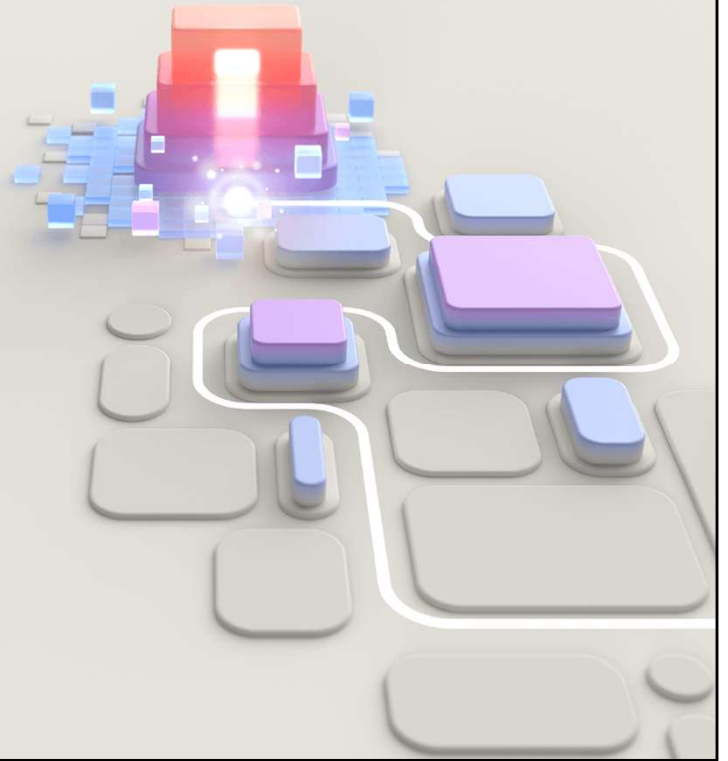





# Develop computer vision solutions with Azure AI Vision



© Copyright Microsoft Corporation. All rights reserved.

# Agenda

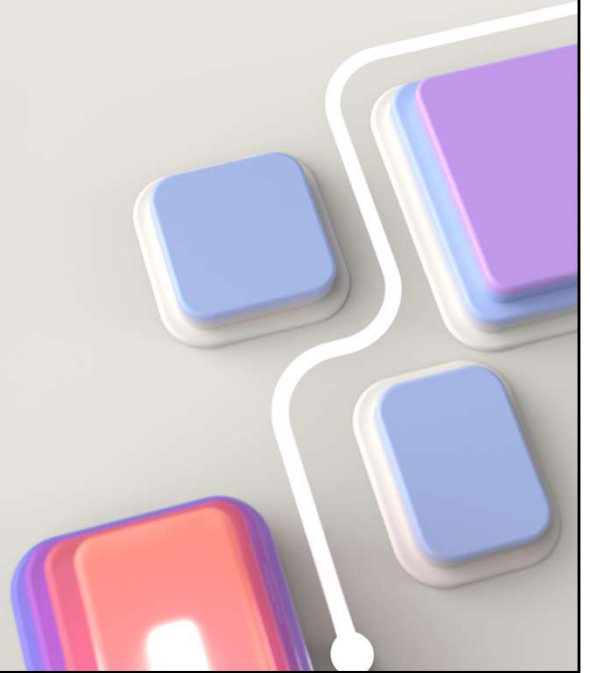


- Analyze and manipulate images
- Create a custom vision model
- Detect and recognize faces
- Analyze video

© Copyright Microsoft Corporation. All rights reserved.

# Analyze and manipulate images

© Copyright Microsoft Corporation. All rights reserved.



## Learning Objectives

After completing this module, you will be able to:

- 1 Understand features and functionality of Image Analysis
- 2 Perform Optical Character Recognition (OCR)
- 3 Connect an app to Image Analysis APIs

© Copyright Microsoft Corporation. All rights reserved.

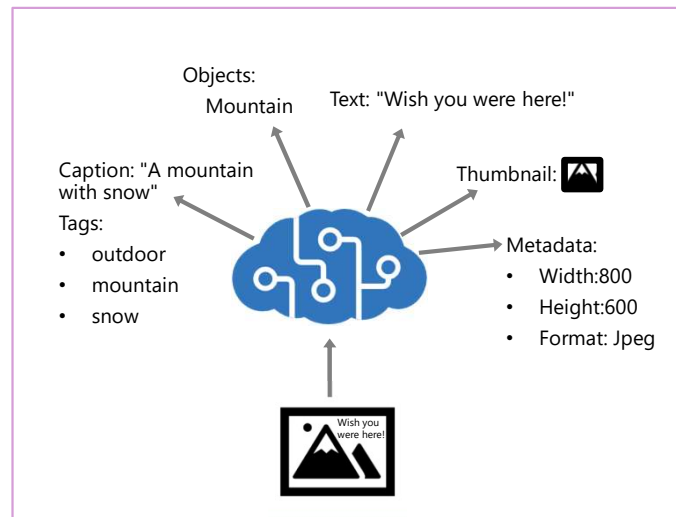
# Azure AI Vision – Image Analysis

## Image analysis:

- Caption and tag generation
- Object detection
- People detection
- Optical character recognition
- Smart crop thumbnails
- Multi-modal embeddings

## Can be used as:

- Standalone **Azure AI Vision** resource
  - Multi-service **Azure AI Services** resource
- \* Some new features are limited to specific regions



© Copyright Microsoft Corporation. All rights reserved.

Image analysis is the core of AI Vision, and should be framed as such. With version 4.0 more features are included within one call, specifically OCR. We'll only be covering 4.0 in this course, but will mention features in 3.2 (namely Face)

## Image Analysis APIs

- Single **Analyze** call to retrieve specified features in **VisualFeatures** enum
  - VisualFeatures.Caption
  - VisualFeatures.DenseCaptions
  - VisualFeatures.Tags
  - VisualFeatures.Objects
  - VisualFeatures.SmartCrops
  - VisualFeatures.People
  - VisualFeatures.Read
- SDKs define the **client**, then call the **analyze()** function from it
  - Client defines endpoint and key of resource
- **Analyze()** needs
  - Image data from file or URL
  - Visual features to analyze for
  - (Optional) Analysis options: Which features, language, and other options for the analysis

### REST

```
https://<endpoint>/computervision/imageanalysis:analyze?
features=caption,people&model-name=latest&
language=en&api-version={version}
```

