

# From Inception to Productization: Hands-on Lab for the Lifecycle of Multimodal Agentic AI in Industry 4.0

Chathurangi Shyalika  
Saumya Ahuja  
Shuxin Lin  
Dr. Dhaval Patel  
Prof. Amit Sheth

*IBM Research  
Artificial Intelligence Institute, University of South Carolina*



**AAAI-26 / IAAI-26 / EAAI-26**  
JANUARY 20-27, 2026 | SINGAPORE

January 21<sup>st</sup>, 2025



# Speakers



**Chathurangi  
Shyalika**

Artificial Intelligence  
Institute, University of  
South Carolina



**Saumya Ahuja**

IBM Research



**Shuxin Lin**

IBM Research



**Dr. Dhaval Patel**

IBM Research



**Prof. Amit Sheth**

Artificial Intelligence Institute,  
University of South Carolina

# Objectives of the lab

- Understand the end-to-end workflow for industrial asset monitoring & maintenance.
- Learn how AssetOpsBench benchmarks multi-agent LLM systems in this setting.
- Practice running, reading, and critiquing agent trajectories in the HF space.
- Analyze why agents fail and how to use those insights to design better systems.

# Outline of the lab

- Beat 1 – World & pain
- Beat 2 – Vision: agentic AI for Industry 4.0
- Beat 3 – AssetOpsBench
- Beat 4 – Agent architectures, orchestration, and failure modes
- Beat 5 – Ecosystem + hands-on

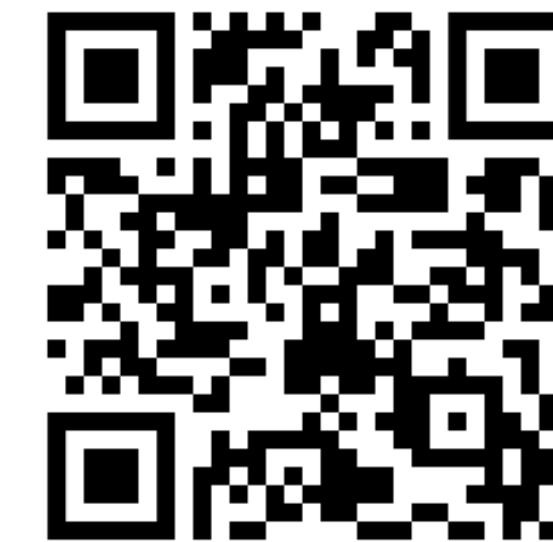
# Resources

AAAI LAB  
Hands-on  
Space

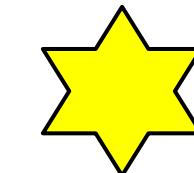


AssetOps-Bench  
Hugging face

AssetOps-Bench  
Guide



AssetOpsBench  
Github Repo



Please star our repo!



AssetOpsBench  
HuggingFace  
Dataset

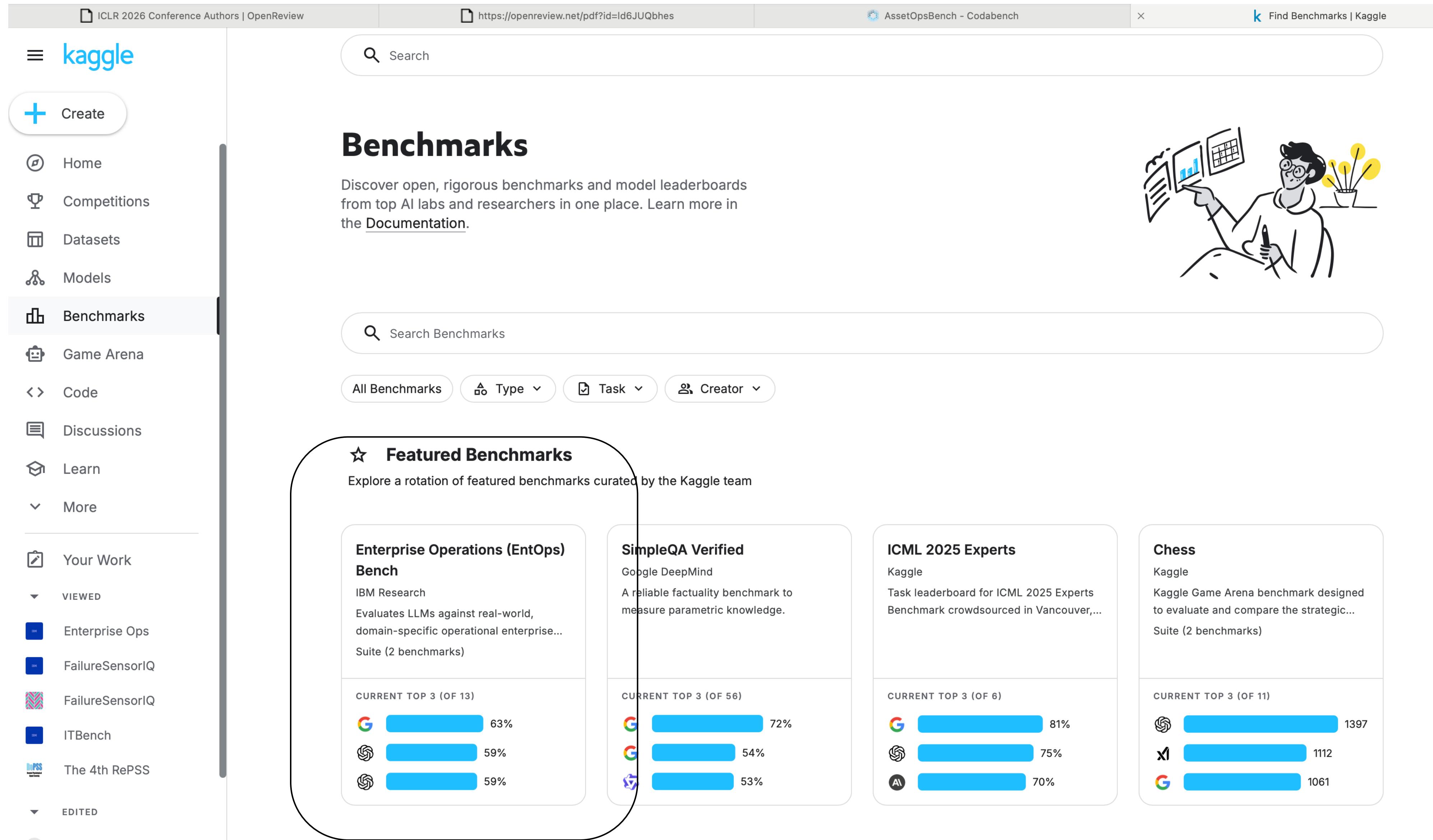


AssetOpsBench  
Codabench  
Competition

# Rise of Enterprise Benchmark

<https://www.kaggle.com/benchmarks>

Special Thanks to Kaggle and IBM Research Team



The screenshot shows the Kaggle Benchmarks page. On the left, a sidebar menu includes options like Home, Competitions, Datasets, Models, Benchmarks (which is selected), Game Arena, Code, Discussions, Learn, and More. Below these are sections for Your Work, VIEWED, and EDITED, each with a list of items such as Enterprise Ops, FailureSensorIQ, and ITBench.

The main content area features a search bar at the top. Below it is a large section titled "Benchmarks" with a sub-section "Featured Benchmarks". This section highlights two benchmarks: "Enterprise Operations (EntOps) Bench" by IBM Research and "SimpleQA Verified" by Google DeepMind. Each card shows a brief description, the current top 3 performers, and their scores. To the right, there are cards for "ICML 2025 Experts" and "Chess", each with its own set of leaderboards.

At the bottom right of the page is a cartoon illustration of a person sitting at a desk, looking at two computer monitors displaying charts and graphs.

CURRENT TOP 3 (OF 13)	
	63%
	59%
	59%

CURRENT TOP 3 (OF 56)	
	72%
	54%
	53%

CURRENT TOP 3 (OF 6)	
	81%
	75%
	70%

CURRENT TOP 3 (OF 11)	
	1397
	1112
	1061

- Beat 1 – World & pain Cold open: why care?
- Beat 2 – Vision: agentic AI for Industry 4.0
- Beat 3 – AssetOpsBench
- Beat 4 – Agent architectures, orchestration, and failure modes
- Beat 5 – Ecosystem + hands-on

# Industrial Assets

We are all surrounded by industrial assets - often without noticing them. These systems form the backbone of critical infrastructure and everyday services:

- **Data Centers:** Chillers, air-handling units, standby generators
- **Hospitals:** Standby generators ensuring uninterrupted operations
- **Energy Generation:** Wind turbines powering renewable electricity
- ....



## Industrial Physical Assets



Chiller



Air-Handing Unit



Wind Turbine



Transportation



Air Production Unit

# Monitoring and Maintenance Tasks

With the emergence of Industry 4.0 applications, these assets are now being monitored using 100s of sensors in real time:

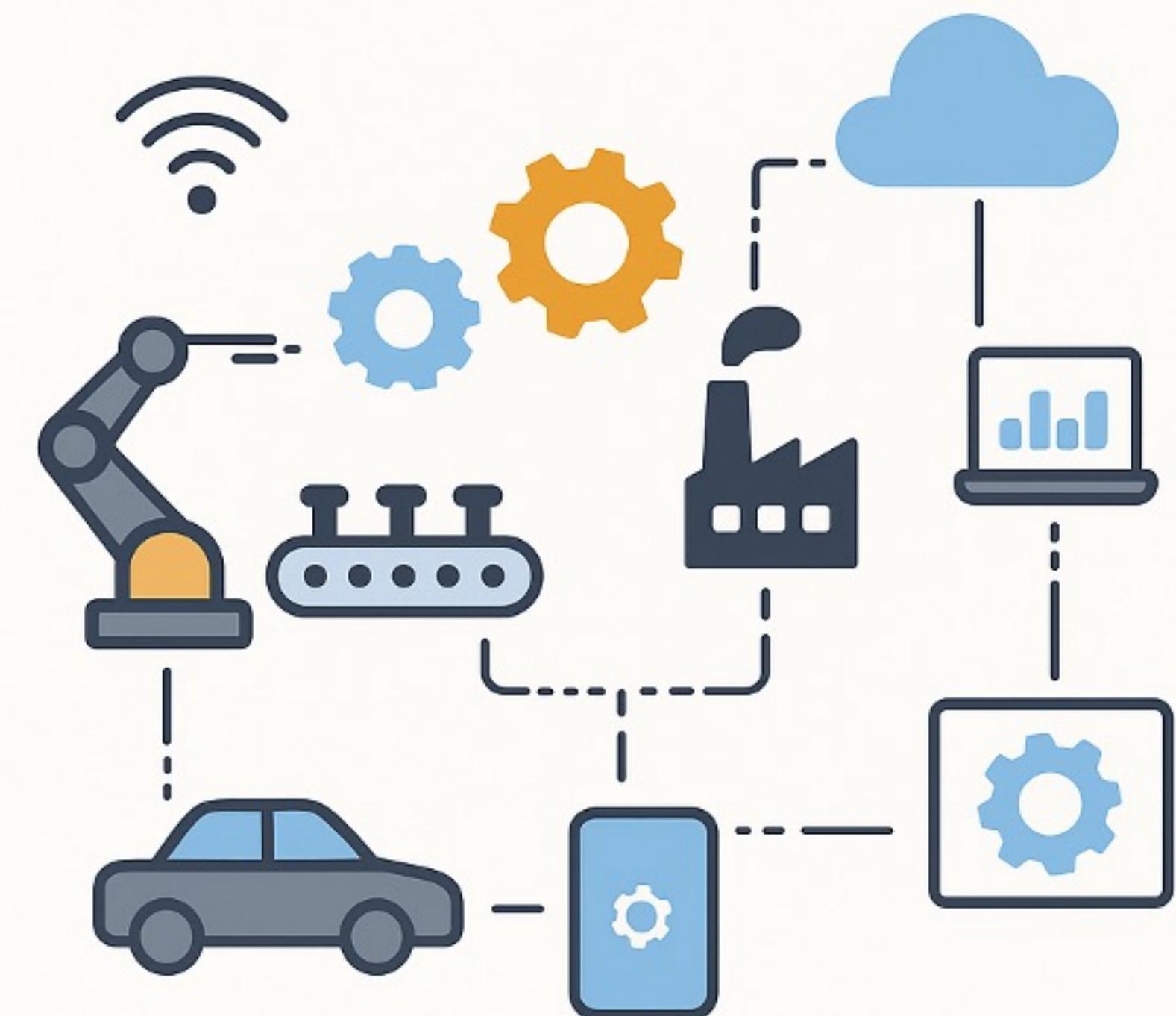
- Temperature,
- Pressure,
- Vibration,
- Power,
- Flow,
- Acoustic,
- Airflow,
- RPM,
- Thermal,
- Gas,
- Visual



*Is Chiller 11's compressor overheating? Generate a service request if needed.*

## Industry 4.0

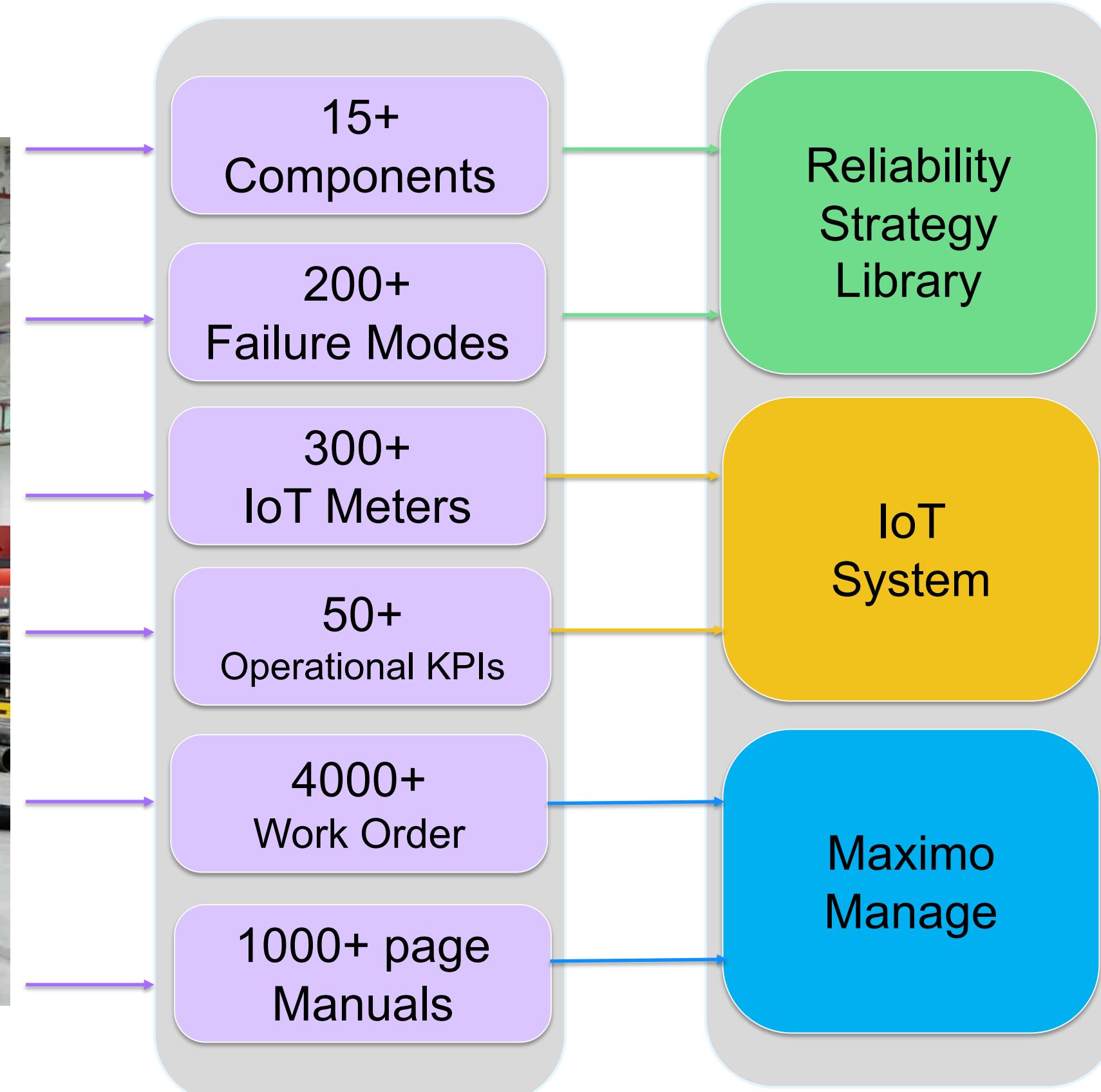
The integration of physical machines with digital technologies like IoT, AI, cloud computing, robotics, and digital twins to create intelligent, automated, and interconnected industrial systems



# AI Agents for Industrial Assets

Industrial assets are complex systems that generate a wealth of data through the monitoring of their lifecycle (ALM)

E.g., Data Center Chiller



This complexity leads to dynamics workflows depending on the alarm/condition of the chiller

Data Diversity and Scale (based on GRE)

