



# Deep Dive – Knowledge in Copilot Studio

**Julie Koesmarno & Noa Ghersin**

Principal Product Manager Leads

Microsoft

[julie.Koesmarno@microsoft.com](mailto:julie.Koesmarno@microsoft.com) , [nghersin@microsoft.com](mailto:nghersin@microsoft.com)

<https://aka.ms/juliek>

Sep 20, 2024

Power Platform Community Conference



# Microsoft Copilot Studio

## Your agent, your way

Copilot Studio is an end-to-end conversational AI product for both building your own agents or and extending Microsoft Copilot with your own enterprise data, APIs, and **knowledge sources**.

The screenshot shows the Microsoft Copilot Studio web interface. The left sidebar has tabs for Home, Create, Copilots, and Library, with 'Create' currently selected. The main area has a heading 'Describe your copilot to create it' with a text input placeholder 'Use everyday words to describe what your copilot should do ...'. Below this is a note 'This AI-powered feature is in preview. See terms.' A 'Recent' section lists two copilots: 'Copilot for Microsoft 365' and 'Copilot for Sales', both created by Mona Kane and Pedro Armijo respectively, with publication dates ranging from 2 weeks ago to 2 months ago. A 'Start with a template' section shows eight pre-built copilot templates: Approval Manager, Job Craft, Organization Navigator, Safe Travels, Store Operations, Sustainability Insights, Weather, and Website Q&A. A 'Learning resources' section at the bottom includes links for Quick start: Create and deploy a copilot, Documentation, Responsible AI FAQs, Extend your copilots with plugins and extension, Quick start: Use Generative AI in a copilot, Support community, Try Copilot Agents, and Security and governance in Copilot Studio.

The background features a large, semi-transparent magenta circle centered on the right side of the slide. Overlaid on the left side of this circle is a smaller, semi-transparent blue circle, creating an overlapping effect.

Organizations run on knowledge.  
Agents do, too.

# Agenda

- 1 Knowledge: Why & what
- 2 Let's build a knowledgeable Agent with:
  - Out of box Knowledge
  - Custom Data
  - Bring your own indexed data
- 3 Knowledge quality
- 4 Recap

# Knowledge is the lifeblood of organizations but is frequently untapped

9 ZB+

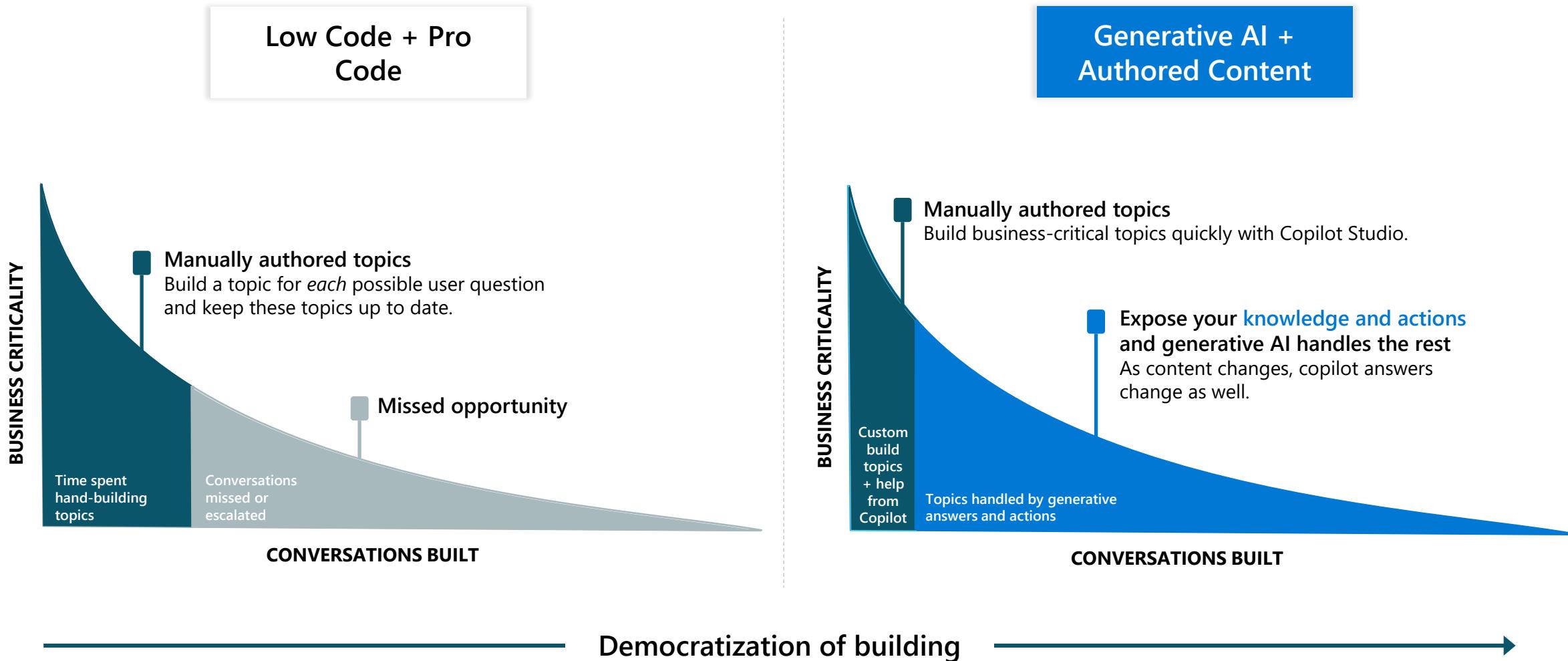
68%

enterprise Data goes unused

1000's

systems where knowledge is captured in an org

# Knowledge + Generative AI is changing conversational AI



# 35,000+ companies use Copilot Studio



1.7  
million+

conversations per month,  
customer Agent helping  
to find the right products



80,000

calls per month supported  
by chat solutions, reduced  
escalation to human  
agent by 50%



40%

cost reduction  
deployed in weeks  
projected for HR/IT  
Agent for employees

Web, SharePoint, Custom  
Data (ServiceNow)



