Agentic AI for Particle-Physics Detector Operations: Towards an AI-Assisted Shifter

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 Workshop on the High Energy CEPC, 23/10/26

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

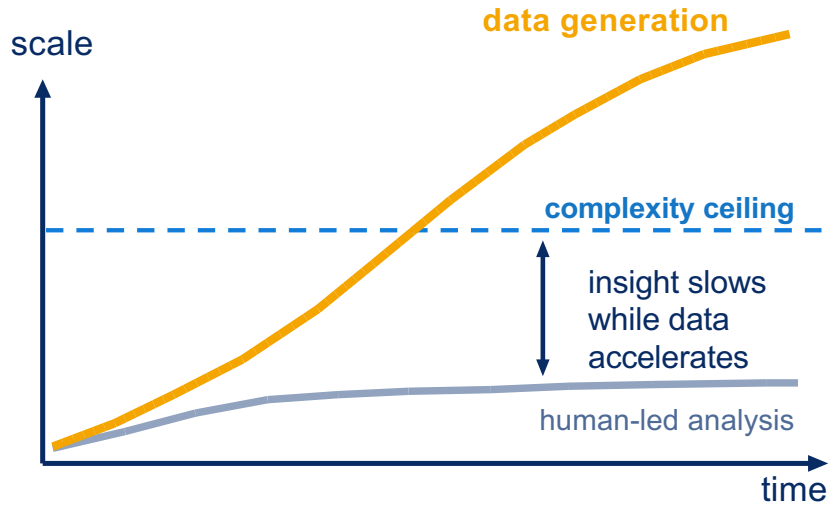
3 Brittle workflows

- 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.

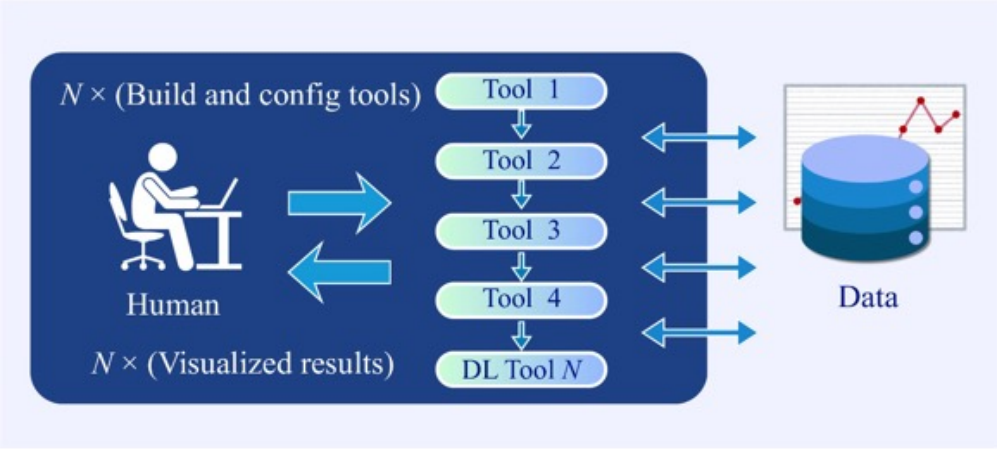
Message

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

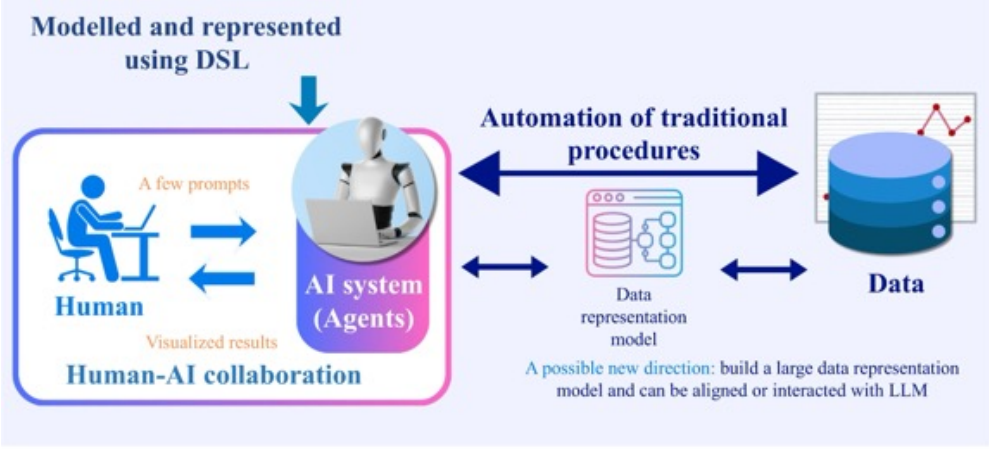


People configure, execute and supervise every stage of the data pipeline.