Applications Diversity

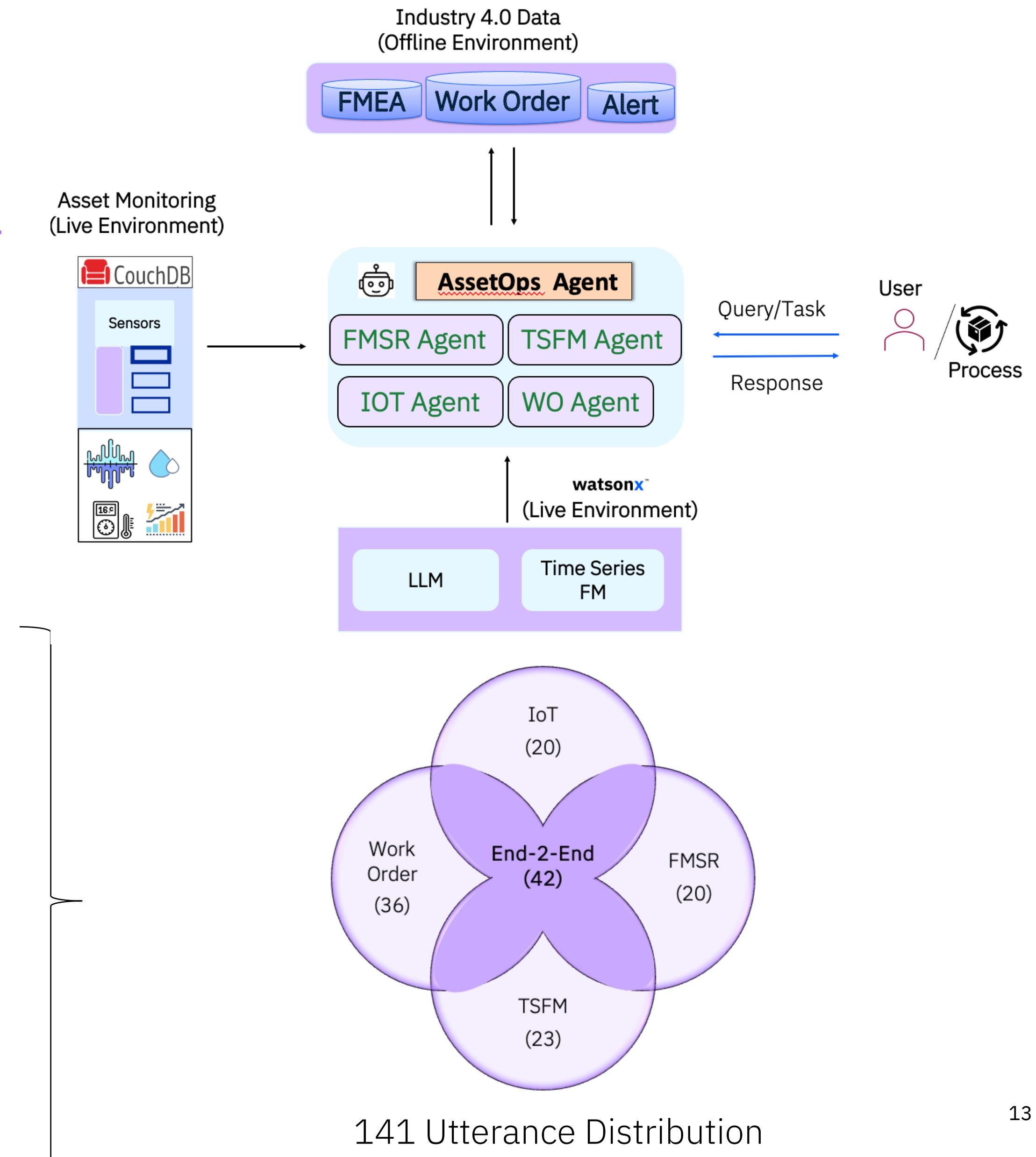
- Beat 1 – World & pain
- Beat 2 – Vision: agentic AI for Industry 4.0
  - From “narrow ML models” to “**agentic workflows**”
- Beat 3 – AssetOpsBench
- Beat 4 – Agent architectures, orchestration, and failure modes
- Beat 5 – Ecosystem + hands-on

# AssetOpsBench : Open-Source Initiative



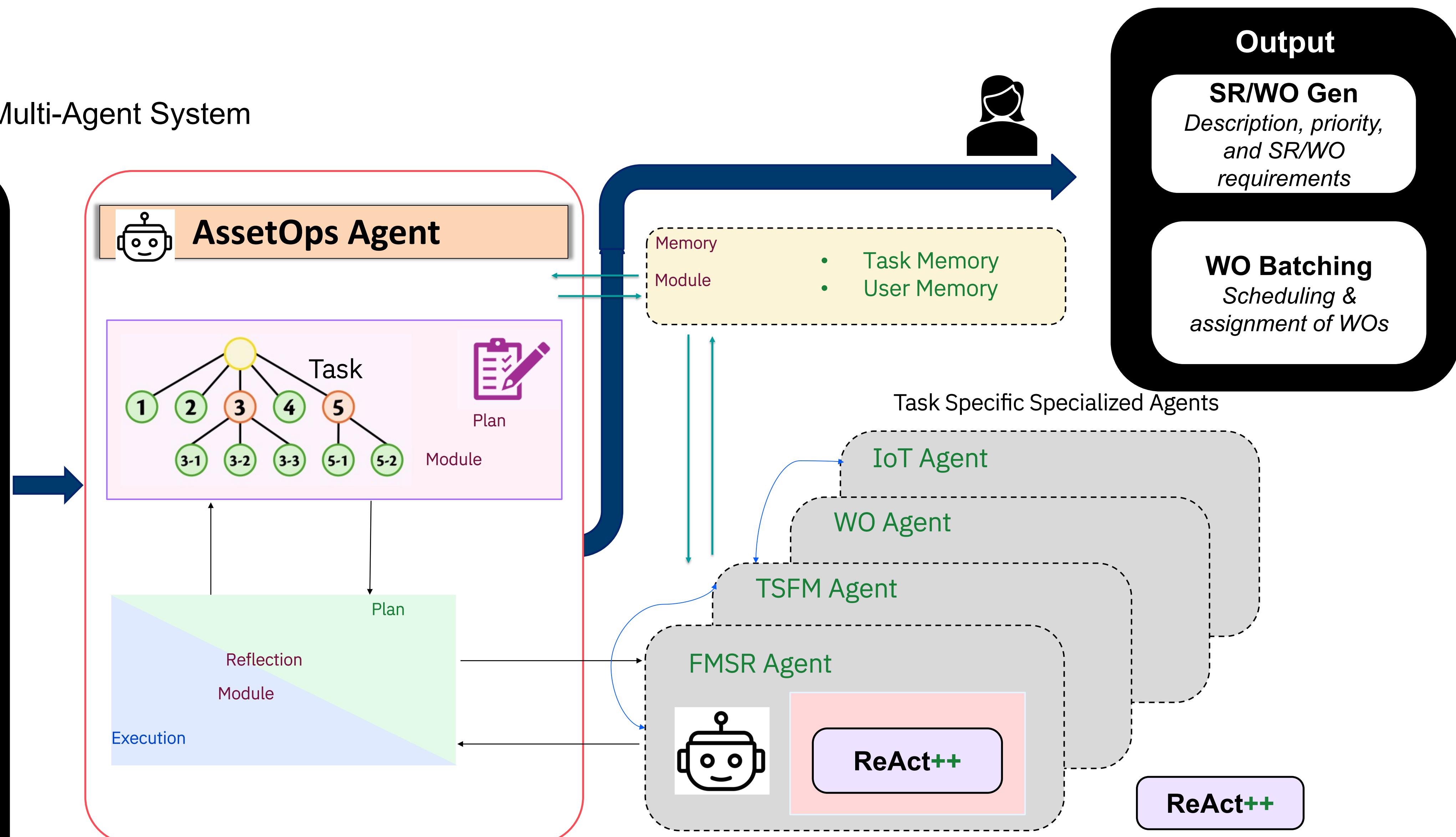
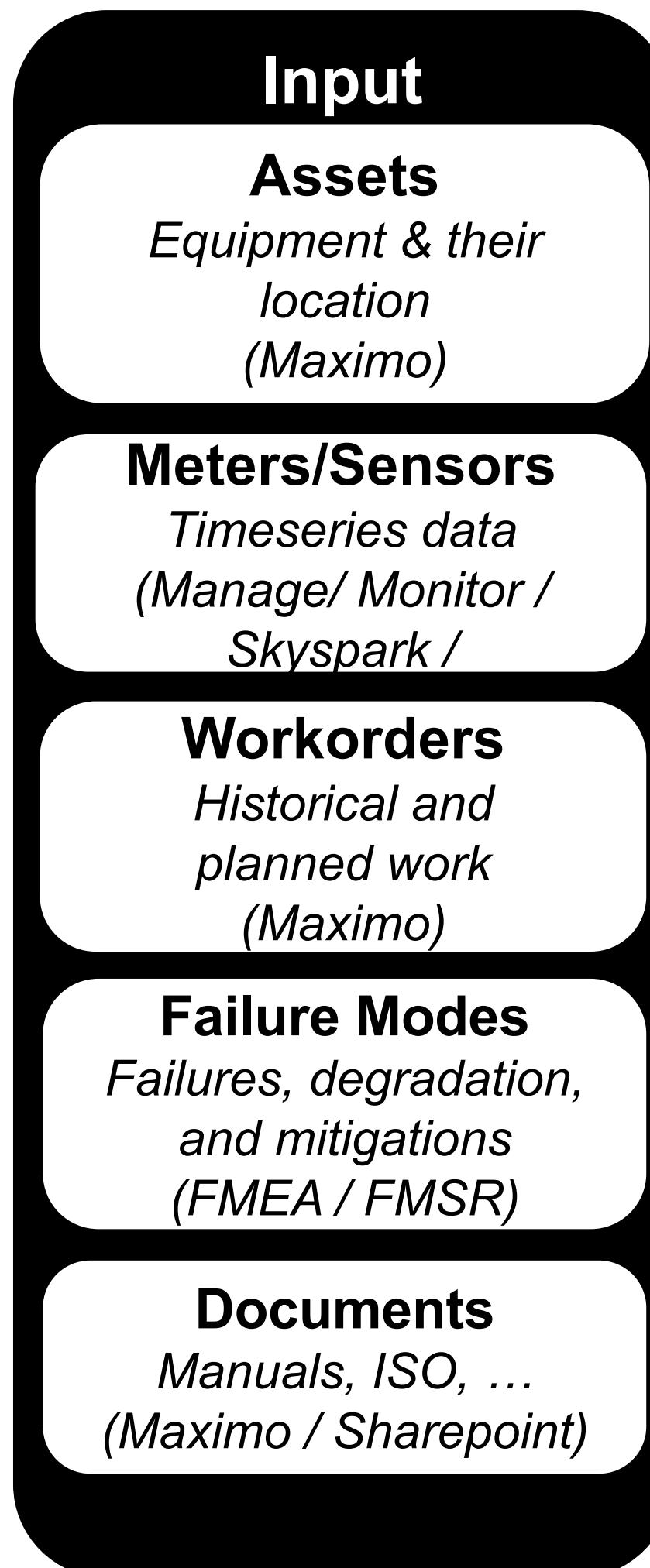
# AssetOpsBench: Open-Source Benchmark for Industry 4.0

- Framework to assess Gen AI solutions' ability to solve Industry 4.0 Automation "Scenarios"
- Simulated industrial environment, 9 multi-source data sets (work orders, FMEAs, timeseries) and 4 domain-specific agents (IoT, data science, work order, failure mode to sensor mapping)
- 140+ expert-authored natural language queries, grounded in enterprise industrial scenarios
- Two Multi-Agent Orchestration Recipes
  - Agent-As-Tool
  - Plan-Execute
- LLM-as-Judge for Rubric-based Agent Evaluation and Reference-based Scoring for Semantic Evaluation
- Agent harness: systematic procedure for automated discovery of emerging failure modes



# Architecture: Agentic Asset Management

## Design Blueprint of Multi-Agent System

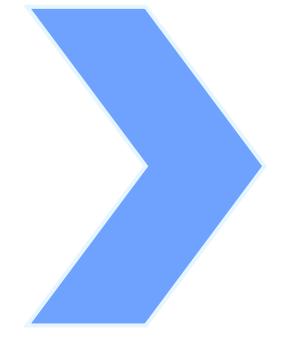


: Novel strategy to enhance ReAct agent

# An Example of Agentic Workflow Patterns in Industrial Setting

Inputs Data:

- FMEAs
- Asset Data
- IoT Meters
- WO History



Agentic  
Workflow



FMSR Agent

*Failure Mode to  
Sensor Relationship  
Mapping*

IoT Agent

*Access to IoT and  
other Time Series  
Data*

TSFM Agent

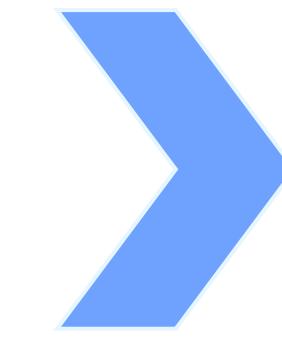
*Automate Time Series  
Analysis using  
pretrained TTM, etc.*

Work Order (WO) Agent

*Automate Workorder  
Analysis & Generation*

Outputs:

- WO Recommendation
- Smart Alerts
- Maintenance Insights



User Query

Example of Tasks → two broad categories of queries:

- Knowledge Query such as “*was there a performance issue over the last month in chiller 9?*”
- Decision Support Query such as – “*recommend a work order based on last three month's health condition of Chiller 9?* (health condition)

Challenges for AI Agent

- Integrate diverse data sources
- Automatically identify key sensors and KPIs.
- Adaptively choose the optimal model for anomaly detection.
- Monitor asset health using historical alerts and anomalies
- Generate actionable maintenance recommendations.

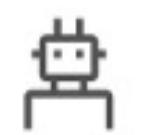
# Task Specific Specialized Agents



IOT Agent

Q\_10\_trajectory\_output.json

Get sensor data for both  
Chiller 6 and Chiller 9's  
Tonnage from last week at  
POKMAIN.



FMSR Agent

Q\_14\_trajectory\_output.json

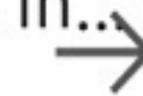
If compressor overheating  
occurs for Chiller 6, which  
sensor should be prioritized  
for monitoring this specific  
failure?



TSFM Agent

Q\_22\_trajectory\_output.json

Explore and use several  
relevant methods to  
analyze time series data  
sensor 'Chiller 9 Condenser  
Water Flow' using data in...



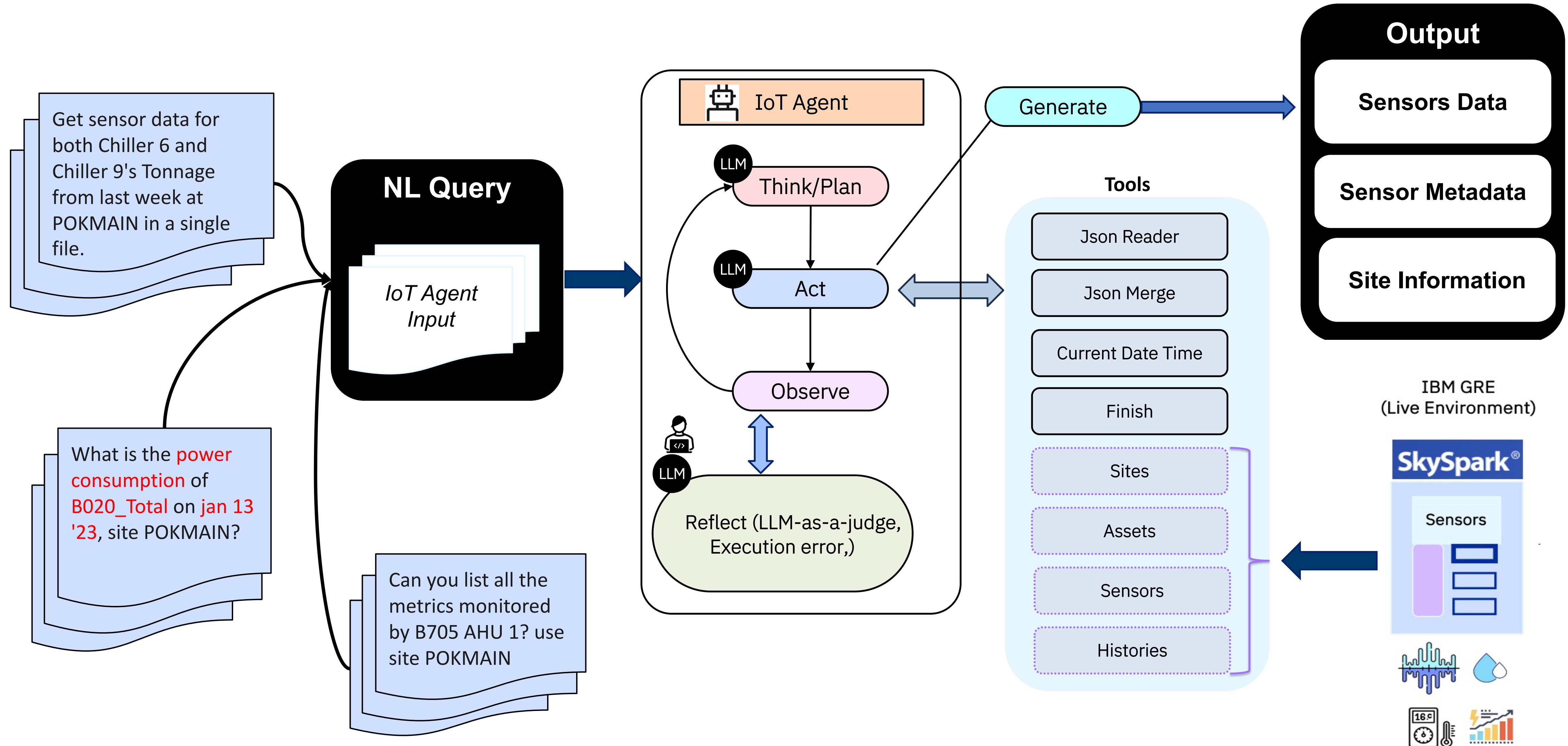
WO Agent

Q\_31\_trajectory\_output.json

Assume today is early of  
July 2020, I would like to  
review the performance of  
chiller 9 with equipment ID  
CU02009 for June 2020...  
→

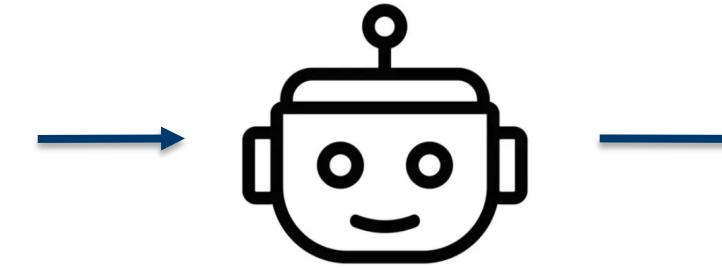
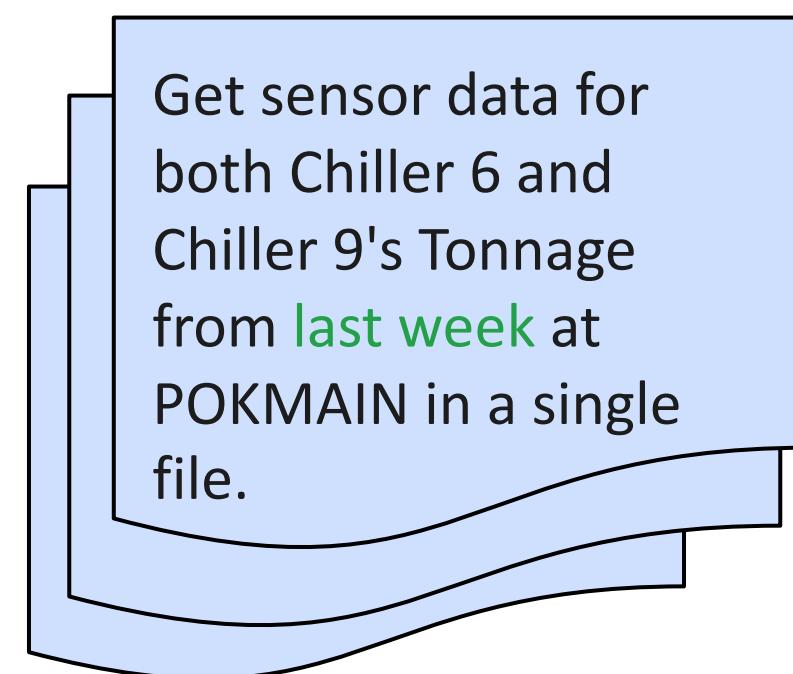
# Architecture: IoT Agent

