

Create a multimodal analysis solution with Azure Content Understanding

1. Introduction

<https://learn.microsoft.com/en-us/training/modules/analyze-content-ai/01-introduction>

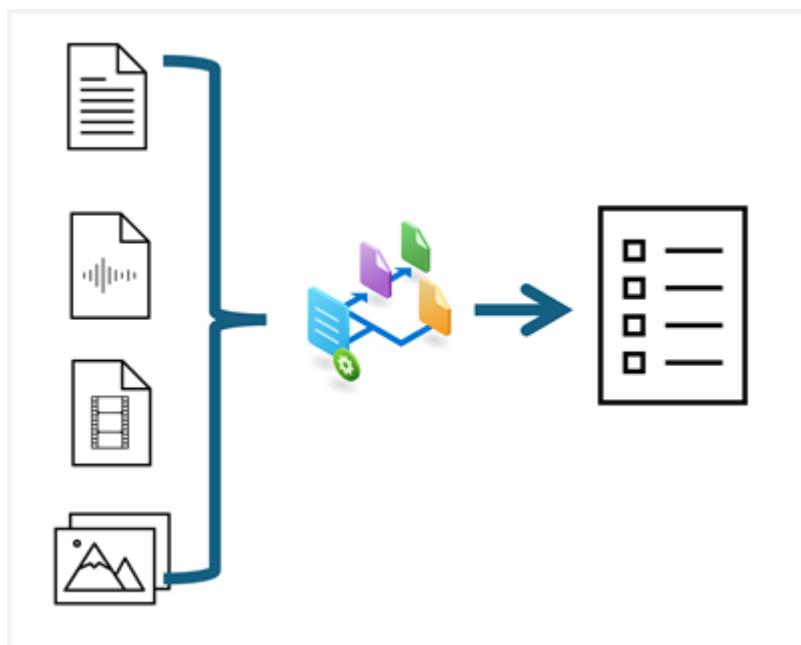
Introduction

Completed

- 1 minute

Organizations today rely on information that is often locked up in content assets such as documents, images, videos, and audio recordings. Extracting information from this content can be challenging, laborious, and time-consuming, and organizations often need to build solutions based on multiple technologies for content analysis depending on the formats being used.

Azure Content Understanding is a multimodal service that simplifies the creation of AI-powered analyzers that can extract information from content in practically any format.



In this module, you'll explore the capabilities of Azure Content Understanding, and learn how to use it to build custom analyzers.

Note

Azure Content Understanding is currently in public preview. Details described in this module are subject to change.

2. What is Azure Content Understanding?

<https://learn.microsoft.com/en-us/training/modules/analyze-content-ai/02-content-understanding>

What is Azure Content Understanding?

Completed

- 5 minutes

Azure Content Understanding is a generative AI service that you can use to extract insights and data from multiple kinds of content. With Content Understanding, you can quickly build applications that analyze complex data and generate outputs that can be used to automate and optimize processes.

Content Understanding is a component of Foundry Tools. To use it, you need to provision a Foundry Tools resource in your Azure subscription. You can develop and manage a Content Understanding solution:

- In the Microsoft Foundry portal
- By using the Content Understanding REST API

Multimodal content analysis

Content Understanding can extract information from common kinds of content, enabling you to use a single service with a straightforward and consistent development process to build multimodal content analysis solutions.

Documents and forms

You can use Content Understanding to analyze documents and forms and retrieve specific field values. For example, you could extract key data values from an invoice to automate payment processing.

