per year)	Projected usage (per month)
Document Intelligence	Custom extraction	\$24,480 \$107,100	1.2M pages 6M pages	100,000 pages 500,000 pages
	Prebuilt	\$8,640 \$38,400	1.2M pages 6M pages	100,000 pages 500,000 pages
	Read	\$12,960 \$45,360	24M pages 96M pages	2M pages 8M pages

Pricing Table : Pricing - Azure AI Document Intelligence | Microsoft Azure

Container	Minimum	Recommended
Read	8 cores, 10-GB memory	8 cores, 24-GB memory
Layout	8 cores, 16-GB memory	8 cores, 24-GB memory
Business Card	8 cores, 16-GB memory	8 cores, 24-GB memory
General Document	8 cores, 12-GB memory	8 cores, 24-GB memory
ID Document	8 cores, 8-GB memory	8 cores, 24-GB memory
Invoice	8 cores, 16-GB memory	8 cores, 24-GB memory
Receipt	8 cores, 11-GB memory	8 cores, 24-GB memory
Custom Template	8 cores, 16-GB memory	8 cores, 24-GB memory

Vision		
	Quota per billing period	Price (billing period)
Document Intelligence Custom API (Trans)	20,000 (POC)	\$459 (Mo)
	1,200,000	\$24,480 (Yr)
	6,000,000	\$107,100 (Yr)
Document Intelligence Prebuilt (Trans)	20,000 (POC)	\$152 (Mo)
	1,200,000	\$8,640 (Yr)
	6,000,000	\$38,400 (Yr)
CV and Document Intelligence Read (Trans)	500,000 (POC)	\$337.50 (Mo)
	24,000,000	\$12,960 (Yr)
	96,000,000	\$45,360 (Yr)

Documentation : Install and run Docker containers for Document Intelligence - Azure AI services | Microsoft Learn

Azure AI Vision Pricing

Instance	Category	Features	Price per year	Max usage per year	Project usage per month
Disconnected container	Computer Vision	Read	\$12,960	24M transactions	2M transactions
			\$45,360	96M transactions	8M transactions

Pricing Table : [Pricing – Computer Vision API | Microsoft Azure](#)

Container	Minimum	Recommended
Read 3.2 2022-04-30	4 cores, 8-GB memory	8 cores, 16-GB memory
Read 3.2 2021-04-12	4 cores, 16-GB memory	8 cores, 24-GB memory

Documentation : [Azure AI Vision 3.2 GA Read OCR container - Azure AI services | Microsoft Learn](#)

Azure AI Translator Pricing

Features	Price per year	Max usage per year	Projected usage per month
Standard	\$513,000	120B Billed Units	10B Billed Units
	\$250,800	48B Billed Units	4B Billed Units

Pricing Table: [Pricing - Translator | Microsoft Azure](#)

Function	Minimum recommended	Notes
Text translation	4 Core, 4-GB memory	
Text transliteration	4 Core, 2-GB memory	
Document translation	4 Core, 6-GB memory	The number of documents that can be processed concurrently can be calculated with the following formula: [minimum of ($n-2$), ($m-6$)/4]. <ul style="list-style-type: none">• n is number of CPU cores.• m is GB of memory.• Example: 8 Core, 32-GB memory can process six(6) concurrent documents [minimum of (8-2), (36-6)/4].

Documentation: [Install and run Translator container using Docker API - Azure AI services | Microsoft Learn](#)

Azure AI Language Pricing

Features	Price per year	Max usage per year	Projected usage per month
Sentiment analysis (and opinion mining), Key phrase extraction, Language detection, Named entity recognition, PII detection	\$33,600	120M text records	10M text records
	\$13,200	36M text records	3M text records
Summarisation	\$67,200	120M text records	10M text records
	\$31,680	36M text records	3M text records

Pricing Table : Pricing – Language service | Microsoft Azure

	Minimum host specs	Recommended host specs	Minimum TPS	Maximum TPS
Language detection	1 core, 5GB memory	1 core, 8GB memory	15	30

Use language detection Docker containers on-premises - Azure AI services | Microsoft Learn