## From NL to Data

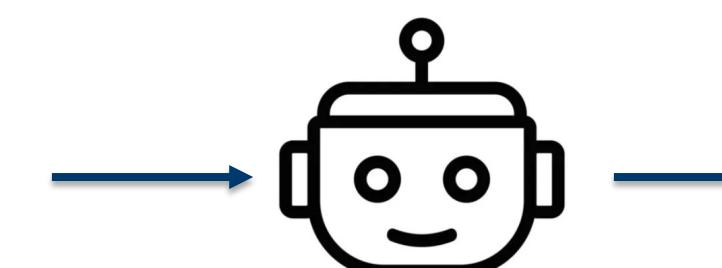
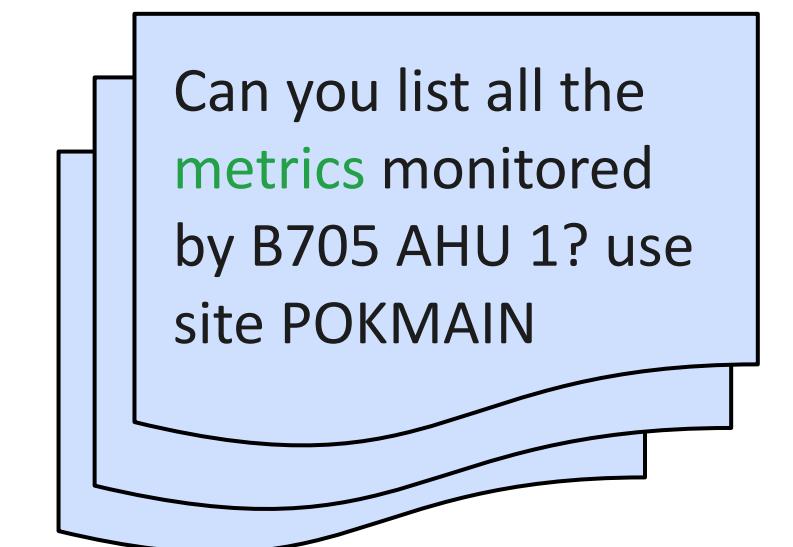


# IoT Agent : Input and Output

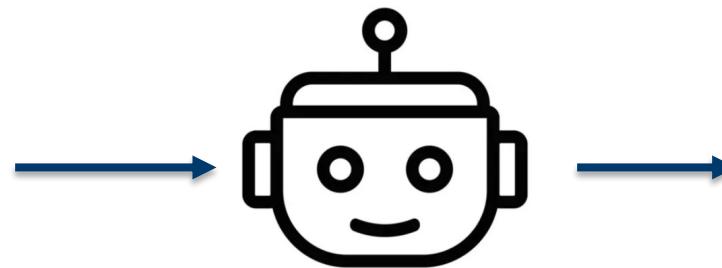
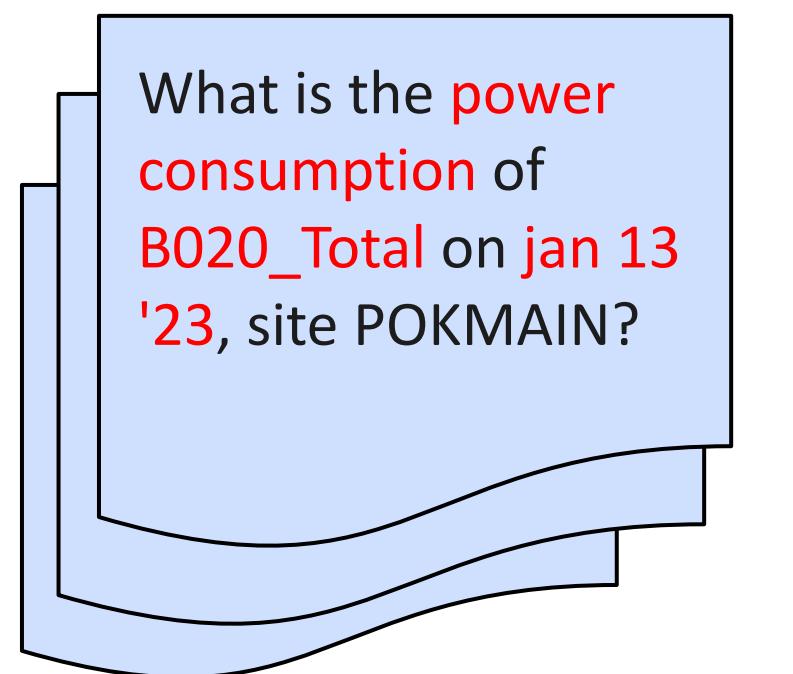
## From NL to Data



The asset history for sensor Chiller 6 Chiller % Loaded on asset Chiller 6 at POKMAIN site from 2020-06-01T00:00:00-04:00 to 2020-06-30T23:59:59-04:00 has been downloaded and is listed in file /var/folders/fz/l1h7gpv96rv

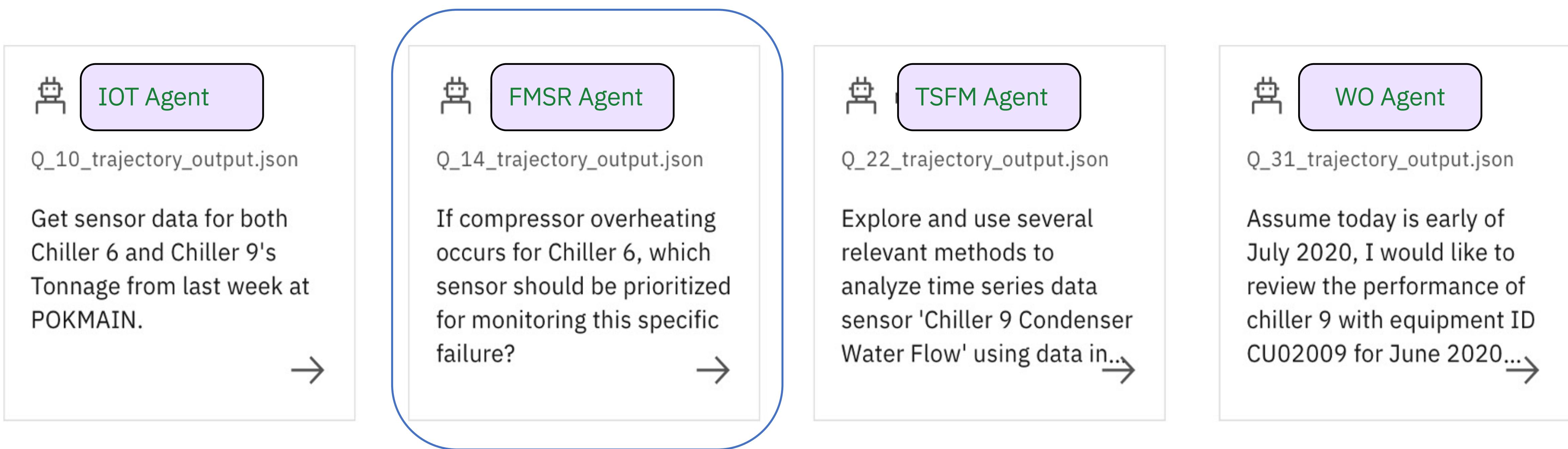


found 27 sensors for asset name B705 AHU 1 and site name POKMAIN. File path contains a JSON array of Sensor data"

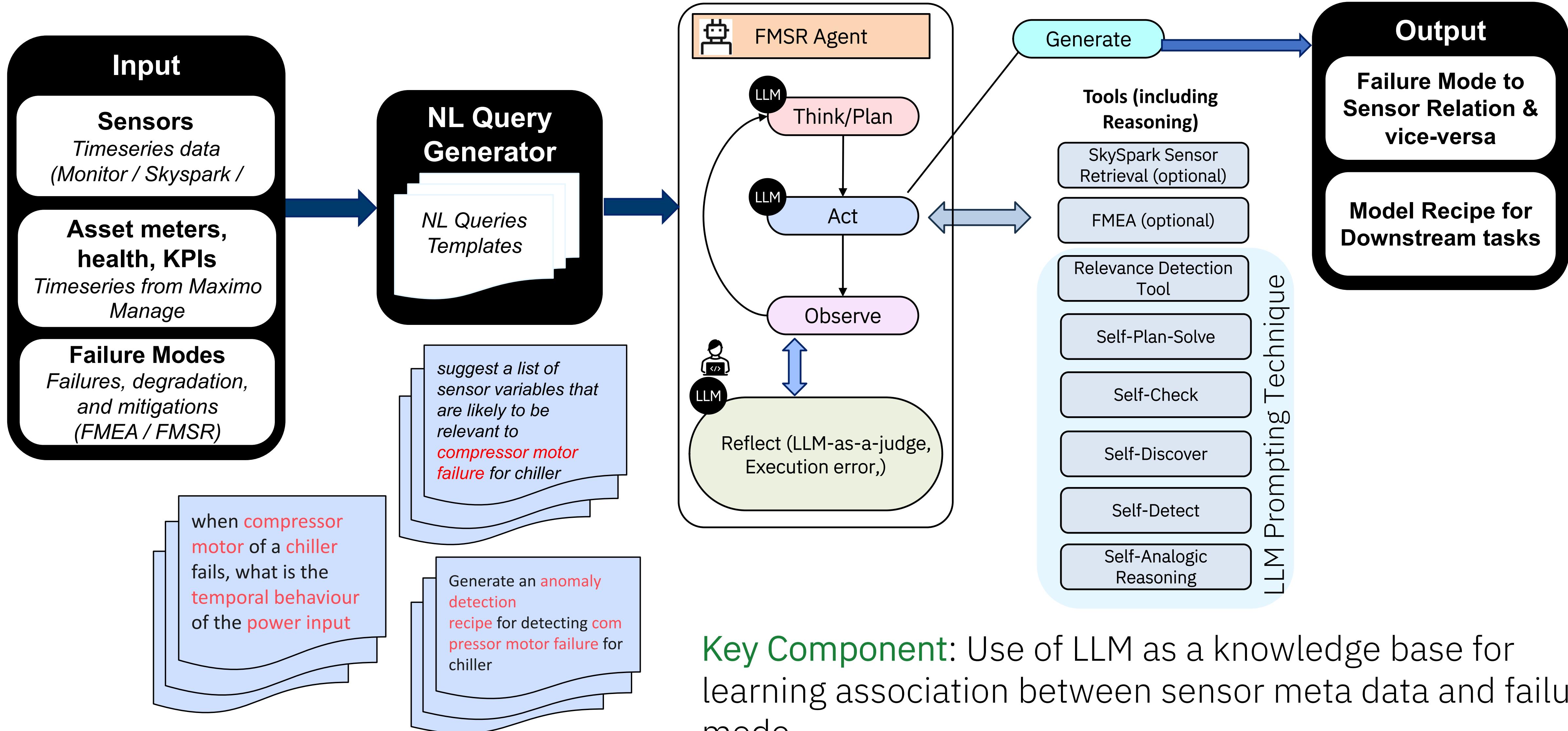


The asset B020 does not exist at site POKMAIN

# Task Specific Specialized Agents

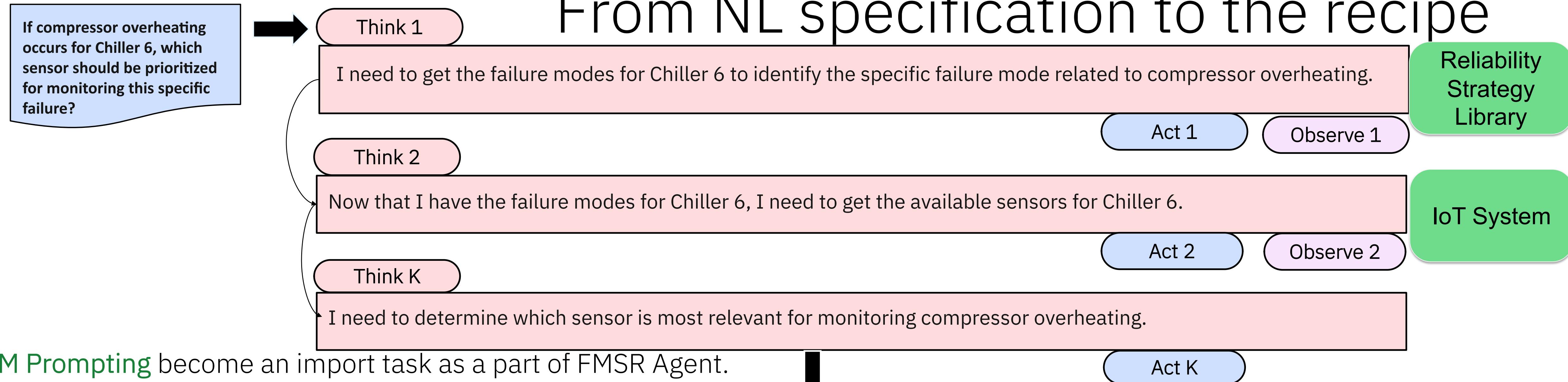


# Architecture: FMSR (Failure Mode 2 Sensor Mapping) Agent



# Illustration of FMSR Agent Working using Example

## From NL specification to the recipe



Prompt 1

Q: For the asset Chiller 6, if the failure Compressor Overheating occurs, can sensor Chiller 6 Chiller Efficiency help monitor or detect the failure for Chiller 6?  
....

LLM

Response 1

A: No, the sensor "Chiller 6 Chiller Efficiency" would not be effective in monitoring or detecting the failure "Compressor Overheating" for Chiller 6.  
...

Prompt 2

Q: For the asset Chiller 6, if the failure Compressor Overheating occurs, can sensor Chiller 6 Return Temperature help monitor or detect the failure for Chiller 6?

LLM

Response 2

A: Yes, the sensor Chiller 6 Return Temperature can help monitor or detect the failure Compressor Overheating for Chiller 6. ....

# Task Specific Specialized Agents



IOT Agent

Q\_10\_trajectory\_output.json

Get sensor data for both  
Chiller 6 and Chiller 9's  
Tonnage from last week at  
POKMAIN.



FMSR Agent

Q\_14\_trajectory\_output.json

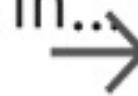
If compressor overheating  
occurs for Chiller 6, which  
sensor should be prioritized  
for monitoring this specific  
failure?



TSFM Agent

Q\_22\_trajectory\_output.json

Explore and use several  
relevant methods to  
analyze time series data  
sensor 'Chiller 9 Condenser  
Water Flow' using data in...

