



SOFIA

AI **global AI
community**

INDUSTRIAL
SUMMIT

20
24



THANKS FOR SUPPORT US TODAY

 awara IT

Platinum Sponsor

 DataArt

 ocado[®]
TECHNOLOGY

Platinum Sponsor

 ICB | SOFTWARE
INNOVATION
A Kongsberg Digital company

<epam>

Silver Sponsor

 Microsoft

 Kinkao_Sofia
THAIFOOD + HOMEMADE

Partner

 SOFIA
TECHPARK

AI-powered data extraction: Azure OpenAI vs Document Intelligence!

Massimo Bonanni - Technical Trainer @ Microsoft





Data extraction from document

Extracting data from documents is a common scenario in many business processes like finance, legal, healthcare, and many others, where decision-making often depends on the information contained within various documents.

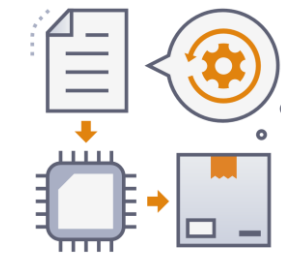
Involving the conversion of unstructured, semi-structured or structured document into a structured data.



Technologies and Tools

OCR Technology

For converting images of text into machine-encoded text. Advanced OCR solutions can handle complex layouts and even handwritten text.



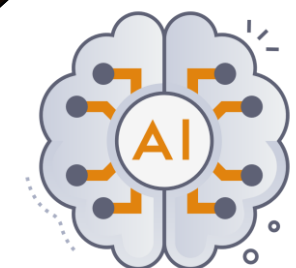
Natural Language Processing (NLP)

For understanding the context of the extracted text, identifying entities, and classifying them appropriately.



Machine Learning and Artificial Intelligence

For improving the accuracy of data extraction over time, learning from corrections and adjustments made during the post-processing phase.



Challenges



Different Document Formats

Document formats like PDFs, Word files, scans, and handwritten notes each need tailored methods for data extraction, adding complexity to the process.

Poor Document Quality

Quality issues like low resolution, skewing, noise, and distortions in scanned documents can impair OCR accuracy, resulting in data extraction errors.

Complex Document Layouts

Extracting data from documents with complex layouts like tables and graphs is challenging and prone to errors, necessitating sophisticated processing techniques.

Unstructured Data

Unstructured documents lack of format complicates data extraction, with NLP and ML offering solutions that demand considerable training data and advanced algorithms for context and semantics.



Document Intelligence





What is Document Intelligence?

Azure Document Intelligence is a cloud-based service that uses **machine learning** to automate the extraction of **text** and **structure** from documents.

Azure Document Intelligence **can be deployed in the cloud or at the edge**, and integrated with other Azure services for data processing, search, and analytics.

It is a powerful tool for enhancing data-driven strategies and enriching document search capabilities.





Document Intelligence Key Features



It can handle various types of documents, such as **Pdf, Images** or **Office documents** (only few models) composed by multiple pages



It offers **prebuilt models** for common scenarios (IDs, receipts, invoices, contracts) as well as custom models for specific document types



It allows users to **customize the extraction** process by training on their own data with minimal effort



It supports both **printed** and **handwritten** text, and provides built-in security and privacy for data and models

Models Overview



Pretrained document-analysis models

Read OCR model

Layout analysis model

Layout analysis model with KeyValuePairs option

Pretrained scenario-specific models

Pretrained models for specific scenario like US tax document, contract, invoice and so on

Custom models

Extraction Models

Classification Models



Add-on capabilities

Use the add-on features to **extend the results** to include more features extracted from your documents.

Some add-on features incur an **extra cost**.

To enable a feature, add the associated feature name to the features **query string** property.

You can enable **more than one add-on feature** on a request by providing a comma-separated list of features.





Custom extraction models

Neural Model

Best for **semi-structured** and **unstructured** documents to extract fields

Supports a subset of the field types supported by custom template models

Template Model

Is an easy-to-train document model that accurately extracts labeled key-value pairs, selection marks, tables, regions, and signatures from documents

Extracts fields from **structured** documents with defined visual templates.

