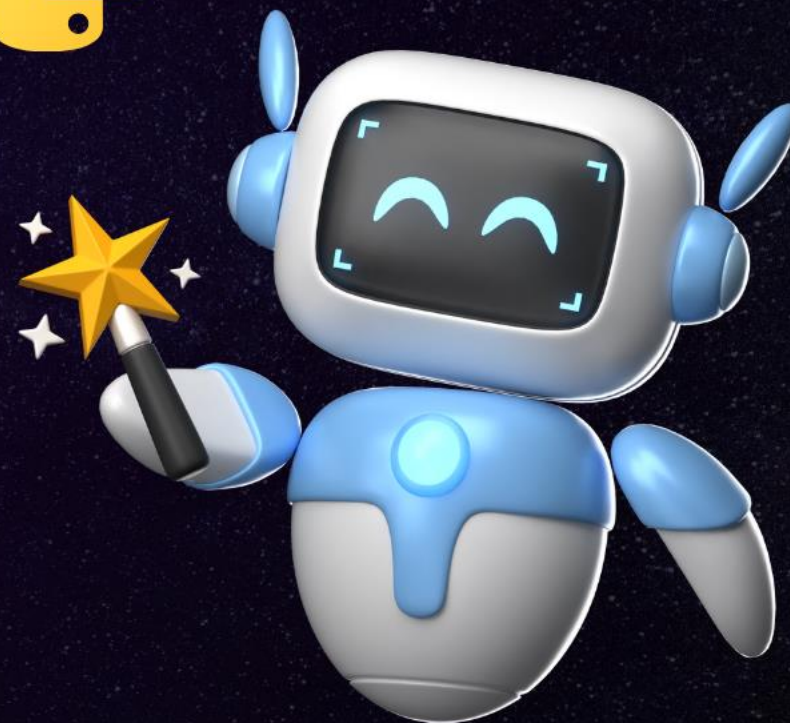






Python + IA

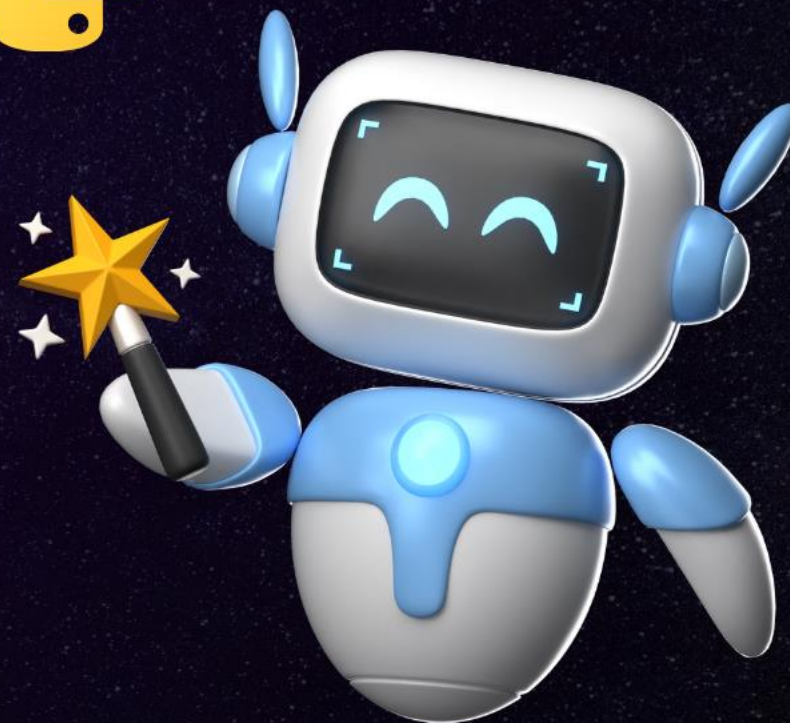


Python + IA

-  3/11: LLMs
- ↩ 3/13: Vector embeddings
-  3/18: RAG
-  3/19: Models de Visión
-  3/25: Salidas Estructuradas
-  3/27: Calidad y Seguridad

Regístrate @ aka.ms/PythonAI/series

Ponte al día @ <http://aka.ms/Pythonia/grabacion>





Python + AI



Modelos de Visión

Gwyneth Peña-Siguenza

Python Cloud Advocate

[Aka.ms/madebygps](https://aka.ms/madebygps)