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**

Ke Li et al., *Front.Phys.* 21 (2026) 9, 096401

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

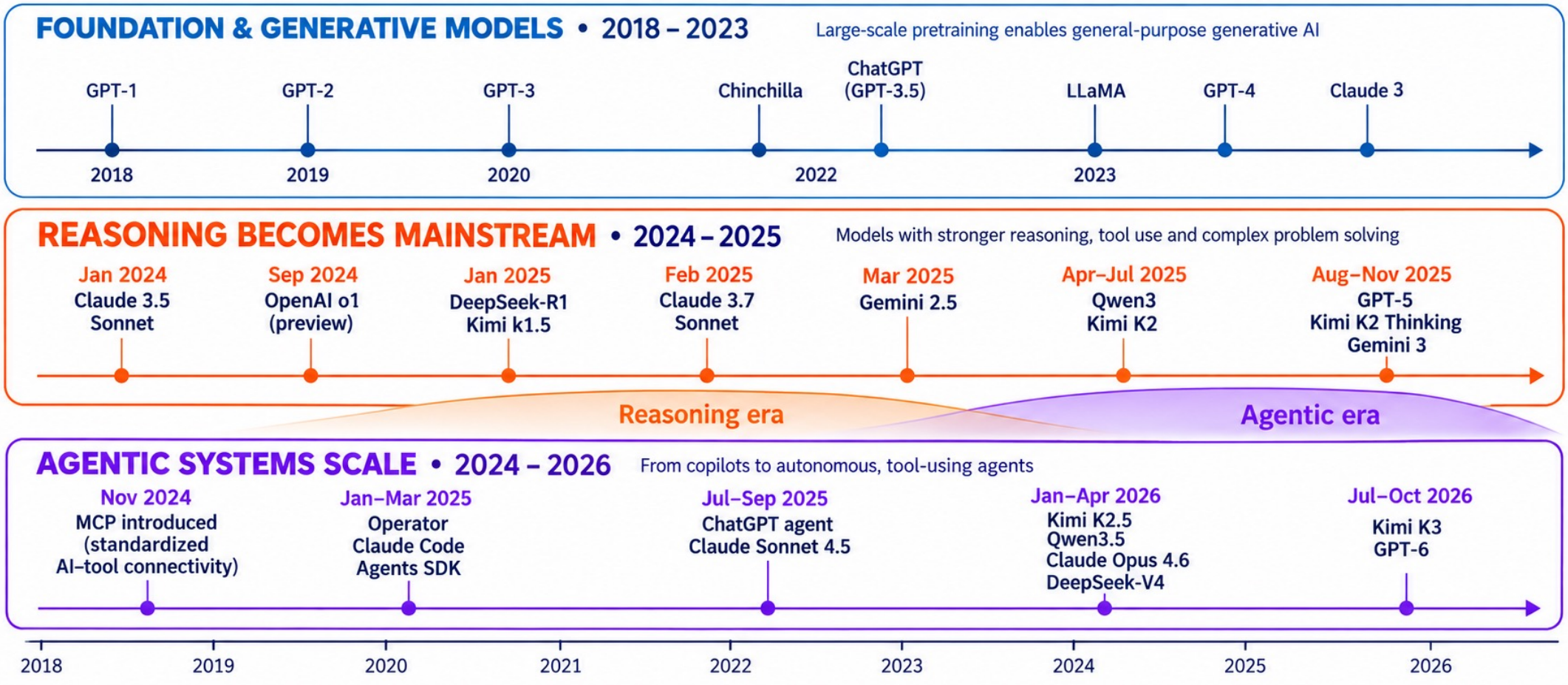
Reasoning: suggests a high-voltage diagnostic.

