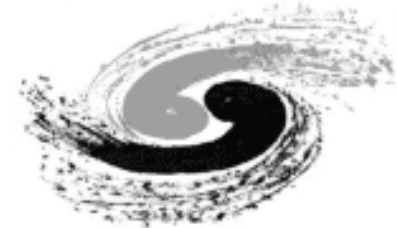


Agentic AI and the Future of Detector Operations in Particle Physics

Bruce Mellado

IHEP at the Chinese Academy of Sciences



中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

**The 2026 International Conference on Instrumentation
for Colliding Beam Physics, Hefei, 15/10/26**

Outline

❑ Introduction

- ❑ The complexity ceiling

- ❑ LLMs: From Reasoning to Agentic AI

❑ Selected Showcases in Instrumentation

- ❑ Telescopes

- ❑ Particle Physics Detector Design

- ❑ Experiments at the CERN

❑ Agentic AI infrastructure

- ❑ CERN

- ❑ IHEP

❑ Summary

The Complexity Ceiling

When data grows faster than understanding

Traditional workflows cannot absorb the scale, diversity and interdependence of modern data.

1 Data explosion



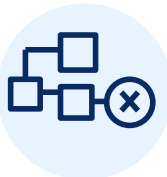
- Petabytes to exabytes
- Continuous streams

2 Heterogeneity



- Sensors, images, text
- Simulation + observation

3 Brittle workflows



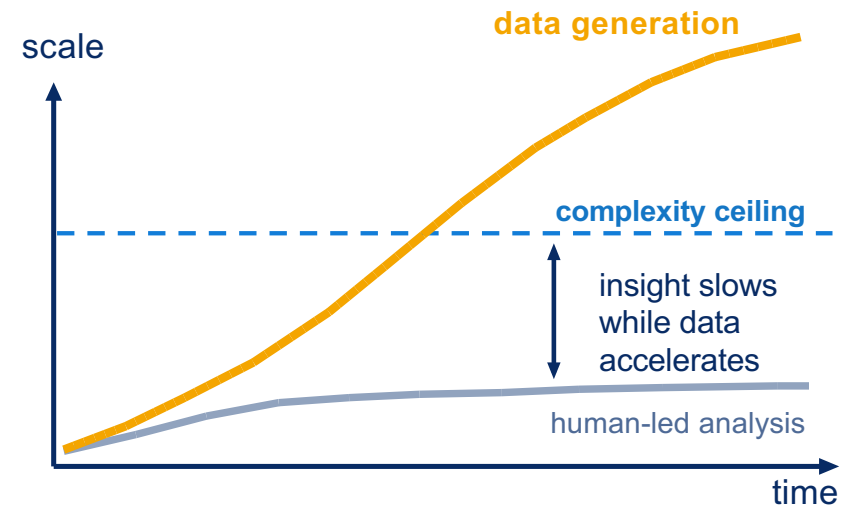
- Bespoke pipelines
- Hard to reuse or scale

4 Human bottleneck



- Scarce expertise
- Rising infrastructure cost

A crisis of scale and interpretation



The challenge is not only computational;
it is epistemic.



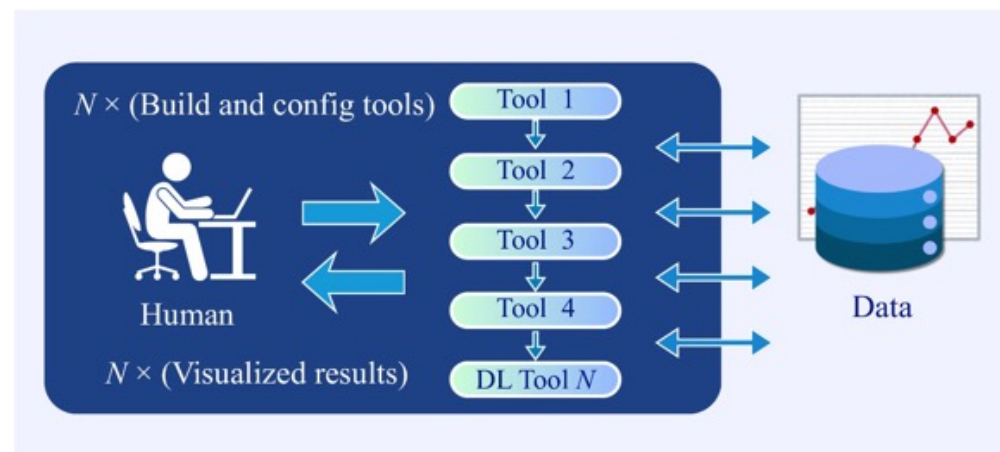
Message

Unless analysis itself evolves, the next generation of experiments may produce insights more slowly, less completely, or not at all.

SCIENTIFIC WORKFLOWS: TWO PARADIGMS

OLD PARADIGM

MANUAL, TOOL-BASED ANALYSIS

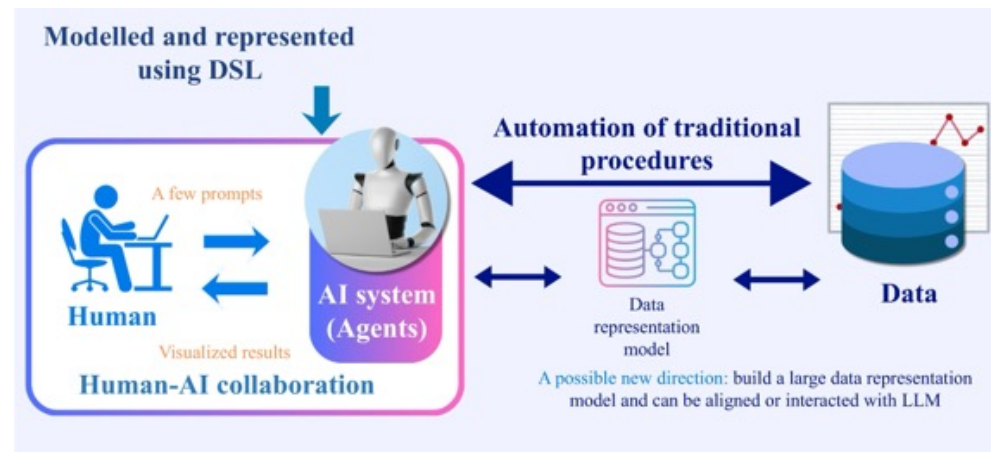


People configure, execute and supervise every stage of the data pipeline.

As data volumes and complexity increase, manual coordination becomes a bottleneck.

NEW PARADIGM

HUMAN-AI COLLABORATION



AI agents orchestrate established analysis tools through domain-specific languages.

Automation scales the workflow while retaining transparency, scientific control and human oversight.

KEY SHIFT From hands-on tool management to human-supervised, agent-orchestrated research workflows

Reasoning in LLMs

How language models work through difficult questions



Reasoning-capable LLMs perform multistep analysis before producing an answer.



1 Break down

Identify steps and constraints



2 Evaluate

Compare possible explanations



3 Refine

Check and improve the answer

**DETECTOR
EXAMPLE**

Noise rising? Compare causes → recommend diagnostics



Reasoning recommends actions; it does not itself execute them.

Agentic AI: Reasoning into Action

Reasoning is the engine; agents add tools, execution and feedback

REASONING LLM

Figures out what to do



AGENTIC SYSTEM

Does it, checks, and adapts

HOW AN AGENT OPERATES

Reason & plan



Use tools



Check result



Adapt & repeat

DETECTOR MONITORING

Reasoning: suggests a high-voltage diagnostic.

Agent: queries DQM → runs checks → alerts the operator.



Agents act within defined permissions and human oversight.