### C#

```
ImageAnalysisResult result = client.Analyze(
    new Uri("<uri-to-image>"),
    VisualFeatures.Caption |
    VisualFeatures.People,
    analysisOptions // Optional
);
```

### Python

```
result = client.analyze(
    image_url="<uri-to-image>",
    visual_features=[
        VisualFeatures.Caption,
        VisualFeatures.People,
    ],
    <analysis_options> # Optional
)
```

Här är en kort beskrivning av de **VisualFeatures** som nämns i bilden:

1. **VisualFeatures.Caption** – Genererar en beskrivande text (caption) för bilden baserat på innehållet.
  2. **VisualFeatures.DenseCaptions** – Skapar flera detaljerade beskrivningar för olika delar av bilden.
  3. **VisualFeatures.Tags** – Identifierar och returnerar nyckelord (tags) relaterade till objekt och scener i bilden.
  4. **VisualFeatures.Objects** – Upptäcker och identifierar specifika objekt i bilden, inklusive deras position.
  5. **VisualFeatures.SmartCrops** – Analyserar bilden och genererar rekommenderade beskärningar för att optimera dess komposition.
  6. **VisualFeatures.People** – Identifierar och markerar personer i bilden, men inte nödvändigtvis deras identitet.
  7. **VisualFeatures.Read** – Extraherar och tolkar text från bilden genom **OCR (Optical Character Recognition)**.
- Dessa funktioner används i **Azure Computer Vision API** för att analysera bilder och extrahera relevant information beroende på vad användaren vill identifiera.

## Image Analysis Options

- Analysis options
  - Cropping aspect ratios
  - Gender neutral caption
  - Language
  - Model version

```
C#
ImageAnalysisClient client = new ImageAnalysisClient(
    new Uri(endpoint),
    new AzureKeyCredential(key));
ImageAnalysisOptions options = new ImageAnalysisOptions {
    GenderNeutralCaption = true,
    Language = "en"
};
ImageAnalysisResult result = client.Analyze(
    imageURL,
    visualFeatures,
    options
);

Python
client = ImageAnalysisClient(
    endpoint=endpoint,
    credential=AzureKeyCredential(key))
result = client.analyze(
    image_url=image_url,
    visual_features=visual_features,
    gender_neutral_caption=True,
    language="en"
)
```

Bilden handlar om **Image Analysis Options** i Azure Computer Vision och listar olika analysalternativ som kan användas vid bildanalys. Här är vad de olika alternativen innebär:

### 1. Cropping Aspect Ratios

- Detta används för att **generera beskärningar av bilden** med vissa förhållanden mellan bredd och höjd (t.ex. 16:9, 4:3).
- Användbart för att optimera bilder för olika enheter eller skärmstorlekar.

### 2. Gender Neutral Caption

- När detta är **true**, genererar systemet **könsneutrala bildbeskrivningar**.
- Istället för "A man is walking" kan det bli "A person is walking".
- Bra för att minska bias i bildigenkänningen.

### 3. Language

- Anger vilket **språk** som ska användas för genererade bildtexter och analysresultat.
- Exempel: "en" för engelska, "fr" för franska.

### 4. Model Version

- Specificerar vilken **modellversion** av bildanalys-API:t som ska användas.
- Standard är ofta den senaste modellen, men du kan välja en äldre version för konsekvens i resultat.

### Kodexempel i bilden

- **C#** och **Python** används för att anropa **Image Analysis API** med dessa options.
- De skapar en klient (ImageAnalysisClient), anger en bild-URL och skickar med options som GenderNeutralCaption och Language.

## Image Analysis Result

- Successful image analysis returns JSON (REST) or an object (SDKs)
- Results may have one or several layers of depth
  - Tags > values[ ] > name
  - Text > lines > words

```
{
  "captionResult": {
    "text": "a man pointing at a screen",
    "confidence": 0.4891590476036072
  },
  "objectsResult": {
    "values": [
      {
        "name": "string",
        "confidence": 0.0
      }
    ]
  },
  "smartCropsResult": {
    "values": [
      {
        "aspectRatio": 0.0,
        "boundingBox": {
          "x": 0,
          "y": 0,
          "w": 0,
          "h": 0
        }
      }
    ]
  },
  ...
}
```

Den här bilden beskriver **Image Analysis Result**, dvs. JSON-strukturen som returneras efter en bildanalys i **Azure Computer Vision API**. Här är vad de olika sektionerna betyder:

### 1. captionResult (Bildbeskrivning)

- "captionResult": { "text": "a man pointing at a screen", "confidence": 0.4891590476036072 } API:t har analyserat bilden och genererat en **beskrivande text**:  
"a man pointing at a screen".

- **Confidence**-värdet (0.49) anger hur säker modellen är på sin analys (skala 0-1, där 1 = 100% säker).

### 2. objectsResult (Objektidentifiering)

- "objectsResult": { "values": [ { "name": "string", "confidence": 0.0 } ] } Identifierar **objekt i bilden**.
- "name" representerar detekterade objektets kategori (ex: "car", "person", "dog").  
(Här står det "string", vilket bara är en placeholder i exemplet).
- "confidence" visar hur säker modellen är på identifieringen (här 0.0, alltså ingen identifiering hittad).

### 3. smartCropsResult (Föreslagna beskärningar)

- "smartCropsResult": { "values": [ { "aspectRatio": 0.0, "boundingBox": { "x": 0, "y": 0, "w": 0, "h": 0 } } ] } **Smart Crops** används för att rekommendera **bästa beskärning av bilden**.
- "boundingBox" anger koordinaterna (x, y, w, h) för den föreslagna beskärningen.
- "aspectRatio" representerar förhållandet mellan bredd och höjd för den föreslagna beskärningen (t.ex. 16:9 eller 4:3).

### Sammanfattning

1. **captionResult** → Genererar en beskrivande text för bilden.

2. **objectsResult** → Identifierar objekt och deras sannolikhet.

3. **smartCropsResult** → Rekommenderar optimala bildbeskärningar.

Detta JSON-svar används oftast för att extrahera information från en bild i en applikation.



## Demo – Explore features in Vision Studio



**Search photos with image retrieval**

**Create smart-cropped images**

© Copyright Microsoft Corporation. All rights reserved.

Instructor demo – nothing set but suggest these two features since we use the rest in an upcoming lab. Open <https://portal.vision.cognitive.azure.com/> and select “Try it out” under these features and use the provided images. This can be skipped.