Composed Models

A composed model **combines multiple custom models** into a single model. When a document is submitted for analysis using a composed model, the service performs a **classification** to determine which custom model best represents the document.

Document Intelligence Studio



Online tool for Document Intelligence service.

Learn, test, experiment, models and features with **your own** or sample documents.

Train custom models for classification and extraction.

Get sample code for different languages.



Pricing



Instance	Document type	Price
Free - Web/Container ¹	All	0 - 500 pages free per month
S0 - Web/Container	Read	0-1M pages - €1.348 per 1,000 pages 1M+ pages - €0.540 per 1,000 pages
	All Prebuilt Models: Document, Layout, Receipt, Invoice, ID, W-2, 1098 Tax forms, Health insurance card, Contract.	€8.985 per 1,000 pages
	Custom classification	€2.696 per 1,000 pages
	Custom extraction	€26.955 per 1,000 pages
	Custom generative extraction	€26.955 per 1,000 pages
	Add-On ²	€5.391 per 1,000 pages
	Query Fields	€8.985 per 1,000 pages





Session Feedback Form with Document Intelligence



[massimobonanni/FeedbackExtractor](https://github.com/massimobonanni/FeedbackExtractor)



Azure OpenAI





What is Azure OpenAI?

Azure OpenAI is a cloud-based platform that offers access to various **OpenAI models**.

The service supports the latest OpenAI models such as GPT-4, GPT-4 Turbo with Vision, **GTP-4o**, GPT-3.5-Turbo, and Embeddings.

The service can be accessed through REST APIs, Python/C# SDK, or the web-based interface in the **Azure OpenAI Studio**.



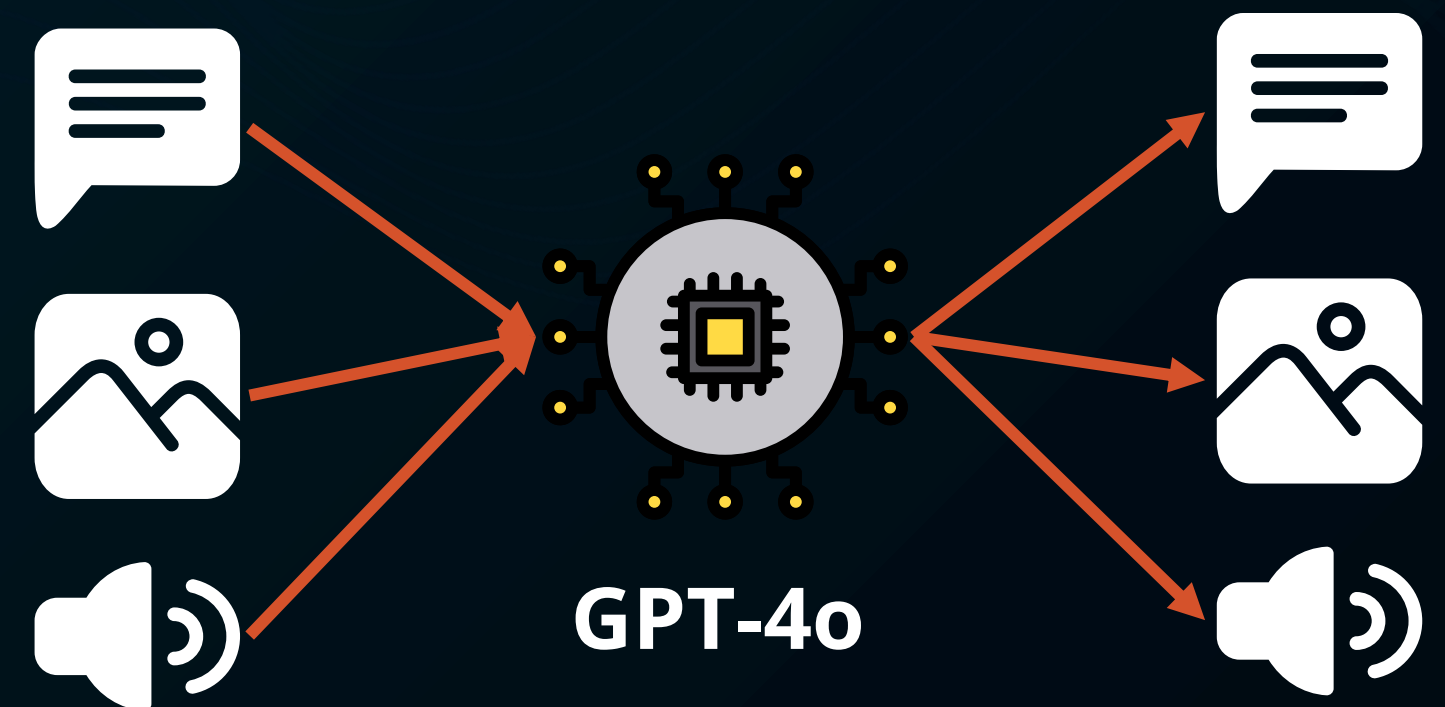
GPT-4o ("o" for "omni")



