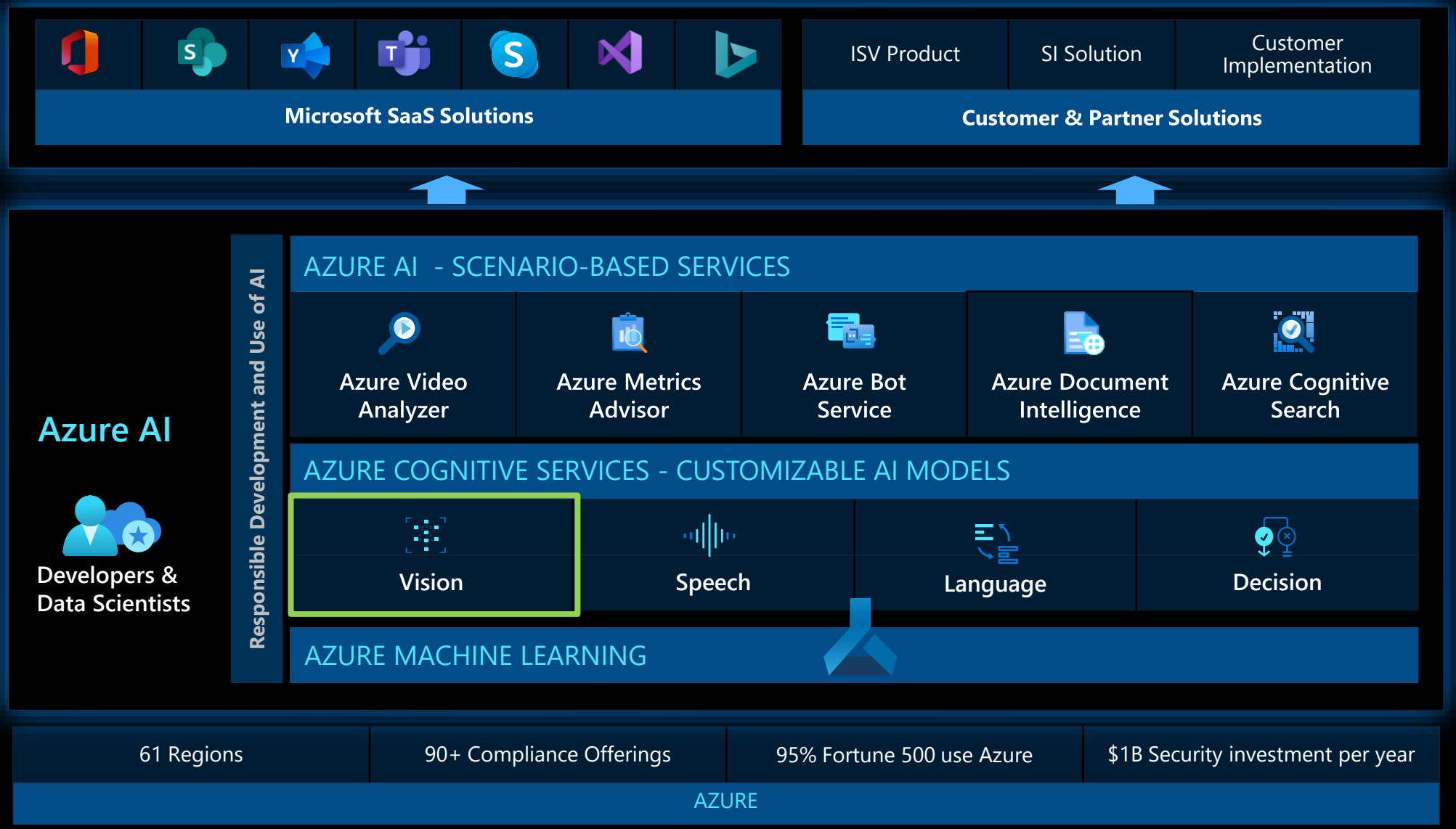
Microsoft Azure

Azure AI:
Vision Service Overview

Azure AI


Microsoft SaaS Solutions
ISV Product
SI Solution
Customer Implementation
Customer & Partner Solutions
Azure AI
Developers & Data Scientists
Responsible Development and Use of AI
AZURE AI - SCENARIO-BASED SERVICES
Azure Video Analyzer
Azure Metrics Advisor
Azure Bot Service
Azure Document Intelligence
Azure Cognitive Search
AZURE COGNITIVE SERVICES - CUSTOMIZABLE AI MODELS
Vision
Speech
Language
Decision
AZURE MACHINE LEARNING
61 Regions
90+ Compliance Offerings
95% Fortune 500 use Azure
$1B Security investment per year
AZURE

Vision Service Overview

Azure's Azure AI Vision service gives you access to advanced algorithms that process images and return information based on the visual features you're interested in.

Azure AI Vision can power many digital asset management (DAM) scenarios. DAM is the business process of organizing, storing, and retrieving rich media assets and managing digital rights and permissions. For example, a company may want to group and identify images based on visible logos, faces, objects, colors, and so on. Or, you might want to automatically generate captions for images and attach keywords so they're searchable.

Vision Services

Optical Character Recognition (OCR)

Image Analysis

Face

Spatial Analysis

https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/overview

Vision Service Features

Optical Character Recognition (OCR)

OCR or Optical Character Recognition is also referred to as text recognition or text extraction. Machine-learning-based OCR techniques allow you to extract printed or handwritten text from images such as posters, street signs and product labels, as well as from documents like articles, reports, forms, and invoices. The text is typically extracted as words, text lines, and paragraphs or text blocks, enabling access to digital version of the scanned text. This eliminates or significantly reduces the need for manual data entry.

Microsoft's Read OCR engine is composed of multiple advanced machine-learning based models supporting global languages. It can extract printed and handwritten text including mixed languages and writing styles. Read is available as **cloud service and on-premises container** for deployment flexibility. With the latest preview, it's also available as a synchronous API for single, non-document, image-only scenarios with performance enhancements that make it easier to implement OCR-assisted user experiences.

input	Examples	Read edition	Benefit
Images: General, in-the-wild images	labels, street signs, and posters	OCR for images (version 4.0 preview)	Optimized for general, non-document images with a performance-enhanced synchronous API that makes it easier to embed OCR in your user experience scenarios.
Documents: Digital and scanned, including images	books, articles, and reports	Document Intelligence read model	Optimized for text-heavy scanned and digital documents with an asynchronous API to help automate intelligent document processing at scale.

Optical Character Recognition (OCR)

Common Features:

The Read OCR model is available in Azure AI Vision and Document Intelligence with common baseline capabilities while optimizing for respective scenarios. The following list summarizes the common features:

- Printed and handwritten text extraction in supported languages
- Pages, text lines and words with location and confidence scores
- Support for mixed languages, mixed mode (print and handwritten)
- Available as Distroless Docker container for on-premises deployment

Vision Service Features

Image Analysis

The Azure AI Vision Image Analysis service can extract a wide variety of visual features from your images. For example, it can determine whether an image contains adult content, find specific brands or objects, or find human faces.

The latest version of Image Analysis, 4.0, which is now in public preview, has new features like synchronous OCR and people detection. We recommend you use this version going forward.

You can use Image Analysis through a client library SDK or by calling the REST API directly.

Image Analysis Concepts

You can analyse images to provide insights about their visual features and characteristics. All of the features in this list are provided by the Analyse Image API.
Transcriptions, captions, or subtitles for pre-recorded audio

- Model Customization
- OCR for Images
- People Detection
- Image Captions
- Object Detection
- Image Tagging
- Smart-copped Thumbnails
- Product Recognition
- Background Removal
- Image retravel
- Content Moderation
- Domain-Specific Content
- Brand detection
- Image Type Detection

Vision Service Features

Image Analysis – Image Tagging

mage Analysis can return content tags for thousands of recognizable objects, living beings, scenery, and actions that appear in images. Tags are not organized as a taxonomy and do not have inheritance hierarchies. A collection of content tags forms the foundation for an image description displayed as human readable language formatted in complete sentences. When tags are ambiguous or not common knowledge, the API response provides hints to clarify the meaning of the tag in context of a known setting.