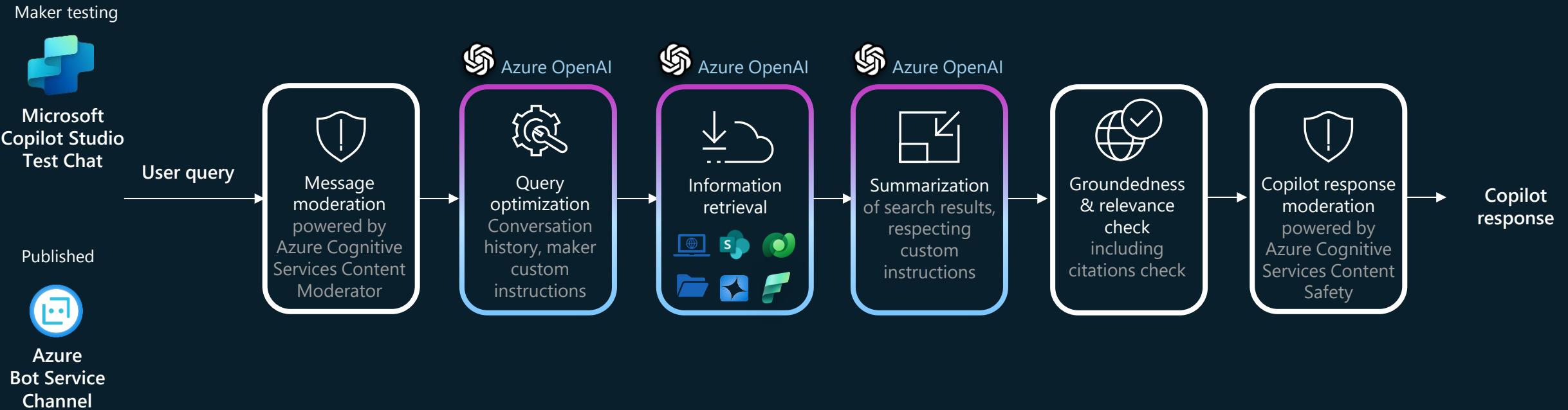
25-40%

internal IT helpdesk  
calls managed by  
employee Agent



Let's look at Knowledge & Generative  
Answers under the hood.

# Knowledge & Generative Answers architecture



# Governing agent Knowledge with data loss prevention policies

Organizational data is the most important asset administrators are responsible for safeguarding.

**Data Loss Prevention** (DLP) policies allow admins to control which Knowledge is used by agents, to which channels agents can be published, and more.

Knowledge DLPs include ability to block publishing of agents configured with:

- **SharePoint**
- **public websites**
- **documents**

The screenshot shows the 'Power Platform admin center' interface for creating a new DLP policy. The left sidebar includes links for Home, Environments, Advisor, Analytics, Billing, Resources, Help + support, Data integration, Data (preview), Policies (selected), Data policies (selected), Billing policies, Admin centers, and a copilot section. The main content area is titled 'DLP Policies > New Policy' and displays a tree structure for policy creation: Policy name, Environments, Connectors (selected), and Review. A note states: 'You can now apply Data Loss Prevention policies on custom connectors. Please ensure that they are classified in the appropriate group.' Below this is a 'Assign connectors' section with tabs for Business (0), Non-business (1657) | Default (selected), Blocked (0), and Exempt (38). The Non-business tab lists various connectors, each with columns for Name, Blockable, Endpoint config., Class, Publisher, and About. The connectors listed are: Studio, Direct Line channels in Copilot Studio, Facebook channel in Copilot Studio, Omnichannel in Copilot Studio, Knowledge source with SharePoint and OneDrive in Copilot Studio, Knowledge source with public websites and data in Copilot Studio, Knowledge source with documents in Copilot Studio, and Application Insights in Copilot Studio. At the bottom are 'Back' and 'Next' buttons.

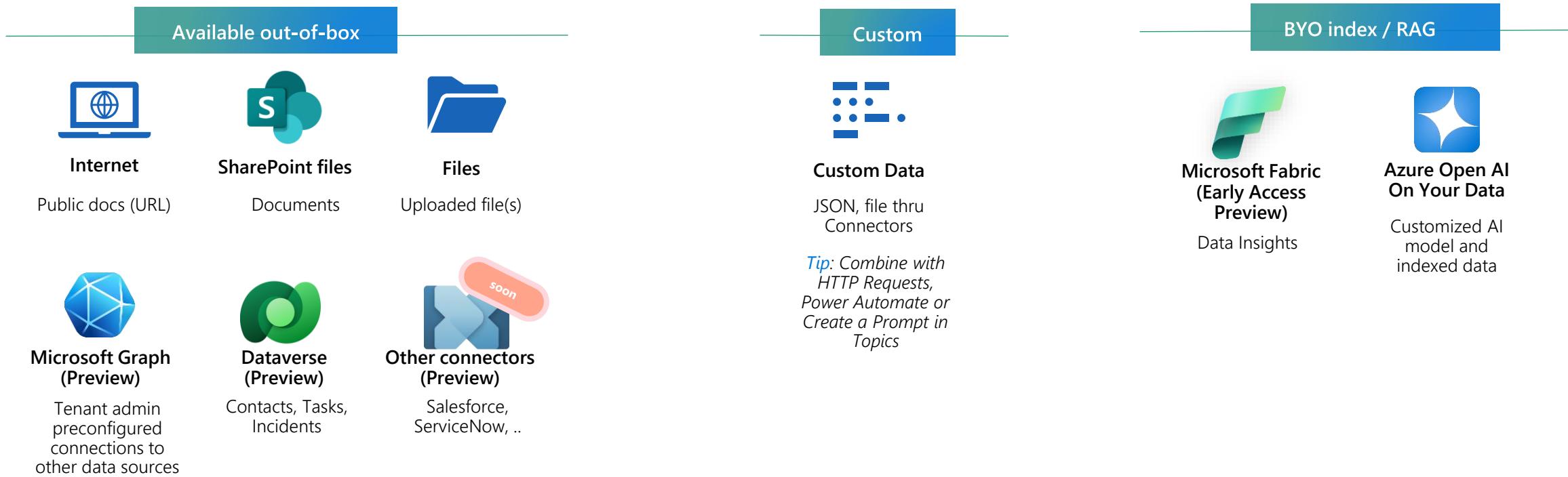
Define which **Graph Connectors** can be added to agents in Microsoft Admin Center

Let's build a knowledgeable Agent.

