The intelligent loop: AI and chilled water systems

Dr. Rand Talib

21 Aug 2025

As power densities rise, driven by AI training clusters, high-frequency trading and HPC workloads, chilled water systems remain the backbone of thermal management in many colocation and hyperscale data centers. While air cooling is still standard for most IT loads, facilities need to prepare for a shift toward liquid-cooled hardware and high-capacity cooling architectures. Chilled water loops, though familiar, now face demands for tighter tolerances, faster response and greater partial-load efficiency.

Various techniques are emerging under the umbrella of AI to help operators manage growing complexity and tighter operational margins. By enhancing predictive control, optimizing pump and chiller sequencing, and detecting inefficiencies before they escalate, AI-driven tools are redefining how chilled water systems are monitored and managed. However, the benefits vary widely depending on infrastructure maturity, sensor coverage and integration capability.

As operators and engineers prepare to tackle the next era of high-density cooling, they are increasingly turning to AI to augment control strategies, stabilize thermal conditions and extract greater efficiency from both legacy and modern loop designs.

In this context, AI refers to applied machine learning, predictive analytics, and optimization algorithms tailored to chilled water systems, tools that can dynamically adjust supply and return temperatures, manage delta-T (ΔT) stability and coordinate subsystems for peak efficiency. This definition does not include large language model (LLM) AI, which is not currently used for these types of data center applications.

The question is no longer whether these tools can support chilled water optimization, but how to deploy them effectively and where the limits of the tools lie.

High density, high complexity

Modern facility water systems are increasingly sophisticated, incorporating variable-speed drives (VSDs), pressure-independent two-way valves, thermal storage tanks and integration with

building management systems (BMS). In recent high-efficiency deployments, chilled water is typically supplied at 17°C to 20°C (63°F to 68°F) and returned at 20°C to 25°C (68°F to 77°F), yielding a ΔT of 5°C to 8°C.

While this range aligns with ASHRAE's recommended thermal guidelines, certain advanced or hybrid liquid-cooled deployments deliberately operate return temperatures toward the upper allowable limit, sometimes approaching 27°C to 30°C (81°F to 86°F), to improve chiller efficiency, extend free cooling hours and reduce pumping energy.

Recent developments in IT have made thermal management a more complex task. Notably, large AI compute clusters generate highly dynamic thermal loads, introducing frequent and unpredictable swings in cooling demand. These fluctuations can drive temperature instability, low ΔT conditions, and excessive chiller cycling, especially in hybrid environments that serve both air-cooled and liquid-cooled loads. As a result, maintaining thermal stability and efficient part-load operation is becoming significantly more complex.

AI and loop performance

Leading data center operators, such as Google, Microsoft and Meta, already embed AI-driven data analytics and control systems to optimize heating, ventilation and air-conditioning (HVAC) and chilled water loop performance. These systems go beyond basic automation by leveraging real-time sensor data, machine learning models and contextual forecasting to adjust parameters autonomously based on IT load, weather conditions and equipment behavior.

In chilled water environments, AI can contribute across four key operational domains:

1. Thermal load forecasting. AI systems ingest IT workload schedules, historical thermal profiles, and live telemetry to anticipate changes in cooling demand. This enables proactive resource staging, such as starting chillers early or adjusting flow, before loads materialize, reducing lag and overshoot.
2. Dynamic setpoint management. By continuously recalibrating supply temperature, differential pressure and valve positions, AI can fine-tune system setpoints to match real-time demand. This improves thermal stability and reduces energy use, particularly during partial-load conditions common in AI and HPC clusters.
3. Fault detection and diagnostics. Anomaly detection models monitor equipment behavior and flag inefficiencies, such as valve hunting, coil fouling or pump cavitation, before they trigger alarms or failures. This supports predictive maintenance strategies and reduces unplanned downtime.
4. Energy optimization across subsystems. AI coordinates chilled water plant subsystems, such as chiller sequencing, VFD pump control, air handlers and storage tanks, to minimize kilowatt (kW)/ton and flatten energy consumption curves. This holistic approach enables greater efficiency than manual or rule-based logic alone.

AI in action: so far

AI-based optimization is already in live use across major data center operators. Google

DeepMind forecasts short-term cooling demand every five minutes, autonomously adjusting chiller staging, pump speeds, and airflow to deliver up to 30% cooling energy savings. Meta uses AI to fine-tune supply water temperature and coordinate chiller/pump staging, improving ΔT stability and reducing water use in AI training clusters. Microsoft's pilots focus on thermal load prediction and zero-water goals in direct-to-chip and hybrid cooling facilities. At Singapore's National Supercomputing Centre (NSCC), a deep reinforcement learning model optimizes chiller loading, setpoints and cooling towers, achieving 11% to 15% cost savings over baseline controls.

These examples show that AI can improve efficiency, water conservation and stability across a variety of operating conditions. The next step is to understand the system-level impacts, how these capabilities translate into more stable ΔT under load swings, improved hydraulic loop performance and smarter, more responsive plant operation.

System-level impacts of AI-based optimization

AI-based control strategies are reshaping key performance parameters (see **Table 1**).

