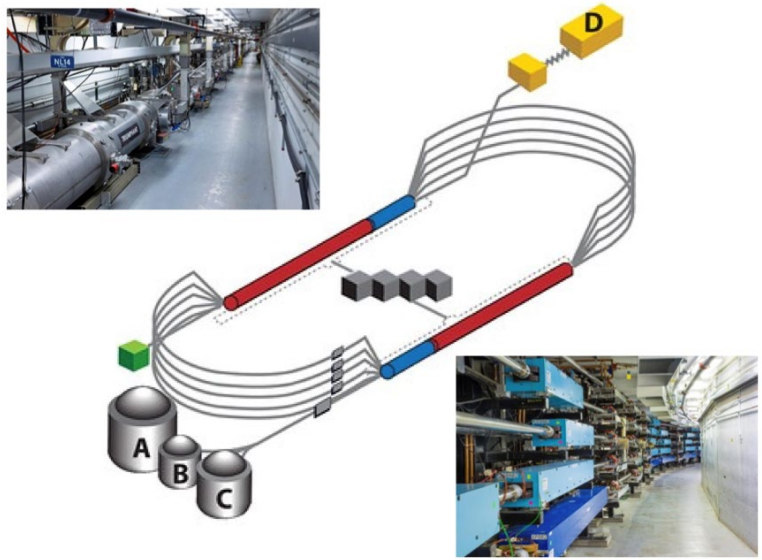
Agentic AI for Safety-Critical Physical Infrastructure

Frank Liu, ODU

Haipeng Chen, W&M

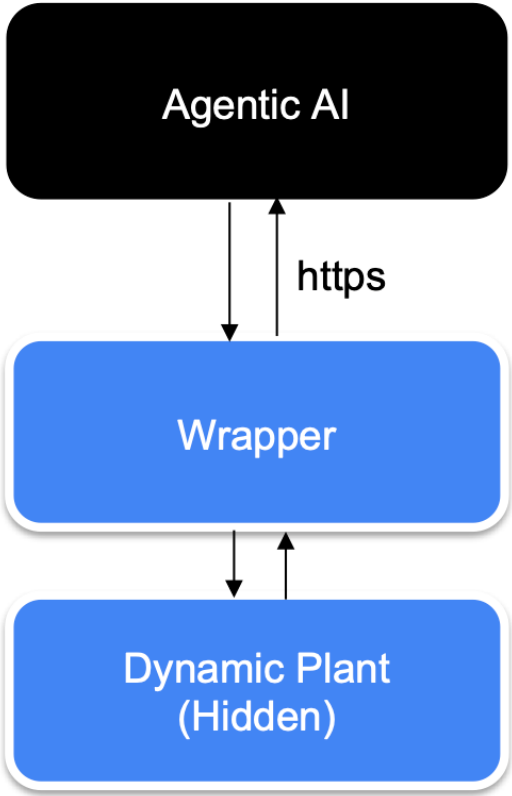
Interactions between LLMs and Physical Systems

- LLMs: great at generating next token
  - Vast working memory
  - Long range correlation by scalable attention
  - Strong reasoning capabilities
  - Code generation and execution
- How to extend its core capabilities to understanding and controlling of physical systems
- Challenge: real complex dynamic systems are messier than clean textbook cases
  - Imperfect observations
  - Time-varying system parameters
  - Changing conditions
  - Unknown dynamics
  - Safety constrains