# Many ways to ground your Agent with Knowledge

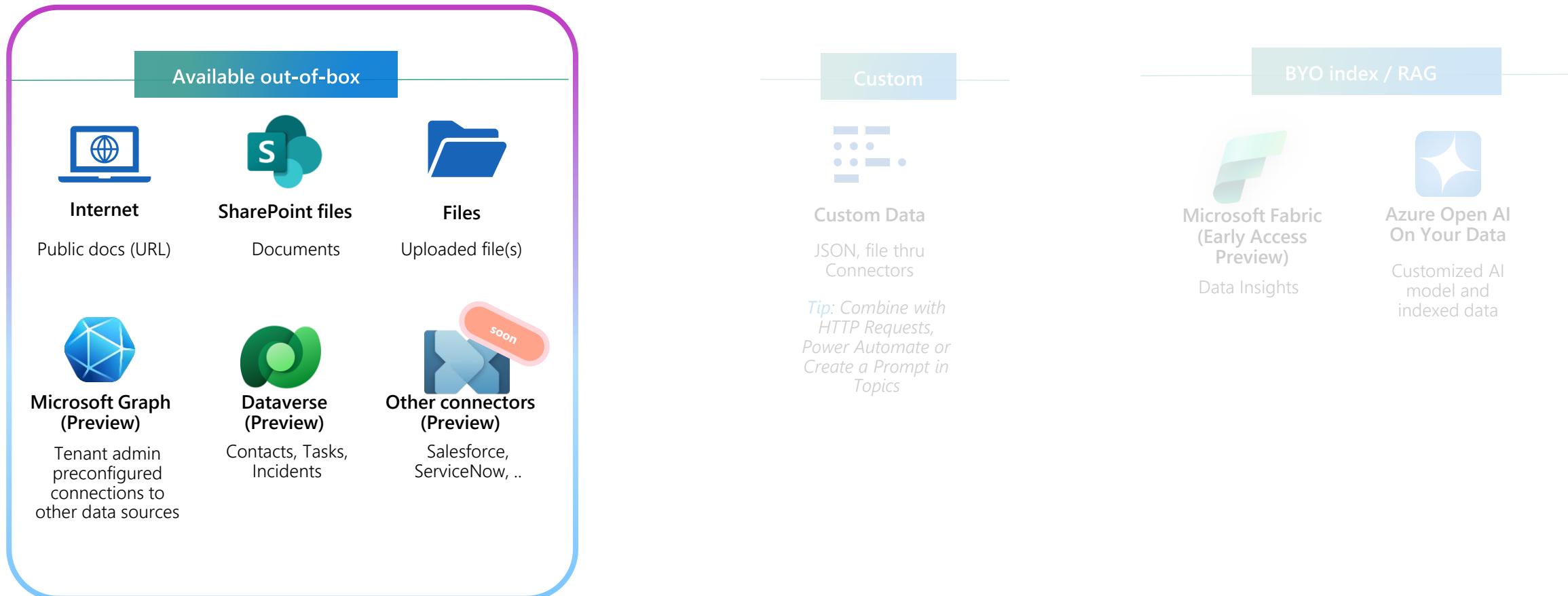
Faster, relevant enterprise knowledge sources

E2E experience for discovering, adding, and managing knowledge sources



... and many more on the way

# Adding out-of-box Knowledge to your Agent





[aka.ms/CopilotStudio  
TabularKnowledge](https://aka.ms/CopilotStudio/TabularKnowledge)

