LC-Opt: Benchmarking Reinforcement Learning and Agentic AI for End-to-End Liquid Cooling Optimization in Data Centers

Avisek Naug[1*], Antonio Guillen-Perez[1*], Vineet Kumar[2], Scott Greenwood[2], Wesley Brewer[2], Sahand Ghorbanpour[1], Ashwin Ramesh Babu[1], Vineet Gundecha[1], Ricardo Luna Gutierrez[1], Soumyendu Sarkar[1*]

1: HPE Labs @ Hewlett Packard Enterprise

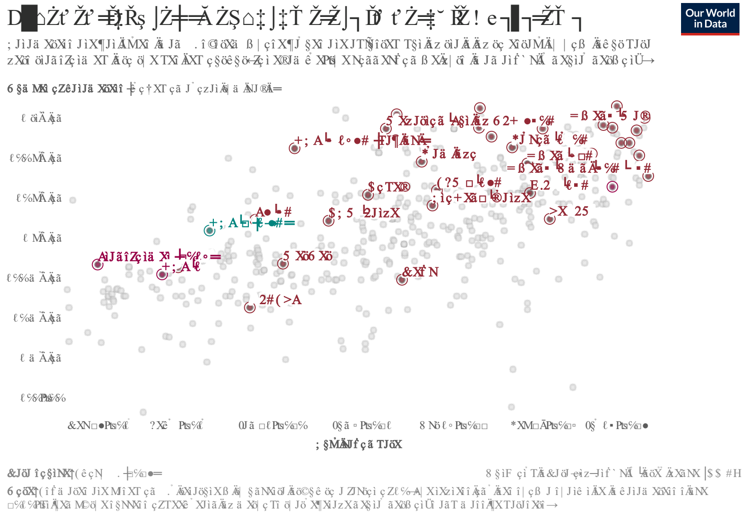
2: Oak Ridge National Laboratory

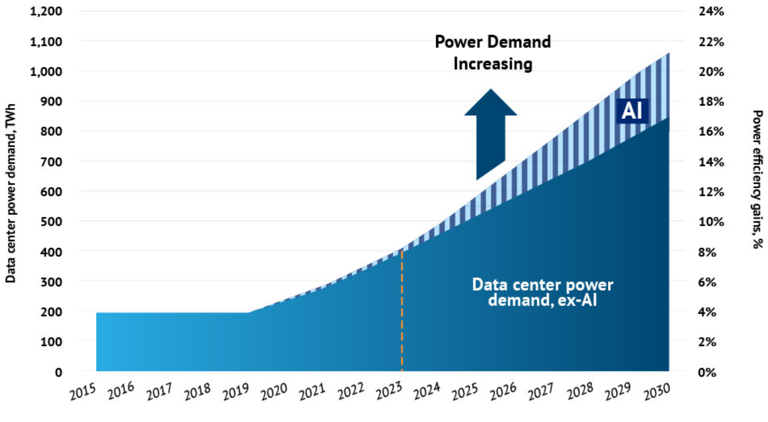
05/12/2025

The AI Scaling Dilemma: More Power, More Heat

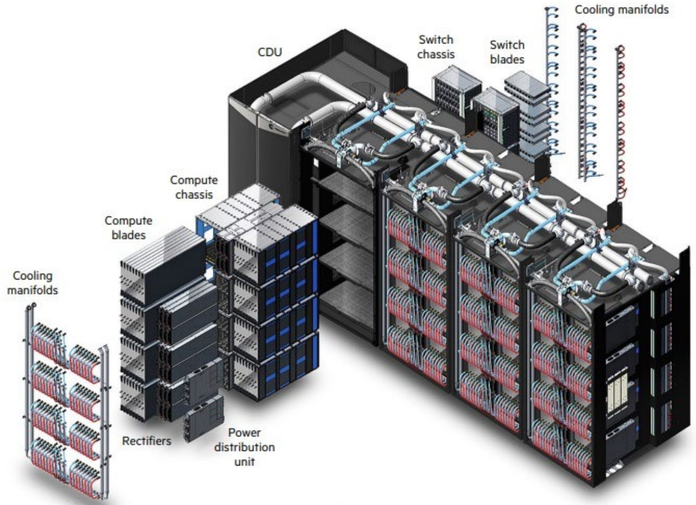
NEURAL INFORMATION PROCESSING SYSTEMS 2025

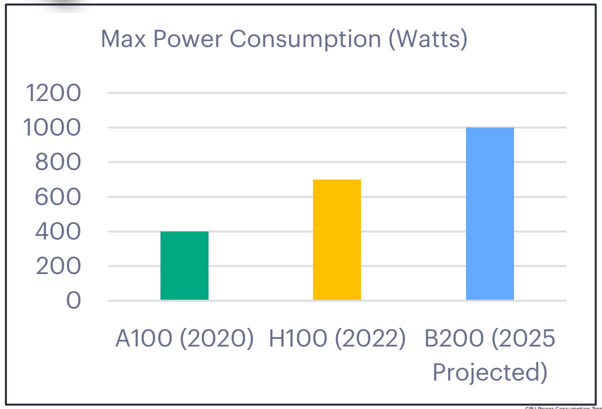
- The demand for large-scale AI is driving exponential growth in data center energy consumption.
- Modern accelerators (GPUs/CPUs) have power densities that traditional air cooling can no longer handle, making liquid cooling (LC) essential.
- LC can cut cooling energy by over 60%, but this potential is currently wasted.
- **The Problem:** Most LC systems today run on static, rule-based controls, not dynamic, AI-driven optimization.

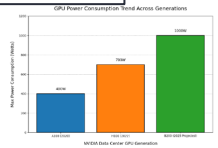




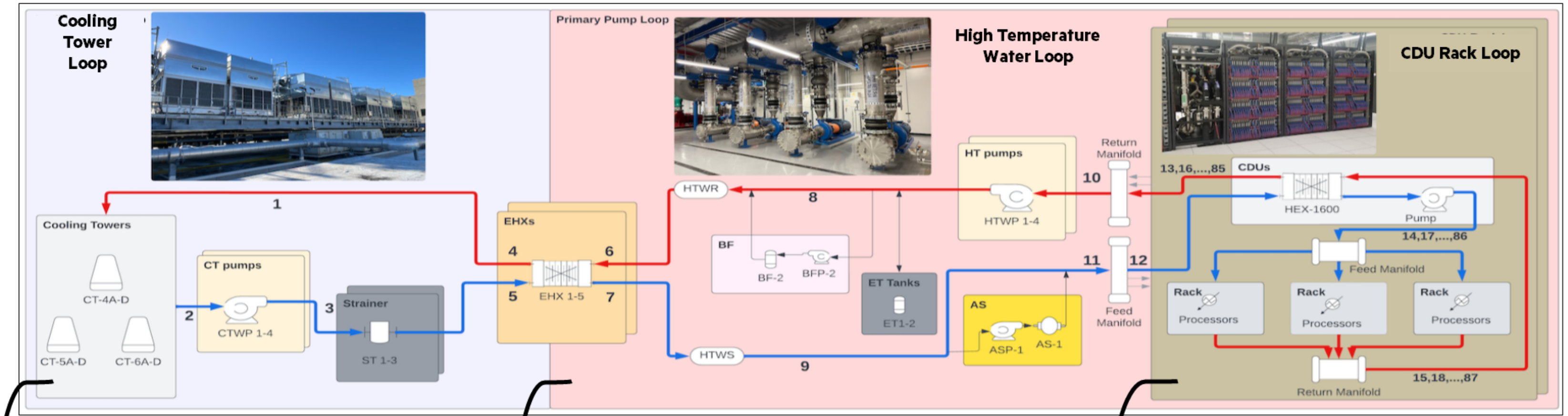
Source: Masanet et al. (2020), Cisco, IEA, Goldman Sachs Global Investment Research







Challenge: End-to-end scalable Data Center Liquid Cooling Optimization



Frontier Cooling System with three loops: CDU, High Temperature Water, and Cooling Tower [1]

A facility loop that carries heat from the intermediate loop to cooling towers for releasing into the air.