After you upload an image or specify an image URL, the Analyze API can output tags based on the objects, living beings, and actions identified in the image. Tagging is not limited to the main subject, such as a person in the foreground, but also includes the setting (indoor or outdoor), furniture, tools, plants, animals, accessories, gadgets, and so on.


```
{
  "tags":[
    {
      "name":"grass",
      "confidence":0.9960499405860901
    },
    {
      "name":"outdoor",
      "confidence":0.9956876635551453
    },
    {
      "name":"building",
      "confidence":0.9893627166748047
    },
    {
      "name":"property",
      "confidence":0.9853052496910095
    },
    {
      "name":"plant",
      "confidence":0.9791355133056641
    },
    {
      "name":"sky",
      "confidence":0.9764555096626282
    },
    {
      "name":"home",
      "confidence":0.9732913970947266
    },
    {
      "name":"house",
      "confidence":0.9726772904396057
    },
    {
      "name":"real estate",
      "confidence":0.972320556640625
    },
    {
      "name":"yard",
      "confidence":0.9480282068252563
    },
    {
      "name":"siding",
      "confidence":0.945357620716095
    },
    {
      "name":"porch",
      "confidence":0.9410697221755981
    },
    {
      "name":"cottage",
      "confidence":0.9143695831298828
    },
    {
      "name":"tree",
      "confidence":0.9111741185188293
    },
    {
      "name":"farmhouse",
      "confidence":0.8988939523696899
    },
    {
      "name":"window",
      "confidence":0.894851565361023
    },
    {
      "name":"lawn",
      "confidence":0.8940501809120178
    },
    {
      "name":"backyard",
      "confidence":0.8931854963302612
    },
    {
      "name":"garden buildings",
      "confidence":0.885913610458374
    },
    {
      "name":"roof",
      "confidence":0.8695329427719116
    },
    {
      "name":"driveway",
      "confidence":0.8670971393585205
    },
    {
      "name":"land lot",
      "confidence":0.8564285039901733
    },
    {
      "name":"landscaping",
      "confidence":0.8540750741958618
    }
  ],
  "requestId":"d60ac02b-966d-4f62-bc24-fbb1fec8bd5d",
  "metadata":{
    "height":200,
    "width":300,
    "format":"Png"
  },
  "modelVersion":"2021-05-01"
}
```

Vision Service Features

Image Analysis – Image Categorization

In addition to tags and a description, Image Analysis can return the taxonomy-based categories detected in an image. Unlike tags, categories are organized in a parent/child hierarchy, and there are fewer of them (86, as opposed to thousands of tags). All category names are in English. Categorization can be done by itself or alongside the newer tags model.

- abstract_
  - abstract_net
  - abstract_nonphoto
  - abstract_rect
  - abstract_shape
  - abstract_texture
- animal_
  - animal_bird
  - animal_cat
  - animal_dog
  - animal_horse
  - animal_panda
- building_
  - building_arch
  - building_brickwall
  - building_church
  - building_corner
  - building_doorwindows
  - building_pillar
  - building_stair
  - building_street
- dark_
  - dark_fire
  - dark_fireworks
  - dark_light
- drink_
  - drink_can
- food_
  - food_bread
  - food_fastfood
  - food_grilled
  - food_pizza
- indoor_
  - indoor_churchwindow
  - indoor_court
  - indoor_doorwindows
  - indoor_marketstore
  - indoor_room
  - indoor_venue
- object_screen
- object_sculpture
- others_
- outdoor_
  - outdoor_city
  - outdoor_field
  - outdoor_grass
  - outdoor_house
  - outdoor_mountain
  - outdoor_oceanbeach
  - outdoor_playground
  - outdoor_pool
  - outdoor_railway
  - outdoor_road
  - outdoor_sportsfield
  - outdoor_stonerock
  - outdoor_street
  - outdoor_water
  - outdoor_waterside
- people_
  - people_baby
  - people_crowd
  - people_group
  - people_hand
  - people_many
  - people_portrait
  - people_show
  - people_swimming
  - people_tattoo
  - people_young
- plant_
  - plant_branch
  - plant_flower
  - plant_leaves
  - plant_tree
- sky_cloud
- sky_object
- sky_sun
- text_
  - text_mag
  - text_map
  - text_menu
  - text_sign
- trans_bicycle
- trans_bus
- trans_car
- trans_trainstation


```
{
   "categories": [
      {
         "name": "people_",
         "score": 0.81640625
      }
   ],
   "requestId": "bae7f76a-1cc7-4479-8d29-48a694974705",
   "metadata": {
      "height": 200,
      "width": 300,
      "format": "Jpeg"
   }
}
```

Vision Service Features

Image Analysis – Image Caption

mage captions in Image Analysis 4.0 (preview) are available through the **Caption and Dense Captions** features.

Caption generates a one sentence description for all image contents. Dense Captions provides more detail by generating one sentence descriptions of up to 10 regions of the image in addition to describing the whole image. Dense Captions also returns bounding box coordinates of the described image regions. Both these features use the latest groundbreaking Florence based AI models.

At this time, image captioning is available in English language only.

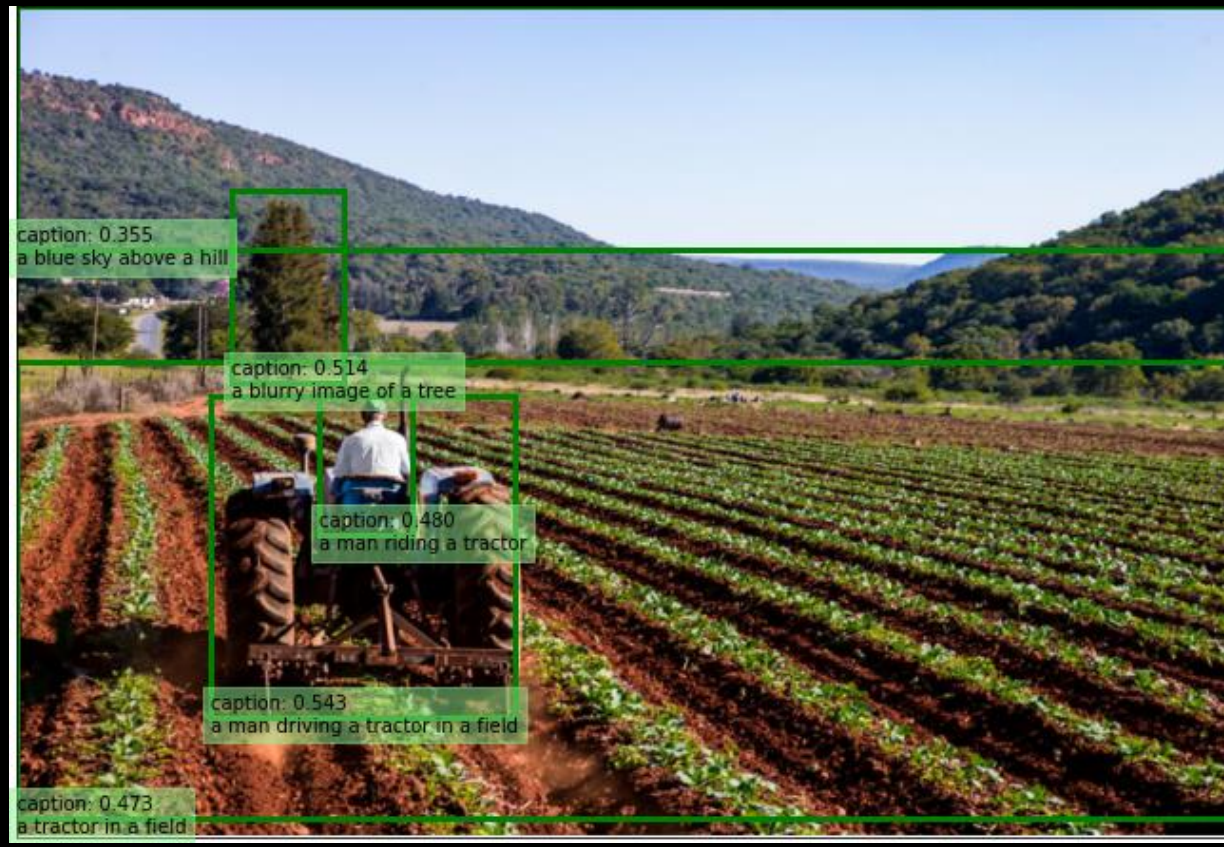
All captions contain gender terms: "man", "woman", "boy" and "girl" by default. You have the option to replace these terms with "person" in your results and receive **gender-neutral captions**. You can do so by setting the optional API request parameter, gender-neutral-caption to true in the request URL.


```
"captions": [
  {
    "text": "a man pointing at a screen",
    "confidence": 0.4891590476036072
  }
]
```

Vision Service Features

Image Analysis – Image Caption


```
{
  "denseCaptionsResult": {
    "values": [
      {
        "text": "a man driving a tractor in a farm",
        "confidence": 0.535620927810669,
        "boundingBox": {
          "x": 0,
          "y": 0,
          "w": 850,
          "h": 567
        }
      },
      {
        "text": "a man driving a tractor in a field",
        "confidence": 0.5428450107574463,
        "boundingBox": {
          "x": 132,
          "y": 266,
          "w": 209,
          "h": 219
        }
      },
      {
        "text": "a blurry image of a tree",
        "confidence": 0.5139822363853455,
        "boundingBox": {
          "x": 147,
          "y": 126,
          "w": 76,
          "h": 131
        }
      },
      {
        "text": "a man riding a tractor",
        "confidence": 0.4799223840236664,
        "boundingBox": {
          "x": 206,
          "y": 264,
          "w": 64,
          "h": 97
        }
      },
      {
        "text": "a blue sky above a hill",
        "confidence": 0.35495415329933167,
        "boundingBox": {
          "x": 0,
          "y": 0,
          "w": 837,
          "h": 166
        }
      },
      {
        "text": "a tractor in a field",
        "confidence": 0.47338250279426575,
        "boundingBox": {
          "x": 0,
          "y": 243,
          "w": 838,
          "h": 311
        }
      }
    ]
  },
  "modelVersion": "2023-02-01-preview",
  "metadata": {
    "width": 850,
    "height": 567
  }
}
```

Vision Service Features

Image Analysis – Object Detection

Object detection is similar to tagging, but the API returns the bounding box coordinates (in pixels) for each object found in the image. For example, if an image contains a dog, cat and person, the object detection operation will list those objects with their coordinates in the image. You can use this functionality to process the relationships between the objects in an image. It also lets you determine whether there are multiple instances of the same object in an image.

