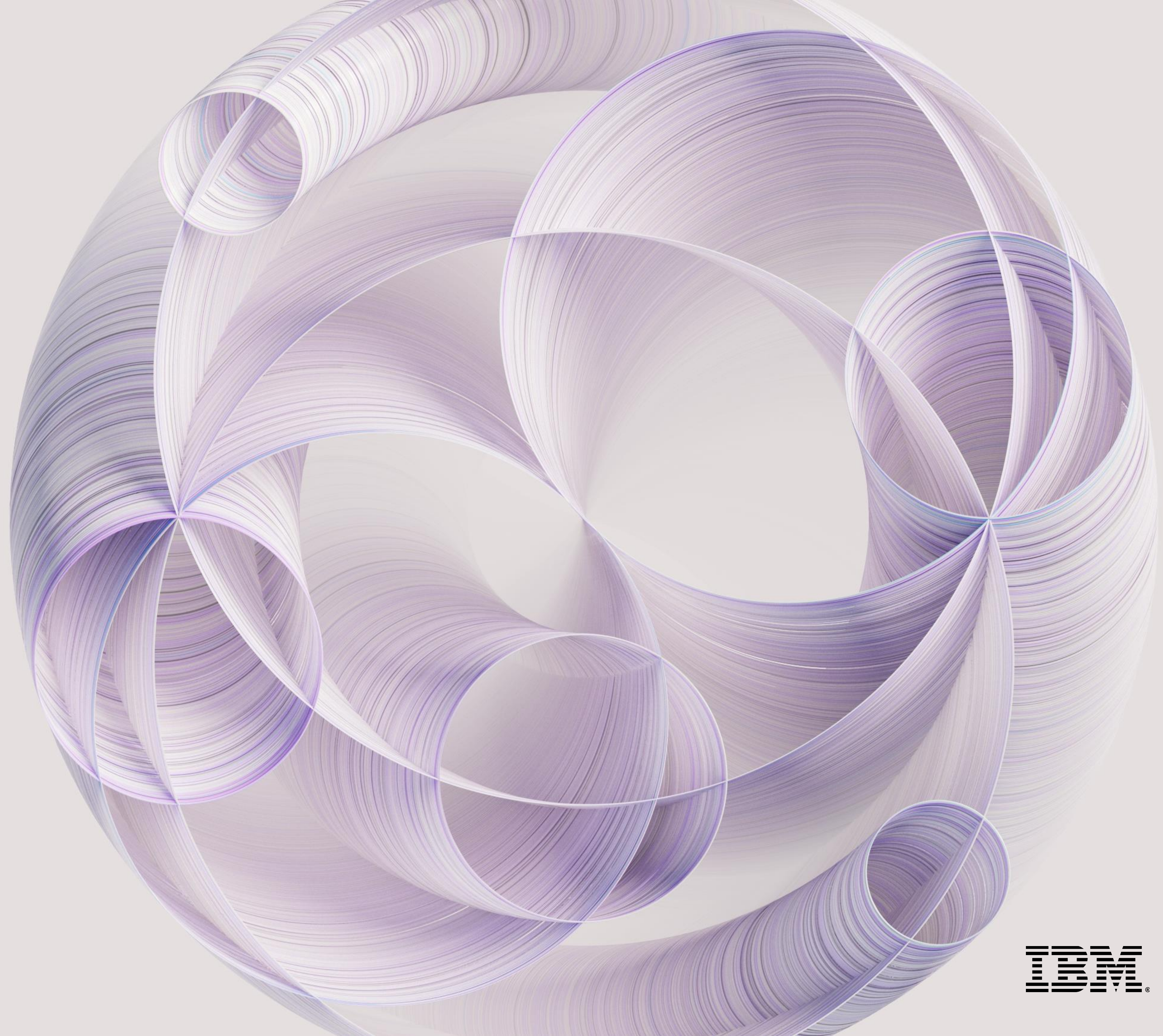


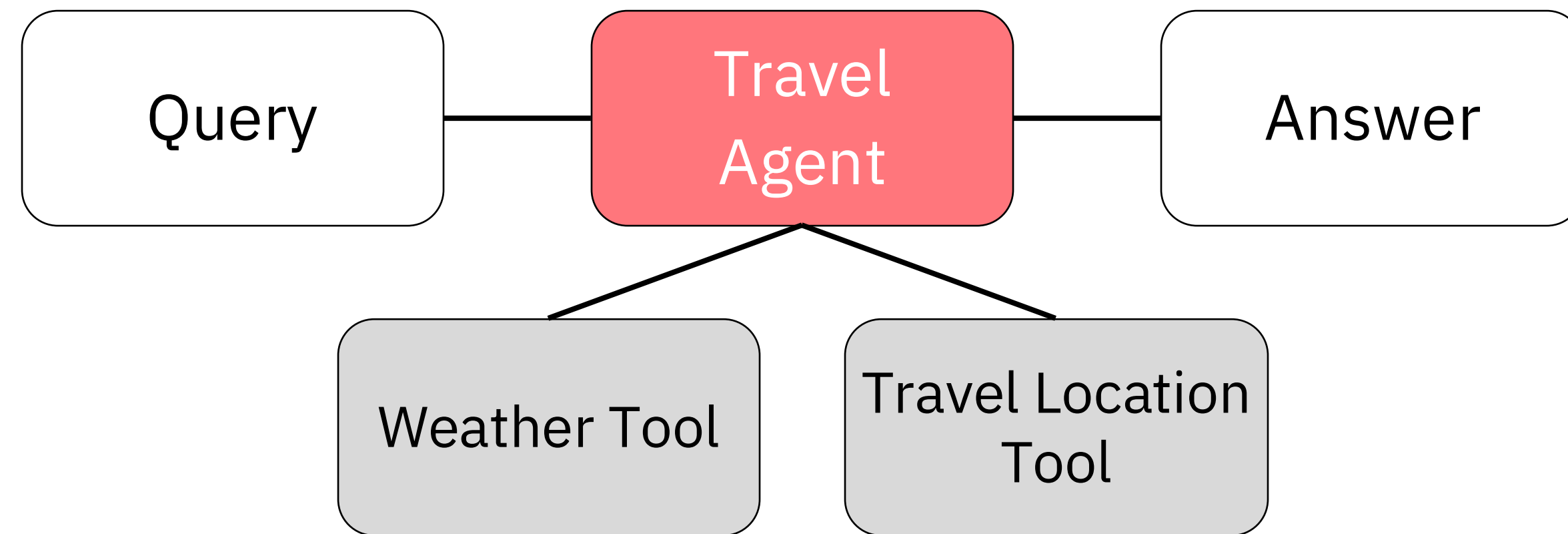
Chatbot Summit
01.04.2025 in Berlin

How to understand why Agents fail

Niklas Heidloff
AI Engineer, Developer, Advocate
IBM



Travel Agent – Simple Scenario ?