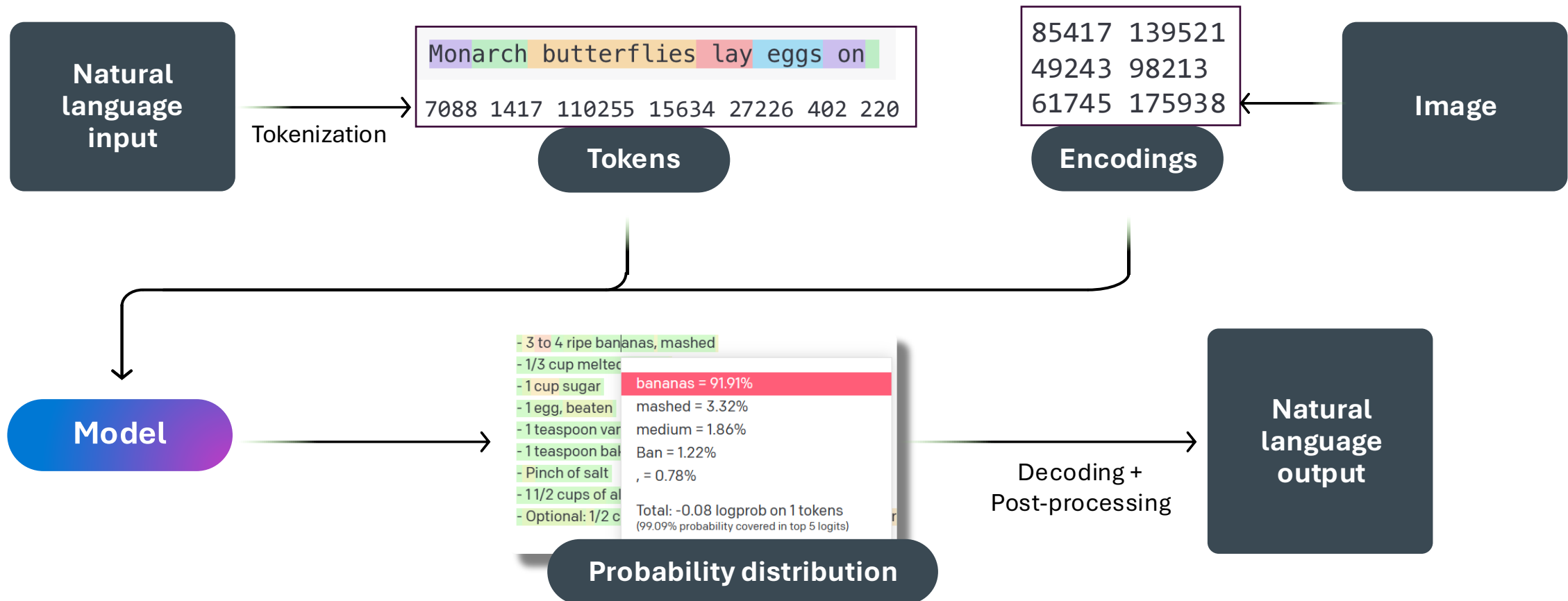
Agenda

- Modelos de lenguaje multimodal
- Casos de uso populares
- Chat con imágenes
- Modelos de embeddings multimodales
- RAG con modelos de visión

LLMs Multimodales

¿Que es un LLM Multimodal?

Además de texto, un LLM multimodal puede aceptar imágenes y, a veces, vídeo/audio.



LLMs Multi Modales en Azure/GitHub

Creator	Models	¿Cómo acceder?
OpenAI	GPT-4o, GPT-4o-mini	Azure OpenAI, GitHub Models
Microsoft	Phi3.5-vision, Phi4-multimodal-instruct	Azure AI Services, GitHub Models
Meta	Llama-3.2-Vision-instruct	Azure AI Services, GitHub Models

<https://azure.microsoft.com/products/ai-services/openai-service>

<https://github.com/marketplace/models>

<https://ai.azure.com/>

Enviando imágenes con el SDK de OpenAI en Python

```
messages = [{
    "role": "user",
    "content": [
        {"type": "text",
         "text": "Who is pictured in this thumbnail?"},
        {"type": "image_url",
         "image_url": {"url": "https://i.ytimg.com/vi/toR644E--w8/hq720.jpg"}}
    ],
}
]
response = openai_client.chat.completions.create(
    model="gpt-4o",
    messages=messages
)
```

<https://aka.ms/chat-vision-app> : notebooks/chat_vision.ipynb

Enviando imágenes con URI codificado en base64

```
def open_image_as_base64(filename):  
    with open(filename, "rb") as image_file:  
        image_data = image_file.read()  
        image_base64 = base64.b64encode(image_data).decode("utf-8")  
        return f"data:image/png;base64,{image_base64}"  
  
messages = [{  
    "role": "user",  
    "content": [  
        {"type": "text",  
         "text": "What animal is pictured?."},  
        },  
        {"type": "image_url",  
         "image_url": {"url": open_image_as_base64("ur.png")}}  
    ],  
}]
```

<https://aka.ms/chat-vision-app> : notebooks/chat_vision.ipynb

Casos de uso populares

Accesibilidad

- Sugerir texto alternativo para imágenes
- Ejemplo: Revisión de accesibilidad para PowerPoint
- Brindar ayuda a usuarios con problemas de vision
- Ejemplo: App móvil "Be My Eyes"

Procesos empresariales más eficientes

- Reclamos de seguros: Identificar reclamos sospechosos
- Análisis de datos: Generar insights basados en gráficos y tablas
- Atención al cliente: Responder preguntas sobre imágenes de productos

<https://aka.ms/chat-vision-app> : notebooks/chat_vision.ipynb

Enviando documentos que (aún) no son imágenes

La mayoría de los modelos de visión solo pueden manejar archivos JPEG, PNG y GIF estáticos, pero pueda que tengas visuales en documentos que no son imágenes, como los PDFs.

Approaches:

1. Convertir todo el documento en una imagen
2. Identificar la parte visual del documento, recortarla y guardarla como imagen: <https://aka.ms/document-intelligence-figure-extraction>

Puedes usar tanto paquetes de Python como servicios alojados en Azure.