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



Towards an AI-Assisted Shifter

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

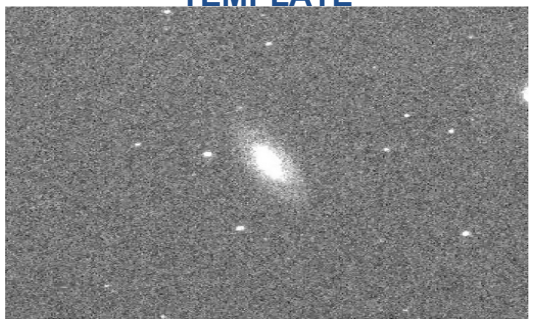
CMOS images in LRGB filters track nearby galaxies and flag new transient sources, including supernovae.

10 telescopes	**4** observing sites	**3,000+** galaxies on clear nights

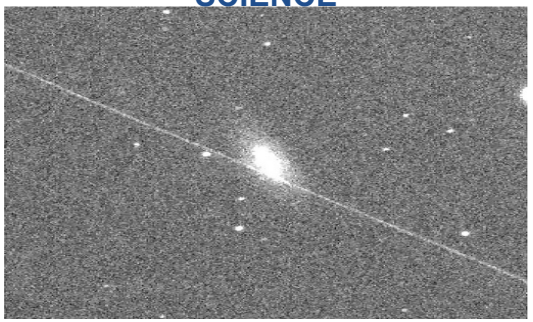
EXAMPLE TRANSIENT DETECTION

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

TEMPLATE



SCIENCE



DIFFERENCE

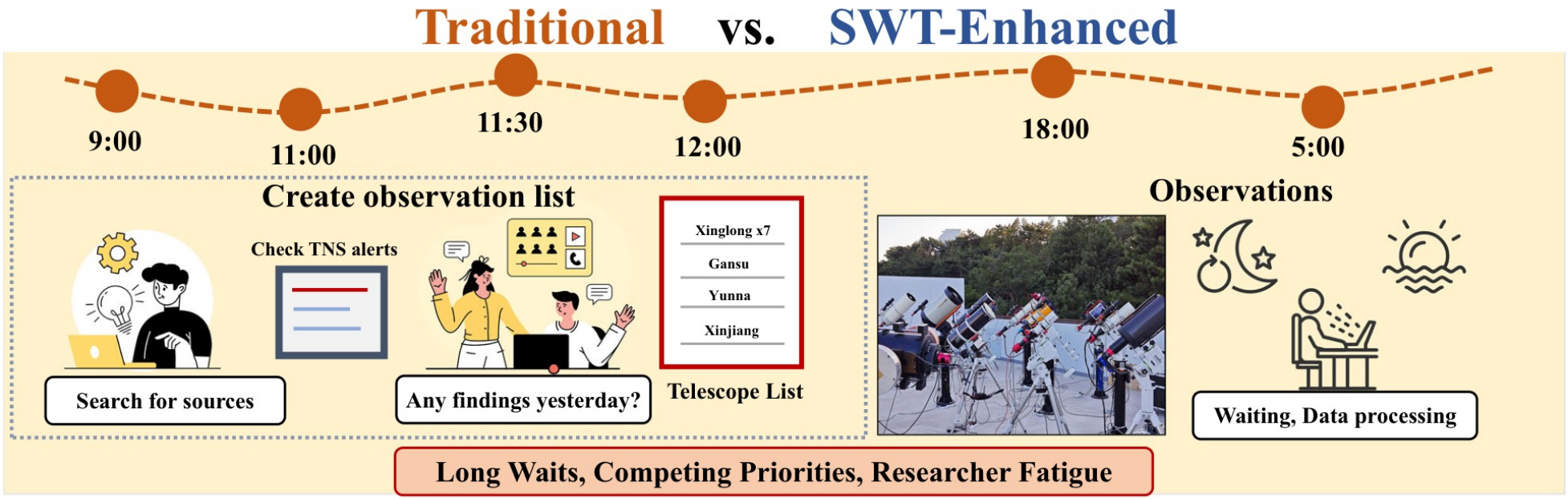


Template → new observation → subtraction reveals a transient candidate

Wang et al., Communications Engineering 4, 184 (2025) DOI: 10.1038/s44172-025-00520-4

