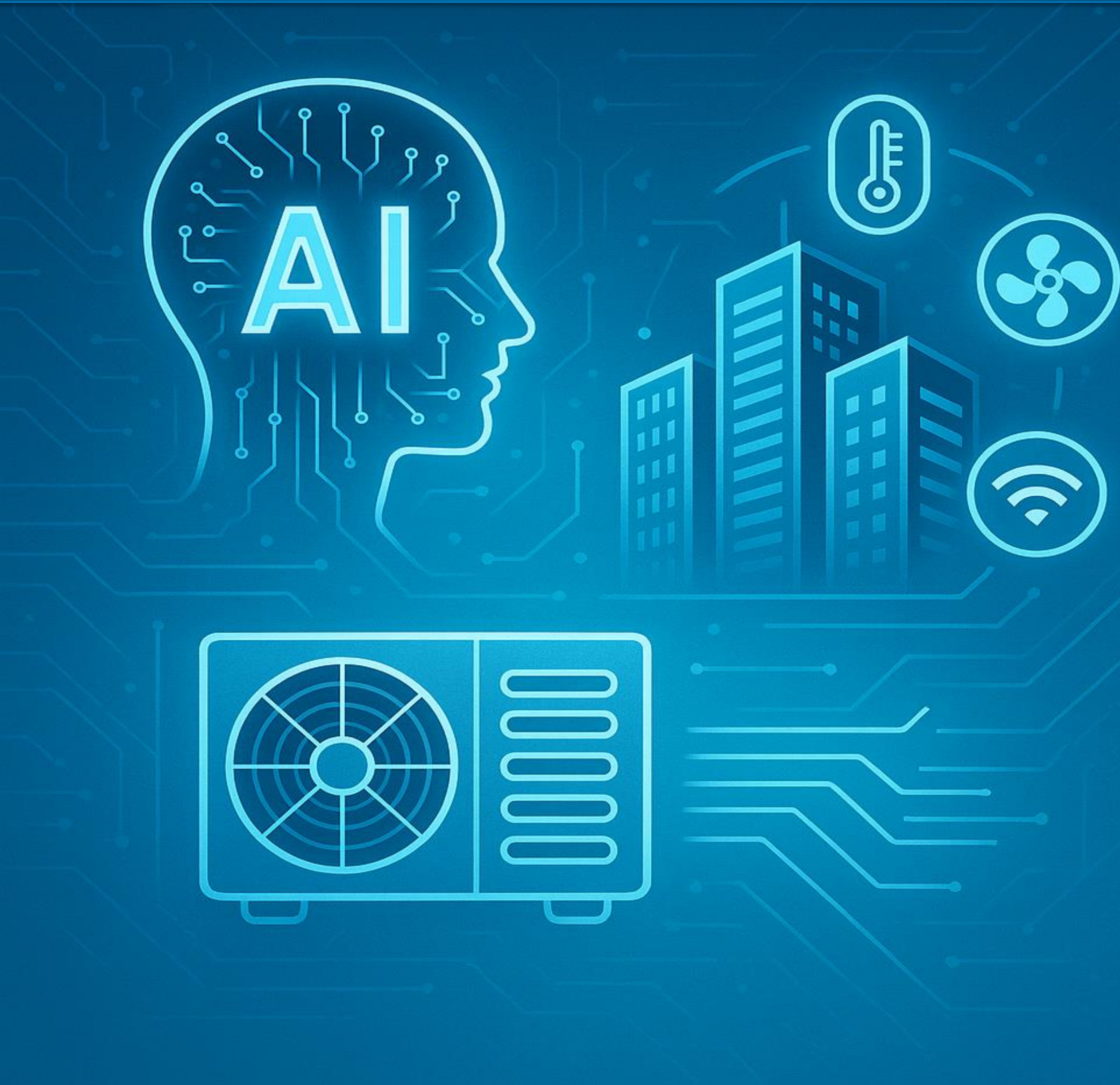
BOSCH

Agentic AI on the Edge:
The Future of HVAC Control

Author: Home Comfort & AI Teams, BGSW



Contents

1. Introduction

Modern smart buildings require intelligent systems that not only react to environmental changes but also anticipate and adapt to them proactively. Conventional rule-based HVAC (heating, ventilation, and air-conditioning) systems often lack the flexibility to optimize across varying comfort demands, occupancy patterns, and dynamic energy pricing. Agentic AI enhances energy efficiency, occupant comfort, and system adaptability by leveraging sensor data, utility grid signals, and human presence patterns. Since HVAC systems also use standalone edge devices, methods, and challenges to deploy agentic AI systems on edge are also described briefly through illustrative workflows and data flow diagrams. This paper details how agent-based AI frameworks can be applied to **Agentic AI-based edge-intelligence for HVAC systems** and smart buildings to improve energy efficiency, sustainability, and occupant satisfaction.

