

INTELLIGENCE UPDATE

AI in facility operations: three applications to watch



Max Smolaks

17 Nov 2025

Long before ChatGPT, Uptime Intelligence observed the rapidly developing capabilities of machine learning (ML) systems and envisioned how they could be applied to data center operations. Over the past three years, many of these theoretical use cases have increasingly become a reality.

At the end of 2023, we noted that innovation in AI for data center operations was being driven primarily by startups rather than established data center software vendors (see [Data center management software is evolving — at last](#)).

The situation has changed markedly by the end of 2025: data center infrastructure management (DCIM) vendors are now leading the charge as they work to enhance their products with AI-based functionality that has been tested in enterprise environments.

This report details three emerging applications of AI in data center operations. Some have already been productized, while others are in various stages of development.

Note: “Emerging” in this context means they have not been previously documented by Uptime Intelligence or the industry press. Any vendors or product names are intended to serve as examples and not recommendations.

AI for IT placement

AI-based IT hardware placement optimization is an example of functionality already on the market. In this application, a specialized ML model uses data from repositories such as a DCIM system, a configuration management database (CMDB) and/or a building management system (BMS) to automate the decisions related to installations, moves, adds and changes (IMACs) of IT equipment.

Such systems use an internal equipment catalogue to determine the specific requirements of each piece of IT equipment and then identify the optimal location for servers or racks, taking into account the availability of power, cooling, space, networking and hardware weight. They

can also ensure compliance with organizational policies and regulatory standards.

These systems are typically able to generate related work orders, describing additional tasks required — for example, the installation of network cards or cable patches.

Adopting AI for IT placement reduces the risk of human error, speeds up deployment and improves resource utilization. This type of software functionality is a relatively safe bet. It uses older and well-understood ML methodologies to solve combinatorial problems — an area in which ML excels. Even in situations where the system issues suboptimal or incorrect guidance, human supervision and the relatively low-risk nature of the task mean it is unlikely to result in any operational disruption.

Currently, Uptime Intelligence is aware of two commercial examples supplied by Nlyte (now owned by Carrier) and RiT Tech:

- Nlyte Software has been experimenting with AI longer than most, having launched an ML product in collaboration with IBM back in 2018. More recently, Nlyte released Placement and Optimization AI within its Asset Optimizer module. It can place up to 100 servers and chassis in a single project based on constraints measured in real-time. The software can conduct “what-if” analyses to simulate infrastructure changes before committing resources and automatically checks for issues with IMACs requests. It runs on-premises on traditional servers — no GPUs required.
- RiT Tech launched an AI-powered IT provisioning module in 2024 as part of the XpediTe DCIM suite. The company says that applying AI to provisioning can reduce the time needed for IMAC planning by up to 95%. RiT is also working on applying ML to model the future state of the data center, looking to further assist in capacity planning, which is regularly reported in Uptime Institute surveys as the most significant challenge facing data center operators.

Data center LLMs

If large language models (LLMs) can be given access to corporate data to perform analytics based on natural-language prompts, why can't the same functionality be extended to operational facility data? At least three DCIM vendors believe LLMs can be made useful in data center facility operations and are currently working to make this a reality.

The idea is that a conversational interface for DCIM will help new staff learn the specifics of the facility and operating procedures. It can also make operational data more accessible and useful to a broader set of non-technical employees who were not previously involved in facilities management.

A well-documented issue with using LLMs in applications where accuracy and consistency matter is their propensity for hallucinations — when they produce confident but inaccurate responses. Data center software vendors believe these issues can be minimized by training models on domain-specific data and by training employees how to prompt the systems effectively. Below are examples from MCIM, RiT Tech and Nlyte:

- MCIM by Fulcrum Collaborations collects anonymized operational data from hundreds of sites. This enables customers to benchmark equipment and facilities against peers — and it also creates a large pool of data for AI training. MCIM is working on a proof of concept designed to let customers interrogate operational data and documents using natural-language queries. Potential applications include root-cause analysis, routine maintenance queries, and questions related to methods of procedures (MOPs) and standard operating procedures (SOPs).
The company is showing its prototype to existing customers and gathering feedback. Any product is at least a few quarters away and will likely be offered as an optional extension to the core MCIM platform.
- RiT Tech is close to releasing an AI analytics agent based on a fine-tuned version of a major open-source LLM. The system can ask clarifying questions in response to queries, generate tables and charts, and provides a library of live dashboards that can be built on demand in real time. RiT Tech is also developing AI-powered data-validation functionality that considers whether operational data is correct and complete. The company plans to continue building AI into its XpediTe system, eventually creating a platform that proactively recommends optimization steps to take.
- Nlyte AI is an upcoming agentic framework based on Microsoft's Copilot and developed in close collaboration with the company. Although the service is cloud-based, all operational data remains on premises. Nlyte is creating industry-specific templates and ontological relationships to help customers derive immediate value from the conversational interface. The system is expected to eventually automate analysis, detect anomalies, predict failures and recommend actions. Pilot customers include several global financial institutions, and the product is slated for launch in the first quarter of 2026.

Vision-language models for data centers

Industries such as construction and manufacturing have started combining image recognition and augmented reality (AR) technologies to assist employees in their day-to-day tasks. The same technologies can be deployed in the data center — for example, enabling staff to quickly diagnose incorrect cable connections through the use of virtual color-coding or automatically looking up equipment specifications, manuals or SOPs based on what appears in their field of view.

Vision-language models (VLMs) offer one solution to staff shortages, lowering the barrier to entry for data center operations and enabling junior staff to work without oversight or assistance.

Unlike other applications described in this report, VLM-based tools require additional devices: cameras for image recognition and digital interfaces, such as smartphones or tablets, for AR imaging. Purpose-built AR headsets can also be used, although these remain prohibitively expensive.

- US startup Visum AI emerged from stealth in 2025 to pursue computer-vision applications for data center employees. The company is developing VLMs that it hopes will guide staff through common data center tasks (e.g., DIMM replacement or switch cabling), support training, verify completed installation or maintenance work,

and help in the creation of digital twins by scanning the facility and replicating it in 3D. Much of this is aspirational, and there are no timelines for when data center VLMs will reach the market. However, interest is there — Visum AI has announced a collaboration with Siemens, which will likely become its first customer.

The Uptime Intelligence View

For many years, Uptime Intelligence has maintained that AI-based functionality in data center operations would be rolled out slowly, with initially conservative and limited applications. This remains the case.

At the same time, the number of available options for operators is increasing and will continue to expand. The fact that traditional DCIM vendors are finally taking AI seriously, and even exploring more adventurous use cases, bodes well for innovation in data center software tooling, which has long been stagnant.

Other related reports published by Uptime Institute include:

[*Lack of trust will hinder adoption of AI-based controls*](#)

ABOUT THE AUTHOR



Max Smolaks

Max is a Research Analyst at Uptime Institute Intelligence. Mr Smolaks' expertise spans digital infrastructure management software, power and cooling equipment, and regulations and standards. He has 10 years' experience as a technology journalist, reporting on innovation in IT and data center infrastructure.

msmolaks@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.