# Tabular knowledge sources (Early Access Preview)



Salesforce



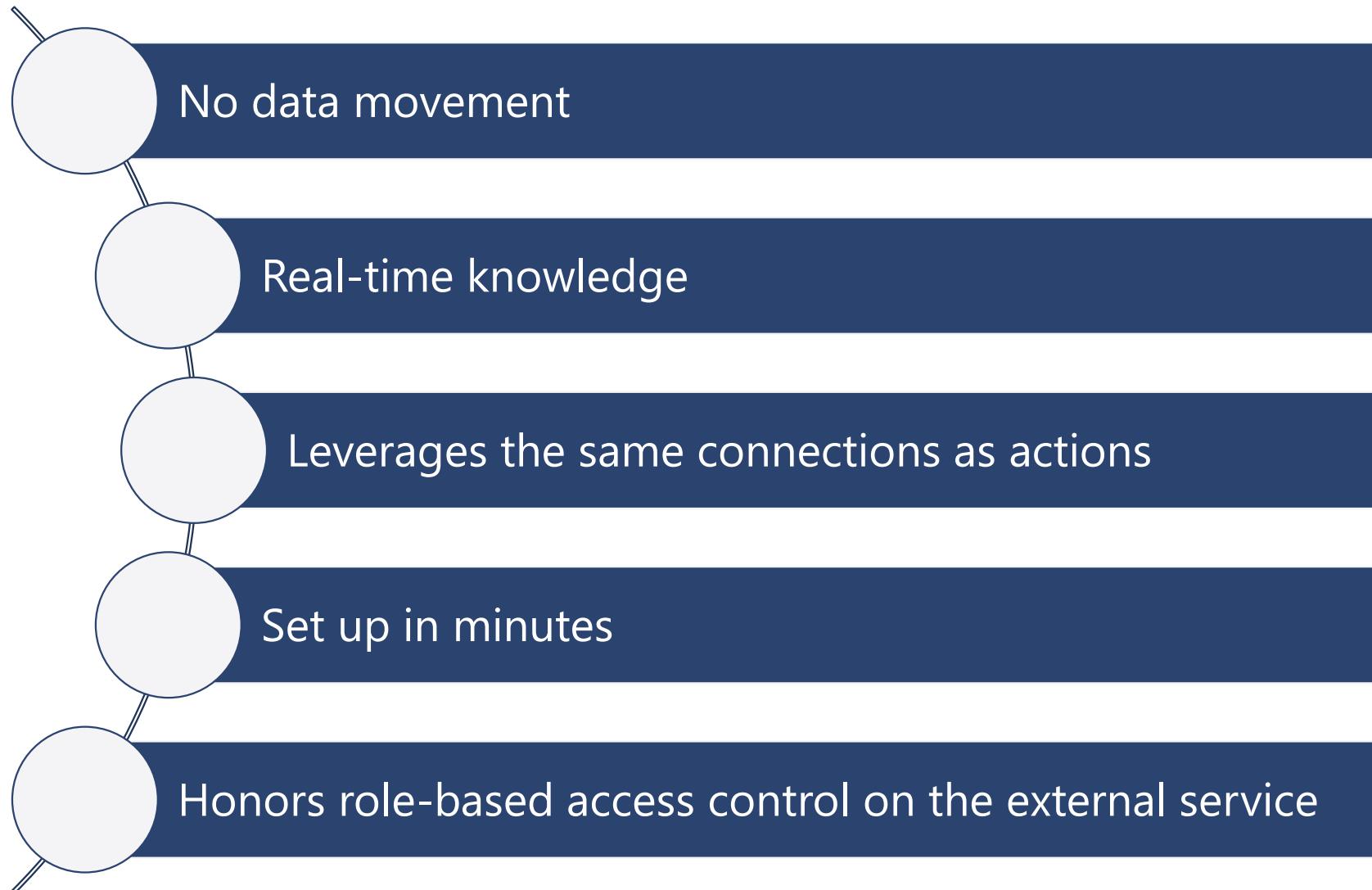
Azure SQL



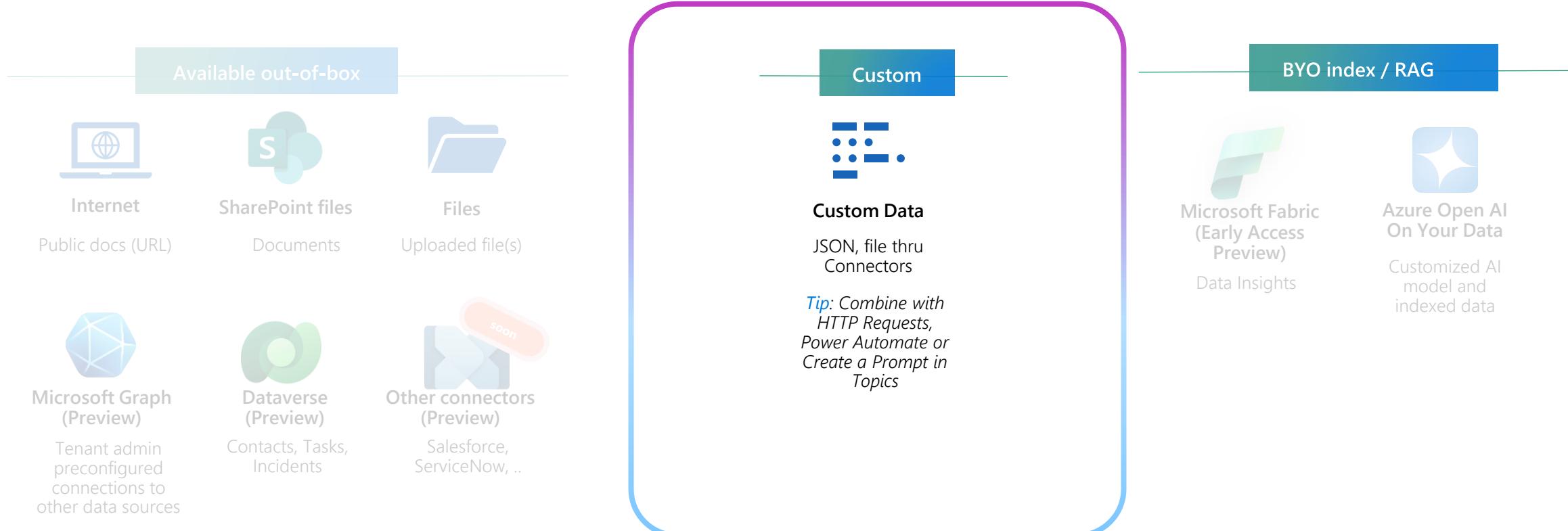
ServiceNow



Zendesk



# Adding custom data to your Agent



# Custom Data

When out-of-the box knowledge source types are insufficient and you want to **use your own data** (e.g. with customized filters)

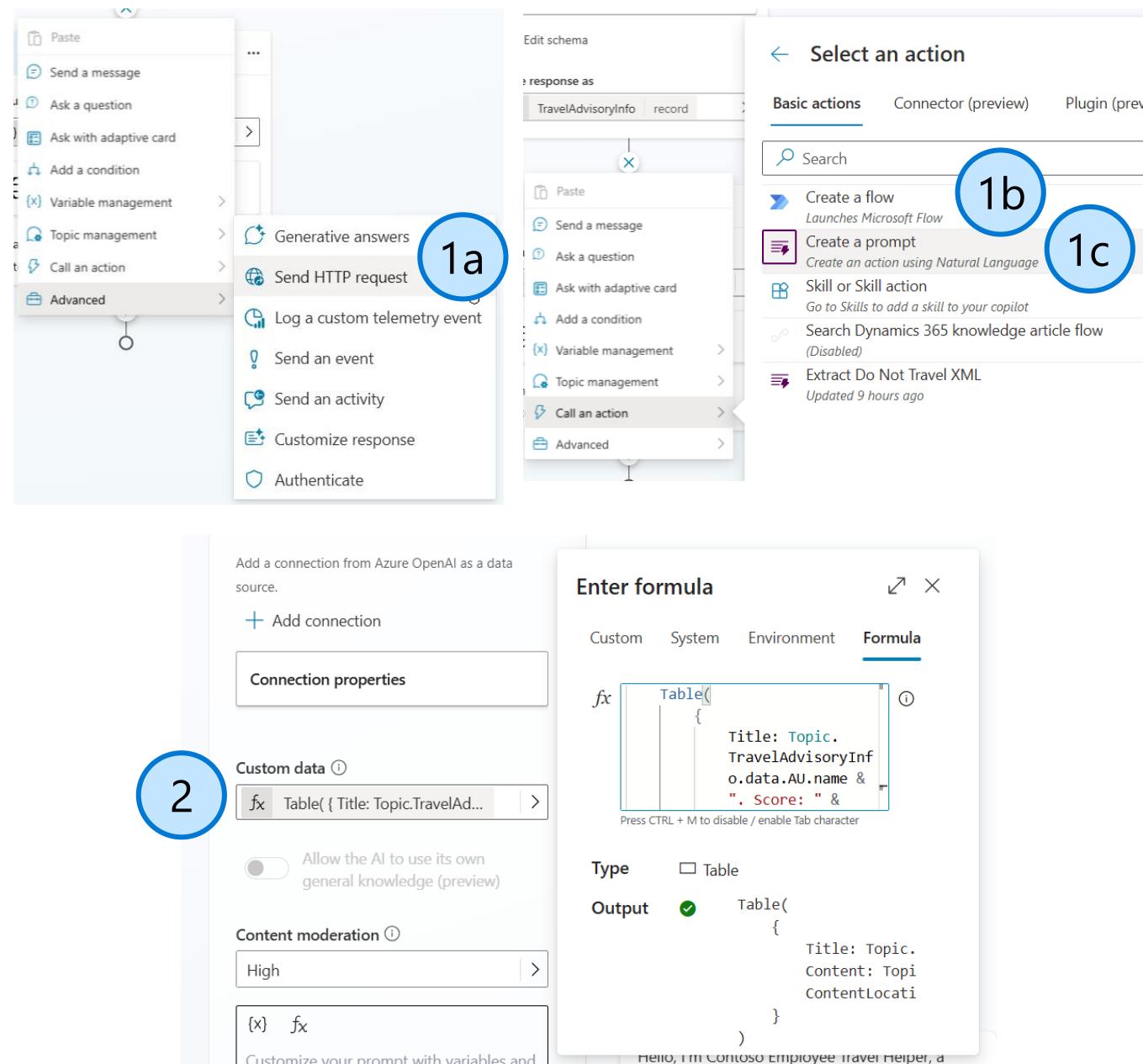
Use it with **Generative Answers** in a topic