1.1. How do existing HVAC systems work?

Traditional building automation systems are often designed with predefined rules, static Time programs, and manual overrides, which limit their ability to respond to real-time dynamics such as occupancy changes (party at home), weather fluctuations, or energy pricing. Current building automation includes a remote controller to set the heating demand of that room. This can be a simple encoder or a traditional room controller with edge intelligence & display or a mobile application which sends the request via a gateway. The gateway may sometimes be connected to internet. The demand sent from the remote controller is provided to by a boiler or a heat pump which uses either non-renewable sources of energy such as oil, gas, or renewable sources of energy such as solar, geo-thermal to provide heat. The heat from boiler or heat pump or cooling unit is transferred to the room directly by means of hot/cool air or more commonly by a liquid medium through a system of closed loop pipes.

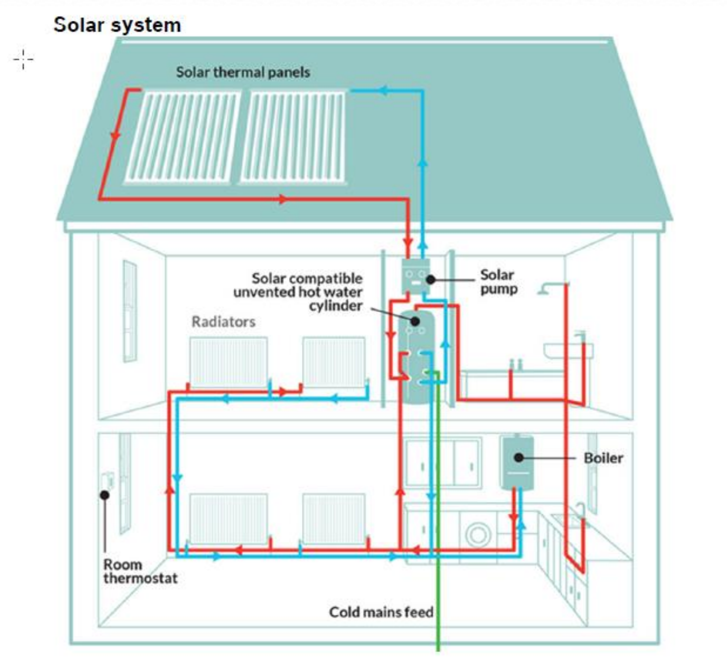


Figure 1 : A typical room heating setup using thermostat as temperature control device and using a boiler and solar as heat source.

1.2. What is Agentic AI?

In contrast, Agentic AI represents a new generation of artificial intelligence capable of independent goal setting, adaptive planning, and autonomous execution—offering a transformative leap for the HVAC and smart building domain. Agentic AI refers to artificial intelligence systems that possess goal-driven autonomy and can perceive their environment, derive meaningful goals, plan actions, and adapt to feedback.

A typical **Agentic AI** refers to AI systems that:

- Possess agency, meaning they act autonomously toward achieving a goal.
- Can perceive their environment through diverse inputs (e.g., sensors, external intelligence, external APIs).
- Are capable of reasoning about their current and future state.
- Can formulate objectives, develop plans, and execute actions independently.
- Learn from the outcomes of their decisions to improve future performance.
- Communicate with other AI Agents for shared intelligence and actions.

2. What is Agentic AI in HVAC?

Traditional building automation systems are often designed with predefined rules, static Time programs, and manual overrides, which limit their ability to respond to real-time dynamics such as occupancy changes (party at home), weather fluctuations, or energy pricing. In contrast, Agentic AI represents a new generation of artificial intelligence capable of independent goal setting, adaptive planning, and autonomous execution—offering a transformative leap for the HVAC and smart building domain. Agentic AI refers to artificial intelligence systems that possess goal-driven autonomy and can perceive their environment, derive meaningful goals, plan actions, and adapt to feedback. In HVAC and building systems, these capabilities translate to autonomous decisions on pre-conditioning spaces, adapting airflow, and minimizing energy costs.

