

Prompt Flow



gollnickdata.de

Introduction

Prompt Flow

What is it?

- Helps to build test, and deploy AI workflows that use LLMs
- Design chains (multi-step LLM workflows)
- Test and debug prompts
- Evaluate performance and quality
- Deploy flows as API or AI services



Prompt Flow

Getting started

+ Create new


Create project

Start by choosing a resource type:




Azure AI Foundry resource

Recommended

Unifies setup, management and access to agents, models and tools. All new capabilities will be introduced through this resource. [Learn more about the Azure AI Foundry resource](#) 



AI hub resource

For advanced scenarios like custom ML training, open-source model hosting, fine-tuning, or Azure Machine Learning integration. [Learn more about the Azure AI Hub resource](#) 

Next

Cancel

Important!!!

Prompt flow is only available in
Hub-resources

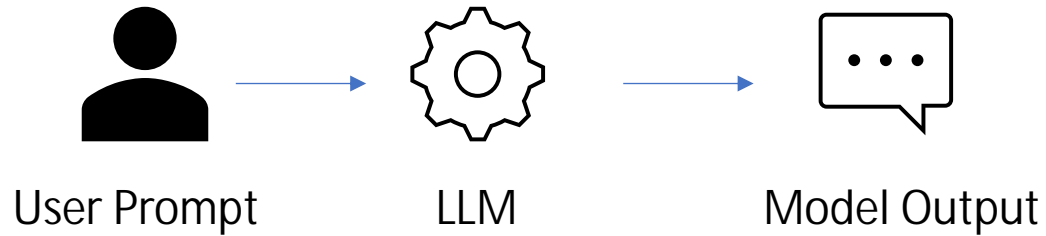


gollnickdata.de

New flow

Prompt Flow

New flow



Prompt Flow

New flow

- Overview
- Model catalog
- Playgrounds
- AI Services
- Build and customize
 - Agents PREVIEW
 - Templates
 - Fine-tuning
 - Content Understanding PREVIEW
 - Prompt flow**

Create a new flow

Create by type

Standard flow

Harness the power of Large Language Models, customized Python code, and more to craft your tailored prompt flow. Test the flow using custom datasets and seamlessly deploy as an endpoint for easy integration.

Create

Chat flow

On top of the standard flow, this option provides the chat history support and a user-friendly chat interface in the authoring/debugging UI.

Create

Evaluation flow

Create an evaluation flow to measure how well the output matches the expected criteria and goals.

Create

Explore gallery

All Standard flow Chat flow Evaluation flow

Chat Multi-Round Q&A on Your Data

Create a chatbot that uses LLM and data from your own indexed files to ground multi-round question and answering capabilities in enterprise chat scenarios.

View detail Clone

Standard Q&A on Your Data

Use LLM and data from your own indexed files to ground multi-round question and answering capabilities.

View detail Clone

Standard Web Classification

Use LLM to classify URLs into multiple categories.

View detail Clone

Chat Chat with Wikipedia

Create a chatbot that leverages Wikipedia data to ground the responses.

View detail Clone

Chat Use GPT Function Calling

Learn how to use GPT function calling to extend the capabilities of GPT models with external data sources.

View detail Clone

Evaluation Classification Accuracy Evaluation

Measuring the performance of a classification system by comparing its outputs to groundtruth.

View detail Clone

Evaluation QnA Groundedness Evaluation

Compute the groundedness of the answer for the given question based on the context.

View detail Clone

Evaluation QnA Relevance Evaluation

Compute the relevance of the answer for the given question based on the context.

View detail Clone

Prompt Flow

New flow

- Running it the first time might result in an error.
- Solution: try it a second time and it will work

The screenshot displays the 'Basic Flow' configuration interface. At the top, there are tabs for 'Clone', 'Save', 'Deploy', and 'Evaluate'. Below the tabs, the 'Flow' section is active, showing a sequence of steps: 'inputs', 'joke', and 'echo'. The 'inputs' step is a simple input box. The 'joke' step is a prompt action. The 'echo' step is a Python action. The 'Outputs' section shows a single output named 'joke' with the value '\${echo.output}'. The 'Code' section at the bottom shows the Python code for the 'echo' step, which is a simple echo function.