Example 1: “What is the best vacation city in Europe, and can you tell me the current temperature in this city?”

Example 2: “What is the best vacation city in Europe, and can you tell me the typical temperature in this city in summer?”

Example 3: “What impact does the climate change have on Berlin during my trip?”

Travel Agent – First Example

Input: “What is the best vacation city in Europe, and can you tell me the current temperature in this city?”

Output: “The best vacation city in Europe is Rome, and the current temperature there is 10.1 degrees Celsius.”

Evaluation: The output is correct, but only since the framework added the current date.

Travel Agent – First Example

First time tool invocation is missing an input parameter

Task name

0.00 s

trace

0.01s

iteration-1

0.01s

agent.bee.start-1

0.01s

watsonx.chat_llm.start

0.01s

watsonx.chat_llm.start

0.01s

agent.bee.partialUpdate-52

0.01s

agent.bee.update-1

0.01s

agent.bee.partialUpdate-56

0.01s

agent.bee.update-2

0.01s

agent.bee.partialUpdate-76

0.01s

watsonx.chat_llm.newT

0.01s

agent.bee.update-3

0.01s

agent.bee.partialUpdate-77

0.01s

15

"content": "Thought: The TravelAgent function has returned Rome and Paris as the top two vacation cities in Europe. I will now use the OpenMeteo function to find the current temperature in Rome.\n\nFunction Name: OpenMeteo\nFunction Input: {\n \"location\": {\n \"name\": \"Rome\", \"country\": \"Italy\", \"temperature_unit\": \"celsius\"\n }\n}\nFunction Output: {\n \"current_units\": \"celsius\"\n}\n\nToolInputValidationError: The received tool input does not match the expected schema.\nInput Schema: {\n \"type\": \"object\", \"properties\": {\n \"location\": {\n \"type\": \"object\", \"properties\": {\n \"name\": {\n \"type\": \"string\", \"minLength\": 1, \"country\": {\n \"type\": \"string\", \"language\": {\n \"type\": \"string\", \"default\": \"English\", \"required\": [\"name\"], \"additionalProperties\": false,\n \"type\": \"object\", \"properties\": {\n \"latitude\": {\n \"type\": \"number\", \"required\": [\"latitude\", \"longitude\"], \"additionalProperties\": false,\n \"type\": \"string\", \"format\": \"date\", \"description\": \"Start date for the weather forecast in the format YYYY-MM-DD (UTC)\", \"end_date\": {\n \"type\": \"string\", \"format\": \"date\", \"description\": \"End date for the weather forecast in the format YYYY-MM-DD (UTC)\", \"temperature_unit\": {\n \"type\": \"string\", \"enum\": [\n \"celsius\", \"fahrenheit\"], \"default\": \"celsius\", \"required\": [\"location\", \"start_date\"], \"additionalProperties\": false, \"\$schema\": \"http://json-schema.org/draft-07/schema#\"}\n }\n }\n }\n }\n }\n }\n }\n }\n }\n }\n}\n\nValidation Errors: [\n {\n \"instancePath\": \"\", \"schemaPath\": \"#/required\", \"keyword\": \"required\", \"params\": {\n \"missingProperty\": \"start_date\", \"message\": \"must have required property 'start_date'\"\n }\n }\n]\n\nHINT: If you're convinced that the input was correct but the function cannot process it then use a different function or say I don't know.\n\",

16

"role": "assistant"

17

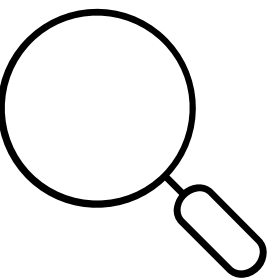
},

18

{

19

"content": "Thought: The OpenMeteo function requires a start date for the weather forecast, even if it's the current date. I will now use the OpenMeteo function again, this time with the required start date.\n\nFunction Name: OpenMeteo\nFunction Input: {\n \"location\": {\n \"name\": \"Rome\", \"country\": \"Italy\", \"start_date\": \"2023-02-18\", \"temperature_unit\": \"celsius\"\n }\n}\nFunction Output: {\n \"current_units\": \"celsius\"\n}\n\nToolInputValidationError: The received tool input does not match the expected schema.\nInput Schema: {\n \"type\": \"object\", \"properties\": {\n \"location\": {\n \"type\": \"object\", \"properties\": {\n \"name\": {\n \"type\": \"string\", \"minLength\": 1, \"country\": {\n \"type\": \"string\", \"language\": {\n \"type\": \"string\", \"default\": \"English\", \"required\": [\"name\"], \"additionalProperties\": false,\n \"type\": \"object\", \"properties\": {\n \"latitude\": {\n \"type\": \"number\", \"required\": [\"latitude\", \"longitude\"], \"additionalProperties\": false,\n \"type\": \"string\", \"format\": \"date\", \"description\": \"Start date for the weather forecast in the format YYYY-MM-DD (UTC)\", \"end_date\": {\n \"type\": \"string\", \"format\": \"date\", \"description\": \"End date for the weather forecast in the format YYYY-MM-DD (UTC)\", \"temperature_unit\": {\n \"type\": \"string\", \"enum\": [\n \"celsius\", \"fahrenheit\"], \"default\": \"celsius\", \"required\": [\"location\", \"start_date\"], \"additionalProperties\": false, \"\$schema\": \"http://json-schema.org/draft-07/schema#\"}\n }\n }\n }\n }\n }\n }\n }\n }\n }\n }\n}\n\nValidation Errors: [\n {\n \"instancePath\": \"\", \"schemaPath\": \"#/required\", \"keyword\": \"required\", \"params\": {\n \"missingProperty\": \"start_date\", \"message\": \"must have required property 'start_date'\"\n }\n }\n]\n\nHINT: If you're convinced that the input was correct but the function cannot process it then use a different function or say I don't know.\n\",



```
[\"celsius\", \"fahrenheit\"], \"default\": \"celsius\"}}, \"required\":  
[\"location\", \"start_date\"], \"additionalProperties\": false, \"$schema\": \"http://json-schema.org/draft-07/schema#\"}\n\nValidation Errors: [\n  {\n    \"instancePath\": \"\", \"schemaPath\": \"#/required\", \"keyword\": \"required\", \"params\": {\n      \"missingProperty\": \"start_date\", \"message\": \"must have required property 'start_date'\"\n    }\n  }\n]\n\nHINT: If you're convinced that the input was correct but the function cannot process it then use a different function or say I don't know.\n\",  
16     \"role\": \"assistant\"  
17   },  
18   {  
19     \"content\": \"Thought: The OpenMeteo function requires a start date for the weather forecast, even if it's the current  
date. I will now use the OpenMeteo function again, this time with the required start date.\n\nFunction Name: OpenMeteo\nFunction  
Input: {\n  \"location\": {\n    \"name\": \"Rome\", \"country\": \"Italy\", \"start_date\": \"2023-02-  
18\", \"temperature_unit\": \"celsius\"\n  }\n}\nFunction Output: {\n  \"current_units\":
```