## Lab – Analyze Images with Azure AI Vision



**Generate captions**

**Create tags**

**Detect objects and people**

© Copyright Microsoft Corporation. All rights reserved.

Use the provided lab VM, or use the instructions here: <https://microsoftlearning.github.io/mslearn-ai-vision/Instructions/Exercises/01-analyze-images.html>

# Azure AI Vision - OCR

Use **Image analysis** with READ feature

Two ways to perform OCR:

- Vision Read: General, non-document images with smaller amounts of text. Synchronous API.
- Document Intelligence: Ideal for larger text heavy documents. Asynchronous API.
  - Document intelligence is covered later in this course

Results in JSON (REST) or object (SDK) of similar structure

```
[{
  "lines": [{
    "text": "You must be the change you",
    "boundingPolygon": [
      {
        "x": 251,
        "y": 265
      },
      {
        "x": 673,
        "y": 260
      },
      {
        "x": 674,
        "y": 308
      },
      {
        "x": 252,
        "y": 318
      }
    ]
  }],
  "words": [
    {
      "text": "You",
      "boundingPolygon": ...,
      "confidence": 0.996
    },
    {
      "text": "must",
      "boundingPolygon": ...
    },
    ...
  ]
}]
```

© Copyright Microsoft Corporation. All rights reserved.

## Azure AI Vision OCR – JSON vs. SDK-Objekt

När du använder **Azure AI Vision OCR (READ)** får du analysresultatet antingen som **JSON via REST API** eller som ett **objekt via SDK**.

### 1. JSON (REST API) – Rådata

REST API returnerar ett JSON-objekt som du måste hantera manuellt:

```
{ "lines": [ { "text": "You must be the change", "boundingPolygon": [...], "words": [ { "text": "You", "confidence": 0.996 }, { "text": "must", "confidence": 0.98 } ] } ] }
```

♦ **JSON kräver parsing** för att komma åt texten.

### 2. SDK-Objekt – Enklare hantering

Med SDK får du ett färdigt **objekt** med inbyggda metoder och egenskaper.

#### Python SDK (kortfattad kod)

```
from azure.ai.vision.imageanalysis import ImageAnalysisClient from azure.core.credentials import
AzureKeyCredential client = ImageAnalysisClient("https://your-endpoint/", AzureKeyCredential("your-key")) result
= client.analyze("https://example.com/image.jpg", features=["read"]) for line in result.read.lines: print(line.text) C#
```

#### SDK (kortfattad kod)

```
using Azure.AI.Vision.ImageAnalysis; using Azure; var client = new ImageAnalysisClient(new Uri("https://your-
endpoint/"), new AzureKeyCredential("your-key")); var result = client.Analyze("https://example.com/image.jpg",
VisualFeatures.Read); foreach (var line in result.Read.Lines) Console.WriteLine(line.Text); ♦ SDK-objektet är
direkt användbart utan att behöva tolka JSON manuellt.
```

#### Sammanfattning

**REST API (JSON)** Returnerar rå JSON. Returnerar ett strukturerat objekt. Kräver parsing (json.loads). Inbyggda metoder och egenskaper. Fler kodsteg för att komma åt data. Kortare och mer lättläst kod. **SDK rekommenderas** för enklare hantering, men REST API är flexibla för plattformsoberoende integration.

The Vision Read feature is covering 4.x, and no longer discusses 3.2 and before (for confusion on sync vs async). Document intelligence hasn't yet been mentioned at this point, but reference that it's covered in an upcoming section.

Vision OCR does better for smaller things, handwriting, notes, etc – most unstructured general OCR. Document intelligence has a higher resolution, but operates better for documents and structured text.

## Lab – Read Text in Images

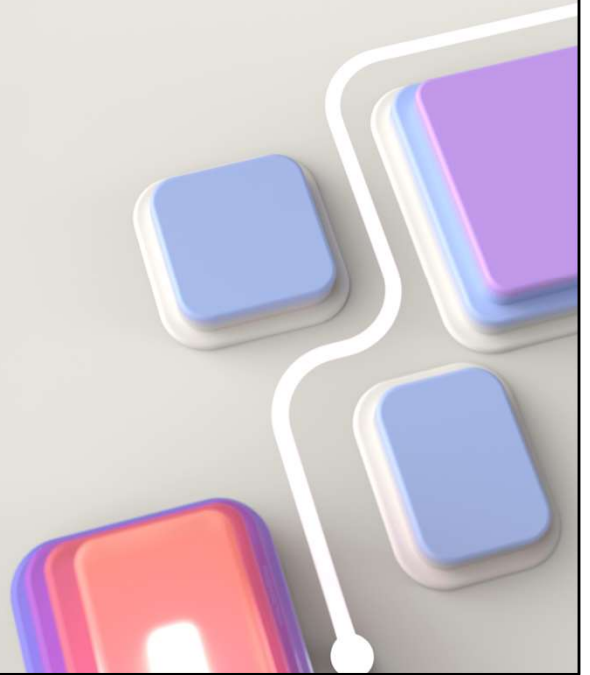


Use the Azure AI Vision for OCR

© Copyright Microsoft Corporation. All rights reserved.

# Detecting Faces with the Azure AI Vision

© Copyright Microsoft Corporation. All rights reserved.



## Learning Objectives

After completing this module, you will be able to:

- 1 Understand features, use cases, and responsibility of the Azure AI Vision Face API
- 2 Use the Face API in an app

© Copyright Microsoft Corporation. All rights reserved.

# Options for Face Detection, Analysis, and Recognition

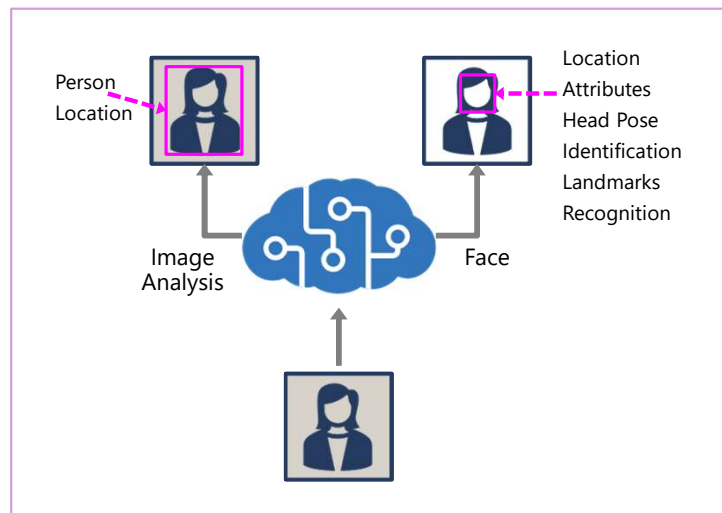
## Image Analysis

- People detection
- Only location provided

## Face Service

- Face detection
- Comprehensive facial feature analysis
- Face comparison and identification\*
- Facial recognition\*

\* Require Limited Access approval



© Copyright Microsoft Corporation. All rights reserved.