Basic Flow View batch runs Clone Save Deploy Evaluate

Flow

+ LLM + Prompt + Python + More tools Raw file mode Wrap text Diff mode

▼ Inputs ⓘ [Show description](#)

Name	Type	Value	Action
topic	string		

+ Add input

▼ Outputs ⓘ

Name	Value	Action
joke	\${echo.output}	

+ Add output

echo python ▶ 🗑️ ↑ ↓ 🔍 🔗

▼ Code ⓘ Referring to: [echo.py](#)

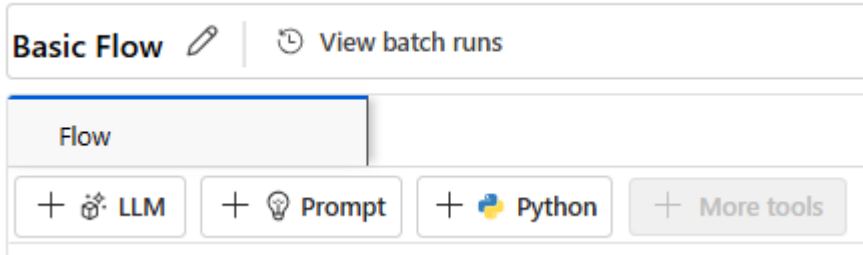
Files

Graph

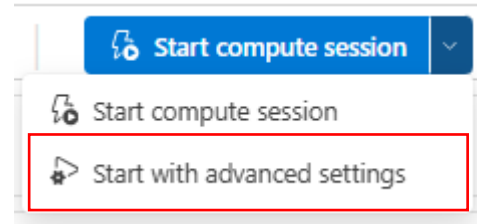
```
graph TD; inputs --> joke; joke --> echo; echo --> outputs;
```

Prompt Flow

New flow



- Tools are hidden – to see them you need to start compute session
- Choose a cheap one



Prompt Flow

New flow

variant_0 Run completed: View outputs

Basic Flow Completed View batch runs View outputs Clone Save Deploy Evaluate Compute session running Run

Flow

+ LLM

+ Prompt

+ Python

+ More tools

Raw file mode

Wrap text

Diff mode

> Advanced

> Function calling

> Prompt Referring to: LLM.jinja2

```
1 # system:
2 You are a helpful assistant that tells 5 fun facts about a particular topic and each one in one or two sentences.
3 # user:
4 {{topic_of_interest}}
```

> Inputs Validate and parse input Validation and parsing input completed successfully.

Name	Type	Value
topic_of_interest	string	\$(inputs.topic_of_interest)

> Activate config

> Outputs Duration 2.22s Completed View full output

Input

Output

Trace

Logs

```
[
  {
    "system_metrics": {
      "duration": 2.223704
    },
    "output":
    "Here are five fun facts about Germans:\n\n1. **Beer Purity Law**: Germany is famous for its Reinheitsgebot, or Beer Purity Law, enacted in 1516, which originally allowed only water, barley, and hops to be used in beer production, ensuring high quality and taste.\n\n2. **Inventio
```

Files

Graph

```
graph TD; inputs([inputs]) --> LLM[LLM]; LLM --> outputs([outputs]);
```

Prompt Flow

New flow

×


View outputs

Outputs

Outputs Logs Metrics Trace

↓ Export ▾

🔍 Search

Details	#	inputs.topic_of_interest	Status	facts
	0	germans	✅ Completed	Here are five fun facts about Germans: 1. Beer Purity Law : Germany is famous for its Reinheitsgebot, or Beer Purity Law, enacted in 1516, which originally allowed only water, barley, and hops to be used in beer production, ensuring high quality and taste. 2. Invention of Items : Germany has been the birthplace of numerous inventions, including the automobile by Karl Benz in 1886, the Christmas tree tradition, and even summer beer , which was invented by Hans Riegel in the

Prompt Flow

New flow



Deploy

Deploy Basic Flow

- 1 Basic settings
- 2 Advanced settings
- 3 Review

Basic settings

Deploy your flow to a managed online endpoint for real-time inference. [Learn more](#)

