
Intelligent Building Automation and Energy Optimization: Exploring How AI and IoT Frameworks Can Dynamically Manage HVAC Systems, Lighting, and Other Building Infrastructure Based on Real-Time Occupancy and Environmental Data

[Demarkus Samuelson](#) *

Posted Date: 9 February 2026

doi: 10.20944/preprints202602.0707.v1

Keywords: intelligent building automation; occupancy prediction; distribution shift; drift detection; uncertainty quantification; graceful degradation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Intelligent Building Automation and Energy Optimization: Exploring How AI and IoT Frameworks Can Dynamically Manage HVAC Systems, Lighting, and Other Building Infrastructure Based on Real-Time Occupancy and Environmental Data

Demarkus Samuelson

Independent Researcher; demarkussamuelson@gmail.com

Abstract

Intelligent building automation uses IoT sensing and AI control to coordinate HVAC, lighting, and ventilation in real-time. Current occupancy prediction and building dynamics models often report strong performance on held-out test sets drawn from the same distribution as the training data. However, operational buildings experience persistent distribution shifts due to hybrid work schedules, seasonal changes, organizational restructuring, sensor drift, and equipment degradation. Under these shifts, models can remain confidently incorrect, causing comfort violations, energy waste, excessive equipment cycling, and loss of operator trust. This study frames intelligent building automation as a cyber-physical deployment problem in which competence boundaries must be detected, communicated, and enforced at runtime. It contributes a six-dimensional human-centered AI framework adapted to building control and uses it to identify gaps in robustness evaluation, uncertainty quantification, drift detection for building time-series, and graceful degradation strategies when AI components become unreliable. The metrics-and-methodology section specifies shift-aware evaluation procedures, calibration and drift indicators, and closed-loop control benchmarks that connect prediction reliability to energy and comfort outcomes. Case analyses illustrate API-integrated occupancy-based HVAC, environmental-sensor occupancy detection, multi-zone occupant-centric control, and reinforcement-learning control under drift, demonstrating how a safe fallback to MPC and conservative modes stabilizes performance.

Keywords: intelligent building automation; occupancy prediction; distribution shift; drift detection; uncertainty quantification; graceful degradation

1. Introduction

1.1. Context and Motivation

Buildings are continuous cyber-physical systems that translate digital decisions into physical outcomes through HVAC, lighting, ventilation, and equipment scheduling systems. The intelligent building trajectory places machine learning and networked sensing inside operational control loops to reduce energy consumption while maintaining indoor comfort and air quality under dynamic occupancy [1]. IoT frameworks enable high-frequency data collection, processing, and analytics across heterogeneous sensors and building subsystems, making real-time inference and control technically feasible at a large scale [2]. Occupancy prediction has become a central control primitive because occupancy drives heating and cooling demand, ventilation requirements, and lighting usage, and the literature shows the broad adoption of machine learning occupancy prediction for energy

efficiency, air quality, and comfort management [3]. Empirical studies have further shown that occupancy prediction quality directly impacts HVAC control performance, linking forecast errors to measurable changes in comfort and energy outcomes in closed-loop settings [4]. Framework work on real-time occupancy-based HVAC integration addresses how occupancy intelligence can be operationalized through service interfaces and building management integration, moving beyond offline modeling toward deployable architectures [5]. Building comfort-focused analysis reinforces that occupancy prediction is not only an energy lever but also a comfort lever when used to avoid over-conditioning while maintaining acceptable comfort bands [6].

1.2. Distribution Shift and Research Objectives

The current state of practice is structurally brittle. Many occupancy and control studies evaluate models using random train-test splits or held-out samples drawn from the same distribution as the training data and then report performance that does not hold under real building operations. Real buildings are not stationary. Hybrid work patterns change occupancy periodicity and spatial distribution of buildings. Seasonal changes alter envelope loads, equipment behavior, and comfort preferences. Organizational restructuring changes zone usage and the meeting density. Sensor drift alters the relationship between physical reality and digital measurements. Equipment degradation changes the system response to control actions. Under these shifts, models can operate outside their competence boundaries without any explicit signal, which converts routine operational variability into systematic control errors.