It accepts as input any combination of **text, audio, image, and video** and generates any combination of **text, audio, and image** outputs.

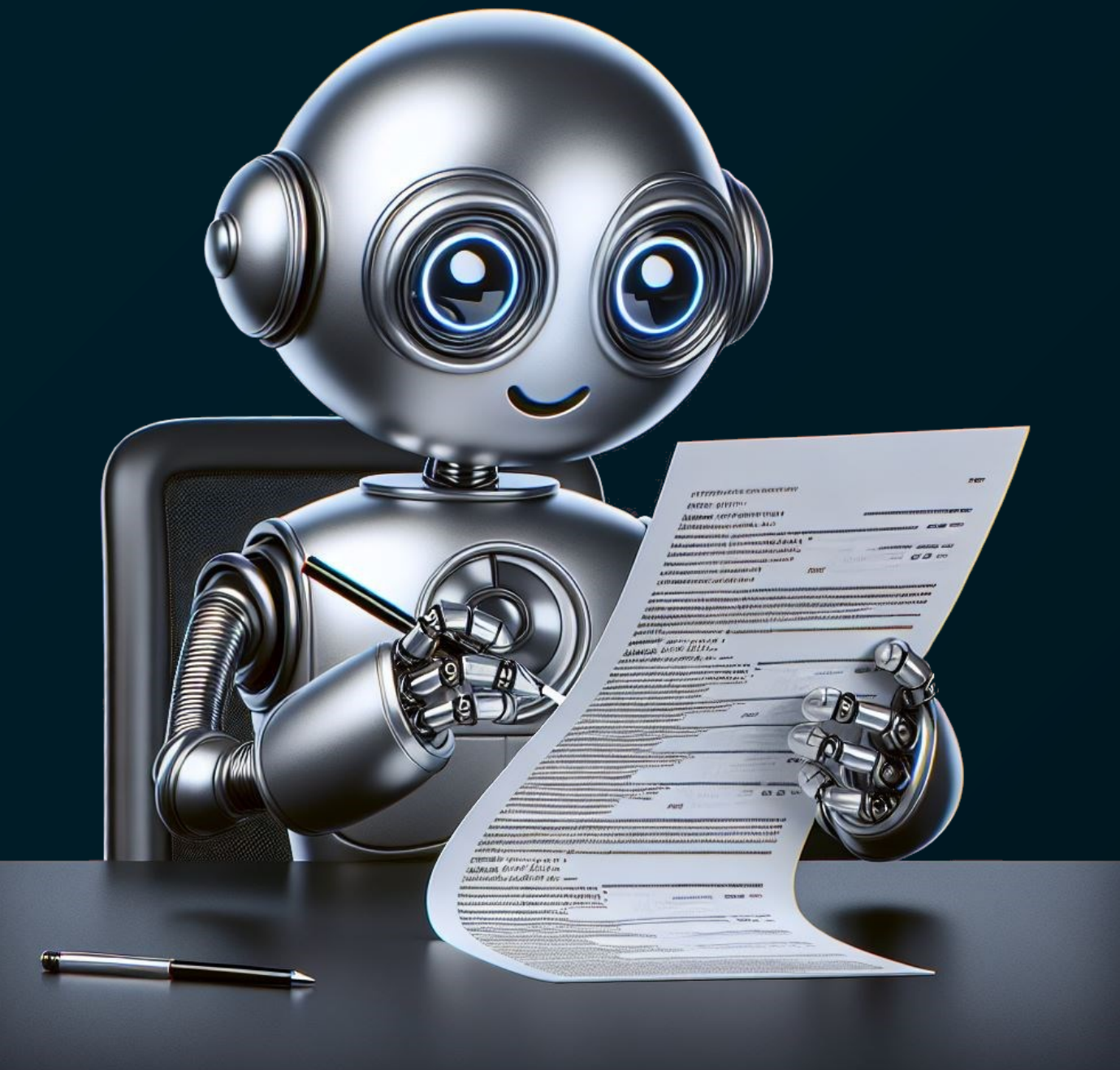
It can respond to audio inputs in as little as **232 milliseconds**, with an average of 320 milliseconds, which is like human response time.