Acts as a *buffer* between the CDU loop and the Cooling tower loop while allowing flexible routing to either heat recovery or cooling towers.

First loop of heat removal from the server racks to protect the hardware.

EXADIGIT
AN OPEN FRAMEWORK FOR DEVELOPING DIGITAL TWINS OF SUPERCOMPUTERS

Global Goal: Minimize total site-level energy consumption.

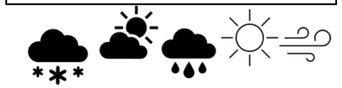
Local Goal: Guarantee thermal safety for every single server (no hotspots).

The Research Gap: The ML community has been unable to tackle this due to a lack of high-fidelity, and open testbeds.

Dynamic CPU and GPU Compute Workload

Volatile Carbon Intensity Trends

Variable Weather Conditions



A data center is a complex, multi-physics system with tightly coupled components: cooling towers, pumps, heat exchangers, and thousands of server blades.

This creates a high-dimensional, multi-objective optimization challenge:

NEURAL INFORMATION PROCESSING SYSTEMS 2025

Our Contribution: LC-Opt

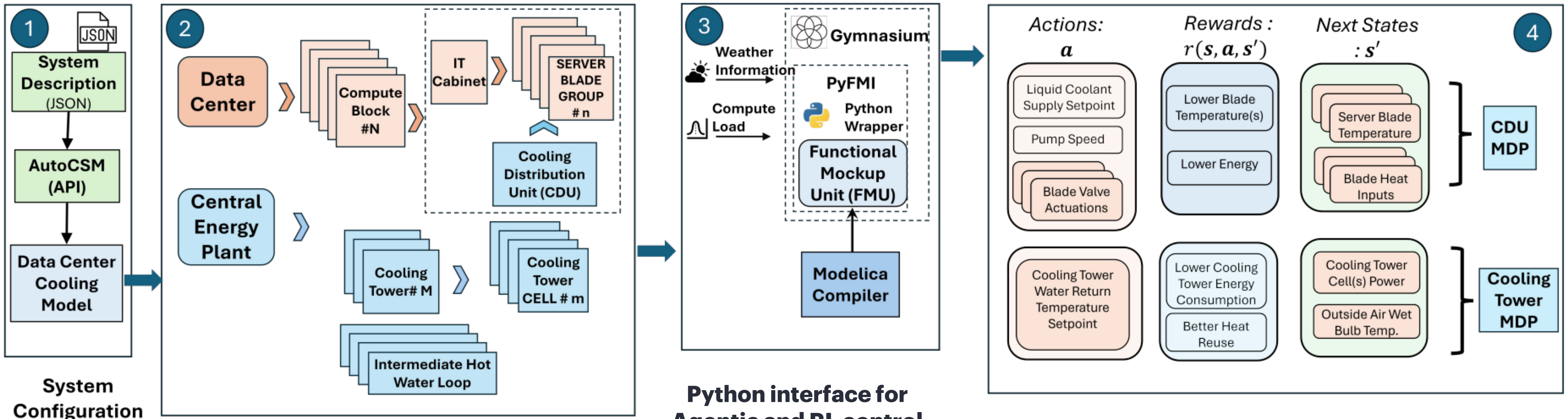
We present **LC-Opt**: The first open-source, high-fidelity benchmark for end-to-end liquid cooling control.

A True Digital Twin: Built on a validated, high-fidelity Modelica model of the **ORNL Frontier Supercomputer's** cooling system.

Landmark Collaboration: A contribution of the **ExaDigiT consortium** (HPE, ORNL, and international supercomputing centers).

ML-Ready: Features a full **Gymnasium interface** for standardized RL and AI agent development.

High-Fidelity, End-to-End Modeling



System Configuration Interface

Modelica Model of the Data Center Liquid Cooling

Python interface for Agentic and RL control

RL Control of the Data Center Liquid Cooling System

System Scope: Spans from site-level cooling towers down to granular server blade-groups.

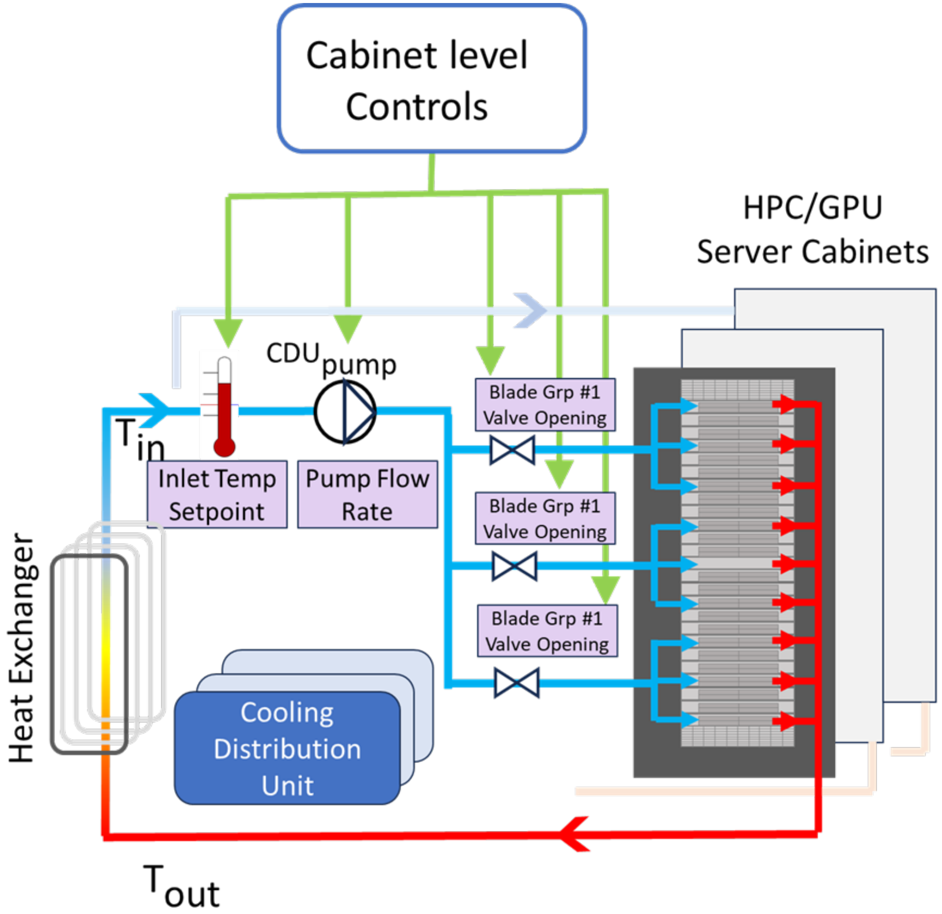
Physics-Based: Captures transient thermo-fluid dynamics using the Modelica modeling language.