Den här bilden jämför **Image Analysis** och **Face Service** i Azure AI för ansiktsdetektering och analys.

### 1. Image Analysis (Grundläggande detektering)

- **People detection** – Identifierar om det finns människor i bilden.
- **Only location provided** – Returnerar endast positionen för personen, **ingen ansiktsanalys**.

### 2. Face Service (Avancerad analys)

- **Face detection** – Identifierar **ansikten** i en bild och deras position.
- **Comprehensive facial feature analysis** – Analyserar detaljer som **ansiktsdrag**, **attribut** och **position**.
- **Face comparison and identification** – Jämför ansikten för att avgöra **om två personer är samma**.
- **Facial recognition\*** – Identifierar personer genom att matcha ansikten mot en databas (*kräver Limited Access-approval*).

### Sammanfattning

- **Image Analysis** är enkel och identifierar **bara var en person finns**.
- **Face Service** ger **mycket mer information** – inklusive ansiktsdrag, jämförelse och igenkänning.

**Limited Access Approval** för **Facial Recognition** i Azure krävs på grund av etiska och juridiska skäl. För att få tillgång:

1. **Ansök via Azure Portal** – Skicka en begäran med affärsanvändning.
2. **Microsoft granskar** – De bedömer **ändamål, sekretess och regelefterlevnad**.
3. **Godkännande krävs** – Endast godkända verksamheter (t.ex. säkerhet, hälsa) får använda tjänsten.
  - ♦ **Utan godkännande** kan du bara använda grundläggande ansiktsdetektering, inte identifiering.

Limited access info: [Microsoft's framework for building AI systems responsibly - Microsoft On the Issues](#) – important to stress this if planning to use the Face service.



# Considerations for Face Detection and Facial Recognition

Principles of responsible AI apply to all kinds of application, but systems that rely on facial data can be particularly problematic. As a safeguard for responsible AI usage, facial recognition, identification, verification, and comparison is behind a Limited Access policy, requiring users to be approved by Microsoft before enabling these features.

## **Data privacy and security**

Systems based on facial data should protect individual privacy, ensuring that personally identifiable data is not accessed inappropriately

## **Transparency**

Users should be informed about how their image will be used, and who will have access to it.

## **Fairness and Inclusiveness**

Facial recognition should not be used in a manner that is prejudicial to individuals based on their appearance, or to unfairly target individuals

© Copyright Microsoft Corporation. All rights reserved.

# The Face Service

Face detection

Face attribute analysis

Facial landmark location

- Nose, eyes, mouth, ...

Face comparison\*

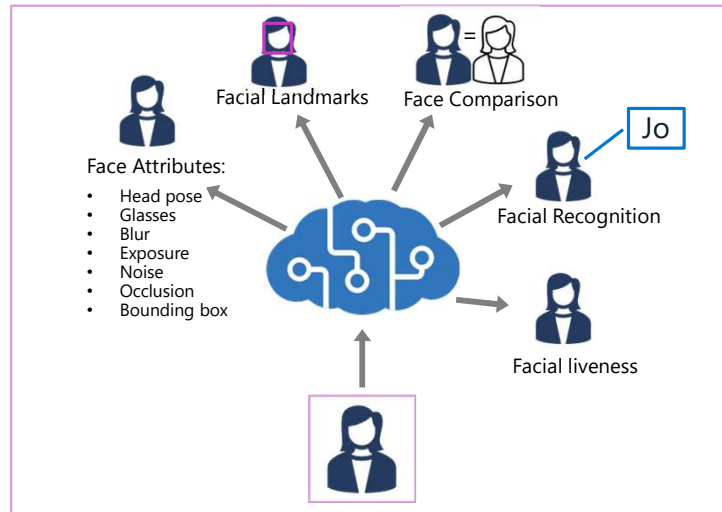
Facial recognition and identification\*

Facial liveness\*

Can be used as:

- Standalone **Face** resource
- Multi-service **Azure AI Services** resource

\* *Require Limited Access approval*



© Copyright Microsoft Corporation. All rights reserved.

Bilden beskriver **Azure Face Service** och dess funktioner för ansiktsanalys, igenkänning och liveness-detektering.

**Funktioner i Face Service**

1. **Face Detection** – Identifierar ansikten i en bild och returnerar deras position.

2. **Face Attribute Analysis** – Analyserar ansiktsattribut, t.ex.:

1. **Head pose** (huvudriktning)
2. **Glasses** (glasögon)
3. **Blur** (oskärpa)
4. **Exposure** (exponering)
5. **Noise** (brus)
6. **Occlusion** (om ansiktet är delvis blockerat)
7. **Bounding box** (ansiktets position i bilden)

3. **Facial Landmark Location** – Identifierar ansiktsdelar som **ögon, näsa och mun**.

4. **Face Comparison\*** – Jämför två ansikten för att avgöra om de tillhör samma person.

5. **Facial Recognition & Identification\*** – Identifierar en person genom att matcha ansiktet mot en databas.

6. **Facial Liveness\*** – Kontrollerar om ansiktet är från en verklig person och inte en bild eller mask.

**Implementering**

- Kan användas som en **fristående Face-resurs**.
- Kan integreras i en **Azure AI Services-miljö** för mer avancerad analys.

**Obs:** Funktioner märkta med \* (Face Comparison, Recognition, Liveness) kräver **Limited Access Approval**.

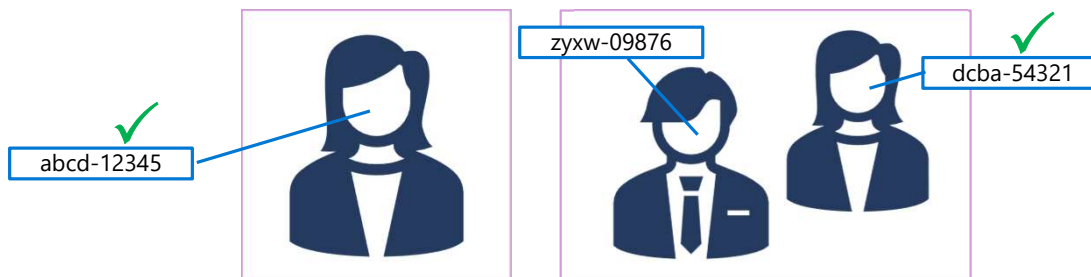
**Sammanfattning**

- ✦ **Face Detection** hittar ansikten, **Face Attributes** analyserar egenskaper.
- ✦ **Face Comparison & Recognition** matchar ansikten för identifiering.
- ✦ **Facial Liveness** förhindrar bedrägerier genom att säkerställa att ansiktet är äkta.