The research problem is not limited to predictions. Control policies amplify prediction errors when they treat occupancy forecasts as deterministic inputs, and they can degrade rapidly if they rely on outdated building dynamics models. The human-in-the-loop HVAC literature shows that building operations are sociotechnical, with occupancy, comfort, and energy efficiency forming a coupled triad that requires structured interaction and oversight rather than blind automation [7]. The requirements and future directions for AI-based thermal comfort systems emphasize multi-objective control, limited labels, non-stationary preferences, and the need for robust operation under uncertainty and constraints [8]. Occupant-centric multi-zone HVAC control illustrates how learning-enabled strategies can coordinate zones in commercial buildings; however, operational robustness still depends on runtime reliability signals and safe fallback pathways when inference becomes unreliable [9]. Model predictive control (MPC) literature frames HVAC control as a constrained optimization problem and highlights opportunities for energy efficiency, but also shows that performance depends on model adequacy and correct constraint handling, which are stressed under distribution shifts and equipment degradation [10]. Reinforcement learning (RL) reviews for building controls describe opportunities and challenges, including generalization, sample efficiency, safety constraints, and operational integration barriers, which are intensified in non-stationary environments [11]. RL reviews for demand response emphasize algorithmic and modeling considerations for sequential decision-making under shifting conditions and external signals, again underscoring that robust deployment is more difficult than offline reward optimization [12]. Classic occupancy detection work using light, temperature, humidity, and CO₂ measurements shows that statistical learning can infer occupancy from environmental sensors; however, it also illustrates the fragility of learned relationships when sensors drift and building usage changes [13].

This study treats competence boundary management as a first-class requirement for intelligent building automation and energy optimization. It defines research questions as scoped deployment objectives, rather than rhetorical prompts. RQ1 targets a framework that connects AI-enabled automation to human-centered constraints, oversight, and runtime reliability. RQ2 targets the systematic evaluation of occupancy prediction and control performance under a distribution shift, including uncertainty quantification that reflects when predictions are unreliable. RQ3 targets drift detection methods suitable for building time-series data and graceful degradation strategies that preserve safe and acceptable performance when models drift or fail.

The contributions are direct in nature. A six-dimensional human-centered AI framework is defined for intelligent building automation, adapted to competence boundaries, drift, and safe degradation. A systematic gap identification is provided using the framework as a mapping tool across sensing, prediction, control, deployment, and oversight domains. The metrics-and-methodology section specifies shift-aware evaluation procedures, calibration and drift indicators, and closed-loop benchmarks that connect prediction reliability to energy, comfort, and equipment outcomes. Case analyses demonstrate how API-integrated occupancy-based HVAC, environmental sensor occupancy detection, occupant-centric multi-zone control, and RL-based building control can be redesigned to detect drift, quantify uncertainty, and degrade gracefully using MPC and conservative operational modes. Human cognitive bias effects are treated as operational failure modes in oversight and adoption, using cognitive bias analysis of fair AI design as a deployment lens for building automation [14]. The risk of overreliance on AI assistance is framed using evidence from AI coding assistants to emphasize competence boundaries and validation behaviors as generalized human-AI interaction requirements that transfer to building operations [15].

2. Background

2.1. CPS and IoT Sensing Infrastructure

Intelligent building automation combines IoT sensing and analytics with control strategies that incorporate learned models of occupancy and the dynamics of buildings. The literature is broad, but a coherent deployment view requires organizing the background around sensing pipelines, occupancy inference, control paradigms and operational integration.

IoT-based smart building infrastructures emphasize sensor data collection, processing, and analysis as prerequisites for adaptive control, with attention to the volume, velocity, and heterogeneity of building data streams [2]. This perspective correctly places data engineering as a control determinant. In buildings, sensing is often opportunistic rather than designed for identification. Many spaces have sparse occupancy sensors. Environmental sensors and equipment telemetry have become proxies for human presence and comfort. This creates a pipeline in which the measurement error and drift are normal rather than exceptional.