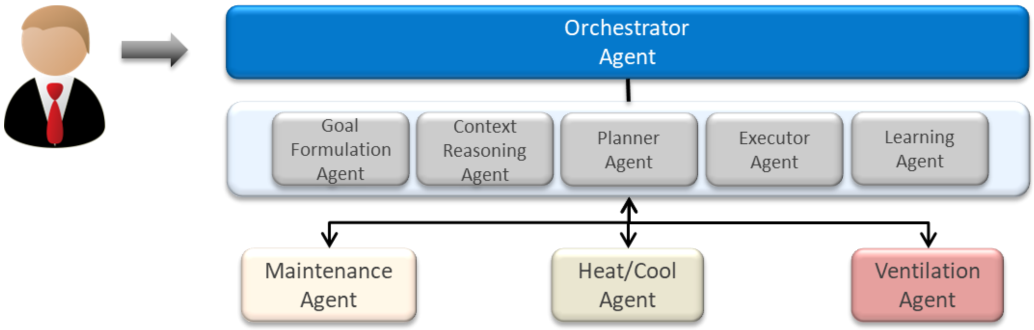


Figure 2 : Diagram highlighting typical agentic communication in HVAC system.

2.1. How Agentic AI is different from traditional solution

Feature	Traditional Automation	Agentic AI
Programming Model	Rule-based logic	Goal-based reasoning
Adaptability	Low	High
Real-Time Decisioning	Limited	Advanced
Learning from Feedback	No	Yes
Scalability	Manual configuration	Distributed intelligence with orchestrators
Example	"Turn on cooling at 8:00 AM"	"Prepare occupied areas for optimal comfort by 9:00 AM using cheapest energy window"

2.2. Enabling Agentic AI in HVAC Edge Network

The deployment of agentic AI—AI systems capable of autonomous decision-making and action—across a continuum from cloud to edge is a pivotal development in distributed intelligence. This analysis, drawing from recent research and industry insights, explores how agentic AI is transforming industries by leveraging both centralized cloud resources and decentralized edge computing, tailored to specific use cases, latency requirements, connectivity constraints, and computational needs.

There are multiple approaches for deploying Agentic AI for HVAC systems such as:

1. Centralized AI Deployment

All Agentic AI components run on a central cloud or on-premises server, receiving data from all zones, and sending back decisions.

2. Distributed/Edge AI Deployment

Agents are deployed closer to field devices (on local edge controllers), enabling low-latency response, resiliency, and offline functionality.

3. Hybrid Architecture

Combines centralized intelligence with local autonomy, offering the benefits of both strategies for large-scale buildings and campuses.

2.3. Key Factors Influencing Agentic AI deployment in Edge HVAC network

Research suggests that the deployment location of agentic AI is determined by several factors, each influencing where along the continuum the AI operates:

- **Use Case**: For tasks requiring real-time decision-making edge deployment is preferable. For instance, computer vision models in smart building detect defects at the edge, ensuring immediate action.
- **Latency Requirements**: Applications needing low-latency responses benefit from edge deployment to minimize delays. For less time-sensitive tasks, cloud deployment is sufficient, as delays in data transmission are less critical.
- **Connectivity Constraints**: In environments with unreliable or no internet connectivity, edge deployment is essential for autonomous operation.
- **Computational Needs**: AI agents requiring significant computational resources, cloud deployment is more appropriate. The advancements in edge computing hardware and techniques like model distillation are enabling more sophisticated models to run on edge devices.

2.4. How Edge AI enables Agentic Behavior in HVAC Systems

Agentic Trait	Edge AI Role in HVAC	Examples
Autonomy	HVAC systems can self-regulate based on internal state + sensors. **Edge AI in HVAC** empowers autonomous operation by allowing systems to • Collect and process sensor data (e.g., temperature, occupancy, CO_2 levels, humidity) locally. • Run pre-trained AI models at the edge to decide when to heat, cool, or ventilate without waiting for cloud input.	A rooftop HVAC unit autonomously adjusts its compressor operation based on real-time solar load.