	Minimum host specs	Recommended host specs	Minimum TPS	Maximum TPS
Key Phrase Extraction	1 core, 2GB memory	1 core, 4GB memory	15	30

Use Docker containers for Key Phrase Extraction on-premises - Azure AI services | Microsoft Learn

	Minimum host specs	Recommended host specs
Named Entity Recognition	1 core, 2GB memory	4 cores, 8GB memory

Use named entity recognition Docker containers on-premises - Azure AI services | Microsoft Learn

	Minimum host specs	Recommended host specs	Minimum TPS	Maximum TPS
Sentiment Analysis	1 core, 2GB memory	4 cores, 8GB memory	15	30

Install and run Docker containers for Sentiment Analysis - Azure AI services | Microsoft Learn

Azure AI Speech Pricing

Category	Features	Price (per year)	Max usage (per year)	Projected usage (per month)
Speech-to-Text ²	Standard	\$74,100 \$285,000 Sign up to get access Learn more	120,000 hours 600,000 hours	10,000 hours 50,000 hours
	Custom	\$88,920 \$342,000 Sign up to get access Learn more	120,000 hours 600,000 hours	10,000 hours 50,000 hours
	Enhanced add-on features: • Language identification • Diarization	\$22,230 \$85,500	120,000 hours 600,000 hours	10,000 hours 50,000 hours
Text to Speech	Neural ¹	\$47,424 \$182,400 Sign up to get access Learn more	4.8B characters 24B characters	400M characters 2,000M characters

Pricing Table : Azure AI Speech Pricing | Microsoft Azure

- Runs a `speech-to-text` container from the container image.
- Allocates 4 CPU cores and 8 GB of memory.
- Exposes TCP port 5000 and allocates a pseudo-TTY for the container.
- Automatically removes the container after it exits. The container image is still available on the host computer.

- Runs a custom speech to text container from the container image.
- Allocates 4 CPU cores and 8 GB of memory.
- Loads the custom speech to text model from the volume input mount, for example, `C:\CustomSpeech`.
- Exposes TCP port 5000 and allocates a pseudo-TTY for the container.
- Downloads the model given the `ModelId` (if not found on the volume mount).
- If the custom model was previously downloaded, the `ModelId` is ignored.
- Automatically removes the container after it exits. The container image is still available on the host computer.

[Speech to text containers - Speech service - Azure AI services | Microsoft Learn](#)

[Custom speech to text containers - Speech service - Azure AI services | Microsoft Learn](#)

- Runs a neural text to speech container from the container image.
- Allocates 6 CPU cores and 12 GB of memory.
- Exposes TCP port 5000 and allocates a pseudo-TTY for the container.
- Automatically removes the container after it exits. The container image is still available on the host computer.

[Neural text to speech containers - Speech service - Azure AI services | Microsoft Learn](#)

Customers Success Stories





An energy supplier

- **Use-case:**

The customer has documents to OCR ranging from 1 page to +10,000 pages. These are highly confidential documents, so 1) the use of a container is necessary and 2) the container must be in offline mode without any data leakage.