Travel Agent – First Example

Output is correct, but how can the agent know the current date?

Task name

0.00 s

trace

0.01s

iteration-1

0.01s

agent.bee.start-1

0.01s

watsonx.chat_llm.start-

0.01s

watsonx.chat_llm.start-

0.01s

agent.bee.partialUpdate-52

0.01s

tables, images, files.\n- Sometimes, things don't go as planned. Functions may not provide useful information on the first few tries. You should always try a few different approaches before declaring the problem unsolvable.\n- When the function doesn't give you what you were asking for, you must either use another function or a different function input.\n- When using search engines, you try different formulations of the query, possibly even in a different language.\n- You cannot do complex calculations, computations, or data manipulations without using functions.\n\n# Role\nYou are a helpful assistant.",

4

"role": "system"

5

},

6

{

7

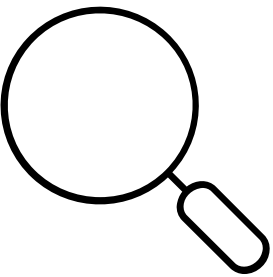
"content": "Message: What is the best vacation city in Europe, and can you tell me the current temperature in this city?"

8

"role": "user"

9

},



7 "content": "Message: What is the best vacation city in Europe, and can you tell me the current temperature in this city?"
\n\nThis message was created at 2025-02-18T10:41:04.111Z",

Travel Agent – Second Example

Input: “What is the best vacation city in Europe, and can you tell me the typical temperature in this city in summer?”

Output: “The typical temperature in Rome in summer is around 25 degrees Celsius.”

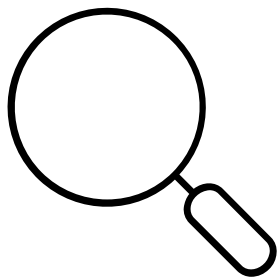
Evaluation: The output is correct, but the weather’s tool output was ignored. Instead, the model returned data from its previous training.

Travel Agent – Second Example

Agent ignores tool output

Task name	0.00 s	
trace	0.01s	
iteration-1	0.01s	
agent.bee.start-1	0.01s	
watsonx.chat_llm.start-	0.01s	
watsonx.chat_llm.start-	0.01s	
agent.bee.partialUpdate-47	0.01s	
agent.bee.update-1	0.01s	

15 "content": "Thought: The best vacation city in Europe is Rome, Italy. Now I need to find the typical temperature in Rome in summer using the OpenMeteo function.\n\nFunction Name: OpenMeteo\nFunction Input: {\n \"location\": {\n \"name\": \"Rome\", \"country\": \"Italy\"},\n \"start_date\": \"2023-06-01\", \"end_date\": \"2023-08-31\", \"temperature_unit\": \"celsius\"}\n }\nFunction Output: {\n \"current_units\": {\n \"time\": \"iso8601\", \"interval\": \"seconds\", \"temperature_2m\": \"°C\", \"rain\": \"mm\", \"relative_humidity_2m\": \"%\", \"wind_speed_10m\": \"km/h\"},\n \"current\": {\n \"time\": \"2025-02-18T12:15\", \"interval\": 900, \"temperature_2m\": 14.8, \"rain\": 0, \"relative_humidity_2m\": 56, \"wind_speed_10m\": 3.4},\n \"daily_units\": {\n \"time\": \"iso8601\", \"temperature_2m_max\": \"°C\", \"temperature_2m_min\": \"°C\", \"rain_sum\": \"mm\"},\n \"daily\": {\n \"time\": [\"2023-06-01\", \"2023-06-02\", \"2023-06-03\", \"2023-06-04\", \"2023-06-05\", \"2023-06-06\", \"2023-06-07\", \"2023-06-08\", \"2023-06-09\", \"2023-06-10\", \"2023-06-11\", \"2023-06-12\", \"2023-06-13\", \"2023-06-14\", \"2023-06-15\", \"2023-06-16\", \"2023-06-17\", \"2023-06-18\", \"2023-06-19\", \"2023-06-20\", \"2023-06-21\", \"2023-06-22\", \"2023-06-23\", \"2023-06-24\", \"2023-06-25\", \"2023-06-26\", \"2023-06-27\", \"2023-06-28\", \"2023-06-29\", \"2023-06-30\", \"2023-07-01\", \"2023-07-02\", \"2023-07-03\", \"2023-07-04\", \"2023-07-05\", \"2023-07-06\", \"2023-07-07\", \"2023-07-08\", \"2023-07-09\", \"2023-07-10\", \"2023-07-11\", \"2023-07-12\", \"2023-07-13\", \"2023-07-14\", \"2023-07-15\", \"2023-07-16\", \"2023-07-17\", \"2023-07-18\", \"2023-07-19\", \"2023-07-20\", \"2023-07-21\", \"2023-07-22\", \"2023-07-23\", \"2023-07-24\", \"2023-07-25\", \"2023-07-26\", \"2023-07-27\", \"2023-07-28\", \"2023-07-29\", \"2023-07-30\", \"2023-07-31\"]\n }\n }\n }\n }\n}