# Detected Face Identification

Every *detected* face is assigned an anonymous ID

- Retained in your service resource for 24 hours
- Can be used to compare faces in multiple images
  - *Verify* faces to determine if they're the same individual
  - *Find similar* faces to identify faces with similar features



© Copyright Microsoft Corporation. All rights reserved.

Den här bilden förklarar hur **Detected Face Identification** fungerar i **Azure Face Service**.

Huvudpunkter:

1. Varje detekterat ansikte får ett anonymt ID

1. När en bild analyseras, tilldelas varje ansikte en **unik identifierare** (t.ex. abcd-12345).
2. ID:t är tillfälligt och lagras bara i **24 timmar** i tjänsten.

2. Jämförelse av ansikten mellan bilder

1. **Verify Faces** – Kontrollera om två ansikten tillhör samma person.
2. **Find Similar** – Identifiera ansikten med liknande drag.

Tillämpning

- ✓ Identifiera samma person i flera bilder
- ✓ Hitta ansikten som liknar ett givet ansikte
- ✓ Används i säkerhet, autentisering och bildsökning

Obs: Identifiering (Face Recognition) kräver **Limited Access Approval**.

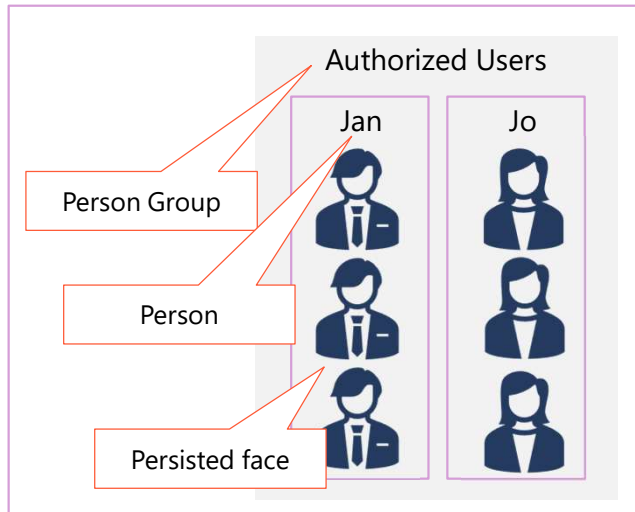
# Persisted Face Recognition

## Train a facial recognition model using face images

1. Create a **Person Group** for the people you want to identify
2. Add a **Person** for each individual
3. Add multiple detected **Faces** to each person
  - These become *persisted* faces
4. Train the model

## Use the model for facial recognition

- *Identify* an individual person
- *Verify* the face of an individual person
- *Find* similar faces to a persisted face



© Copyright Microsoft Corporation. All rights reserved.

Den här bilden förklarar **Persisted Face Recognition** i **Azure Face Service**, vilket innebär att man tränar en modell för ansiktsgenkänning med sparade ansikten.

Hur det fungerar:

### 1. Skapa en Person Group

1. En grupp för personer som ska identifieras (exempelvis "Anställda" eller "Kunder").

### 2. Lägg till en Person för varje individ

1. Varje individ får en unik post i systemet.

### 3. Lägg till flera detekterade ansikten (Persisted Faces) för varje person

1. Dessa bilder används för att förbättra igenkänning.

### 4. Träna modellen

1. Efter att ansiktsdata har laddats upp tränas modellen för att känna igen personerna.

Vad kan modellen göra?

- ✓ **Identify** – Hitta en person i en databas baserat på ett ansikte.
- ✓ **Verify** – Kontrollera om ett ansikte matchar en viss person.
- ✓ **Find similar** – Leta efter ansikten som liknar ett sparat ansikte.

Sammanfattning

- ✦ **Person Group** innehåller flera **Persons**, där varje **Person** har sparade **Persisted Faces**.
- ✦ Modellen används för **identifiering, verifiering och likhetsanalys** av ansikten.
- ✦ Kräver **Limited Access Approval** för ansiktsgenkänning.

# Face Detection with Azure AI Vision

Use the **Face** endpoint, specifying **Faces** as a visual feature  
*Note: Most features like recognition and identification are not enabled for new users*

## Optional request parameters

returnFaceId	recognitionModel
returnFaceLandmarks	returnRecognitionModel
returnFaceAttributes	detectionModel



**Request:** `https://{endpoint}/face/v1.0/detect[?options]`  
**Body:** `{ "url": "http://path-to-image" }`

**Response:**

```
[
  {
    "faceId": "c5c24a82-6845-4031-9d5d-978df9175426",
    "recognitionModel": "recognition_03",
    "faceRectangle": {
      "width": 78,
      "height": 78,
      "left": 394,
      "top": 54
    },
    "faceLandmarks": {
      "pupilLeft": {
        "x": 412.7,
        "y": 78.4
      },
      "pupilRight": {
        "x": 446.8,
        "y": 74.2
      },
      ...
    }
  }
]
```

© Copyright Microsoft Corporation. All rights reserved.

Detail that there are other endpoints, such as Identify, Verify, Group, and Find Similar. Encourage students to go to the docs to explore more about Face, and apply for access. Exercises/examples are difficult without getting that access.

Den här bilden beskriver **Face Detection med Azure AI Vision** och visar hur man använder **Face API** för att identifiera ansikten i en bild.

## Huvudfunktioner:

- Använder **Face Endpoint** för att upptäcka ansikten i en bild genom att skicka en HTTP-begäran.
- **Valfria parametrar** kan läggas till för att få mer detaljerad analys:
  - **returnFaceId** – Returnerar ett unikt ID för varje ansikte.
  - **returnFaceLandmarks** – Returnerar koordinater för ansiktets nyckelpunkter (t.ex. ögon, näsa, mun).
  - **returnFaceAttributes** – Hämtar ytterligare ansiktsattribut (t.ex. ålder, kön, känslor).
  - **recognitionModel** – Anger vilken igenkänningsmodell som används.
  - **returnRecognitionModel** – Returnerar vilken modell som använts.
  - **detectionModel** – Anger vilken detektionsmodell som används.

