

Azure AI Content Understanding documentation

Content Understanding is a solution that analyzes and comprehends various media content—such as audio, video, text, and images—transforming it into structured, organized, and searchable data.

About Azure AI Content Understanding

OVERVIEW

[What is Azure AI Content Understanding?](#)

[Azure AI Content Understanding REST API](#)

[FAQ](#)

Capabilities

CONCEPT

[Audio](#)

[Video](#)

[Image](#)

[Document](#)

Get started

QUICKSTART

[Try Content Understanding REST API](#)

[Try Azure AI Foundry portal](#)

How-to guides



HOW-TO GUIDE

[Create a multi-service resource](#)

What's new?



WHAT'S NEW

[What's new in Azure AI Content Understanding?](#)

Reference



REFERENCE

[Service limits](#)

[Language and region support](#)

What is Azure AI Content Understanding (preview)?

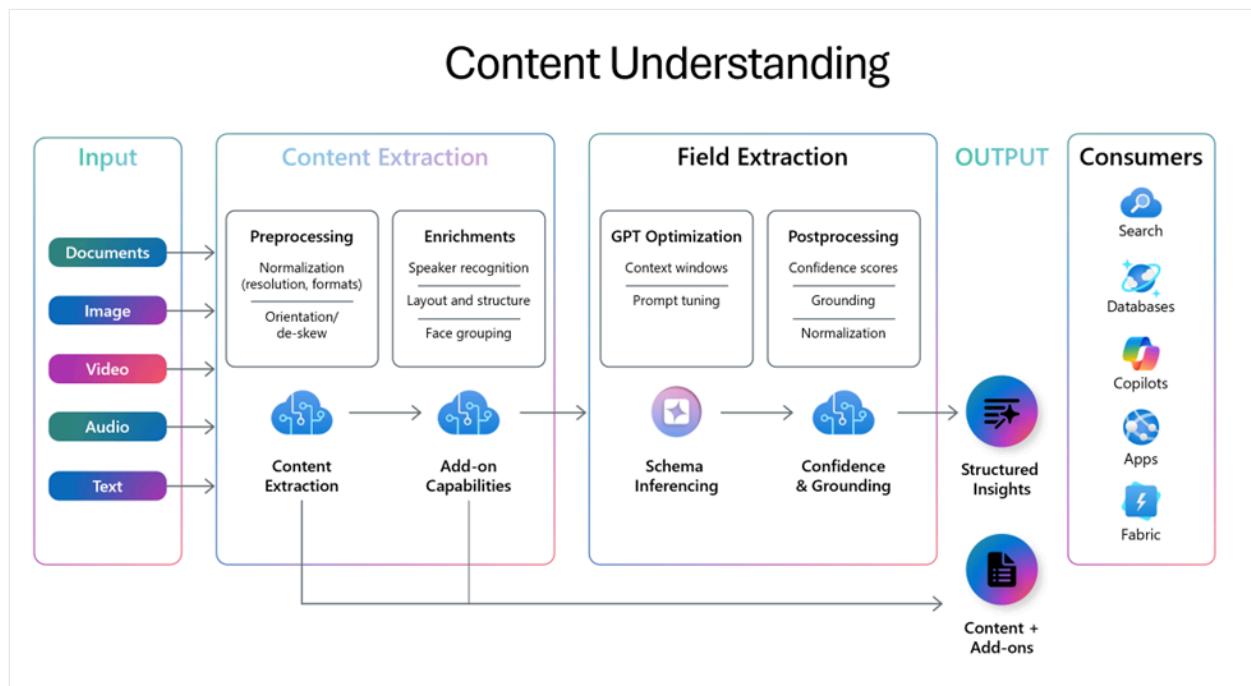
Article • 11/19/2024

ⓘ Important

- Azure AI Content Understanding is available in preview. Public preview releases provide early access to features that are in active development.
- Features, approaches, and processes may change or have constrained capabilities, prior to General Availability (GA).
- For more information, see [Supplemental Terms of Use for Microsoft Azure Previews](#).

Azure AI Content Understanding is a new Generative AI based [Azure AI Service](#), designed to process/ingest content of any types (documents, images, videos, and audio) into a user-defined output format.

Content Understanding offers a streamlined process to reason over large amounts of unstructured data, accelerating time-to-value by generating an output that can be integrated into automation and analytical workflows.



Why process with Content Understanding?

- **Simplify and streamline workflows.** Azure AI Content Understanding standardizes the extraction of content, structure, and insights from various content types into a unified process.
- **Simplify field extraction.** Content Understanding's field extraction makes it easier to generate structured output from unstructured content. Define a schema to extract, classify, or generate field values with no complex prompt engineering
- **Enhance accuracy.** Content Understanding employs multiple AI models to analyze and cross-validate information simultaneously, resulting in more accurate and reliable results.

Content Understanding use cases

- **Automation.** Content Understanding supports automation scenarios by converting unstructured content into structured data, which can be integrated into various workflows and applications. Confidence scores minimize human review and lower costs. For example, automate procurement and payment processes by extracting fields from invoices.
- **Search and retrieval augmented generation (RAG).** Content Understanding enables ingestion of content of any modality into the search index. The structured output representation improves the relevance for RAG scenarios.
- **Analytics and reporting:** Content Understanding's extracted field outputs enhance analytics and reporting, allowing businesses to gain valuable insights, conduct deeper analysis, and make informed decisions based on accurate reports.

Applications

Common applications for Content Understanding include:

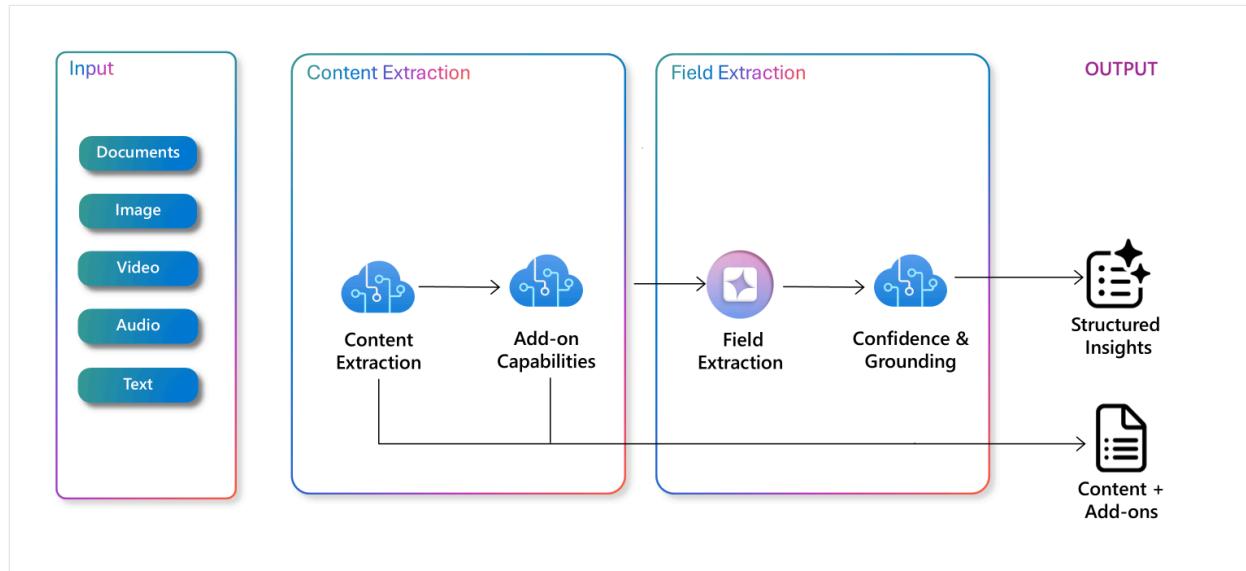
 Expand table

Application	Description	Quickstart
Post-call analytics	Businesses and call centers can generate insights from call recordings to track key KPIs, improve product experience, generate business insights, create differentiated customer experiences, and answer queries faster and more accurately.	Post-call analytics quickstart

Application	Description	Quickstart
Media asset management	Software and media vendors can use Content Understanding to extract richer, targeted information from videos for media asset management solutions.	Media asset management quickstart
Tax automation	Tax preparation companies can use Content Understanding to generate a unified view of information from various documents and create comprehensive tax returns.	Tax automation quickstart
Chart understanding	Businesses can enhance chart understanding by automating the analysis and interpretation of various types of charts and diagrams using Content Understanding.	Chart understanding quickstart

See [Quickstart](#) for more examples.

Components



[Expand table](#)

Component	Description
Analyzer	The analyzer is the core component of Content Understanding. It allows customers to configure content extraction settings and field extraction schema. Once configured, the analyzer consistently applies these settings to process all incoming data.
Content extraction	Content extraction enables users to specify the types of information to be identified and extracted from incoming content. User-specified information includes options such as <code>OCR</code> for text, layout analysis, barcodes, tables, and more, allowing users to focus on the most relevant content elements.

Component	Description
Add-ons	Content Understanding add-ons enhance content extraction by incorporating added elements like barcodes, tables, and detected faces.
Field extraction	<p>Field extraction allows users to define the structure and schema of the desired fields to extract from input files. See service limits for a complete list of field types supported. Fields can be generated via one of the following methods:</p> <ul style="list-style-type: none"> • Extract: Directly extract values as they appear in the input content, such as dates from receipts or item details from invoices. • Classify: Classify content from a predefined set of categories, such as call sentiment or chart type. • Generate: Generate values freely from input data, such as summarizing an audio conversation or creating scene descriptions from videos.
Grounding source	Content Understanding identifies the specific regions in the content where the value was generated from. Source grounding allows users in automation scenarios to quickly verify the correctness of the field values, leading to higher confidence in the extracted data.
Confidence score	Content Understanding provides confidence scores from 0 to 1 to estimate the reliability of the results. High scores indicate accurate data extraction, enabling straight-through processing in automation workflows.

Responsible AI

Azure AI Content Understanding is designed to guard against processing harmful content. For more information, see our [Transparency Note](#) and our [Code of Conduct](#).

Data privacy and security

Developers using the Content Understanding service should review Microsoft's policies on customer data. For more information, visit our [Data, protection and privacy](#) page.

Important

If you are using Microsoft products or services to process Biometric Data, you are responsible for: (i) providing notice to data subjects, including with respect to retention periods and destruction; (ii) obtaining consent from data subjects; and (iii) deleting the Biometric Data, all as appropriate and required under applicable Data Protection Requirements. "Biometric Data" will have the meaning set forth in Article

4 of the GDPR and, if applicable, equivalent terms in other data protection requirements. For related information, see [Data and Privacy for Face](#).

Getting started

Our quickstart guides help you quickly start using the Content Understanding service:

- [Azure AI Foundry portal Quickstart](#)
- [Rest API Quickstart](#)

Feedback

Was this page helpful?



Yes



No

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

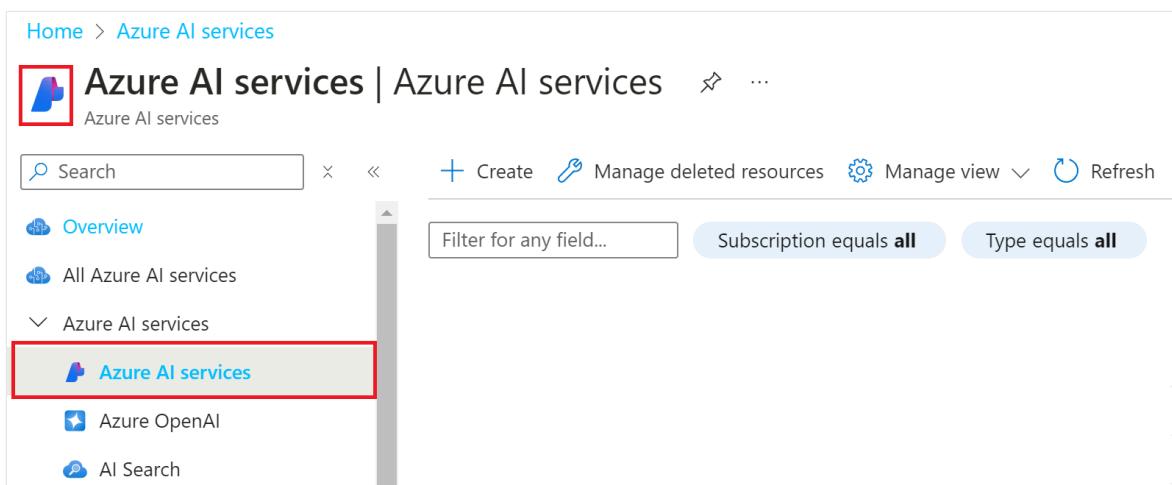
Create an Azure AI services multi-service resource

Article • 11/19/2024

To use Content Understanding, you need an Azure AI Services resource. This multi-service resource enables access to multiple Azure AI services with a single set of credentials.