The screenshot shows the Azure AI Foundry interface for 'Invoice Analysis'. On the left, there's a sidebar with various icons and two uploaded PDF files: 'invoice-1234.pdf' and 'invoice-1235.pdf'. The main area is titled 'Test analyzer' and displays the contents of an invoice from 'Contoso Ltd' with address '2 Main St, Bigtown, England, EH1 234' and phone 'Tel: 555 123-4567'. The invoice number is 1235 and the date is 03/07/2025. The details section shows 'Customer Name: Ava Jones' and 'Address: 321 Pond Lane, Waterville, England, GL1 010'. Below is a table of items:

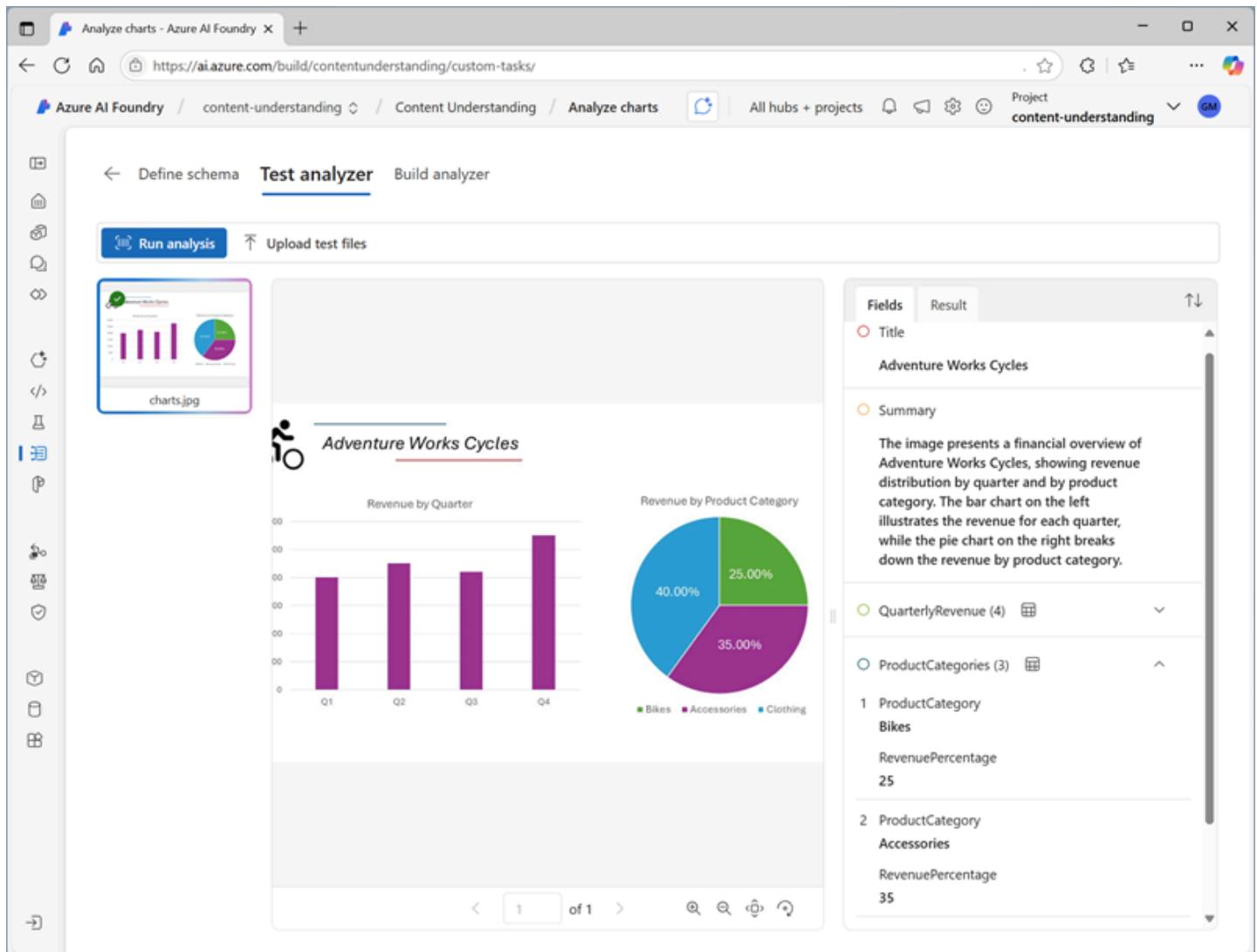
Item	Price	Quantity	Item Total
42mm Widget	26.50	3	79.50
8mm screws pack	5.40	2	11.80
			Subtotal 91.48
			Tax 9.14
			Shipping 15.00
			Total Due 115.62