Agentic Trait	Edge AI Role in HVAC	Examples
	• Operate even if cloud or internet connectivity is lost — crucial for resilience in critical environments (e.g., hospitals, industrial plants).	
Reactivity	Responds instantly to environmental changes (e.g., sudden crowd). • HVAC systems can sense and respond to local disturbances (e.g., door opens, people enter a room) with millisecond latency. • AI models embedded in controllers adjust fan speeds or damper positions instantly based on microenvironment shifts.	A classroom's HVAC increases ventilation within seconds when CO_2 levels spike due to a sudden influx of students.
Proactivity	Anticipates needs (e.g., pre-heats a room before residents arrive). • Predictive control models that learn patterns (e.g., occupancy schedules, weather trends) and act *before* conditions shift. • Integration with user calendars, sensors, and historical data to pre-condition environments.	A smart home starts pre-cooling/heating a residence before a resident arrives home.
Goal-Oriented	Balances comfort and energy goals through local decision loops. • Use optimization algorithms running locally to balance competing goals like energy cost, indoor air quality, and user comfort. • Continuously evaluate trade-offs using edge-deployed ML models (e.g., reinforcement learning).	An HVAC unit shifts to eco-mode during peak grid pricing hours unless critical temperature thresholds are at risk, aligning with the building's energy cost reduction goals.
Adaptability	Learns from environment/user behavior and evolves control logic. • Local learning loops — systems adjust control strategies over time using reinforcement or continual learning models. • Edge-based anomaly detection that evolves as the system observes new data patterns (e.g., drift in sensor behavior, occupancy anomalies).	A hotel room's HVAC learns a guest's comfort preferences across stays and automatically sets the room to their favored climate conditions upon check-in.

3. Architecting Agentic AI at the Edge for HVAC Systems

The system architecture for applying Agentic AI to HVAC and smart building environments follows a **modular** and **layered design**. Each layer serves a distinct role, enabling scalability, robustness, and adaptability across various building types and control strategies. Below is a high-level breakdown of the architecture:

Below picture captures different components of Agentic AI for HVAC system

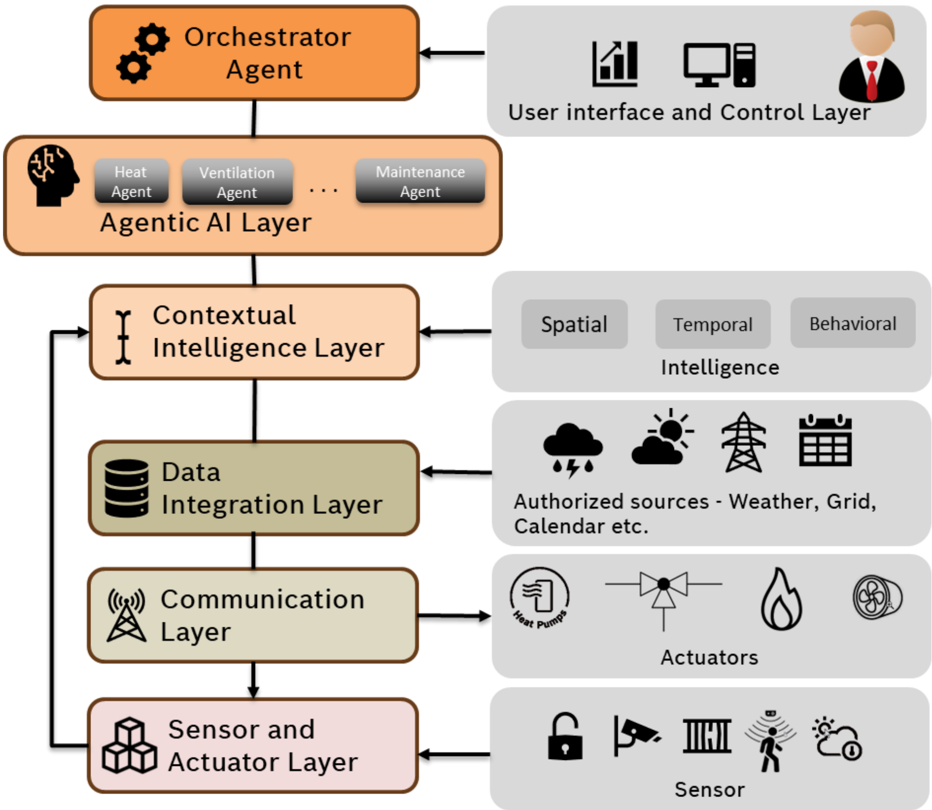


Figure 3 : Different components of Agentic AI for HVAC system

3.1. Architecture Layers

1. Sensor and Actuator Layer

- Includes field-level devices: temperature sensors, humidity sensors, CO_2 detectors, motion sensors, window/door sensors, and actuators like dampers, valves, pumps, fans, and VAV boxes.