Prerequisites

1. To get started, you need an active [Azure account](#). If you don't have one, you can [create a free subscription](#).
2. Once you have your Azure subscription, create an [Azure AI services multi-services resource](#) in the Azure portal. The Azure AI services multi-service resource is listed under Azure AI services → Azure AI services in the portal:



The screenshot shows the Azure portal interface. At the top, there's a navigation bar with 'Home' and 'Azure AI services'. Below it is a search bar and a toolbar with 'Create', 'Manage deleted resources', 'Manage view', and 'Refresh' buttons. On the left, there's a sidebar with 'Overview', 'All Azure AI services', and a collapsed section 'Azure AI services' which is expanded to show three items: 'Azure AI services' (highlighted with a red box), 'Azure OpenAI', and 'AI Search'. On the right, there are filters for 'Subscription equals all' and 'Type equals all'.

Important

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

3. Select the **Create** button.

Create a resource

To use the Azure AI Content Understanding service, you must create your Azure AI Service resource in a supported region. The Content Understanding features are available in the following regions:

 Expand table

Geography	Region	Region Identifier
US	West US	westus
Europe	Sweden Central	swedencentral
Australia	Australia East	australiaeast

1. Complete the **Azure AI Services** fields with the following values:

- **Subscription.** Select one of your available Azure subscriptions.
- **Resource group.** The [Azure resource group](#) that contains your resource. You can create a new group or add it to an existing group.
- **Name.** Enter a name for your resource. We recommend using a descriptive name, for example *YourNameAIResourcesResource*.

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

3. Azure will run a quick validation check, after a few seconds you should see a green banner that says **Validation Passed**.

4. Once the validation banner appears, select the **Create** button from the bottom-left corner.

5. After you select create, you'll be redirected to a new page that says **Deployment in progress**. After a few seconds, you'll see a message that says, **Your deployment is complete**.

Feedback

Was this page helpful?

 Yes

 No

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

What's new in Azure AI Content Understanding?

Article • 11/19/2024

The Content Understanding service is continuously updated. Bookmark this page to stay informed about the latest features and samples.

November 2024

Welcome! The Azure AI Content Understanding API version `2024-12-01-preview` is now in public preview. This version allows you to generate a structured representation of content tailored to specific tasks from various modalities or formats. Content Understanding uses a defined schema to extract content suitable for processing by large language models and subsequent applications.

Feedback

Was this page helpful?

 Yes

 No

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

Content Understanding service quotas and limits

Article • 11/19/2024

This article offers a quick reference of the quotas and limits for the Azure AI Content Understanding service.

Resource limits

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

Quota	Standard (S0)
Max analyzers	100k
Max analysis/min	1000 pages/images Four hours of audio Four hours of video
Max operations/min	3000

Input file limits

Document and text

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

Supported File Types	File Size	Length
✓ .pdf ✓ .tiff ✓ .jpg ✓ .png ✓ .bmp ✓ .heif	≤ 200 MB	≤ 300 pages
✓ .txt	≤ 1 MB	≤ 1M characters

Image

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

Supported File Types	File Size	Resolution
✓ .jpg	≤ 200 MB	Min: 50 x 50 pixels
✓ .png		Max: 10k x 10k pixels
✓ .bmp		
✓ .heif		

Audio

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

Supported File Types	File Size	Length
✓ .wav (PCM, A-law, μ-law)	≤ 200 MB	≤ 2 hours
✓ .mp3		
✓ .opus, .ogg (Opus)		
✓ .flac		
✓ .wma		
✓ .aac		
✓ .amr (AMR-NB, AMR-WB)		
✓ .webm (Opus, Vorbis)		
✓ .m4a (AAC, ALAC)		
✓ .spx		

Video

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

Supported File Types	File Size	Resolution	Length
✓ .mp4, .m4v	≤20 GB †	Min: 320 x 240 pixels	≤4 hours †
✓ .flv (H.264 and AAC)		Max: 1920 x 1,080 pixels	
✓ .wmv, .asf			
✓ .avi			
✓ .mkv			
✓ .mov			

Note

The file size limit is 200 MB and the duration limit is 30 minutes if the video file is included directly in the analysis request.

Field schema limits

Content Understanding supports both basic field value types and nested structures, including lists, groups, tables, and fixed tables.

- **Basic field value types:** *string, date, time, number, integer, and boolean.*
- **List field:** A sequence of values of the same type, represented as an array of basic fields in the API.
- **Group field:** A set of semantically related fields, represented as an object of basic fields in the API.
- **Table field:** A variable number of items with fixed subfields, represented as an array of objects of basic fields in the API.
- **Fixed table field:** A group of fields with shared subfields, represented as an object of objects of basic fields in the API.

The following limits apply as of version 2024-12-01-preview.

Basic limits

[+] Expand table

Property	Document	Image	Text	Audio	Video
Max fields	50	10	10	10	10
Max classify field categories	300	300	300	300	300
Supported generation methods	extract	generate classify	generate classify	generate classify	generate classify

- The *Max fields* limit includes all named fields. For example, a list of strings counts as one field, while a group with string and number subfields counts as three fields. To extract beyond default limits, contact us at cu_contact@microsoft.com.
- The *Max classify field categories* limit is the total number of categories across all fields using the `classify` generation method.
- The generation method currently applies only to basic fields.

Field type limits

[+] Expand table

Field type	Document	Image	Text	Audio	Video
Basic	No <code>boolean</code>	No <code>date, time</code>	<code>string</code>	<code>string</code>	No <code>date, time</code>
List	N/A	No <code>date, time</code>	<code>string</code>	<code>string</code>	No <code>date, time</code>
Group	N/A	No <code>date, time</code>	N/A	N/A	No <code>date, time</code>
Table	No <code>boolean</code>	No <code>date, time</code>	<code>string</code>	<code>string</code>	No <code>date, time</code>
Fixed table	No <code>boolean</code>	N/A	N/A	N/A	N/A

Classification fields

Classification fields can be defined to return either a single category (single-label classification) or multiple categories (multi-label classification).

- **Single-label classification:** Defined using a string field with the `classify` method. It can be a top-level basic field or a subfield within a group or table.
- **Multi-label classification:** Represented as a list of string fields with the `classify` method. In the [REST API](#), `method=classify` and `enum` are specified on the inner string field and can only be a top-level field.

Note: Document analyzers currently don't support classification fields.

Training limits

[] Expand table

File type	Max training data
Document	1 GB total 50k pages/images

Feedback

Was this page helpful?

 Yes

 No

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

Content Understanding region and language support

Article • 11/19/2024

Azure AI Content Understanding schemas provide multilingual multimodal processing support. Our language and regional support capabilities enable users to communicate with Content Understanding applications in natural ways and empower global outreach. The following tables list the available region, language, and locale support by modality.

Region support

To use the Azure AI Content Understanding service, you must create your Azure AI Service resource in a supported region. The Content Understanding features are available in the following regions:

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

Geography	Region	Region Identifier
US	West US	westus
Europe	Sweden Central	swedencentral
Australia	Australia East	australiaeast

Language support

Azure AI Content Understanding enables you to process and integrate multiple data modalities simultaneously. Our language support capabilities enable users to communicate with your applications in natural ways and empower global outreach.

To learn more Content Understanding language support, select a capability from the following tabs.

Document

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

Language	Language code and locale
English	<ul style="list-style-type: none">• en-AU (Australia)• en-CA (Canada)• en-GB (United Kingdom)• en-IN (India)• en-US (United States)

Feedback

Was this page helpful?

 Yes

 No

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

Frequently asked questions

FAQ

Find answers to commonly asked questions about Azure AI Content Understanding

What is Content Understanding

Content Understanding is a new Azure AI Service designed to generate structured insights from unstructured content using artificial intelligence. It provides consistent experience to extract content or a structured schema from audio, video, images, documents, or text inputs.

How does Content Understanding work?

Content Understanding utilizes Generative AI models to analyze and interpret various forms of unstructured content. It integrates data from different modalities (for example, text, images, audio) to generate a cohesive and structured output. The service uses machine learning models trained on diverse datasets and generative AI models to ensure high accuracy and relevance in the insights provided.

What types of unstructured content can Content Understanding process?

Content Understanding can process a wide range of unstructured content, including but not limited to:

- Audio recordings
- Video content
- Documents
- Text content
- Images

What are the key benefits of using Content Understanding?

The key benefits of using Content Understanding include:

- Confidence scores: Ensure the accuracy of extracted values while minimizing the cost of human review.
- Defined schema: Define a schema to ensure the extracted values align with intended use.
- Quality improvements over time: The service provides capabilities to improve the quality of the schema extracted.
- Improved decision-making: Structured insights help organizations make informed decisions quickly and effectively.
- Increased efficiency: Automating the analysis of unstructured content saves time and reduces the manual effort required.
- Scalability: The service can handle large volumes of data, making it suitable for organizations of all sizes.

How can businesses use Content Understanding?

Businesses can use Content Understanding in various ways, such as:

- Automation: Automate processing of content to extract a defined schema. Call center, documents, and other similar scenarios.
- Content cataloging: managing a large corpus of digital assets.
- Customer sentiment analysis: Understanding customer feedback from reviews, social media, and support interactions.
- Market research: Analyzing trends and patterns from diverse data sources to inform business strategies.
- Operational insights: Gain insights from internal documents, emails, and other unstructured data to improve operations.

Is Content Understanding easy to integrate with existing systems?

Yes, Content Understanding easily integrates with existing systems and workflows. The service offers a set of easy-to-use APIs that can be integrated into any application.

What security measures are in place to protect data processed by Content

Understanding?

Azure AI Services, including Content Understanding, adheres to strict security and compliance standards to ensure data protection. These measures include data encryption, secure access controls, and compliance with industry regulations such as GDPR and HIPAA. The service also adheres to Microsoft's responsible use of AI.

What base models does Azure AI Content Understanding use?

Content Understanding uses various models and capabilities from Azure OpenAI, Azure AI Speech, Vision, and Language to support single- modality and multi-modal scenarios. The service determines the selection of base models appropriate for each scenario.

What are the pricing tier options for Content Understanding

Content Understanding only supports Standard S0 pricing tier. See more details on the [Content Understanding pricing page](#).

Feedback

Was this page helpful?

 Yes

 No

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

Content understanding terminologies

Article • 11/19/2024

 Expand table

Term	Description
File	Any type of data, including text, documents, images, videos, and audio.
File type	The MIME type of a file, such as text/plain, application/pdf, image/jpeg, audio/wav, and video/mp4. Generic categories like <i>document</i> refer to all corresponding MIME types supported by the service.
Analyzer	A component that processes and extracts content and structured fields from files. Content Understanding offers a few analyzer templates for common scenarios.
Analyzer template	A predefined configuration and field schema for an analyzer. It simplifies creating analyzers by allowing modifications to a template instead of starting from scratch. This feature is available only in AI Foundry, not via REST API/SDKs.
Analyzer result	The output generated by an analyzer after processing input data. It typically includes extracted content in Markdown, extracted fields, and optional modality-specific details.
Add-ons	Added features that enhance content extraction results, such as layout elements, barcodes, and figures in documents.
Fields	List of structured key-value pairs derived from the content, as defined by the field schema. Learn more about supported field value types .
Field schema	A formal description of the fields to extract from the input. It specifies the name, description, value type, generation method, and more for each field.
Generation method	The process of determining the extracted value of a specified field. Content Understanding supports: <ul style="list-style-type: none">• Extract: Directly extract values from the input content, such as dates from receipts or item details from invoices.• Classify: Classify content into predefined categories, such as call sentiment or chart type.• Generate: Generate values from input data, such as summarizing an audio conversation or generating scene descriptions from videos.
Span	A reference indicating the location of an element (for example, field, word) within the extracted Markdown content. A character offset and length represent a span. Different programming languages use various character encodings, which can affect the exact offset and length values for Unicode text. To avoid confusion, spans are only returned if the desired encoding is explicitly specified in the

Term	Description
	request. Some elements can map to multiple spans if they aren't contiguous in the markdown (for example, page).
Grounding source	<p>The specific regions in content where a value was generated. It has different representations depending on the file type:</p> <ul style="list-style-type: none">• Image - A polygon in the image, often an axis-aligned rectangle (bounding box).• PDF/TIFF - A polygon on a specific page, often a quadrilateral.• Audio - A start and end time range.• Video - A start and end time range with an optional polygon in each frame, often a bounding box.
Confidence score	The level of certainty that the extracted data is accurate.