Convirtiendo PDFs a imágenes

```
import pymupdf
from PIL import Image

filename = "plants.pdf"
doc = pymupdf.open(filename)
for i in range(doc.page_count):
    doc = pymupdf.open(filename)
    page = doc.load_page(i)
    pix = page.get_pixmap()
    original_img = Image.frombytes("RGB",
    [pix.width, pix.height], pix.samples)
    original_img.save(f"page_{i}.png")
```



PDF



pymupdf

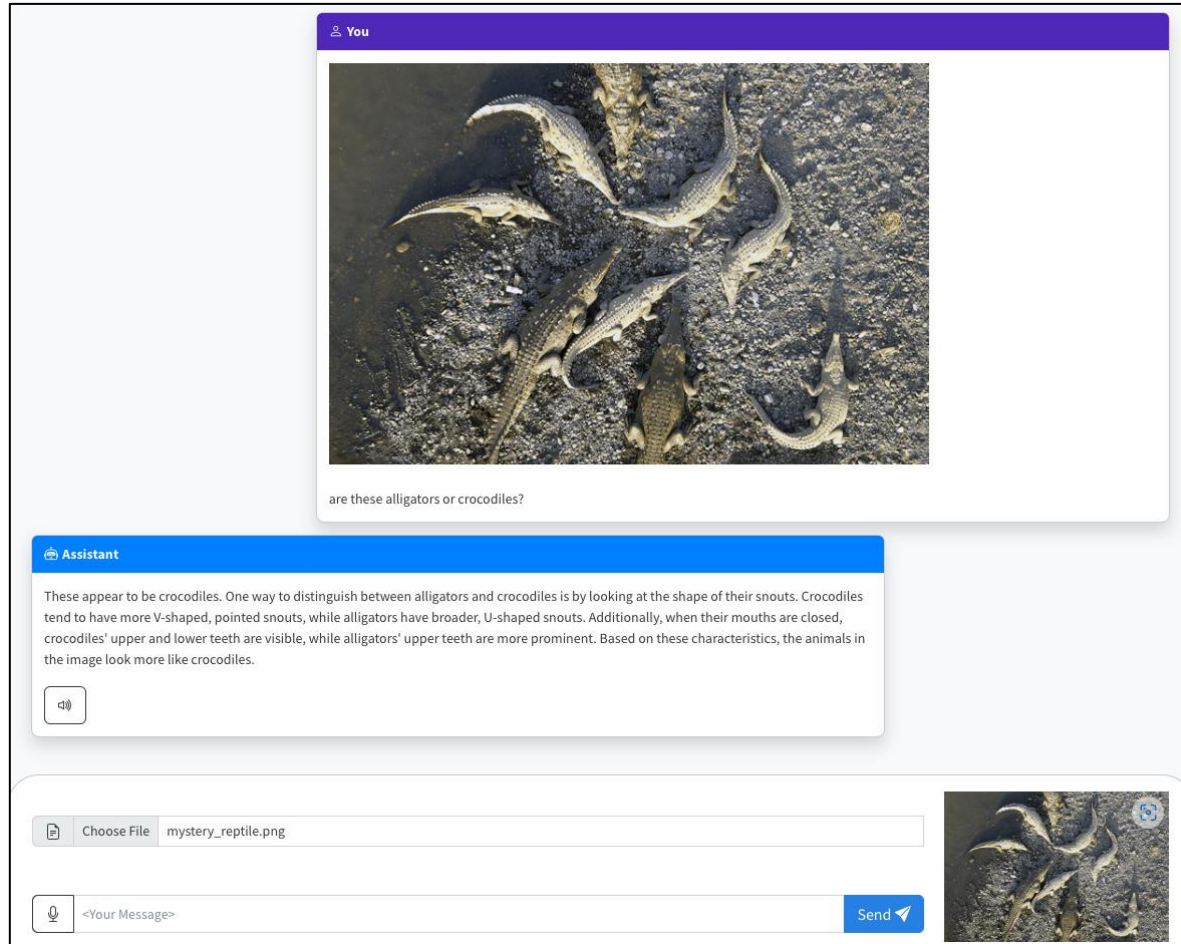


pillow

<https://aka.ms/chat-vision-app> : notebooks/chat_pdf_images.ipynb

Construyendo apps con modelos de visión

Open-source template: chat con visión



Azure OpenAI con gpt-4o
Python backend (Quart)

Repo:

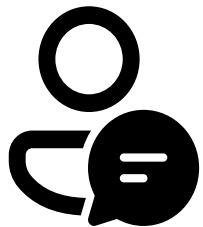
<https://aka.ms/chat-vision-app>

Demo:

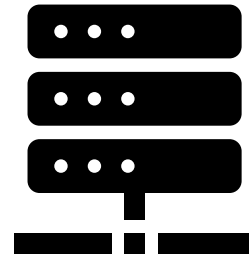
<https://aka.ms/chat-vision-app/demo>

Chat con visión: Flow

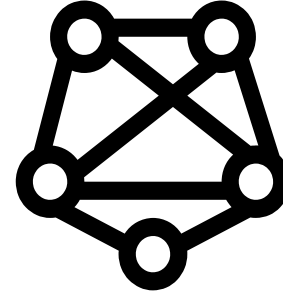
¿Caimanes o cocodrilos?



User
Question



Backend

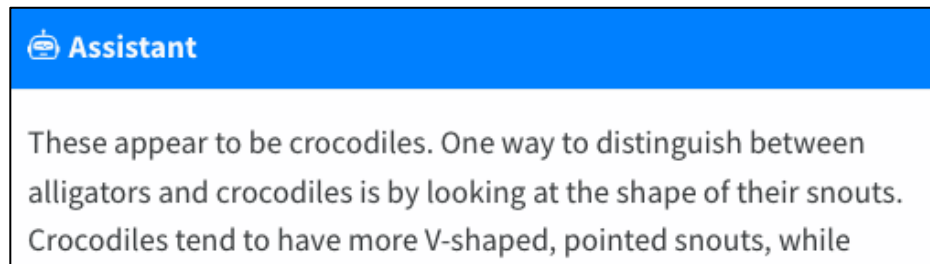


LLM

Esos parecen ser
cocodrilos, por sus
hocicos en forma de V.