The object detection function applies tags based on the objects or living things identified in the image. There is currently no formal relationship between the tagging taxonomy and the object detection taxonomy. At a conceptual level, the object detection function only finds objects and living things, while the tag function can also include contextual terms like "indoor", which can't be localized with bounding boxes.

Limitations:

It's important to note the limitations of object detection so you can avoid or mitigate the effects of false negatives (missed objects) and limited detail.

- Objects are generally not detected if they're small (less than 5% of the image).
- Objects are generally not detected if they're arranged closely together (a stack of plates, for example).
- Objects are not differentiated by brand or product names (different types of sodas on a store shelf, for example). However, you can get brand information from an image by using the Brand detection feature.

Vision Service Features

Image Analysis – Object Detection


```
{
  "metadata":
  {
    "width": 1260,
    "height": 473
  },
  "objectsResult":
  {
    "values":
    [
      {
        "name": "kitchen appliance",
        "confidence": 0.501,
        "boundingBox": {"x":730,"y":66,"w":135,"h":85}
      },
      {
        "name": "computer keyboard",
        "confidence": 0.51,
        "boundingBox": {"x":523,"y":377,"w":185,"h":46}
      },
      {
        "name": "Laptop",
        "confidence": 0.85,
        "boundingBox": {"x":471,"y":218,"w":289,"h":226}
      },
      {
        "name": "person",
        "confidence": 0.855,
        "boundingBox": {"x":654,"y":0,"w":584,"h":473}
      }
    ]
  }
}
```


Vision Service Features

Image Analysis – Optical Character Recognition (OCR)

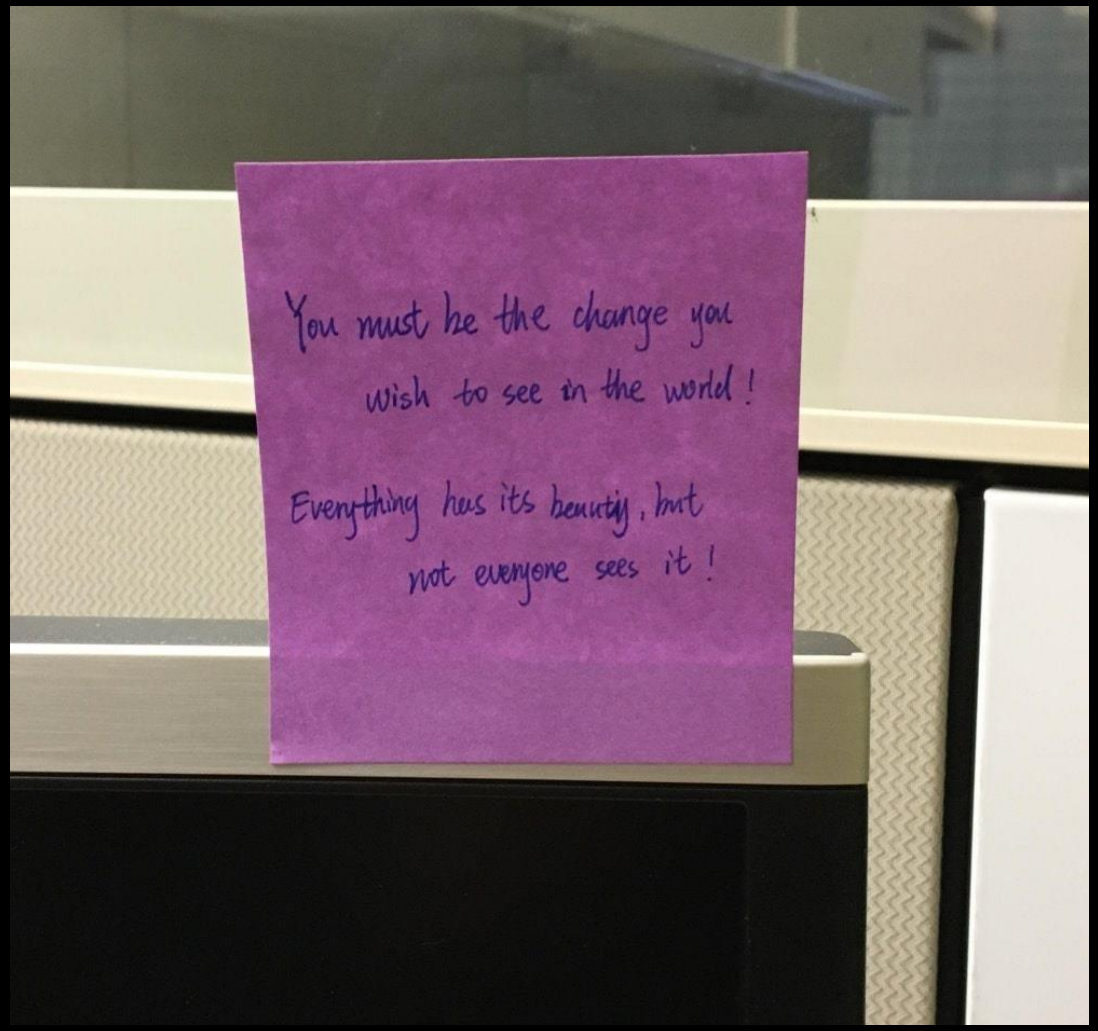
OCR traditionally started as a machine-learning-based technique for extracting text from in-the-wild and non-document images like product labels, user-generated images, screenshots, street signs, and posters. For several scenarios, such as single images that aren't text-heavy, you need a fast, synchronous API or service. This allows OCR to be embedded in near real-time user experiences to enrich content understanding and follow-up user actions with fast turn-around times.

The new Computer Vision Image Analysis 4.0 REST API offers the ability to extract printed or handwritten text from images in a unified performance-enhanced synchronous API that makes it easy to get all image insights including OCR results in a single API operation. The Read OCR engine is built on top of multiple deep learning models supported by universal script-based models for global language support.