{\n \"name\": \"Rome\", \"country\": \"Italy\"},\n \"start_date\": \"2023-06-01\", \"end_date\": \"2023-08-31\",\n \"temperature_unit\": \"celsius\"}\nFunction Output: {\n \"current_units\": {\n \"time\": \"iso8601\", \"interval\": \"seconds\", \"temperature_2m\": \"°C\", \"rain\": \"mm\", \"relative_humidity_2m\": \"%\", \"wind_speed_10m\": \"km/h\"},\n \"current\": {\n \"time\": \"2025-02-18T12:15\", \"interval\": 900, \"temperature_2m\": 14.8, \"rain\": 0, \"relative_humidity_2m\": 56, \"wind_speed_10m\": 3.4},\n \"daily_units\": {\n \"time\": \"iso8601\", \"temperature_2m_max\": \"°C\", \"temperature_2m_min\": \"°C\", \"rain_sum\": \"mm\"},\n \"daily\": {\n \"time\": [\"2023-06-01\", \"2023-06-02\", \"2023-06-03\", \"2023-06-04\", \"2023-06-05\", \"2023-06-06\", \"2023-06-07\", \"2023-06-08\", \"2023-06-09\", \"2023-06-10\", \"2023-06-11\", \"2023-06-12\", \"2023-06-13\", \"2023-06-14\", \"2023-06-15\", \"2023-06-16\", \"2023-06-17\", \"2023-06-18\", \"2023-06-19\", \"2023-06-20\", \"2023-06-21\", \"2023-06-22\", \"2023-06-23\", \"2023-06-24\", \"2023-06-25\", \"2023-06-26\", \"2023-06-27\", \"2023-06-28\", \"2023-06-29\", \"2023-06-30\", \"2023-07-01\", \"2023-07-02\", \"2023-07-03\", \"2023-07-04\", \"2023-07-05\", \"2023-07-06\", \"2023-07-07\", \"2023-07-08\", \"2023-07-09\", \"2023-07-10\", \"2023-07-11\", \"2023-07-12\", \"2023-07-13\", \"2023-07-14\", \"2023-07-15\", \"2023-07-16\", \"2023-07-17\", \"2023-07-18\", \"2023-07-19\", \"2023-07-20\", \"2023-07-21\", \"2023-07-22\", \"2023-07-23\", \"2023-07-24\", \"2023-07-25\", \"2023-07-26\", \"2023-07-27\", \"2023-07-28\", \"2023-07-29\", \"2023-07-30\", \"2023-07-31\"]\n }\n }\n }\n }\n}

tool.start-1	0.01s	
watsonx.llm.start-	0.01s	
watsonx.llm.succe	0.01s	
watsonx.llm.finish-	0.01s	
watsonx.llm.start-	0.01s	

16 "role": "assistant"
17 },
18 {
19 "content": "The typical temperature in Rome in summer is around 25 degrees Celsius.",
20 "role": "assistant"
21 }
22]

Travel Agent – Second Example

Agent ignores tool output

Task name	0.00 s	
trace	0.01s	
iteration-1	0.01s	
agent.bee.start-1	0.01s	
watsonx.chat_llm.start	0.01s	
watsonx.chat_llm.start	0.01s	
agent.bee.partialUpdate-47	0.01s	
agent.bee.update-1	0.01s	
agent.bee.partialUpdate-51	0.01s	
agent.bee.update-2	0.01s	

```
19     "content": "The typical temperature in Rome in summer is around 25 degrees Celsius.",
20     "role": "assistant"
```

[illegible]

Travel Agent – Third Example

Input: “What impact does the climate change have on Berlin during my trip?”

Output: “Based on the weather forecast, your trip to Berlin will be quite cold, with temperatures ranging from -8.6°C to 12.1°C. There will be some rainfall, with a maximum of 4.7mm on February 25. Unfortunately, I don't have enough information about the historical weather patterns in Berlin to determine the impact of climate change on your trip.”

Evaluation: The agent didn't realize this is a trick question and used the weather tool anyway.

Travel Agent – Third Example

Agent invokes weather tool

Task name

0.00 s

trace

0.01s

iteration-1

0.01s

agent.bee.start-1

0.01s

watsonx.chat_llm.start

0.01s

watsonx.chat_llm.start

0.01s

agent.bee.partialUpdate-45

0.01s

agent.bee.update-1

0.01s

agent.bee.partialUpdate-51

0.01s

agent.bee.update-2

0.01s

agent.bee.partialUpdate-10

0.01s

watsonx.chat_llm.newT

0.01s

agent.bee.update-3

0.01s

11

"content": "Thought: The user wants to know the impact of climate change on Berlin during their trip. I can use the OpenMeteo function to retrieve the weather forecast for Berlin and then analyze it to determine the impact of climate change.\nFunction Name: OpenMeteo\nFunction Input: {\"location\": {\"name\": \"Berlin\", \"country\": \"Germany\"}, \"start_date\": \"2025-02-18\", \"end_date\": \"2025-02-25\", \"temperature_unit\": \"celsius\"}\nFunction Output: {\"current_units\": {\"time\": \"iso8601\", \"interval\": \"seconds\", \"temperature_2m\": \"°C\", \"rain\": \"mm\", \"relative_humidity_2m\": \"%\", \"wind_speed_10m\": \"km/h\"}, \"current\": {\"time\": \"2025-02-18T12:30\", \"interval\": 900, \"temperature_2m\": -1, \"rain\": 0, \"relative_humidity_2m\": 49, \"wind_speed_10m\": 7.6}, \"daily_units\": {\"time\": \"iso8601\", \"temperature_2m_max\": \"°C\", \"temperature_2m_min\": \"°C\", \"rain_sum\": \"mm\"}, \"daily\": {\"time\": [\"2025-02-18\", \"2025-02-19\", \"2025-02-20\", \"2025-02-21\", \"2025-02-22\", \"2025-02-23\", \"2025-02-24\", \"2025-02-25\"], \"temperature_2m_max\": [-0.5, 1.1, 1.2, 3.5, 5.4, 11.3, 12.1, 11.4], \"temperature_2m_min\": [-8.6, -5.8, -5.5, -2, -0.2, 1.5, 7.3, 7], \"rain_sum\": [0, 0, 0, 1.3, 0.4, 3.8, 3.7, 4.7]}}\n\",

12

"role": "assistant"