Chat con visión: App architecture

Frontend
(HTML, JavaScript)



Python backend
(Quart, Uvicorn)

```
@bp.post("/chat/stream")  
async def chat_handler()
```

base64 image +
user question

Model

Streamed response

```
Transfer-Encoding: Chunked  
{ "content": "He"  
{ "content": "llo"  
{ "content": "It's"  
{ "content": "me"
```


Codificando imágenes en el frontend (simplificado)

```
const toBase64 = file => new Promise((resolve, reject) => {
  const reader = new FileReader();
  reader.readAsDataURL(file);
  reader.onload = () => resolve(reader.result);
  reader.onerror = reject;
});

form.addEventListener("submit", async function(e) {
  const file = document.getElementById("file").files[0];
  const fileData = file ? await toBase64(file) : null;

  // Get all messages and send with file to backend
  const result = await client.getStreamedCompletion(messages, {
    context: {file: fileData, file_name: file ? file.name : null}
  });

  // ...
});
```

<https://aka.ms/chat-vision-app> : src/quartapp/templates/index.html

Manejando imágenes en el backend (simplificado)

```
@bp.post("/chat/stream")
async def chat_handler():
    request_json = await request.get_json()
    request_messages = request_json["messages"]
    image = request_json["context"]["file"]
    all_messages = request_messages[0:-1]
    if image:
        all_messages.append(
            {"role": "user",
             "content": [
                 {"text": request_messages[-1]["content"], "type": "text"},
                 {"image_url": {"url": image, "detail": "auto"}, "type": "image_url"}]])
    else:
        all_messages.append(request_messages[-1])

chat_coroutine = bp.openai_client.chat.completions.create(
    model=os.environ["OPENAI_MODEL"], messages=all_messages)
```

<https://aka.ms/chat-vision-app> : src/quartapp/chat.py

Otras formas de manejar el upload de imágenes

- Envía una solicitud POST desde el frontend usando multipart form data <https://aka.ms/chat-vision-multipart>
- Carga imágenes por separado:
 1. Almacena la imagen en un almacenamiento de archivos (por ejemplo, Azure Blob Storage)
 2. Envía el identificador del archivo al frontend
 3. El frontend envía el identificador del archivo de vuelta al backend

⚠ ¡Asegúrate de que otros usuarios no puedan acceder a los archivos de los demás!

Modelos de embeddings multimodales

Azure AI Vision: Multimodal embeddings API

Usa la API de Azure AI Vision para generar embeddings con el modelo Florence.



/vectorizeImage



```
[  
  3.3652344,  
  0.8413086,  
  1.2783203,  
  ...],
```

"a beach-themed tealight
candle holder"



/vectorizeText



```
[  
 -0.027022313,  
 -0.011945606,  
  0.019690325,  
  ...],
```