2. Communication Layer

- Facilitates secure, low-latency data exchange via protocols like Modbus, BACnet, Zigbee, BLE, RS-485, Wi-Fi, or Ethernet. It enables reliable connectivity between field devices and higher control layers.

3. Data Integration Layer

- Aggregates and normalizes data from various sources:
  - Internal (building sensors, equipment feedback).
  - External (weather APIs, electricity grid signals, occupancy calendars).
- Also manages time-series storage and real-time streaming.

4. Contextual Intelligence Layer

- The Contextual Intelligence Layer acts as the brainstem of a smart building's decision-making framework. It is responsible for fusing data from multiple sources, extracting operational context, and guiding downstream Agentic AI decisions. This layer allows the system to shift from rule-based reactions to awareness-driven optimization, significantly improving responsiveness, adaptability, and efficiency.
- Derives spatial, temporal, and behavioral context from integrated data:
  - Spatial intelligence
    1. Determines how zones are physically used.
    2. Adjusts HVAC intensity based on room size, exposure (east vs west), and function (server room vs meeting room).

            3. Coordinates between zones to avoid cross-over inefficiencies (e.g., cooling adjacent to heating zones).
        - Temporal intelligence
            1. Aligns building operation with **time-based patterns** like:
                a. Weekday vs weekend behavior
                b. Morning vs evening load spikes
                c. Scheduled events (e.g., meetings, shifts, holidays)
            2. Enables **predictive control**, such as pre-cooling or staggered equipment start-up.
        - Behavioral intelligence
            1. Learns human usage patterns:
                a. Which rooms fill up at what times?
                b. What are comfort preferences per season or tenant?
            2. Adapts system behavior over time using AI feedback loops.
        - Identifies occupancy patterns.
        - Predicts solar gain.
        - Maps airflow and comfort needs across different zones.
    - Supports dynamic goal adaptation and operational awareness.

5. Agentic AI Layer

- Comprises multiple autonomous agents, each responsible for a specific zone, system, or objective (e.g., temperature regulation, ventilation control).
- Core components:
    - **Goal Formulation Agent**: Sets short- and long-term control objectives.
    - **Context Reasoning Agent**: Evaluates inputs and adapts strategies.
    - **Planner Agent**: Builds operational schedules.
    - **Executor Agent**: Issues actionable commands.
    - **Learning Agent**: Continuously improves decision quality through reinforcement learning.
- Additional intelligent Agents:
    - **Thermal Comfort Agent** for individual zones
    - **Ventilation Agent** for air quality and circulation
    - **Energy Optimization Agent** for utility interactions
    - **Maintenance Agent** for anomaly detection and predictive service
    - **Building supply chain Agent**
    - **Elevator Agent**

6. Orchestrator Agent (Coordinator Layer)

- Supervises collaboration across agents, ensuring consistency in goal alignment.
- Resolves conflicts (e.g., shared air handling resources across zones).
- Adjusts priorities during special events or emergencies.

7. User Interface and Control Layer

- Offers facility managers access to:
    - Real-time dashboards.
    - Override and scheduling options.
    - AI explainability (why a certain action was taken).
    - Reports on energy savings and comfort trends.

4. Applications of Agentic AI at the Edge for HVAC Systems

Agentic AI transforms HVAC systems into adaptive, intelligent ecosystems. Each component (e.g., thermostat, boiler controller, chiller) becomes an **intelligent agent** capable of optimizing its behavior while coordinating with others.

Agents can in many places replace human actions and in some places support human actions. Agentic AI applications can improve.