On the right, a 'Fields' panel lists extracted fields with their confidence scores:

Fields	Result	Score
CustomerAddress	p.1	86.90%
CustomerName	p.1	96.80%
InvoiceDate	p.1	99.80%
InvoiceId	p.1	97.20%
InvoiceTotal	p.1	97.60%
SubTotal	p.1	98.90%
TotalTax	p.1	98.40%
VendorAddress	p.1	97.80%

Images

You can analyze images to infer information from visuals such as charts, identify physical defects in products or other items, detect the presence of specific objects or people, or determine other information visually.



Audio

Analysis of audio enables you to automate tasks like summarizing conference calls, determining sentiment of recorded customer conversations, or extracting key data from telephone messages.

The screenshot shows the Azure AI Foundry interface for Content Understanding, specifically the 'Test analyzer' section. On the left, there's a sidebar with various icons. In the center, a file named 'call.mp3' is selected. A play button and a timeline from 00:00 to 00:24 are visible. Below the timeline is a 'Transcript' section containing the following text:

```
00:02.560 --> 00:05.120
<v Speaker 1>Hi, this is Ava from Contoso.

00:05.520 --> 00:08.000
<v Speaker 1>Just calling to follow up on our meeting last week.

00:08.320 --> 00:12.880
<v Speaker 1>I wanted to let you know that I've run the numbers and I think we can
meet your price expectations.

00:13.040 --> 00:21.520
<v Speaker 1>Please call me back on 555-12345 or send me an e-mail at
Ava@contoso.com and we'll discuss next steps.

00:21.960 --> 00:23.280
<v Speaker 1>Thanks, bye.***
```

To the right, there are two tabs: 'Fields' and 'Result'. The 'Fields' tab is active, showing the following extracted data:

- Caller**: Ava from Contoso
- Summary**: Ava from Contoso called to follow up on a meeting and mentioned that they can meet the price expectations. She requested a callback or an email to discuss the next steps.
- Actions**: Call back or send an email to discuss next steps.
- CallbackNumber**: 555-12345
- AlternativeContacts (1)**: Ava@contoso.com

Video

Video accounts for a large volume of the data captured today, and you can use Content Understanding to analyze and extract insights from video to support many scenarios. For example, to extract key points from video conference recordings, to summarize presentations, or to detect the presence of specific activity in security footage.

The screenshot shows the Azure AI Foundry Content Understanding interface. On the left, there's a sidebar with various icons. In the center, a video player displays two frames: one of a man named Harry and one of a woman named Lisa. Below the frames is a bar chart titled "Projected user adoption". A timeline slider is positioned between the frames, with markers at 00:33 and 00:37. To the right of the video player is a results panel. The "Fields" tab is selected, showing a summary of the analysis. The summary states: "Harry shares a slide with projected user adoption metrics for the first three months, which impresses Lisa. She requests him to email the slide, and Harry agrees, encouraging Lisa to impress the board with her presentation." Below this, under "participants", it lists "1 Harry" and "2 Lisa". Under "assignedActions", it lists "assignedTo Harry task Send the presentation slide to Lisa". Under "sharedSlides", it lists "Projected user adoption metrics for the first three months".