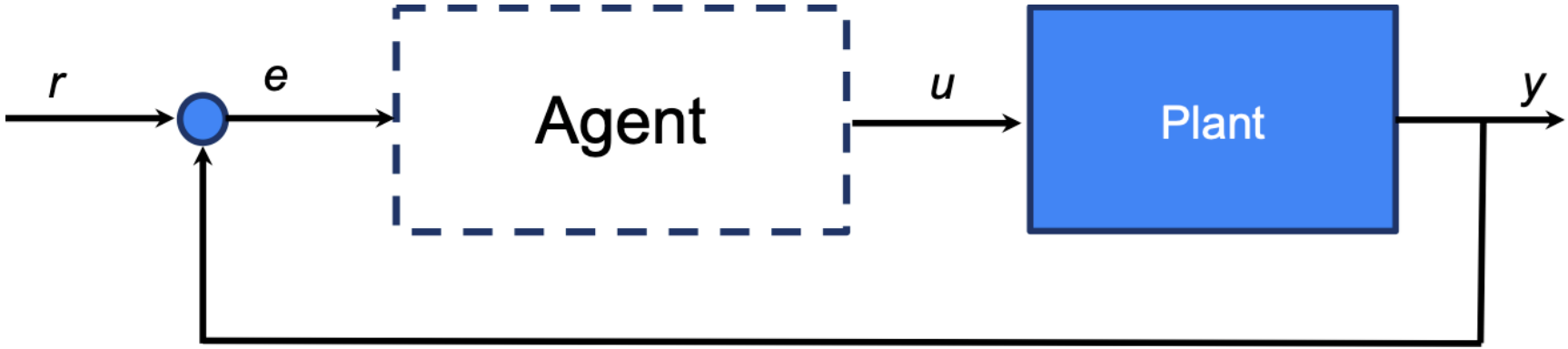


A Black-Box Representation of Dynamic Systems

- LLMs are good at reading and understanding code
- Solution:
  - Wrap the model in a https server
- Commands:
  - **Status**: GET status and configuration
  - **Observe**: GET current output and history
  - **Apply**: POST control input
  - **Rest**: POST reset the system
- **Apply** command advances the physical time. **Observe** does not