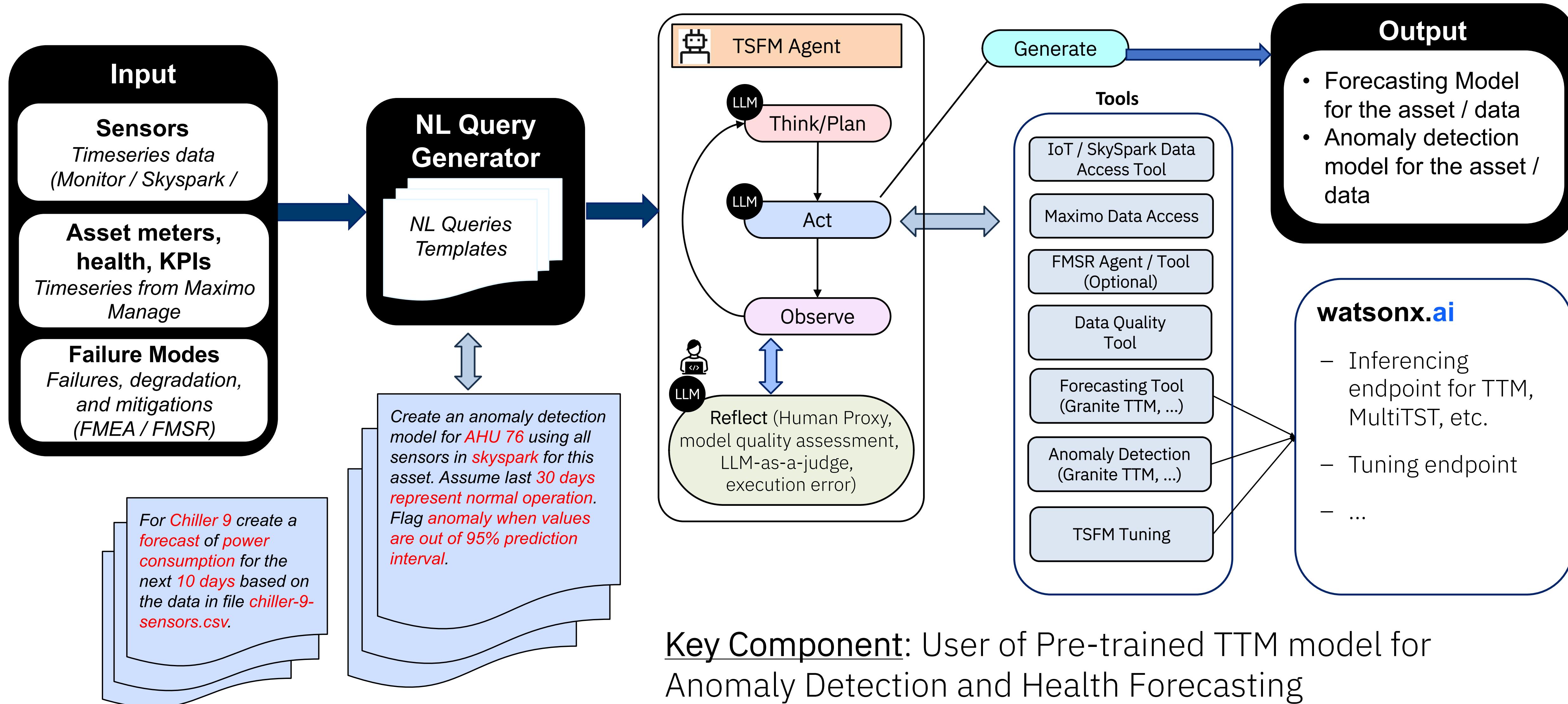


WO Agent

Q\_31\_trajectory\_output.json

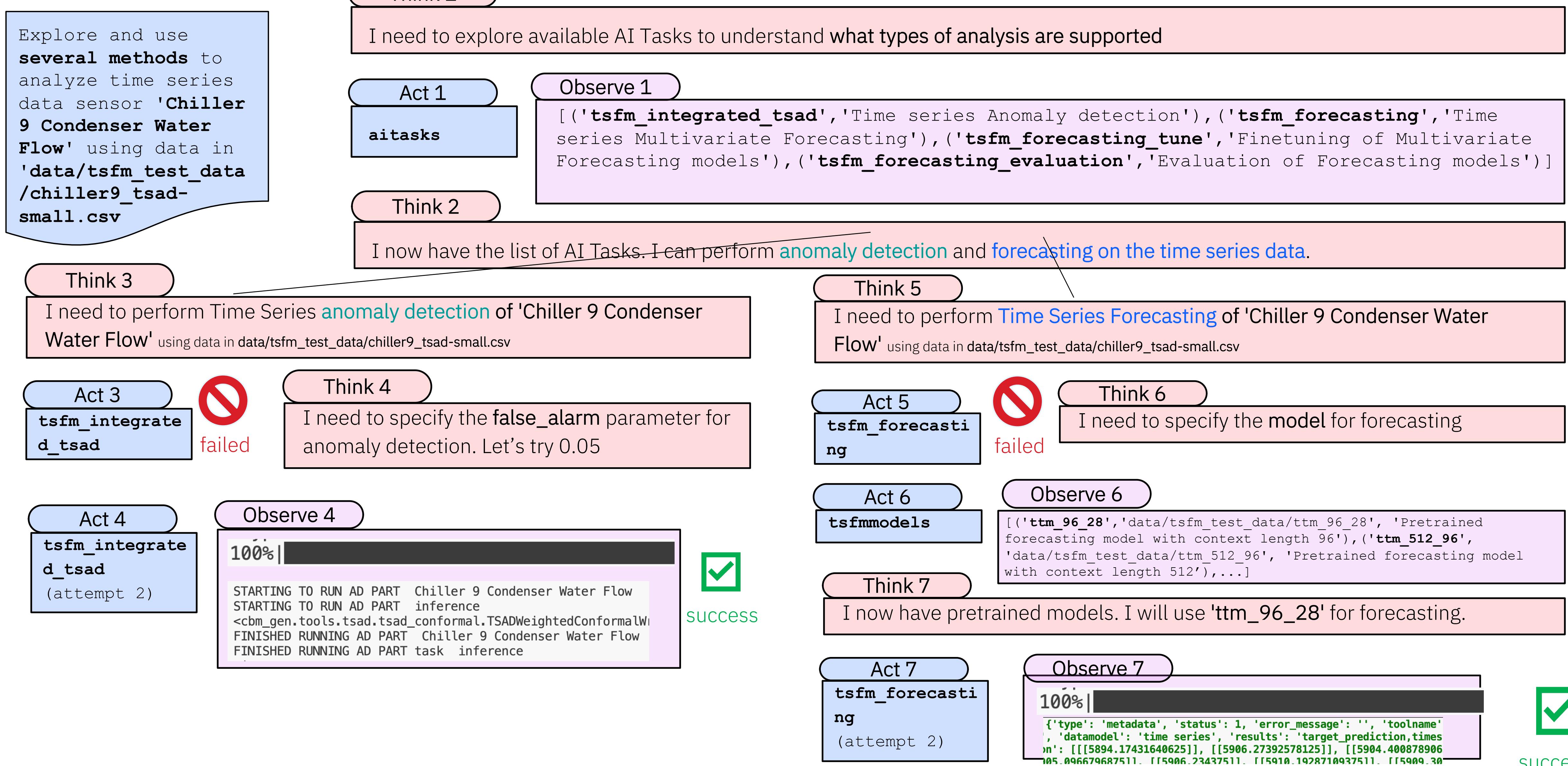
Assume today is early of  
July 2020, I would like to  
review the performance of  
chiller 9 with equipment ID  
CU02009 for June 2020...  
→

# Architecture: TSFM (Time Series Foundation Model) Agent

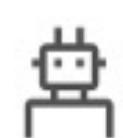


# Architecture: TSFM (Time Series Foundation Model) Agent

## Interesting Trajectory of TSFM Agent



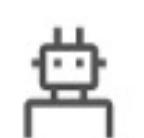
# Task Specific Specialized Agents



IOT Agent

Q\_10\_trajectory\_output.json

Get sensor data for both  
Chiller 6 and Chiller 9's  
Tonnage from last week at  
POKMAIN.



FMSR Agent

Q\_14\_trajectory\_output.json

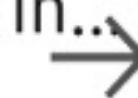
If compressor overheating  
occurs for Chiller 6, which  
sensor should be prioritized  
for monitoring this specific  
failure?



TSFM Agent

Q\_22\_trajectory\_output.json

Explore and use several  
relevant methods to  
analyze time series data  
sensor 'Chiller 9 Condenser  
Water Flow' using data in...

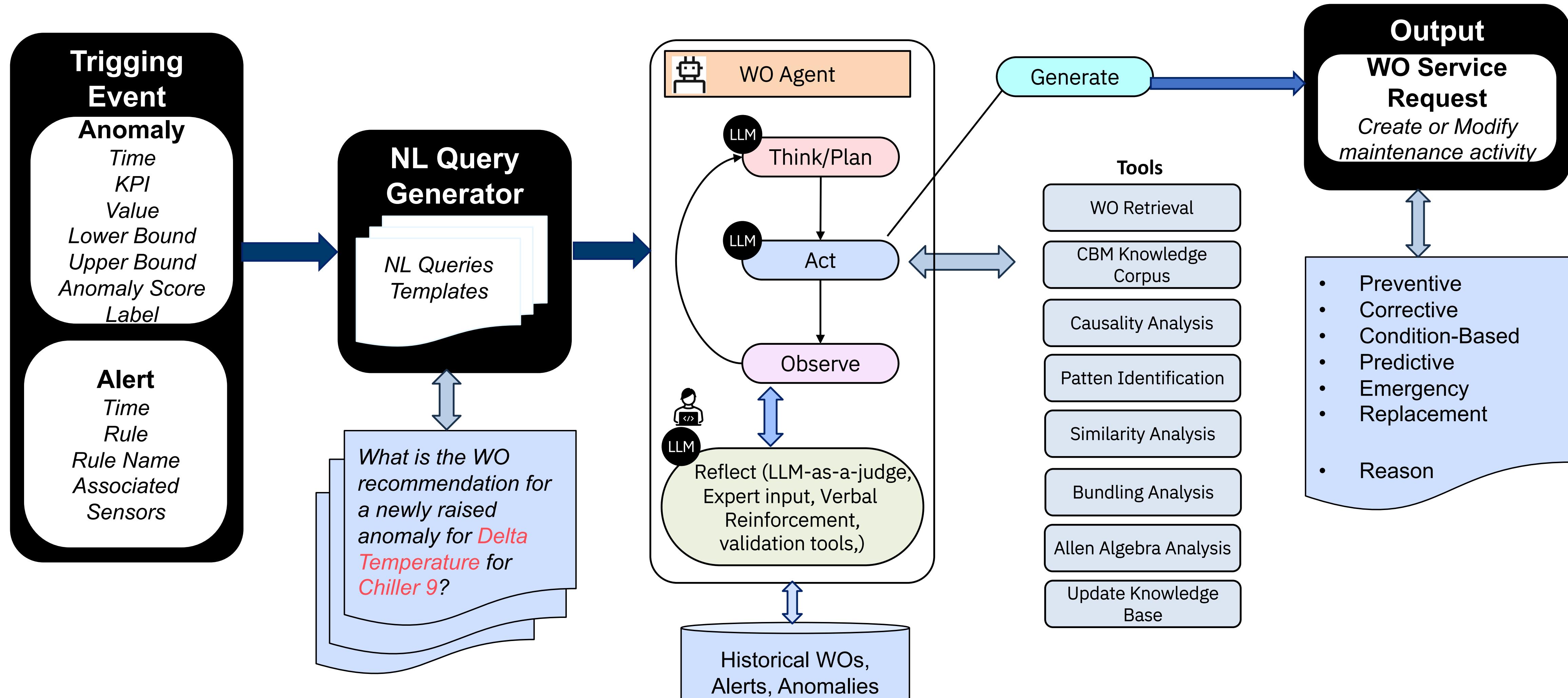


WO Agent

Q\_31\_trajectory\_output.json

Assume today is early of  
July 2020, I would like to  
review the performance of  
chiller 9 with equipment ID  
CU02009 for June 2020...  
→

# Architecture: WO (Work Order) Agent



Key Component: LLM's ability to inspect structured output and code for doing temporal event analysis on Alert, Anomaly and Work-Orders

# Interesting Trajectory of WO Agent

Monitor Chiller 9 and identify anomalous behaviors and provide a work order recommendation if any.

## Think 1

To recommend a work order for Chiller 9 at the POKMAIN site, I need to first identify any anomalous behaviors within the specified date range. I will use the `get\_events` API to retrieve all events for Chiller 9 during this period and then filter for anomalies .....

## Act 1

```
```python
# Define the equipment and date range
equipment = Equipment(equipment_id="CU02009", site="POKMAIN")
date_range = DateRange(start_date=Date("2020-06-01"), end_date=Date("2020-06-01"))

# Retrieve the events
events_pickle_path = get_events(equipment, date_range)

# Load the work orders data from the pickle file
with open(work_orders_pickle_path, "rb") as file:
    work_orders = pickle.load(file)

# Filter corrective work orders
corrective_work_orders = filter_corrective_work_orders(work_orders)

# Display corrective work orders
for cwo in corrective_work_orders:
    print(cwo)```
```

## Observe 1

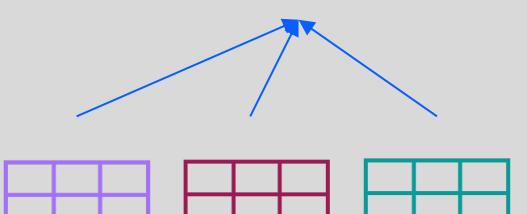
```
Event(event_id=al_d460a5db-bfde-420c-9309-0a6d8bffb1b4, event_group=ALERT, event_category=ALERT, event_type=CR00008,
Event(event_id=al_6c7a9e83-9434-4f2d-890f-87c056f0a2e8, event_group=ALERT, event_category=ALERT, event_type=CR00008,
Event(event_id=al_d01ecec9-b076-4f36-85f6-77e7e5220e3f, event_group=ALERT, event_category=ALERT, event_type=CR00008,
Event(event_id=al_c6deb4a0-b1ff-4a0d-acb6-12a19a5cc404, event_group=ALERT, event_category=ALERT, event_type=CR00008,
Event(event_id=an_c8c1176e-4464-4c85-bda9-5dd891494222, event_group=ANOMALY, event_category=ANOMALY, event_type=KPI
Event(event_id=an_e83c9bc0-2046-4a0c-a3f8-bc1c72f0d19b, event_group=ANOMALY, event_category=ANOMALY, event_type=KPI
Event(event_id=al_6c86d0aa-4ee9-4c7d-8fbf-67063276fb9d, event_group=ALERT, event_category=ALERT, event_type=CR00008,
Event(event_id=al_6b05fb2d-df7c-45f3-8a79-624a8e5ec1c2, event_group=ALERT, event_category=ALERT, event_type=CR00008,
```

# Interesting Trajectory of WO Agent

## From NL specification to model

Monitor Chiller 9 and identify anomalous behaviors and provide a work order recommendation if any.

Thanks to Data Classes and Structured input to LLM; thinking process improved



Data Classes

### Think 1

To recommend a work order for Chiller 9 at the POKMAIN site, I need to first identify any anomalous behaviors within the specified date range. I will use the `get\_events` API to retrieve all events for Chiller 9 during this period and then filter for anomalies .....

### Think 2

The data shows multiple anomalies and alerts related to the chiller's evaporator approach being high and various KPI anomalies. There is one corrective work order scheduled for a vibration analysis on June 16, 2020. Given the frequency and nature of the alerts and anomalies, it is likely that additional corrective work orders may be needed. To determine if new work orders should be generated, I will analyze the trend of these anomalies and alerts. Specifically, I will check if the frequency of these events is increasing over time.

### Act 2

```
# Count the number of anomalies and alerts per day
from collections import defaultdict

daily_event_count = defaultdict(int)
for event in anomaly_alert_events:
    event_date = event.event_time.strftime("%Y-%m-%d")
    daily_event_count[event_date] += 1