FROM GENERATIVE AI TO REASONING AND AGENTS

Selected model and agent milestones | 2018 – October 2026

FOUNDATION & GENERATIVE MODELS • 2018 – 2023

Large-scale pretraining enables general-purpose generative AI



REASONING BECOMES MAINSTREAM • 2024 – 2025

Models with stronger reasoning, tool use and complex problem solving

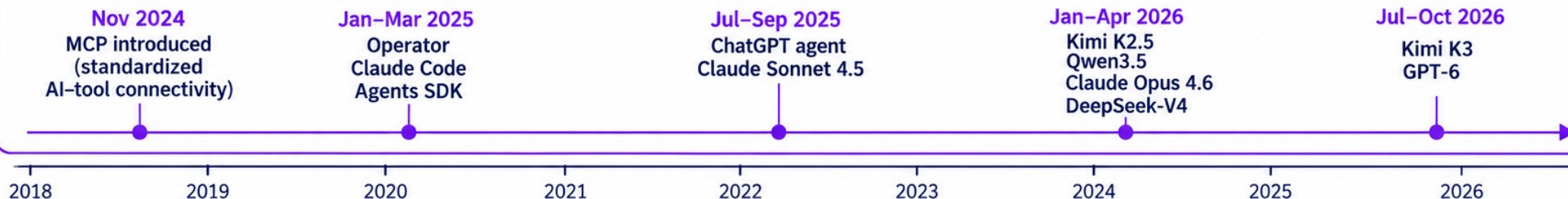


Reasoning era

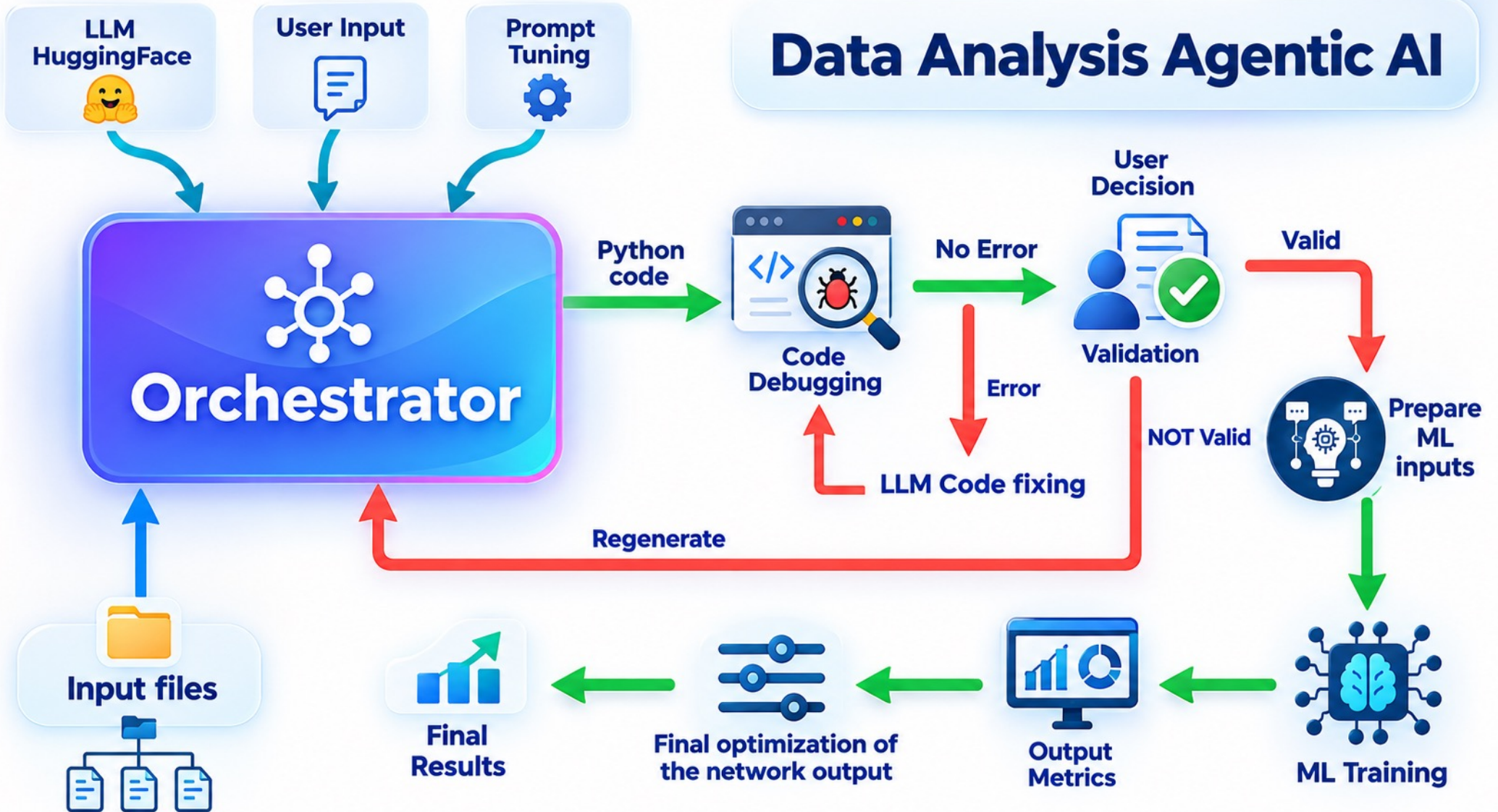
Agentic era

AGENTIC SYSTEMS SCALE • 2024 – 2026

From copilots to autonomous, tool-using agents



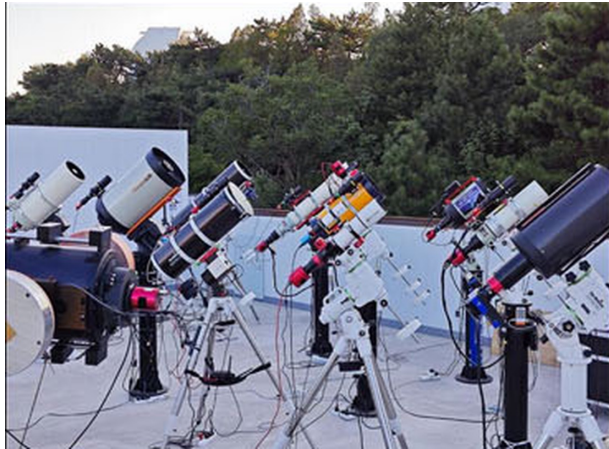
Data Analysis Agentic AI



**Selected
Showcases of
Agentic AI for
Instrumentation**

StarWhisper: Telescopes on Autopilot

Nearby Galaxy Supernovae Survey (NGSS) • optical transient monitoring



THE EXPERIMENT

A network of 10 optical telescopes at four sites in China repeatedly surveys nearby galaxies.

WHAT IT MEASURES

CMOS images in LRGB filters track nearby galaxies and flag new transient sources, including supernovae.