Retrieve data with Actions

- HTTP Request **1a**
- Create a Flow (Power Automate) **1b**
- Create a Prompt (AI Builder) **1c**

**2** Data is in table format with these columns:

- **Content** – used by Generative Answers
- **Title** – name of citation
- **ContentLocation** – URL link for citation



The screenshot shows the Microsoft AI Builder interface with three main sections highlighted by blue circles:

- 1a**: A context menu for "Advanced" actions, with "Generative answers" highlighted.
- 1b**: The "Basic actions" section of the "Select an action" pane, with "Create a flow" highlighted.
- 1c**: The "Basic actions" section of the "Select an action" pane, with "Create a prompt" highlighted.

Below these, a "Enter formula" dialog is open, showing a table structure with columns for "Title" and "Score". The formula tab is selected, displaying the following code:

```
Table( { Title: Topic.TravelAd... } )
```

The formula editor also shows the output type as "Table" and the output content as:

```
Table( { Title: Topic. Content: Topi ContentLocati... } )
```

# Velrada

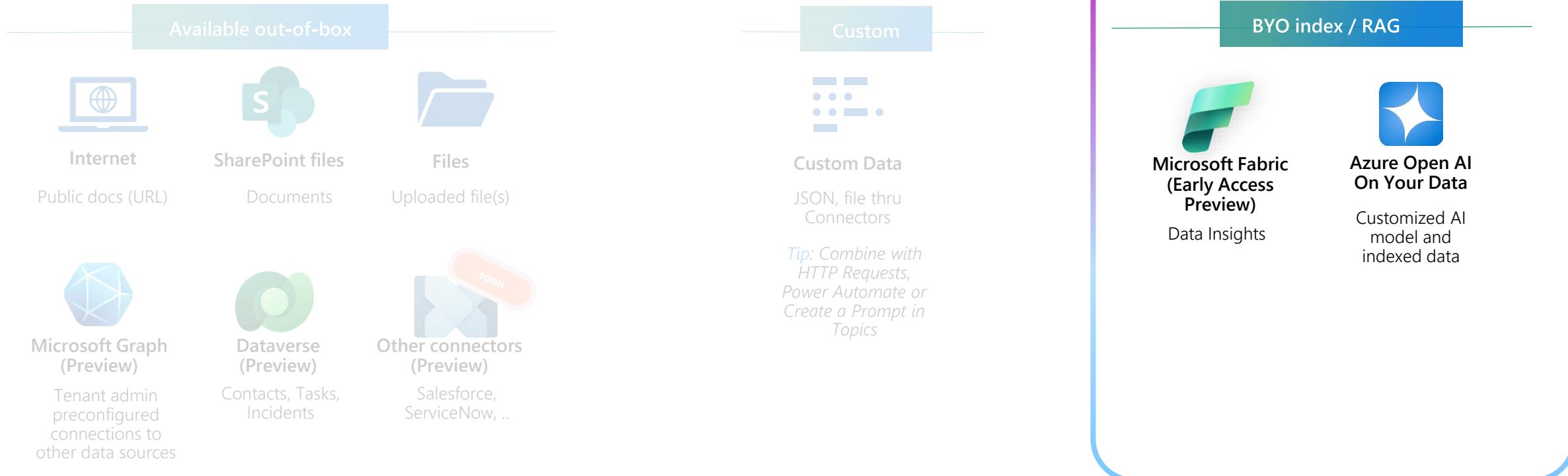


G'day I'm  
**Matthew Pontel**  
Perth, Western Australia

**Velrada**  
Microsoft Cross Platform Director  
& Global Practice Lead – Power Platform

<https://www.linkedin.com/in/mattpontel/>

# Adding your own indexed data to your Agent



# Fabric AI Skills

(Early Access Preview)

Conversational Q&A  
Systems with Gen AI in  
Fabric for large enterprise  
data in OneLake.

## Enhanced accuracy

- **Customization:** instructions and examples to guide the AI to the correct answers
- **Contextual Understanding:** AI skills capture the context and nuances of your data systems

The screenshot shows the Microsoft Fabric AI Skills interface. On the left, the 'Explorer' pane lists tables from the 'Costoso\_Customer\_Sales' database, including dimaccount, dimcurrency, dimcustomer, dimdate, dimdepartmentgroup, dimemployee, dimgeography, dimorganization, dimproduct, dimproductcategory, dimproductssubcategory, dimpromotion, dimreseller, dimsalesreason, dimsalesterritory, and dimscenario. A specific table, 'dimcustomer', is selected. In the center, a query card displays the results of a SQL query: "show 5 top cities that have the most sales?". The results are a table with columns 'City' and 'TotalSales', showing data for London, Paris, Wollongong, Warmambool, and Bendigo. Below the results is the raw SQL query: 

```
SELECT TOP 5 g.city,
    SUM(fs.SalesAmount) AS TotalSales
FROM [dbo].[factinternetsales] fs
JOIN [dbo].[dimcustomer] c ON fs.CustomerKey = c.CustomerKey
JOIN [dbo].[dimgeography] g ON c.GeographyKey = g.GeographyKey
```

. At the bottom, there is a text input field for entering a question or request to test the model's response, and a 'Preview data' button.