StarWhisper: Why Agentic AI Matters

From manual observing to autonomous operations and science-ready data



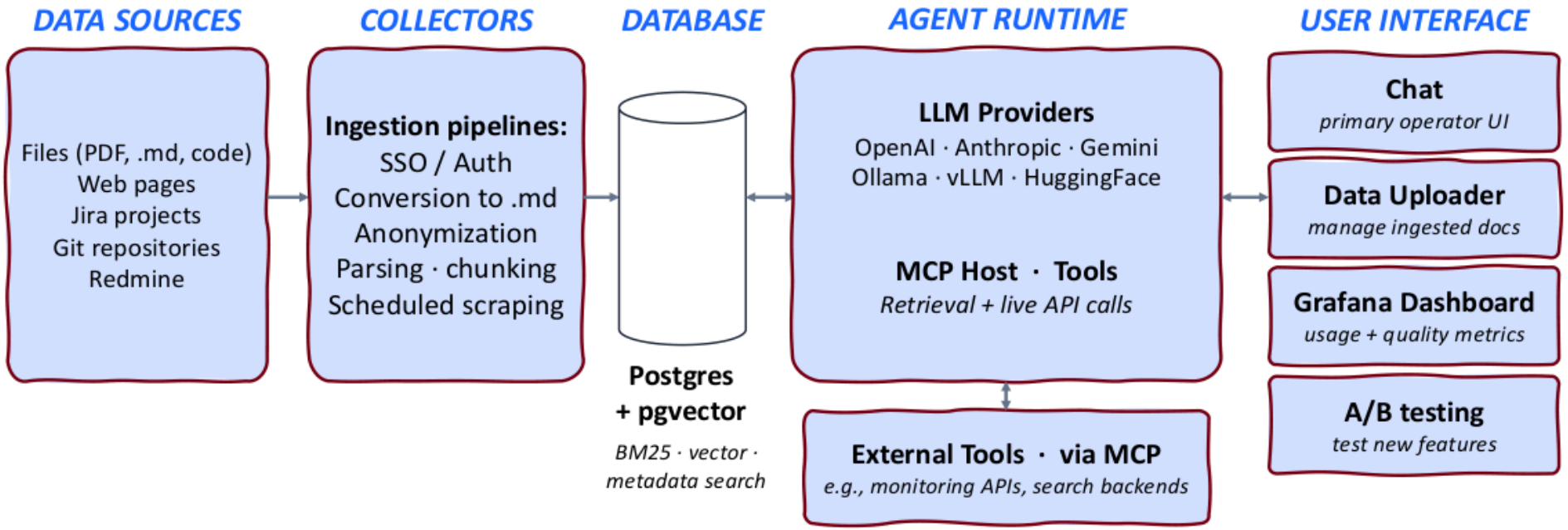
Agentic AI automates planning, execution and data handoff — researchers focus on science.

Fig. 1: Wang et al., Communications Engineering 4, 184 (2025) DOI: 10.1038/s44172-025-00520-4

CMS Archi: Agentic Computing Operations

CMS collision data → reconstruction, production and worldwide distribution

ARCHI • DOCUMENTATION + LIVE MONITORING + TOOL-ASSISTED REASONING



COMPUTING CHAIN

Tier-0, simulation and >1 EB across >60 storage sites

AGENT DIAGNOSTICS

~10,000 documents + read-only Rucio / HTCondor tools

IN PRODUCTION

393 conversations; 20 users (Feb–Apr 2026)

✓ **Agents investigate failures and retrieve runbooks; operators decide.**

arXiv:2606.04755 ↗

CMS Archi: Operational Diagnosis in Practice

A deployed agent combines live computing data with documented operational knowledge