10 telescopes

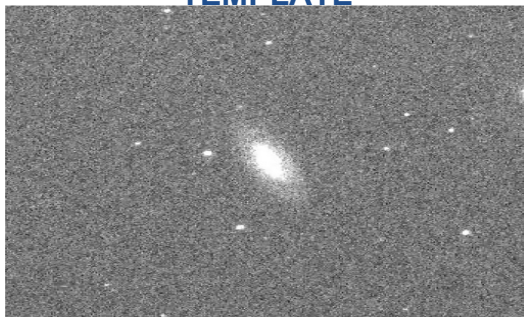
4 observing sites

3,000+ galaxies on clear nights

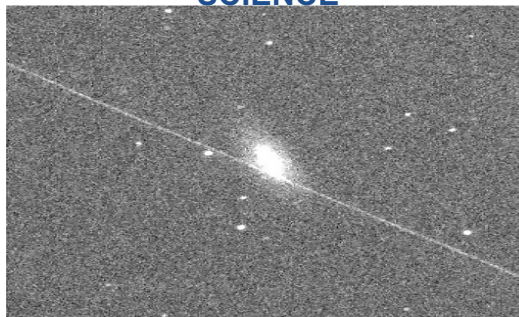
EXAMPLE TRANSIENT DETECTION

Reference image, new observation and difference image for AT2025pk near NGC 0615

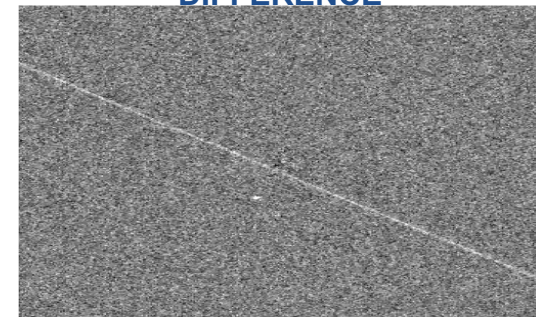
TEMPLATE



SCIENCE



DIFFERENCE

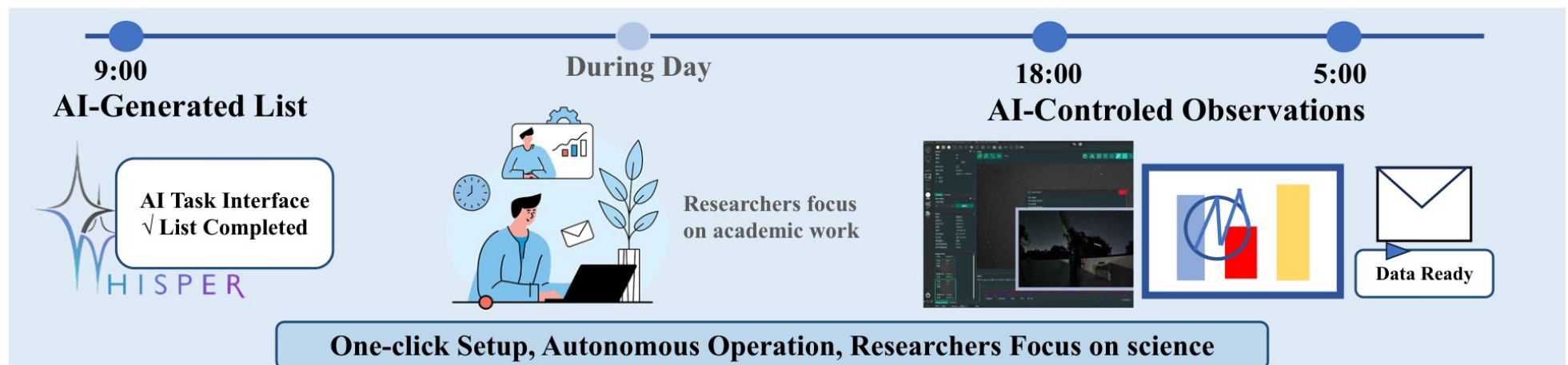
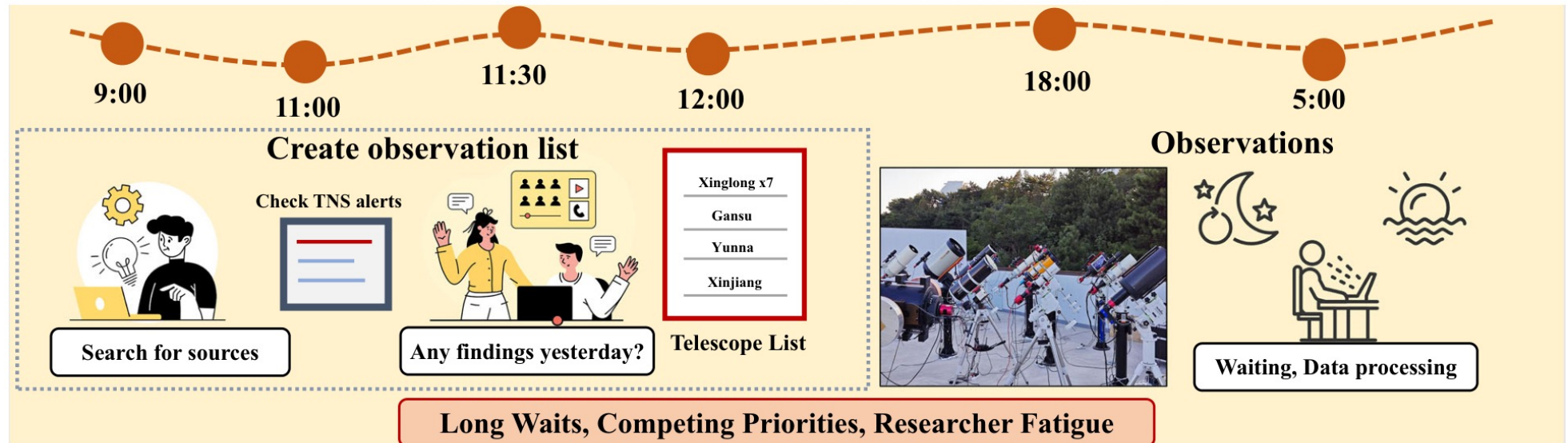


Template → new observation → subtraction reveals a transient candidate

StarWhisper: Why Agentic AI Matters

From manual observing to autonomous operations and science-ready data

Traditional vs. SWT-Enhanced

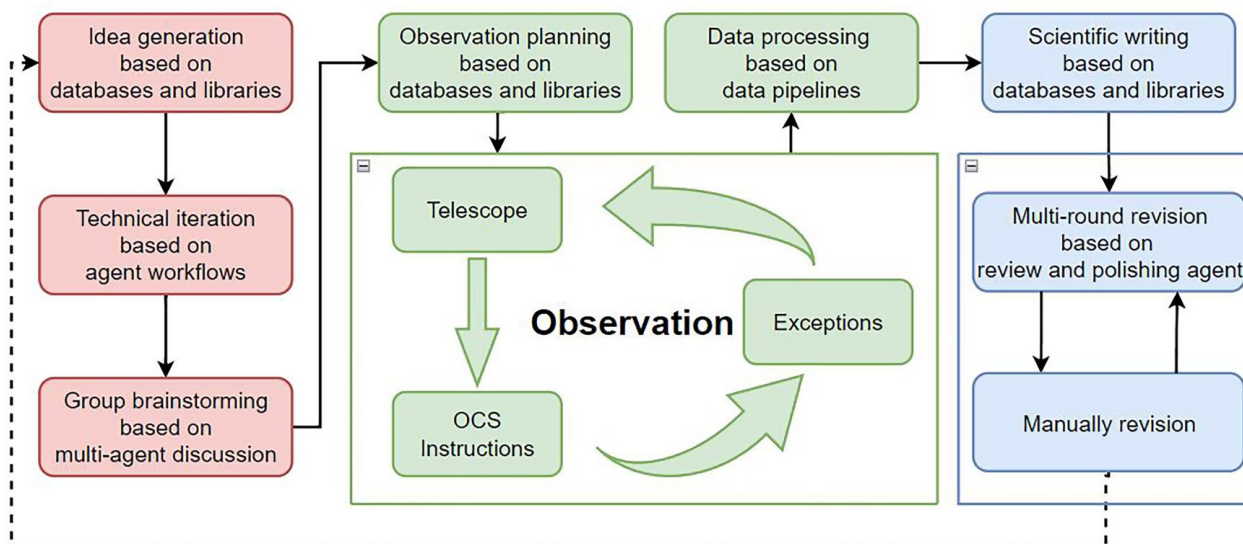


Agentic AI automates planning, execution and data handoff — researchers focus on science.

StarWhisper: Toward an AI Astronomer

A proposed closed loop: scientific ideas, observing, data and new insights

AI ASTRONOMER • PROPOSED WORKFLOW



1 GENERATE & PLAN

Agents use literature and datasets to develop ideas and observation strategies.

2 OBSERVE & PROCESS

Telescopes observe; control handles exceptions; pipelines process data.

3 WRITE & ITERATE

Scientific writing and human review feed results back into the next round of ideas.

ALREADY DEPLOYED

NGSS observation plans, telescope control, image pipelines and transient alerts.

FUTURE EXTENSION

Hypothesis generation, coordinated follow-up and writing with feedback.

Vision: hypotheses → observations → results → new hypotheses

JW-ASTClaw: The Solar Telescope

Solar Full-disk Multi-layer Magnetograph (SFMM) • Chinese Meridian Project



SFMM telescope • Ganyu, China

THE EXPERIMENT

At Ganyu, the SFMM monitors the full solar disk from the photosphere to the chromosphere for space-weather research.