## Exempel på API-begäran:

En POST-begäran skickas till `/face/v1.0/detect`, där en bild-URL skickas i body:

`{ "url": "http://path-to-image" }` **Exempel på API-svar:**

API:t svarar med en JSON-struktur som innehåller:

- **faceId** – Ett unikt ID för detekterat ansikte.
- **faceRectangle** – Positionen och storleken på ansiktet i bilden.
- **faceLandmarks** – Koordinater för specifika punkter som pupiller, nästipp och mungipor.

Exempel på JSON-svar:

```
{ "faceId": "c5c24a82-6845-4031-9d5d-978df9175426", "recognitionModel": "recognition_03", "faceRectangle": {
  "width": 78, "height": 78, "left": 394, "top": 54 }, "faceLandmarks": { "pupilLeft": { "x": 412.7, "y": 78.4 }, "pupilRight": {
  "x": 446.8, "y": 74.2 } } }
```

**Användningsområden:**

- Ansiktsdetektering** för foton och videoströmmar.
  - Ansiktsanalys** för att extrahera detaljer som huvudriktning och ansiktsuttryck.
  - Förberedelse för ansiktsigenkänning**, där **faceId** används för att jämföra ansikten.
- Azure Face API gör det möjligt att snabbt identifiera och analysera ansikten i en bild, men funktioner som **igenkänning** och **identifiering** är begränsade för nya användare och kräver **Limited Access Approval**.

## Demo – Detect and Analyze Faces



Detect faces with the Azure AI Vision service

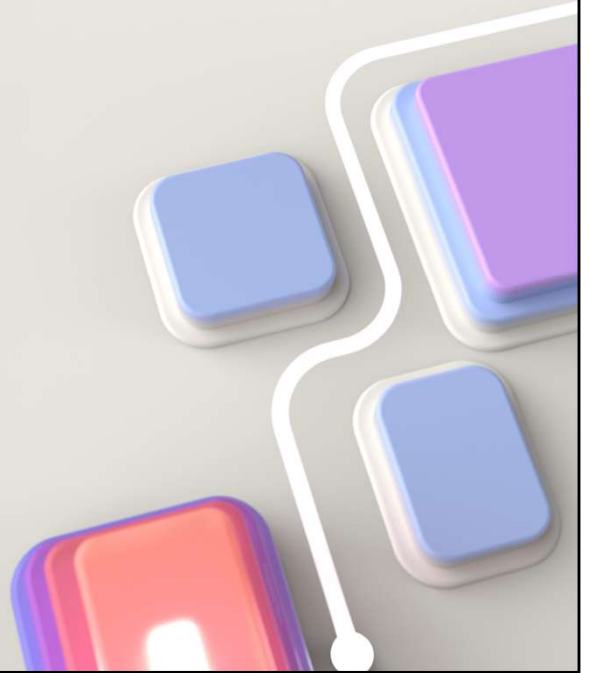
Detect and analyze faces with the Face service

© Copyright Microsoft Corporation. All rights reserved.

<https://microsoftlearning.github.io/mslearn-ai-vision/Instructions/Exercises/04-face-service.html>

# Custom vision models with Azure AI Custom Vision

© Copyright Microsoft Corporation. All rights reserved.





## Learning Objectives

After completing this module, you will be able to:

- 1 Understand use cases of Custom Vision models
- 2 Label data for both image classification and object detection
- 3 Use a custom vision model in an app

© Copyright Microsoft Corporation. All rights reserved.

## Two types of custom vision models

### Azure AI Custom Vision (previous service)

- Portal: **customvision.ai**
- Base model:
  - Convolutional neural network (CNN)
- Tasks:
  - Image classification
  - Object detection
- Labeling:
  - Customvision.ai
- Minimum training data needed:
  - 15 images per category
- Training data storage
  - Uploaded to Custom Vision service

### Custom AI Vision models (new Florence model)

- Portal: **Vision Studio**
- Base model:
  - Transformer (multi modal)
- Tasks:
  - Image classification
  - Object detection
  - Product recognition
- Labeling:
  - AML Studio or COCO file
- Minimum training data needed:
  - 2-5 images per category
- Training data storage
  - In user's blob storage account

© Copyright Microsoft Corporation. All rights reserved.

### Skillnader mellan de två Custom Vision-modellerna:

**Azure AI Custom Vision (gammal tjänst)** **Custom AI Vision (Florence, ny tjänst)** **Portal:** customvision.ai **Portal:** Vision Studio **Modelltyp:** CNN (Convolutional Neural Network) **Modelltyp:** Transformer (multi-modal) **Uppgifter:** Bildklassificering, objektidentifiering **Uppgifter:** Bildklassificering, objektidentifiering, produktigenkänning  
**Märkning:** Customvision.ai **Märkning:** AML Studio eller COCO-fil **Minsta träningsdata:** 15 bilder per kategori **Minsta träningsdata:** 2–5 bilder per kategori **Lagring:** Custom Vision-tjänsten **Lagring:** Användarens blob storage  
 Den nya Florence-modellen är mer flexibel, kräver mindre träningsdata och lagrar data direkt i användarens egen blob storage.

Hiding but leaving here for reference – we're moving back to Custom Vision for these workflows, so there is no "new" service anymore. Much of this is heading the direction of multimodal genAI, but how that looks in the future is still undecided.

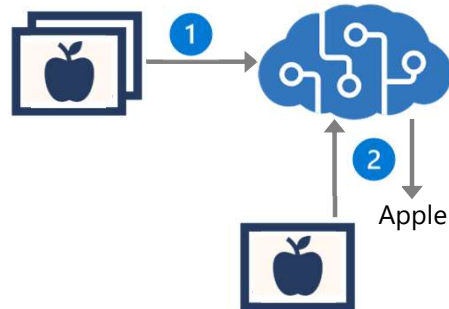
Great unit to pull more information on the differences between base models: [Machine learning for computer vision - Training | Microsoft Learn](#)

Highlight that we'll be focusing on the Custom AI Vision models with Vision Studio. Note that there is a Migration page for moving from Custom Vision to the new custom AI Vision models on the docs page

# Azure AI Custom Vision model

## Train custom models with your own images

- Upload your images
- Label your images
- Train your model
- Query your model with new images to predict labels



- 1 Use your own images to train a model
- 2 Use the model to predict labels for new images

© Copyright Microsoft Corporation. All rights reserved.