3. Create a Content Understanding analyzer

<https://learn.microsoft.com/en-us/training/modules/analyze-content-ai/03-create-analyzer>

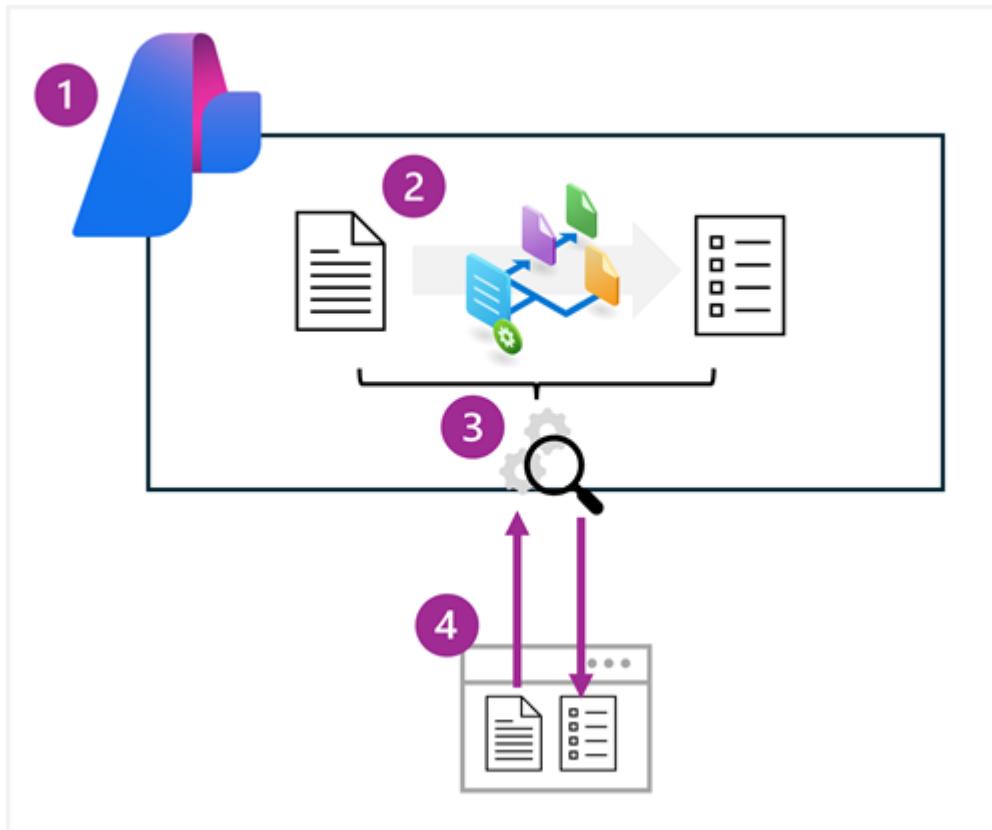
Create a Content Understanding analyzer

Completed

- 5 minutes

Content Understanding solutions are based on the creation of an *analyzer*, which is trained to extract specific information from a particular type of content based on a *schema* that you define.

The high-level process for creating a Content Understanding solution includes the following steps:



1. Create a Foundry Tools resource.
2. Define a Content Understanding schema for the information to be extracted. This can be based on a content sample and an analyzer template.
3. Build an analyzer based on the completed schema.
4. Use the analyzer to extract or generate fields from new content.

Numerous analyzer templates are provided to help you develop an appropriate analyzer for your needs quickly. Additionally, because of the generative AI capabilities of Content Understanding, you can use minimal training data to define a schema by example. In many cases, the service accurately identifies the data values in the sample content that map to the schema elements automatically, though you can also explicitly label fields in content such as documents to improve the performance of your analyzer.