- **Solution:**

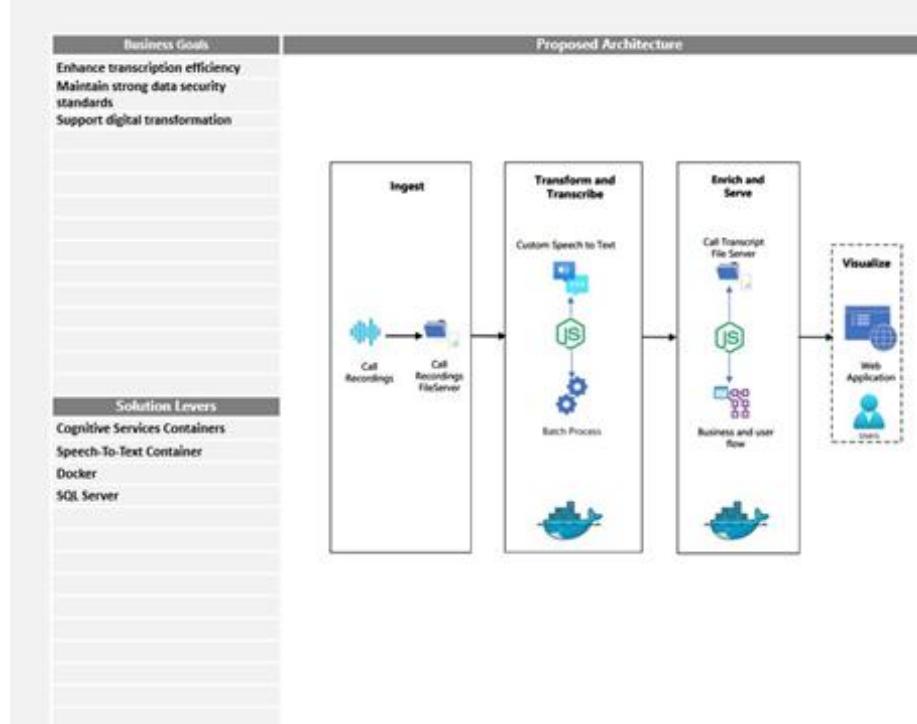
Azure Document intelligence running on a disconnected container.

- **Benefits:**

They process now +100,000 pages per day with a goal of reaching 1 million pages digitized per day with Document Intelligence on disconnected containers.

The quality of the results with Document Intelligence in an offline container is excellent. Perfect recognition of text, handwriting, formulas and plans.

Saudi National Information Center



- **Headquarters**

Riyadh, Saudi Arabia

- **Industry**

Public Sector

- **Azure Workloads**

Azure Cognitive Services Disconnected Container for Speech-To-Text in Saudi with 10k hours



Lenovo

- **Success Case Short Description:**

The incorporation of advanced AI into smart devices, particularly AI PCs, is transforming our interaction with technology, enriching user experiences with personalized, intuitive interfaces. Through this agreement, Lenovo delivers a groundbreaking AI platform at the edge, secured by Microsoft's state-of-the-art AI Safety technology. The business model for these newly developed hybrid solutions has revealed new revenue streams and set a precedent for collaborating with customers in various smart-device scenarios. As AI continues to progress, it will unveil unprecedented opportunities for creating value, fostering growth, and redefining industries worldwide.

- **Microsoft Solution:**

Azure AI Content Safety on disconnected containers and devices: along with AI PC endeavor, Lenovo AINow project, which is a new AI assistant powered by Small Language Model installed on its PC shipments. AINow integrated our Azure AI Content Safety to best serve its customers with improved user experience and compliance. According to Lenovo scenarios, the deployment of Disconnected Container will be primary mode. The embedded version is required to be installed on PC shipments to handle some situations where network connectivity is bad.

Customization for the features on devices: to comply with local AI content regulation and quick response to the potential social concerns, to provide customized mechanisms to help the customer to manage AI content safety on devices.

A considerate commercial solution: new EA with special terms to cover product upgrades, and unified support contract, also include a unprecedent hybrid license agreement to address the customer's concerns on different AI usage scenarios.

Lenovo

Lenovo AI Now

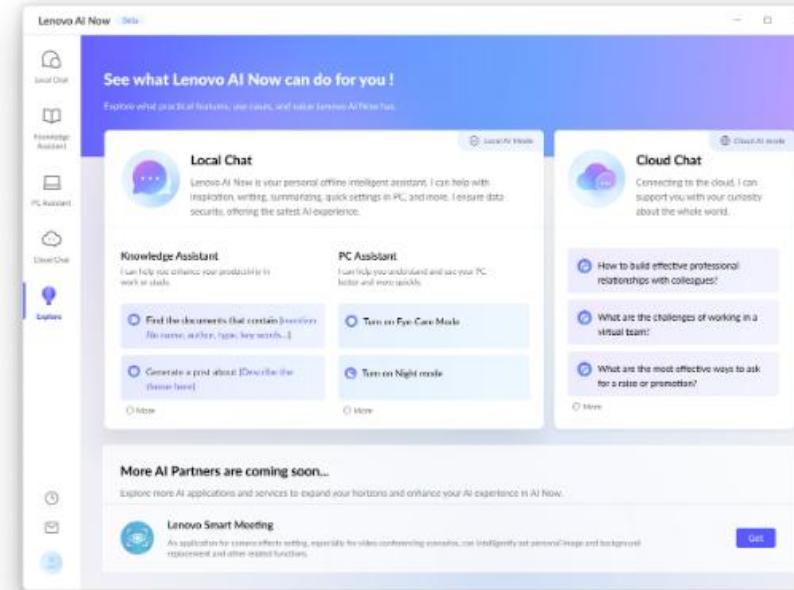
Download Lenovo AI Now for a personalized, productive, and protected AI PC experience on [select Lenovo PCs](#).