# Display daily event counts
for date, count in sorted(daily_event_count.items()):
    print(f"{date}: {count} events")
```

### Observe 2

2020-06-01: 1 events	2020-06-16: 9 events
2020-06-02: 1 events	2020-06-17: 1 events
2020-06-03: 1 events	2020-06-18: 3 events
2020-06-04: 1 events	2020-06-19: 2 events
2020-06-05: 1 events	2020-06-21: 1 events
2020-06-06: 1 events	2020-06-22: 2 events
2020-06-08: 15 events	2020-06-23: 1 events
2020-06-09: 5 events	2020-06-24: 1 events
2020-06-10: 3 events	

# Interesting Trajectory of WO Agent

## From NL specification to model

Monitor Chiller 9 and identify anomalous behaviors and provide a work order recommendation if any.

### Think 1

To recommend a work order for Chiller 9 at the POKMAIN site, I need to first identify any anomalous behaviors within the specified date range. I will use the `get\_events` API to retrieve all events for Chiller 9 during this period and then filter for anomalies .....

### Think 2

The data shows multiple anomalies and alerts related to the chiller's evaporator approach being high and various KPI anomalies. There is one corrective work order scheduled for a vibration analysis on June 16, 2020. Given the frequency and nature of the alerts and anomalies, it is likely that additional corrective work orders may be needed. To determine if new work orders should be generated, I will analyze the trend of these anomalies and alerts. Specifically, I will check if the frequency of these events is increasing over time.

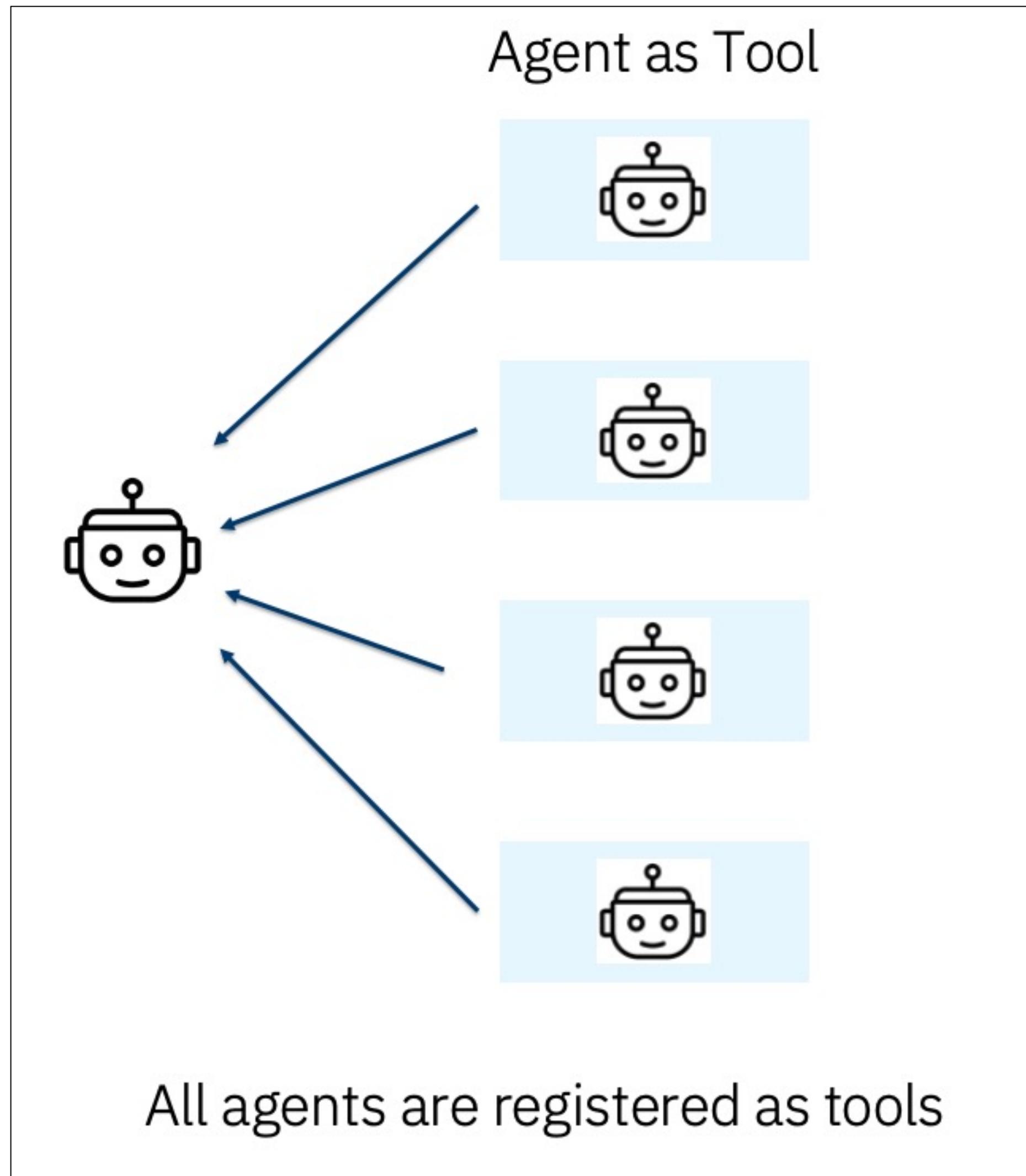
### Think 3

The daily event counts show a significant spike in anomalies and alerts on June 8 and June 16, with 15 and 9 events respectively. This indicates a potential issue that may require additional corrective work orders. To further analyze the situation, I will check if there are any pending work orders related to these issues. If not, I will recommend generating new work orders based on the nature of the anomalies and alerts.

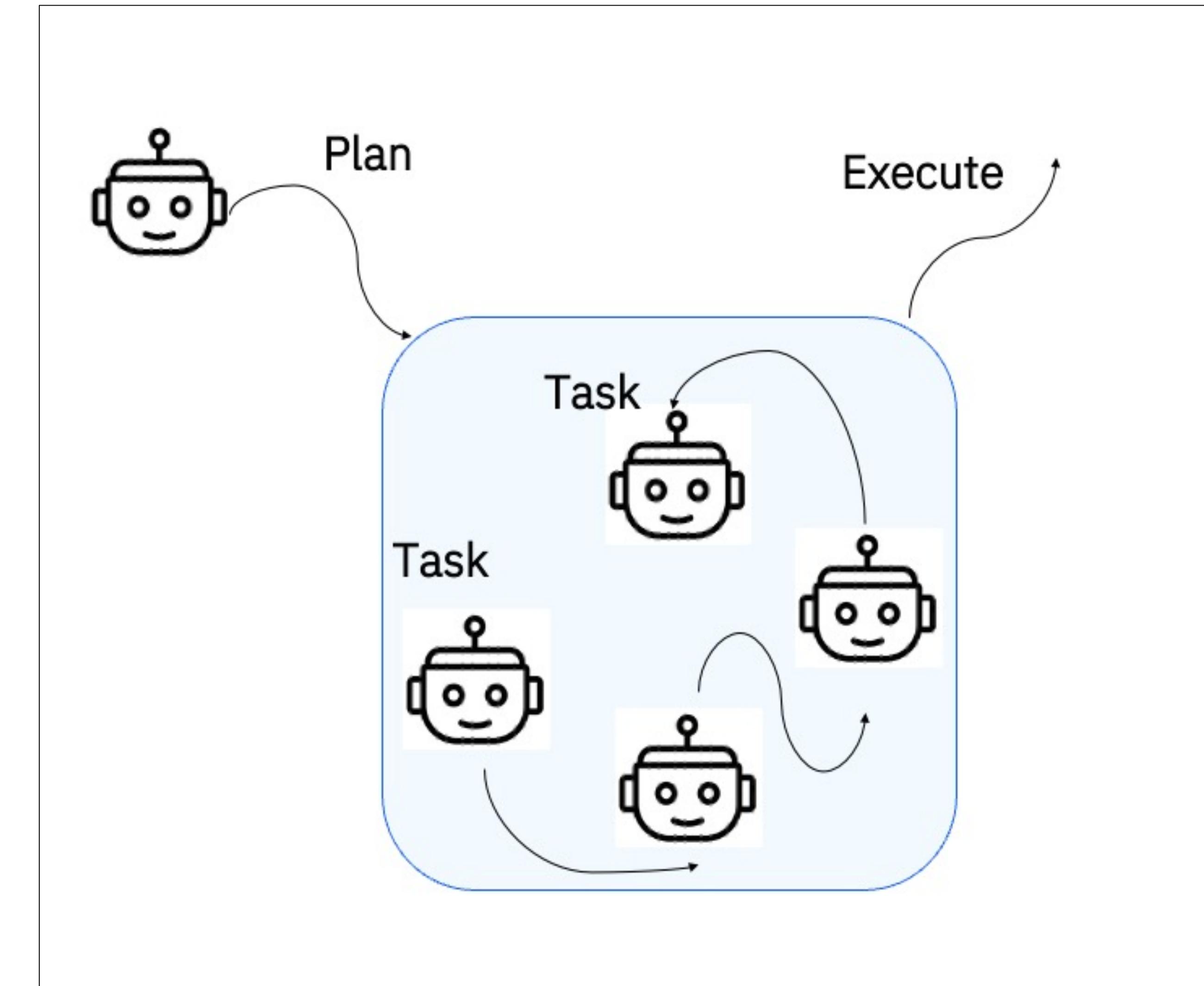
```
Final Answer:  
{  
    "corrective_work_orders": [  
        {  
            "wo_id": "L355477",  
            "wo_description": "RESCAN CHILLER # 9",  
            "collection": "entire chiller system",  
            "primary_code": "MT013",  
            "primary_code_description": "Vibration Analysis",  
            "secondary_code": "MT013c",  
            "secondary_code_description": "Post-Maintenance Vibration Analysis",  
            "equipment_id": "CU02009",  
            "equipment_name": "Chiller 9",  
            "preventive": false,  
            "work_priority": 1,  
            "actual_finish": "2020-06-16 19:30:00",  
            "duration": "0:00",  
            "actual_work_hour": "0:00"  
        }  
    ],  
    "recommendations": [  
        {  
            "recommendation": "Generate new work orders to address the chiller's  
            issues."  
        }  
    ]  
}
```

- Beat 1 – World & pain
- Beat 2 – Vision: agentic AI for Industry 4.0
- Beat 3 – AssetOpsBench
- Beat 4 – Agent architectures, orchestration, and failure modes
- Beat 5 – Ecosystem + hands-on

# AssetOpsBench : Multi-Agent Implementation Strategy

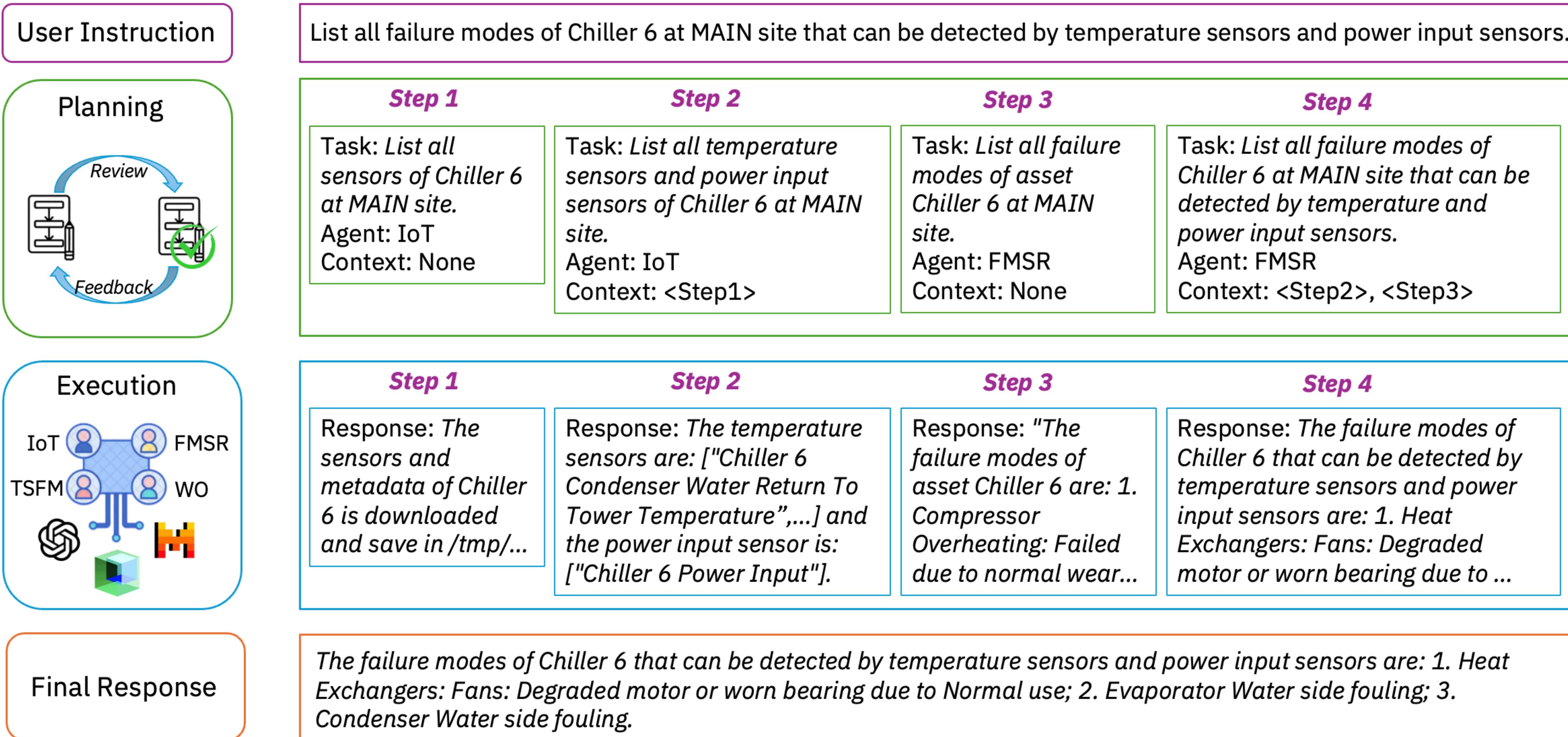


Agent-As-Tool Approach



Plan-Execute Approach

# AssetOpsBench : Illustrative Workflow



# AssetOpsBench : Evaluation



# Automatic Evaluation of Agentic Workflow

## Ground Truth Preparation

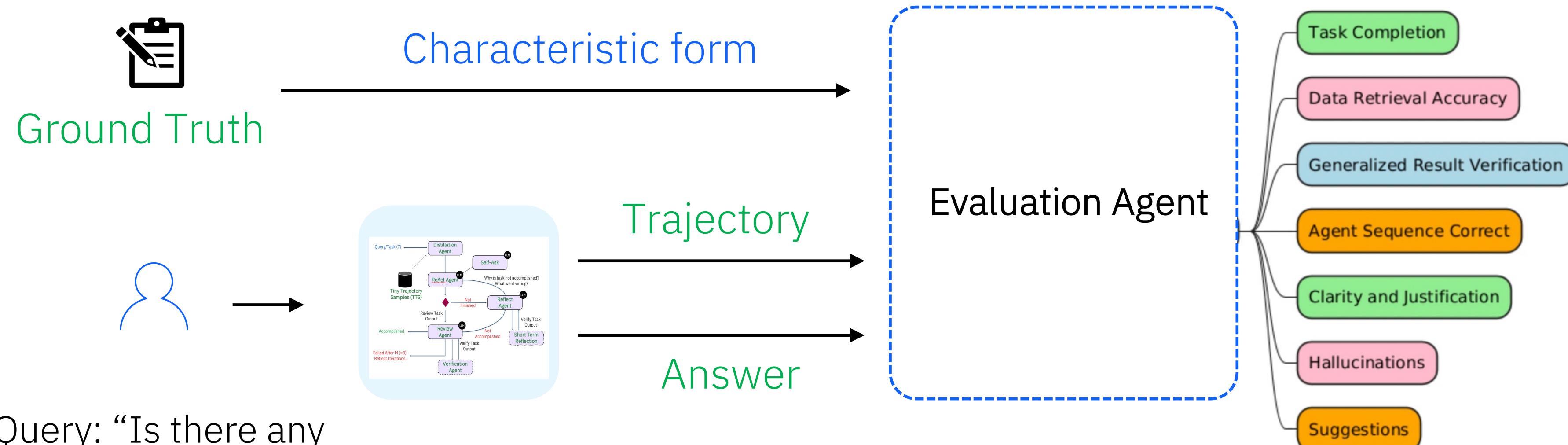
- **Characteristic form** describe the final output along with the process/step to be conducted. As an example:

“The expected response should confirm the successful execution of all required actions while ensuring the correct variables, including the asset (Chiller 6), location (POKMAIN), and time range (week of 2020-04-27), were used for data retrieval and analysis. It should specify that **IoT Agent** was called to request and download the data, and **TSFM Agent** was properly utilized to perform Time Series anomaly detection on the Tonnage parameter. The response must also verify that the data was accurately stored in the designated file location, and that the analysis results were saved to a new file. Additionally, the response should explicitly confirm the detection of anomalies in Chiller 6's Tonnage during the specified timeframe at the POKMAIN site, as these anomalies were anticipated.”