WHAT IT MEASURES • FOUR SPECTRAL CHANNELS

Fe I 5324 Å

Photospheric magnetic field

H β 4861 Å

Doppler velocity

Ca II 8542 Å

Chromospheric activity

H α 6563 Å

Filaments and flares

AGENTIC CONTROL • EDITABLE ARCHITECTURE

PERCEIVE

DQM, sky conditions, flares



REASON

Balance goals and constraints



EXECUTE

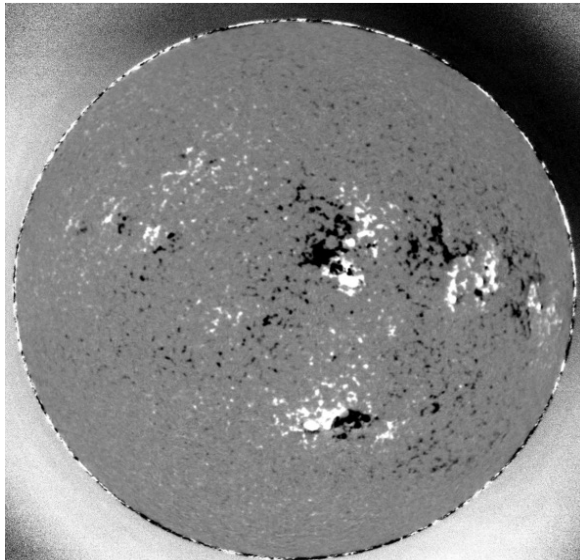
MCP commands + safety checks

Agentic decisions adapt how the solar instrument observes.

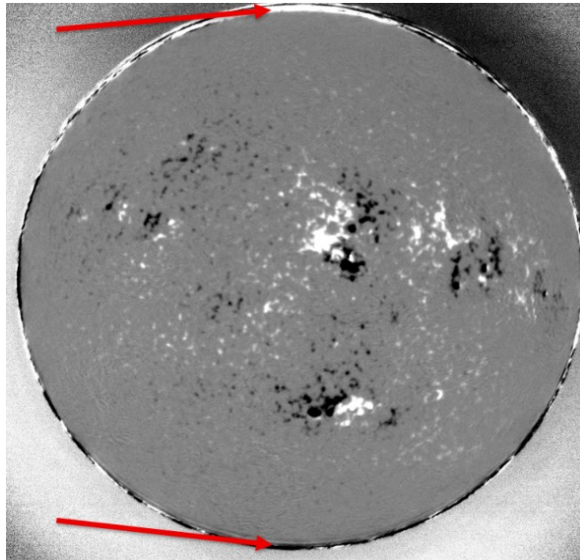
JW-ASTClaw: DQM Triggers Action

Real-time image quality and sky conditions feed autonomous observation decisions

NORMAL MAGNETOGRAM



WIND-INDUCED JITTER



WIND / DQM

Detect motion artifacts and compromised exposures.

CLOUDS / SCHEDULING

Adjust observing plans or pause for approaching clouds.

ALL-SKY CAMERA • SOLAR DISK, MOTION AND CLOUD ASSESSMENT



FLARE / FAST MODE

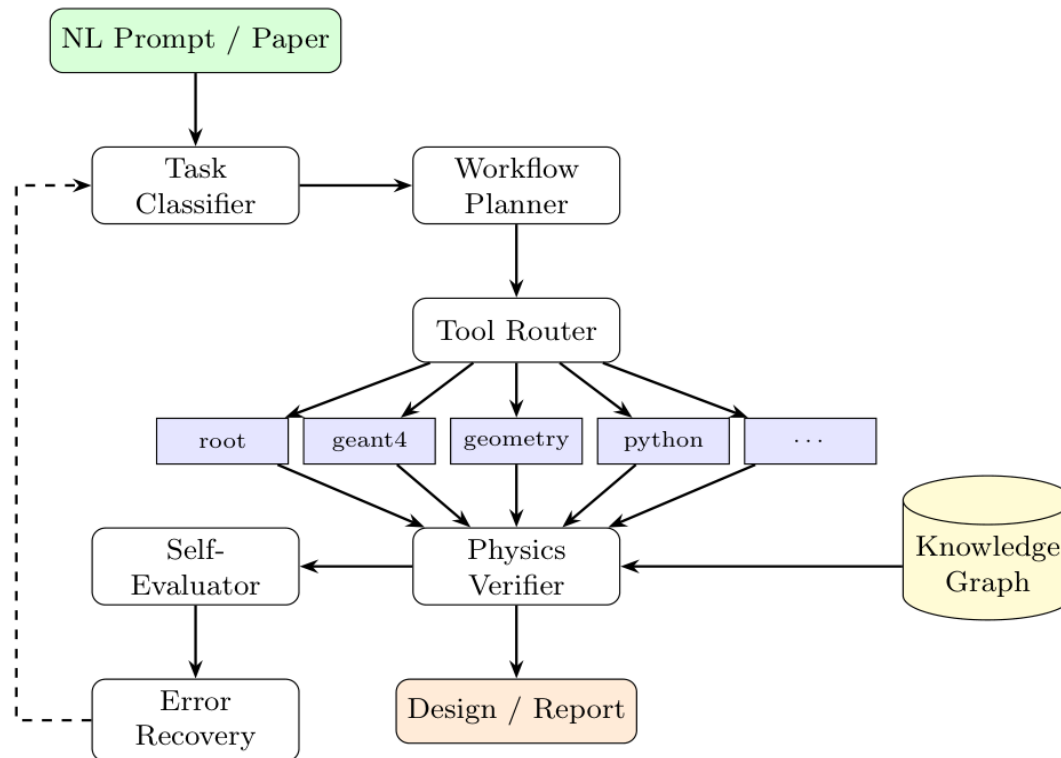
Switch to rapid 2 s H α observations with HDR (High Dynamic Range) exposures.

MCP actions pass safety checks; rule-based and manual fallbacks remain.

GRACE: Agentic AI Designs Detectors

From a physics brief or paper to simulated, verified design alternatives

GRACE • AGENT ARCHITECTURE



WHAT IT DESIGNS

Detector geometry, materials and sensor configurations, guided by physics and practical constraints.

HOW THE AGENT WORKS

Builds simulation workflows, calls Geant4 and other tools, verifies physics, and iterates when needed.

DESIGN EXAMPLES

Electromagnetic calorimeters, muon systems and optical detectors.

CALORIMETER DEMO • ENERGY RESOLUTION (SIMULATED)

2.26% → 1.38% Projective PbWO₄ towers

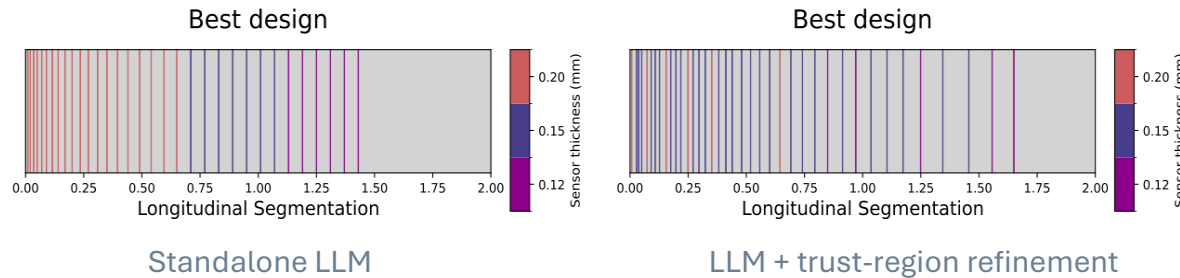
Toy-study result; the paper notes omitted real-world effects.

✓ Agents propose geometry, simulate detector response and verify physics.

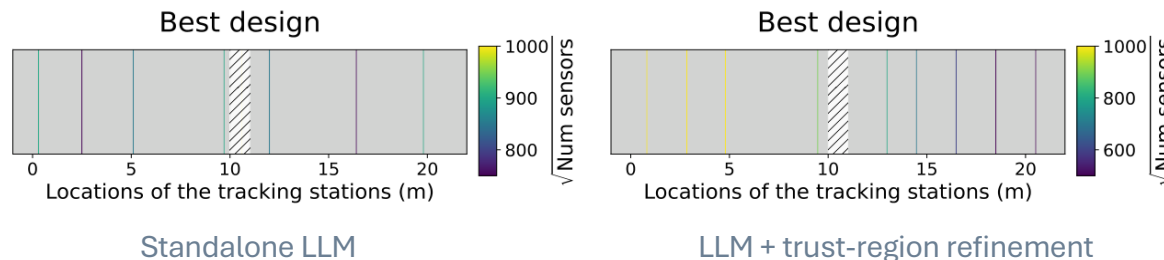
LLMs as Detector-Design Meta-Planners

Two physics testbeds: calorimeter layers and spectrometer stations

SAMPLING CALORIMETER • LAYER PLACEMENT & THICKNESS



MAGNETIC SPECTROMETER • TRACKING-STATION DESIGN



PROPOSE

LLMs suggest layer types, sensor positions and tracking layouts.