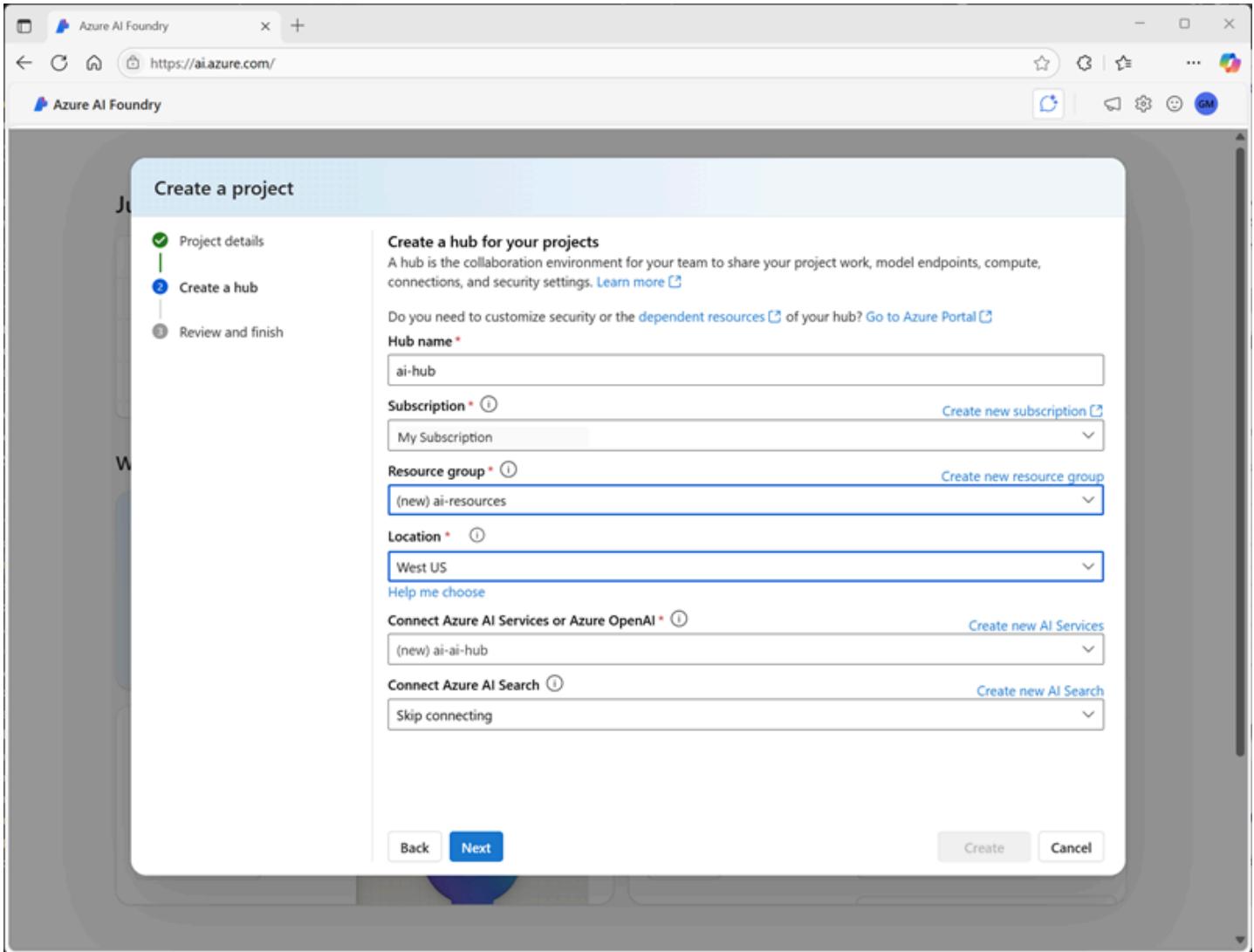
Creating an analyzer with Microsoft Foundry

While you can provision a Foundry Tools resource and develop a complete Content Understanding solution through the REST API, the preferred approach for AI development projects is to use Microsoft Foundry. Specifically, you can use the Microsoft Foundry portal to create a project, define a Content Understanding schema, and build and test an analyzer.

Creating a Content Understanding project

In Microsoft Foundry, you can create a project in an existing AI hub, or you can create a new hub as you create the project. In addition to the AI hub itself, creating a hub provisions the Azure resources

needed to support one or more projects; including a Foundry Tools resource, storage, and a key vault resource to store sensitive details like credentials and keys.