AI-Enabled Interfaces: We augmented the model with control interfaces for:

- Cooling Tower (CT) setpoints
- CDU liquid supply temperature & flow rate
- Granular blade-group valve actuation

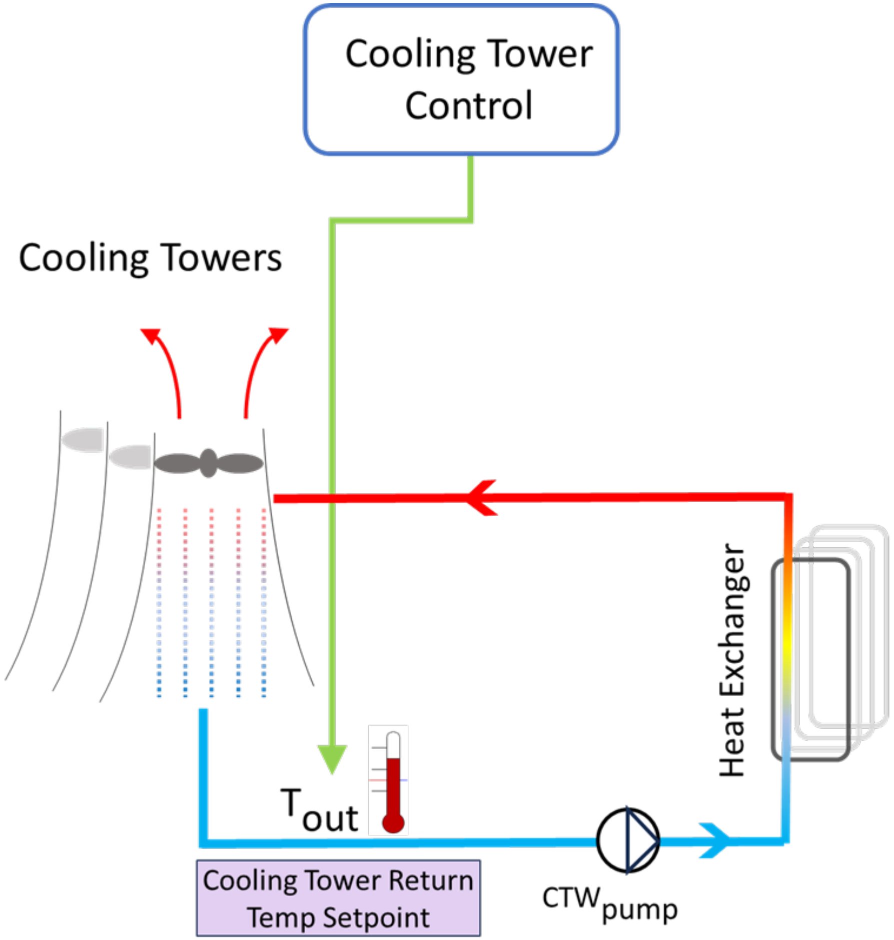
Modular API: An AutoCSM API allows users to build custom data center models from simple JSON descriptions.

IT Cabinets: Scalable Data Center Liquid Cooling Optimization



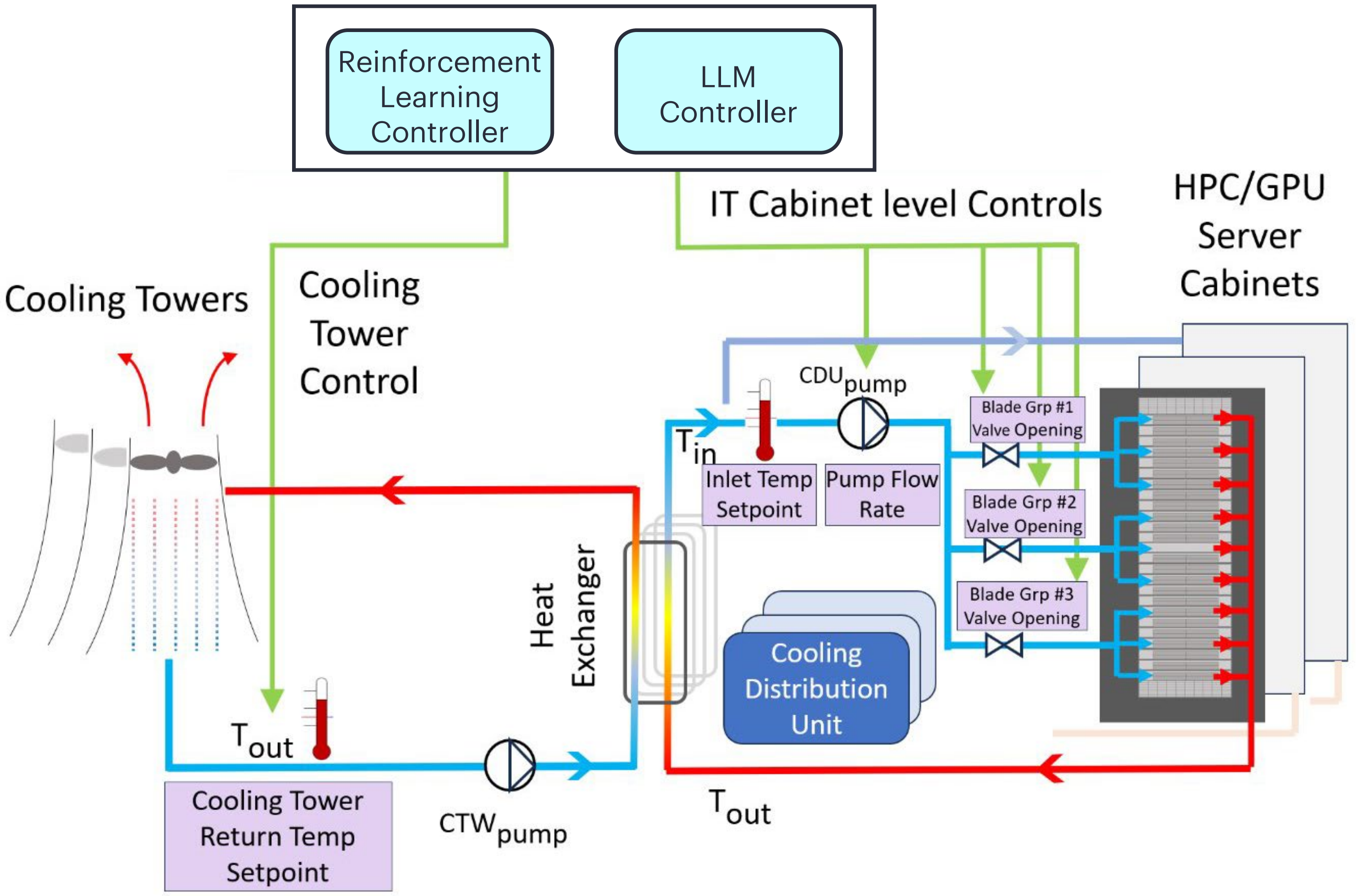
- Control Action:
  - Temperature setpoint of the coolant fluid that is goes into the Cabinet
  - Differential Pressure setpoint of the pump that determines flow rate of the coolant fluid that goes into the cabinet
  - Valve Actuations for Branches 1
  - Valve Actuations for Branches 2
  - Valve Actuations for Branches 3
- Objective:
  - Temperature Stability
  - Minimize Cooling Energy