CHECK

Enforce hard geometry and budget limits; score with physics simulation.

OPTIMISE

Local refinement improves placements; RL (Reinforcement Learning) achieves stronger final designs.

100 GeV MOMENTUM RESOLUTION • EXAMPLE BENCHMARK

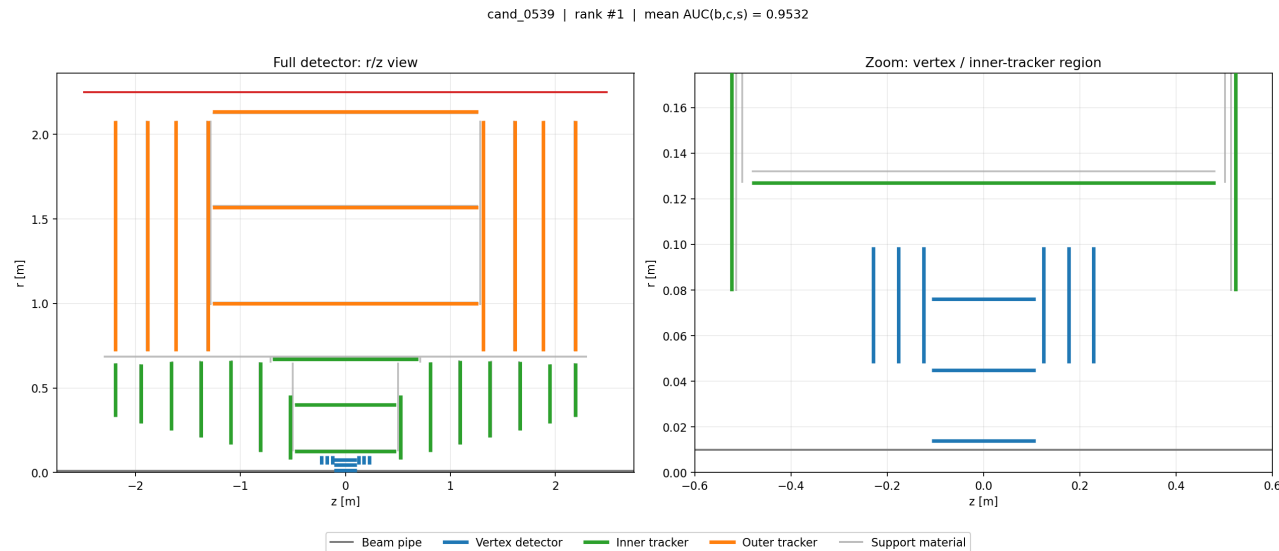
Baseline momentum resolution 13.86% → LLM+TR 7.97% ≈ RL 7.95%

✓ LLMs find feasible layouts; dedicated optimisers still drive the best performance.

FCC-ee: AI Optimises Vertex Detector Design (Not LLM)

11-parameter silicon tracker • closed-loop, multi-fidelity Bayesian optimisation

DETECTOR GEOMETRY • SIMULATED OPTIMISED LAYOUT



DESIGN SPACE

3 barrel layers + 3 endcap disk pairs

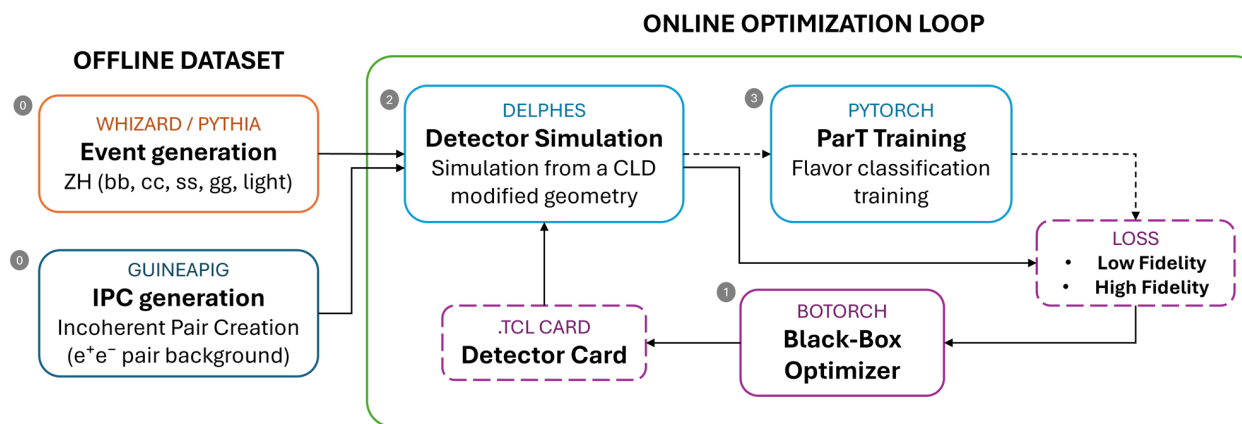
11 VARIABLES

Positions, spacing, thickness and resolution

HARD CONSTRAINTS

Beam pipe, no overlaps, ≥ 6 tracker hits

AUTONOMOUS DESIGN LOOP • BAYESIAN OPTIMISATION



MULTI-FIDELITY SEARCH

~1 min low-fidelity IP proxy

~1 h simulation + ParT tagger

BEST MACRO-AUC 0.9532

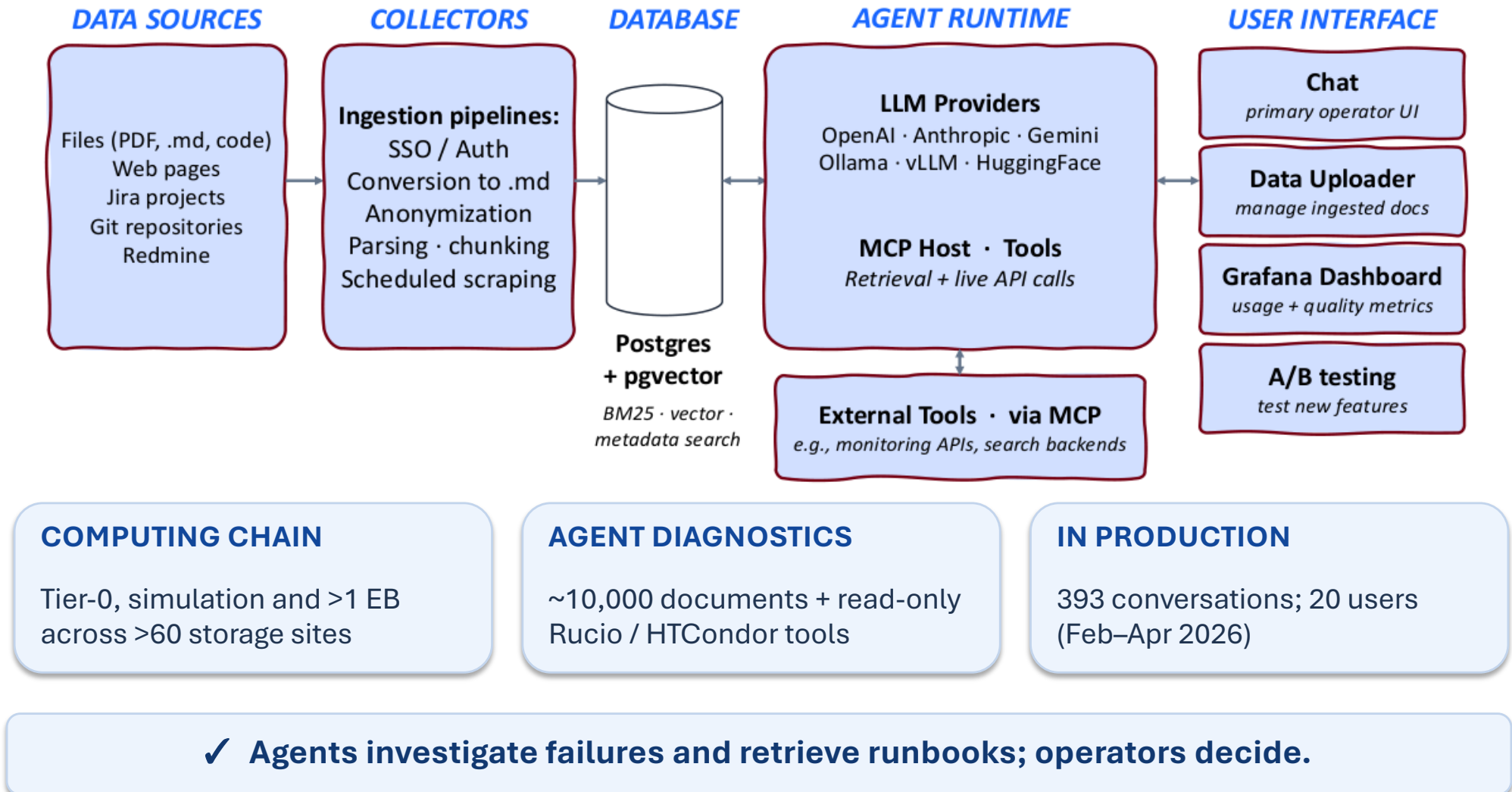
154 high-fidelity + 1,682 low-fidelity evaluations