Note

Content Understanding schemas can only be created in Azure locations where the service is supported. For more information, see [Content Understanding region and language support](#).

Defining a schema

After creating a project, the first step in building an analyzer is to define a schema for the content the analyzer will process, and the information it will extract. Microsoft Foundry provides a schema editor interface in which you can upload a file (document, image, audio, or video) on which the schema should be based. You can then apply an appropriate schema template and define the specific fields you want the analyzer to identify.

Contoso Ltd
2 Main St, Bigtown, England, EH1 234
Tel: 555 123-4567

Customer Name:	John Smith
Address:	123 River Street Marshtown England GL1 234
Item	Price
38mm Widget	24.50
3.5mm screws pack	4.99
Left-handed screwdriver	7.49

Note

The templates and field types available in a schema depend on the content type of the file on which the schema is based. Some content types support additional optional functionality, such as extracting barcodes and formulae from text in documents. For more information about using Content Understanding with different content types, see the following articles in the product documentation:

- [Content Understanding document solutions](#)
- [Content Understanding image solutions](#)
- [Content Understanding audio solutions](#)
- [Content Understanding video solutions](#)

Testing

You can test the analyzer schema at any time during the development process by running analysis on the sample file used to define the schema or other uploaded files. The test results include the extracted field values and the JSON format output returned by the analyzer to client applications.

The screenshot shows the Azure AI Foundry Content Understanding Test analyzer interface. On the left, there's a sidebar with various icons and two uploaded PDF files: "invoice-1234.pdf" and "invoice-1235.pdf". The main area displays the content of "invoice-1235.pdf", which includes a header for "Contoso Ltd" with address "2 Main St, Bigtown, England, EH1 234" and phone number "Tel: 555 123-4567". It also shows a table of items with columns for Item, Price, Quantity, and Item Total, along with sub-totals for Tax and Shipping, and a final Total Due. To the right, a results table lists extracted fields with their values and confidence scores: CustomerAddress (86.90%), CustomerName (96.80%), InvoiceDate (99.80%), InvoicedId (97.20%), InvoiceTotal (97.60%), SubTotal (98.90%), TotalTax (98.40%), and VendorAddress (97.80%).

Building an analyzer

When you're satisfied with the performance of your schema, you can build your analyzer. Building an analyzer makes it accessible to client applications through Content Understanding endpoint for the Foundry Tools resource associated with your project.

A screenshot of the Azure AI Foundry portal. The top navigation bar shows 'Invoice Analysis - Azure AI Foundry' and the URL 'https://ai.azure.com/build/contentunderstanding/custom-tasks/'. The main content area is titled 'Build analyzer'. It features a table with columns: Name, Description, Date built, and Status. One row is listed: 'Name' is 'invoice-analyzer', 'Description' is 'An analyzer for invoices', 'Date built' is '5/15/2025, 3:21 PM', and 'Status' is 'Ready'. Below the table, there are two numbered steps: 1. 'View sample code and get key, AAD token and endpoint configuration' with a 'View code' button, and 2. 'Follow the quickStart' with a 'Use REST API' button.

After building your analyzer, you can continue to test it in the Microsoft Foundry portal, and refine the schema to create new named versions with different capabilities.

4. Use the Content Understanding REST API

<https://learn.microsoft.com/en-us/training/modules/analyze-content-ai/04-use-api>

Use the Content Understanding REST API

Completed

- 5 minutes

The Content Understanding REST API provides a programmatic interface that you can use to create, manage, and consume analyzers.

To use the REST API, your client application submits HTTP calls to the Content Understanding endpoint for your Foundry Tools resource, passing one of the authorization keys in the header. You can obtain the endpoint and keys in the Azure portal or in the Microsoft Foundry portal. You can also use the Microsoft Foundry API to connect to the project and retrieve the endpoint and key for your Foundry Tools resource programmatically.