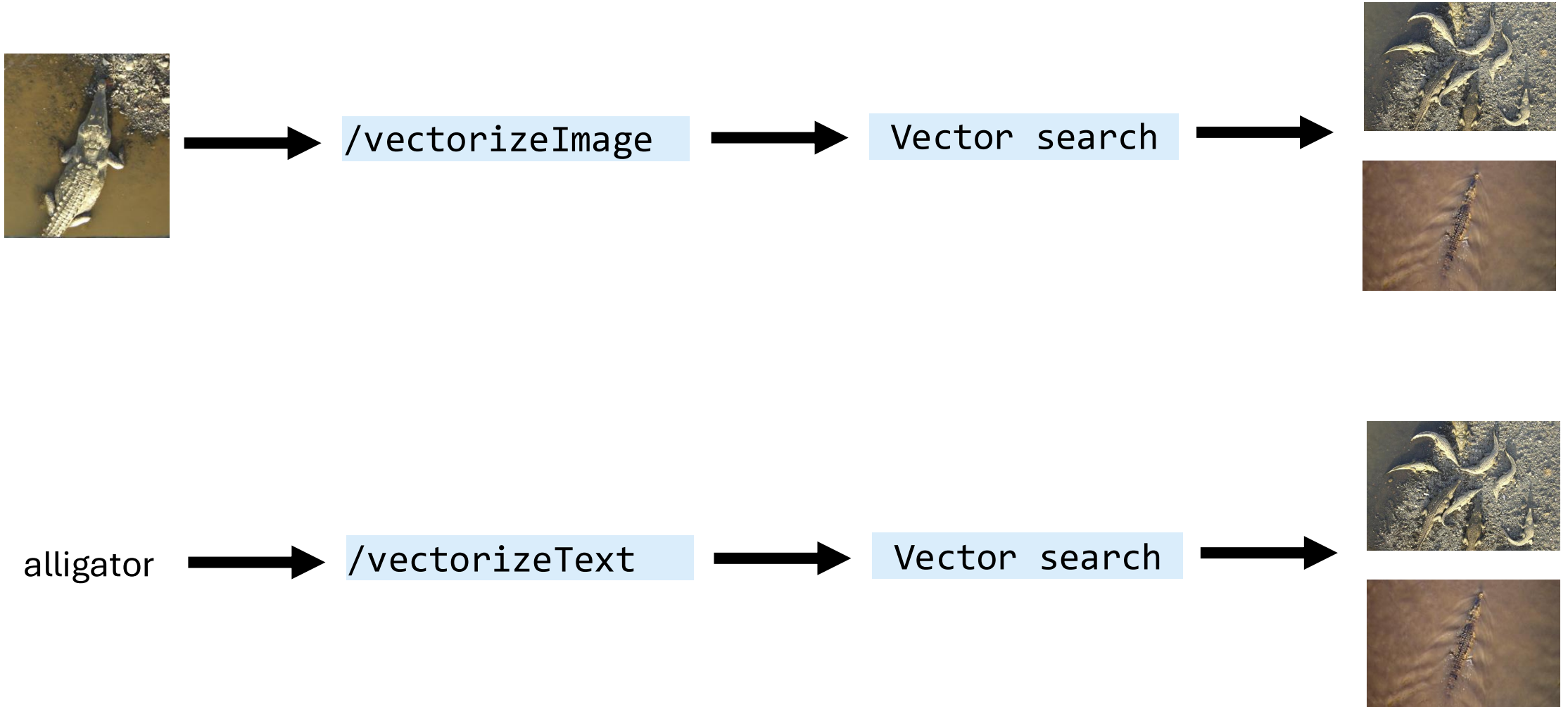
Azure AI Vision: Usando la API con Python

```
def get_image_embedding(image_file):
    mimetype = mimetypes.guess_type(image_file)[0]
    url = f"{AZURE_AI_VISION_URL}:vectorizeImage"
    headers = get_auth_headers()
    headers["Content-Type"] = mimetype
    response = requests.post(url, headers=headers,
                             params=get_model_params(), data=open(image_file, "rb"))
    return response.json()["vector"]

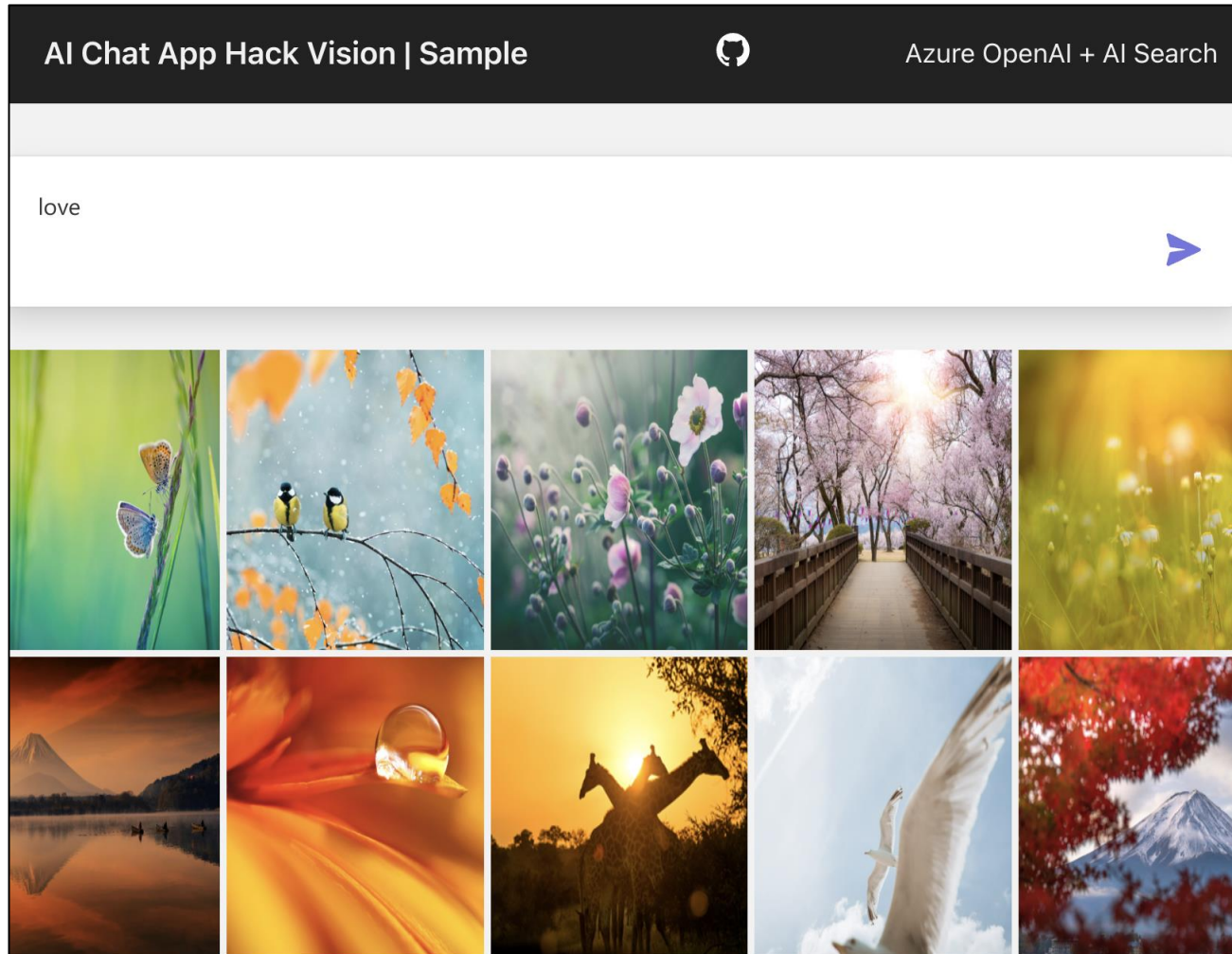
def get_text_embedding(text):
    url = f"{AZURE_AI_VISION_URL}:vectorizeText"
    response = requests.post(url, headers=get_auth_headers(),
                             params=get_model_params(), json={"text": text})
    return response.json()["vector"]
```

Notebook: [multimodal_vectors.ipynb](#)

Vector search con multimodal embeddings



Open source template: Image search



Azure AI Vision +
Azure AI Search

Code:

<https://aka.ms/aisherech-images-app>

Demo:

<https://aka.ms/aisherech-images-app/dem>

RAG con modelos de visión

Open-source template: RAG con vision support

Can you identify any correlation between oil prices and stock market trends?