Cooling Tower: Data Center Liquid Cooling Optimization

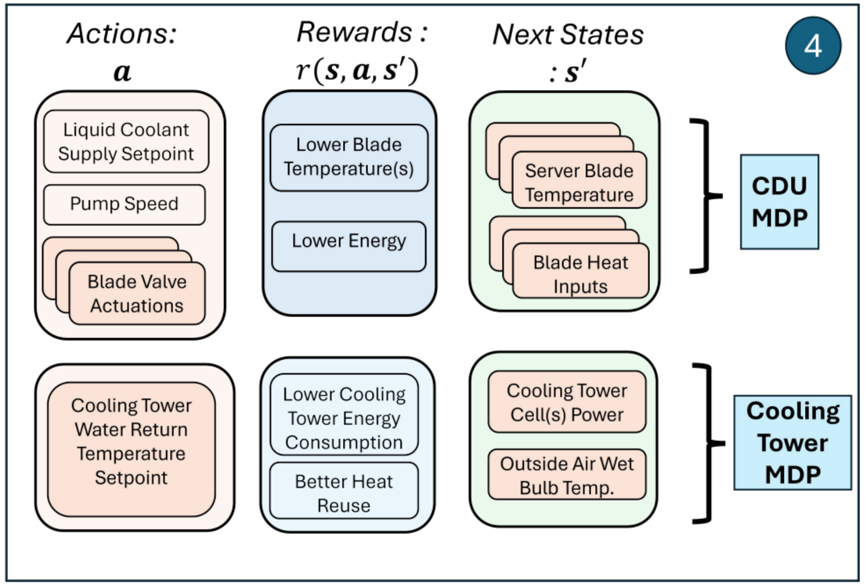


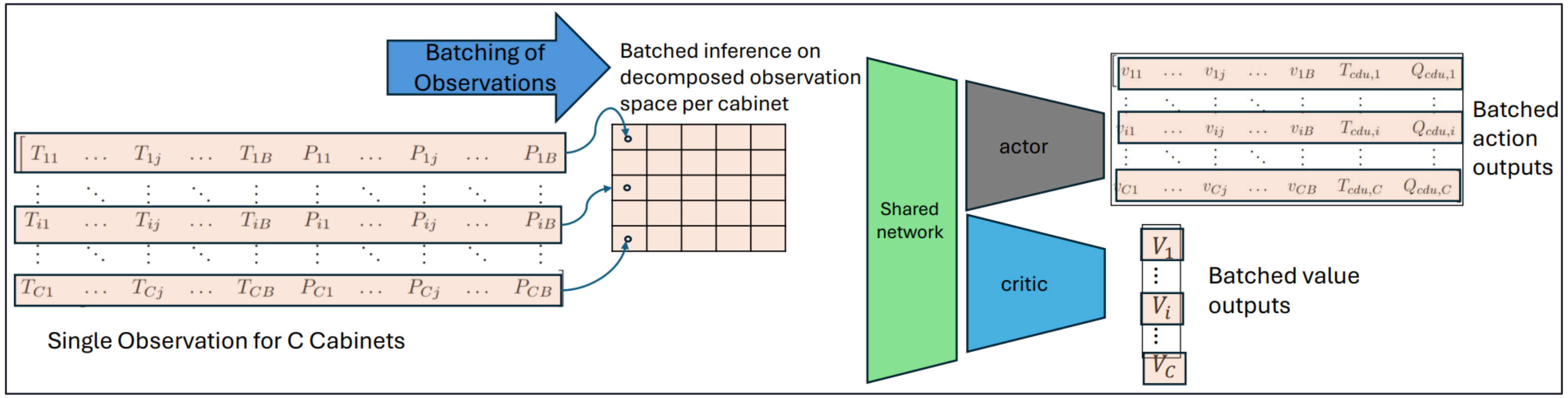
- Action:
  - Change in the cooling tower setpoint from previous time point δ_i for $i = 1 \ldots t$ where t is the number of cells. Here t = 2
- Objective:
  - -(Aggregate/Sum Power of the Cooling Tower Cells): $-\sum_{i=1\ldots 2} P_i$

End-to-end scalable Data Center Liquid Cooling Optimization



Approach 1: Multi-Agent RL for Optimal Control





The Goal: Train agents to find an optimal policy that minimizes energy while ensuring thermal compliance.

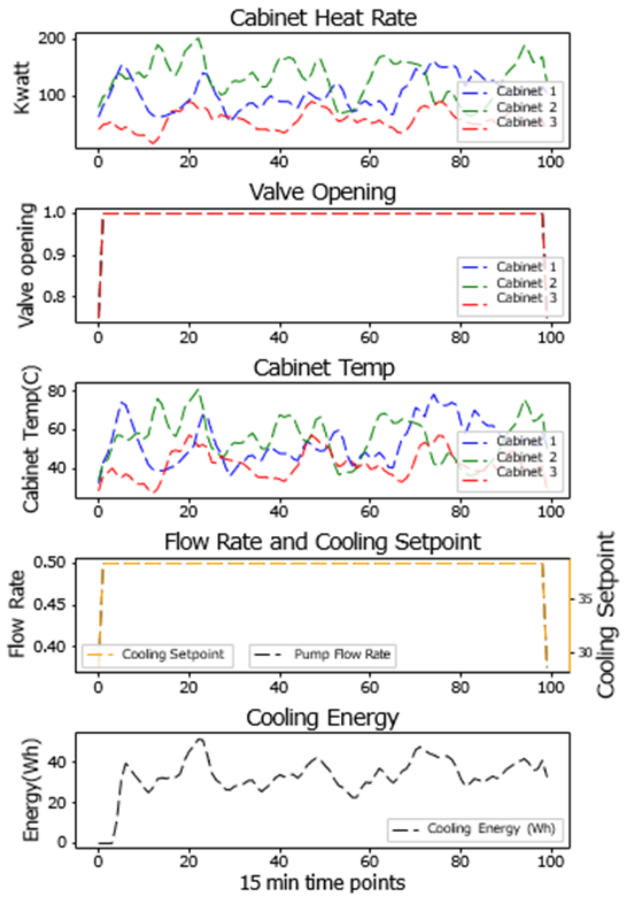
Multi-Agent Hierarchy: We define two sets of independent agents:

- **Cooling Tower (CT) Agents:** Manage global energy. Goal: Minimize total power consumption.
- **Blade Group (BG) Agents:** Manage local safety. Goal: Keep all blade temperatures in the ideal range.

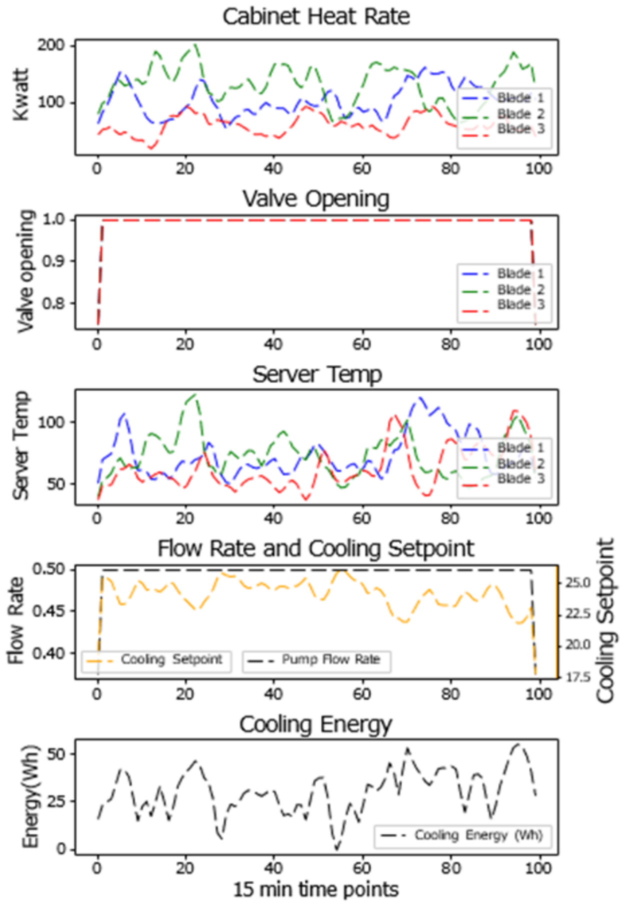
Innovation for Scale:

- **Centralized Action Execution (CA):** We use a shared policy with batched inference to scale to thousands of control points.
- **Multi-Head Policy:** A novel actor network that disentangles control actions (e.g., flow rate vs. valve position) for stable learning.