It matches GPT-4 Turbo performance on text in English and code, with significant **improvement on text in non-English languages**.





Extract Data with Prompt Engineering



Prompt Engineering is a technique that leverages **natural language processing** and artificial intelligence to create effective prompts for various tasks and domains.

A **prompt** is a text input that guides an AI model to produce a desired output, such as a summary, a classification, a translation, or a generation.

One of the applications of Prompt Engineering is to **extract data from texts and documents**, such as names, dates, locations, keywords, sentiments, etc.



Extract Data with Prompt Engineering

Session Feedback Form

Event Name: GLOBAL AI BOOTCAMP

Session Code: Session 1

How useful did you think this event was for you?
(please check the appropriate number where 1=not at all useful and 5=extremely useful)

1	2	3	4	5
☐	☐	☐	☒	☐

How would you rate the session?
(please check the appropriate number where 1=not at all useful and 5=extremely useful)

1	2	3	4	5
☐	☐	☐	☐	☐

Do you agree with the feedback?
(please check the appropriate number where 1=not at all useful and 5=extremely useful)

1	2	3	4	5
☐	☐	☐	☐	☐

Comment: Amazing

Analyze the image and extract the information about feedback in the following JSON format:

```
{  
  "EventName": "",  
  "SessionCode": "",  
  "EventQuality": 1,  
  "SessionQuality": 2,  
  "SpeakerQuality": 5,  
  "Comment": ""  
}
```

The EventQuality field is an integer value from 1 to 5 and it is represented by the checked square in the second row of the first table: starting from the left, if the first square is checked then the value is 1; if the second square is checked then the value is 2; if the third square is checked then the value is 3; if the fourth square is checked then the value is 4; If the fifth square is checked then the value is 5. if you don't find any checked squares in the row, set the value 0.

.....
If you are not able to extract data, put the string "<error>" in the field.
If you are not able to extract the entire set of data, return empty string.

```
{
```

In a GPT prompt, you would use:

- **primary** to direct the main focus of the conversation,
- **supporting** to add detail and nuance,
- **grounding** to ensure that the response is well-informed and relevant to the topic.

Primary

Supporting

Grounding

Pricing



Models	Context	Input (Per 1,000 tokens)	Output (Per 1,000 tokens)
			
			
GPT-4o Regional API	128K	€0.0045	€0.0135

The number of tokens you spend in the GTP-4o calls depends on the images, audios and videos size you are using.





Session Feedback Form with GTP-4o and Prompt Engineering



[massimobonanni/FeedbackExtractor](https://github.com/massimobonanni/FeedbackExtractor)



Conclusions



Document Intelligence supports PDF and DOCX files up to 2000 pages and run on containers



GPT-4o can extract data from unstructured documents



You can orchestrate both calling Document Intelligence first and GPT-4o after to refine data.