[aka.ms/FabricAISSkillMCS](https://aka.ms/FabricAISSkillMCS)

# Fabric AI Skills

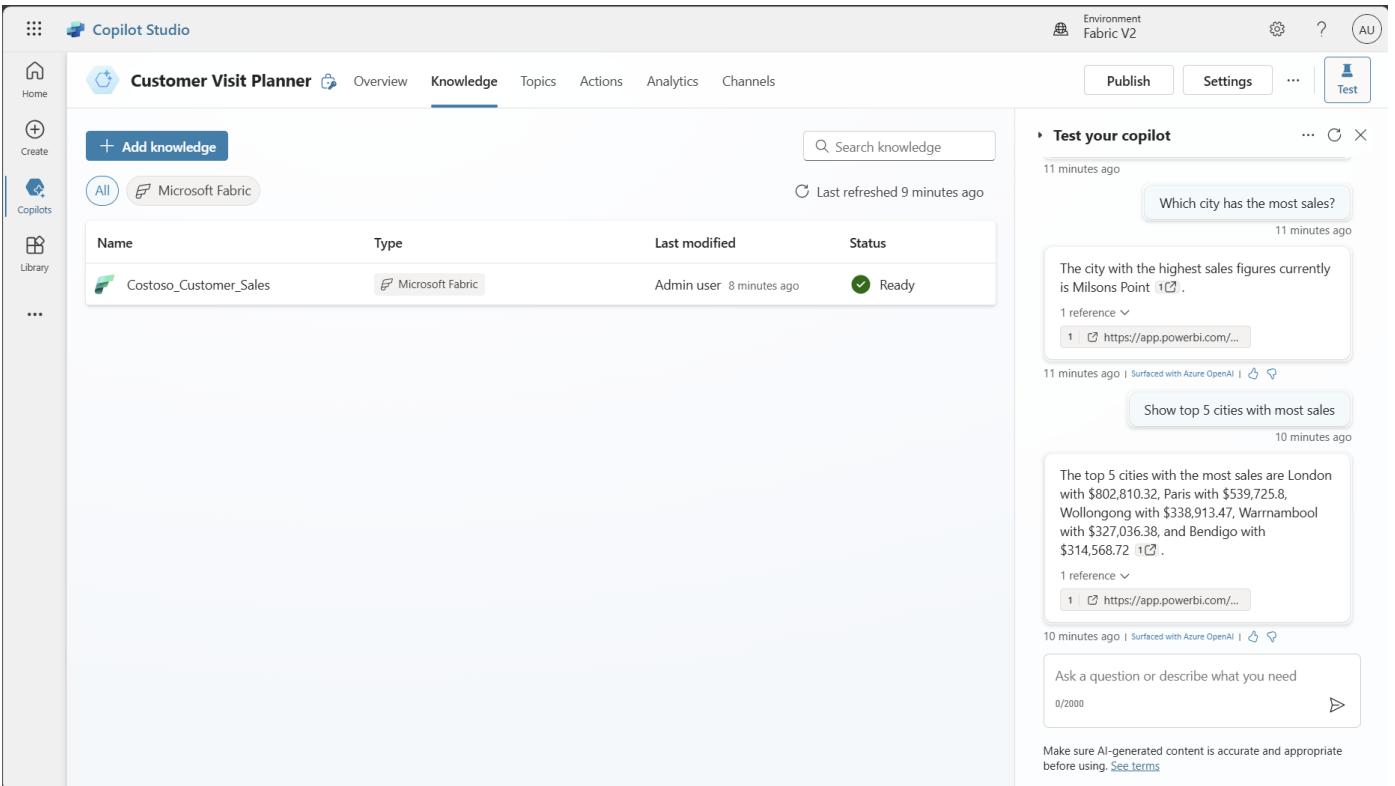
(Early Access Preview)



[aka.ms/FabricAISkillMCS](https://aka.ms/FabricAISkillMCS)

## Seamless integration with Copilot Studio

- Increased Productivity to streamline workflows and enhance data-driven decision-making.



The screenshot shows the Copilot Studio interface with the 'Customer Visit Planner' knowledge base selected. The 'Knowledge' tab is active, displaying a table of existing knowledge. One entry is shown: 'Costoso\_Customer\_Sales' (Type: Microsoft Fabric, Last modified: Admin user 8 minutes ago, Status: Ready). To the right, a sidebar titled 'Test your copilot' shows a series of AI-generated responses to user queries:

- Which city has the most sales? (Answer: The city with the highest sales figures currently is Milsons Point.)
- Show top 5 cities with most sales (Answer: The top 5 cities with the most sales are London with \$802,810.32, Paris with \$539,725.8, Wollongong with \$338,913.47, Warrnambool with \$327,036.38, and Bendigo with \$314,568.72.)
- Ask a question or describe what you need (Placeholder text: 0/2000)

At the bottom, a note states: 'Make sure AI-generated content is accurate and appropriate before using. [See terms](#)'.

How we're improving Agents' Knowledge search quality.

# Searching knowledge is a challenging task.

## Challenge

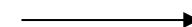
Online evaluation: performance is difficult to define for many real-world tasks



## What we're doing about it

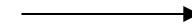
- Before we ship: Automated offline testing and adherence to ship bar
- After we ship: online monitoring across multiple dimensions, e.g. answer rate, response relevance, query rephrase rate, GPT as judge

Responsible AI: Harmful content, jailbreak risk, ungroundedness



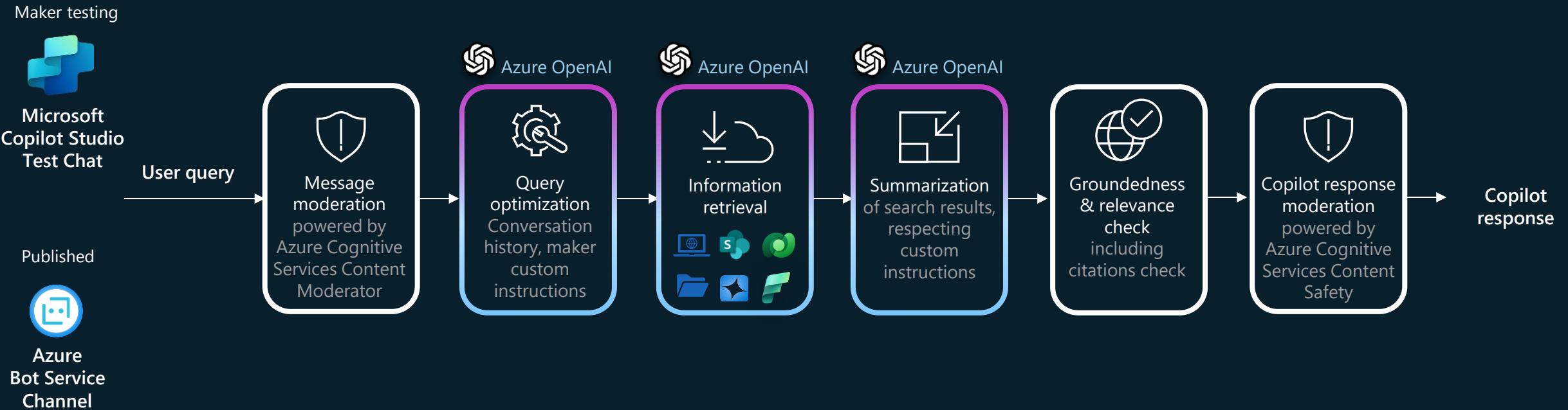
- Separating information retrieval from search result summarization as means to increase groundedness
- Citation check
- Multiple Responsible AI checks along the pipeline for redundancy
- Red teaming