EXAMPLE • FAILED DATA TRANSFER

1 OPERATOR ASKS

Provides the Rucio ID of a failed transfer

↓

2 AGENT QUERIES LIVE SYSTEMS

Checks failed files, sites and dominant errors

↓

3 AGENT RECALLS EXPERIENCE

Finds matching JIRA issues and relevant runbooks

↓

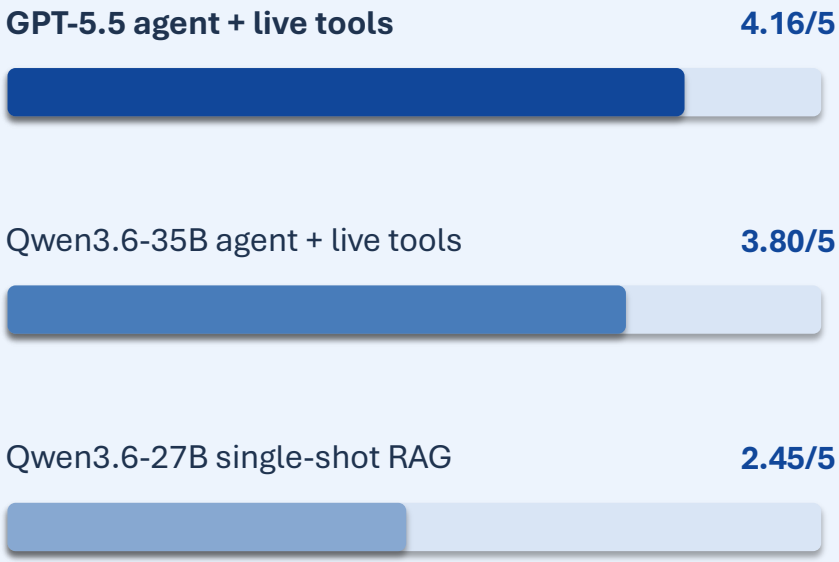
4 AGENT REPORTS FINDINGS

Returns likely causes, precedents and next steps

Live tools are read-only; diagnostic calls are traceable.

EVIDENCE • HUMAN EVALUATION

Answer quality (mean of correctness + usefulness)



63-question expert panel; RAG scored on 53 questions

Production trade-off: 107 s mean response time

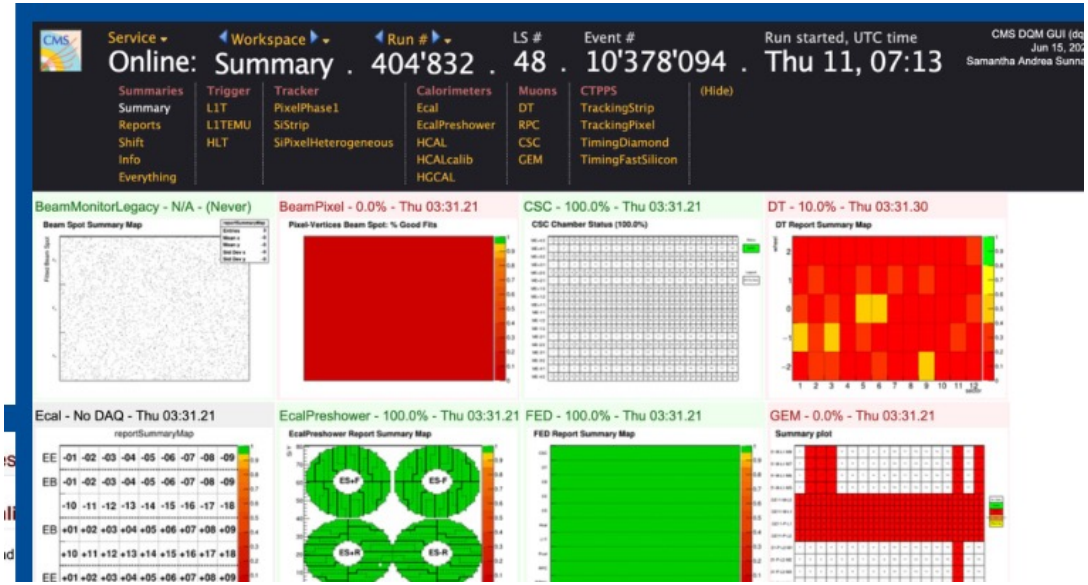
✓ **Agents diagnose with live data and runbooks; operators remain in control.**

arXiv:2606.04755 ↗

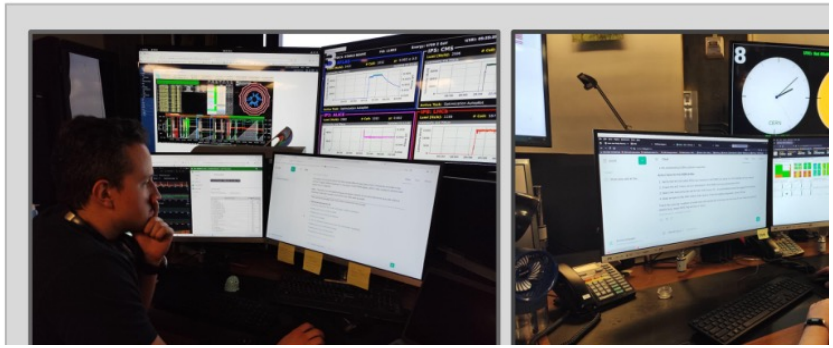
CMS WisDQM: Agentic Support for Shifters