13

},

14

{

15

"content": "Based on the weather forecast, your trip to Berlin will be quite cold, with temperatures ranging from -8.6°C to 12.1°C. There will be some rainfall, with a maximum of 4.7mm on February 25. Unfortunately, I don't have enough information about the historical weather patterns in Berlin to determine the impact of climate change on your trip. However, it is generally expected that climate change will lead to more extreme weather events, so you may want to be prepared for unexpected changes in the weather.",

16

"role": "assistant"

17

}

18

]

11

"content": "Thought: The user wants to know the impact of climate change on Berlin during their trip. I can use the OpenMeteo function to retrieve the weather forecast for Berlin and then analyze it to determine the impact of climate change.\nFunction Name: OpenMeteo\nFunction Input: {\"location\": {\"name\": \"Berlin\", \"country\": \"Germany\"}, \"start_date\": \"2025-02-18\", \"end_date\": \"2025-02-



11 "content": "Thought: The user wants to know the impact of climate change on Berlin during their trip. I can use the OpenMeteo function to retrieve the weather forecast for Berlin and then analyze it to determine the impact of climate change.\nFunction Name: OpenMeteo\nFunction Input: {\"location\": {\"name\": \"Berlin\", \"country\": \"Germany\"}, \"start_date\": \"2025-02-18\", \"end_date\": \"2025-02-

Demo – Simplified Architecture



Cloud

Desktop



BeeAI Framework

Open-source project from IBM Research

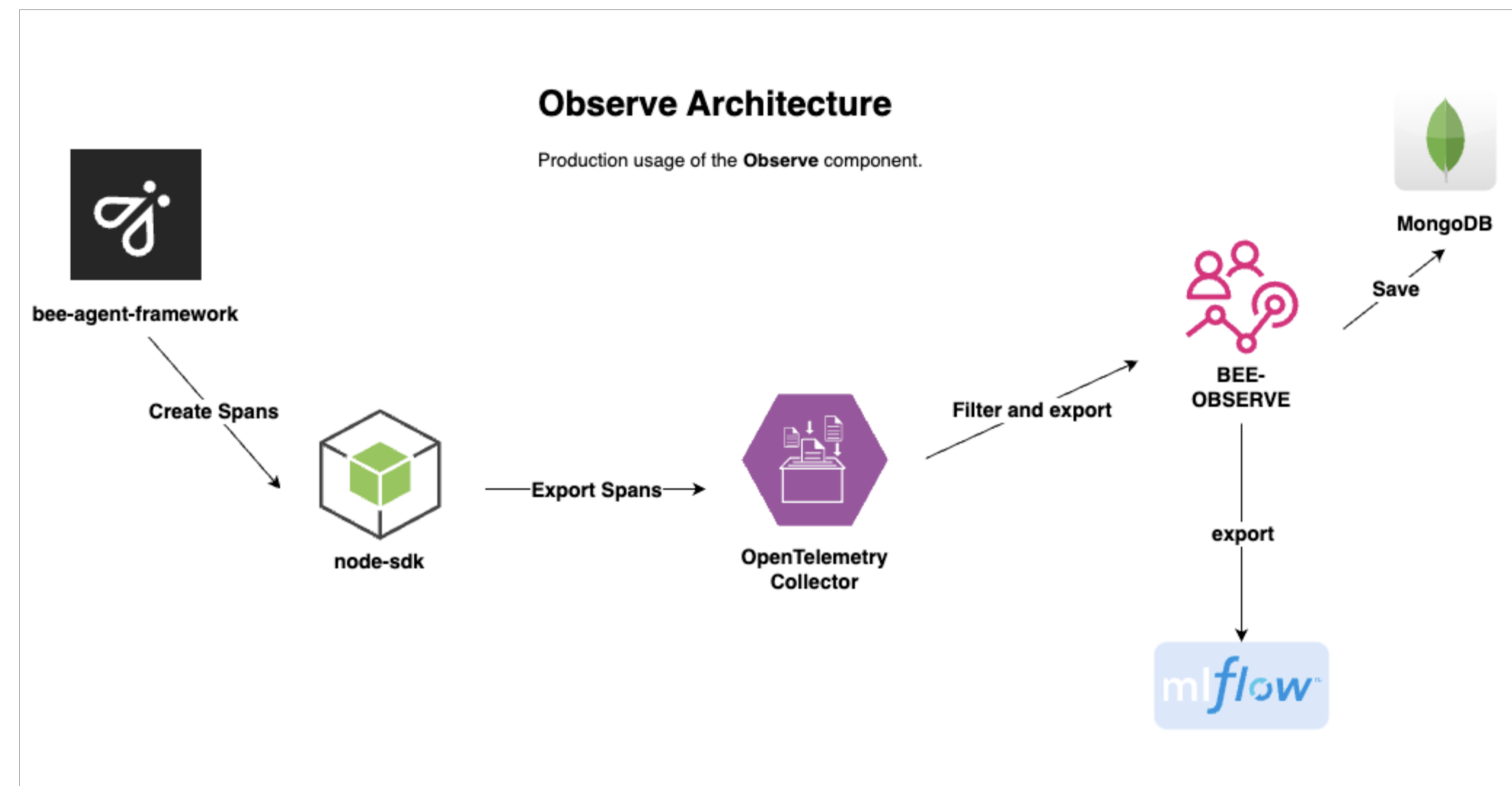
Support for Python and TypeScript

Strength: Observability

OpenTelemetry, Arize Phoenix, MLflow
and more



Powering the future of open-source AI agent development.



Agent Lab in IBM watsonx.ai – Trajectory

IBM watsonx

Upgrade ⓘ 🔔 2956389 - watsonx-dach-... Dallas NH

Projects / L3 / Agent Lab

Share feedback Unsaved ⌵ Deploy ↗ New agent +

Build

AI Model: llama-3-3-70b-instruct ⌵ ⚙️ <

Setup ⌵

Configuration ⌶

Framework

LangGraph ⌵

Architecture

ReAct ⌵

Instructions

You are a helpful assistant that uses tools to answer questions in detail.
When greeted, say "Hi, I am watsonx.ai agent. How can I help you?"

Tools

Add a tool

Create custom tool

Added tools (1)

Weather

Retrieve the weather of a city.

Agent preview

NH You: 02:25 PM

How will the weather be in Berlin this weekend?

🤖 watsonx Agent 02:25 PM

The weather in Berlin this weekend is expected to be cold with a temperature of 0.3°C, no rain, and a relative humidity of 49%. The wind speed will be 8.3km/h.