Powered by a local large language model (LLM) based on Meta's Llama 3, it handles a wide range of tasks, from knowledge management (Q&A, document search, document summary, etc.) to device control, all while safeguarding your data and privacy through local AI functions. You also have full control and the option to access cloud services with your consent.

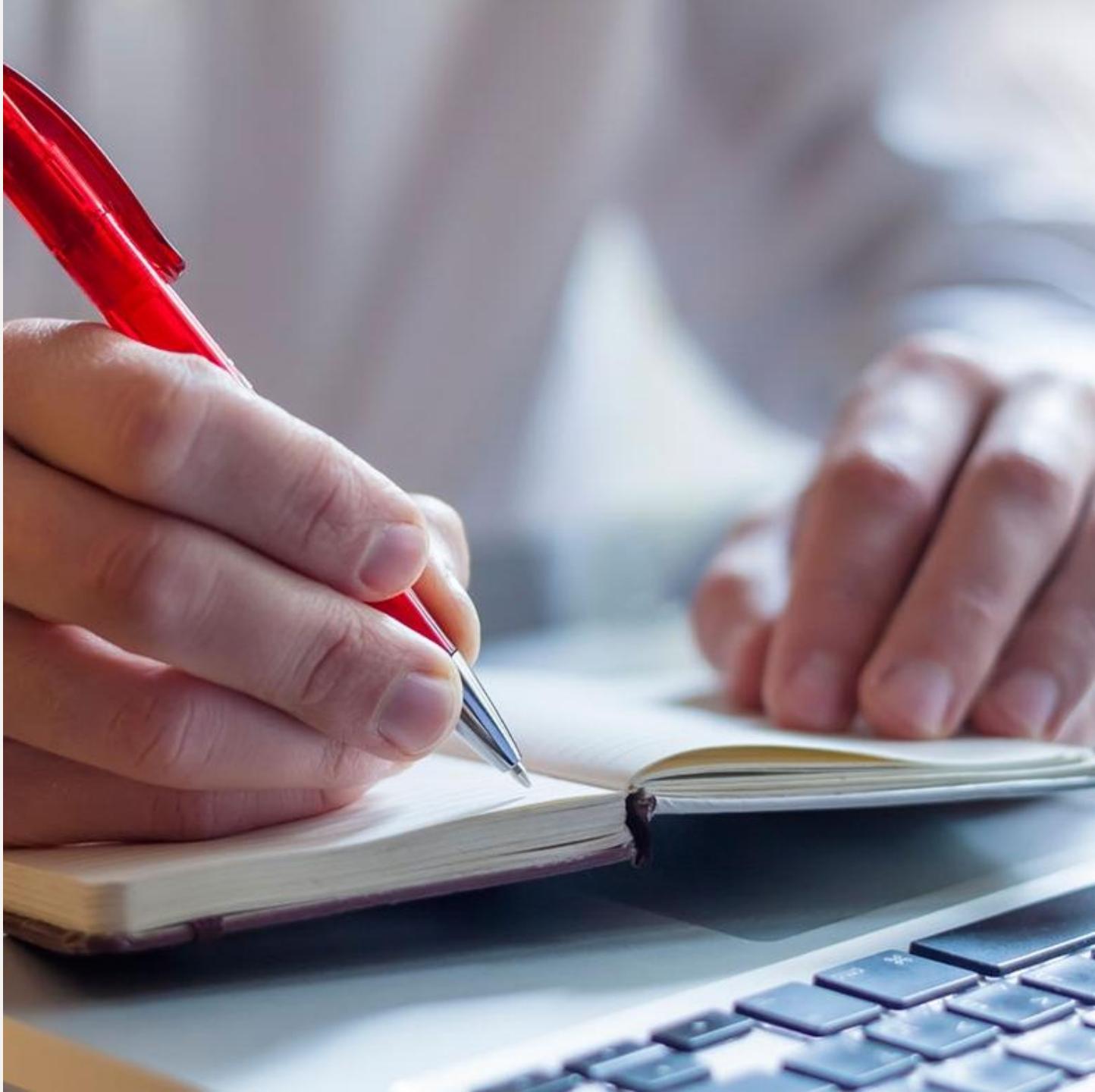
Click to start your journey:

[Download now](#)

1. **Lenovo AI Now only supports select PCs.**
Please see the following for details.
2. **Lenovo AI Now runs on Windows 11,**
and is English only.
3. **Lenovo AI Now currently doesn't support**
Qualcomm(ARM) platform.



Documentation



Documentation

Azure AI Containers documentation:

[Azure AI Services containers - Azure AI Services | Microsoft Learn](#)

FAQ for Azure AI Containers:

[Azure AI Services FAQ - Azure AI Services | Microsoft Learn](#)

Application for Disconnected Containers:

[Azure AI Services Application for Disconnected containers \(microsoft.com\)](#)

FAQ for Azure AI Disconnected containers:

<https://learn.microsoft.com/en-us/azure/ai-services/containers/disconnected-container-faq>

List of Microsoft containers:

<https://mcr.microsoft.com/en-us/>

Deploying Azure AI Services in Containers:

<https://www.youtube.com/watch?v=rkTpJphfq4>

Training



Deploy Azure AI services in containers

49 min • Module • 6 Units

[Feedback](#)

Intermediate AI Engineer Developer DevOps Engineer Solution Architect Student Azure AI services

Azure Container Instances

Learn about Container support in Azure AI services allowing the use of APIs available in Azure and enable flexibility in where to deploy and host the services with Docker containers.

Learning objectives

After completing this module, learners will be able to:

- Create containers for reuse
- Deploy to a container and secure a container
- Consume Azure AI services from a container

[Start >](#) [+ Add](#)

Prerequisites

- Learners should have an Azure account and be familiar with navigating the Azure portal.
- Learners should also have experience with Docker and containers.

