

Five visual capabilities

What must happen to the image?

Possible outputs

- Assign a label
- Find and locate objects
- Read visible text
- Interpret visual content
- Create a new image

Key message

The required output determines the capability.

Image classification

Assigns one or more labels to an image.

Input · “Product photograph”

Output

- damaged laptop

Common scenarios

- Identify an animal species
- Categorize a product
- Detect whether an image contains damage
- Classify manufacturing defects

Classification answers: “What is this image?”

Exam signal

Determine which category best describes this image.

Object detection

Identifies objects and returns their locations within an image.

Typical output

- Object label
- Confidence score
- Bounding-box coordinates

Example

- Laptop: [120, 80, 640, 470]
- Crack: [390, 130, 510, 320]

Label the overall image → classification. Find objects and positions → object detection.

Exam signal

Identify where each object appears.

Optical character recognition

OCR converts text visible in an image or scanned document into machine-readable text.

Input · “Photograph of a shipping label”

Output

“Order 10482”

Common scenarios

- Read signs and labels
- Digitize scanned pages
- Capture serial numbers
- Recognize printed or handwritten text

Exam signal

Read or digitize the text shown in an image.

Image understanding

Uses visual context to describe, interpret, or answer questions about an image.

Possible tasks

- Describe an image
- Explain what is happening
- Compare visual elements
- Answer questions about the scene
- Reason across images and text

Example

- Interpretation: The laptop screen appears cracked near the upper-right corner, but the shipping box shows no visible exterior damage.

Exam signal

Describe, interpret, compare, or answer a question about visual content.

Image generation

Creates a new image from a prompt or other input.

Example prompt · “Create a clean product-support illustration of a laptop with a cracked screen.”

The system produces

- A new image
- An edited or extended image
- Visual variations

Common scenarios

- Create marketing artwork
- Produce product concepts
- Generate illustrations
- Edit or extend an existing image
- Create visual variations

Key distinction

Analyze an existing image → computer vision. Produce a new image → image generation.

Same image, different objective

“Input: Photograph of an invoice”

Desired output	Capability
Read all visible words	OCR
Describe the document	Image understanding
Determine whether it is an invoice	Image classification
Return vendor, date, and total as fields	Information extraction

Example structured output

```
{
  "vendor": "Contoso",
  "invoice_date": "2026-09-12",
  "total": 847.50
}
```

Key message

Understanding explains the image. Extraction maps its content into a schema.

What output is required?

The fastest exam decision tree

Classify the output

- Category → Image classification
- Objects plus locations → Object detection
- Visible characters → OCR
- Description or answer → Image understanding
- Structured fields → Information extraction
- New visual → Image generation

Key message

Classify the output—not the fact that an image was provided.

A claims system receives photographs of vehicle damage. It must identify each damaged component and mark its position in the photograph.

Which capability is most appropriate?

A. Image classification

B. Object detection

C. Optical character recognition

D. Image generation

Answer reveal

B. Object detection

The system must identify multiple items and return their locations. Classification would label the image but would not show where each damaged component appears.