▼ How did I get this answer?

1: {"name":"Berlin","country":"Germany"} ✓

Tool: Weather

Input: {"name":"Berlin","country":"Germany"}

Output:

Current weather in Berlin, Germany:
Temperature: 0.3°C
Rain: 0mm
Relative humidity: 49%
Wind: 8.3km/h

View full trace details

2: Generating final answer ✓

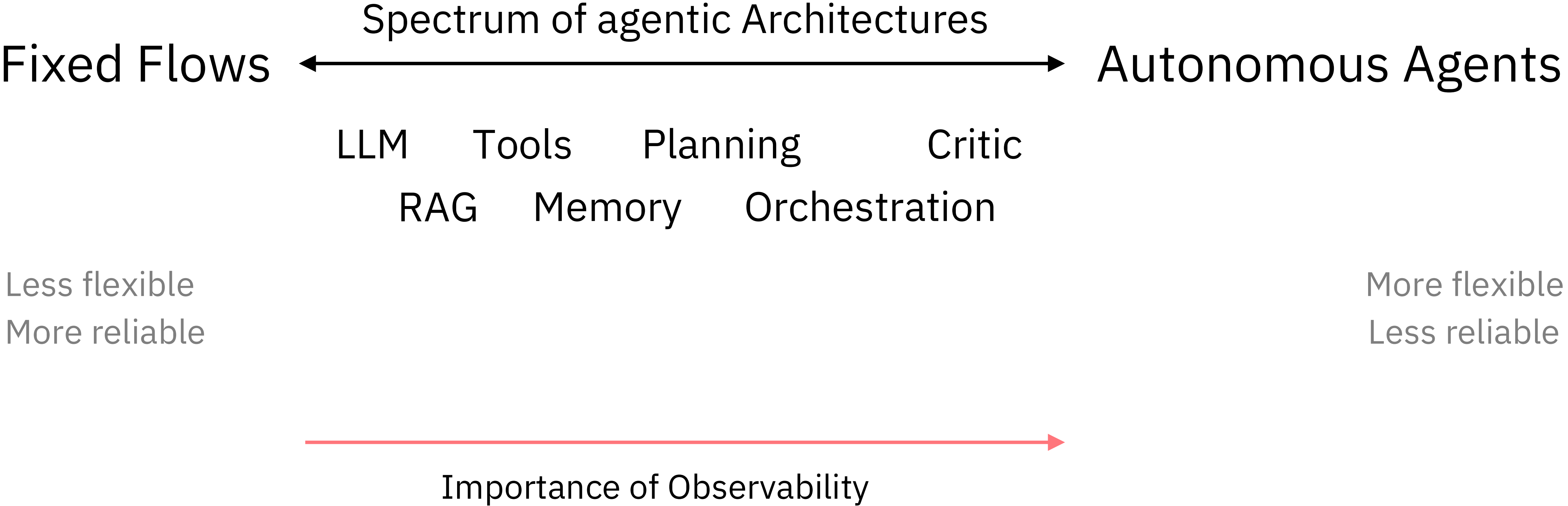
Tool: meta-llama/llama-3-3-70b-instruct

Input: Generating final answer

Steps created by Agent

Type something...

Definition of Agents and agentic Systems



Observability Topics

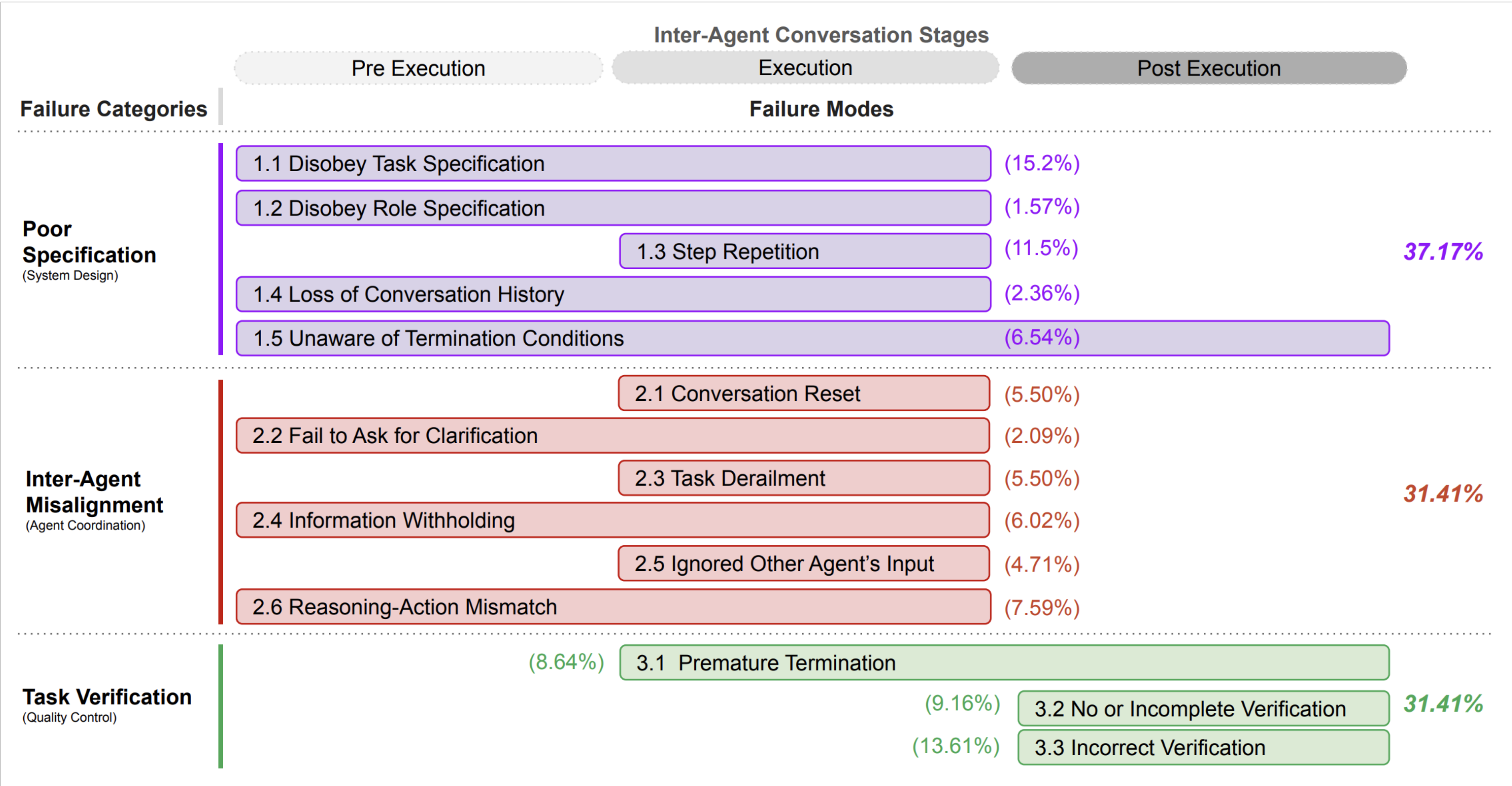


Why do Agents and agentic Systems fail?