Yes, there is a correlation between oil prices and stock market trends. Specifically, the correlation between oil prices and the S&P 500 is 0.6, and the correlation between oil prices and the NASDAQ is 0.65. This indicates a moderate positive correlation, meaning that as oil prices increase, stock indices like the S&P 500 and NASDAQ also tend to increase, and vice versa ¹.

Citations: 1. Financial Market Analysis Report 2023-6.png

SourceFileName: Financial Market Analysis Report 2023-6.png

Interplay Between Different Financial Markets

Correlation Between Stock Indices, Cryptocurrencies, and Oil

	S&P 500	NASDAQ	Bitcoin
S&P 500	1		
NASDAQ	0.95	1	
Bitcoin	0.3	0.4	1
Ethereum	0.35	0.45	0.9
Oil	0.6	0.65	0.2
Gold	-0.2	-0.15	-0.1

Financial markets are interconnected, with movements in one market often influencing others. This section examines the correlations between stock indices, cryptocurrencies, and oil, revealing how changes in one market can have ripple effects across others.

SourceFileName: Financial Market Analysis Report 2023-4.png

Cryptocurrency Market Price Fluctuations

Price Fluctuations of Bitcoin and Ethereum (Last 12 Months)

Azure OpenAI +
Azure AI Search +
Azure AI Vision

Main repo:

<https://aka.ms/ragchat>

Setup guide:

<https://aka.ms/ragchat/vision>

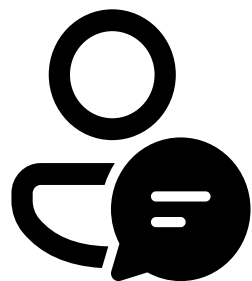
Demo:

<https://aka.ms/ragchat/vision/demo>

Enable "GPT vision" in Settings

RAG con visión: Flow

¿Hay alguna correlación entre los precios del petróleo y las tendencias del mercado de valores?



User Question

“Hay alguna...”

OpenAI
text embedding

“Hay alguna...”

[[0.0014615238, -
0.015594152, -
0.0072768144, -
0.012787478,...]]



Azure AI
Search

“Hay alguna...”

AI vision API
/vectorizeText

[[0.0021338, -
0.01123152, -
0.0238144, -
0.0123478,...]]



“Hay alguna...”



OpenAI
gpt-4o

Sí, hay una correlación entre los precios del petróleo y las tendencias del mercado de valores.

Financial Market Analysis 2023-6.png
Esta sección examina las correlaciones entre los índices bursátiles, los precios de las criptomonedas...



Interplay between Different Market Segments

Correlation between Stock Indices, Cryptocurrency Prices, and Commodity Prices

	SP500	NASDAQ	Bitcoin	Ethereum	Oil	Gold
S&P 500	1.00	0.95	0.12	0.15	-0.05	0.08
NASDAQ	0.95	1.00	0.15	0.20	-0.08	0.10
Bitcoin	0.12	0.15	1.00	0.85	0.02	-0.01
Ethereum	0.15	0.20	0.85	1.00	0.03	-0.02
Oil	-0.05	-0.08	0.02	0.03	1.00	0.15
Gold	0.08	0.10	-0.01	-0.02	0.15	1.00

Financial markets are interconnected, with movements in one segment often influencing others. This table examines the correlation between stock indices, cryptocurrency prices, and commodity prices, revealing how changes in one market can ripple across the financial ecosystem.

Ingesta de datos con modelos de visión

Estos pasos se hacen además de los pasos habituales de ingestión de texto.



Python

Dividir PDFs en
páginas
Y

Generar imagen
de cada página



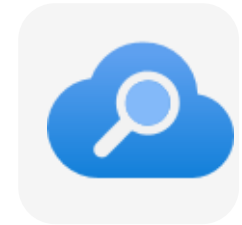
Azure Blob
Storage

Upload
imágenes
Necesario para
renderizar las
citas en la
interfaz de
usuario de la
app.