Vision Service Features

Image Analysis – Optical Character Recognition (OCR)


```
{
   "metadata":
   {
      "width": 1000,
      "height": 945
   },
   "readResult":
   {
      "stringIndexType": "TextElements",
      "content": "You must be the change you\nWish to see in the world !\nEverything has its beauty , but\nnot everyone sees it !",
      "pages":
      [
         {
            "height": 945,
            "width": 1000,
            "angle": -1.099,
            "pageNumber": 1,
            "words":
            [
               {
                  "content": "You",
                  "boundingBox": [253,268,301,267,304,318,256,318],
                  "confidence": 0.998,
                  "span": {"offset":0,"length":3}
               },
               {
                  "content": "must",
                  "boundingBox": [310,266,376,265,378,316,313,317],
                  "confidence": 0.988,
                  "span": {"offset":4,"length":4}
               },
               {
                  "content": "be",
                  "boundingBox": [385,264,426,264,428,314,388,316],
                  "confidence": 0.928,
                  "span": {"offset":9,"length":2}
               },
               {
                  "content": "the",
                  "boundingBox": [435,263,494,263,496,311,437,314],
                  "confidence": 0.997,
                  "span": {"offset":12,"length":3}
               },
               {
                  "content": "change",
                  "boundingBox": [253,267,670,262,671,307,254,318],
                  "confidence": 0.998,
                  "span": {"offset":16,"length":6}
               },
               {
                  "content": "you",
                  "boundingBox": [326,343,691,332,693,369,327,382],
                  "confidence": 0.998,
                  "span": {"offset":23,"length":3}
               },
               {
                  "content": "Wish",
                  "boundingBox": [254,443,640,438,641,485,255,493],
                  "confidence": 0.998,
                  "span": {"offset":27,"length":4}
               },
               {
                  "content": "to",
                  "boundingBox": [364,512,658,496,660,534,365,549],
                  "confidence": 0.998,
                  "span": {"offset":32,"length":2}
               },
               {
                  "content": "see",
                  "boundingBox": [599,334,655,333,658,371,601,373],
                  "confidence": 0.998,
                  "span": {"offset":35,"length":3}
               },
               {
                  "content": "in",
                  "boundingBox": [663,333,687,333,690,370,666,371],
                  "confidence": 0.915,
                  "span": {"offset":39,"length":2}
               },
               {
                  "content": "the",
                  "boundingBox": [255,446,371,441,372,490,256,494],
                  "confidence": 0.97,
                  "span": {"offset":42,"length":3}
               },
               {
                  "content": "world",
                  "boundingBox": [599,334,655,333,658,371,601,373],
                  "confidence": 0.998,
                  "span": {"offset":46,"length":5}
               },
               {
                  "content": "!",
                  "boundingBox": [663,333,687,333,690,370,666,371],
                  "confidence": 0.915,
                  "span": {"offset":52,"length":1}
               },
               {
                  "content": "Everything",
                  "boundingBox": [255,446,371,441,372,490,256,494],
                  "confidence": 0.97,
                  "span": {"offset":54,"length":10}
               },
               {
                  "content": "has",
                  "boundingBox": [380,441,421,440,421,488,381,489],
                  "confidence": 0.793,
                  "span": {"offset":65,"length":3}
               },
               {
                  "content": "its",
                  "boundingBox": [430,440,471,439,471,487,431,488],
                  "confidence": 0.998,
                  "span": {"offset":69,"length":3}
               },
               {
                  "content": "beauty",
                  "boundingBox": [480,439,552,439,552,485,481,487],
                  "confidence": 0.296,
                  "span": {"offset":73,"length":6}
               },
               {
                  "content": ",",
                  "boundingBox": [561,439,571,439,571,485,562,485],
                  "confidence": 0.742,
                  "span": {"offset":79,"length":1}
               },
               {
                  "content": "but",
                  "boundingBox": [254,443,640,438,641,485,255,493],
                  "confidence": 0.998,
                  "span": {"offset":81,"length":3}
               },
               {
                  "content": "not",
                  "boundingBox": [364,512,658,496,660,534,365,549],
                  "confidence": 0.998,
                  "span": {"offset":86,"length":3}
               },
               {
                  "content": "everyone",
                  "boundingBox": [326,343,691,332,693,369,327,382],
                  "confidence": 0.998,
                  "span": {"offset":90,"length":8}
               },
               {
                  "content": "sees",
                  "boundingBox": [599,334,655,333,658,371,601,373],
                  "confidence": 0.998,
                  "span": {"offset":99,"length":4}
               },
               {
                  "content": "it",
                  "boundingBox": [663,333,687,333,690,370,666,371],
                  "confidence": 0.915,
                  "span": {"offset":104,"length":2}
               },
               {
                  "content": "!",
                  "boundingBox": [255,446,371,441,372,490,256,494],
                  "confidence": 0.97,
                  "span": {"offset":106,"length":1}
               }
            ],
            "lines":
            [
               {
                  "content": "You must be the change you",
                  "boundingBox": [253,267,670,262,671,307,254,318],
                  "spans": [{"offset":0,"length":26}]
               },
               {
                  "content": "Wish to see in the world !",
                  "boundingBox": [326,343,691,332,693,369,327,382],
                  "spans": [{"offset":27,"length":26}]
               },
               {
                  "content": "Everything has its beauty , but",
                  "boundingBox": [254,443,640,438,641,485,255,493],
                  "spans": [{"offset":54,"length":31}]
               },
               {
                  "content": "not everyone sees it !",
                  "boundingBox": [364,512,658,496,660,534,365,549],
                  "spans": [{"offset":86,"length":22}]
               }
            ]
         }
      ],
      "styles":
      [
         {
            "isHandwritten": true,
            "spans":
            [
               {
                  "offset": 0,
```

Vision Service Features

Image Analysis – Smart-cropped Thumbnails

A thumbnail is a reduced-size representation of an image. Thumbnails are used to represent images and other data in a more economical, layout-friendly way. The Azure AI Vision API uses smart cropping to create intuitive image thumbnails that include the most important regions of an image with priority given to any detected faces.

The Azure AI Vision smart-cropping utility takes one or more aspect ratios in the range [0.75, 1.80] and returns the bounding box coordinates (in pixels) of the region(s) identified. Your app can then crop and return the image using those coordinates.

This feature uses face detection to help determine important regions in the image. The detection does not involve distinguishing one face from another face, predicting or classifying facial attributes, or creating a facial template (a unique set of numbers generated from an image that represents the distinctive features of a face).

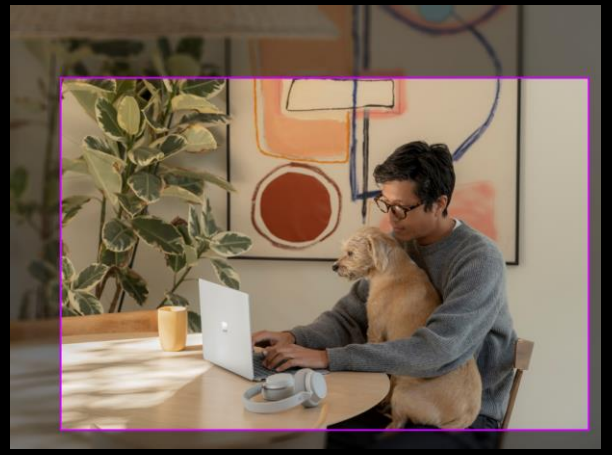
Vision Service Features

Image Analysis – Smart-cropped Thumbnails

0.75

1.00

1.50



Vision Service Features

Image Analysis – Background Removal

The Image Analysis service can divide images into multiple segments or regions to help the user identify different objects or parts of the image. Background removal creates an alpha matte that separates the foreground object from the background in an image.

This feature provides two possible outputs based on the customer's needs:

- The foreground object of the image without the background. This edited image shows the foreground object and makes the background transparent, allowing the foreground to be placed on a new background.

- An alpha matte that shows the opacity of the detected foreground object. This matte can be used to separate the foreground object from the background for further processing.

This service is currently in preview, and the API may change in the future.

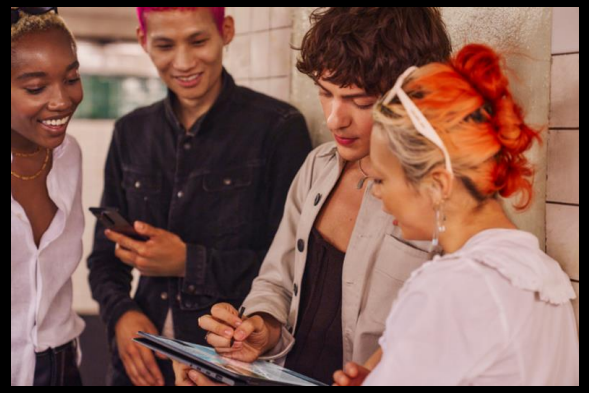
Limitations

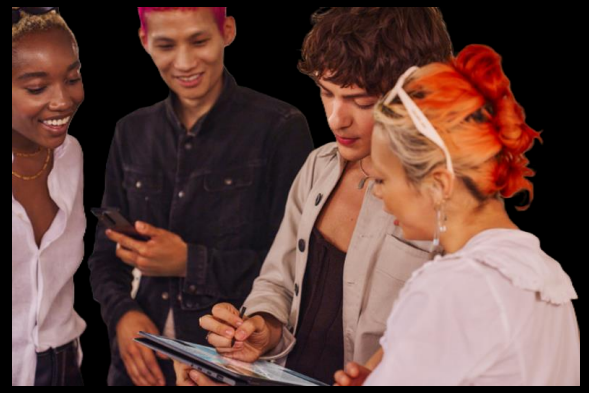
It's important to note the limitations of background removal:

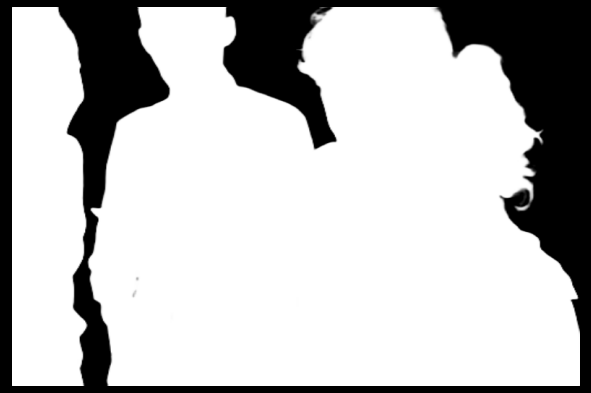
- Background removal works best for categories such as people and animals, buildings and environmental structures, furniture, vehicles, food, text and graphics, and personal belongings.
- Objects that aren't prominent in the foreground may not be identified as part of the foreground.
- Images with thin and detailed structures, like hair or fur, may show some artifacts when overlaid on backgrounds with strong contrast to the original background.
- The latency of the background removal operation will be higher, up to several seconds, for large images. We suggest you experiment with integrating both modes into your workflow to find the best usage for your needs (for instance, calling background removal on the original image versus calling foreground matting on a downsampled version of the image, then resizing the alpha matte to the original size and applying it to the original image).