Top 5 Reasons from a Development Perspective

1. AI != Classic programming
2. Insufficient observability and transparency
3. Tool calling failures
4. Planning/reasoning failures and infinite looping
5. Missing critic agent

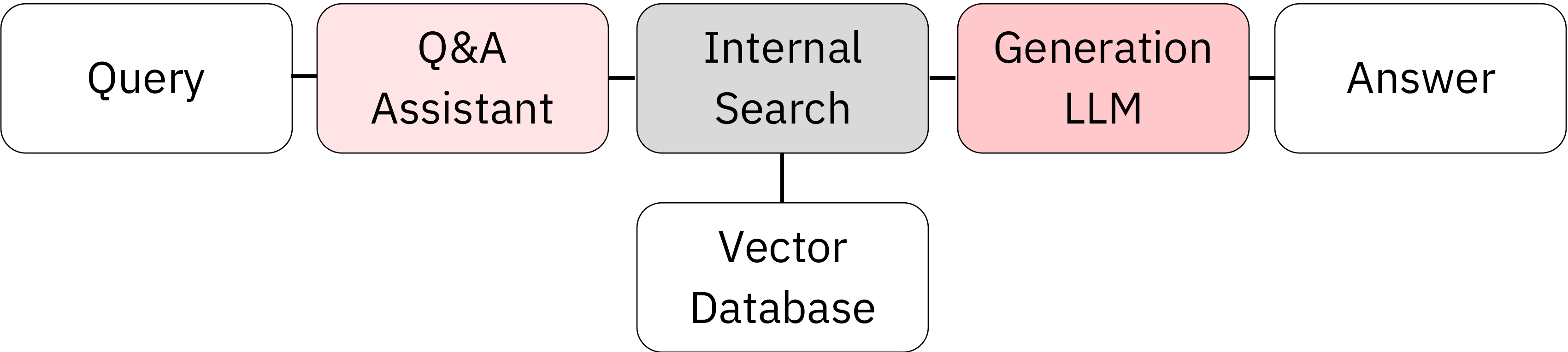
Agentic System Failures in different Stages



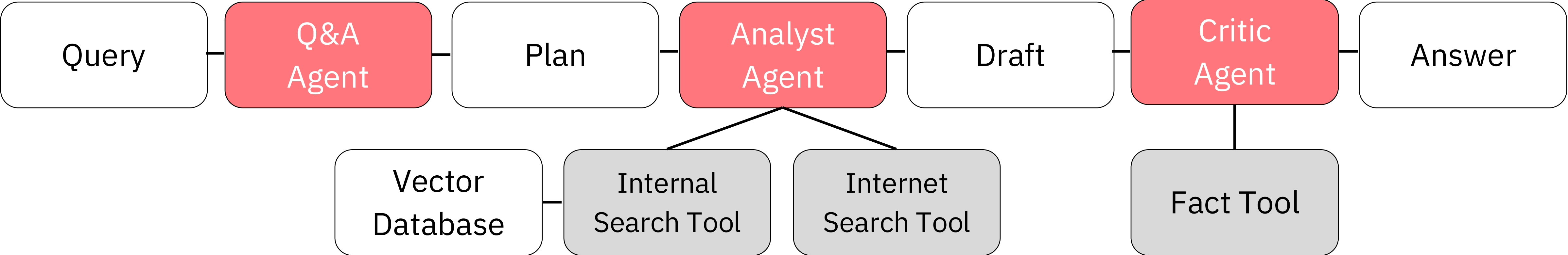
Why Do Multi-Agent LLM Systems Fail?
University of California, Berkeley
17.03.2025
<https://arxiv.org/pdf/2503.13657>