Notebooks

 **azure-ai-containers-samples** Public

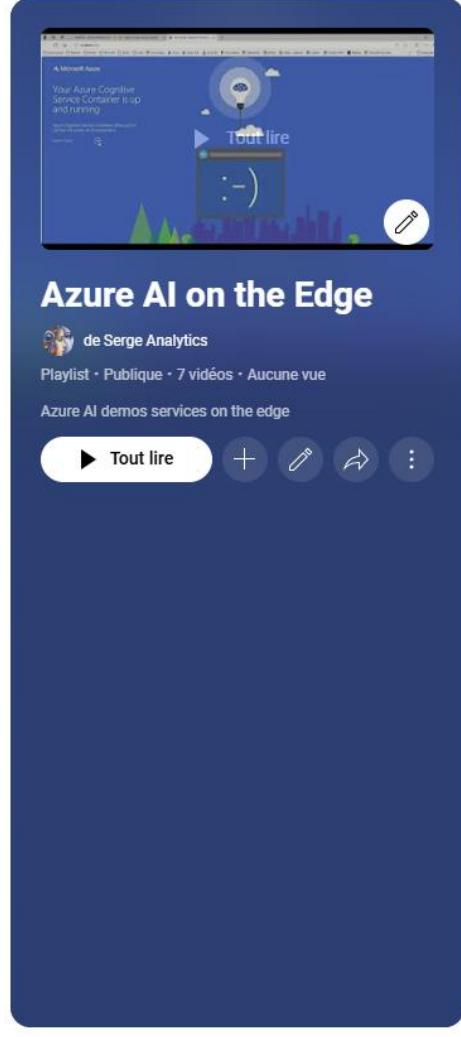
Pin Unwatch 1

main 1 Branch 0 Tags Go to file Add file Code

retkowsky	Update README.md	10b8eb1 · 1 minute ago	37 Commits
content-safety	Delete content-safety/test	1 hour ago	
custom-vision	Delete custom-vision/test	1 hour ago	
document-intelligence	Add files via upload	11 minutes ago	
speech	Add files via upload	15 minutes ago	
translator	Delete translator/test	1 hour ago	
vision	Delete vision/test	1 hour ago	
Azure AI services Containers.pdf	Add files via upload	51 minutes ago	
README.md	Update README.md	1 minute ago	
container.png	Add files via upload	1 minute ago	

<https://github.com/retkowsky/azure-ai-containers-samples/>

Demos videos



The image shows a video player interface with a dark blue background. On the left, there is a large thumbnail for a video titled "Azure AI on the Edge" by "de Serge Analytics". The thumbnail features a brain icon and a speech bubble. Below the thumbnail, the title "Azure AI on the Edge" is displayed in white, along with the author's name "de Serge Analytics", the status "Publique", "7 vidéos", and "Aucune vue". A "Tout lire" button is highlighted in white, and other controls like a plus sign, edit, and share are visible. To the right of this main area, there is a list of six smaller video thumbnails, each with a number (1 through 6) and a title. Each thumbnail includes a timestamp in the bottom right corner.

- 1 Azure AI Speech - Speech to Text on container
Serge Analytics • Aucune vue • il y a 29 secondes
0:40
- 2 Azure AI Custom Vision - Object detection model training
Serge Analytics • Aucune vue • il y a 1 minute
1:14
- 3 Azure AI Document Intelligence and Phi3 Layout API - English demo
Serge Analytics • Aucune vue • il y a 1 minute
1:56
- 4 Azure AI Custom Vision - Image classification Docker deployed model
Serge Analytics • Aucune vue • il y a 2 minutes
1:54
- 5 Azure AI Custom Vision - Object detection Docker deployed model
Serge Analytics • Aucune vue • il y a 3 minutes
1:06
- 6 Azure AI Document Intelligence and Phi3 Layout API - French demo
Serge Analytics • Aucune vue • il y a 5 minutes
2:06