✓ AI proposes feasible geometries; simulation and performance feedback refine each design.

CMS Archi: Agentic Computing Operations

CMS collision data → reconstruction, production and worldwide distribution

ARCHI • DOCUMENTATION + LIVE MONITORING + TOOL-ASSISTED REASONING

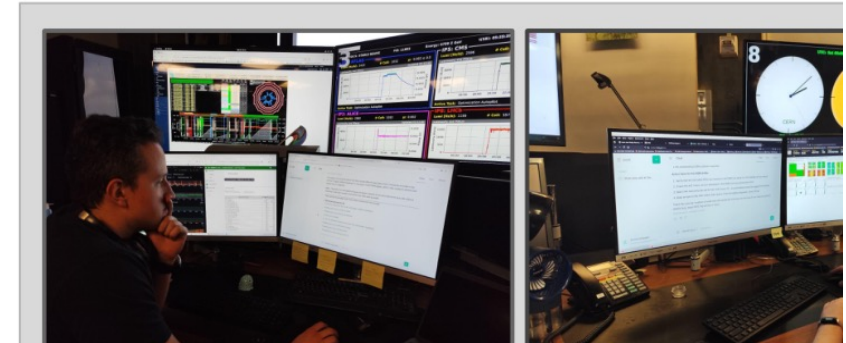


CMS WisDQM: Agentic Support for Shifters

Online DQM, Online Trigger and Offline Tracker DQM • FNAL Remote Operations Center

CMS DQM • >200 PLOTS PER RUN

SHIFTERS USING WisDQM AT FNAL



ASK



RETRIEVE



GUIDE

SHIFT WORKLOAD

Three 8-hour shifts daily; shifters inspect detector status and call experts

LIVE PILOT

Deployed at FNAL (April 2026); GPT-OSS 120B and daily CMS run reports

SHIFTER FEEDBACK

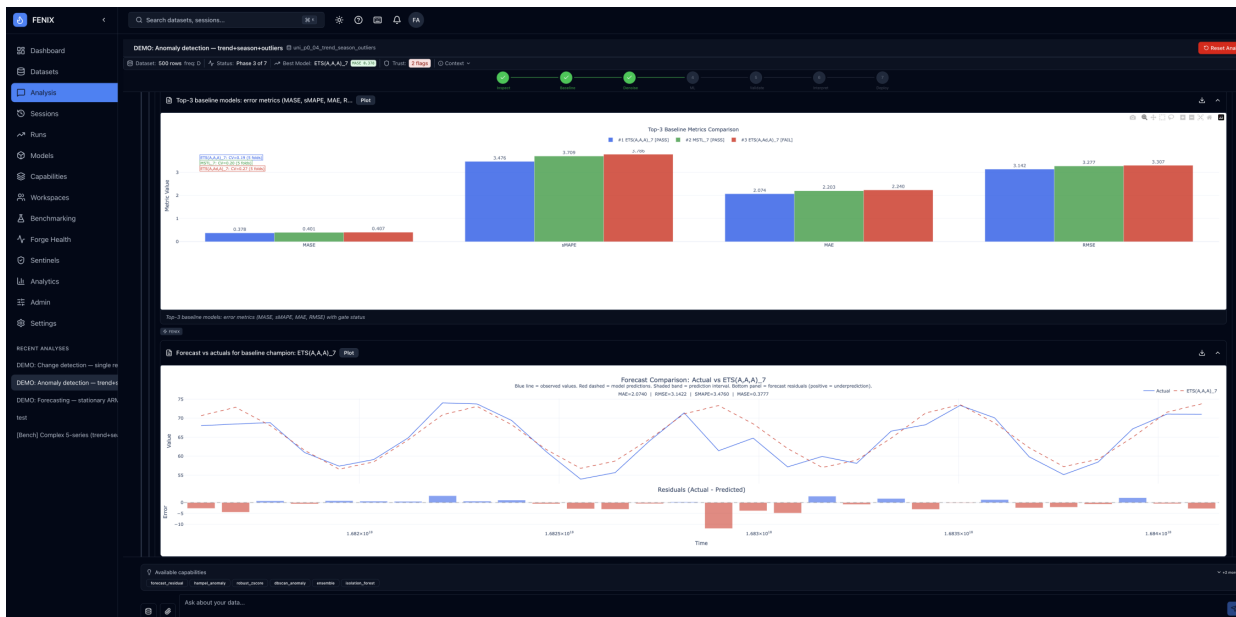
13.9 s mean response; useful for procedures and shift priorities

✓ The shifter retains DQM decisions; direct GUI plot access is planned.

Fenix: Agents for Detector Infrastructure

CERN EP-DT • predictive maintenance across critical control and safety services

FENIX FORGE • TIME-SERIES INTELLIGENCE



FORGE | ANALYZE

Forecasting, anomaly and change detection

SENTINEL | MONITOR

Data quality, drift checks and model fallback

EMBER | RECOMMEND

Agent reasons over evidence and maintenance context

**MAY 2026: Sentinel initial version;
Ember MVP**

7-PHASE MODEL DEVELOPMENT

Inspect → Model → Validate → Deploy

Evidence ledger • policy gates • reproducible results

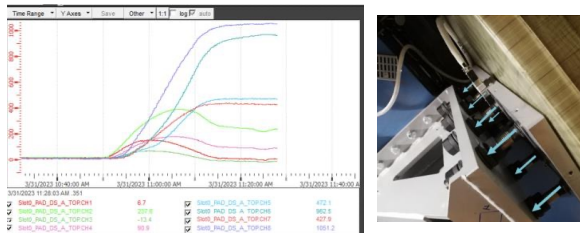
✓ **From anomalies to maintenance advice — operators retain control.**

Genesis: Specialized Agents, One Platform

CERN EP-DT • three May 2026 pilots share retrieval, tools, access control and knowledge

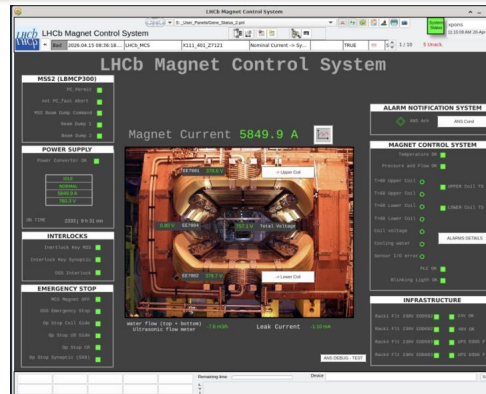
THREE OPERATIONAL PILOTS • EXPERTISE REMAINS DOMAIN-SPECIFIC

SPARK Magnet strain



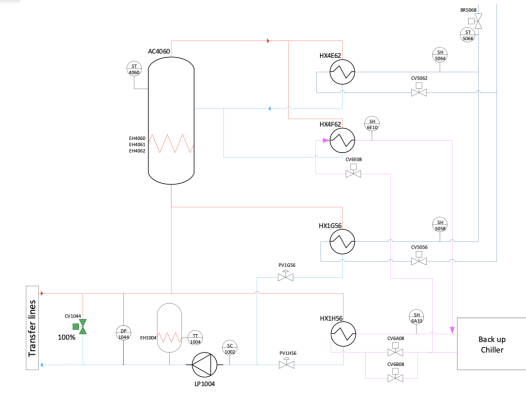
Selects strain sensors, bridge configurations and signal chains.