Azure AI
Vision

Vectorizar
imágenes



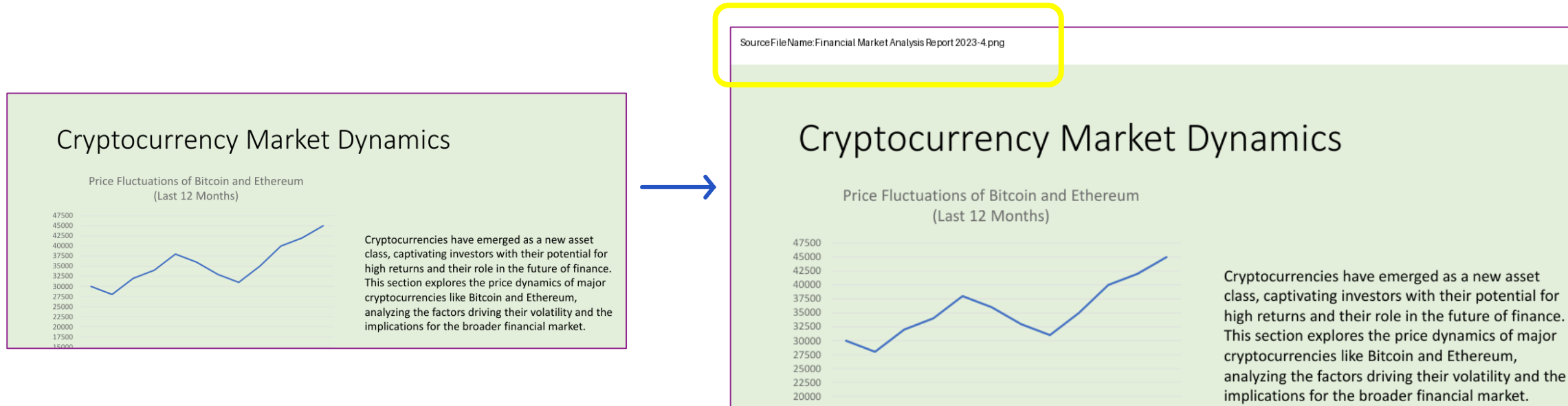
Azure
AI Search

Indexar imágenes

- Guarda el embedding en el campo imageEmbedding
- Guarda el nombre del archivo de la imagen en el campo sourcepage

Ejemplo de ingestión: Generar imágenes de páginas

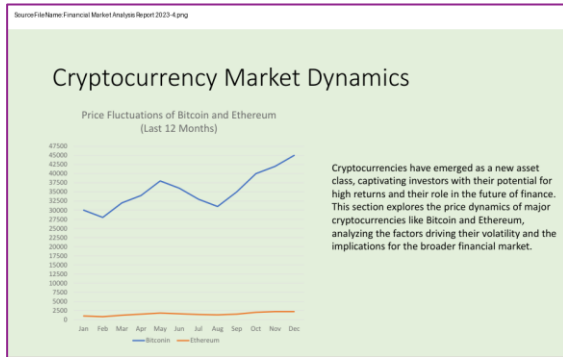
Para que el modelo dé una respuesta con citas, debemos incorporar el nombre del archivo en la imagen:



```
draw = ImageDraw.Draw(new_img)
text = f"SourceFileName:{blob_name}"
draw.text((10, 10), text)
```

<https://aka.ms/ragchat>: blobmanager.py

Ejemplo de ingestion: Almacenar imágenes de páginas



Azure Blob Storage



<https://stfvid7hrxoifmi.blob.core.windows.net/content/Keystone-Plant-Signs-Sunflower-8.5x11-1.png?st=2025-01-22T21%3A30%3A27Z&se=2025-01-23T21%3A59%3A59Z&sp=r&sv=2024-08-04&sr=b&skoid=c589476e-841sdfs7-4bb9-bc33-15d1eb1c2503&sktid=c37ef95c-b0dsfcd-4200-a919-2e04c2f8c2cfdsd>

```
output = io.BytesIO()
new_img.save(output, format="PNG")
output.seek(0)
```

```
container_client.upload_blob(
    blob_name, output)
```

<https://aka.ms/ragchat>: blobmanager.py

Ejemplo de ingestión: Vectorizar imágenes

<https://stfvid7hrxoifmi.blob.core.windows.net/content/Keystone-Plant-Signs-Sunflower-8.5x11-1.png?st=2025-01-22T21%3A30%3A27Z&se=2025-01-23T21%3A59%3A59Z&sp=r&sv=2024-08-04&sr=b&skoid=c589476e-841sdfs7-4bb9-bc33-15d1eb1c2503&sktid=c37ef95c-b0dsfcd-4200-a919-2e04c2f8c2cfsd>



Azure AI
Vision

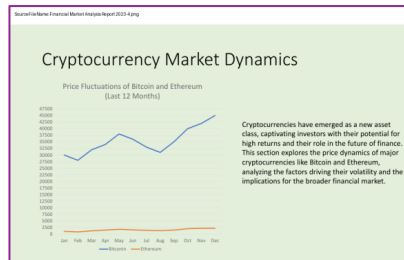


[-3.203125,
1.5576172
...+1022 more]

```
endpoint = urljoin(self.endpoint, "computervision/retrieval:vectorizeImage")
embeddings: List[List[float]] = []
async with aiohttp.ClientSession(headers=headers) as session:
    for blob_url in blob_urls:
        body = {"url": blob_url}
        async with session.post(url=endpoint, params=params, json=body) as resp:
            resp_json = await resp.json()
            embeddings.append(resp_json["vector"])
```

<https://aka.ms/ragchat>: embeddings.py

Ejemplo de ingestión: Indexar imagen & texto

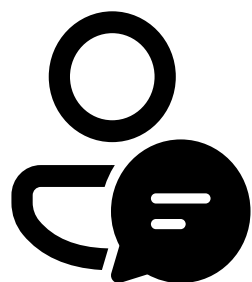


```
{
  "id": "file-Financial_Market_Analysis_Report_2023_pdf-46696E616E6369616C204D61726B657420416E616C79736973205265706F727420323032332E706466-page-2",
  "content": "Cryptocurrency Market Dynamics\nPrice Fluctuations of Bitcoin and Ethereum (Last 12 Months)\n47500\n45000\n42500\n40000\n37500\n35000\n32500\n30000\n27500\n25000\n22500\n20000\n17500\n15000\n12500\n10000\n7500\n5000\n2500\n0\nJan\nFeb\nMar\nApr\nMay\nJun\nJul\nAug\nSep\nOct\nNov\nDec\n-Bitcoin -Ethereum\nCryptocurrencies have emerged as a new asset class, captivating investors with their potential for high returns and their role in the future of finance. This section explores the price dynamics of major cryptocurrencies like Bitcoin and Ethereum...",
  "embedding": "[-0.009356536, -0.035459142 ...+1534 more]",
  "imageEmbedding": "[-0.94384766, 1.5185547 ...+1022 more]",
  "sourcepage": "Financial Market Analysis Report 2023-4.png",
  "sourcefile": "Financial Market Analysis Report 2023.pdf"
}
```

<https://aka.ms/ragchat>: searchmanager.py

RAG con visión: Flow (revisitado)

¿Hay alguna correlación entre los precios del petróleo y las tendencias del mercado de valores?