AI

THANK
YOU

20
24



Massimo Bonanni

Technical Trainer @ Microsoft

massimo.bonanni@microsoft.com



aka.ms/maxlinkedin

NEXT CONFERENCE



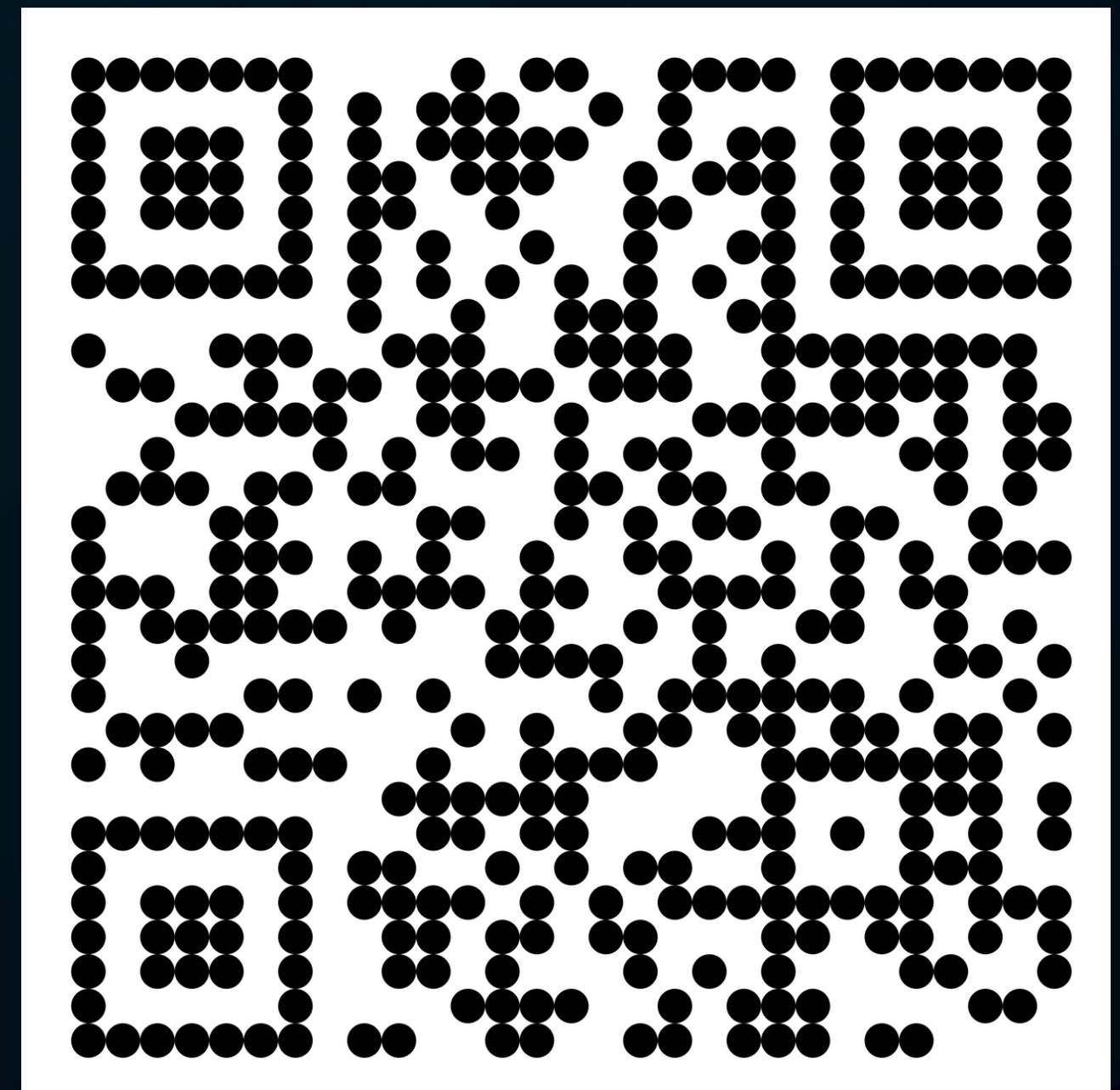
Data Saturday Sofia 2024

05 | October | 2024



Free Ticket

Labs building, Sofia Tech Park





References

[Azure AI Document Intelligence](#)

[Document Intelligence documentation - Quickstarts, Tutorials, API Reference](#)

[Document Intelligence Studio](#)

[Pricing - Azure AI Document Intelligence](#)

[Azure OpenAI Service – Advanced Language Models](#)

[Azure OpenAI Studio](#)

[Azure OpenAI Service – Pricing](#)

[Using Azure AI Document Intelligence and Azure OpenAI to extract structured data from documents](#)