Feedback

Was this page helpful?

 Yes

 No

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

Quickstart: Azure AI Content Understanding REST APIs

Article • 11/19/2024

- Start using the latest preview version of the Azure AI Content Understanding [REST API \(2024-12-01-preview\)](#).
- Azure AI Content Understanding is a new generative AI-based [Azure AI Service](#) that analyzes files of any modality (documents, images, videos, and audio) and extracts structured output in user-defined field formats.
- Integrate the Content Understanding service into your workflows and applications easily by calling our REST APIs.
- This quickstart guides you through using the [Content Understanding REST API](#) to create a custom analyzer and extract content and fields from your input.

Prerequisites

To get started, you need [An active Azure subscription](#). If you don't have an Azure account, you can [create a free subscription](#).

- Once you have your Azure subscription, create an [Azure AI Services resource](#) in the Azure portal. This multi-service resource enables access to multiple Azure AI services with a single set of credentials.
 - This resource is listed under Azure AI services → Azure AI services in the portal.

Important

Azure provides more than one resource type named Azure AI services. Make certain that you select the one listed under Azure AI services → Azure AI services as depicted in the following image. For more information, see [Create an Azure AI Services resource](#).

The screenshot shows the Azure AI services portal. At the top, there's a navigation bar with 'Home > Azure AI services'. Below it is a search bar and several navigation links: '+ Create', 'Manage deleted resources', 'Manage view', and 'Refresh'. A sidebar on the left lists categories: 'Overview', 'All Azure AI services', 'Azure AI services' (which is expanded), 'Azure AI services' (highlighted with a red box), 'Azure OpenAI', and 'AI Search'. There are also filter options: 'Filter for any field...', 'Subscription equals all', and 'Type equals all'.

- In this quickstart, we use the cURL command line tool. If it isn't installed, you can download a version for your dev environment:
 - [Windows](#)
 - [Mac or Linux](#)

Create a custom analyzer

Document

To create a custom analyzer, you need to define a field schema that describes the structured data you want to extract. In the following example, we define a schema for extracting basic information from an invoice document.

First, create a JSON file named `request_body.json` with the following content:

JSON

```
{  
  "description": "Sample invoice analyzer",  
  "scenario": "document",  
  "config": {  
    "returnDetails": true  
  },  
  "fieldSchema": {  
    "fields": {  
      "VendorName": {  
        "type": "string",  
        "method": "extract",  
        "description": "Vendor issuing the invoice"  
      },  
      "Items": {  
        "type": "array",  
        "method": "extract",  
        "items": {  
          "type": "object",  
          "properties": {  
            "name": {  
              "type": "string",  
              "method": "extract",  
              "description": "Item name"  
            },  
            "quantity": {  
              "type": "number",  
              "method": "extract",  
              "description": "Item quantity"  
            },  
            "unitPrice": {  
              "type": "number",  
              "method": "extract",  
              "description": "Item unit price"  
            }  
          }  
        }  
      }  
    }  
  }  
}
```

```
        "Description": {
            "type": "string",
            "method": "extract",
            "description": "Description of the item"
        },
        "Amount": {
            "type": "number",
            "method": "extract",
            "description": "Amount of the item"
        }
    }
}
}
}
}
```

Before running the following `cURL` commands, make the following changes to the HTTP request:

1. Replace `{endpoint}` and `{key}` with the endpoint and key values from your Azure portal Azure AI Services instance.
2. Replace `{analyzerId}` with the name of the new analyzer and create, such as `myInvoice`.

PUT Request

Bash

```
curl -i -X PUT "{endpoint}/contentunderstanding/analyzers/{analyzerId}?api-version=2024-12-01-preview" \
-H "Ocp-Apim-Subscription-Key: {key}" \
-H "Content-Type: application/json" \
-d @request_body.json
```

PUT Response

The 201 (`Created`) response includes an `Operation-Location` header containing a URL that you can use to track the status of this asynchronous creation operation.

```
201 Created
Operation-Location:
```

```
{endpoint}/contentunderstanding/analyzers/{analyzerId}/operations/{operation  
Id}?api-version=2024-12-01-preview
```

Upon completion, performing an HTTP GET on the URL returns "status": "succeeded".

Bash

```
curl -i -X GET "  
{endpoint}/contentunderstanding/analyzers/{analyzerId}/operations/{operation  
Id}?api-version=2024-12-01-preview" \  
-H "Ocp-Apim-Subscription-Key: {key}"
```

Analyze a file

You can analyze files using the custom analyzer you created to extract the fields defined in the schema.

Before running the cURL command, make the following changes to the HTTP request:

Document

1. Replace `{endpoint}` and `{key}` with the endpoint and key values from your Azure portal Azure AI Services instance.
2. Replace `{analyzerId}` with the name of the custom analyzer created earlier.
3. Replace `{fileUrl}` with a publicly accessible URL of the file to analyze, such as a path to an Azure Storage Blob with a shared access signature (SAS) or the sample URL <https://github.com/Azure-Samples/cognitive-services-REST-api-samples/raw/master/curl/form-recognizer/rest-api/invoice.pdf>.

POST request

Bash

```
curl -i -X POST "  
{endpoint}/contentunderstanding/analyzers/{analyzerId}:analyze?api-  
version=2024-12-01-preview" \  
-H "Ocp-Apim-Subscription-Key: {key}" \  
-H "Content-Type: application/json" \  
-d "{\"url\": \"{fileUrl}\")"
```

POST response

The 202 (Accepted) response includes an `Operation-Location` header containing a URL that you can use to track the status of this asynchronous analyze operation.

```
202 Accepted
Operation-Location:
{endpoint}/contentunderstanding/analyzers/{analyzerId}/results/{resultId}?
api-version=2024-12-01-preview
```

Get analyze result

Use the `resultId` from the `Operation-Location` header returned by the previous `POST` response and retrieve the result of the analysis.

1. Replace `{endpoint}` and `{key}` with the endpoint and key values from your Azure portal Azure AI Services instance.
2. Replace `{analyzerId}` with the name of the custom analyzer created earlier.
3. Replace `{resultId}` with the `resultId` returned from the `POST` request.

GET request

Bash

```
curl -i -X GET "
{endpoint}/contentunderstanding/analyzers/{analyzerId}/results/{resultId}?
api-version=2024-12-01-preview" \
-H "Ocp-Apim-Subscription-Key: {key}"
```

GET response

The 200 (`ok`) JSON response includes a `status` field indicating the status of the operation. If the operation isn't complete, the value of `status` is `running` or `notStarted`. In such cases, you should call the API again, either manually or through a script. Wait an interval of one second or more between calls.

Sample response

Document

JSON

```
{  
    "id": "bcf8c7c7-03ab-4204-b22c-2b34203ef5db",  
    "status": "Succeeded",  
    "result": {  
        "analyzerId": "sample_invoice_analyzer",  
        "apiVersion": "2024-12-01-preview",  
        "createdAt": "2024-11-13T07:15:46Z",  
        "warnings": [],  
        "contents": [  
            {  
                "markdown": "CONTOSO LTD.\n\n# INVOICE\n\nContoso  
Headquarters...",  
                "fields": {  
                    "VendorName": {  
                        "type": "string",  
                        "valueString": "CONTOSO LTD.",  
                        "spans": [ { "offset": 0, "length": 12 } ],  
                        "confidence": 0.941,  
                        "source":  
                            "D(1,0.5729,0.6582,2.3353,0.6582,2.3353,0.8957,0.5729,0.8957)"  
                    },  
                    "Items": {  
                        "type": "array",  
                        "valueArray": [  
                            {  
                                "type": "object",  
                                "valueObject": {  
                                    "Description": {  
                                        "type": "string",  
                                        "valueString": "Consulting Services",  
                                        "spans": [ { "offset": 909, "length": 19 } ],  
                                        "confidence": 0.971,  
                                        "source":  
                                            "D(1,2.3264,5.673,3.6413,5.673,3.6413,5.8402,2.3264,5.8402)"  
                                    },  
                                    "Amount": {  
                                        "type": "number",  
                                        "valueNumber": 60,  
                                        "spans": [ { "offset": 995, "length": 6 } ],  
                                        "confidence": 0.989,  
                                        "source":  
                                            "D(1,7.4507,5.6684,7.9245,5.6684,7.9245,5.8323,7.4507,5.8323)"  
                                    }  
                                }  
                            },  
                            ...  
                        ]  
                    }  
                },  
                "kind": "document",  
            ]  
        ]  
    }  
}
```

```
"startPageNumber": 1,
"endPageNumber": 1,
"unit": "inch",
"pages": [
  {
    "pageNumber": 1,
    "angle": -0.0039,
    "width": 8.5,
    "height": 11,
    "spans": [ { "offset": 0, "length": 1650 } ],
    "words": [
      {
        "content": "CONTOSO",
        "span": { "offset": 0, "length": 7 },
        "confidence": 0.997,
        "source":
"D(1,0.5739,0.6582,1.7446,0.6595,1.7434,0.8952,0.5729,0.8915)"
      }, ...
    ],
    "lines": [
      {
        "content": "CONTOSO LTD.",
        "source":
"D(1,0.5734,0.6563,2.335,0.6601,2.3345,0.8933,0.5729,0.8895)",
        "span": { "offset": 0, "length": 12 }
      }, ...
    ]
  }
],
"paragraphs": [
  {
    "content": "CONTOSO LTD.",
    "source":
"D(1,0.5734,0.6563,2.335,0.6601,2.3345,0.8933,0.5729,0.8895)",
    "span": { "offset": 0, "length": 12 }
  }, ...
],
"sections": [
  {
    "span": { "offset": 0, "length": 1649 },
    "elements": [ "/sections/1", "/sections/2" ]
  },
  {
    "span": { "offset": 0, "length": 12 },
    "elements": [ "/paragraphs/0" ]
  }, ...
],
"tables": [
  {
    "rowCount": 2,
    "columnCount": 6,
    "cells": [
      {
        "kind": "columnHeader",
        "RowIndex": 0,
        "ColumnIndex": 0,
        "content": "C1"
      },
      {
        "kind": "columnHeader",
        "RowIndex": 0,
        "ColumnIndex": 1,
        "content": "C2"
      },
      {
        "kind": "columnHeader",
        "RowIndex": 0,
        "ColumnIndex": 2,
        "content": "C3"
      },
      {
        "kind": "columnHeader",
        "RowIndex": 0,
        "ColumnIndex": 3,
        "content": "C4"
      },
      {
        "kind": "columnHeader",
        "RowIndex": 0,
        "ColumnIndex": 4,
        "content": "C5"
      },
      {
        "kind": "columnHeader",
        "RowIndex": 0,
        "ColumnIndex": 5,
        "content": "C6"
      }
    ],
    "rows": [
      {
        "RowIndex": 0,
        "ColumnIndex": 0,
        "content": "R1C1"
      },
      {
        "RowIndex": 0,
        "ColumnIndex": 1,
        "content": "R1C2"
      },
      {
        "RowIndex": 0,
        "ColumnIndex": 2,
        "content": "R1C3"
      },
      {
        "RowIndex": 0,
        "ColumnIndex": 3,
        "content": "R1C4"
      },
      {
        "RowIndex": 0,
        "ColumnIndex": 4,
        "content": "R1C5"
      },
      {
        "RowIndex": 0,
        "ColumnIndex": 5,
        "content": "R1C6"
      }
    ],
    "rowIndex": 0
  }
]
```

```
        "columnIndex": 0,
        "rowSpan": 1,
        "columnSpan": 1,
        "content": "SALESPERSON",
        "source": "D(1,0.5389,4.5514,1.7505,4.5514,1.7505,4.8364,0.5389,4.8364)",
        "span": { "offset": 512, "length": 11 },
        "elements": [ "/paragraphs/19" ]
    },
    ...
],
"source": "D(1,0.4885,4.5543,8.0163,4.5539,8.015,5.1207,0.4879,5.1209)",
"span": { "offset": 495, "length": 228 }
},
...
]
}
}
```

Next steps

- In this quickstart, you learned how to call the [REST API](#) to create a custom analyzer. For a user experience, try [Azure AI Foundry portal](#).