RISEN PLC controls

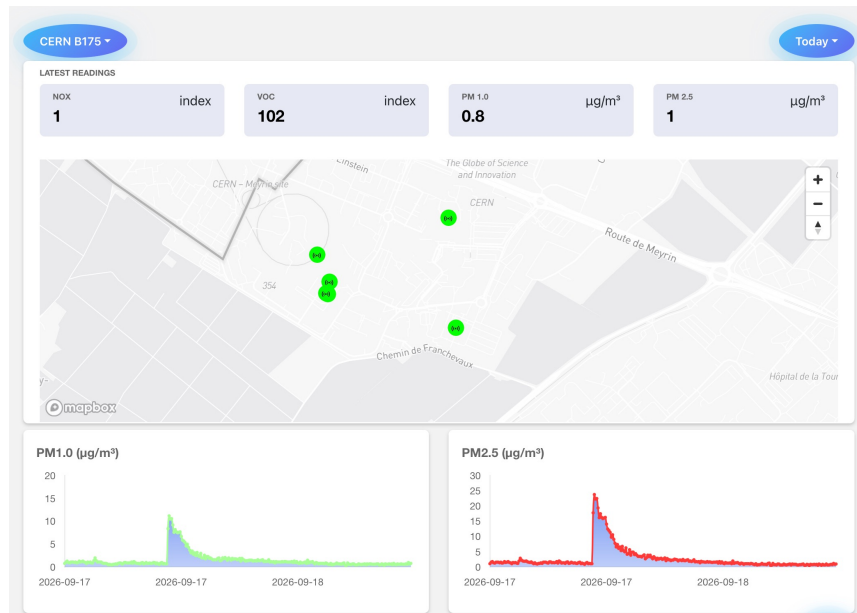


Drafts functional analyses from P&IDs; answers PLC and spec-file questions.

APEX CO₂ cooling



Environmental sensing at CERN

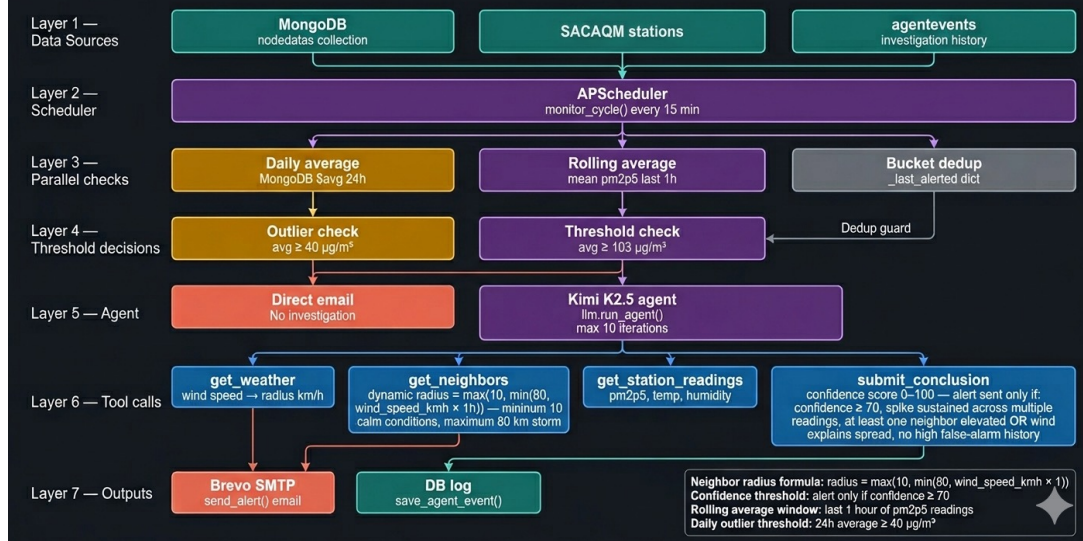


Environmental spikes and alerts handled by Agentic AI

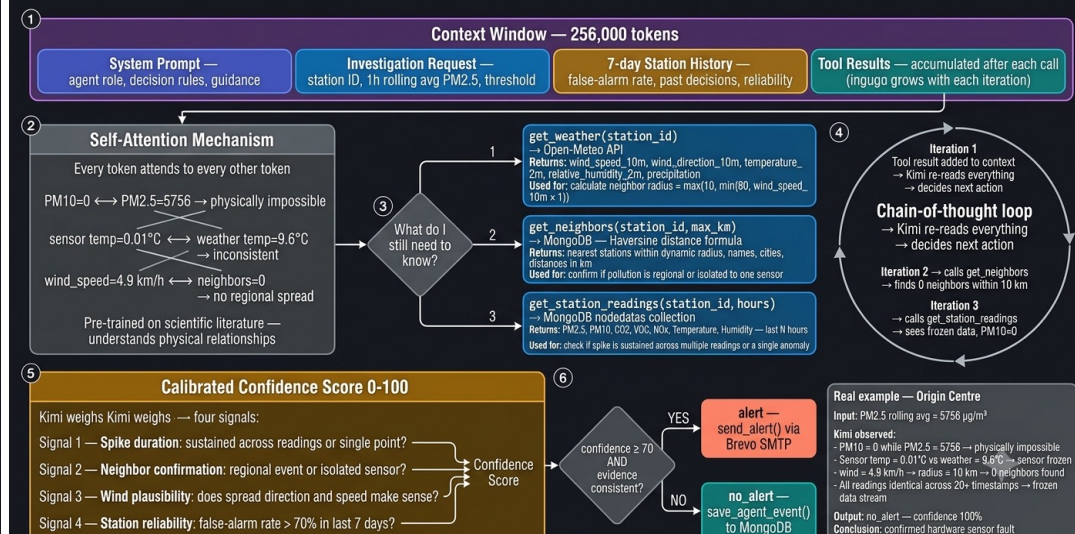
Use several models, such as Kimi, Claude and others.