Endpoint

☒ New ☐ Existing

Endpoint name * ⓘ

aihub-based-project-ocuzq

Deployment name * ⓘ

aihub-based-project-ocuzq-1

Virtual machine * ⓘ

Standard_D2as_v4 2 Cores, 8 GB (RAM), 16 GB (Disk), \$0.10/hr

Instance count * ⓘ

1

Inferencing data collection ⓘ

☒ Enabled

Review + Create

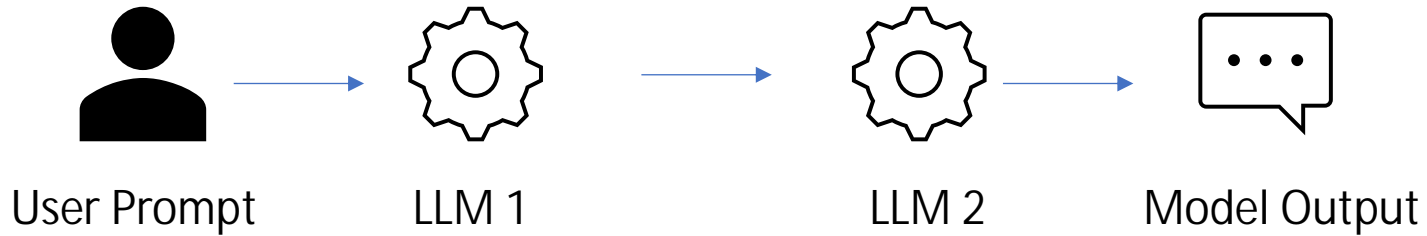


gollnickdata.de

Flow with multiple LLMs

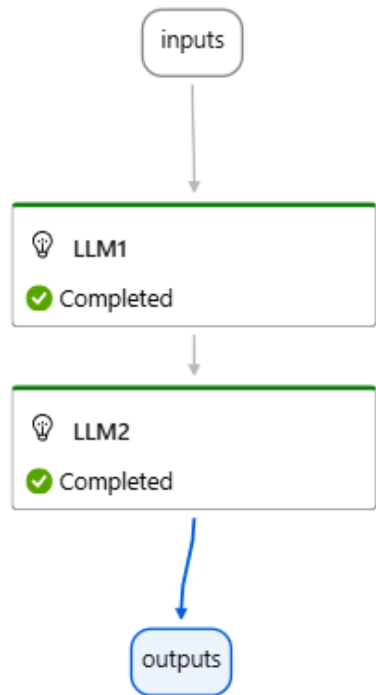
Prompt Flow

Flow with multiple LLMs



Prompt Flow

Flow with multiple LLMs



system:

You are a helpful assistant that explains a particular topic in 15-20 sentences

user:

{{topic_of_interest}}

system:

You are a helpful assistant that turns hard explanations into simple words for a 5-year old.

user:

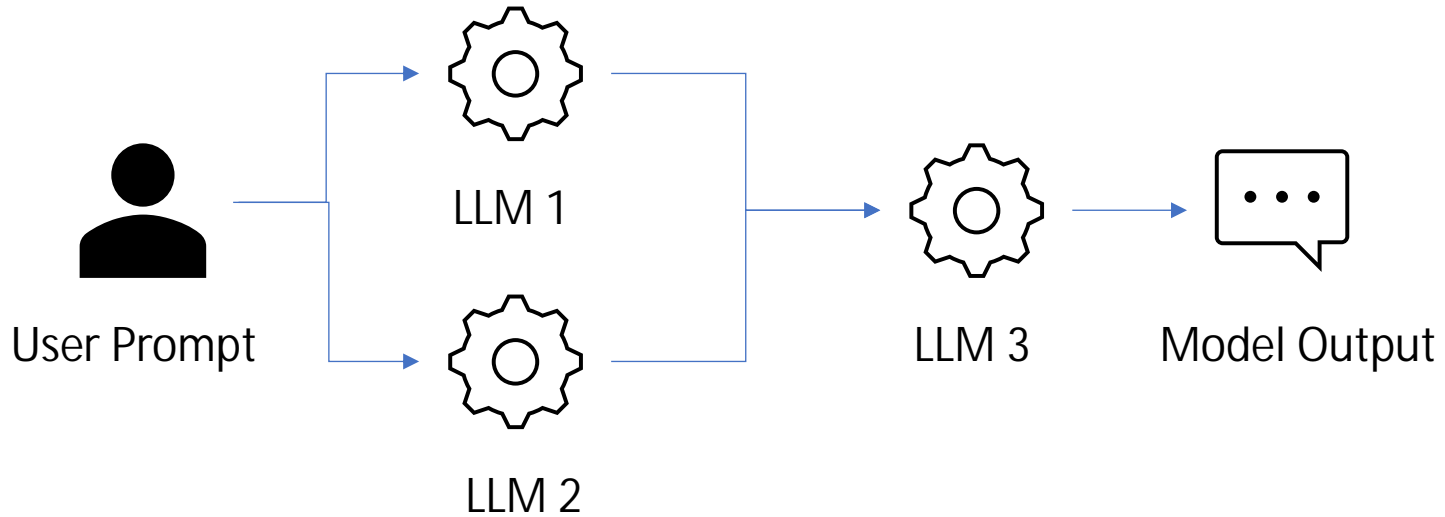
{{explanation}}



Parallel Flow

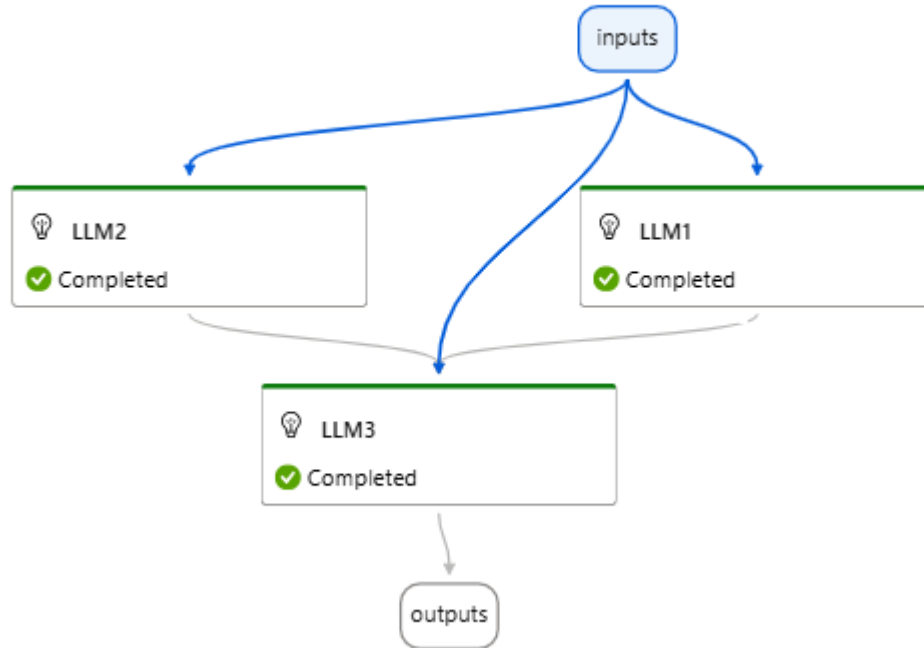
Prompt Flow

Flow with multiple LLMs



Prompt Flow

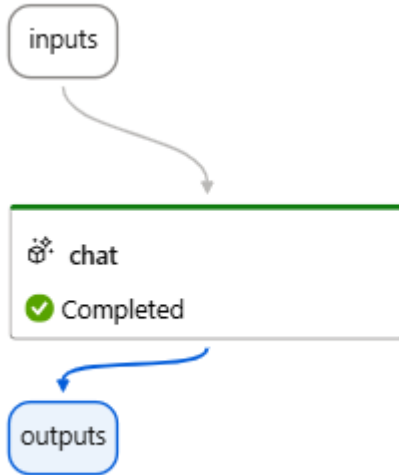
Flow with multiple LLMs



Chat Flow

Prompt Flow

Chat Flow



Chat History

```
1  # system:
2  You are a helpful assistant.
3
4  {% for item in chat_history %}
5  # user:
6  {{item.inputs.question}}
7  # assistant:
8  {{item.outputs.answer}}
9  {% endfor %}
10
11 # user:
12 {{question}}
13
```

RAG with Prompt Flow

Prompt Flow

RAG with Prompt Flow

Create Data

Azure AI Foundry / aihub_based_project / Data + indexes / Kurse:1

← Kurse:1

Data details

Name
Kurse

Current version
Version: 1 (latest) + New version

Latest version
1

Type
Folder

Number of files 3	Total size ⓘ 1.149 MiB
Created on Oct 28, 2025, 5:36:12 PM	Modified on Oct 28, 2025, 5:36:12 PM

Created by
 Bert Gollnick

Tags
+ Add tag

Overview

Model catalog

Playgrounds

AI Services

Build and customize ^

Agents PREVIEW

Templates

Fine-tuning

Content Understanding PREVIEW

Prompt flow

Observe and optimize ^

Tracing PREVIEW

Monitoring

Protect and govern ^

Evaluation

Guardrails + controls

⏪

- ▼ a7034941-2a20-4073-917a-a70...
- ▼ UI
 - ▼ a7034941-2a20-4073-917a-a70...
 - 2025-10-28_163533.UTC
 - Kursangebot_AiEngineer...
 - Kursangebot_Generative...
 - Kursangebot_KifürJobUn...



gollnickdata.de

Prompt Flow

RAG with Prompt Flow

Create Index

Create a vector index PREVIEW

- ✓ Source location
- ✓ Index configuration
- 3 Search settings
- 4 Review and finish

Configure search settings

Combining hybrid retrieval with semantic ranking (Hybrid + Semantic) gives most accurate search results for generative AI applications. To generate vector index, embedding model is required.