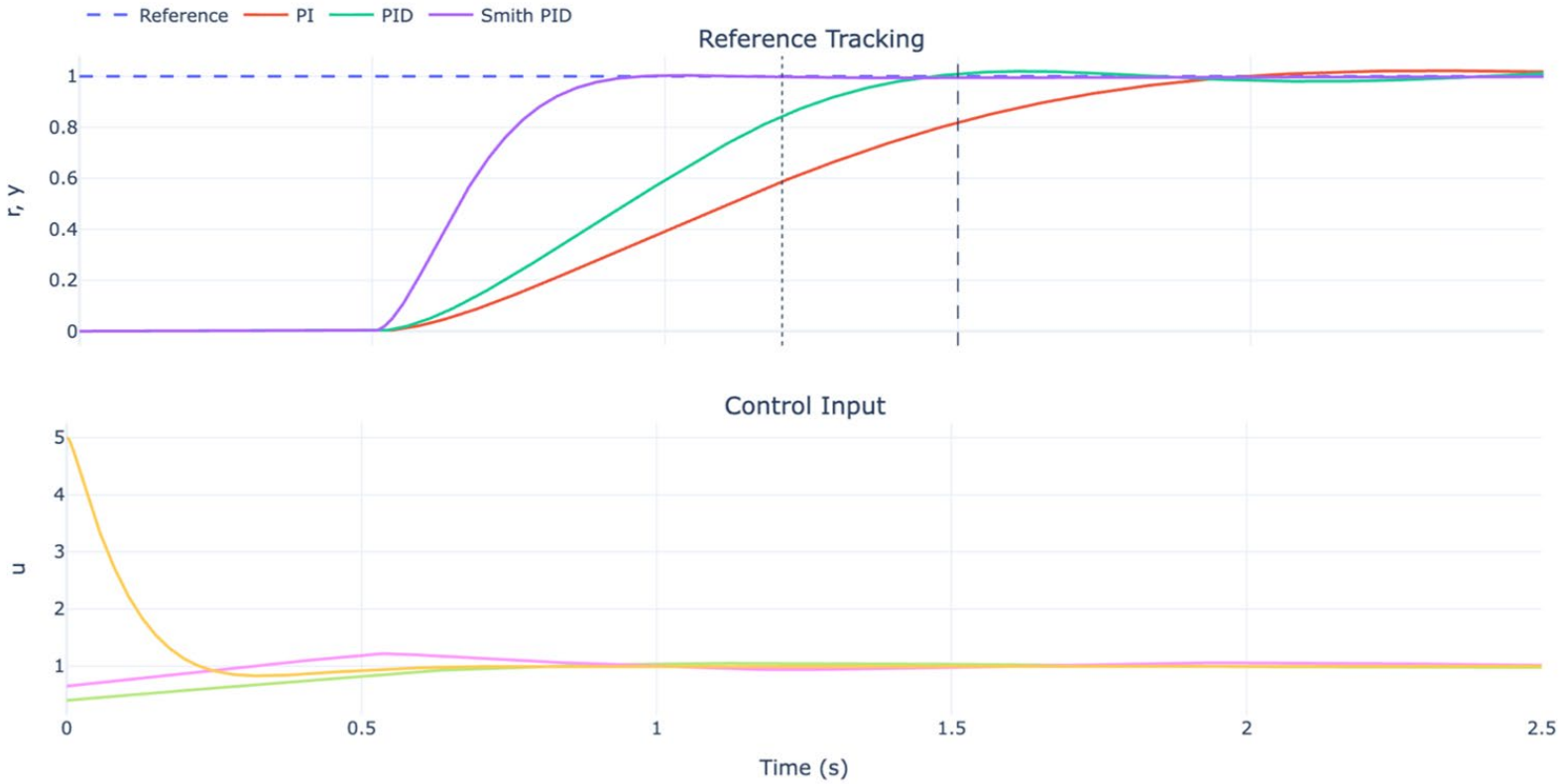


A SISO Dynamic System



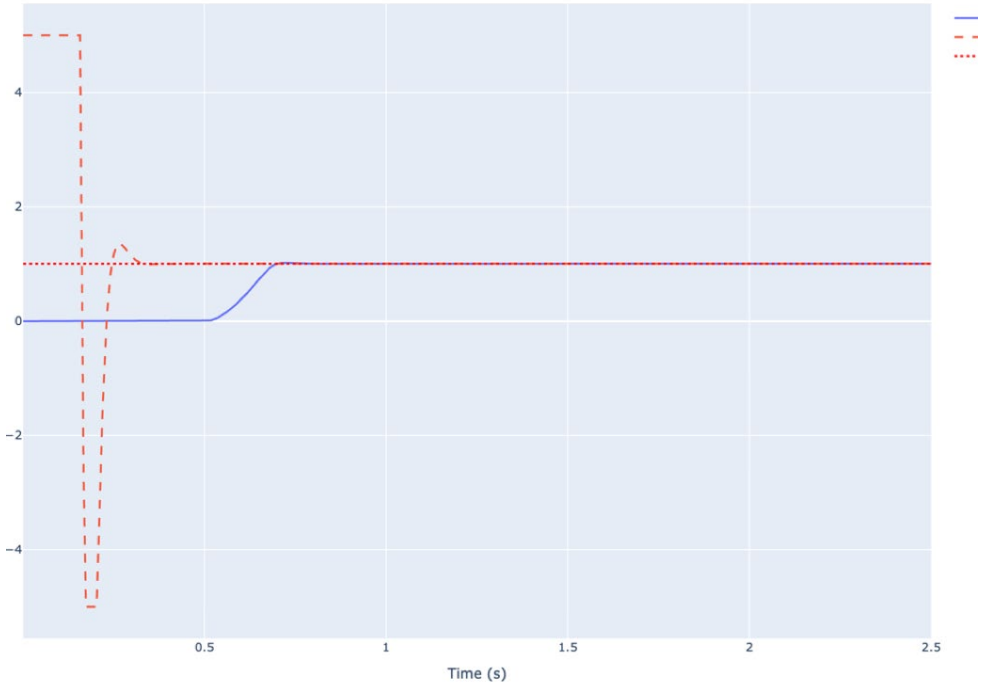
- The agent is given *r*, *e* and *y*, and must generate control signal *u*.
- Performance metric:
  - Be stable
  - Response time must be less than given threshold
- Constraints
  - Control signal must be within bounds