Agent constrained to a limited number of queries and tasks

SACAQM AI Agent — Technical Architecture



How Kimi K2.6 Reasons — Internal Mechanism

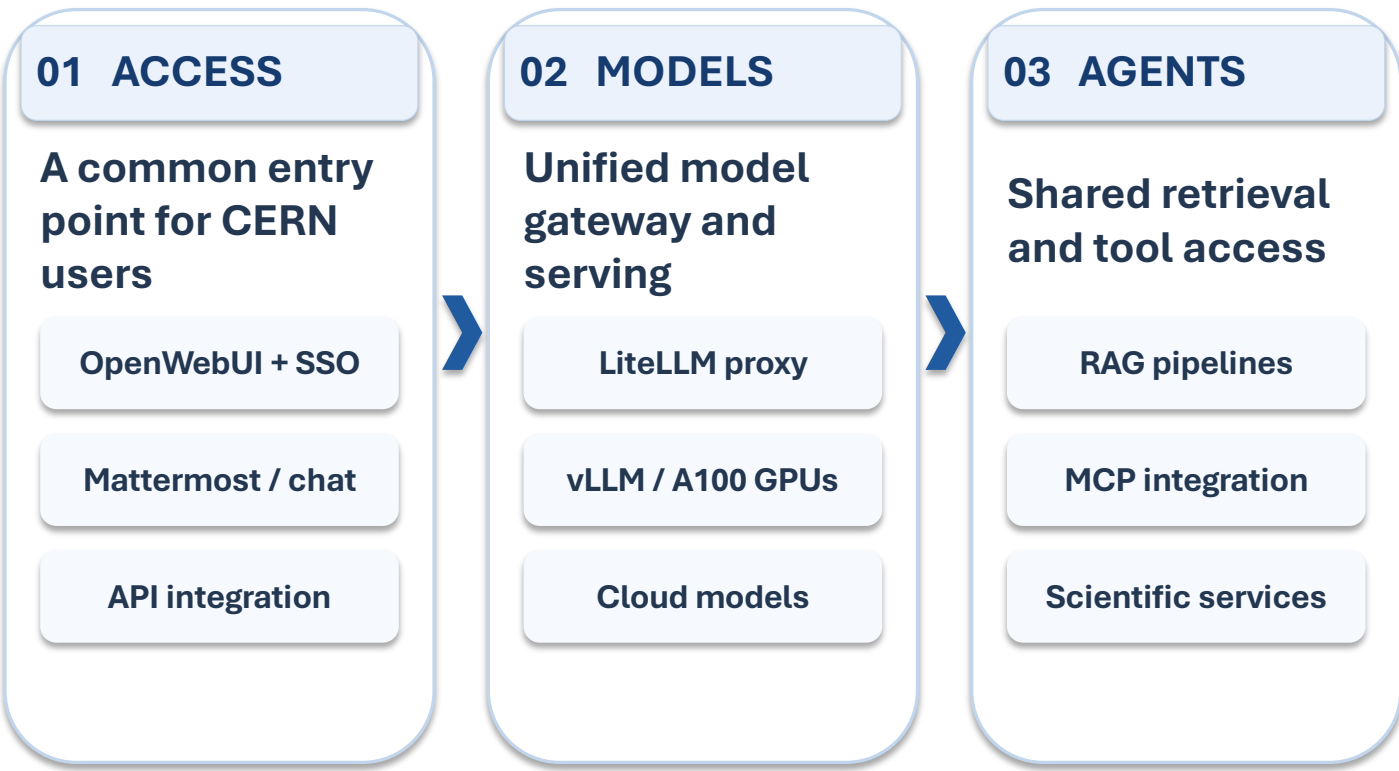


Agentic Infrastructures

CERN: Shared Agentic AI Infrastructure

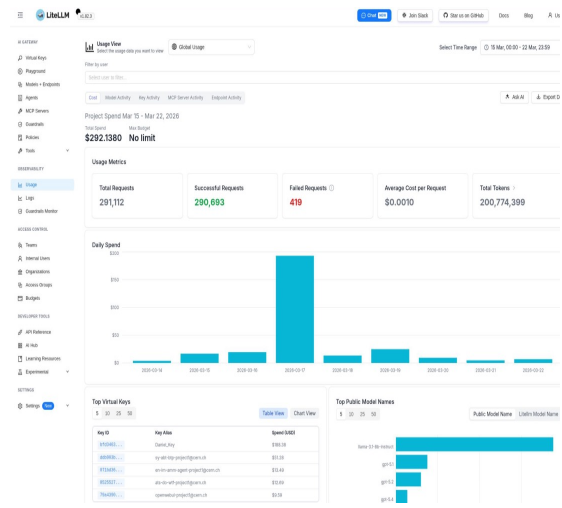
ItGPT / (AI)² building blocks are being extended into the central CERN ML service

ESTABLISHED BUILDING BLOCKS • ITGPT / (AI)²



MCP tools: ROOT • CDS • Zenodo • Indico

REAL CERN GATEWAY



LiteLLM operations dashboard

600+ AccGPT users

80+ LLM use cases identified

100+ ItGPT participants

✓ One shared AI stack for CERN experiments — not isolated pilots.

CERN IT: From Pilots to a Central AI Service

The 2026 plan separates shared infrastructure from experiment-specific science

ROADMAP PRESENTED AT HEPIX SPRING 2026

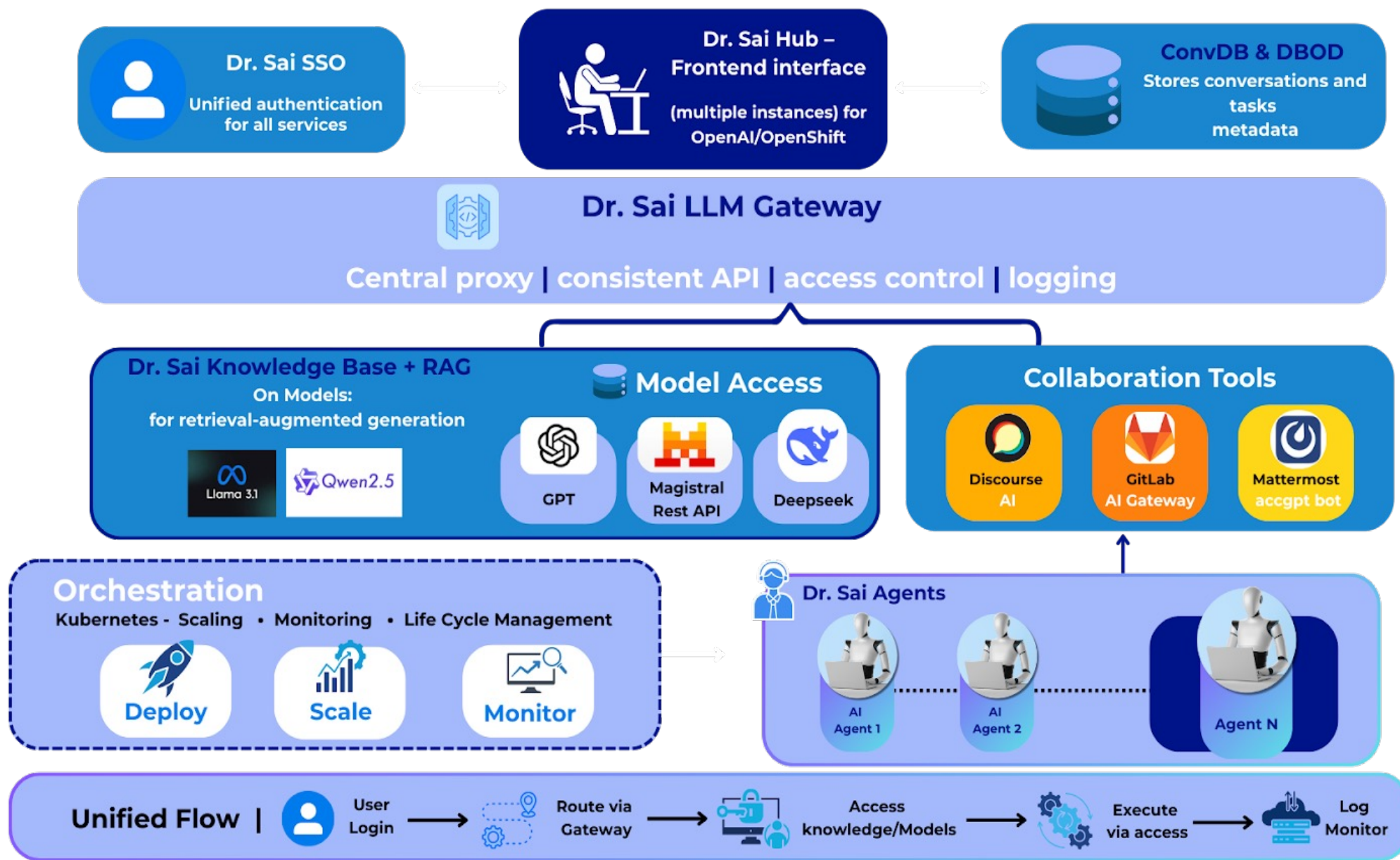


ONE PLATFORM • DISTINCT OWNERSHIP



✓ **CERN IT runs the platform; experiments own the science.**

Dr Sai AI Agent Infrastructure



Agentic AI: The Emerging Instrumentation Layer

Across detector R&D, monitoring and operations, a common pattern is taking shape

THREE TAKEAWAYS

01 DESIGN

Explore detector configurations

Propose new layouts; test them with simulation and physics constraints.

EXAMPLES IN THIS TALK

GRACE • FCC-ee

02 OPERATE

Diagnose and adapt in real time

Combine DQM, telemetry and tools to alert, advise or execute bounded actions.

EXAMPLES IN THIS TALK

Telescopes • CMS • Fenix

03 SCALE

Make agents reusable and trusted

Shared models, RAG and MCP tools with clear domain ownership.

EXAMPLES IN THIS TALK

CERN IT • Genesis • Dr Sai

ESSENTIAL GUARDRAILS

Physics checks • Controlled tools • Audit trails • Human oversight

✓ **From isolated pilots to physics-grounded, human-supervised instruments.**