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.**

Archi framework: arXiv:2606.04755 ↗

ATLAS TileCal: How Data Quality Monitoring Works

Online reconstruction → automatic DQMF tests → shifter intervention → offline certification

DATA-TO-DECISION CHAIN • PHYSICS AND CALIBRATION RUNS

1 DATA IN

Collision events and laser / CIS / Cs runs plus detector conditions

→

2 HISTOGRAMS

Occupancy, noise, pedestals, timing and digital readout errors

→

3 DQMF TESTS

Reference comparisons and numeric thresholds; flags propagate in a tree

→

4 SHIFTER ACTION

Inspect DQMD / OHP; check known issues; log or escalate anomalies

FROM TESTS TO DETECTOR STATUS

READOUT INTEGRITY	🔴 RED
ELECTRONIC NOISE	🟢 GREEN
LASER TIMING	🟡 YELLOW

Channel / DMU → module → partition status

SHIFTER TRIAGE AND OFFLINE REVIEW

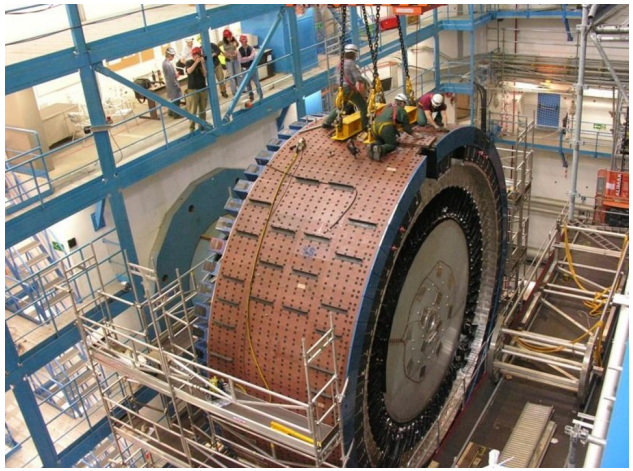
1	**INSPECT**	Drill down into failed histogram tests
2	**CROSS-CHECK**	DCS, whiteboard, MCWS bad-channel list
3	**DECIDE**	E-log / expert; offline defects and GRL

✓ **DQMF raises flags; shifters diagnose; offline DQ certifies the data for physics.**

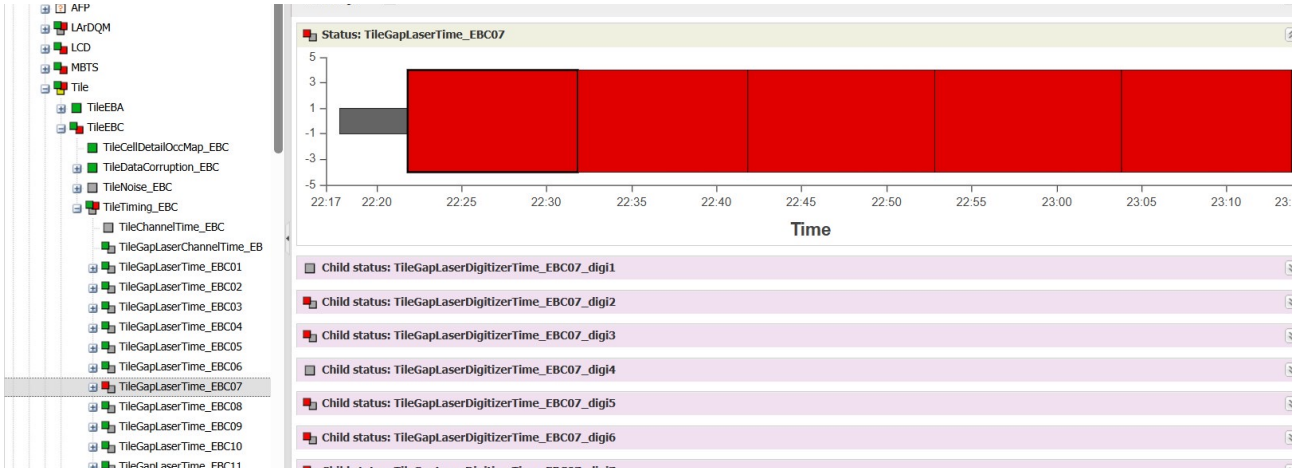
Calo DQ on Autopilot, Shifter in Command (work in progress)

Agentic AI for ATLAS TileCal Data Quality Monitoring • DQ shifter assistant

ATLAS TILE CALORIMETER



DQMF • AUTOMATED DATA QUALITY MONITORING



256 modules • ~10,000 channels

~50,000 histograms per run

6-step judgement checklist

PROPOSED AGENTIC WORKFLOW • DQMF ALERT → REASONING → DRAFT E-LOG

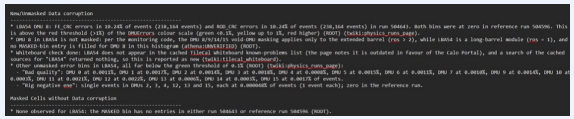
1 DQMF FLAGS

Green / yellow / red statuses highlight anomalies for DQ review.