Feedback

Was this page helpful?

 Yes

 No

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

Use Content Understanding in Azure AI Foundry

Article • 12/02/2024

Azure AI Foundry [↗](#) is a comprehensive platform for developing and deploying generative AI applications and APIs responsibly. This guide shows you how to use Content Understanding and build an analyzer, either by creating your own schema from scratch or by using a suggested analyzer template.

Steps to create a Content Understanding analyzer

Azure AI Foundry enables you to build a Content Understanding analyzer tailored to your specific needs. An analyzer can extract data from your content based on your scenario.

Follow these steps to create your own analyzer:

1. Upload a sample data file.
2. Select an analyzer template or build your own schema from scratch.
3. Customize the schema to fit your specific scenario.
4. Test the analyzer on your data and validate its accuracy.
5. Build the analyzer to integrate it into your applications.

Build a schema

To follow is an example of building an analyzer to extract key data from an invoice document:

1. Upload a sample file of an invoice document or any other data relevant to your scenario.

Define schema

Review fields from selected template and customize for your own scenario.

Upload a test file *

Drag and drop one file here or [Browse file](#)

Select one template to start

Upload your data to see the scenario templates that apply to your data type.

Upload a file to get started
Once upload a file, you can preview the prebuilt schemas.
For more information, please see [Content Understanding Documentation](#).

[Create](#)

2. Content Understanding suggests analyzer templates based on your content type. For this example, select **Document analysis** and build your own schema tailored to the invoice scenario. When using your own data, select the analyzer template that best fits your needs, or create your own. See Analyzer templates for a full list of available templates.

3. Select Create.

Define schema

Review fields from selected template and customize for your own scenario.

Upload a test file *

Drag and drop one file here or [Browse file](#)

Document analysis

Analyze documents to extract text, layout, structured fields, and more.

1. Define your schema by specifying the fields you want to extract from the input files. Choose clear and simple [field names](#). Use [field descriptions](#) to provide explanations, exceptions, rules of thumb, and other details to clarify the desired behavior.

2. For each field, indicate the [value type](#) of the desired output. Besides basic types like strings, dates, and numbers, you can define more complex structures such as [tables](#) (repeated items with subfields) and [fixed tables](#) (groups of fields with common subfields).

[Create](#)

4. Add fields to your schema:

- Specify clear and simple field names. Example fields: **vendorName**, **items**, **price**.

- Indicate the value type for each field (strings, dates, numbers, lists, groups). To learn more, see [supported field types](#).
- [Optional] Provide field descriptions to explain the desired behavior, including any exceptions or rules.
- [Optional] Specify the method to generate the value for each field.

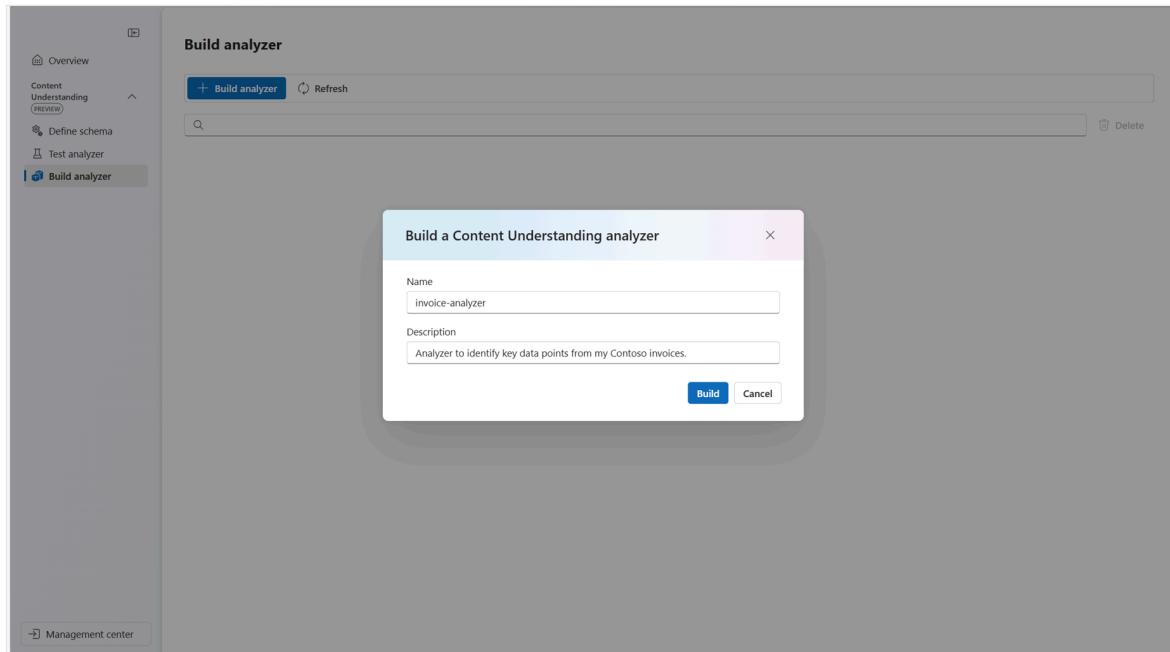
5. Select Save.

The screenshot shows the 'Define schema' page in the Azure AI Foundry interface. On the left, there's a sidebar with 'Overview', 'Content Understanding (PREVIEW)', 'Define schema' (which is selected), 'Test analyzer', and 'Build analyzer'. The main area has a title 'Define schema' and a sub-instruction 'Review fields from selected template and customize for your own scenario.' Below this is a table with columns 'Field name', 'Field description', 'Value type', and 'Method'. The table contains 11 rows of fields. To the right of the table is a preview of an 'INVOICE' document for 'CONTOSO LTD.' with various sections like 'BILL TO', 'SHIP TO', 'SERVICE ADDRESS', and a detailed 'ITEMS' section with a table of items and totals. At the bottom right of the preview is a note 'THANK YOU FOR YOUR BUSINESS!'

6. Content Understanding generates the output based on your schema. Test the analyzer's accuracy on added data or make changes to the schema if needed.

The screenshot shows the 'Test analyzer' page in the Azure AI Foundry interface. The sidebar is identical to the previous screen. The main area has a title 'Test analyzer' and a sub-instruction 'Run analysis' with a progress bar. Below this is a preview of an invoice document for 'CONTOSO LTD.' and a preview of an 'invoice.png' file. To the right is a 'Fields' table with columns 'Fields' and 'Result' showing analysis results for various fields. The table includes rows for 'amountdue', 'customername', 'duedate', 'invoiceID', 'invoicedate', 'previousunpaidbalance', 'salestax', and 'serviceaddress', each with its corresponding value and a confidence score.

7. Once you're satisfied with the quality, select **Build analyzer**. This action creates an analyzer that you can integrate into your applications. You also receive an analyzer ID, which you can use to call the analyzer from your code.



Analyzer templates

Content Understanding analyzer templates give you a head start by allowing you to build your analyzer without creating schemas from scratch. They're fully customizable, allowing you to adjust any fields in the schemas to better fit your needs.

The following analyzer templates are available for use in the [Azure AI Foundry Content Understanding experience](#).

Template	Description
Document analysis	Analyze documents to extract text, layout, structured fields, and more.
Text analysis	Analyze texts and extract structured fields.

Azure AI Foundry / sample-project / Define schema

All hubs + projects Project sample-project

Overview Model benchmarks Content understanding (PREVIEW) Define schema Test analyzer Build analyzer Management center

Define schema

Review fields from selected template and customize for your own scenario.

Upload a test file *

Sample_Image.png

Drag and drop one file here or Browse file

Select one template to start

Defect detection Identify potential defects in provided images of metal plates.

Document analysis Analyze documents to extract text, layout, structured fields, and more.

Image analysis Analyze images to extract structured fields.

Create

Next steps

- In this quickstart, you learned how to create an analyzer in Azure AI Foundry. To use [REST API](#), see the [REST API quickstart](#).

Feedback

Was this page helpful?

Yes

No

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

Content Understanding document solutions (preview)

Article • 01/15/2025

ⓘ Important

- Azure AI Content Understanding is available in preview. Public preview releases provide early access to features that are in active development.
- Features, approaches, and processes can change or have limited capabilities, before General Availability (GA).
- For more information, see [Supplemental Terms of Use for Microsoft Azure Previews](#).

Content Understanding is a cloud-based [Azure AI Service](#) designed to efficiently extract content and structured fields from documents and forms. It provides a comprehensive suite of APIs and an intuitive UX experience for optimal efficiency.

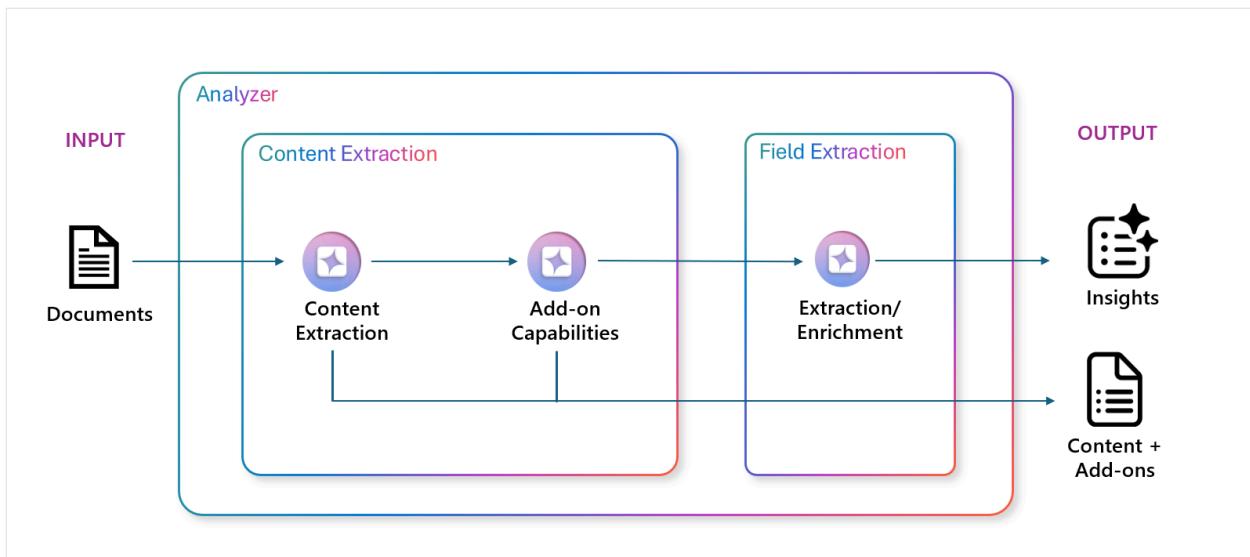
Content Understanding enables organization to streamline data collection and processing, enhance operational efficiency, optimize data-driven decision making, and empower innovation. With customizable analyzers, Content Understanding allows for easy extraction of content or fields from documents and forms, tailored to specific business needs.

Business use cases

Document analyzers can process complex documents in various formats and templates:

- **Contract lifecycle management:** Extract key fields, clauses, and obligations from various contract types.
- **Loan and mortgage applications:** Automate processing to enable quicker handling by banks, lenders, and government entities.
- **Financial services:** Analyze complex documents like financial reports and asset management reports.
- **Expense management:** Parse receipts and invoices from various retailers to validate expenses across different formats and templates.

Document analyzer capabilities



Content extraction enables the extraction of both printed and handwritten text from forms and documents, delivering business-ready content that is immediately actionable, usable, or adaptable for further development within your organization.

Add-on capabilities

Enhance your document extraction with optional add-on features, which can incur added costs. These features can be enabled or disabled based on your needs. Currently supported add-ons include:

- **Layout:** Extracts layout information such as paragraphs, sections, tables, and more.
- **Barcode:** Identifies and decodes all barcodes in the documents.
- **Formula:** Recognizes all identified mathematical equations from the documents.