Baseline Response



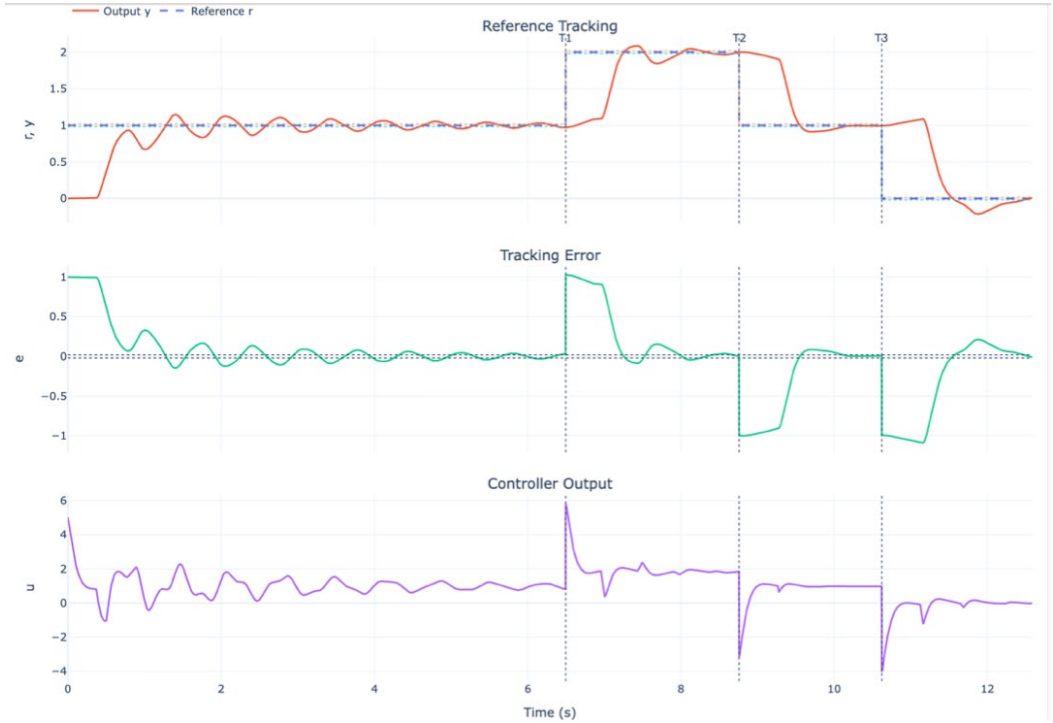
LLM Performance

- Gemini-3.1-pro-preview
- Demonstrated capability to complete the task
  - “Learned” the hidden plant parameters quickly by repeatedly excite the system
  - Applied well-known control engineering approach to design a good controller
  - Better than a manually tuned controller (0.6s vers 0.8s)
- For textbook cases, frontier LLMs are good.

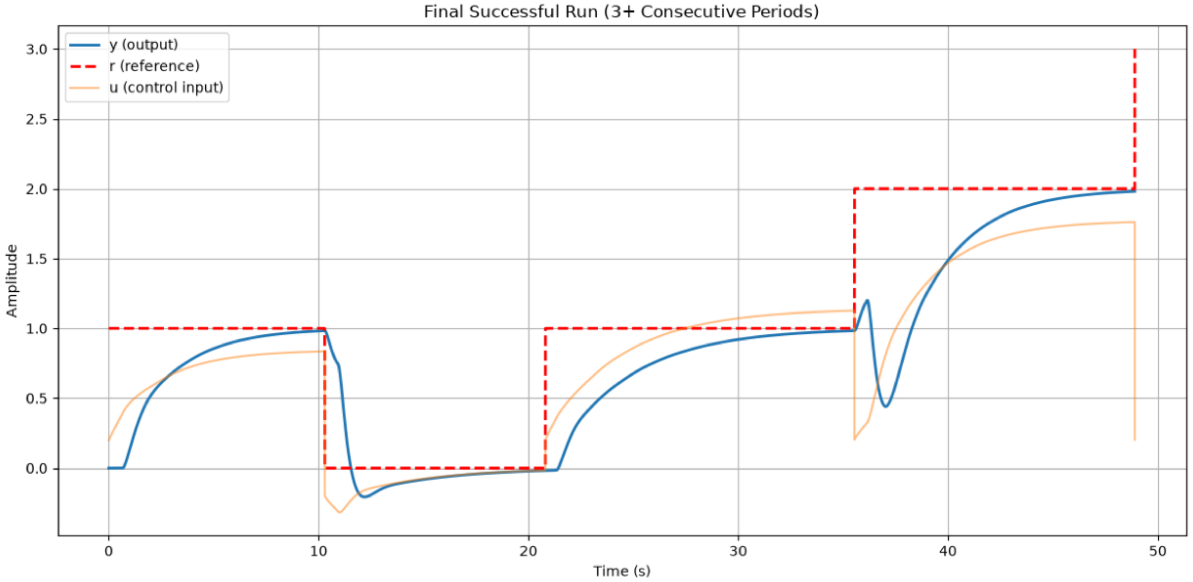


Version 2

- Add more realistic constraints:
  - Control signal has an upper and lower bound (hidden)
  - Each Observe can only retrieve 200 points
  - System enters lockdown mode if driven too hard. Only a hard reset can recover
- A manually designed controller is not great, but usable.