Table 1 Traditional versus AI-driven system functions

System function	Manual or traditional method	AI-driven enhancement
Delta-T stability	Typical delta-T of 5-6°C, often degraded by load volatility.	Predictive AI maintains 8-10°C by anticipating thermal shifts and tuning flow.
Thermal surge management	Short thermal surges strain chiller response during transitions.	AI-integrated thermal storage preconditions loops and buffers transient load changes.
Sensor granularity	Sensors are deployed sparsely, often at the room or CRAH level.	High-resolution RTDs, differential pressure sensors, and flow meters every 5-10 racks enable granular control.
Valve configuration	Three-way bypass valves allow return/supply mixing.	Two-way throttling valves allow precise zone-level cooling and eliminate mixing.
Chiller low-flow Resilience	Chiller trips at low load due to insufficient flow.	AI maintains minimum chiller flow via VFD control and storage coordination.
Operator trust and explainability	Operators resist black-box automation.	Site-trained AI with explainable models (e.g., SHAP, decision trees) supports operator oversight.

CRAH, computer room air-handler; RTD, resistance temperature detector; SHAP, Shapley additive explanations; VFD, variable frequency drive

UPTIME INSTITUTE 2025

uptime
INTELLIGENCE

Two capabilities highlighted in **Table 1** stand out in practice. Firstly, stabilizing ΔT under large and unpredictable load variations, common with AI training and HPC workloads, prevents efficiency losses from low ΔT syndrome and allows chillers and coils to operate at peak effectiveness.

Secondly, improved hydraulic loop management through two-way valve configurations and AI-driven modulation reduces mixing between supply and return water, preserves thermal stratification in storage and minimizes pump energy. Together, these advances translate directly

into more efficient cooling plant operation, lower operating costs and greater system resilience under stress.

The question remains: how ready are operators to hand over critical decisions to an algorithm? If, like Google, AI adjusts chiller staging and pump speeds every five minutes, who stays in charge? And how do we ensure the system remains safe?

More complex control systems can introduce more complex failures. These may range from unstable loop behavior due to sensor faults or insufficient data, control oscillations from poorly tuned algorithms and reduced situational awareness when decisions are made inside opaque "black box" models.

Addressing these risks starts with ensuring that the supporting infrastructure, integration approach and operational safeguards are strong enough to handle both planned and unplanned disruptions.

Integration requirements and limitations

Implementing AI for chilled water loop optimization requires more than just algorithms; it demands strong infrastructure, seamless integration and operator confidence. These capabilities carry real costs, because expanding sensor networks, adding control automation and integrating with supervisory platforms increase both capital outlay and operational complexity. More active components also mean more potential points of failure, making robust fallback design essential. In practice, deploying AI effectively depends on addressing several requirements and limitations:

- **High-resolution sensor infrastructure.** Accurate, granular data is essential. Facilities need high-resolution resistance temperature detectors (RTDs), differential pressure sensors, and electromagnetic or ultrasonic flow meters across supply and return lines and branch loops, ideally at every five to 10 rack intervals. This level of coverage improves AI performance, but adds to cost and maintenance requirements.
- **Control platform compatibility.** AI needs to integrate with BMS or supervisory platforms using open protocols such as BACnet/IP, Modbus TCP or via middleware. Real-time communication is essential for adjusting setpoints, pumps and valves, while preserving operator override. This integration layer itself should be resilient, because a failure could disrupt plant operation.
- **Operators trust and transparency.** To gain acceptance, AI models need to provide explainable outputs, for example, Shapley additive explanations (SHAP) values or decision trees, as well as clear diagnostics. Without transparency, operators may limit AI to advisory roles or disable automation entirely.
- **Resilience and fallback logic.** AI should be prepared for failures, power outages, communication loss and hardware issues. Systems need built-in fallback sequences or manual control options to ensure thermal safety and continuity. This raises a strategic question: should AI be mission-critical or should the plant be able to operate safely, though less efficiently, without it?
- **Site-specific adaptability.** AI models should be trained and tuned for local infrastructure, load profiles and climate. Generic models risk underperformance without ongoing validation.

The Uptime Intelligence View

AI is shifting from experimental pilots to a core part of chilled water plant strategy. Its strength lies less in chasing "perfect" efficiency and more in delivering stability, adaptability, and real-time insight at a scale that humans alone cannot match. By anticipating load swings, maintaining ΔT under stress and coordinating subsystems, AI is changing how high-density facilities approach cooling. But more automation brings more interdependence and, without a robust infrastructure, more points of failure. Success will come from balancing ambition with resilience: investing in the right sensors, seamless integration and keeping operators in the loop before AI becomes mission-critical.

ABOUT THE AUTHOR

Dr. Rand Talib

Dr. Rand Talib is a Research Analyst at Uptime Institute with expertise in energy analysis, building performance modeling, and sustainability. Dr. Talib holds a Ph.D. in Civil Engineering with a concentration in building systems and energy efficiency. Her background blends academic research and real-world consulting, with a strong foundation in machine learning, energy audits, and high-performance infrastructure systems.

rtalib@uptimeinstitute.com

About Uptime Institute

Uptime Institute is the Global Digital Infrastructure Authority. Its Tier Standard is the IT industry's most trusted and adopted global standard for the proper design, construction, and operation of data centers – the backbone of the digital economy. For over 25 years, the company has served as the standard for data center reliability, sustainability, and efficiency, providing customers assurance that their digital infrastructure can perform at a level that is consistent with their business needs across a wide array of operating conditions.

With its data center Tier Standard & Certifications, Management & Operations reviews, broad range of related risk and performance assessments, and accredited educational curriculum completed by over 10,000 data center professionals, Uptime Institute has helped thousands of companies, in over 100 countries to optimize critical IT assets while managing costs, resources, and efficiency.