This characteristic form serves as the ground truth for evaluating responses using a **rubric-based performance assessment**.

# Automatic Evaluation of Agentic Workflow

## Evaluation Agent: Rubric-based Performance Assessment

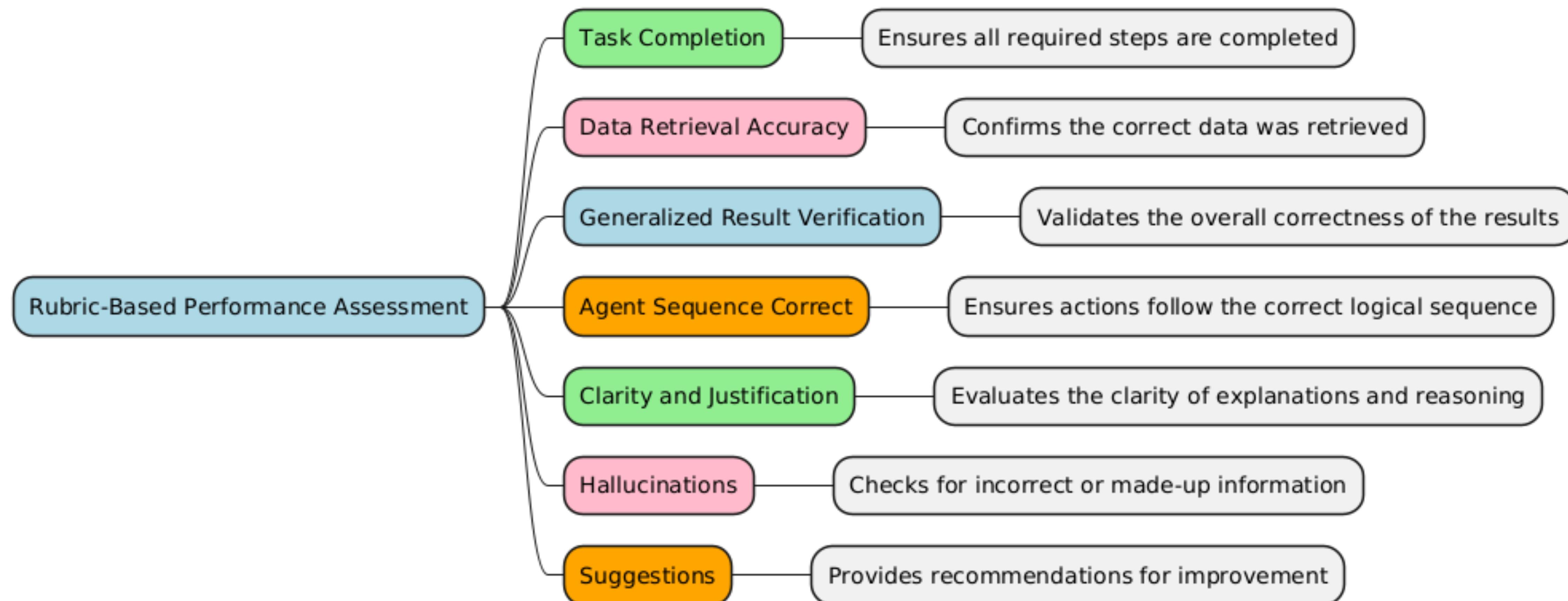


Query: “Is there any anomaly detected in Chiller 6's Tonnage in the week of 2020-04-27 at the POKMAIN site?”

# Automatic Evaluation of Agentic Workflow

## Evaluation Agent: Rubric-based Performance Assessment

We captured performance across 6 metrics to better understand the failure mode of agents.



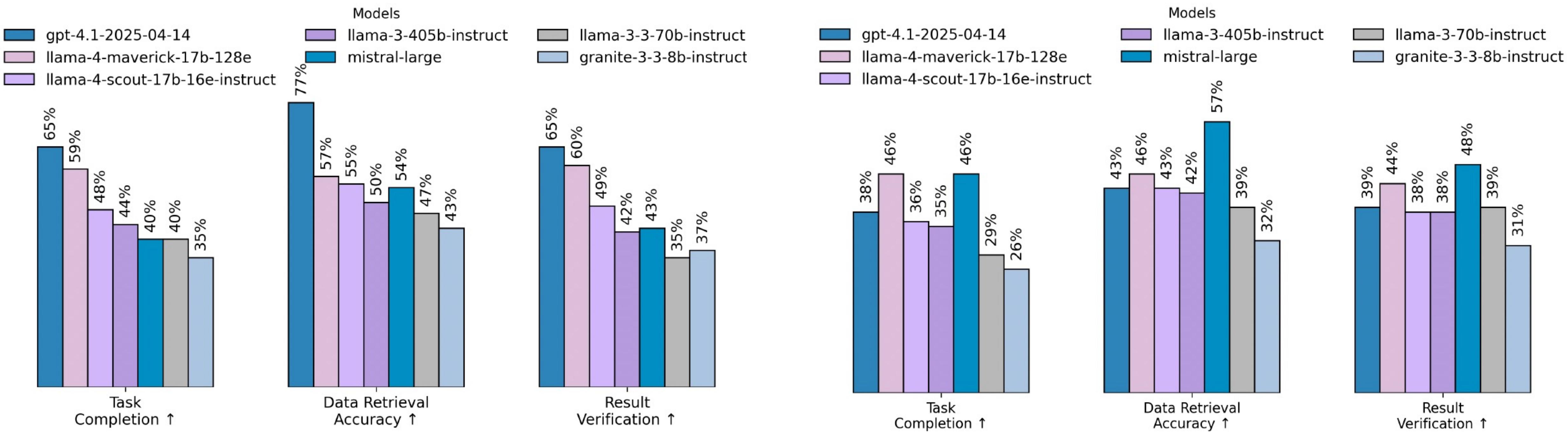
# AssetOpsBench : Summary of Result



# Leaderboard

We conducted an extensive benchmark to compare the two orchestration recipes:

- Agent-As-Tool consistently outperforms Plan-Execute



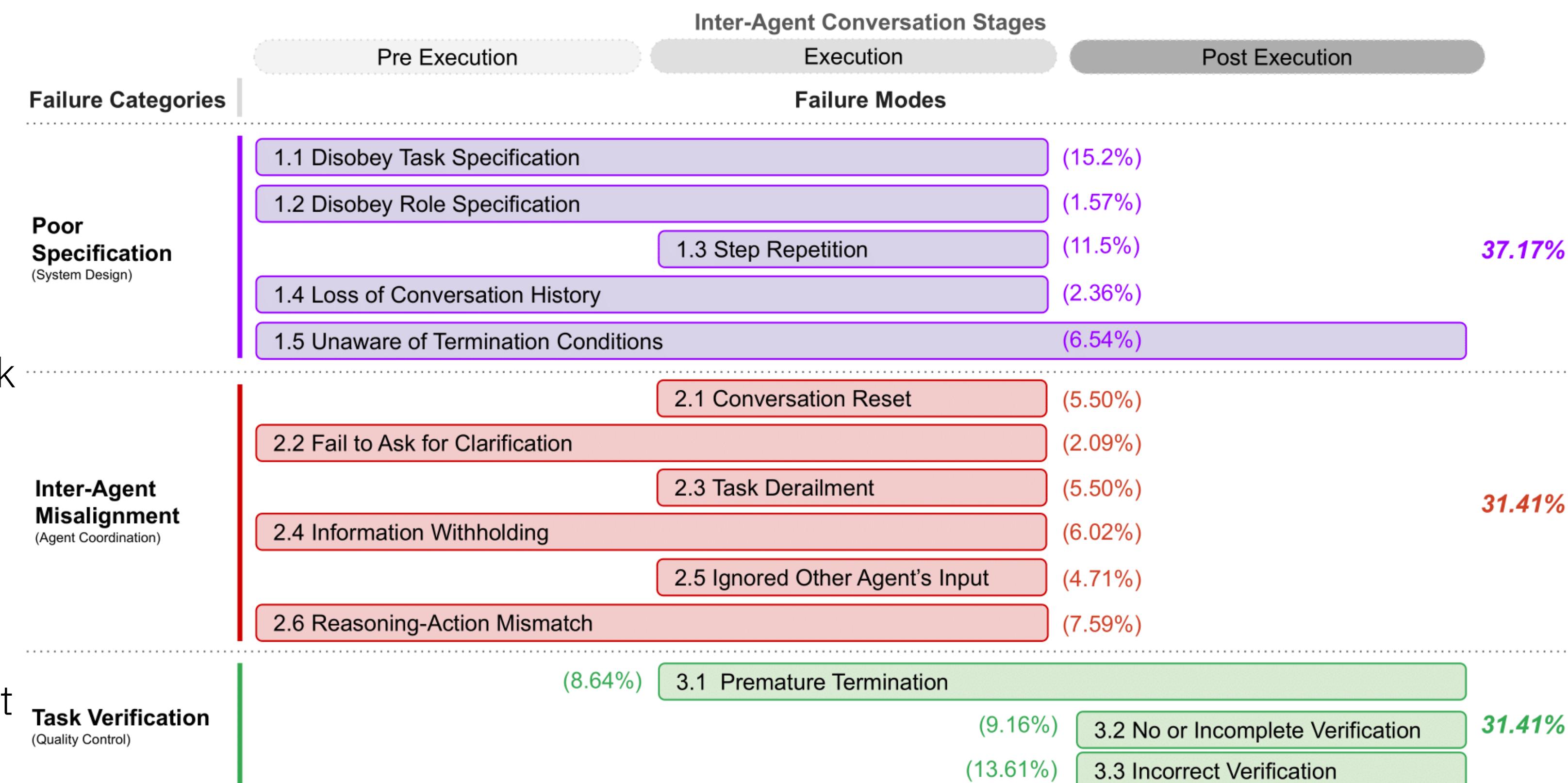
((a)) **Agent-As-Tool** Approach

((b)) **Plan-Execute** Approach

# Multi-Agent System Taxonomy's Core Failure Categories

Three Main Failure Categories:

- **Specification Failures**: Misunderstood tasks, ambiguous goals, or unclear success metrics.
  - Example: Agent reports success, but output doesn't match user intent.
- **Inter-Agent Failures**: Poor coordination or lack of shared context.
  - Example: One agent updates a plan, but others act on outdated data.
- **Task Verification Failures**: Missing validation, poor execution tracking.
  - Example: Action taken, but no check if effect occurred.



# Emergent Failure Mode Discovery

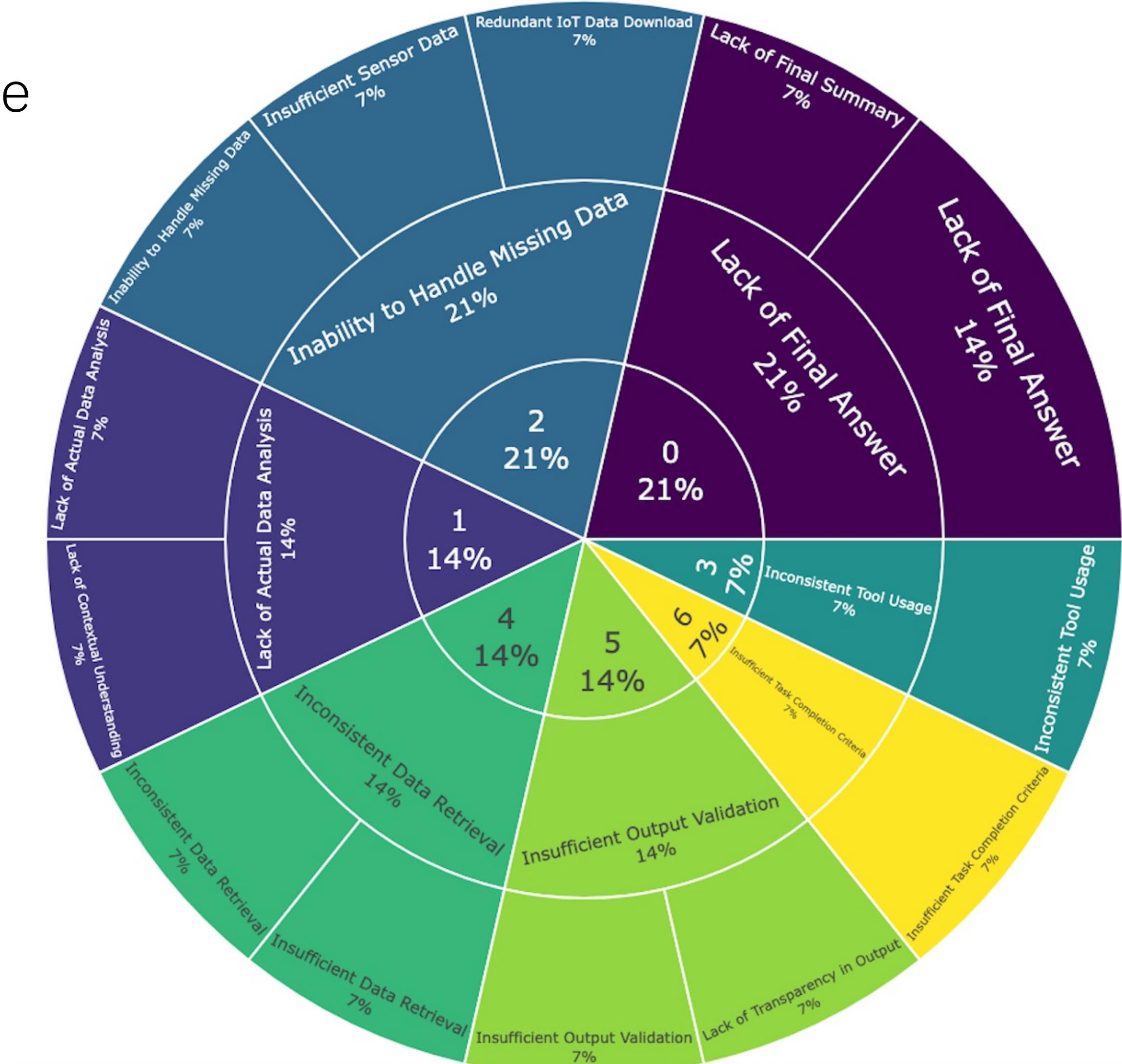
We have an automated way of trajectory introspection to discover novel, emerging failure modes automatically

## Agent Coordination (Total 27.52%)

Conversation Reset	Execution: 0.00%
Fail to Ask for Clarification	Execution: 10.22%
Task Derailment	Execution: 4.34%
Information Withholding	Execution: 2.22%
Ignored Agent's Input	Execution: 2.06%
Action Mismatch	Execution: 8.68%

We added a new method called “Self-Ask” to demonstrate the performance improvement

Model	enable_agent_ask=True	enable_agent_ask=False
gpt-4.1-2025-04-14	63%	65%
lama-4-maverick	66%	59%
llama-3-405b-instruct	61%	44%
mistral-large	58%	40%
llama-3-3-70b-instruct	35%	40%
granite-3-3-8b-instruct	32%	35%



# AssetOpsBench Extensive Research: FailureSensorIQ

- FailureSensorIQ introduces a **dataset** and **benchmark** that tests whether LLMs can reason about sensors, assets, and failure modes beyond data-driven correlations. It benchmarks *sensor-failure relationships*, which is the primary capability targeted by the Failure Mode Sensor Relation (FMSR) agent in AssetOpBench.
- FailureSensorIQ is accepted in NeurIPS 2025.

The screenshot shows the HuggingFace Dataset page for FailureSensorIQ. At the top, it displays basic dataset statistics: Tasks (Question Answering), Modalities (Text), Formats (json), Languages (English), Size (1K - 10K), and ArXiv (None). Below this, there are sections for Libraries (Datasets, pandas, Croissant), License (apache-2.0), and navigation links (Dataset card, Data Studio, Files and versions, xet, Community).

The screenshot shows the Dataset Viewer for the FailureSensorIQ dataset. It displays a table with columns: subject, id, question, options, and option\_ids. The table has two rows under the subset "multi\_true\_multi\_choice\_qa". The first row corresponds to the "failure\_mode\_sensor\_analysis" entry in the main table below. The second row is a header for the split "train". The table includes a search bar and a "Search this dataset" button.

subject	id	question	options	option_ids
failure_mode_sensor_analysis	1	For electric motor, if a...	[ "oil debris", ... ]	[ "A", "B", "C", "D", "E" ... ]