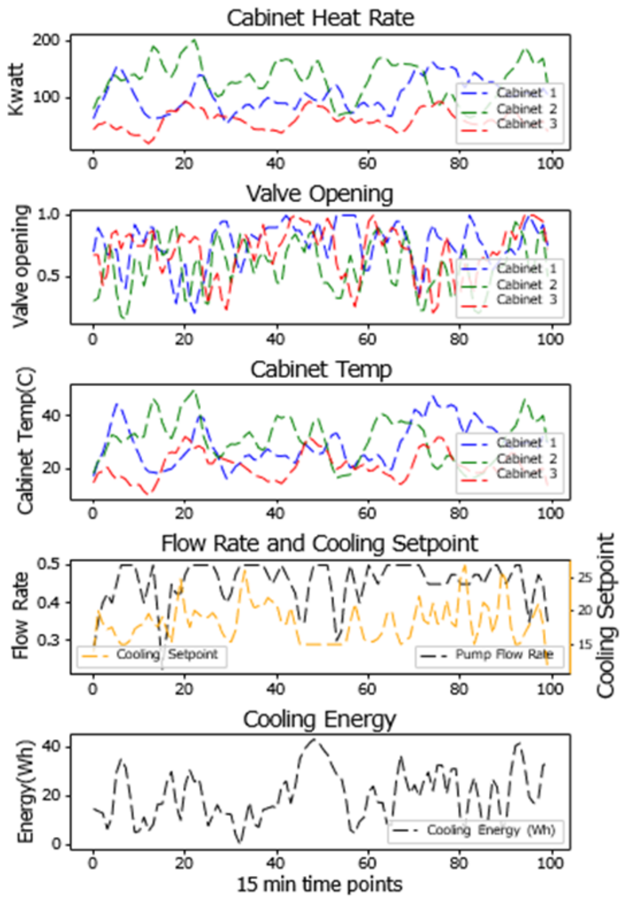
IT Cabinets: How it works



Non-supervisory PID Control



ASHRAE G36 Rule Based Trim & Respond Control



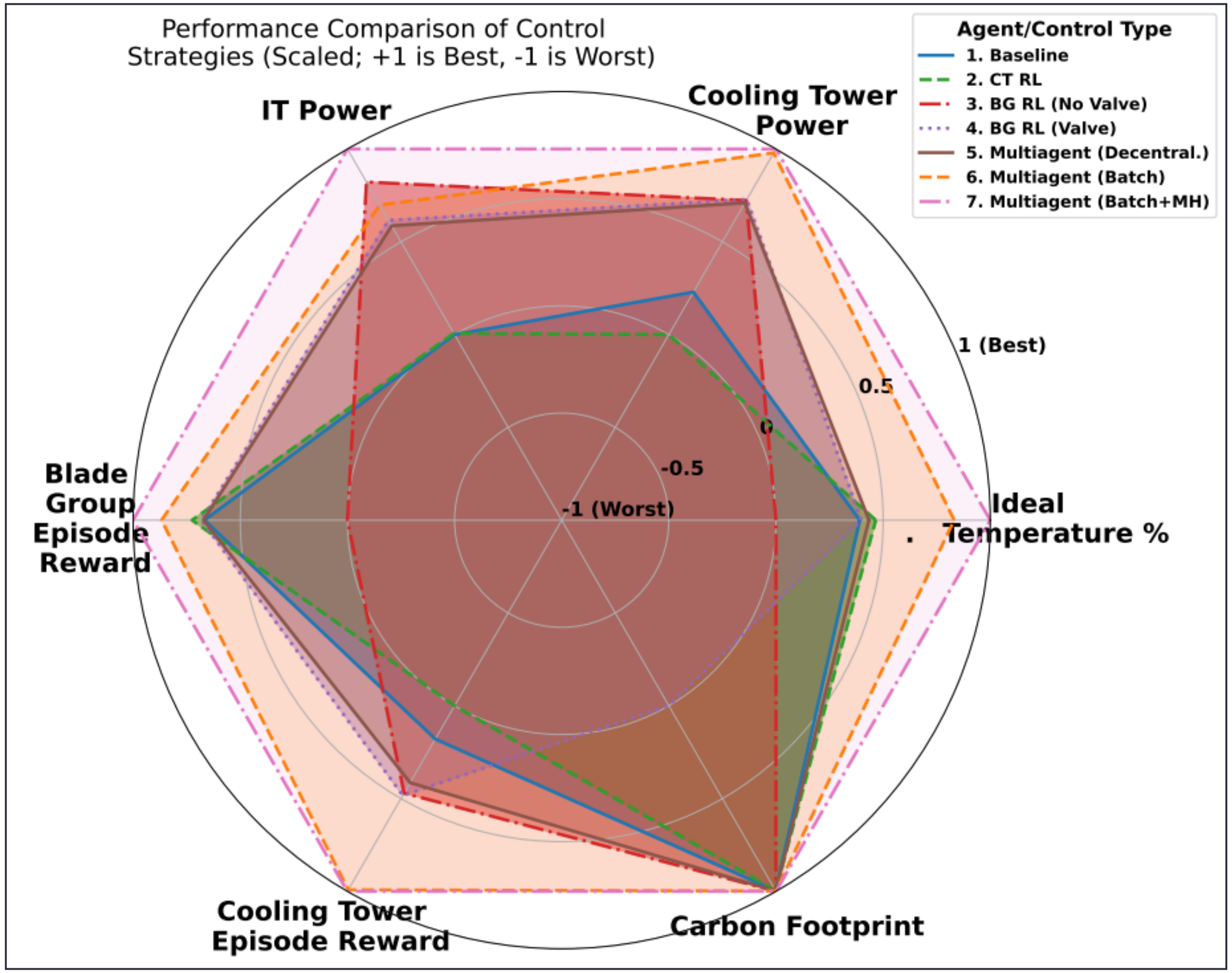
RL Control - with Blade Group Level Control

Results (1): Multi-Agent RL Dominates Baselines

NEURAL INFORMATION PROCESSING SYSTEMS 2025

- We benchmarked 7 strategies, from the industry-standard ASHRAE baseline (Case 1) to our advanced RL agent (Case 7).
- **Key Finding:** Granular valve control is critical. RL *without* it (Case 3) fails to maintain temperature (64.9% compliance).
- **The Winner:** Our **Multi-Agent RL (CA + Multi-Head)** policy (Case 7) is the best on all fronts.
- **Generalization:** The RL policy, trained on 5 cabinets, successfully scales to an unseen 25-cabinet system, maintaining 92.6% compliance while the baseline's performance collapses.

Metric → Agent/Control Type ↓	Control Details	$D_{blade,avg}$% (% of time Temp within ideal range)	P_{ij}(kW) (Cooling Tower Avg Power)	Q_i (IT Level Avg Cooling Power)	Avg Episode Reward per Cabinet	Avg Episode Reward per Cooling Tower
1. Baseline Control	ASHRAE G36	76.92	237.31	235.28	1697.08	360.17
2. CT RL + BG Baseline	Only CT RL control	79.21	246.46	235.03	1702.16	352.28
3. CT Baseline + BG RL	No Valve Control	64.91	217.6	203.96	1638.48	372.97
4. CT Baseline + BG RL	With Valve Control	77.13	217.37	211.83	1698.36	373.52
5. Multiagent RL	Decentralized Action	78.24	218.11	212.94	1697.49	370.51
6. Multiagent RL	Centralized Action (CA)	90.46	207.37	208.69	1714.65	395.88
7. Multiagent RL	CA & Multihead policy	**95.63**	**206.52**	**197.18**	**1726.31**	**396.24**