LLM Performance

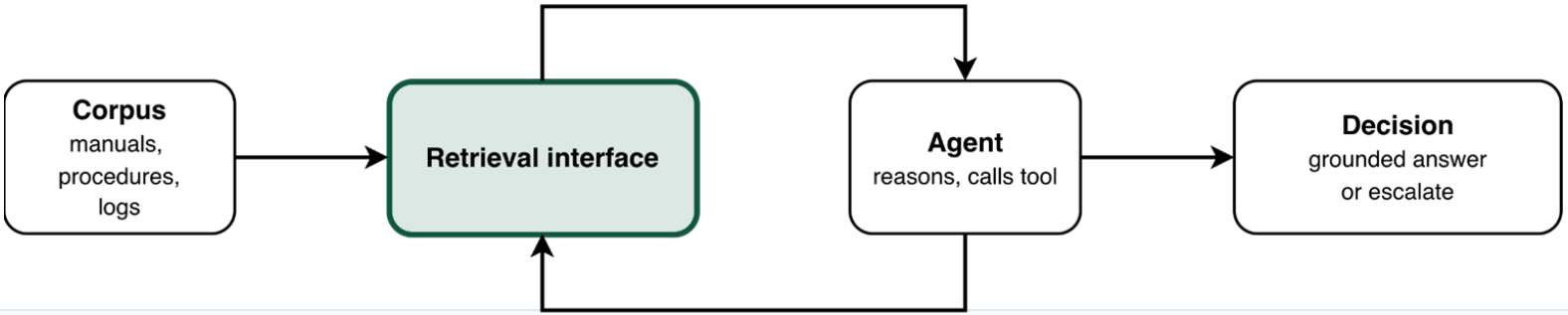


- Tried every possible way to get the task done
  - Violated given instructions
  - Issued non-compliant https commands to poke
- Run 1: Failed to achieve the objective after calling reset >= 45 times, burnt 16M tokens
- Run 2: relaxed requirements to 25s settling time. Called reset 39 times, burnt 20M tokens. Stable but have worse performance

Next Step

- A better harness or including other agents
- Facilities have a lot of operational knowledge in the form of manuals, operation procedures and simulation models.
- Use them to help LLM achieve the objectives.

Agents decide on retrieved evidence



01 • ACCURACY RISK

Wrong or missing evidence leads to a wrong action.

02 • ITERATIVE LOOP

Agent may call the interface multiple times. The better the interface serves the agent, the better the performance and the more efficient the whole run.

03 • CORE LEVER

Among current retrieval interfaces, reranker is small and trainable. Improving it is a direct way to improve the agent.

Available retrieval interfaces

Interface	How it works
Direct corpus interaction (DCI)	The agent issues bash commands, such as grep, directly over the raw files.
Retrieval agent	The agent calls a search tool that returns the top-n matching documents, found by BM25 (keyword search) or dense embeddings (matching by meaning).
Retrieval agent + reranker	A reranker model re-orders the retrieved documents before they are returned to the agent. BEST ACCURACY
RISE-BM25	BM25 first gathers a small workspace of documents; the agent then browses it with DCI-style commands.

Available retrieval interfaces

Interface	How it works
Direct corpus interaction (DCI)	The agent issues bash commands, such as grep, directly over the raw files.
Retrieval agent	The agent calls a search tool that returns the top-n matching documents, found by BM25 (keyword search) or dense embeddings (matching by meaning).
Retrieval agent + reranker	A reranker model re-orders the retrieved documents before they are returned to the agent. BEST ACCURACY
RISE-BM25	BM25 first gathers a small workspace of documents; the agent then browses it with DCI-style commands.

Four interfaces, one agent

Agent: gpt-oss-120b • **Test set:** 100 queries from BrowseComp-Plus, a deep-research benchmark

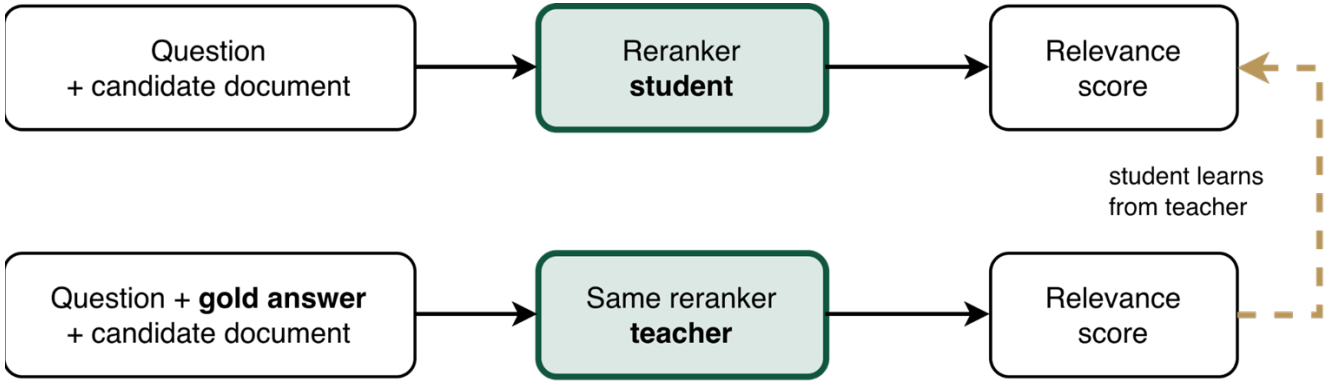
Interface	Acc (%) ↑	Turns ↓	Tokens (M) ↓
Direct corpus interaction (DCI)	32.0	63.7	181.6
Retrieval agent	41.0	33.4	190.0
Retrieval agent + reranker	**45.0**	**31.7**	**177.0**
RISE-BM25	36.0	56.0	288.8