The screenshot shows the 'Overview - Azure AI Foundry' page in a browser. The URL is <https://ai.azure.com/build/overview>. The page title is 'content-understanding'. On the left, there's a sidebar with sections like 'Overview', 'Model catalog', 'Playgrounds', 'AI Services', 'Build and customize', 'Agents', 'Templates', 'Fine-tuning', 'Content Understanding', 'Prompt flow', 'Assess and improve', 'Tracing', 'Evaluation', 'Safety + security', 'My assets', 'Models + endpoints', 'Data + indexes', and 'Web apps'. The main content area has tabs for 'Endpoints and keys' (selected), 'Included capabilities', and 'Project details'. Under 'Endpoints and keys', there's an 'API Key' field containing a redacted value. Under 'Included capabilities', there are sections for 'Azure AI inference', 'Azure OpenAI', and 'Azure AI Services'. The 'Azure AI Services' section contains fields for 'Azure AI Services endpoint' (set to <https://xxxxxxxxxxxxxxxxxxxxxxxxxxxx.cognitiveservices.azure.com/>), 'Speech to text endpoint' (set to <https://westus.stt.speech.microsoft.com>), 'Text to speech endpoint' (set to <https://westus.tts.speech.microsoft.com>), and a link to '(x) API documentation'. Under 'Project details', there's a 'Project connection string' field containing a redacted value, and a 'Subscription' section set to 'My Subscription'.

Using the REST API to analyze content

One of the most common uses of the REST API is to submit content to an existing analyzer that you have previously built, and retrieve the results of analysis. The analysis request returns an operation ID value that represents an asynchronous task. Your client application must then use another request to pass the operation ID back to the endpoint and retrieve the operation status - potentially polling multiple times until the operation is complete and the results are returned in JSON format.

For example, to analyze a document, a client application might submit a POST request to the `analyze` function containing the following JSON body:

```
POST {endpoint}/contentunderstanding/analyzers/{analyzer}:analyze?api-version={api version}
{
```

```
        "url": "https://host.com/doc.pdf"
    }
```

Note

You can specify a URL for the content file location, or you can include the binary contents of the file.

Assuming the request is authenticated and initiated successfully, the response will be similar to this example:

```
Operation-Id: 1234abcd-1234-abcd-1234-abcd1234abcd
Operation-Location: {endpoint}/contentunderstanding/analyzers/{analyzer}/results/1234abcd-1234-abcd-1234-abcd1234abcd
{
    "id": "1234abcd-1234-abcd-1234-abcd1234abcd",
    "status": "NotStarted"
}
```

Your client application must then use the operation ID that has been returned to check the status of the operation until it has succeeded (or failed) by submitting a GET request to the `results` function.

```
GET {endpoint}/contentunderstanding/analyzers/{analyzer}/results/1234abcd-1234-abcd-1234-abcd1234abcd
```

When the operation has completed successfully, the response contains a JSON payload representing the results of the analysis. The specific results depend on the content and schema.

Note

For more information about the Content Understanding REST API, see the [reference documentation](#).

5. Exercise - Extract information from multimodal content

<https://learn.microsoft.com/en-us/training/modules/analyze-content-ai/05-exercise>

Exercise - Extract information from multimodal content

Completed

- 40 minutes

Now it's time to put what you've learned about Content Understanding into practice!

In this exercise, you use Azure Content Understanding to extract information from documents, images, audio files, and videos.

Note

To complete this lab, you need an [Azure subscription](#) in which you have administrative access.

Launch the exercise and follow the instructions.

[Launch Exercise](#)

6. Module assessment

<https://learn.microsoft.com/en-us/training/modules/analyze-content-ai/06-knowledge-check>

Module assessment

Completed

- 3 minutes

7. Summary

Summary

Completed

- 1 minute

Azure Content Understanding is a multimodal AI service that enables you to extract information from many different kinds of content. In this module, you learned how to use the Microsoft Foundry portal to create a Content Understanding project and build an analyzer.

Note

For more information about Azure Content Understanding, see [**Azure Content Understanding documentation**](#).