# What is Image Classification?

Train a model to predict the class label for the image

In other words, what is this a picture of?



Apple



Banana



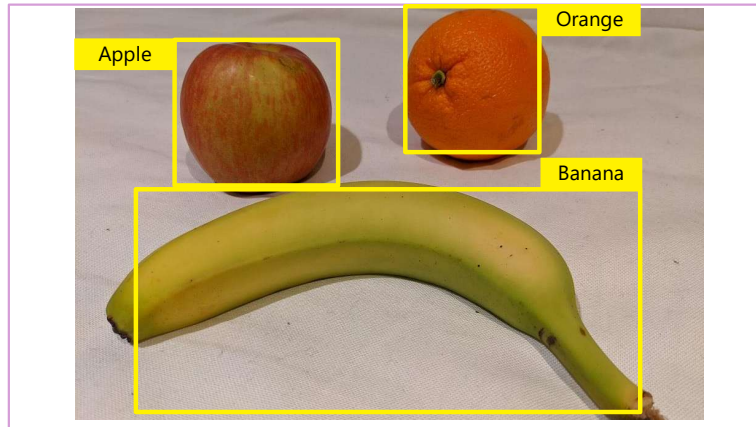
Orange

© Copyright Microsoft Corporation. All rights reserved.

# What is Object Detection?

Train a model to detect and locate specific classes of object in images

In other words, what objects are in this image, and where?

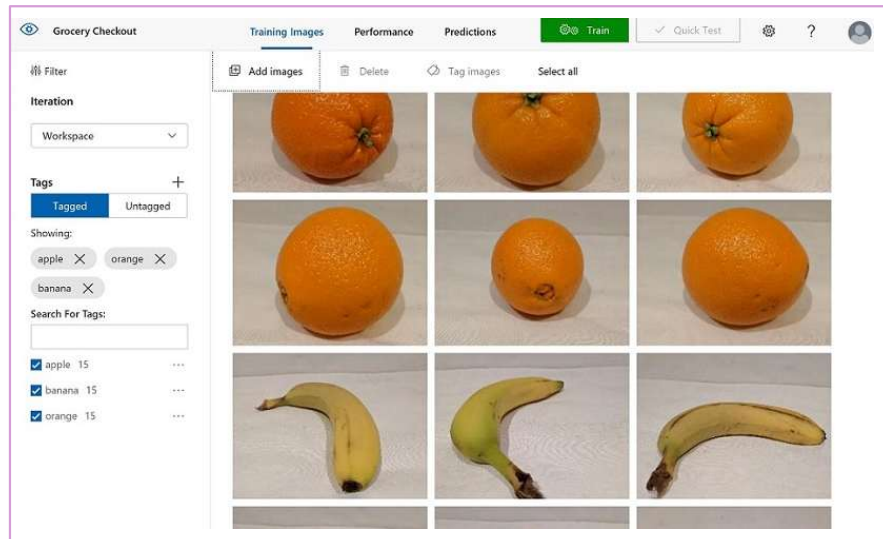


© Copyright Microsoft Corporation. All rights reserved.

# Train a custom model

## Use the Azure Custom Vision Studio

1. Create a custom model project, or retrieve an existing one
2. Select your resource, if necessary
3. Add your dataset(s) and specify model type
  1. Image classification: Assign label to each image
  2. Object detection: Define bounding boxes for objects in each image
4. Label your images
5. Train a new model



© Copyright Microsoft Corporation. All rights reserved.

## Lab – Classify Images with Azure AI Vision custom model



**Create a custom model**

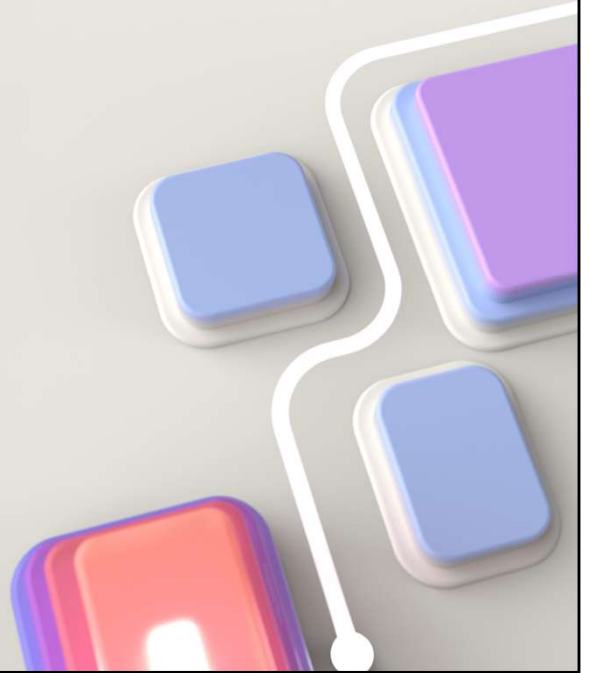
**Integrate custom model into app**

© Copyright Microsoft Corporation. All rights reserved.

<https://microsoftlearning.github.io/mslearn-ai-vision/Instructions/Exercises/07-custom-vision-image-classification.html>

# Analyzing Videos

© Copyright Microsoft Corporation. All rights reserved.

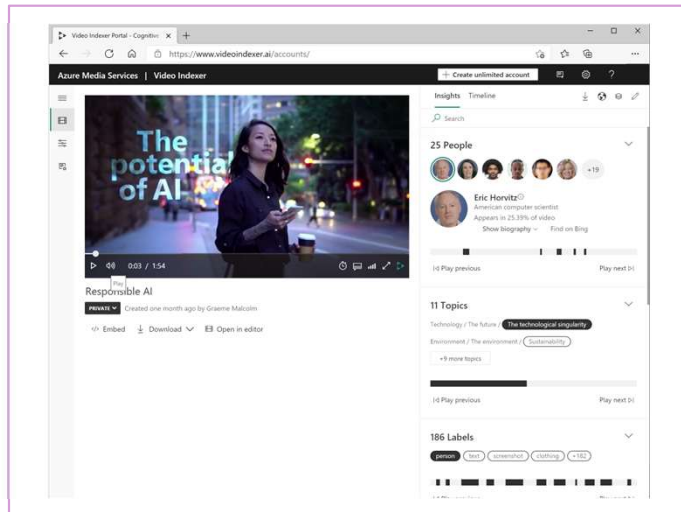




# Video Indexer

## Video analysis:

- Facial recognition (limited access)
- Optical character recognition
- Speech transcription
- Topics
- Sentiment
- Labels
- Content moderation
- Scene segmentation



© Copyright Microsoft Corporation. All rights reserved.