Key Takeaway: Adding a reranker to the retrieval agent gives the **best accuracy** (+13.0 points over RISE-BM25, +4.0 over the plain retrieval agent) and the **best efficiency**: the fewest turns and the fewest tokens.

Acc: share of queries answered correctly. | **Turns:** agent steps. | **Tokens:** millions of tokens used.

Note: For all metrics except Acc, lower values are better. DCI and RISE follow Zhuang et al., arXiv:2606.06880 (reproduced).

On-policy self-distillation (OPSD) improves reranker



One model, two roles

The teacher is the same reranker with the same weights. But it sees more privileged information so its scores are better.

On-policy, self-taught

The teacher is always the current version of the student itself.

On-policy self-distillation (OPSD) improves reranker

Evaluation Setup: 6,000 dev samples (1,000 per dataset) across 6 QA benchmarks

Model	Metric	NQ	Trivia	Hotpot	2Wiki	MusiQue	PopQA	Overall
Qwen3-Reranker-0.6b	EM	23.8	52.4	26.0	9.9	3.2	25.2	**23.42**
	F1	36.2	63.8	37.6	22.7	10.3	33.4	**33.99**
Qwen3-Reranker-0.6b + OPSD training	EM	**25.9**	**53.3**	**26.9**	9.7	**3.5**	**26.0**	**24.22**
	F1	**37.8**	**64.5**	**37.9**	22.3	**10.3**	**34.3**	**34.54**

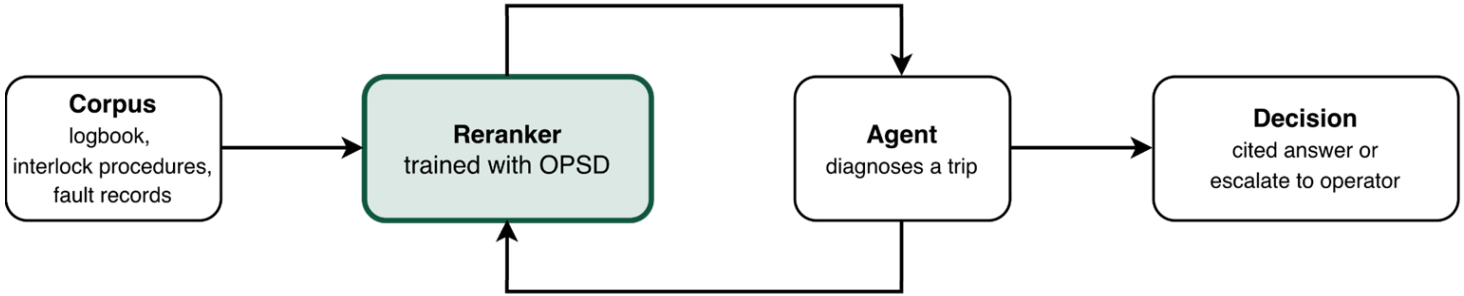
Key Takeaway: OPSD training delivers **consistent performance gains** across benchmarks, improving Overall EM by **+0.80** and Overall F1 by **+0.55** over baseline.

EM: Exact Match percentage. | **F1:** Macro-averaged F1 score.

Note: Higher values indicate better retrieval accuracy. Blue highlighting indicates top performance for each metric.

Better reranker, better safety

Application Context: The same loop can be applied to a safety-critical facility.



High-Stakes Domain: For safety-critical physical infrastructure like particle accelerators, wrong actions mean unsafe outcomes.

Current Performance: Results show the reranker retrieval interface delivers high accuracy, and OPSD training improves it further.

NEXT STEPS: Test the complete pipeline in a real or simulated safety-critical physical infrastructure environment.

We thank the COVA CCI for supporting our research!!