Vision Service Features

Image Analysis – Background Removal







Vision Service Features

Image Analysis – Image Description

Azure AI Vision can analyze an image and generate a human-readable phrase that describes its contents. The algorithm returns several descriptions based on different visual features, and each description is given a confidence score. The final output is a list of descriptions ordered from highest to lowest confidence.

At this time, English is the only supported language for image description.


```
{
  "description":{
    "tags":[
      "outdoor",
      "city",
      "white"
    ],
    "captions":[
      {
        "text":"a city with tall buildings",
        "confidence":0.48468858003616333
      }
    ]
  },
  "requestId":"7e5e5cac-ef16-43ca-a0c4-02bd49d379e9",
  "metadata":{
    "height":300,
    "width":239,
    "format":"Png"
  },
  "modelVersion":"2021-05-01"
}
```

Vision Service Features

Image Analysis – Domain Specific Content

In addition to tagging and high-level categorization, Azure AI Vision also supports further domain-specific analysis using models that have been trained on specialized data.

There are two ways to use the domain-specific models: by themselves (scoped analysis) or as an enhancement to the categorization feature.

Scoped Analysis

You can analyze an image using only the chosen domain-specific model by calling the Models/<model>/Analyze API.

Enhanced categorization analysis

You can also use domain-specific models to supplement general image analysis. You do this as part of high-level categorization by specifying domain-specific models in the details parameter of the Analyze API call.
In this case, the 86-category taxonomy classifier is called first. If any of the detected categories have a matching domain-specific model, the image is passed through that model as well and the results are added.
The following JSON response shows how domain-specific analysis can be included as the detail node in a broader categorization analysis.

Model Name	Description
celebrities	Celebrity recognition, supported for images classified in the people_ category
landmarks	Landmark recognition, supported for images classified in the outdoor_ or building_ categories

Vision Service Features

Image Analysis – Domain Specific Content

Scoped Analysis Example


```
{
  "result": {
    "celebrities": [{
      "faceRectangle": {
        "top": 391,
        "left": 318,
        "width": 184,
        "height": 184
      },
      "name": "Satya Nadella",
      "confidence": 0.99999856948852539
    }]
  },
  "requestId": "8217262a-1a90-4498-a242-68376a4b956b",
  "metadata": {
    "width": 800,
    "height": 1200,
    "format": "Jpeg"
  }
}
```


Vision Service Features

Image Analysis – Domain Specific Content

Enhanced categorization analysis Example


```
"categories":[
  {
    "name":"abstract_",
    "score":0.00390625
  },
  {
    "name":"people_",
    "score":0.83984375,
    "detail":{
      "celebrities":[
        {
          "name":"Satya Nadella",
          "faceRectangle":{
            "left":597,
            "top":162,
            "width":248,
            "height":248
          },
          "confidence":0.999028444
        }
      ],
      "landmarks":[
        {
          "name":"Forbidden City",
          "confidence":0.9978346
        }
      ]
    }
  }
]
```


Vision Service Features

Face

The Azure AI Face service provides AI algorithms that detect, recognize, and analyze human faces in images. Facial recognition software is important in many different scenarios, such as identity verification, touchless access control, and face blurring for privacy.

You can use the Face service through a client library SDK or by calling the REST API directly

Face detection is required as a first step in all the other scenarios. The Detect API detects human faces in an image and returns the rectangle coordinates of their locations. It also returns a unique ID that represents the stored face data. This is used in later operations to identify or verify faces.

Optionally, face detection can extract a set of face-related attributes, such as head pose, age, emotion, facial hair, and glasses. These attributes are general predictions, not actual classifications. Some attributes are useful to ensure that your application is getting high-quality face data when users add themselves to a Face service. For example, your application could advise users to take off their sunglasses if they're wearing sunglasses.

Example Use Cases

Identity verification: Verify someone's identity against a government-issued ID card like a passport or driver's license or other enrollment image. You can use this verification to grant access to digital or physical services or to recover an account. Specific access scenarios include opening a new account, verifying a worker, or administering an online assessment. Identity verification can be done once when a person is onboarded, and repeated when they access a digital or physical service.

Touchless access control: Compared to today’s methods like cards or tickets, opt-in face identification enables an enhanced access control experience while reducing the hygiene and security risks from card sharing, loss, or theft. Facial recognition assists the check-in process with a human in the loop for check-ins in airports, stadiums, theme parks, buildings, reception kiosks at offices, hospitals, gyms, clubs, or schools.

Face redaction: Redact or blur detected faces of people recorded in a video to protect their privacy.

Vision Service Features

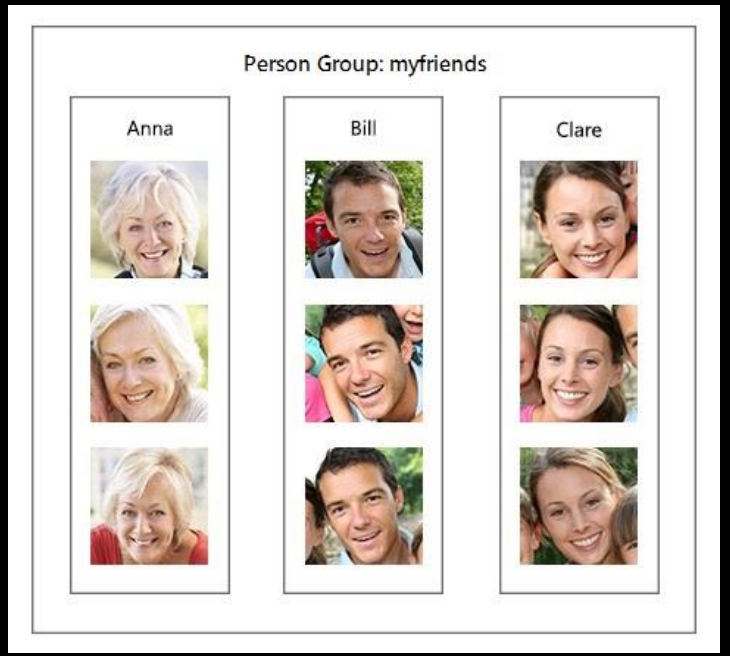
Face - Identity verification

Face identification can address "one-to-many" matching of one face in an image to a set of faces in a secure repository. Match candidates are returned based on how closely their face data matches the query face. This scenario is used in granting building or airport access to a certain group of people or verifying the user of a device.

The following image shows an example of a database named "myfriends". Each group can contain up to 1 million different person objects. Each person object can have up to 248 faces registered.

The verification operation answers the question, "Do these two faces belong to the same person?".

Verification is also a "one-to-one" matching of a face in an image to a single face from a secure repository or photo to verify that they're the same individual. Verification can be used for Identity Verification, such as a banking app that enables users to open a credit account remotely by taking a new picture of themselves and sending it with a picture of their photo ID.



Vision Service Features

Face – Find Similar Faces

The Find Similar operation does face matching between a target face and a set of candidate faces, finding a smaller set of faces that look similar to the target face. This is useful for doing a face search by image.

The service supports two working modes, matchPerson and matchFace. The matchPerson mode returns similar faces after filtering for the same person by using the Verify API. The matchFace mode ignores the same-person filter. It returns a list of similar candidate faces that may or may not belong to the same person.

Vision Service Features

Face – Group Faces

The Group operation divides a set of unknown faces into several smaller groups based on similarity. Each group is a disjoint proper subset of the original set of faces. It also returns a single "messyGroup" array that contains the face IDs for which no similarities were found.

All of the faces in a returned group are likely to belong to the same person, but there can be several different groups for a single person. Those groups are differentiated by another factor, such as expression, for example.

Vision Service Features

Spatial Analysis

You can use Azure AI Vision Spatial Analysis to detect the presence and movements of people in video. Ingest video streams from cameras, extract insights, and generate events to be used by other systems. The service can do things like count the number of people entering a space or measure compliance with face mask and social distancing guidelines. By processing video streams from physical spaces, you're able to learn how people use them and maximize the space's value to your organization.