Video Indexer analyserar videor med AI och erbjuder följande funktioner:

- **Facial recognition** (begränsad åtkomst) – Identifierar personer i videon.
- **Optical Character Recognition (OCR)** – Extraherar text från videobilder.
- **Speech transcription** – Omvandlar tal till text.
- **Topics** – Identifierar ämnen som diskuteras.
- **Sentiment analysis** – Avgör känslor i tal och text.
- **Labels** – Märker objekt, personer och scener.
- **Content moderation** – Identifierar olämpligt innehåll.
- **Scene segmentation** – Delar upp videon i logiska sektioner.

Används för medianalys, innehållsmoderering och sökbarhet i videor.

Facial recognition is now behind a limited access policy, which users need to apply to but takes time to get approved. Not available immediately to new customers: [Responsible AI investments and safeguards for facial recognition | Azure Blog and Updates | Microsoft Azure](#)

# Custom Insights

Pre-defined models for recognizing language, well-known celebrities, brands, ...

Create your own models for:



## People

Train facial recognition\* from sample images



## Language

Recognize industry/organization-specific terms



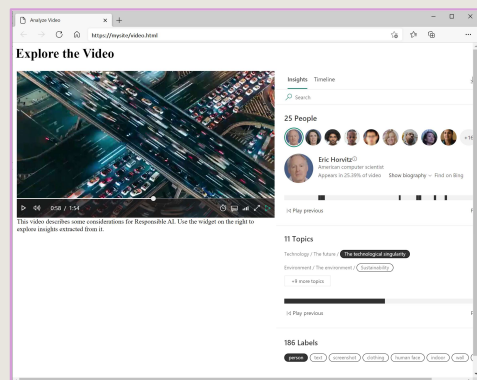
## Brands

Recognize brand names for products, companies, projects,...

\* Facial recognition and celebrity identification requires Limited Access approval, in accordance with our Responsible AI Standard

© Copyright Microsoft Corporation. All rights reserved.

## Video Indexer Widgets and API



Share insights and analysis in web pages

[https://api.videoindexer.ai.../Videos?accessToken=\\$token](https://api.videoindexer.ai.../Videos?accessToken=$token)

```
{
  "results": [
    {
      "accountId": "1234abcd-9876fghi-0156kihb-00123",
      "id": "a12345bc6",
      "name": "Responsible AI",
      "description": "Microsoft Responsible AI video",
      "created": "2021-01-05T15:33:58.918+00:00",
      "lastModified": "2021-01-05T15:50:03.123+00:00",
      "lastIndexed": "2021-01-05T15:34:08.007+00:00",
      "processingProgress": "100%",
      "durationInSeconds": 114,
      "sourceLanguage": "en-US",
    }
  ],
}
```

Automate video analysis with the REST API

Bilden visar **Video Indexer Widgets och API**, som används för att analysera och dela insikter från videor via webbsidor och REST API.

•**Vänster sida:** Exempel på en webbapplikation där videoanalysdata, såsom identifierade personer och ämnen, visas i en interaktiv widget.

•**Höger sida:** En JSON-respons från Video Indexer API som innehåller metadata om en analyserad video, inklusive:

- Video-ID, namn och beskrivning
- Skapad och modifierad tid
- Bearbetningsstatus och varaktighet
- Källspråk

API:t möjliggör **automatisk videotranskription, analys och indexering**, vilket kan integreras i appar och tjänster.

## Demo – Analyze Video



Analyze a video

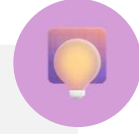
Use Video Indexer widgets

Use the Video Indexer API

© Copyright Microsoft Corporation. All rights reserved.

Instructor demo, can use the exercises here: [mslearn-ai-vision/Instructions/Exercises at main · MicrosoftLearning/mslearn-ai-vision \(github.com\)](https://github.com/microsoft/mslearn-ai-vision/blob/main/Instructions/Exercises%20at%20main%20-%20MicrosoftLearning.md)

# Knowledge check

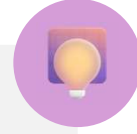


- 1** You want to use the Azure AI Vision Analyze Image function to generate an appropriate caption for an image. Which visual feature should you specify?
  - ☐ Tags
  - ☒ Caption
  - ☐ Text
  
- 2** What is the effect of the *Smart Cropping* option when using Azure AI Vision to generate a thumbnail?
  - ☐ The aspect ratio of the original image is maintained.
  - ☐ The thumbnail is skewed to fit the specified proportions.
  - ☒ The region of interest is centered in the thumbnail.
  
- 3** You want Video Analyzer to recognize colleagues in videos recorded from conference calls. What should you do?
  - ☒ Create a custom model containing a *Person* for each colleague, with example images of their faces.
  - ☐ Edit the conference call videos to include a caption of each person's name on their first appearance.
  - ☐ Embed the Video Analyzer widgets in a custom web site that employees access using their own user credentials.

© Copyright Microsoft Corporation. All rights reserved.

Use the slide animation to reveal the correct answers.

# Knowledge check



- 4** Which of the following facial attributes does the Azure AI Face service not predict?
- ☒ Brands, deepfake
  - ☐ Glasses, mask, facialHair
  - ☐ Accessories, occlusion, headPose
- 5** You need to create a facial recognition solution that can identify named employees. Which service should you use?
- ☐ Vision
  - ☐ Personalizer
  - ☒ Face
- 6** You need to detect if a specific dangerous item is in your company's inspection photos. What should you do?
- ☐ Create an image classification custom model.
  - ☐ Use the Object feature in Image Analysis.
  - ☒ Train a custom model to detect that item.

© Copyright Microsoft Corporation. All rights reserved.

Use the slide animation to reveal the correct answers.

## Learning Path Recap

### In this learning path, we:

Used Image Analysis to analyze images, extract insights, remove background, and perform OCR.

Detected faces and facial recognition.

Created custom vision models trained on your own images.

Extracted insights from videos with Azure Video Indexer.

© Copyright Microsoft Corporation. All rights reserved.



© Copyright Microsoft Corporation. All rights reserved.