<https://aka.ms/azureai-edge-demosvideos>

Q&A



Microsoft a 50 ans



Celebrating Microsoft's 50 years
- The Official Microsoft Blog

programmez!
Le magazine des dévs - CTO & Tech Lead

1975-2025

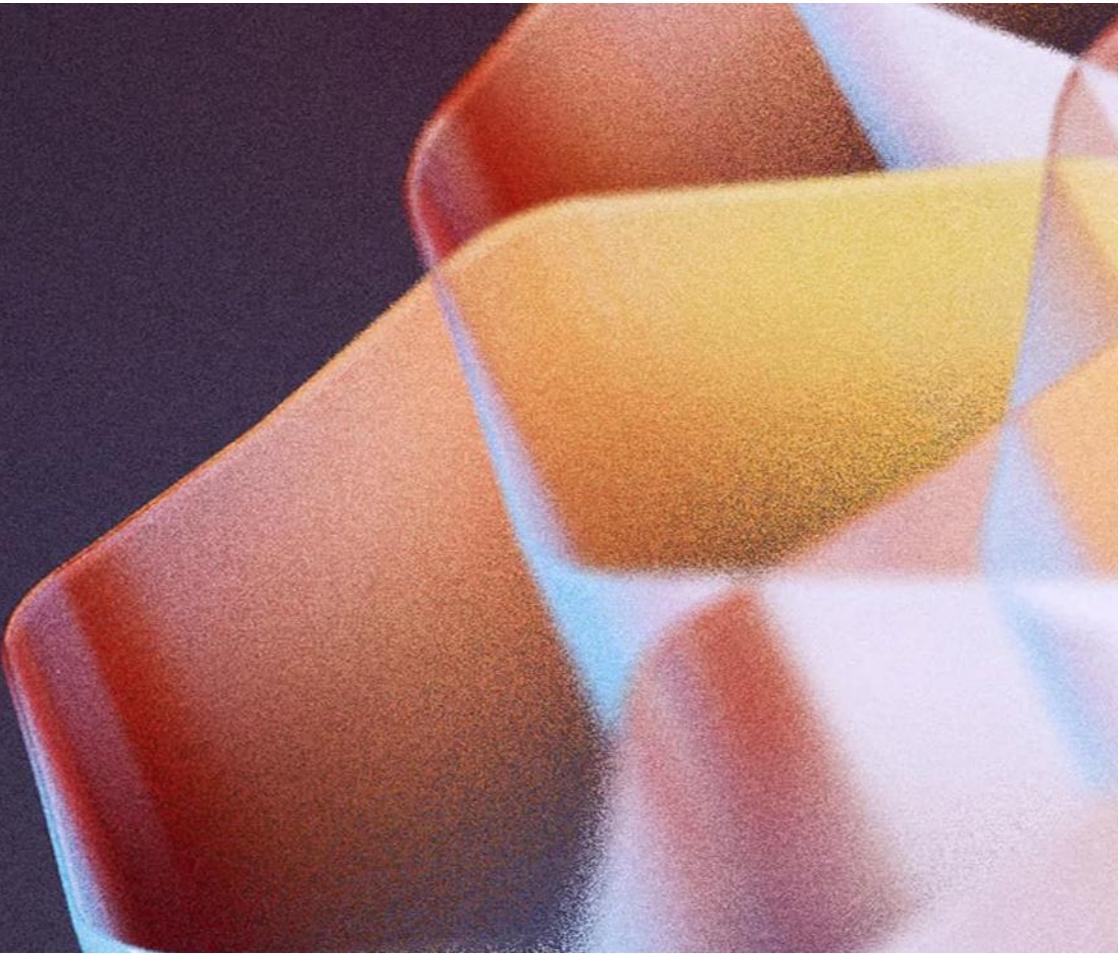
SPÉCIAL PRINTEMPS 2025
HORS-SÉRIE #18



The future is yours @Build

In Seattle and online
May 19-22, 2025

[Register now >](#)



Thu, Apr 24, 1:00 PM CEST



Fine-Tuning des modèles fondationnels avec Azure AI Foundry

Les modèles de GenAI sont extrêmement puissants, mais peuvent manquer de précision pour certaines tâches spécifiques. Le fine-tuning permet d'adapter un modèle fondationnel LLM ou SLM à vos besoins spécifiques en exploitant vos propres données.



Serge RETKOWSKY

France AI Global Black Belt
@Microsoft



Farid EL ATTAOUI

France AI Global Black Belt
@Microsoft



Nicolas ROBERT

AI Architect @Expertime
Microsoft MVP IA



Live jeudi **24/04/2025** à partir de 13h

[GAIC-vf] Fine-Tuning des modèles fondationnels avec Azure AI Foundry

Event by Global AI Community Virtuelle Francophone

Thu, Apr 24, 2025, 1:00 PM - 2:00 PM (your local time) [Add to calendar](#) ▾



∞ Nicolas Robert and 6 other attendees

[Share](#) ▾



Fine-tuning avec Azure AI Foundry le 24 avril 2025

(7) [GAIC-vf] [Fine-Tuning des modèles fondationnels avec Azure AI Foundry](#) | [LinkedIn](#)



MERCI

TACK

ขอบคุณครับ

KIITOS

MULȚUMESC

DANKE

SALMAT PO

GRACIAS

감사합니다

شُكْرًا

DZIĘKUJĘ

OBRIGADO

MAHALO

TEŞEKKÜRLER

ευχαριστώ

ধন্যবাদ

СПАСИБО

ഡാൻറി

ה תְּהִלָּה

Thank you!

多謝晒

PALDIES

ДЯКУЮ

ĎAKUJEM

مُتَشَكِّرُ م

TERIMA KASIH

ТАК

HVALA

شُكْرِيَّه

DĚKUJI

謝謝

CÁM ƠN

DANKON

GRAZIE

KÖSZÖNÖM

БЛАГОДАРЯ

ありがとうございます