Vector settings

☒ Add vector search to this search resource

Azure OpenAI connection * ⓘ

ai-bertgollnick9527ai080637128723_aoi

Embedding model * ⓘ

text-embedding-ada-002

Embedding model deployment ⓘ

Select an embedding model deployment

ⓘ No embedding model deployments found.

ⓘ This resource requires an embedding model. If you don't have one already, **text-embedding-ada-002:2** will be deployed for you. Using vector embeddings will incur usage to your account. [View Azure OpenAI pricing](#)

Back

Next

Create vector index

Cancel

Prompt Flow

RAG with Prompt Flow

Check Index

← shy-gyro-t0hw5mt8ny-index

Status

● Running

Refresh

Version

-

Embed with model

No

Source type

Azure AI On Your Data

Vector store

-

Total indexing time

-

Compute

Serverless compute

Created on

Oct 28, 2025, 5:42:35 PM

Created by

Bert Gollnick

Job details

Test data

Source data

Name	Type	Size
Kursangebot_AiEngineer_v01....	.pdf	613.03 KB
Kursangebot_GenerativeKI_All...	.pdf	211.31 KB
Kursangebot_KiFürJobUndAllta...	.pdf	352.16 KB

Prompt Flow

RAG with Prompt Flow

Index Lookup

Generate












Name	Type	Value
index_type	string	Azure AI Search ⓘ ▾
acs_index_connection	Azure AI Search	search20251002 ▾
acs_index_name	string	shy-gyro-t0hw5mt8ny-index ⓘ ▾
acs_content_field	string	content ⓘ ▾
acs_embedding_field	string	▾
acs_metadata_field	string	meta_json_string ⓘ ▾
semantic_configuration	string	azureml-default ⓘ ▾
embedding_type	string	Azure OpenAI ⓘ ▾
aoai_embedding_connection	Azure OpenAI	ai-bertgollnick9527ai080637128723_aoai ▾
embedding_deployment	string	text-embedding-ada-002 ⓘ ▾

Prompt Flow

RAG with Prompt Flow

Index Lookup

 **indexlookup**   Index Lookup      

▼ Inputs

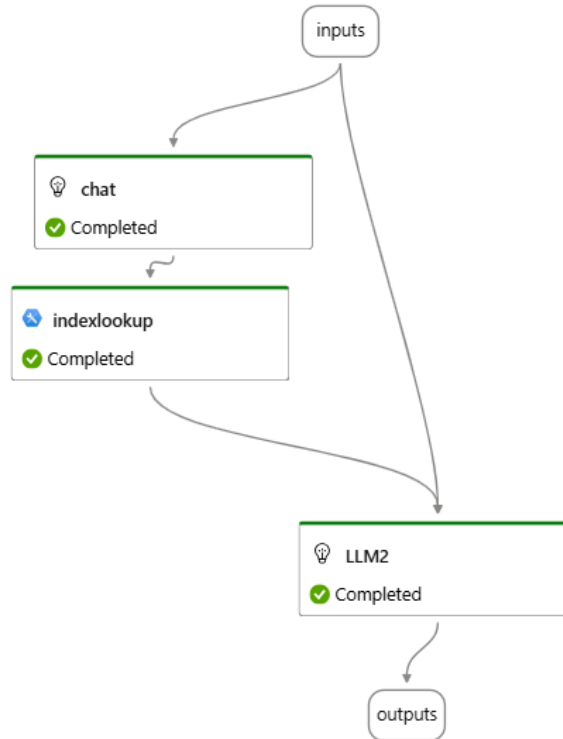
Name	Type	Value
mlindex_content	string	<input type="text" value="embeddings: api_base: https://ai-bertgollnick9527ai080637128723.openai.azure"/>
queries	object	<input type="text" value="\${chat.output}"/>
query_type	string	<input type="text" value="Semantic"/> ⓘ ▼
top_k	int	<input type="text" value="3"/>

> Activate config

Prompt Flow

RAG with Prompt Flow

Prompt Flow



gollnickdata.de

Prompt Flow

RAG with Prompt Flow

Test

weissst du wer bert gollnick ist?