Field extraction

Field extraction enables the extraction of structured data from various forms and documents tailored to your specific needs. For instance, you can extract customer names, billing addresses, and line items from invoices; or parties, renewal date, and payment clause from contracts. You can start field extraction right after defining the schema or enhance it by labeling more sample documents to improve extraction quality.

Key Benefits

- **Accuracy and reliability:** Ensure precise data extraction, reducing errors and boosting efficiency.
- **Scalability:** Seamlessly scale out document processing to meet business demands.
- **Customizable:** Adapt document analyzer to fit specific workflows.
- **Grounding source:** Localize extracted data for human review workflows.

- **Confidence scores:** Enhance automation with estimated confidence scores to maximize efficiency and minimize costs.

Input requirements

For detailed information on supported input document formats, refer to our [Service quotas and limits](#) page.

Supported languages and regions

For a detailed list of supported languages and regions, visit our [Language and region support](#) page.

Data privacy and security

Developers using Content Understanding should review Microsoft's policies on customer data. For more information, visit our [Data, protection, and privacy](#) page.

Next step

- Try processing your document content using Content Understanding in [Azure](#).
- Learn to analyze document content [analyzer templates](#).
- Review code samples: [visual document search](#).
- Review code sample: [analyzer templates](#).

Feedback

Was this page helpful?



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

Content Understanding image solutions (preview)

Article • 01/15/2025

ⓘ Important

- Azure AI Content Understanding is available in preview. Public preview releases provide early access to features that are in active development.
- Features, approaches, and processes can change or have limited capabilities, before General Availability (GA).
- For more information, see [Supplemental Terms of Use for Microsoft Azure Previews](#).

Azure AI Content Understanding standardizes the extraction of data from images, making it easier to analyze large volumes of unstructured data. Standardized extraction speeds up time-to-value and simplifies integration into downstream analytical workflows. With the Content Understanding APIs, you can define schema to specify the fields, descriptions, and output types for extraction. The service then analyses the images and provides structured data, which can be applied in various use cases, such as:

- **Retrieval-augmented generation (RAG) applications:** Extract key details from your images to build a robust index that powers user-facing chat experiences. This index enables users to ask questions and receive accurate answers based on the content of your images.
- **Financial analysis and business intelligence:** Analyze business performance charts and trends to generate real-time reports that help analysts, managers, and executives make faster, more informed decisions.
- **Manufacturing quality control:** Automate the detection of defects and anomalies, such as scratches, cracks, or misalignments, in production lines and manufacturing environments.
- **Shelf analysis and inventory management:** Detect, count, and extract specific details about retail products, optimizing operations, and improving customer satisfaction by ensuring products are well-stocked and properly organized.

Key Benefits

Content Understanding offers several key benefits for extracting information from images, including,

- **Enhanced data usability and structure:** By providing structured data, Content Understanding simplifies integration with databases, spreadsheets, and systems like Customer Relationship Management (CRM) or Enterprise Resource Planning (ERP) tools.
- **Improved accuracy for specific use cases:** Content Understanding enables targeted data extraction that aligns directly with your unique requirements, helping to improve model accuracy by focusing on the most important data points.
- **Faster and more cost-effective automation:** Extracting only the necessary fields enables Content Understanding to streamline automation. Thus allowing organizations to scale their data processing workflows efficiently and reduce the storage and processing of irrelevant data.

Input requirements

For detailed information on supported input file formats, refer to our [Service quotas and limits](#) page.

ⓘ Note

For best results, image schema should only be used to process non-document-based images. Text heavy images of documents should be processed using a document schema. Use cases that require extraction of text from document images or scanned documents should be processed using a document field extraction schema.

Supported languages and regions

For a detailed list of supported languages and regions, visit our [Language and region support](#) page.

Supported field types

For detailed information on supported field types, refer to our [Service quotas and limits](#) page.

Data privacy and security

As with all the Azure AI services, developers using the Content Understanding service should be aware of Microsoft's policies on customer data. See our [Data, protection and privacy](#) page to learn more.

ⓘ Important

If you're using Microsoft products or services to process Biometric Data, you're responsible for: (i) providing notice to data subjects, including with respect to retention periods and destruction; (ii) obtaining consent from data subjects; and (iii) deleting the Biometric Data, all as appropriate, and required under applicable Data Protection Requirements. "Biometric Data" has the meaning articulated in Article 4 of the GDPR and, if applicable, equivalent terms in other data protection requirements. For related information, see [Data and Privacy for Face](#).

Next steps

- Try processing your video content using Content Understanding in [Azure AI Foundry portal](#).
- Learn to analyze video content [analyzer templates](#).
- Review code samples: [image, text, and table, content extraction](#).
- Review code sample: [analyzer templates](#).

Feedback

Was this page helpful?

 Yes

 No

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

Content Understanding audio solutions (preview)

Article • 01/15/2025

ⓘ Important

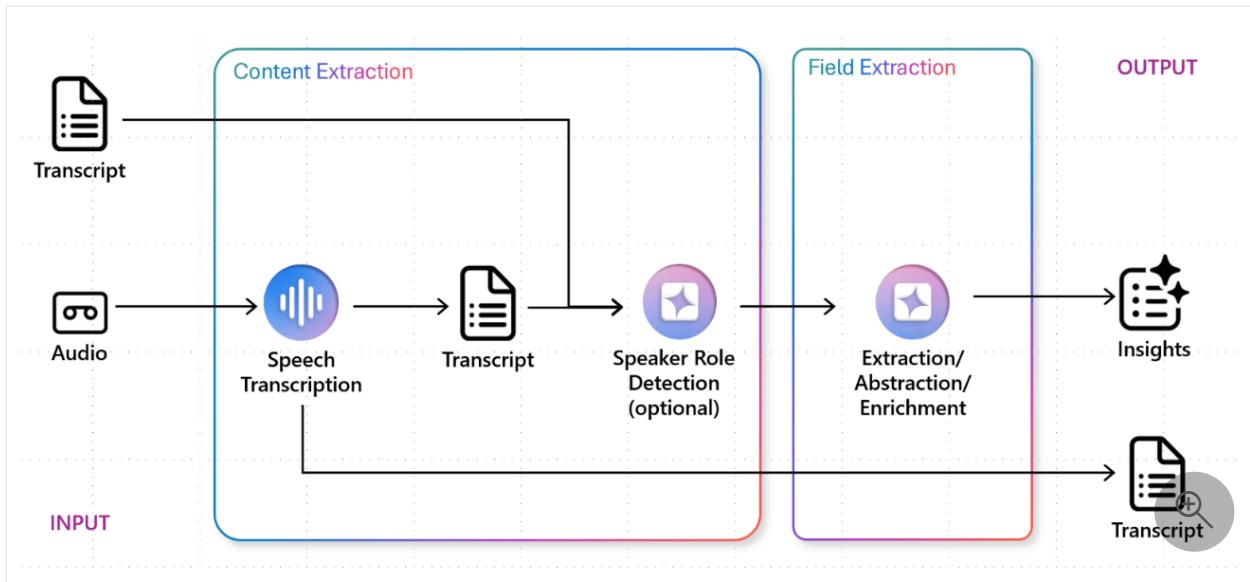
- Azure AI Content Understanding is available in preview. Public preview releases provide early access to features that are in active development.
- Features, approaches, and processes can change or have limited capabilities, before General Availability (GA).
- For more information, see [Supplemental Terms of Use for Microsoft Azure Previews](#).

Content Understanding audio analyzers enable transcription and diarization of conversational audio, extracting structured fields such as summaries, sentiments, and key topics. Customize an audio analyzer template to your business needs using [Azure AI Foundry portal](#) to start generating results.

Here are common scenarios for using Content Understanding with conversational audio data:

- Gain customer insights through summarization and sentiment analysis.
- Assess and verify call quality and compliance in call centers.
- Create automated summaries and metadata for podcast publishing.

Audio analyzer capabilities



Content Understanding serves as a cornerstone for Media Asset Management solutions, enabling the following capabilities for audio files:

Content extraction

- **Transcription.** Converts conversational audio into searchable and analyzable text-based transcripts in WebVTT format. Customizable fields can be generated from transcription data. Sentence-level and word-level timestamps are available upon request.
- **Diarization.** Distinguishes between speakers in a conversation, attributing parts of the transcript to specific speakers.
- **Speaker role detection.** Identifies agent and customer roles within contact center call data.
- **Language detection.** Automatically detects the language in the audio or uses specified language/locale hints.

Field extraction

Field extraction allows you to extract structured data from audio files, such as summaries, sentiments, and mentioned entities from call logs. You can begin by customizing a suggested analyzer template or creating one from scratch.

Key Benefits

Content Understanding offers advanced audio capabilities, including:

- **Customizable data extraction.** Tailor the output to your specific needs by modifying the field schema, allowing for precise data generation and extraction.
- **Generative models.** Utilize generative AI models to specify in natural language the content you want to extract, and the service generates the desired output.
- **Integrated pre-processing.** Benefit from built-in preprocessing steps like transcription, diarization, and role detection, providing rich context for generative models.
- **Scenario adaptability.** Adapt the service to your requirements by generating custom fields and extract relevant data.

Content Understanding audio analyzer templates

Content Understanding offers customizable audio analyzer templates:

- **Post-call analysis.** Analyze call recordings to generate conversation transcripts, call summaries, sentiment assessments, and more.
- **Conversation analysis.** Generate transcriptions, summaries, and sentiment assessments from conversation audio recordings.

Start with a template or create a custom analyzer to meet your specific business needs.

Input requirements

For a detailed list of supported audio formats, refer to our [Service limits and codecs](#) page.

Supported languages and regions

For a complete list of supported regions, languages, and locales, see our [Language and region support](#) page.

Data privacy and security

Developers using Content Understanding should review Microsoft's policies on customer data. For more information, visit our [Data, protection, and privacy](#) page.

Next steps

- Try processing your audio content using Content Understanding in [Azure AI Foundry portal](#).
 - Learn how to analyze audio content [analyzer templates](#).
 - Review code sample: [audio content extraction](#).
 - Review code sample: [analyzer templates](#).
-

Feedback

Was this page helpful?

 Yes

 No

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

Azure AI Content Understanding video solutions (preview)

Article • 01/15/2025

ⓘ Important

- Azure AI Content Understanding is available in preview. Public preview releases provide early access to features that are in active development.
- Features, approaches, and processes can change or have limited capabilities, before General Availability (GA).
- For more information, see [Supplemental Terms of Use for Microsoft Azure Previews](#).

Azure AI Content Understanding allows you to extract and customize video metadata. Content Understanding helps efficiently manage, categorize, retrieve, and build workflows for video assets. It enhances your media asset library, supports workflows such as highlight generation, categorizes content, and facilitates applications like retrieval-augmented generation (RAG).

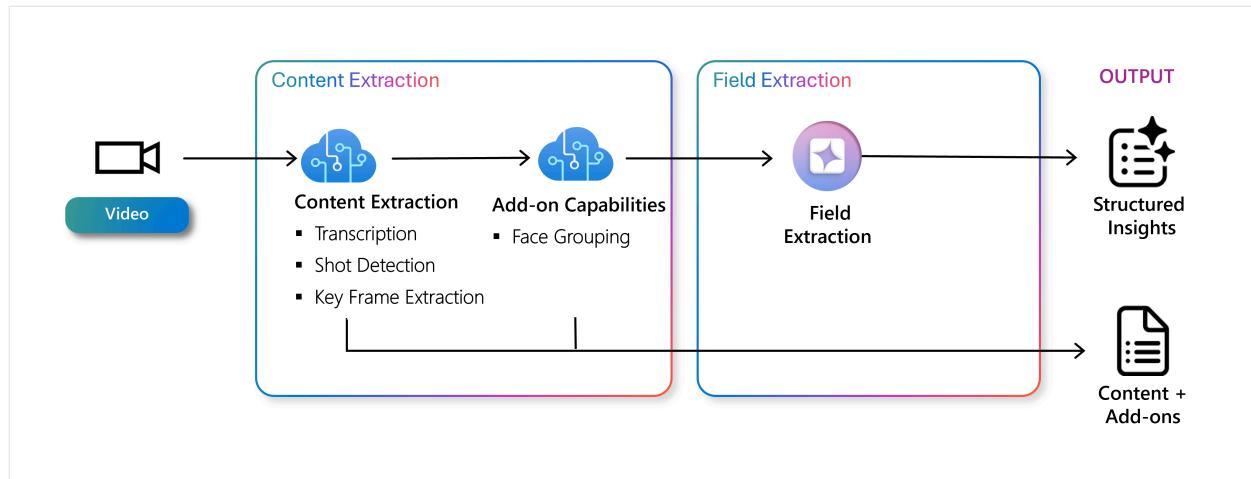
Content understanding for video has broad potential uses. For example, you can customize metadata to tag specific scenes in a training video, making it easier for employees to locate and revisit important sections. You can also use metadata customization to identify product placement in promotional videos, which helps marketing teams analyze brand exposure.