Approach 2: Agentic AI for Explainable Control

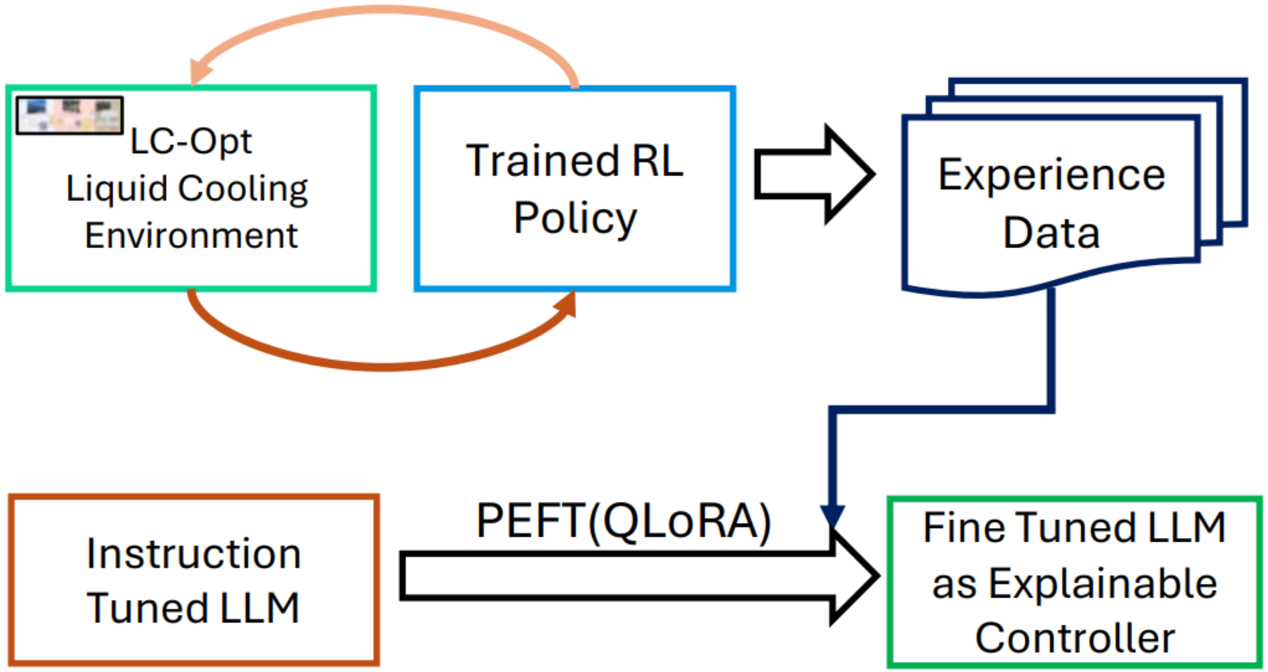
The Problem with RL: Deep RL policies are "black boxes." A data center operator will *never* deploy a controller they cannot understand or trust.

A Paradigm Shift: Can we use a Large Language Model as the control policy?

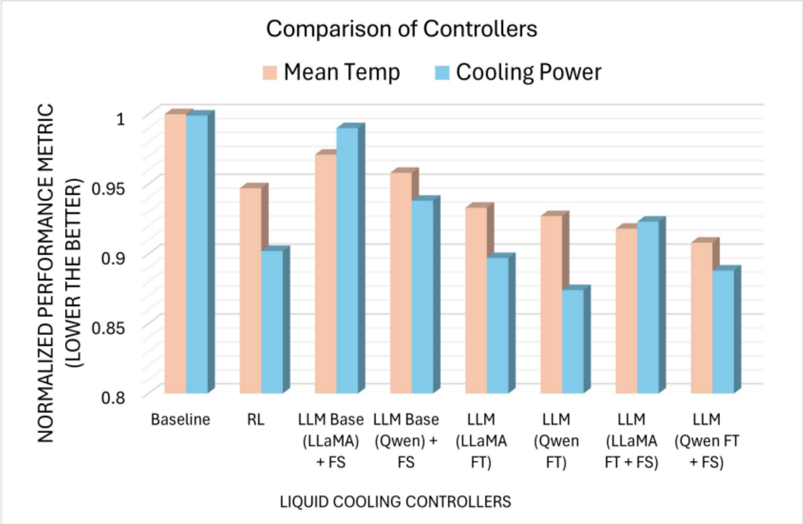
Method: Policy Distillation

- First, train a high-performance (PPO) RL agent (the "oracle").
- Generate "Experience Data" from this expert policy.
- Use PEFT (QLoRA) to fine-tune an instruction-tuned LLM on this expert data.

The Result: An "Agentic LLM" that can both *take control actions* and *explain why* it's taking them in natural language.



Results (2): LLMs match RL Performance



LLM Explainability and Guidance Example

Obs: (307.2, 306.74, 309.77, 21567.55, 44779.22, 33672.41), Act: (0.19, 0.61, 0.2, 22.39, 25.52)
Explanation: Branch 2 remains relatively cool, allowing the controller to redistribute cooling capacity toward more loaded or hotter branches. A setpoint of 22.39 and a controlled flow of 25.52 signify this adaptive behavior. The decision not only maximizes performance-per-watt but also aligns with global decarbonization efforts. It is recommended to reduce branch 1 coolant allocation to 0.1 and to allocate the rest to branch 3 since it has a higher workload
Validity: 1) Human Expert Evaluation: "This is one of the better examples where the agent is explaining the actions for all three blade groups. It also, provides a good way of reifining the RL valve actions"

The Surprise: Fine-tuned LLM controllers (distilled from RL) *outperformed* the original RL policy *and* the baseline.

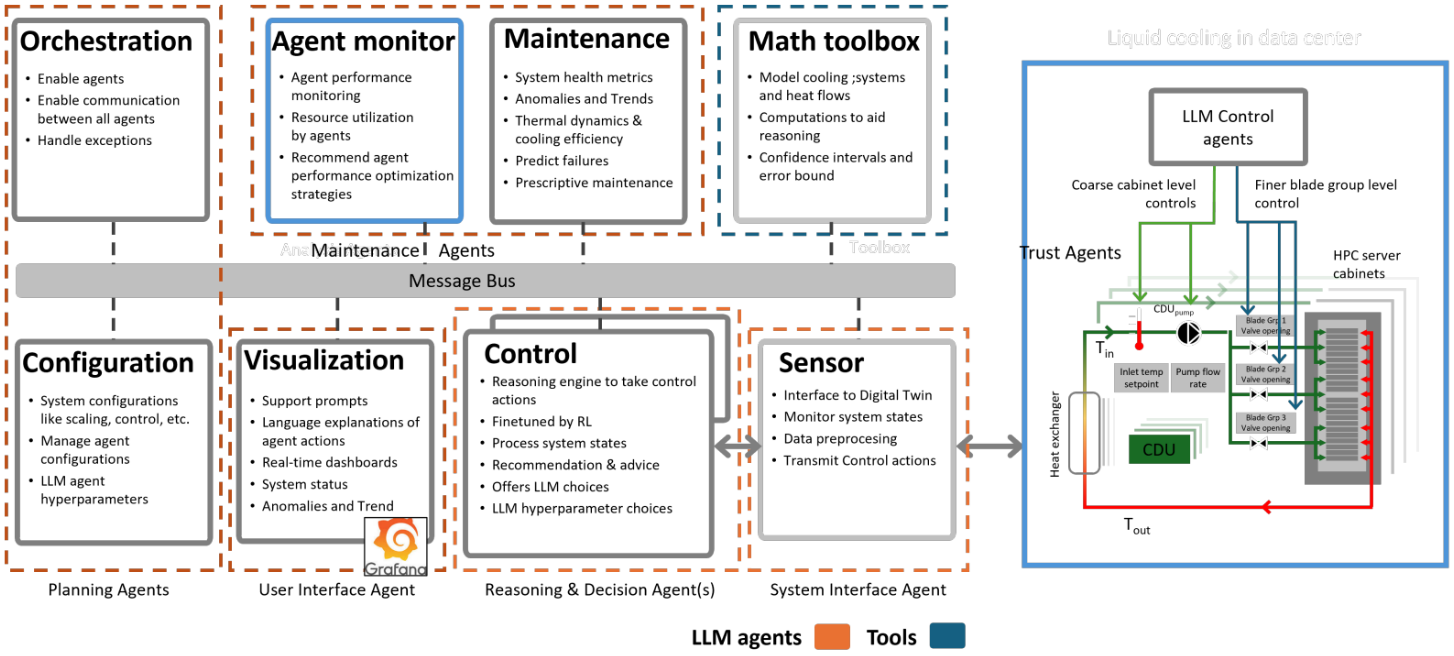
Why? The distillation process appears to smooth out and stabilize the original, high-variance RL policy, leading to more robust control.

Explainability: The LLM agent generates real-time, natural language explanations for its control actions.

Human-in-the-Loop Validation: We had domain experts validate these explanations for:

- **Trust:** Does the explanation match the action?
- **Comprehensiveness:** Does it explain *all* parts of the action?