failure_mode_sensor_analysis	2	For electric motor, if a...	[ "resistance", ... ]	[ "A", "B", "C", "D", "E" ... ]
failure_mode_sensor_analysis	3	For electric motor, if a...	[ "coast down time", ... ]	[ "A", "B", "C", "D", "E" ... ]
failure_mode_sensor_analysis	4	For electric motor, if a...	[ "partial discharge", ... ]	[ "A", "B", "C", "D", "E" ... ]
failure_mode_sensor_analysis	5	For electric motor, if a...	[ "temperature", ... ]	[ "A", "B", "C", "D", "E" ... ]

HuggingFace Dataset

The screenshot shows the FailureSensorIQ Kaggle Benchmark page. At the top, it displays basic benchmark statistics: Tasks (Question Answering), Modalities (Text), Formats (json), Languages (English), Size (1K - 10K), and ArXiv (None). Below this, there are sections for Libraries (Datasets, pandas, Croissant), License (apache-2.0), and navigation links (Leaderboard, Discussion (0)).

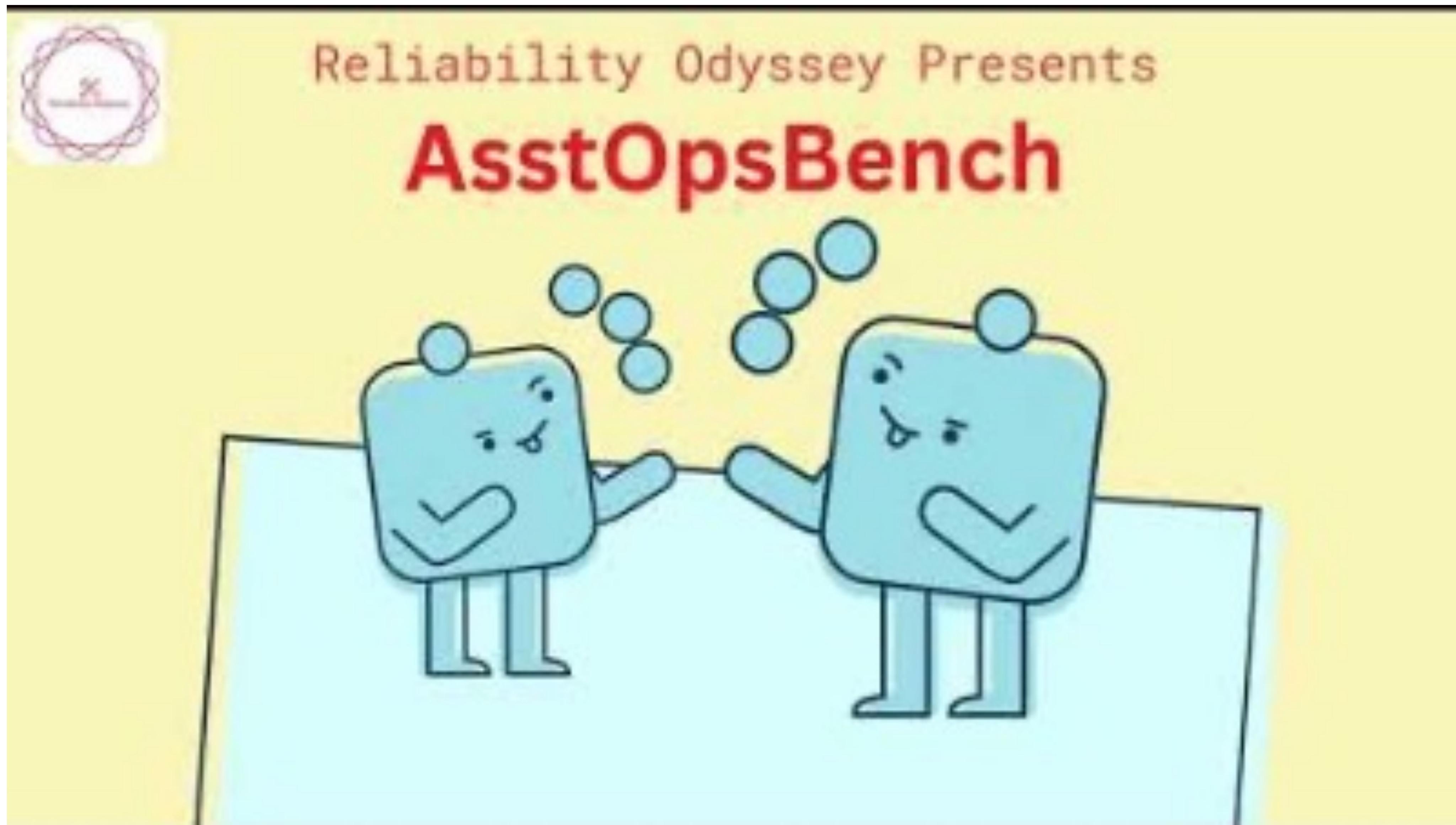
The screenshot shows the Leaderboard for the FailureSensorIQ Kaggle Benchmark. It lists eight models with their scores and other metrics. The table includes columns: #, Model, Score, Consistency, F-Score, and Elimination Accuracy.

#	Model	Score	Consistency	F-Score	Elimination Accuracy
1	Gemini-3-Pro-Preview	69.1%	63.0%	69.1%	70.8%
2	O3-2025-04-16	67.6%	59.1%	67.4%	69.4%
3	Gpt-5-2025-08-07	67.2%	59.2%	67.3%	69.4%
4	Gemini-2.5-Pro	67.0%	57.5%	67.0%	68.8%
5	Gemini-2.5-Flash	65.5%	56.1%	65.8%	68.3%
6	Gpt-5-Mini-2025-08-07	65.3%	56.8%	65.5%	68.2%
7	O4-Mini-2025-04-16	64.8%	56.7%	65.0%	67.4%
8	Grok-4.1-Fast-Reasoning	64.6%	57.9%	64.9%	66.4%

Kaggle Benchmark



# AssetOpsBench : User Guide



# Power Forecast Assistant with Agent

**XX Power** is a corporation responsible for building, maintaining and operating an electricity network connecting customers to traditional and renewable energy sources.

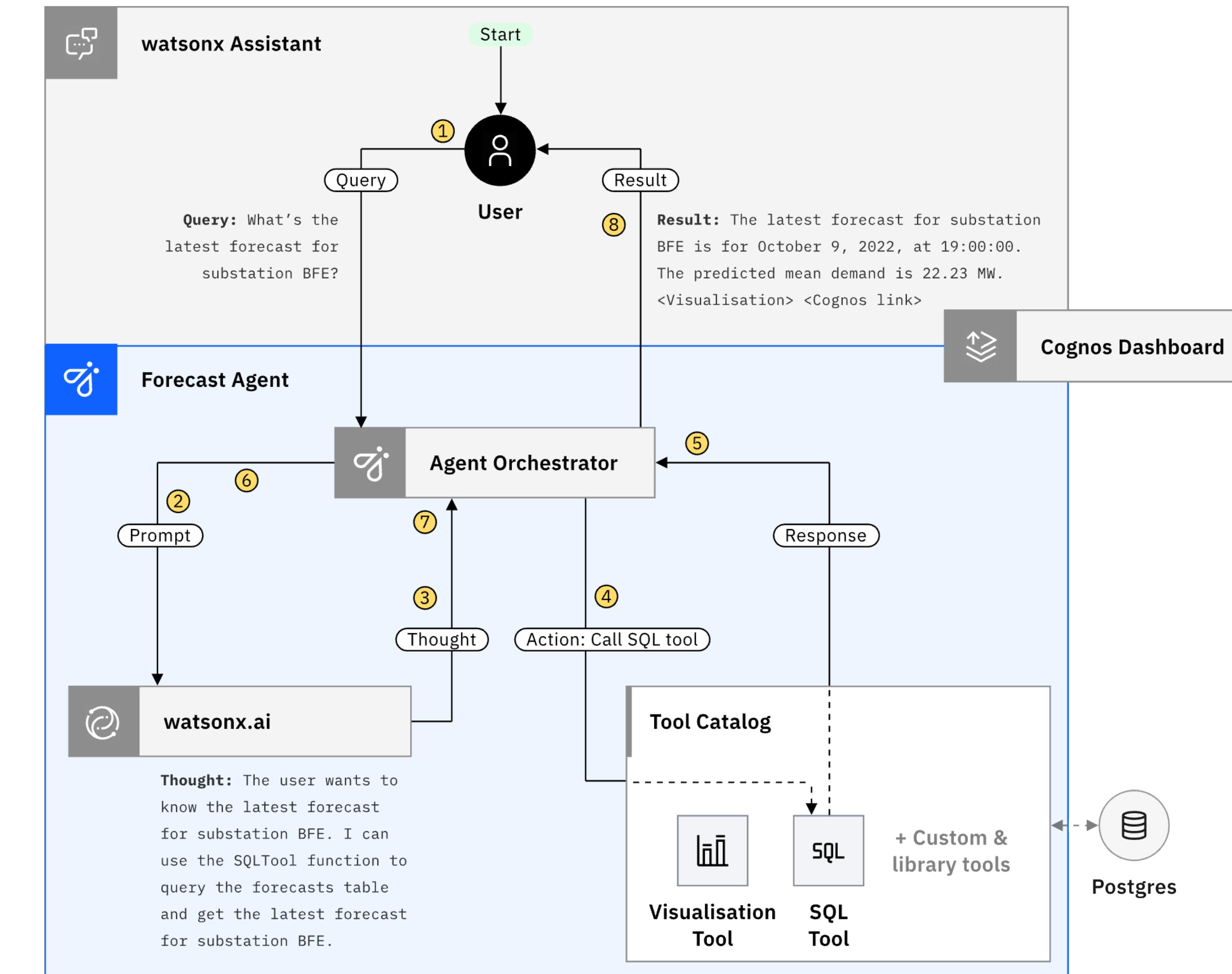
**Use Case:** Use of virtual assistant as a new way to interact with model forecasts of power stations to provide AI transparency and end-to-end lifecycle management.

## Focus:

The focus for this use case is on Forecast Consumption and governance of the forecast model and assistant.

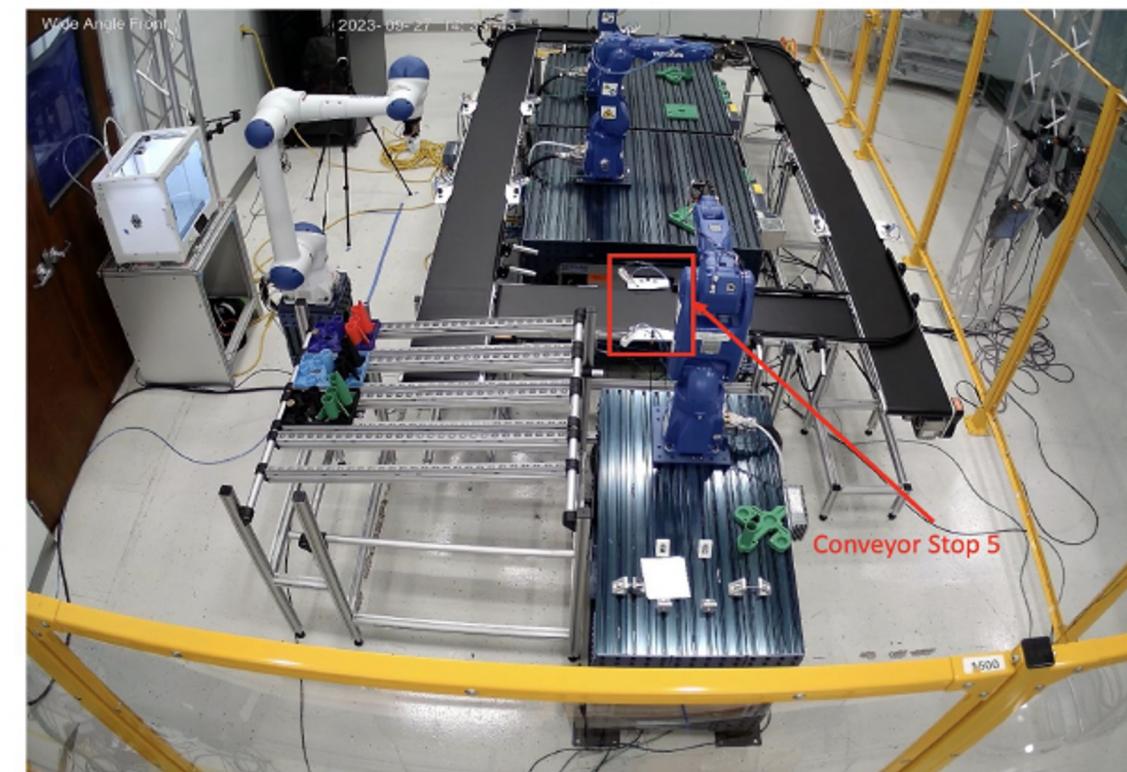
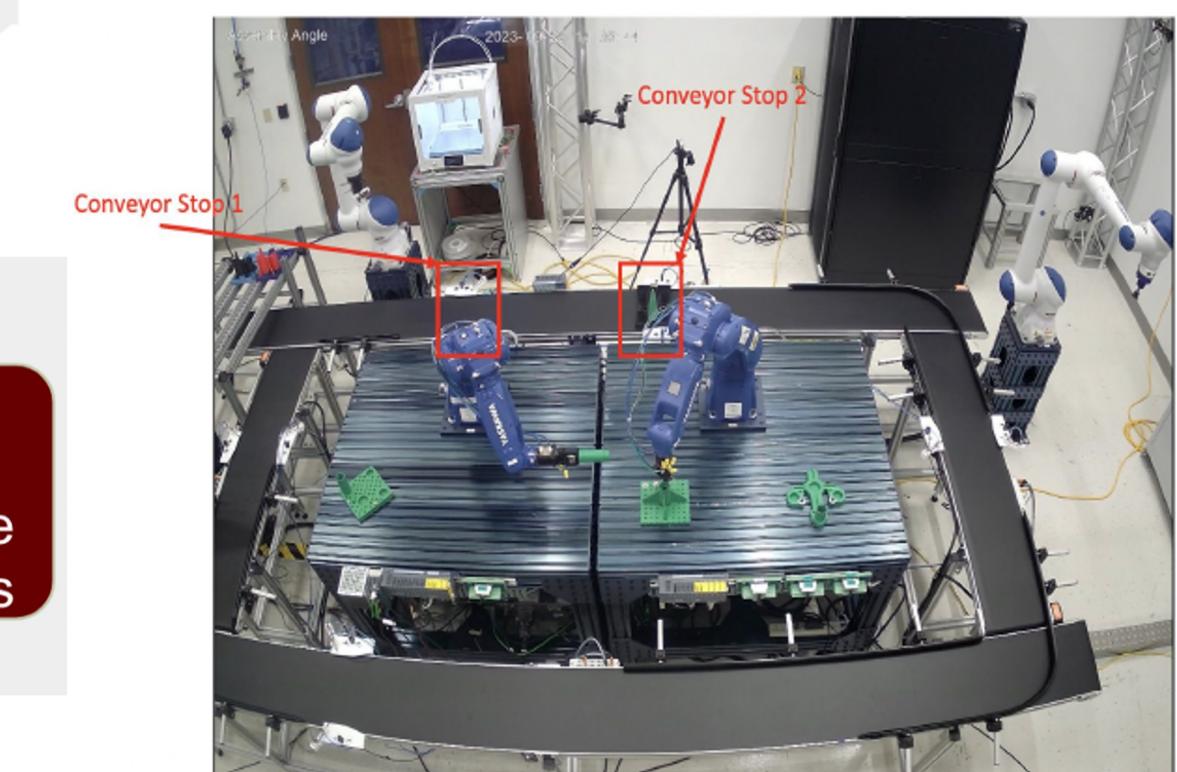
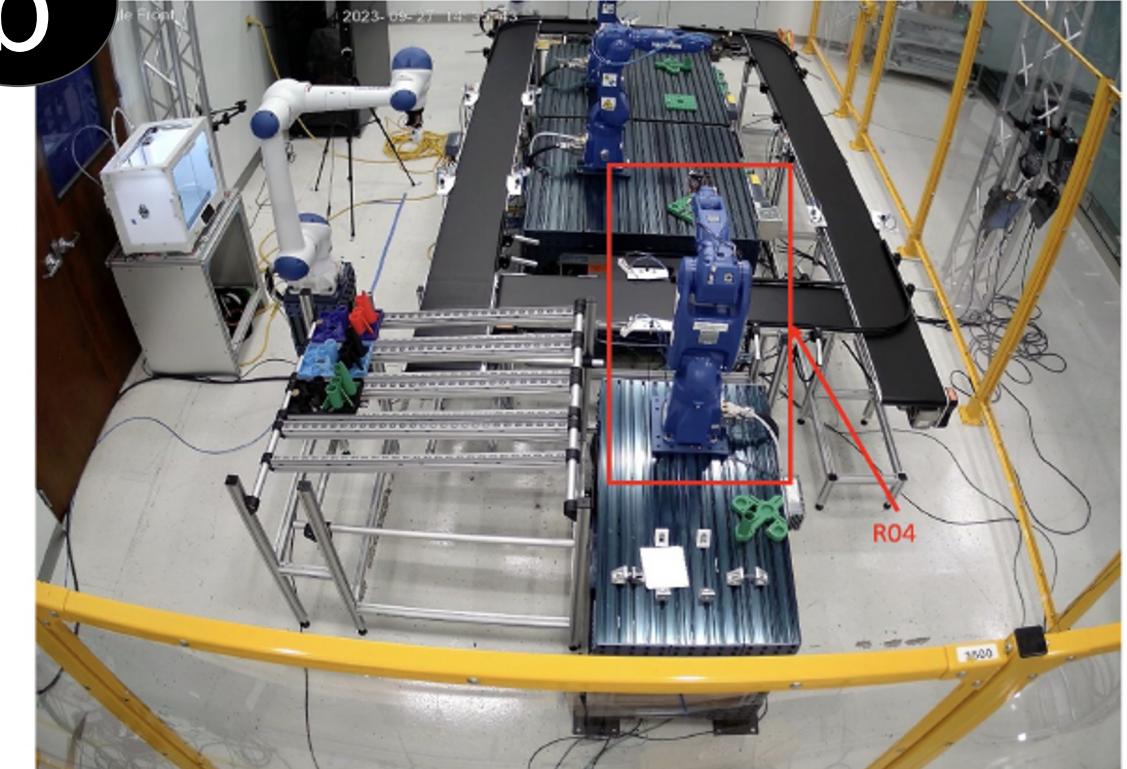
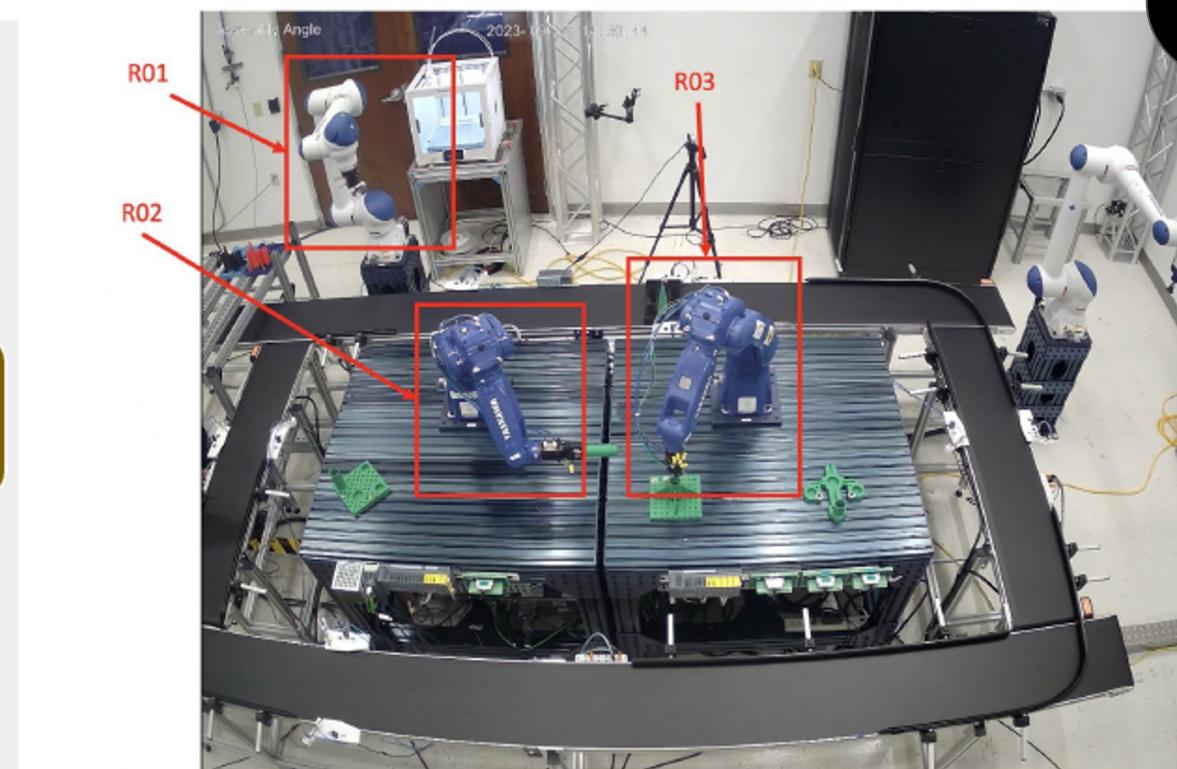
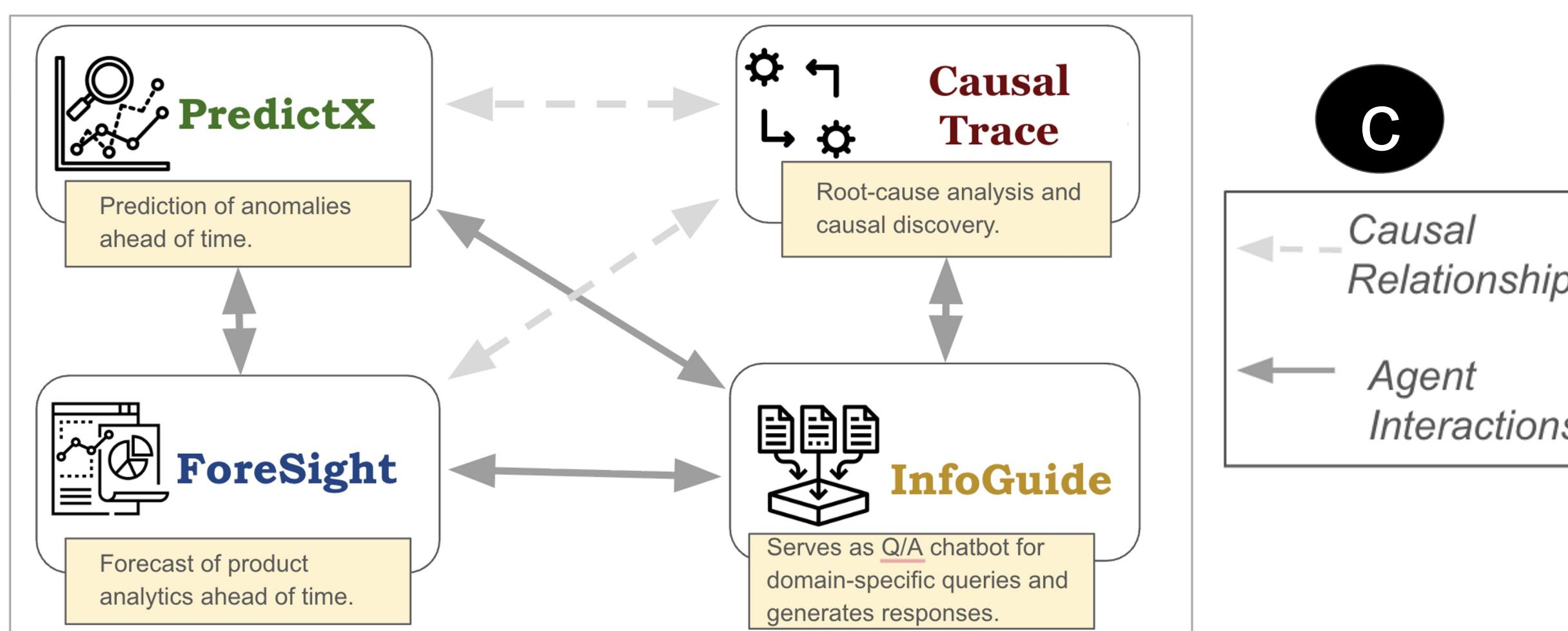
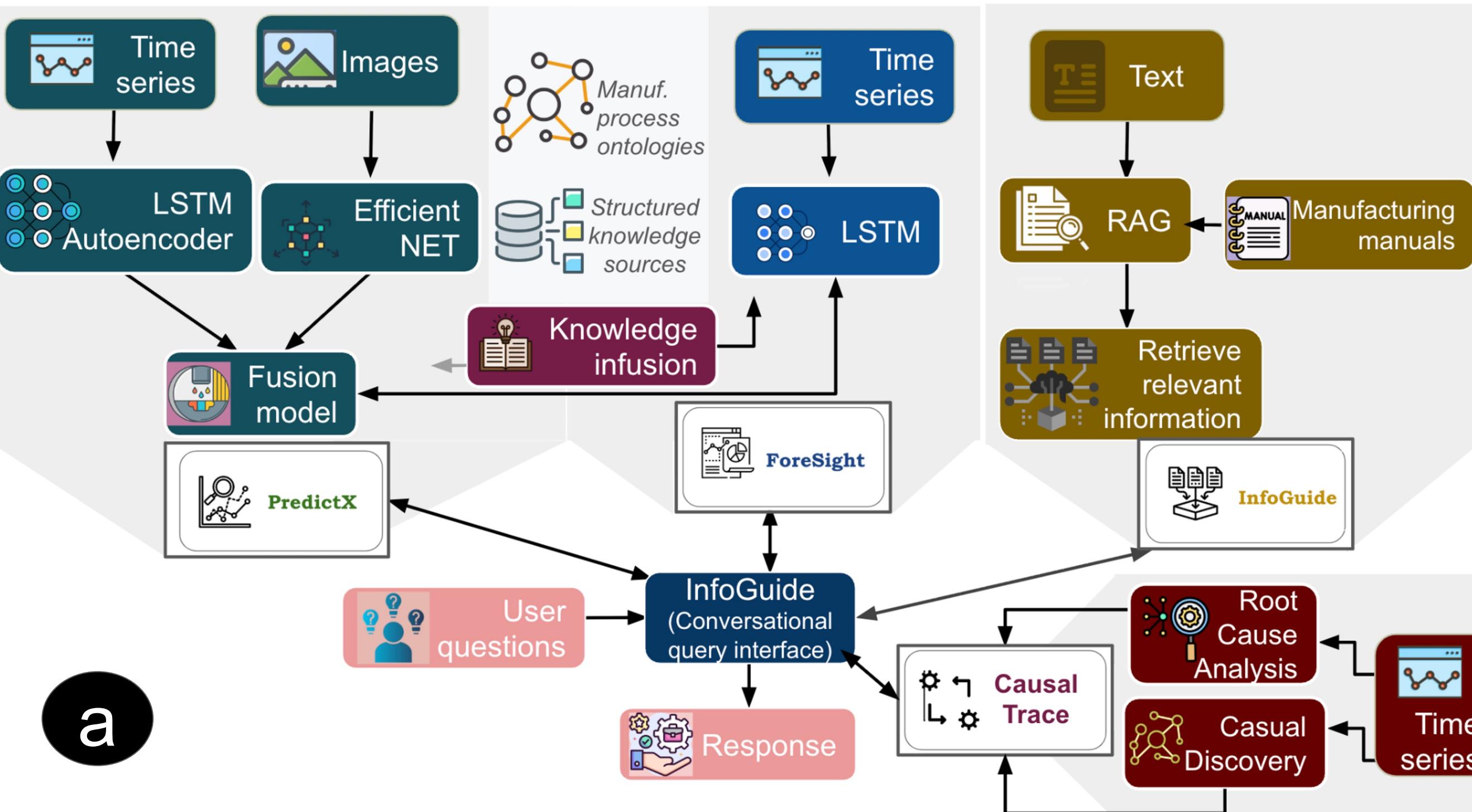
## IBM Portfolio Products Used:

- watsonx Assistant
- Cognos
- watsonx.ai
- Bee Agent (open source)
- watsonx.governance





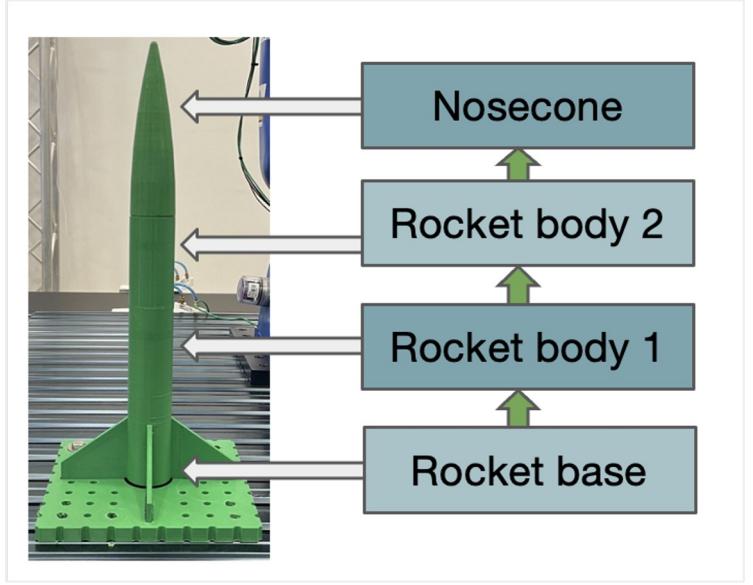
# SmartPilot: Multiagent CoPilot for Adaptive and Intelligent Manufacturing



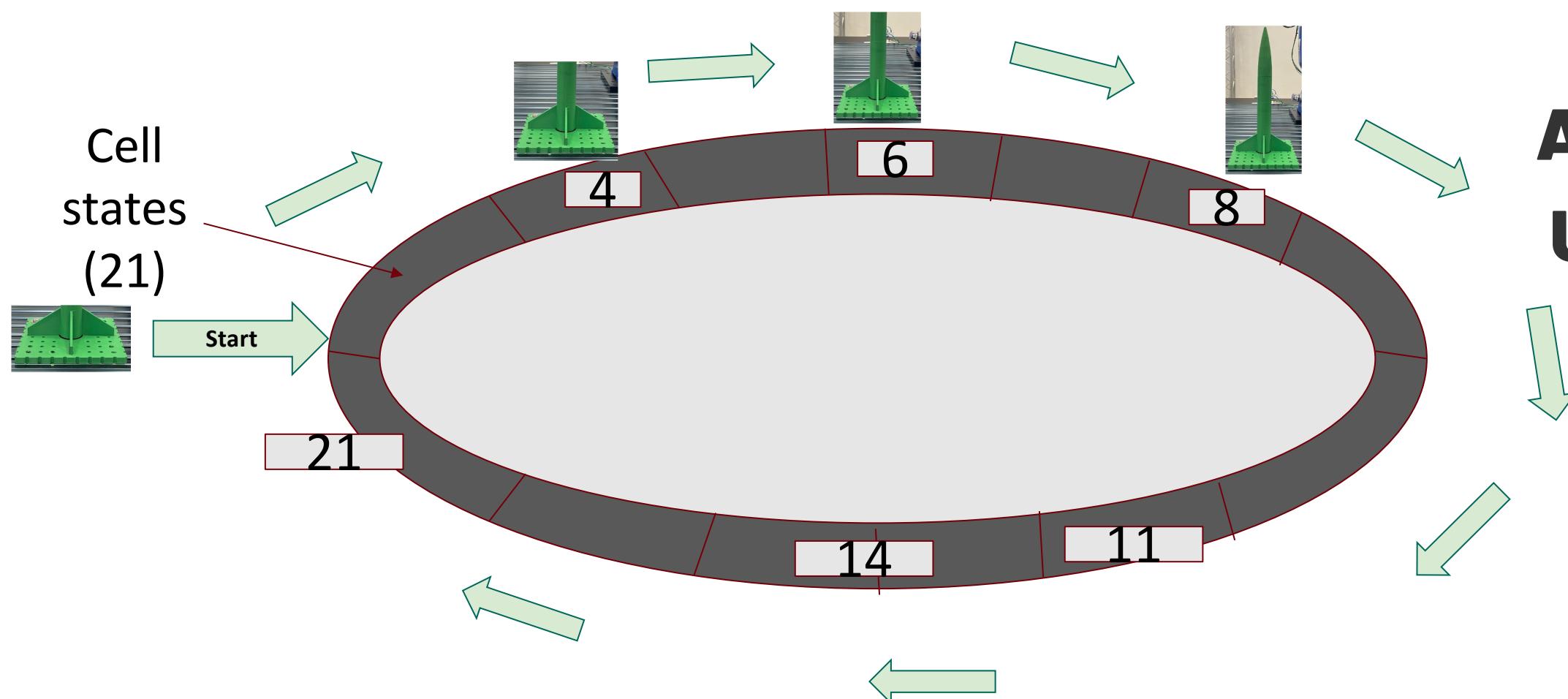
**a)** The C3AN-driven architecture integrates multimodal data: **time series**, **images**, and **text** through agents (**PredictX**, **ForeSight**, **InfoGuide**, and **CausalTrace**) for predictive, diagnostic, and causal reasoning in smart manufacturing.

**b)** Real-world smart manufacturing cell setup with collaborative robots and conveyor systems capturing multimodal sensor data.

**c)** Composite AI ecosystem illustrating **agent interactions and causal relationships** among predictive, forecasting, and reasoning modules enabling autonomous manufacturing intelligence.



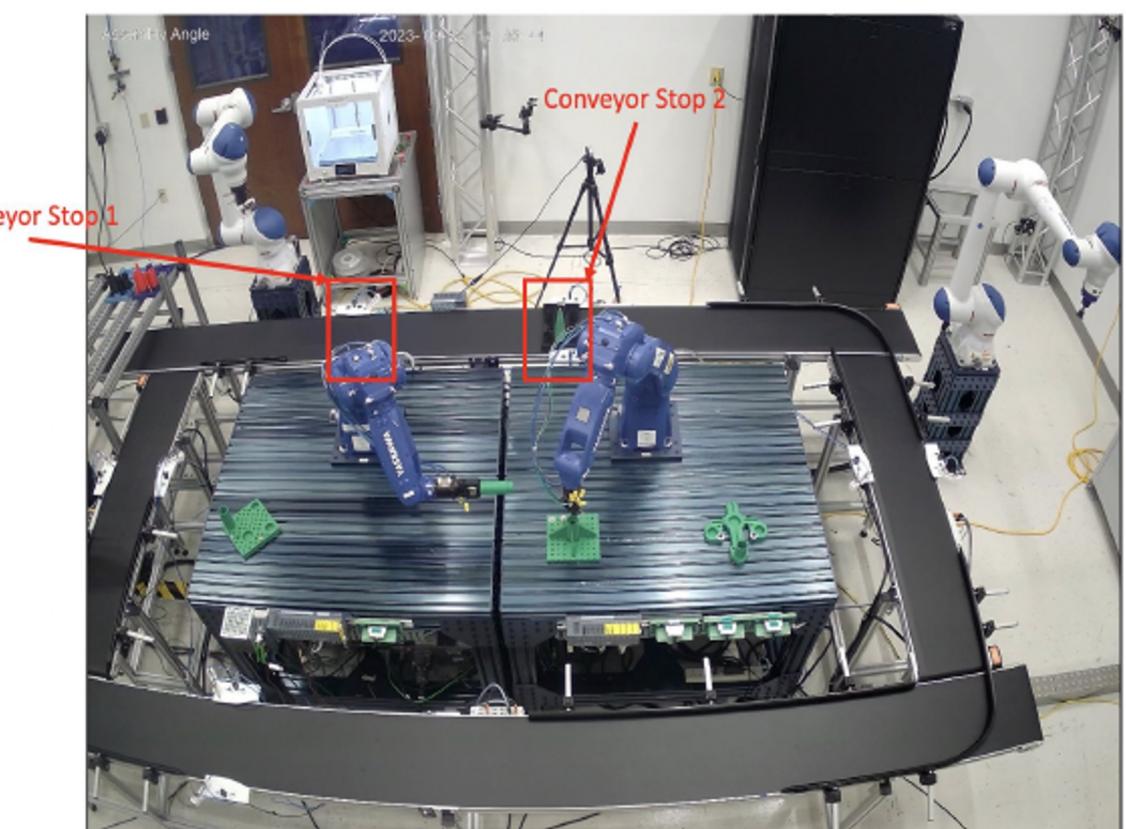
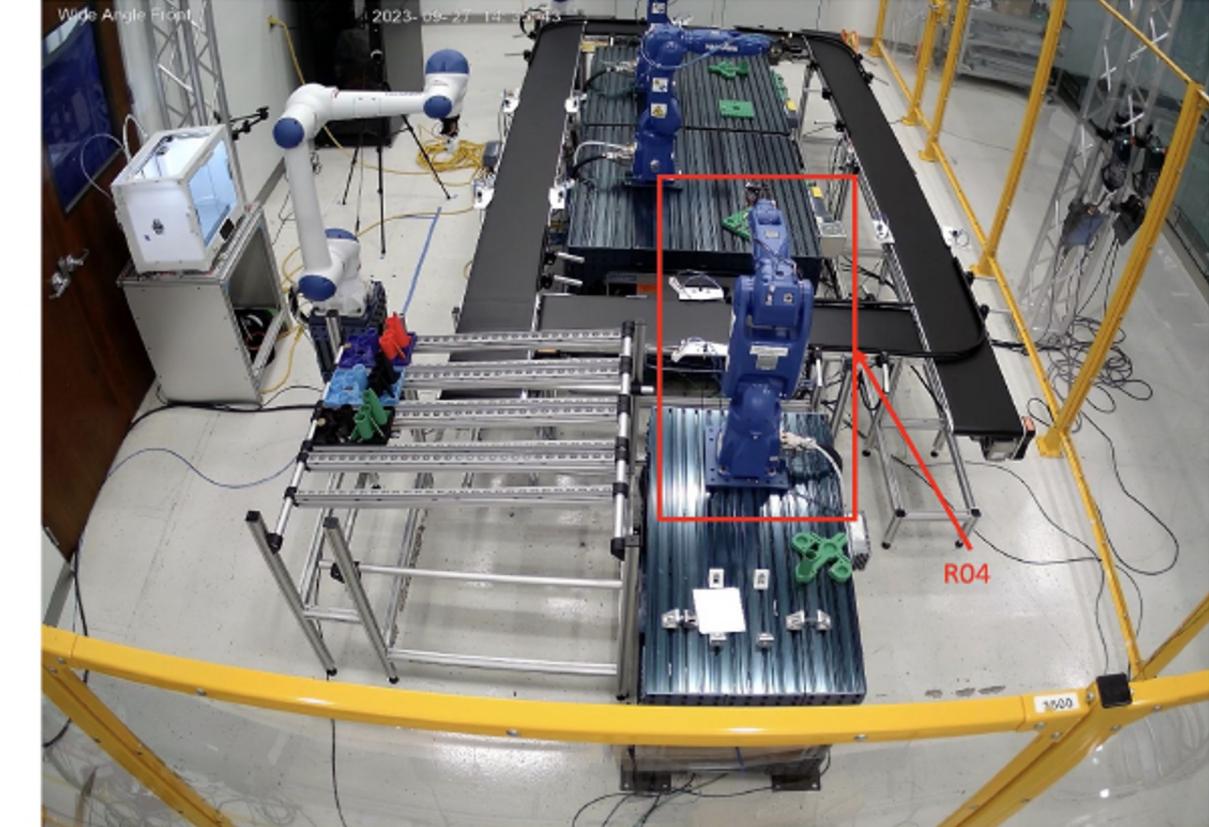
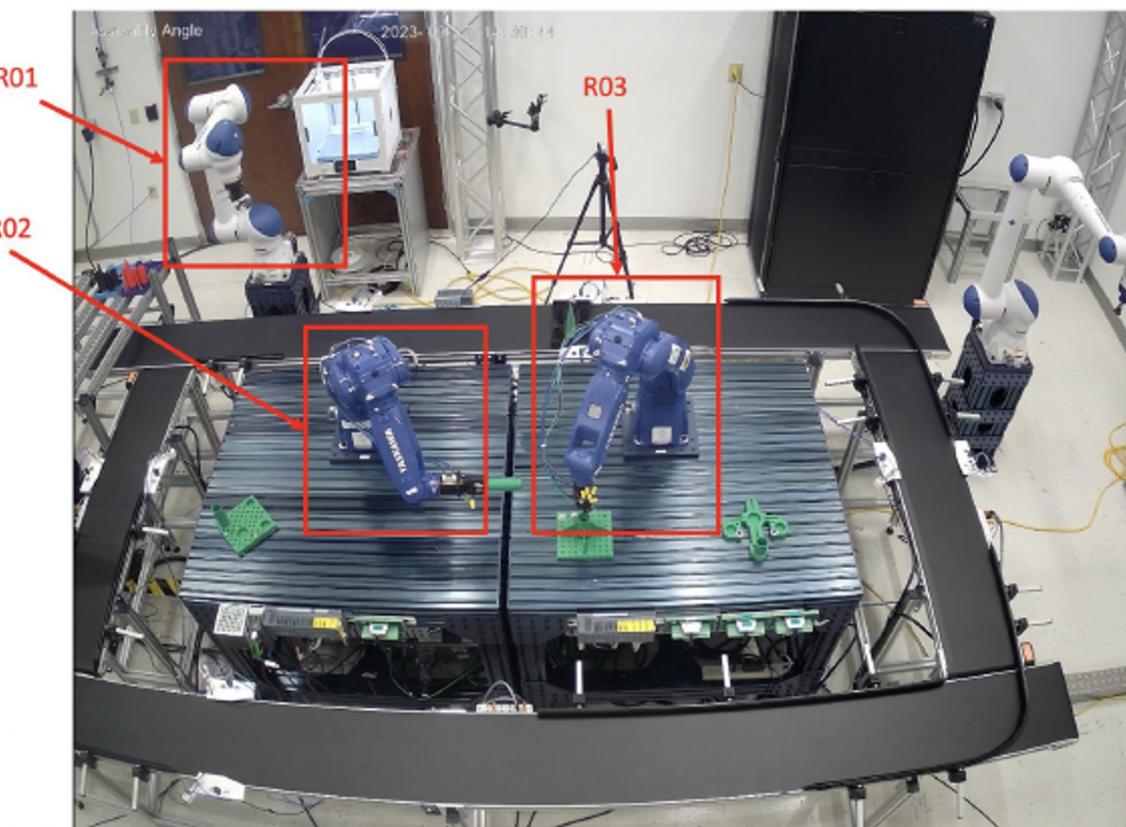
Rocket assembled by the Future Factories Lab. Any missing part is considered an anomaly: absence of Rocket body 1 - "NoBody1," absence of both Rocket body 1 and body 2 - "NoBody2, NoBody1."



The Future Factories (FF) cell consists of 21 distinct states:

- State 4: Body Part 1 is assembled.
- State 6: Body Part 2 is assembled.
- State 8: The Nose cone is assembled.
- States 11-14: The rocket is being disassembled.
- States 15-21: Rocket parts are returned to their original trays.

## Rocket Assembly Use Case



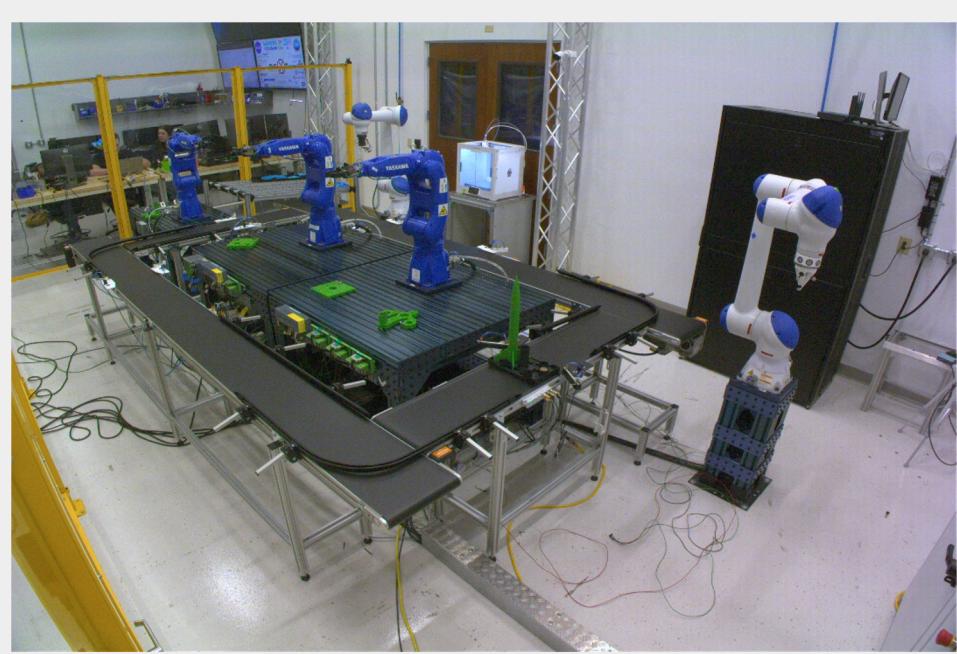
Future Factories Setup. R01-Robot 1, R02-Robot 2, R03-Robot 3, R04-Robot 4

## Multimodal Data (Time series + Images)