Business use cases

Azure AI Content Understanding provides a range of business use cases, including:

- **Broadcast media and entertainment:** Manage large libraries of shows, movies, and clips by generating detailed metadata for each asset.
- **Education and e*Learning:** Index and retrieve specific moments in educational videos or lectures.
- **Corporate training:** Organize training videos by key topics, scenes, or important moments.
- **Marketing and advertising:** Analyze promotional videos to extract product placements, brand appearances, and key messages.

Video understanding capabilities



Content Understanding processes video files through a customizable pipeline that can perform both **content extraction** and **field extraction** tasks. Content Extraction focuses on analyzing the video to generate foundational metadata, while Field Extraction uses that metadata to create more detailed, custom insights tailored to specific use cases. To follow is an overview of each capability.

Content extraction

Content extraction for video includes transcription, shot detection, key frame extraction, and face grouping. These operations are performed over sampled frames from the entire video and generate a structured text output representing the video. Content extraction also serves as grounding data for generative capabilities of Field Extraction by providing context on what is contained in the video.

Specific capabilities of content extraction:

- **Transcription:** Converts speech to structured, searchable text via Azure AI Speech, allowing users to specify recognition languages.
- **Shot detection:** Identifies segments of the video aligned with shot boundaries where possible, allowing for precise editing and repackaging of content with breaks exactly on shot boundaries.
- **Key frame extraction:** Extracts key frames from videos to represent each shot completely, ensuring each shot has enough key frames to enable Field Extraction to work effectively.
- **Face grouping:** Groups faces appearing in a video to extract one representative face image for each person and provides segments where each one is present. The grouped face data is available as metadata and can be used to generate customized metadata fields.

- This feature is limited access and involves face identification and grouping; customers need to register for access at [Face Recognition](#).

Field extraction

Field extraction enables the generation of structured data for each segment of the video, such as tags, categories, or descriptions, using a customizable schema tailored to your specific needs. This structured data makes it easier to organize, search, and automatically process video content efficiently. Field extraction uses a multimodal generative model to extract specific data from the video, using key frames and text output from Content Extraction as input. Field extraction enables the generative model to make detailed insights based on the visual content captured from shots, providing detailed identification.

Examples of fields for different industries:

- **Media asset management:**
 - **Shot type:** Helps editors and producers organize content, simplifying editing, and understanding the visual language of the video. Useful for metadata tagging and quicker scene retrieval.
 - **Color scheme:** Conveys mood and atmosphere, essential for narrative consistency and viewer engagement. Identifying color themes helps in finding matching clips for accelerated video editing.
- **Advertising:**
 - **Brand:** Identifies brand presence, critical for analyzing ad impact, brand visibility, and association with products. This capability allows advertisers to assess brand prominence and ensure compliance with branding guidelines.
 - **Ad categories:** Categorizes ad types by industry, product type, or audience segment, which supports targeted advertising strategies, categorization, and performance analysis.

Key benefits

Content Understanding provides several key benefits when compared to other video analysis solutions:

- **Segment-based multi-frame analysis:** Identify actions, events, topics, and themes by analyzing multiple frames from each video segment, rather than individual frames.
- **Customization:** Customize the metadata you generate by modifying the schema in accordance with your specific use case.

- **Generative models:** Describe in natural language what content you want to extract, and Content Understanding uses generative models to extract that metadata.
- **Optimized preprocessing:** Perform several content extraction preprocessing steps, such as transcription and scene detection, optimized to provide rich context to AI generative models.

Input requirements

For detailed information on supported input document formats, refer to our [Service quotas and limits](#) page.

Supported languages and regions

For a detailed list of supported languages and regions, visit our [Language and region support](#) page.

Data privacy and security

As with all the Azure AI services, developers using the Content Understanding service should be aware of Microsoft's policies on customer data. See our [Data, protection and privacy](#) page to learn more.

Important

Users of Content Understanding can enable features like Face Grouping for videos, which involved processing Biometric Data. If you're using Microsoft products or services to process Biometric Data, you're responsible for: (i) providing notice to data subjects, including with respect to retention periods and destruction; (ii) obtaining consent from data subjects; and (iii) deleting the Biometric Data, all as appropriate, and required under applicable Data Protection Requirements. "Biometric Data" has the meaning articulated in Article 4 of the GDPR and, if applicable, equivalent terms in other data protection requirements. For related information, see [Data and Privacy for Face](#).

Next steps

- Try processing your video content using Content Understanding in [Azure portal](#).
- Learn to analyze video content [analyzer templates](#).

- Review code sample: [video content extraction](#).
 - Review code sample: [video search with natural language queries](#).
 - Review code sample: [analyzer templates](#)
-

Feedback

Was this page helpful?



Yes



No

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

Transparency note: Content Understanding

Article • 11/19/2024

What is a Transparency Note?

An AI system includes not only the technology, but also the people who will use it, the people who will be affected by it, and the environment in which it is deployed. Creating a system that is fit for its intended purpose requires an understanding of how the technology works, what its capabilities and limitations are, and how to achieve the best performance. Microsoft's Transparency Notes are intended to help you understand how our AI technology works, the choices system owners can make that influence system performance and behavior, and the importance of thinking about the whole system, including the technology, the people, and the environment. You can use Transparency Notes when developing or deploying your own system or share them with the people who will use or be affected by your system.

Microsoft's Transparency Notes are part of a broader effort at Microsoft to put our AI Principles into practice. To find out more, see the [Microsoft AI principles](#).

The basics of Azure AI Content Understanding

Introduction

Azure AI Content Understanding ingests unstructured content in any modality such as documents, images, videos, and audio to produce structured outputs from prebuilt or user-defined schemas to best represent task specific scenarios from the content. This output can then be consumed by downstream applications, such as saving it in a database, sending the output to a customer-developed system for reasoning with LLMs (i.e, Retrieval Augmented Generation or RAG), building specific AI/ML models on the data, or used in workflows to automate business processes. Content Understanding will expand the scope of [Azure AI Document Intelligence](#) and leverage capabilities from [Azure Open AI Service](#), [Azure AI Speech](#), and [Azure AI Vision](#) to support single-modal and multimodal scenarios.

Key terms

Term	Definition
Classify	This is a type of field kind. The field will classify a value from the input data using the field name. An example would be classifying whether the image has defect, or a face has glasses on or not.
Confidence value	All Content Understanding output returns confidence values in the range between 0 and 1 for all extracted words and key-value mappings. This value represents the estimate percentage of how many times it correctly extracts the word out of 100 or correctly maps the key-value pairs. For example, a word that's estimated to be extracted correctly 82% of the time results in a confidence value of 0.82.
Diarization	Diarization distinguishes between individual speakers in each audio recording by assigning a temporary, anonymous label to each speaker (e.g., GUEST1, GUEST2, GUEST3, etc.) to denote which speaker is speaking in the audio file.
	All Content Understanding APIs that support transcription also support diarization.
Extract	This is a type of field kind. The field will directly extract a value from input data. An example would be extracting dates from invoices or signatures from documents.
Face detection	Finds human faces in an image and returns bounding boxes indicating where the faces are. Face detection models alone do not find individually identifying features, only a bounding box marking the entire face. For all the faces detected, Face ID is assigned based on embeddings. Please refer to the Face detection concept documentation for more information.
Face grouping	After the faces are detected, the identified faces are filtered into local groups. If a person is detected more than once, more observed face instances are created for this person. Please refer to [Face grouping documentation](/azure/ai-services/computer-vision/overview-identity"\ "group-faces) for more information.
Generate	This is a type of field kind. The field will generate a value from parent field content. An example would be generating scene description from videos or summarizing from a call audio.
Schema	Schema is the term we use for field names and descriptions that customers need to provide for us to extract values from the input. Content Understanding provides a set of prebuilt schemas to fit your scenarios. Depending on the scenario, Content Understanding has a pre-defined list of fields that will be filled out based on the input. You can use these prebuilt schemas to get started on your project faster without having to define the fields yourself.
Transcription	Content Understanding's automated speech-to-text output feature, sometimes called machine transcription or automated speech recognition (ASR). Transcription

Term	Definition
	uses Azure AI Speech and is fully automated. All Content Understanding APIs that support transcription also support diarization.

Capabilities

System behavior

Azure AI Content Understanding is a cloud-based Azure AI service that uses a variety of AI/ML models (such as those available through Azure OpenAI Service, Azure Face Service, and Azure Speech) to extract, classify and generate fields from a customer's input file. Content Understanding does not support integrating any models that customers bring in.

Content Understanding first extracts the content into a structured output. It then uses a large language model (LLM) to generate fields and assign confidence scores to applicable fields.

Currently, Content Understanding can ingest data of the following types: document, image, text, video, and audio. Depending on the type of data the user uploads, Content Understanding will automatically suggest common prebuilt schemas users can get started with. Users also have the choice to customize the schema themselves, allowing for a fuller data ingestion capability. In the case where user uploads harmful content, Content Understanding will issue a warning in the output to let the users know that the input file contains harmful content, but it will still output the fields.

The goal of the service is to provide a normalized, task specific representation of the input data to enable extractive and generative scenarios for customers while providing a consistent experience across modalities. Note that Content Understanding is not intended to support ungrounded inferencing, and it will only generate output based on the information and context given by the users.

ⓘ Note

Face blurring

For inputs to GPT-4 Turbo with Vision and GPT-4o that contain images or videos of people, the system will first blur faces prior to processing to return the requested results. Blurring helps protect the privacy of the individuals and groups involved.

Blurring shouldn't affect the quality of your completions, but you might see the system refer to the blurring of faces in some instances.

ⓘ Important

Any identification of an individual is neither the result of facial recognition nor the generation and comparison of facial templates. The identification is a result of training the model to associate images of an individual with the same name through image tagging, whereby the model returns the name with any subsequent image inputs of that individual. The model can also take contextual cues other than the face, which is how the model can still associate the image with an individual even if the face is blurred. For example, if the image contains a photo of a popular athlete wearing their team's jersey and their specific number, the model can still detect the individual based on the contextual cues.

Limited Access to Azure AI Content Understanding

The Face grouping feature in Content Understanding is a Limited Access service and registration is required for access to it. For more information, see [Microsoft's Limited Access Policy](#) and access the [Face API registration](#). Certain features are only available to Microsoft managed customers and approved partners, and only for certain use cases selected at the time of registration. Note that facial detection, facial attributes, and facial redaction use cases do not require registration.

ⓘ Note

On June 11, 2020, Microsoft announced that it will not sell facial recognition technology to police departments in the United States until strong regulation, grounded in human rights, has been enacted. As such, customers may not use facial recognition features or functionality included in Azure Services—such as Face, Video Indexer, or Content Understanding—if a customer is, or is allowing use of such services by or for, a police department in the United States.

Use cases

Intended uses

Here are some examples of when you might use Content Understanding.

- **Tax process automation:** You can utilize Content Understanding's document extraction feature to extract fields from tax forms. Regardless of different templates, you will be able to extract key data from tax forms to generate a unified view of information that results in tax process automation.
- **Call center post-call analytics:** Businesses can generate insights from call recordings. Audio input will be transformed into text transcription output, which can be used to extract valuable insights that leads to improved call center efficiency and customer experience.
- **Marketing Automation and DAM (Digital Asset Management):** To build a media asset management solution, you can use Content Understanding to extract fields defined in schema from images and videos to extract insights to enhance the relevance of targeted advertising.
- **Content search and discovery with RAG (Retrieval Augmented Generation):** Customers who need to search and discover content of any modality (such as text, images, audio, video, or mixed media,) based on their content, metadata, or features can use the structured output from Content Understanding to enable downstream RAG scenarios.
- **Content or media summarization:** For example, a media company could use Content Understanding to generate a summary and highlights of the sports event.
- **Chart and graph understanding:** Financial forms or academic journals that contain charts and graphs are usually hard to understand when only the text is being extracted. Content Understanding solves the problem by interpreting the charts and graphs in the context of the given document or image itself, and users can easily extract information they want such as the type of chart or graph, summary, and overall meaning.