Vision: An Agentic AI Ecosystem for Physical Systems



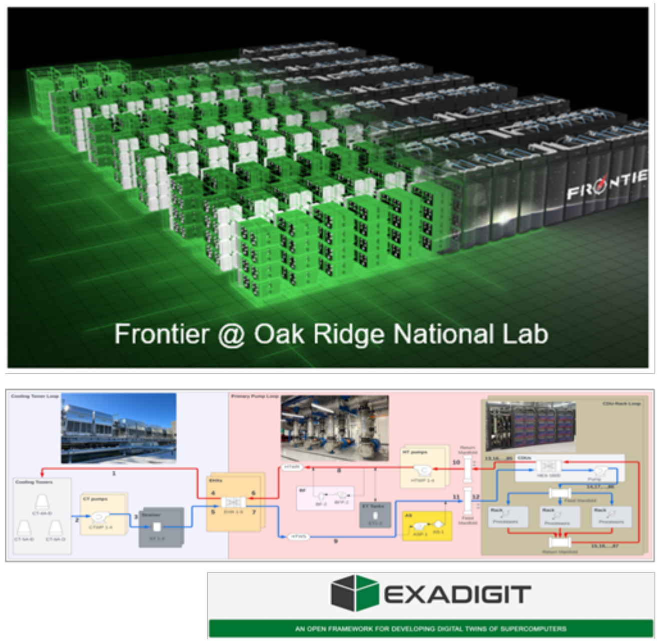
The LLM controller is just one piece of a larger **Agentic AI architecture** designed for trustworthy autonomy.

We propose a multi-agent system for holistic, explainable management:

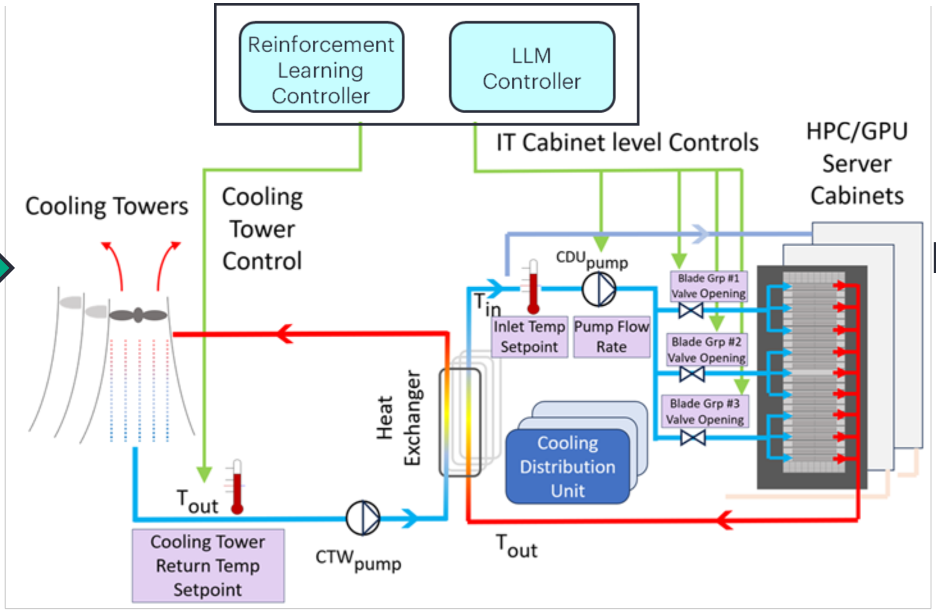
- **Control Agent:** The fine-tuned LLM policy making decisions.
- **Sensor Agent:** Interfaces with the digital twin (or real hardware).
- **Maintenance Agent:** Monitors health and predicts failures.
- **User Interface Agent:** Provides natural language explanations to the human operator.

A New Research Frontier: LC-Opt is a benchmark for studying **reasoning, control, and safety** of AI agents in mission-critical physical systems.

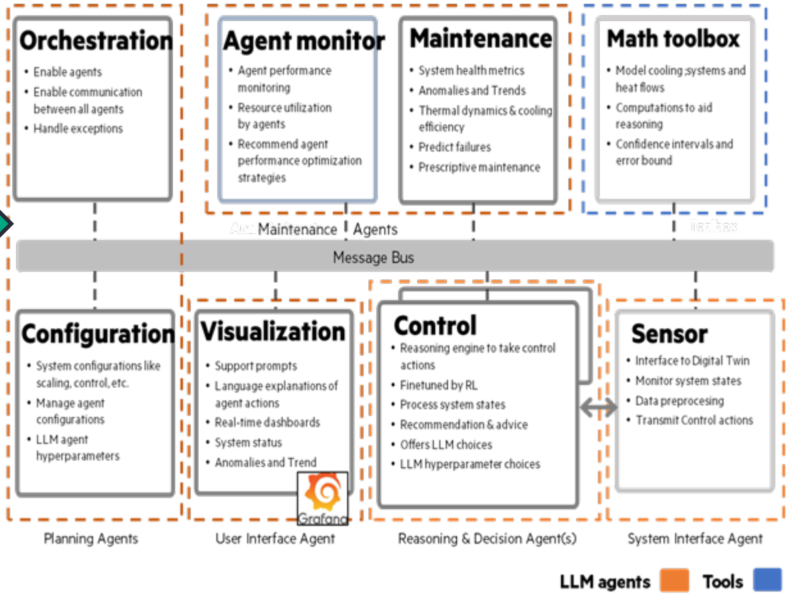
Putting it all together



Data Center Liquid Cooling Modelica Model



Reinforcement Learning & LLM Control



Agentic Mesh with LLM Agents

Impact: Agentic AI-Driven Sustainability

NEURAL INFORMATION PROCESSING SYSTEMS 2025

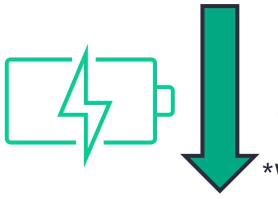
Environmental: Our best agent (Case 7) shows a **~24% reduction in carbon footprint** over the baseline in our 2-day experiment.

Efficiency: The benchmark supports **Heat Recovery Units (HRUs)**, which our tests show can cut cooling tower power by another **21%**.

Economic: This translates directly to reduced opex (energy costs) and better capex (designing more efficient facilities).

Deployable (Sim-to-Real): We provide a clear, 4-phase pathway from this benchmark to "shadow mode" and full production deployment .

-24%



-21%

*With heat recovery

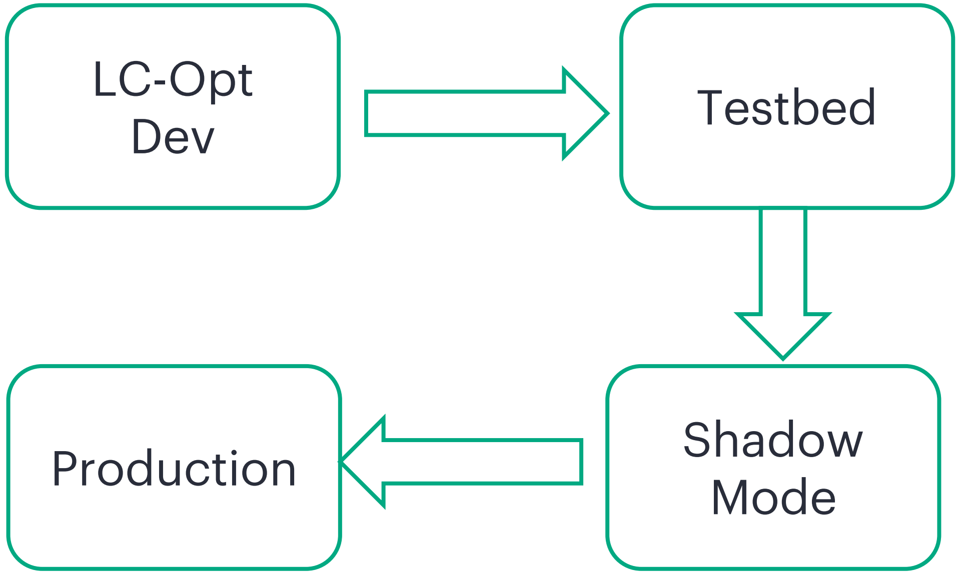


Table 3: Ablation of RL Agent Design with PPO. We incrementally replace the static baseline (Case 1) with RL controllers for: Cooling Tower (Case 2), CDU coolant setpoint/flow (Case 3), and Blade Group valves (Case 4). Case 5 introduces a single multi-agent RL controller, Case 6 adds batching for state space reduction, and Case 7 uses a multi-head policy. Experiments use N=2 towers, m=2 cells, C=5 cabinets with B=3 blade groups each, and are evaluated on an unseen exogenous trace. Blade-group temperature compliance $D_{blade,avg}$ is computed with $\mathcal{U}_T$=40°C and $\mathcal{L}_T$=20°C.