2 AGENT CHECKS

Whiteboard, persistence and prior shifts inform its assessment.

3 DRAFT E-LOG



For shifter review

✓ Agent investigates and drafts; the DQ shifter validates and decides.

Agentic Infrastructure at CERN

CERN: Shared Agentic AI Infrastructure

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

ESTABLISHED BUILDING BLOCKS • ITGPT / $(AI)^2$

01 ACCESS

A common entry point for CERN users

- OpenWebUI + SSO
- Mattermost / chat
- API integration

02 MODELS

Unified model gateway and serving

- LiteLLM proxy
- vLLM / A100 GPUs
- Cloud models

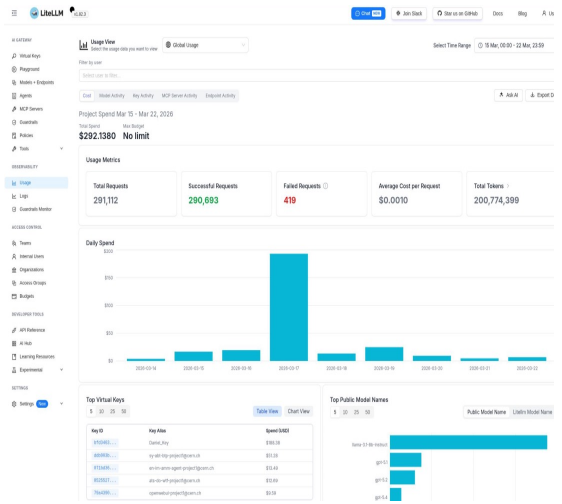
03 AGENTS

Shared retrieval and tool access

- RAG pipelines
- MCP integration
- Scientific services

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

Q2 2026

01 FIRST PoC

Gateway, model catalog, monitoring and accounting

Initial use-case onboarding

Q4 2026

02 AGENTIC AI

Agent hosting, MCP integration and shared RAG pipelines

Selected agentic use cases

Q2 2027

03 GENERAL AVAILABILITY

Sustainable operations, resource portal and costing

Documentation and training

ONE PLATFORM • DISTINCT OWNERSHIP

CENTRAL CERN IT SERVICE

Model catalog + serving • agent hosting
Identity, RBAC, MCP + RAG • evaluation

EXPERIMENT / DOMAIN TEAMS

Scientific prompts, workflows and integrations
Curated knowledge • validation + accountability

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

Summary & Conclusions

From large-scale complexity to a reliable, human-supervised AI-assisted shifter

WHAT THE PRESENTED SYSTEMS DEMONSTRATE

STARWHISPER

Autonomous observing

Plans and executes telescope observations; delivers science-ready data.

FIELD DEPLOYMENT

CMS ARCHI

Operational diagnosis

Combines ~10k documents with live read-only tools to assist computing operators.

IN PRODUCTION

CMS WisDQM

Shifter assistance

FNAL pilot retrieves procedures and helps operators prioritize DQM checks.

LIVE PILOT

ATLAS TILECAL

DQM-to-e-log assistant

Proposed: DQMF flags trigger checks, evidence gathering and draft e-logs.

WORK IN PROGRESS

THE AI-ASSISTED SHIFTER: A TRACEABLE DATA-TO-DECISION PATH

1 MONITOR

Histograms, alarms and DQMF status

→

2 GROUND

Runbooks, history and live APIs

→

3 INVESTIGATE

Cross-check causes; draft the e-log

→

4 HUMAN DECIDES

Validate, escalate and certify data

Shared CERN AI services provide models, RAG/MCP and access control; experiments own scientific validation.

✓ **Agents investigate and recommend; shifters remain in command.**

Examples: StarWhisper • CMS Archi / WisDQM • ATLAS TileCal • CERN IT (references on preceding slides)