Considerations when choosing other use cases

Please consider the following factors when you choose a use case:

- **Avoid scenarios where use or misuse could result in physical or psychological harm.** For example, using Content Understanding to diagnose patients or prescribe medications can cause significant harm.

⊗ Caution

Content Understanding is not designed, intended, or made available as a medical device, and is not designed or intended to be a substitute for professional medical advice, diagnosis, treatment, or judgment, and should

not be used to replace or substitute professional medical advice, diagnosis, treatment, or judgment.

- **Not suitable for biometric identification or verification.** For example, Content Understanding was not designed or intended for the unique identification or verification of individuals based on their facial geometry, voice patterns, or other physical, physiological, or behavioral characteristics.

Important

If you are using Microsoft products or services to process Biometric Data, you are responsible for: (i) providing notice to data subjects, including with respect to retention periods and destruction; (ii) obtaining consent from data subjects; and (iii) deleting the Biometric Data, all as appropriate and required under applicable Data Protection Requirements. "Biometric Data" will have the meaning set forth in Article 4 of the GDPR and, if applicable, equivalent terms in other data protection requirements. For related information, see Data and Privacy for Face.

- **Avoid use for tracking people in real world contexts.** Examples include using Content Understanding for surveillance of individuals in real world contexts or using Content Understanding to verify that individuals pictured in separate locations are the same person. This recommendation does not apply to using Context Understanding for creative purposes, like to find different scenes of a move with the same actor.
- **Avoid scenarios where use or misuse of the system could have a consequential impact on life opportunities or legal status.** Examples include scenarios where the use of Content Understanding could affect an individual's legal status, legal rights, or their access to credit, education, employment, healthcare, housing, insurance, social welfare benefits, services, opportunities, or the terms on which they're provided. Consider incorporating meaningful human review and oversight to help reduce the risk of harmful outcomes.
- **Carefully consider use cases in high stakes domains or industry.** Examples include but are not limited to healthcare, medicine, finance, or legal.
- **Avoid use for task-monitoring systems that can interfere with privacy.** Content Understanding's underlying AI models were not designed to monitor individual patterns to infer intimate personal information, such as an individual's sexual or political orientation.
- **Avoid scenarios in which use, or misuse of the system could spread false narratives about sensitive topics or people.** Examples include the creation and

distribution of misinformation about highly sensitive events or generation of information about real people in circumstances that reflect a false narrative.

- **Carefully consider the supported locales and languages:** Content Understanding model has different supported locales and languages. For example, within English language itself, there are different locales such as US, UK and Australia, which has differences in how the time is formatted, as well as spellings for some words. Be sure to carefully check the officially supported locales and languages for each modality.
- **Avoid use where a human in the loop or secondary verification method is not available.** Fail-safe mechanisms (e.g., a secondary method being available to the end user if the technology fails) help prevent denial of essential services or other harms due to errors in output.
- **Not suitable for scenarios where up-to-date, factually accurate information is crucial** unless you have human reviewers or are using the models to search your own documents and have verified suitability for your scenario. Content Understanding does not have information about events that occur after its training date, likely has missing knowledge about some topics, and might not always produce factually accurate information.
- **Conversation transcription with speaker recognition:** Content Understanding is not designed to provide diarization with speaker recognition, and it cannot be used to identify individuals. In other words, speakers will be presented as Guest1, Guest2, Guest3, and so on, in the transcription. These will be randomly assigned and may not be used to identify individual speakers in the conversation. For each conversation transcription, the assignment of Guest1, Guest2, Guest3, and so on, will be random.
- **Legal and regulatory considerations.** Organizations need to evaluate potential specific legal and regulatory obligations when using Content Understanding. Content Understanding is not appropriate for use in every industry or scenario. Always use Content Understanding in accordance with the applicable terms of service and the relevant codes of conduct, including the Generative AI Code of Conduct.
- **Legal and regulatory considerations:** Organizations need to evaluate potential specific legal and regulatory obligations when using any AI services and solutions, which may not be appropriate for use in every industry or scenario. Additionally, AI services or solutions are not designed for and may not be used in ways prohibited in applicable terms of service and relevant codes of conduct.

Limitations

Technical limitations, operational factors, and ranges

As with all AI systems, there are some limitations to Content Understanding that customers should be aware of.

If highly disturbing input files are uploaded to Content Understanding, it can return harmful and offensive content as part of the results. To mitigate this unintended result, we recommend that you control access to the system and educate the people who will use it about appropriate use.

Face grouping

Faces are blurred before the image or video is sent to the model for analysis thus inference on faces, such as emotion, won't work in either image or video. Only video modality supports face grouping which only provides groups of similar faces without any additional analysis.

Important

Face grouping feature in Content Understanding is limited based on eligibility and usage criteria. in order to support our Responsible AI principles. Face service is only available to Microsoft managed customers and partners. Use the [Face Recognition intake form](#) to apply for access. For more information, see the [Face limited access page](#).

Document

Document extraction capability is heavily dependent on the way you name the fields and description of the fields. Also, the product forces grounding – anchoring outputs in the text of the input documents – and will not return answers if they cannot be grounded. Therefore, in some cases, the value of the field may be missing. Due to the nature of the grounded extraction, the system will return content from the document even if the document is incorrect or the content is not visible to the human eye. Documents should also have a reasonable resolution, where the text is not too blurry for the [Layout model](#) to recognize.

Video

Content Understanding is not intended to replace the full viewing experience of videos, especially for content where details and nuances are crucial. It's also not designed for

summarizing highly sensitive or confidential videos where context and privacy are paramount.

- **Video quality:** Always upload high-quality video and audio content. The recommended maximum frame size is HD and frame rate is 30 FPS. A frame should contain no more than 10 people. When outputting frames from videos to AI models, only send around one frame per second. Processing 10 or more frames might delay the AI result. At least 1 minute of spontaneous conversational speech is required to perform analysis. Detecting non-speech audio signals like sound effects and singing is not supported.
- Lower accuracy of the generated insights might occur when faces recorded by cameras that are high-mounted, down-angled or with a wide field of view (FOV) may have fewer pixels.
- Detectors may misclassify objects in videos that are in an overhead view as they were trained with a frontal view of objects.
- **Non-English languages:** Content Understanding was primarily tested and optimized for the English language. When applied to non-English languages, the accuracy and quality of the summaries may vary. To mitigate this limitation, users employing the feature for non-English languages should verify the generated summaries for accuracy and completeness.
- **Videos with multiple languages:** If a video incorporates speech in multiple languages, the Textual Video Summary may struggle to accurately recognize all the languages featured in the video content. Users should be aware of this potential limitation when utilizing the Textual Video Summarization feature for multilingual videos.
- **Highly specialized or technical videos:** Video Summary AI models are trained on a wide variety of videos, including news, movies, and other general content. If the video is highly specialized or technical, the model might not be able to accurately extract the summary of the video.
- **Videos with poor audio quality nor (optical character recognition) OCR:** Textual Video Summary AI models rely on audio and other insights to extract the summary from the video, or on OCR to extract the text appearing on screen. If the audio quality is poor and there's no identified text, the model might not be able to accurately extract the summary from the video.
- **Videos with low lighting or fast motion:** Videos that are shot in low lighting or have fast motion might be difficult for the model to process the insights, resulting in poor performance.
- **Videos with uncommon accents or dialects:** AI models are trained on a wide variety of speech, including different accents and dialects. However, if the video contains speech with an accent or dialect that isn't well represented in the training data, the model might struggle to accurately extract the transcript from the video.

Audio

For audio files, you may need to specify a locale for each audio input. The locale must match the actual language that's spoken in an input voice. Content Understanding supports automatic language detection as well for some use cases. For more information, see the list of [supported locales](#).

- **Acoustic quality:** Speech to text–enabled applications and devices may use a wide variety of microphone types and specifications. Unified speech models have been trained on various voice audio device scenarios, such as telephones, mobile phones, and speaker devices. Voice quality might be degraded by the way a user speaks into a microphone, even if they use a high-quality microphone. For example, if a speaker is located far from the microphone, the input quality may be too low. A speaker who is too close to the microphone could also cause audio quality deterioration. These cases, as well as any cases where it causes the audio file quality to be degraded can adversely affect the accuracy of speech to text.
- **Non-speech noise:** If an input audio contains a certain level of noise, accuracy is affected. Noise that comes from audio devices that are used to make a recording, or audio input itself might contain noise, such as background or environmental noise.
- **Overlapped speech:** There might be multiple speakers within range of an audio input device, and they might speak at the same time. Audio files that have other speakers voice recorded in the background while the main speaker is recording also results in overlapped speech file. In addition, although there is no limitation on the numbers of speakers in the conversation, the system performs better when the number of speakers is under 30.
- **Vocabularies:** If a word that doesn't exist in a model appears in the audio, the result is an error in transcription.
- **Accents:** Even within one locale, such as in English - United States (en-US), many people have different accents. Very specific accents might also lead to an error in transcription.
- **Mismatched languages or locales:** If you specified English - United States (en-US) for an audio input, but a speaker spoke in Swedish, for example, accuracy would be reduced.
- **Insertion errors:** At times, the model can produce insertion errors in the presence of noise or soft background speech.

Image

- **Object Recognition:** Recognition of certain ambiguous products may not be accurate if it cannot be recognized by the model. Abstract concepts that do not

correspond to the image, for example, gender and emotion, may not be recognized as well.

System performance

Performance metrics differ for each modality within Content Understanding. Each modality will have different industry standards for measuring AI performance.

One common metric we provide in Content Understanding across all modality is the confidence score for the fields. As of now, only fields kind of "extract" and "generate" will have confidence scores.

A distinctive feature of Content Understanding is its support for grounding and confidence scores, currently available only for the document modality but planned for future expansion. Grounding in documents includes page numbers and bounding boxes for extracted values, aiding the user experience by highlighting locations for human review and correction. Confidence scores, ranging from 0 to 1, estimate the accuracy of extracted values based on analyzed or training documents, with higher scores indicating greater confidence. For guidelines on using confidence scores, refer to the Evaluation section of Content Understanding.

Below are general performance metrics you can utilize for each modality:

Document

Accuracy

Text is composed of lines and words at the foundational level and entities such as names, prices, amounts, company names, and products at the document understanding level.

Word-level accuracy

A measure of accuracy for OCR is word error rate (WER), or how many words were incorrectly output in the extracted results. The lower the WER, the higher the accuracy.

WER is defined as:

$$WER = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C}$$

Where:

Term	Definition	Example
S	Count of incorrect words ("substituted") in the output.	"Velvet" gets extracted as "Veivet" because "l" is detected as "i."
D	Count of missing ("deleted") words in the output.	For the text "Company Name: Microsoft," Microsoft isn't extracted because it's handwritten or hard to read.
I	Count of nonexistent ("inserted") words in the output.	"Department" gets incorrectly segmented into three words as "Dep artm ent." In this case, the result is one deleted word and three inserted words.
C	Count of correctly extracted words in the output.	All words that are correctly extracted.
N	Count of total words in the reference ($N=S+D+C$) excluding I because those words were missing from the original reference and were incorrectly predicted as present.	Consider an image with the sentence, "Microsoft, headquartered in Redmond, WA announced a new product called Velvet for finance departments." Assume the OCR output is ", headquartered in Redmond, WA announced a new product called Veivet for finance dep artm ents." In this case, S (Velvet) = 1, D (Microsoft) = 1, I (dep artm ents) = 3, C (11), and N = S + D + C = 13. Therefore, WER = $(S + D + I) / N = 5 / 13 = 0.38$ or 38% (out of 100).

Document and entity-level accuracy At the document level, for example, in the case of an invoice or receipt, an error of only one character in the entire document might be rated insignificant. If that error is in the text that represents the paid amount, the entire invoice or receipt might get flagged as incorrect.