Metric →		$D_{blade,avg}\%$	$\sum P_{ij}(kW)$	$\sum Q_i$	**Avg Episode Reward**	
Agent/Control Type ↓	**Control Details**	(% of time Temp within ideal range)	(Cooling Tower Avg Power)	(IT Level Avg Cooling Power)	**per** Cabinet	**per** Cooling Tower
1. Baseline Control	ASHRAE G36	76.92	237.31	235.28	1697.08	360.17
2. CT RL + BG Baseline	Only CT RL control	79.21	246.46	235.03	1702.16	352.28
3. CT Baseline + BG RL	No Valve Control	64.91	217.6	203.96	1638.48	372.97
4. CT Baseline + BG RL	With Valve Control	77.13	217.37	211.83	1698.36	373.52
5. Multiagent RL	Decentralized Action	78.24	218.11	212.94	1697.49	370.51
6. Multiagent RL	Centralized Action (CA)	90.46	207.37	208.69	1714.65	395.88
7. Multiagent RL	CA & Multihead policy	**95.63**	**206.52**	**197.18**	**1726.31**	**396.24**

The Future: Trustworthy AI for Critical Infrastructure

NEURAL INFORMATION PROCESSING SYSTEMS 2025

LC-Opt is a foundational platform bridging industrial-scale systems and cutting-edge AI research.

It enables a new wave of research on **AI safety and alignment** in the physical world—where mistakes have real consequences.

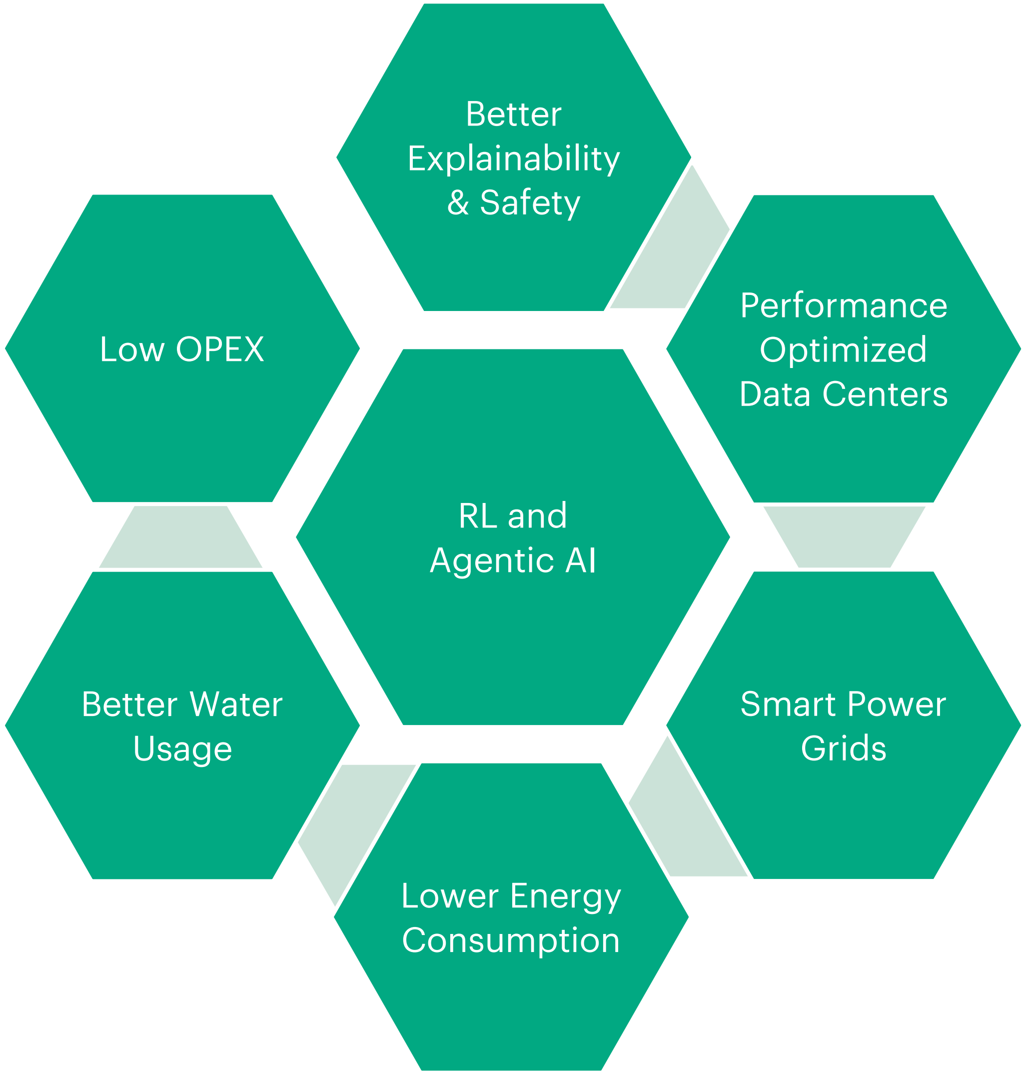
We are open-sourcing the benchmark and models via the **ExaDigiT consortium** to foster global collaboration.

Get involved:

- Paper: https://arxiv.org/abs/2511.00116
- Code: github.com/HewlettPackard/sustain-lc

Paper

Code



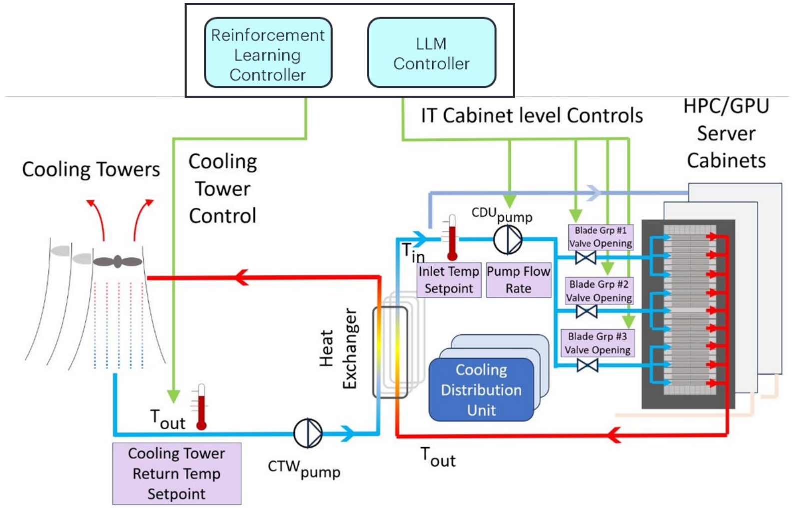
HPE

Oak Ridge National Laboratory

NeurIPS 2025

Neural Information Processing Systems

Thank You




Avisek Naug


Antonio Guillen


Vineet Kumar


Scott Greenwood


Wesley Brewer

Paper

Code


Sahand Ghorbanpour


Ashwin R Babu


Vineet Gundecha


Ricardo Luna


Soumyendu Sarkar

LC-Opt: Benchmarking Reinforcement Learning and Agentic AI for End-to-End Liquid Cooling Optimization in Data Centers

soumyendu.sarkar@hpe.com