Main Features

Spatial Analysis ingests video then detects people in the video. After people are detected, the system tracks the people as they move around over time then generates events as people interact with regions of interest. All operations give insights from a single camera's field of view.

- People Counting
- Entrance Counting
- Social Distancing and Face Mask Detection

Input Requirements

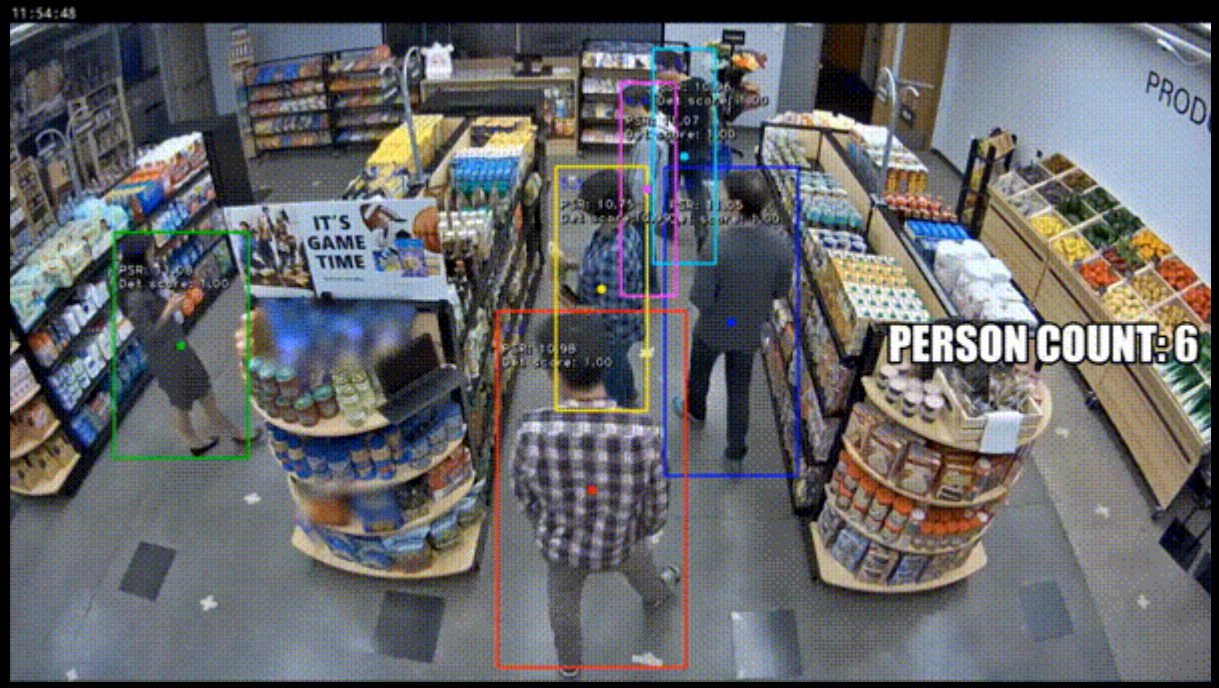
Spatial Analysis works on videos that meet the following requirements:

- The video must be in RTSP, rawvideo, MP4, FLV, or MKV format.
- The video codec must be H.264, HEVC(H.265), rawvideo, VP9, or MPEG-4.

Vision Service Features

Spatial Analysis – People Counting

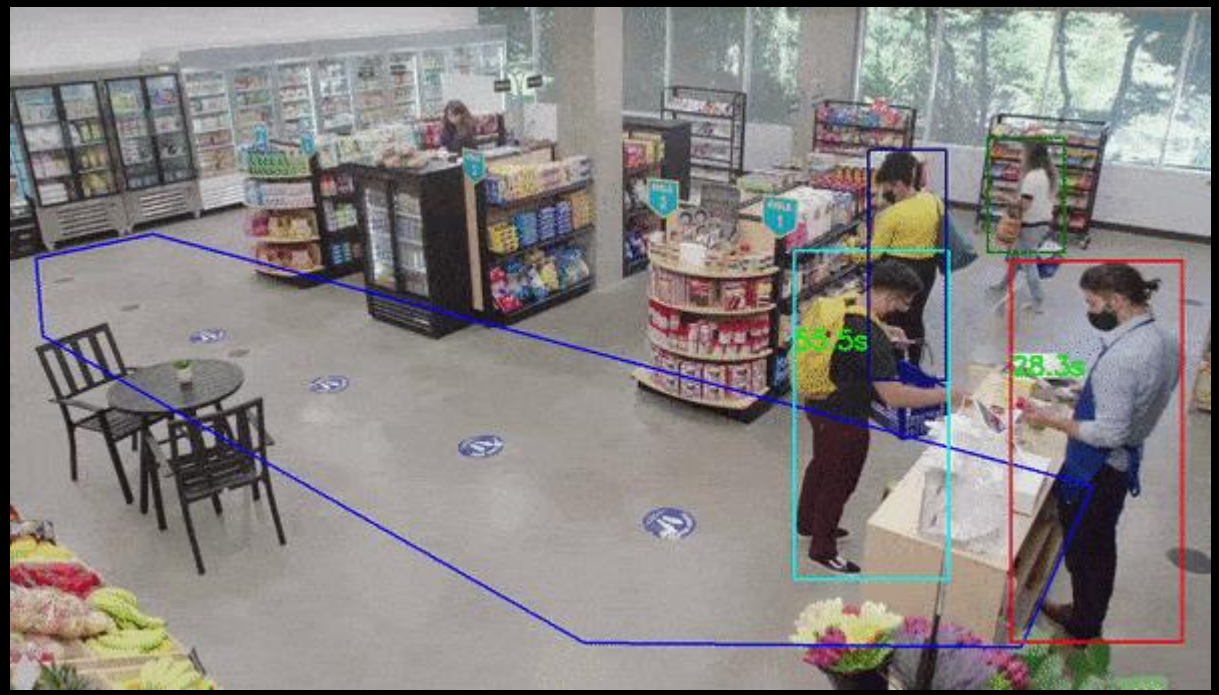
his operation counts the number of people in a specific zone over time using the PersonCount operation. It generates an independent count for each frame processed without attempting to track people across frames. This operation can be used to estimate the number of people in a space or generate an alert when a person appears.



Vision Service Features

Spatial Analysis – Entrance Detection

This feature monitors how long people stay in an area or when they enter through a doorway. This monitoring can be done using the PersonCrossingPolygon or PersonCrossingLine operations. In retail scenarios, these operations can be used to measure wait times for a checkout line or engagement at a display. Also, these operations could measure foot traffic in a lobby or a specific floor in other commercial building scenarios.



Vision Service Features

Spatial Analysis – Social Distancing and Face Mask Detection

This feature analyzes how well people follow social distancing requirements in a space. The system uses the PersonDistance operation to automatically calibrates itself as people walk around in the space. Then it identifies when people violate a specific distance threshold (6 ft. or 10 ft.).

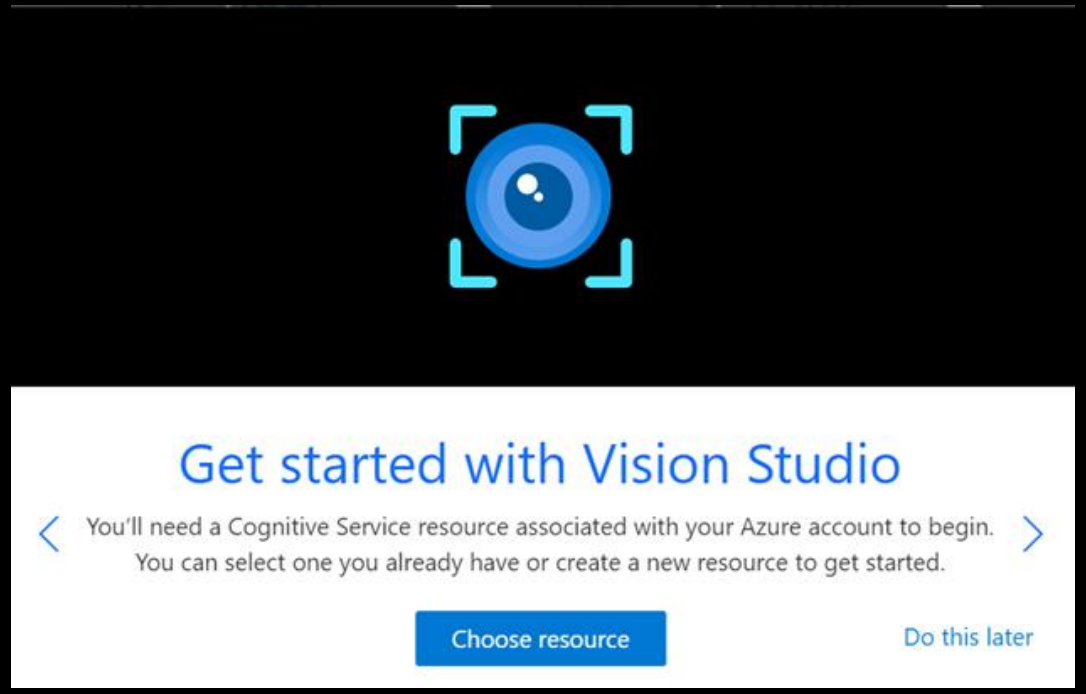
Spatial Analysis can also be configured to detect if a person is wearing a protective face covering such as a mask. A mask classifier can be enabled for the PersonCount, PersonCrossingLine, and PersonCrossingPolygon operations by configuring the ENABLE_FACE_MASK_CLASSIFIER parameter.