No universally good chunking, retrieval, and re-ranking strategy

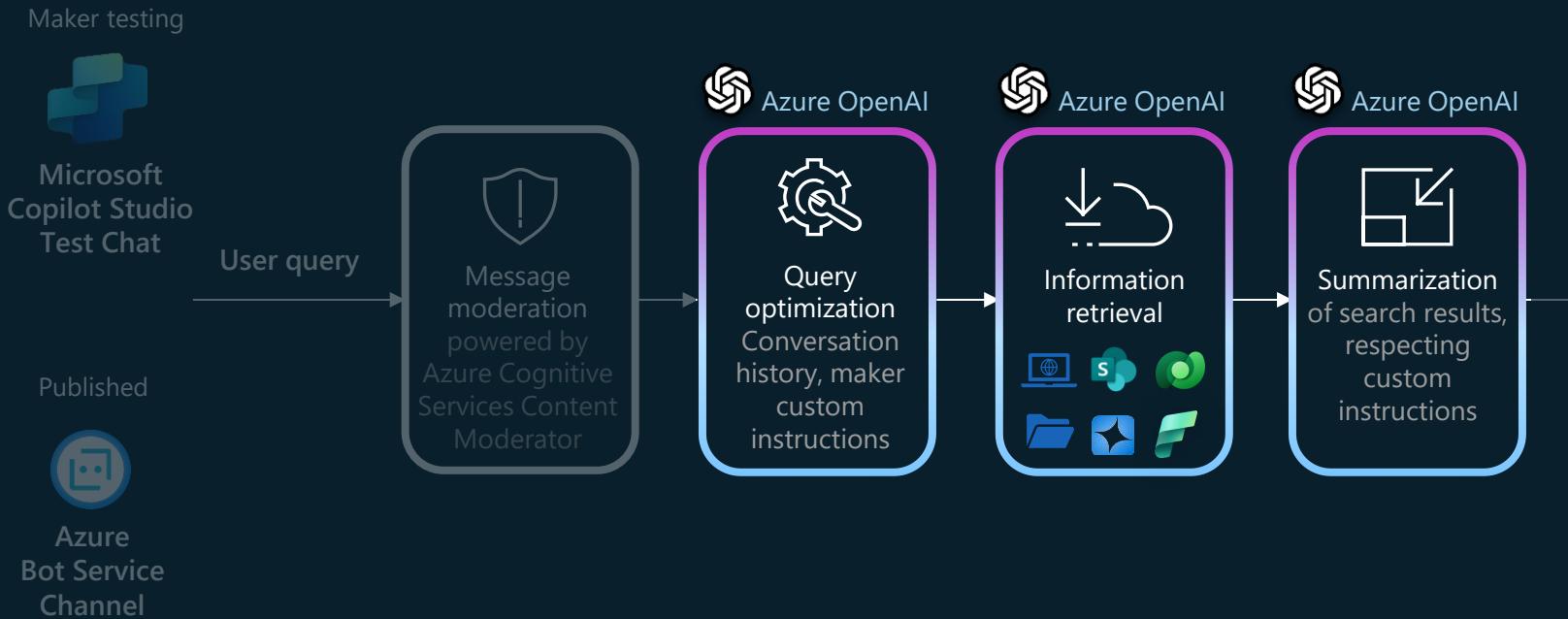


- Knowledge source-specific logic and optimizations (e.g. chunk size, embedding model) to improve retrieval relevance and recall
- Pipelines for structured data and unstructured data

# Knowledge & Generative Answers architecture



# Newest models



Moved to GPT-4o and GPT-4o mini

More relevant search results by **5-10%**

Ability to reason over **images and tables** in files

**13%** better multi- language support

https://learn.microsoft.com/en-us/microsoftsearch/semantic-index-for-copilot

# Microsoft Ignite

Nov 18–22, 2024

Register now >

Learn | Discover | Product documentation | Development languages | Topics

Filter by title

Learn / Semantic index for Copilot

Article • 08/28/2024 • 4 contributors

Feedback

## Semantic index for Copilot

### In this article

- What is an index?
- How the semantic index helps manage your data
- Features
- How the semantic index works

Show 15 more

Semantic index is generated from content in Microsoft Graph. It's used to aid in the production of contextually relevant responses to user queries. It allows organizations to search through billions of vectors (mathematical representations of features or attributes) and return related results. Combined with enhancements across the Microsoft Graph, semantic index connects you with relevant information in your organization. It's built on Microsoft's comprehensive approach to security, compliance, privacy, and respects all organizational boundaries within your tenant.

## What is an index?

The concept of indexing data is well established in Microsoft 365. Indexing is one of the important ways that Microsoft 365 services access the tremendous amount of data in Microsoft Graph,

Lower latency & higher quality responses

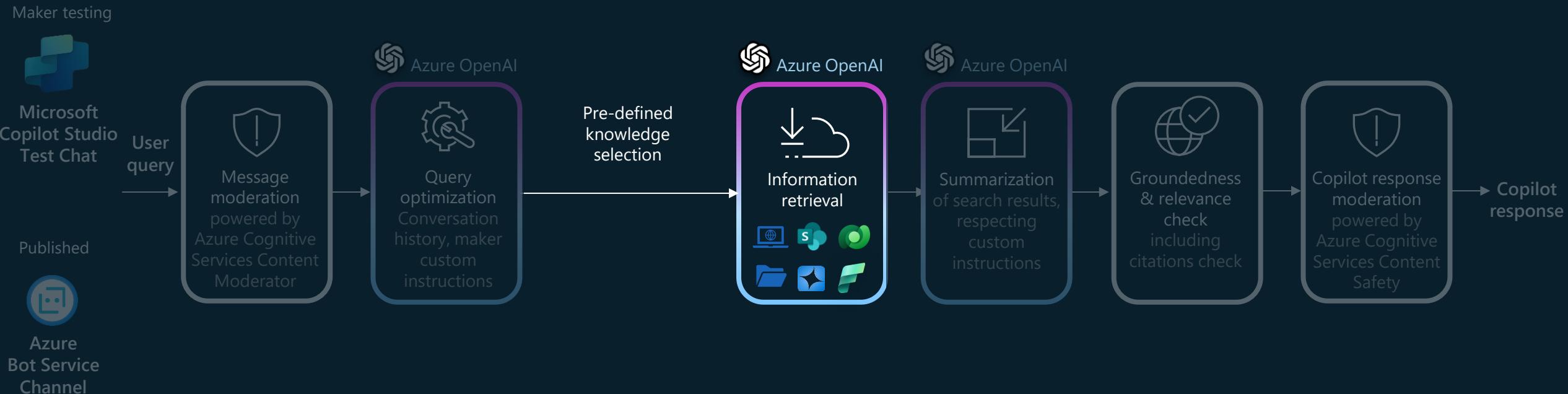
New PDF parsing mechanism reduces latency by **45%**