```

"records": [
    {
        "time": "50:52.1",
        "Q_VFD1_Temperature": 73.13000488,
        "Q_VFD2_Temperature": 72.13002014,
        "Q_VFD3_Temperature": 73.13002014,
        "Q_VFD4_Temperature": 73.36352539,
        "M_Conv1_Speed_mmmps": 111,
        "M_Conv2_Speed_mmmps": 83,
        "M_Conv3_Speed_mmmps": 111,
        "M_Conv4_Speed_mmmps": 111,
        "L_R01_Gripper_Pot": 5269,
        "I_R01_Gripper_Load": 462,
        "L_R02_Gripper_Pot": 11688,
        "I_R02_Gripper_Load": 1297,
        "L_R03_Gripper_Pot": 2413,
        "I_R03_Gripper_Load": 1384,
        "I_R04_Gripper_Pot": 13243,
        "I_R04_Gripper_Load": 681,
        "L_SafetyDoor1_Status": true,
        "L_SafetyDoor2_Status": true,
        "L_Cell_CycleCount": 1,
        "L_MHS_GreenRocketTray": true,
        "Cam1": "Dataset/BATCH1000/000000_0.png",
        "Cam2": "Dataset/BATCH1000/000000_1.png",
        "Description": "NoNose",
        "CycleState": 2,
        "Cycle_Count_New": 1,
        "actual_state": "Normal"
    }
]

```



Sample images produced for the corresponding sample time series (left)

# SmartPilot Resources



1. Shyalika, C., Prasad, R., Al Ghazo, A., Eswaramoorthi, D.L., Shree Muthuselvam, S. and Sheth, A., 2025, May. SmartPilot: Agent-Based CoPilot for Intelligent Manufacturing. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems* (pp. 3053-3055).
2. Shyalika, C., Prasad, R., Al Ghazo, A., Eswaramoorthi, D., Kaur, H., Muthuselvam, S.S. and Sheth, A., 2025, May. Smartpilot: A multiagent copilot for adaptive and intelligent manufacturing. In *2025 IEEE Conference on Artificial Intelligence (CAI)* (pp. 1-8). IEEE.
3. C. Shyalika, R. Prasad, F. E. Kalach, R. Venkataraman, R. Zand, R. Harik, and A. Sheth. 2025. NSF-MAP: Neurosymbolic Multimodal Fusion for Robust and Interpretable Anomaly Prediction in Assembly Pipelines. In *34th International Joint Conference on Artificial Intelligence (IJCAI)*, August 2025.
4. C. Shyalika, A. Sharma, F. E. Kalach, U. Jaimini, C. Henson, R. Harik, and A. Sheth. 2025. CausalTrace: A Neurosymbolic Causal Analysis Agent for Smart Manufacturing. In *AAAI-IAAI Conference on Artificial Intelligence 40*, January 2026.

# Learn More

## Contact Information:

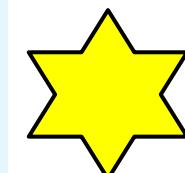
Dhaval Patel  
[pateldha@us.ibm.com](mailto:pateldha@us.ibm.com)

Shuxin Lin  
[shuxin.lin@ibm.com](mailto:shuxin.lin@ibm.com)

Chathurangi Shyalika  
[jayakodc@email.sc.edu](mailto:jayakodc@email.sc.edu)

Saumya Ahuja  
[sauyma.ahuja@ibm.com](mailto:sauyma.ahuja@ibm.com)

[AssetOpsBench Github Repo](#)



Please star our repo!



[AssetOpsBench HuggingFace Dataset](#)



[FailureSensorIQ HuggingFace Dataset](#)



[AssetOpsBench Arxiv Paper](#)



[AssetOpsBench Codabench Competitions](#)



[FailureSensorIQ Kaggle Benchmark](#)



## Join us at our AAAI events

- 22<sup>nd</sup> Paper: Spiral-World Model for Efficient LLM Planning
- 23<sup>rd</sup> Studio Spotlight: Invited talk on AssetOpsBench @IBM Industry 4.0 Studio
- 24<sup>th</sup>, AssetOpsBench-Live Demo, Hall 2 (Booth 10), 12-2 PM
- 24<sup>th</sup>, Agent Benchmarking, Paper 1466:Work Order Agent

Thank you

Any Questions?