Vision Service Features

Vision Studio

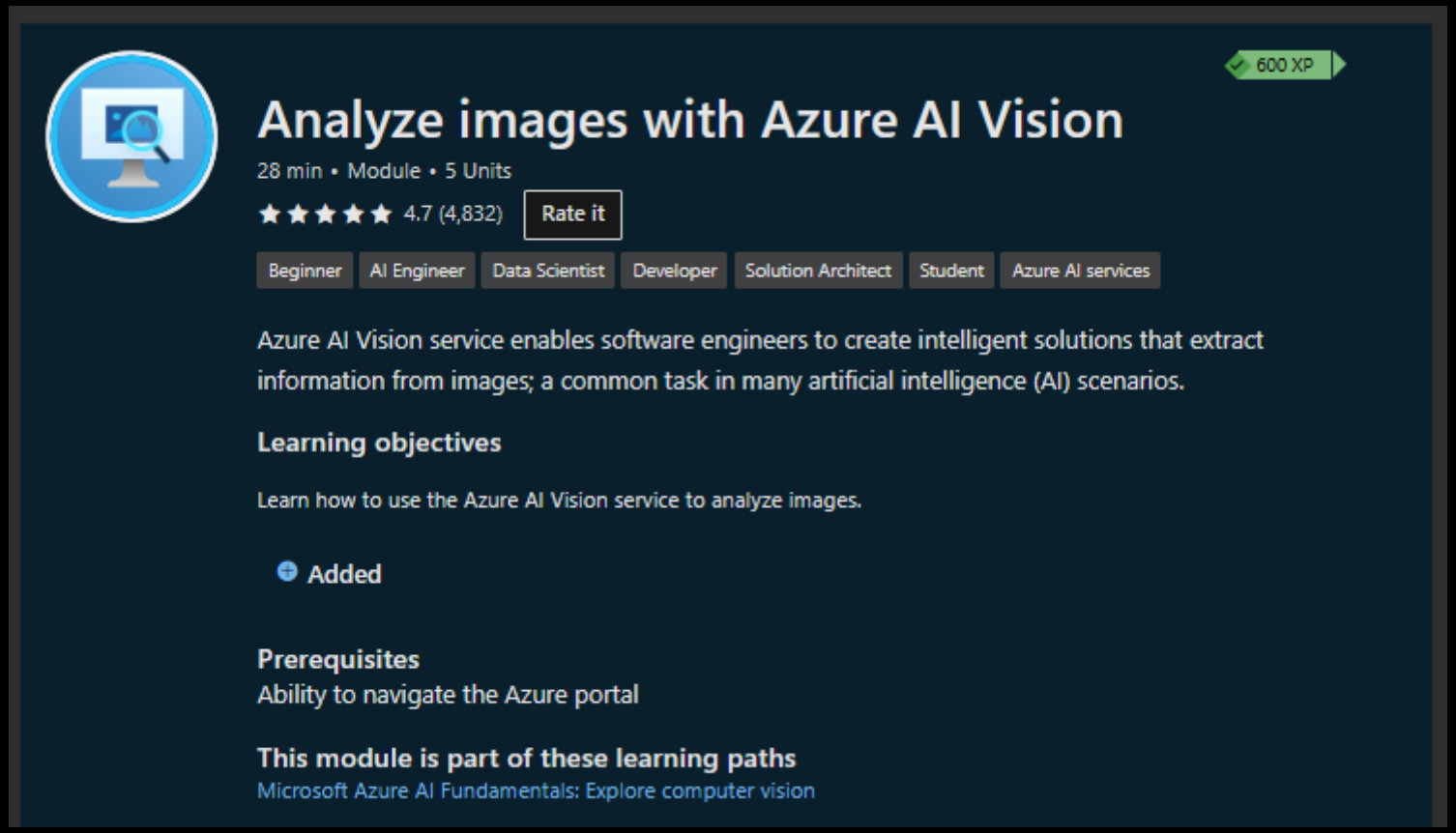
Vision Studio is a set of UI-based tools that lets you explore, build, and integrate features from Azure AI Vision.

Vision Studio provides you with a platform to try several service features and sample their returned data in a quick, straightforward manner. Using Studio, you can start experimenting with the services and learning what they offer without needing to write any code. Then, use the available client libraries and REST APIs to get started embedding these services into your own applications.



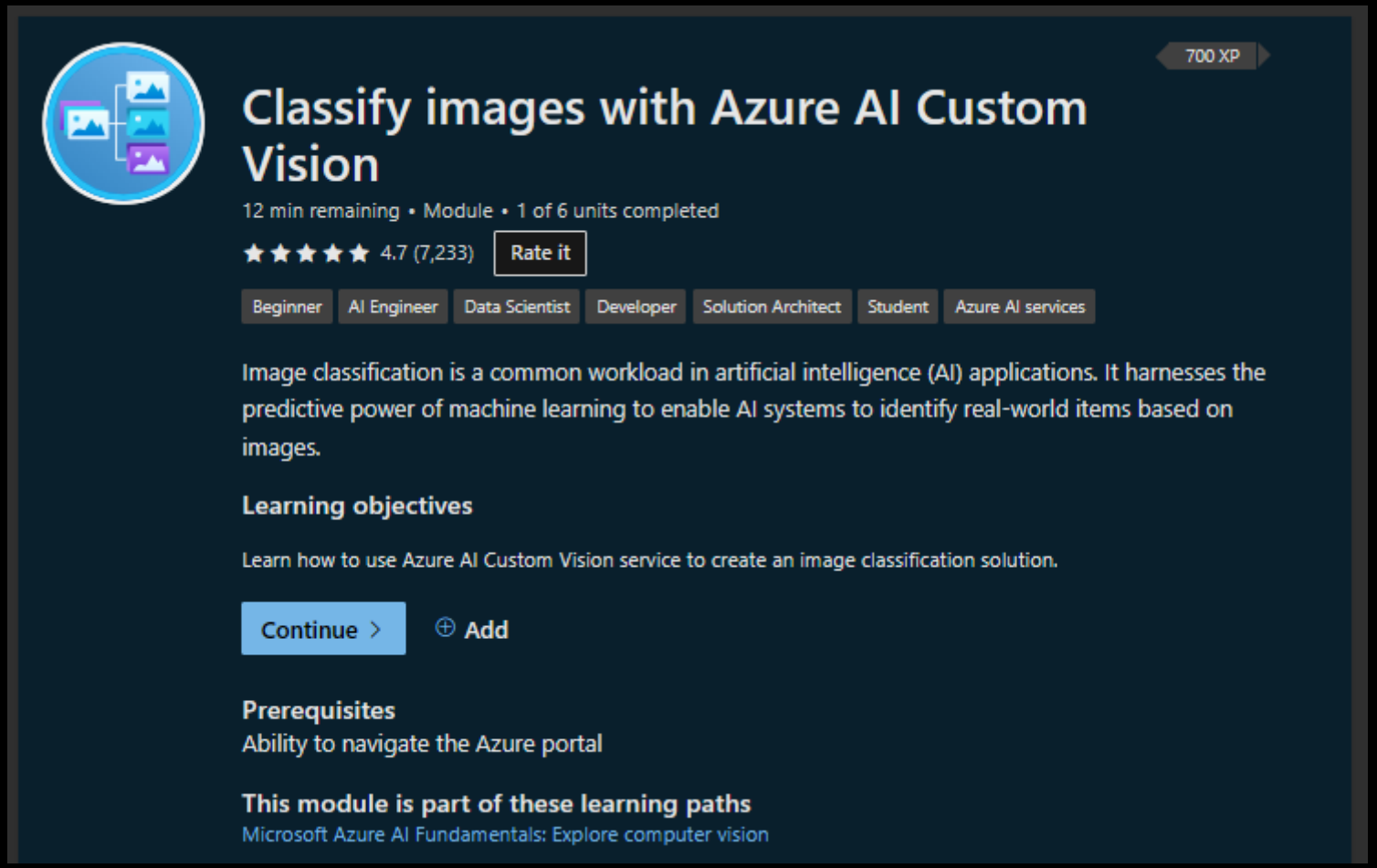
https://portal.vision.cognitive.azure.com/