Another metric is entity error rate (EER). It's the percentage of incorrectly extracted entities, such as names, prices, amounts, and phone numbers, out of the total number of the corresponding entities in one or more documents. For example, for a total of 30 words representing 10 names, 2 incorrect words out of 30 equals 0.06 (6%) WER. But if that results in 2 names out of 10 as incorrect, the Name EER is 0.20 (20%), which is much higher than the WER.

Measuring both WER and EER is a useful exercise to get a full perspective on document understanding accuracy.

Video

Accuracy of video analysis depends on several factors including camera placement and the interpretation of the system's output. The accuracy should be assessed by how

closely the mode's field value results align with the actual content of the video. For instance, when a user looks for entities within a video, it is expected to return a full list of entities found in the video. To evaluate accuracy, specific test datasets, representative of various real-world scenarios and conditions, are used. These datasets include a wide range of video content types and user interaction scenarios.

[+] [Expand table](#)

Term	Definition
True Positive	The system-generated output correctly corresponds to a real event.
True Negative	The system correctly does not generate an event when a real event has not occurred.
False Positive	The system incorrectly generates/extracts/classifies an output when no real event has occurred.
False Negative	The system incorrectly fails to generate an output when a real event has occurred.

Audio

System performance is measured by these key factors:

- Word error rate (WER)
- Token error rate (TER)
- Runtime latency

A model is considered better only when it shows significant improvements (such as a 5% relative WER improvement) in all scenarios (like transcription of conversation speech, call center transcription, dictation, and voice assistant) while being in line with the resource usage and response latency goals.

For diarization, we measure the quality by using word diarization error rate (WDER). The lower the WDER, the better the quality of diarization.

Image

The accuracy of image analysis is a measure of how well the outputs correspond to actual visual content that is present in images. To measure accuracy for image analysis, you might evaluate the image with your ground-truth data and compare the output of the AI model. By comparing the ground truth with AI-generated results, you can classify

events into two kinds of correct ("true") results and two kinds of incorrect ("false") results:

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

Term	Definition
True Positive	The system-generated output correctly corresponds to ground-truth data. For example, the system correctly tags an image of a dog as a dog.
True Negative	The system correctly does not generate results that are not present in the ground-truth data. For example, the system correctly does not tag an image as a dog when no dog is present in the image.
False Positive	The system incorrectly generates an output that is absent in the ground-truth data. For example, the system tags an image of a cat as a dog.
False Negative	The system fails to generate results that are present in the ground-truth data. For example, the system fails to tag an image of a dog that was present in the image.

These event categories are used to calculate precision and recall:

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

Term	Definition
Precision	A measure of the correctness of the extracted content. From an image that contains multiple objects, you find out how many of those objects were correctly extracted.
Recall	A measure of the overall content extracted. From an image that contains multiple objects, you find out how many objects were detected overall, without regard to their correctness.

The precision and recall definitions imply that, in certain cases, it can be hard to optimize for both precision and recall at the same time. Depending on your scenario, you might need to prioritize one over the other. For example, if you are developing a solution to detect only the most accurate tags or labels in the content, such as to display image search results, you would optimize for higher precision. But if you're trying to tag all possible visual content in the images for indexing or internal cataloging, you would optimize for higher recall.

Best practices for improving system performance

In most cases, improving system performance is heavily dependent on the user providing data that is reasonably understandable for Content Understanding to extract values from.

Make sure that the fields generated from the content are relevant to your downstream intended uses. For example, if you want to search for "dogs playing in the backyard," make sure that your field output includes these concepts and update the schema definition like field name and descriptions of the fields to correct it if it does not.

For images, see the following [documentation](#) for specific input requirements. Images should have reasonable quality, light exposure and contrast.

For audio, mismatching locales reduces accuracy, so it is important to match input locales to the speakers in the file. Use audio files with reasonable acoustic conditions and avoid files with background noise, side speech, distance to microphone and speaking styles that may adversely affect the accuracy.

Taking into consideration the limitations of each modality with regards to currently supported inputs, languages and locales, and scenarios will also help improve system performance.

For document extraction, however, there are ways to improve the analyzer quality, which is to update or correct the field label results as needed with each document that you are adding to the dataset. Document extraction feature supports in-context learning, so more dataset and accurate field labels will lead to a better system performance in general. For filled-in forms, it is also recommended to use examples that have all the fields filled in and use real-world values that you expect to see for each field.

Evaluation of Content Understanding

Evaluation methods

To create Content Understanding, we prepared datasets that target common customer use cases. These are independently prepared by Microsoft, and we do not use customer data sent to our services for any training or evaluation purposes.

Content Understanding's effectiveness will depend on the specific applications it's used for. Customers should perform their own tests to guarantee the best results.

For example, in document extraction, the service assigns a confidence value from 0 to 1 for each word and field. Running a pilot can help customers determine confidence value ranges and extraction quality. They can then set thresholds, like sending results with confidence values of 0.80 or higher for straight-through processing, and those below for human review.

Evaluation results

To ensure service performance, we regularly conduct evaluations and error analysis, using the results to enhance our offerings. Many of these evaluations are tailored to customer scenarios and help determine constraints like field numbers and training data sizes. These constraints are documented for customer reference. Due to numerous possible scenarios, we can't test everything. For instance, we frequently test financial domains but have less coverage in medical fields.

Fairness considerations

One important dimension to consider when using AI systems, is how well the system performs for different groups of people. Research has shown that without conscious effort focused on improving performance for all groups, AI systems can exhibit varying levels of performance across different demographic factors such as race, ethnicity, gender, and age.

As part of our evaluation of Azure AI Content Understanding, we have conducted an analysis to assess potential fairness harms. We have examined the system's performance across different demographic groups, aiming to identify any disparities or differences that may exist and could potentially impact fairness.

In some cases, there may be remaining performance disparities. It is important to note that these disparities may exceed the target, and we are actively working to address and minimize any potential biases or performance gaps and seek diverse perspectives from a variety of backgrounds.

Regarding representational harms, such as stereotyping, demeaning, or erasing outputs, we acknowledge the risks associated with these issues. While our evaluation process aims to mitigate such risks, we encourage users to consider their specific use cases carefully and implement additional mitigations as appropriate. Having a human in the loop can provide an extra layer of oversight to address any potential biases or unintended consequences.

We are committed to continuously improving our fairness evaluations to gain a deeper understanding of the system's performance across various demographic groups and potential fairness concerns. The evaluation process is ongoing, and we are actively working to enhance fairness and inclusivity and mitigate any identified disparities. You can find more fairness testing related to speech in this [documentation](#).

Evaluating and integrating Image Analysis for your use

When integrating Content Understanding for your use case, knowing that Content Understanding is subject to the [Microsoft Generative AI Services Code of Conduct](#), and [Code of Conduct for Azure AI Vision Face](#) will ensure a successful integration.

When you're getting ready to integrate Content Understanding to your product or features, the following activities help to set you up for success:

- **Understand what it can do:** Fully assess the potential of Content Understanding to understand its capabilities and limitations. Understand how it will perform in your scenario and context. For example, if you're using audio content extraction, test with real-world recordings from your business processes to analyze and benchmark the results against your existing process metrics.
- **Respect an individual's right to privacy:** Only collect data and information from individuals from whom you have obtained consent, and for lawful and justifiable purposes.
- **Legal and regulatory considerations.** Organizations need to evaluate potential specific legal and regulatory obligations when using Content Understanding. Content Understanding is not appropriate for use in every industry or scenario. Always use Content Understanding in accordance with the applicable terms of service and the [Microsoft Generative AI Services Code of Conduct](#).
- **Human-in-the-loop:** Keep a human in the loop, and include human oversight as a consistent pattern area to explore. This means ensuring constant human oversight of the AI-powered product or feature and to maintain the role of humans in decision-making. Ensure that you can have real-time human intervention in the solution to prevent harm. A human in the loop enables you to manage situations when Content Understanding does not perform as required.
- **Security:** Ensure your solution is secure and that it has adequate controls to preserve the integrity of your content and prevent unauthorized access.

Learn more about responsible AI

- [Microsoft AI principles ↗](#)
- [Microsoft responsible AI resources ↗](#)
- [Microsoft Azure Learning courses on responsible AI](#)

Learn more about Content Understanding

- [Azure OpenAI overview](#)
- [Azure AI Document Intelligence overview](#)
- [Azure AI Speech overview](#)
- [Azure AI Vision overview](#)

- Azure AI Face service overview
- Azure AI Video Indexer overview

Additional transparency notes for underlying services

- Azure OpenAI
- Azure AI Document Intelligence
- Azure AI Speech
- Azure AI Vision
- Azure AI Face
- Azure AI Video Indexer

Codes of conduct

- Azure OpenAI
- Azure AI Vision
- Azure AI Speech

Data, privacy, and security for Content Understanding

Article • 11/19/2024

Content Understanding builds upon the functionality of Document Intelligence, Speech to Text, Image Analysis, Face, Video and Azure OpenAI, each designed with compliance, privacy, and security at its core. This combined service processes various types of customer-provided data, such as documents, audio, images, biometric data (face), text, and video to deliver powerful analysis and intelligence capabilities. Importantly, users are responsible for ensuring that their use of this service complies with all applicable laws and regulations in their jurisdiction, including data protection, privacy, and communications laws, as well as any specific requirements around biometric data when leveraging facial recognition features. It is essential to acquire all necessary permissions, licenses, or third-party rights for the content and data submitted for processing.

Since the data processed in this integrated service may involve personal or sensitive information, including biometric identifiers and human speech content, users must follow all jurisdictional requirements related to data protection. For instance, when using biometric technologies, it is crucial to provide clear, conspicuous disclosure to individuals, particularly in regions with strict biometric data governance. Data provided to the Azure OpenAI service is stored and processed to monitor compliance with product terms, and [Microsoft's Products and Services Data Protection Addendum](#) applies to all data handling within the Azure OpenAI framework. By combining these technologies, our service offers robust insights while ensuring users maintain responsibility for adhering to legal and regulatory standards.

What data does Content Understanding process?

Content Understanding can process audio input or voice audio, image files, document files and video files. Each input type has different file limits, such as file type, size, length and resolution. The limits are outlined in the [service quotas and limits documentation](#).

How does Content Understanding process data?

Authenticate

Content Understanding first requires users to authenticate access to Content Understanding API by using Azure AI services API key. Each request to the service URL must include an authentication header. This header passes along an API key (or token if applicable), which is used to validate your subscription for a service. Apart from authenticating access with API Key, Content Understanding also supports AAD and Entra ID Authentication. For more information, see [Authenticate requests to Azure AI services](#), which has additional information on AAD, Entra ID, and authorizing access to managed identities.

Secure data in transit

All Azure AI services endpoints use HTTPS URLs for encrypting data during transit. The client operating system needs to support Transport Layer Security (TLS) 1.2 for calling the end points. For more information, see [Transport Layer Security](#). The incoming data is processed in the same region where the Azure resource was created.

Encrypts input data for processing

The incoming data is processed in the same region where the Content Understanding resource was created. When you submit your files to a Content Understanding operation, it starts the process of analyzing the input. Your data and results are then temporarily encrypted and stored in Azure Storage before it is sent to Azure OpenAI for further processing.

Retrieve the results

The "Get Result" operation is authenticated against the same API key that was used to call the "Analyze" operation to ensure no other customer can access your data. It returns the analysis job completion status, When the status shows as succeeded, the operation also returns the extracted results in JSON format.

Data retention

For all the input, they are processed data as soon as possible, and the input files are not retained or stored in the service after processing. Analysis result may be stored for up to 24 hours. The name of the analyzers will be logged for reporting and debugging.

Regional Process

Data does not get stored outside the designated region that the user selected for the Content Understanding resource, even temporarily. However, based on OpenAI availability, we may route traffic to another region within the same geo moving forward.

Face

Face is a gated feature as it processes biometric data. We detect faces in the input files and group them by their similarity. All intermediate data do not persist beyond the processing of the request. The face groupings associated with analysis results are persisted for 48 hours unless the user explicitly deletes face data. For more information, please refer to the [Data and Privacy for Face documentation](#).

Azure OpenAI

Content Understanding also utilizes Azure OpenAI model once each modality input is processed through the underlying AI services. Please refer to the [Azure OpenAI Data, privacy, and security documentation](#) for more information.

Data Plane

Article • 11/19/2024

REST Operation Groups

 Expand table

Operation Group
Analyzers