Pattern: Critic Agent

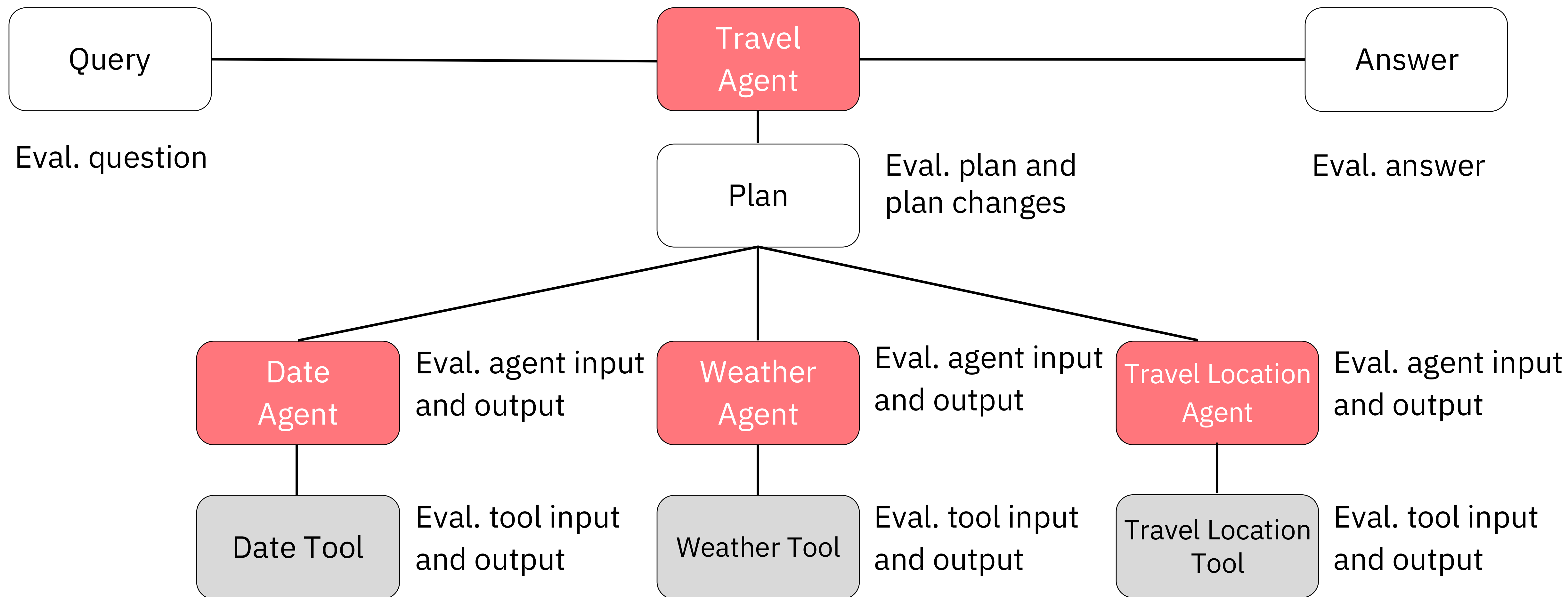
Classic RAG



Agentic RAG



Pattern: Evaluate intermediate Results and final Answer



What’s next?

Agentic system metrics and traceability

Catalogs of agents and tools

Agent Communication Protocol

Production monitoring

Governance for agents

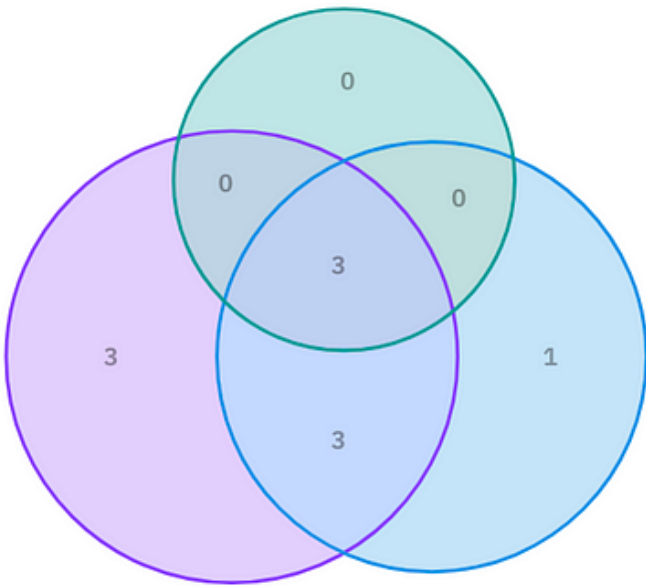
More to come ...

Answer quality

Answer quality metrics evaluates the quality of the model answer in terms of faithfulness, relevance and success ratio. [Learn more](#)

Evaluation at: Dec 4, 2024, 9:04 PM

Threshold violation distribution



Select metrics

3 × Faithfulness, Answer relevance, Unsuccessful requests

Selected metrics

- ☒ Faithfulness
- ☒ Answer relevance
- ☒ Unsuccessful requests

17 out of 17 transactions were used for calculation

Question	Answer	context	Faithfulness	Answer relevance
From the list of beneficiaries, I am not able to see	Please use 'raise a query' option available in ...	Conditions of Failure: Digital Banking. My ben...	0.8	1
If a transaction is not credited to the beneficiary a	Yes. If the beneficiary's bank is unable to cre...	Conditions of Failure: Digital Banking 1 My be...	0.8	1
How do I enroll in paperless statements in the mo	To update your billing preferences, please na...	35 Is there a limit on using mobile check dep...	0	1
What are the risks in investing mutual funds?	I don't know the answer to this question.	35 Is there a limit on using mobile check dep...	0.2	0.2
What are floating charges?	Floating charges refer to the handling fees th...	How can I inquire about account balance afte...	0.2	0.2

Items per page: 5 1–5 of 17 items

Call to Action and Questions

Call to Action

Try out agentic frameworks and observability tools

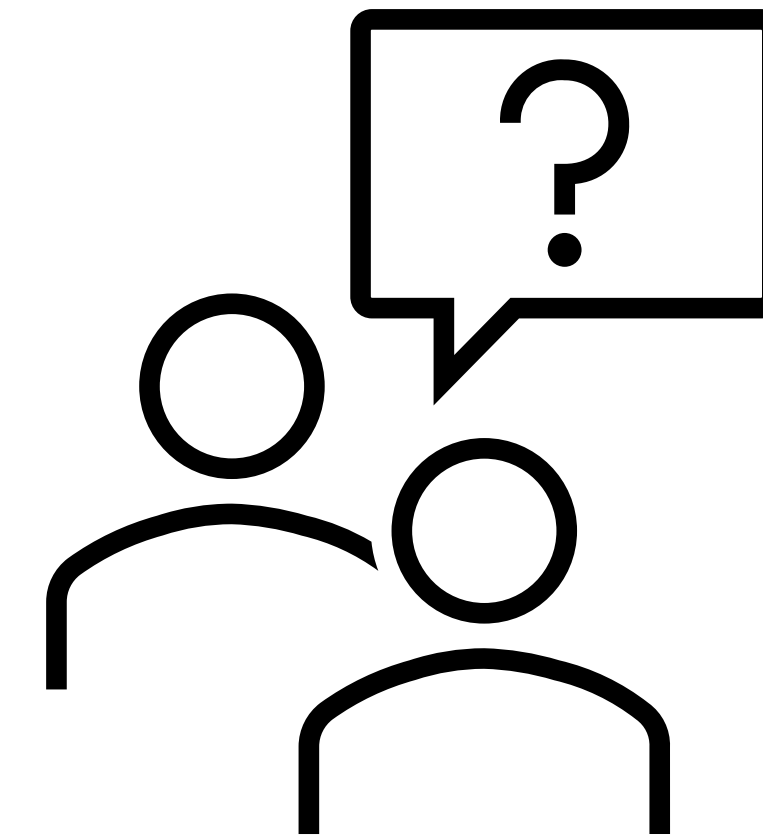
Build your own use cases, start simple

Contact IBM to define use cases, build pilots and deploy solutions

Read “Agents – Expectations vs. reality”

<https://ibm.biz/agents-expectations-reality>

Thank you!



niklas_Heidloff@de.ibm.com

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

<https://heidloff.net/>