Improved document & passage selection logic improves response quality by **20%**

More quality improvements to the search itself at Ignite with transition from keyword to semantic index

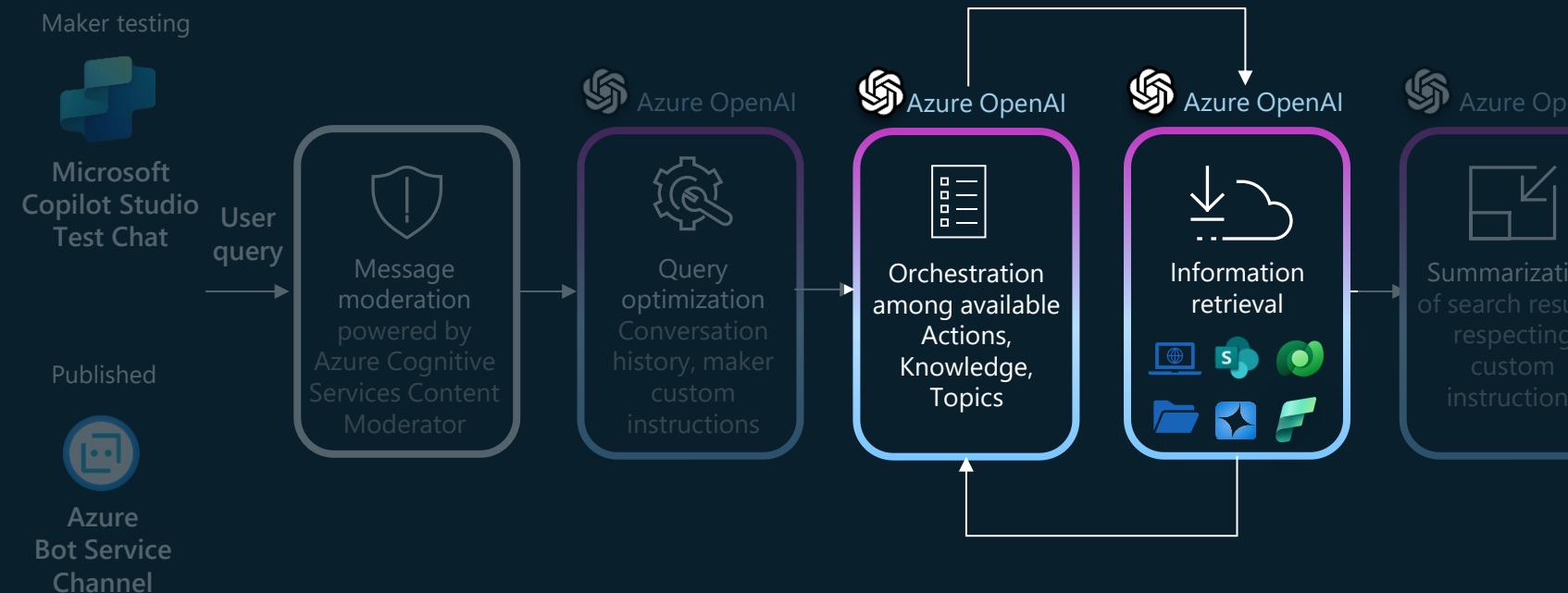
# Deterministic to dynamic conversation flow

Knowledge source selected according to pre-determined maker configuration



# Deterministic to dynamic conversation flow

Knowledge source selected at runtime; handles compound questions, chained searches



using GPT-4o to identify when and what knowledge Agent should search

Through **Generative** mode

Can handle **complex questions** (ambiguous, require reasoning)

Better **maker experience**: No need to embed knowledge sources in a Generative Answers node– simply add to the Knowledge page

# What you can do to improve knowledge search

## Tip

Knowledge choice should match the agent's job-to-be-done

Add metadata during configuration of knowledge sources, e.g. synonyms and glossary when using Dataverse tables.

Knowledge descriptions impacts orchestration quality:  
Simple, informative, short, unique.

## What good looks like

An HR Copilot should likely include information about vacation policy, employee performance, healthcare coverage, etc.

Synonym: Column name: ABCD\_123 → conference code, conf code  
Glossary: preso → presentation

"This knowledge source can answer questions" → "This SharePoint site can answer questions about HR policies like time off, promotions, and compensation."

# What you can try now -

## Available out-of-box

- Public Websites
- SharePoint
- File uploads
- Dataverse (*Preview*)
- Graph connections (*Preview*)
- Tabular Knowledge (*Early Access Preview*)

## Bring your custom data (add via Topics)

- Custom Data, via actions:
  - [HTTP Request](#)
  - [Power Automate](#)
  - [Create a Prompt](#)

Reference -

[Use a custom data source for generative answers - Microsoft Copilot Studio | Microsoft Learn](#)

## Bring your own index / RAG

- Fabric AI Skills (*Early Access Preview*)
- [Azure Open AI On Your Data](#)



## Learn more:

<https://aka.ms/trycopilotstudio>

<https://aka.ms/copilotstudiotoknowledge>

<https://www.youtube.com/@Dewain27>

Admin controls: [Configure data loss prevention policies for copilots](#)

# Sign up for Early Access Previews

[aka.ms/CopilotStudioTabularKnowledge](https://aka.ms/CopilotStudioTabularKnowledge)

Copilot Studio Tabular knowledge:  
Early Access sign-up form



Salesforce, ServiceNow, Zendesk, Azure SQL

[aka.ms/FabricAISkillIMCS](https://aka.ms/FabricAISkillIMCS)

Early access - Fabric AI skills in  
Copilot Studio



# Thank you

**Julie Koesmarno and Noa Ghersin**

[julie.Koesmarno@microsoft.com](mailto:julie.Koesmarno@microsoft.com) , [nghersin@microsoft.com](mailto:nghersin@microsoft.com)

<https://aka.ms/juliek>