Exercice 1



https://learn.microsoft.com/en-us/training/modules/analyze-images-computer-vision/

Exercice 2

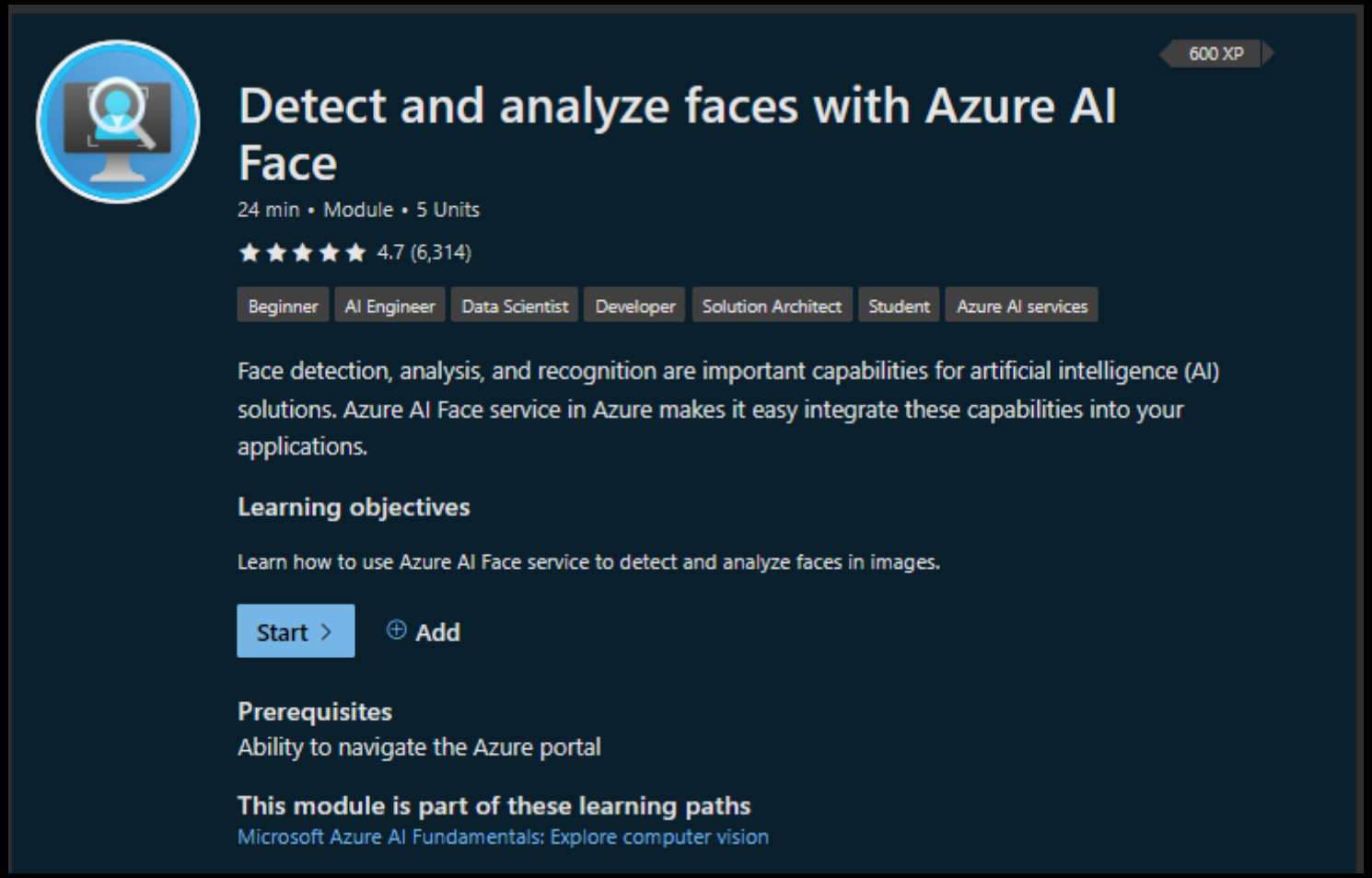


https://learn.microsoft.com/en-us/training/modules/classify-images-custom-vision/

Exercice 3

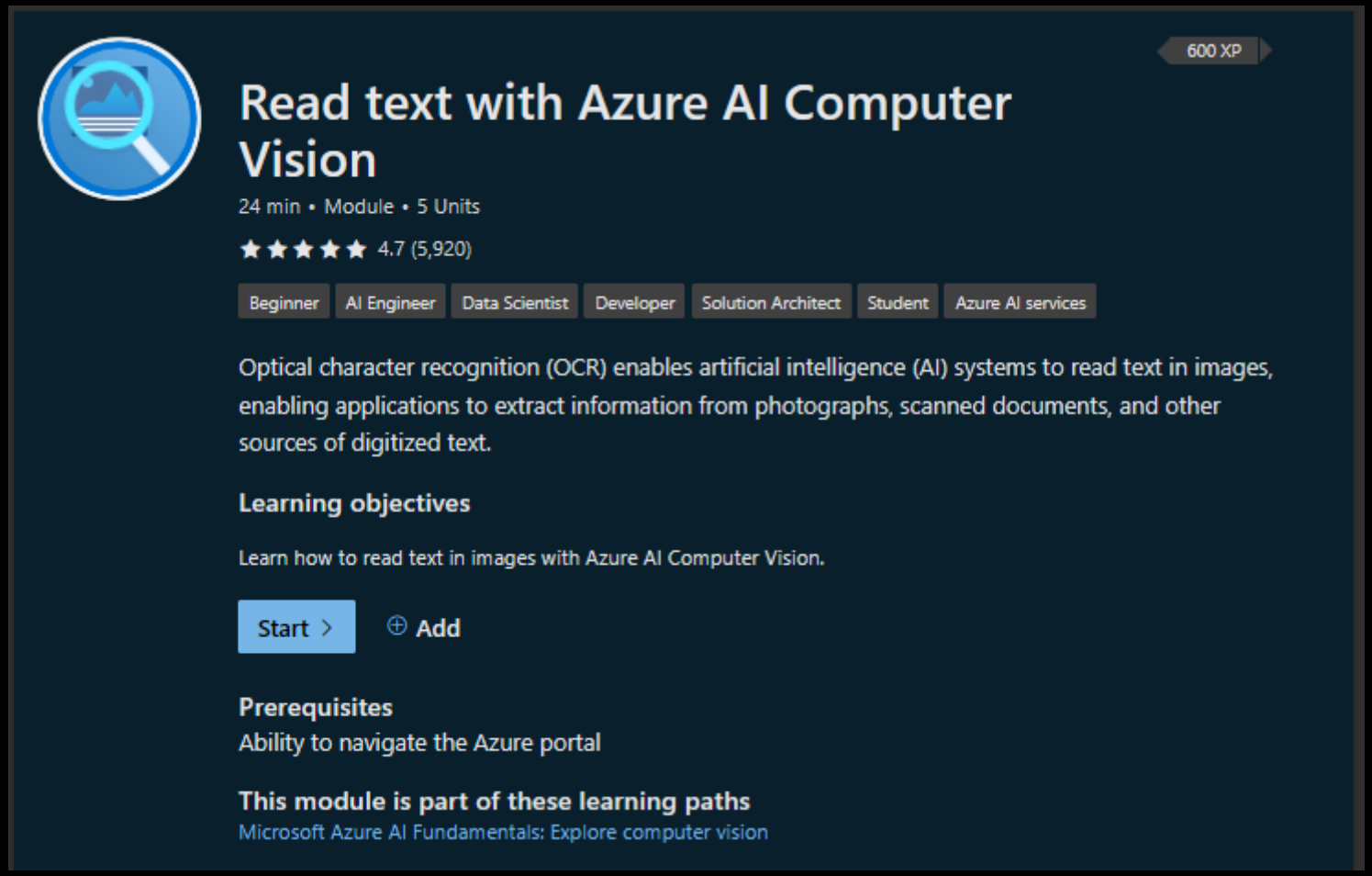
https://learn.microsoft.com/en-us/training/modules/detect-objects-images-custom-vision/

Exercice 4



https://learn.microsoft.com/en-us/training/modules/detect-analyze-faces/

Exercice 5



https://learn.microsoft.com/en-us/training/modules/read-text-computer-vision/

Microsoft Azure

Invent with purpose.

Thank you