1. Human comfort.
2. Energy efficiency.
3. Prevent/notify breakdown of systems.

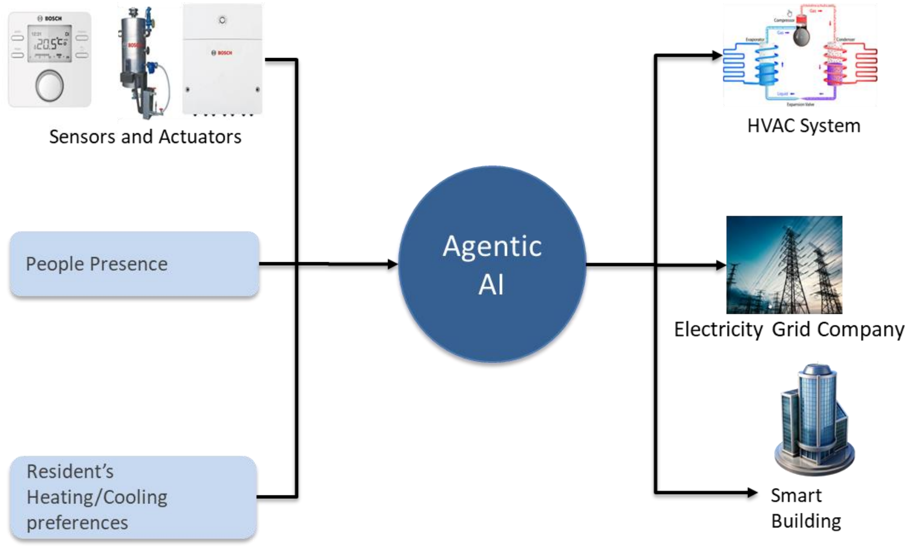


Figure 4 :Agentic AI in HVAC

Some possible Agentic AI applications are as follows:

Grid-Interactive Thermal demand

AI Agents running on HVAC systems can optimize HVAC operation by pre-cooling or pre-heating spaces during off-peak electricity hours when rates are lower, reducing energy costs during peak periods.

AI agents can get pricing data directly from the internet or when the user connects his phone through a secure mobile application to the HVAC system.

By following this approach, the user can save costs and energy company can also cope with better energy demands.

Predictive Maintenance – identify carbon buildup in the boiler

Agentic AI systems equipped with predictive maintenance capabilities can monitor equipment health in real time and anticipate failures before they occur. In the case of boilers, the system analyzes data such as burner efficiency, exhaust gas temperature, vibration, and combustion air ratios to detect early signs of carbon buildup—a common issue that reduces efficiency and increases the risk of breakdowns. By identifying these patterns proactively, the system can trigger alerts or schedule maintenance during low-demand periods, preventing costly downtime and extending equipment life. This minimizes unplanned service calls, reduces energy waste, and helps maintain optimal performance across the HVAC infrastructure.

Occupancy sensors detect room usage in real time and adjust flow accordingly

By using various occupancy detection sensors such as Passive Infra-Red (PIR) sensors Agentic AI systems can be fed with this data that allows agents to turn off room conditioning when not needed.

They can also co-ordinate between different AI agents in various rooms especially when a house is occupied by a set of people, and they are moving between rooms.

Edge-Based Agents Adjust Airflow and Temperature Locally

Rather than relying solely on a centralized control system, edge-based agents are deployed directly at the HVAC endpoints—such as air handling units, variable air volume (VAV) boxes, and thermostats. These agents use localized data (e.g., temperature, humidity, CO_2 levels, and occupancy) to make immediate, context-aware decisions. For example, if an office becomes unexpectedly crowded, the local agent can increase ventilation and reduce temperature on the spot—without waiting for instructions from a central server. This decentralization not only improves responsiveness but also ensures resilience and operational continuity even if network connectivity is lost.

5. Conclusion

The integration of **Agentic AI** and **Edge AI** is redefining HVAC and smart building systems—transforming them from passive infrastructure into intelligent, self-optimizing ecosystems. Unlike traditional control systems, these buildings *think locally* and *act autonomously*, with edge-deployed agents making real-time decisions based on context, occupancy, and environmental shifts.

Key benefits include:

- **Energy Efficiency**: Dynamic optimization driven by real-time, on-site intelligence.
- **Comfort Personalization**: Adaptive comfort tailored to individual behavior.
- **Operational Resilience**: Instant response to disruptions without cloud dependency.
- **Cost Savings**: Reduced energy waste, manual oversight, and latency.

This amalgamation of Agentic AI at the Edge for HVAC Systems reduces dependence on cloud infrastructure, improving latency, data privacy, and system resilience. It also allows for better scalability and coordination across distributed components. With autonomous capabilities, HVAC systems can achieve predictive maintenance, energy efficiency, and enhanced user comfort—all with reduced human intervention.

1 References

- **IBM Research – Agentic AI**
  What Is Agentic AI? | IBM
- **IEEE Smart Buildings Standards**
  IEEE P2784™ – *Guide for the Integration of Intelligent Building Systems*
  https://standards.ieee.org/
- **International Energy Agency (IEA)**
  Digitalization and Energy
  https://www.iea.org/reports/digitalisation-and-energy
- **Nguyen, T. A., & Aiello, M. (2013).**
  Energy intelligent buildings based on user activity: A survey.
  Energy and Buildings, 56, 244–257.
  https://doi.org/10.1016/j.enbuild.2012.09.005