User
Question

“Hay alguna...”



OpenAI
text embedding

“Hay alguna...”

[[0.0014615238, -
0.015594152, -
0.0072768144, -
0.012787478,...]



Azure AI
Search

“Hay alguna...”

**Financial Market
Analysis 2023-6.png**
This section examines
the correlations
between stock indices,
cryptocurrency prices...



OpenAI
gpt-4o

Sí, hay una correlación entre los precios del petróleo y las tendencias del mercado de valores.

“Hay alguna...”



AI vision API
/vectorizeText

[[0.0021338, -
0.01123152, -
0.0238144, -
0.0123478,...]



Interplay Between Different Market Segments

Correlation Between Stock Indices, Cryptocurrency Prices, and Commodity Prices

Asset Class	Index	Correlation	Volatility	Return	Risk
Stocks	S&P 500	0.75	0.15	0.10	0.10
Cryptocurrency	Bitcoin	0.60	0.20	0.15	0.15
Commodities	Oil	0.50	0.10	0.05	0.05

RAG con visión: System prompt

You are an intelligent assistant helping analyze the Annual Financial Report of Contoso Ltd. The documents contain text, graphs, tables and images.

Each image source has the file name in the top left corner of the image with coordinates (10,10) pixels and is in the format SourceFileName:<file_name>

Each text source starts in a new line and has the file name followed by colon and the actual information

Always include the source name from the image or text for each fact you use in the response in the format: [filename]

Answer the following question using only the data provided in the sources below.

If asking a clarifying question to the user would help, ask the question.

Be brief in your answers.

The text and image source can be the same file name, don't use the image title when citing the image source, only use the file name as mentioned.

If you cannot answer using the sources below, say you don't know. Return just the answer without any input texts.

RAG con visión: System prompt

Eres un asistente inteligente que ayuda a analizar el Informe Financiero Anual de Contoso Ltd.
Los documentos contienen texto, gráficos, tablas e imágenes.

Cada fuente de imagen tiene el nombre del archivo en la esquina superior izquierda de la imagen con coordenadas (10,10) píxeles y está en el formato:
SourceFileName:<nombre_de_archivo>

Cada fuente de texto comienza en una nueva línea y muestra el nombre del archivo seguido de dos puntos, y luego la información correspondiente.

Incluye siempre el nombre de la fuente de la imagen o el texto para cada dato que uses en tu respuesta, con el formato: [nombre_de_archivo].

Responde la pregunta usando únicamente la información proporcionada en las fuentes de abajo.
Si crees que una pregunta de aclaración al usuario ayudaría, hazla.

Sé breve en tus respuestas.

El texto y la imagen pueden tener el mismo nombre de archivo; no uses el título de la imagen cuando cites la fuente, solo el nombre de archivo tal como se menciona.

Si no puedes responder con la información de las fuentes proporcionadas, di que no sabes. Devuelve solo la respuesta, sin ningún texto adicional de entrada.

RAG con visión: Chat completion messages

Envía un mensaje de usuario con múltiples partes que incluya tanto texto como imágenes:

```
{ "role": "user",  
  "content": [  
    { "type": "text", "text": "how have bitcoin and ethereum been fluctuating?" },  
    { "type": "image_url",  
      "image_url": { "url": "data:image/png;base64,iVBORw0KGgoAAAANSUhEUgA..." } },  
    { "type": "image_url",  
      "image_url": { "url": "data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAA..." } },  
    { "type": "image_url",  
      "image_url": { "url": "data:image/png;base64,iVBORw0KGgoAAAAA8AAAAJECAI..." } },  
    { "type": "text",  
      "text": "Sources:\n\nFinancial Market Analysis Report 2023-3.png: Commodities  
The global financial market is a vast and intricate network of exchanges,  
instruments, and assets, ranging from traditional stocks and bonds to modern  
cryptocurrencies and commodities.... "  
  ]  
}
```



RAG con visión: Consideraciones

Beneficios

- El paso de búsqueda va a encontrar todo lo que coincida semánticamente con el texto o las imágenes del documento.
- El modelo va a tener acceso a la imagen completa en tiempo de inferencia, así que puede usar detalles de la imagen que no estén en el texto.

Desventajas

- Mayor latencia y costo durante el flujo RAG (búsqueda vectorial adicional, más tokens).
- Limita tu elección de modelo solo a aquellos con soporte multimodal.

-  Aprende más
- Lee el blog de Pamela sobre el proceso: <https://aka.ms/ragchat/vision/blog>
- Mira la charla de Pamela sobre RAG multimedia: <https://aka.ms/ragdeepdive/watch>

Próximos pasos

¡próximos streams! →

Ven a las horas de oficina:

[aka.ms/pythonia/hor](https://aka.ms/pythonia/horario)

AI Agents Hackathon:

aka.ms/agentshack/es

Más recursos de Python AI

aka.ms/thesource/Python_AI



3/11: LLMs



3/13: Vector embeddings



3/18: RAG



3/19: Models de Vision



3/25: Salidas estructuradas



3/27: Calidad y Seguridad

Regístrate @ aka.ms/PythonIA/series

The background is a light purple gradient. It is decorated with stylized, darker purple clouds at the top and bottom edges. In the corners, there are four-pointed starburst shapes in a darker purple with yellow highlights.

Thank you!