Ja, Bert Gollnick ist ein Trainer und Data Scientist mit einem Hintergrund in

Time spent: 5.55 sec

in welchem kurs geht es um pytorch

Der Kurs, der sich mit PyTorch beschäftigt, ist der AI-Engineering Kurs

Time spent: 5.6 sec



gollnickdata.de

Flow with Web Search

Prompt Flow

Flow with Web Search

Management center

All resources

Quota

Hub (bertgollnick-9527_ai) ▾

Overview

Users

Models + endpoints

Connected resources

Compute

Project (aihub_based_project) ▾

Overview

Users

Models + endpoints

Connected resources

Go to project

+ New connection

Add a connection to external assets

serp

Azure AI

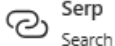
Data

Agent Knowledge Tools

Indexes

Azure Databricks

Other resource types



gollnickdata.de

Prompt Flow

Flow with Web Search

+ More tools

Search more tools

- Content Safety (Text Analyze)
- Embedding
- Open Model LLM
- Serp API**
- Index Lookup
- Azure OpenAI GPT-4 Turbo with Vision Preview
- OpenAI GPT-4V Preview
- Rerank Preview

websearch Serp API

Inputs

Name	Type	Value
connection	Serp	SERPAPI
engine	string	google
location	string	
num	int	5
query	string	\$(inputs.question)
safe	string	off




> Activate config

> Outputs

Duration 2.94s Completed View full output


Prompt Flow


Flow with Web Search


 **LLM**   llm


Show variants


Generate variants














Connection * ai-bertgollnick9527ai080637128723_aiai Api * chat



deployment_name * gpt-4o-mini temperature 1 stop max_tokens response_format [{"type":"text"}]



> Advanced

> Function calling

< Prompt ⓘ Referring to: LLM.jinja2


```
1 # system:
2 You are a helpful assistant.
3 # user:
4 answer the question based on the information provided.
5 information: {{information}}
6 question: {{question}}
```

< Inputs  Validate and parse input  Validation and parsing input completed successfully.

Name	Type	Value
information	string	<div>\$(websearch.output)</div>
question	string	<div>\$(inputs.question)</div>

> Activate config

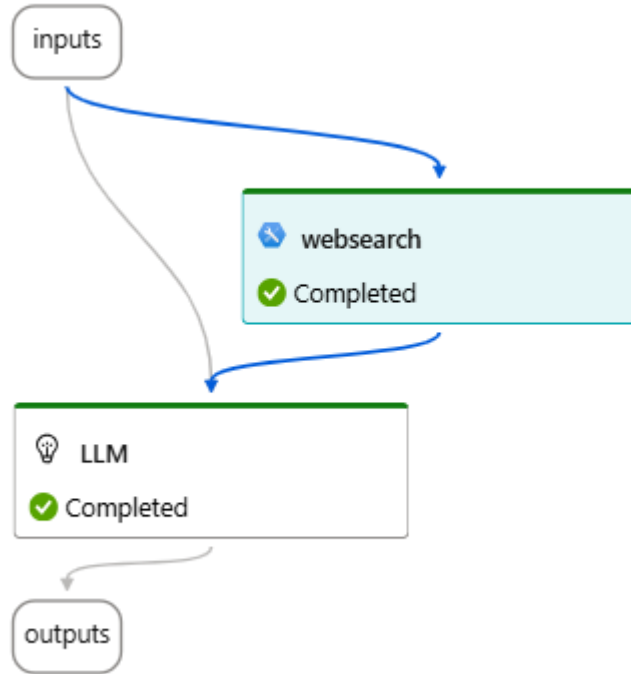
> Outputs

Duration 1.31s  Completed

View full output

Prompt Flow

Flow with Web Search



Search				
Details	#	inputs.question	Status	answer
	0	wieviele einwohner hat hamburg im oktober 2025.	Completed	Im Oktober 2025 wird die Bevölkerung Hamburgs auf etwa 1.912.000 geschätzt. Diese Zahl stammt aus einer Prognose des Bundesinstituts für Bau-, Stadt- und Raumforschung (BBSR) und bezieht sich auf das Jahr 2025.