2.2. Occupancy Inference and Prediction for Control

Occupancy inference involves detection and prediction. Occupancy detection work has demonstrated that statistical learning models can infer office room occupancy from environmental measurements such as light, temperature, humidity, and CO₂ [13]. This result is operationally attractive because it uses sensors that are already present in many buildings. This is also a robustness warning. The relationship between CO₂ and occupancy depends on the ventilation settings, infiltration, sensor calibration, and occupant activity. The relationship between light and occupancy depends on daylight, blinds, and lighting. When these conditions shift, a model that performs well on held-out data can become systematically incorrect while producing plausible outputs.

Occupancy prediction literature consolidates modeling approaches and applications across energy efficiency, air quality, and comfort, documenting methods from classical machine learning to more complex time-series models [3]. The review landscape reflects strong performance under controlled evaluation but also reveals a recurring pattern: evaluation is often limited to in-distribution test splits rather than longitudinal deployment conditions. Impact studies show why this is important. When occupancy prediction is used to drive HVAC control, errors affect both comfort and energy, and the same prediction model can have different operational consequences depending on how the controller uses the forecast [4]. A comfort-focused analysis reinforces the coupling between occupancy prediction and HVAC control performance, positioning occupancy forecasting as a key mechanism for balancing comfort with energy optimization [6]. These studies establish a technical base that justifies occupancy-based control while exposing the missing layer of competence monitoring under drift.

2.3. Control and Integration Paradigms

The control paradigms for building automation include rule-based schedules, MPC, and RL. MPC literature frames building control as a constrained optimization, providing problem formulations, applications, and opportunities for improving HVAC energy efficiency [10]. MPC is relevant to robustness because it explicitly represents constraints and can incorporate uncertainty through robust or stochastic variants; however, its performance depends on model adequacy and the correct representation of dynamics and disturbances, which shift over time. RL for building controls provides a contrasting paradigm in which a policy is learned through interaction, promising adaptation to complex environments but facing challenges, including safety, generalization, and deployment constraints [11]. RL for demand response extends this view to sequential decision-making under external signals and variable price incentives, highlighting the modeling and algorithmic choices that determine stability and performance under changing conditions [12]. In practice, MPC and RL are often hybridized with learned predictors for disturbances such as occupancy, weather, and internal gains; thus, the distribution shift in predictors becomes a control stability risk rather than a pure forecasting issue.

Operational integration determines whether AI advances are deployable. An intelligent API framework for real-time occupancy-based HVAC integration addresses the integration architecture for serving occupancy intelligence and connecting it to smart building management systems [5]. This shifts the focus from isolated model performance to system behavior under latency, missing data, and integration constraints. The integration architecture also determines whether the uncertainty and drift signals can be propagated to the controllers and dashboards. If integration treats forecasts as point values and discards the metadata, competence boundaries cannot be enforced.

2.4. Human Oversight and Adoption Dynamics

Occupant-centric control and human-in-the-loop operation are essential in buildings because comfort is subjective and operational goals are multi-objective in nature. AI-based occupant-centric HVAC control for multi-zone commercial buildings illustrates the integration of occupant-centric objectives into coordinated control strategies [9]. Human-in-the-loop HVAC reviews provide a quantitative synthesis that positions occupancy, comfort, and energy efficiency as central dimensions and stresses that effective systems integrate human feedback and maintain oversight pathways rather than displacing operators and occupants [7]. AI for thermal comfort systems further identifies requirements and future directions, including the need to handle uncertainty, personalization, constraints, and robustness in real deployments [8]. These perspectives are directly connected to competence boundary management because a system that does not reveal its uncertainty cannot support appropriate human oversight, and a system that does not degrade safely under uncertainty will produce sharp failures that trigger rejection.

Human and organizational factors determine whether intelligent automation is adopted and trusted by organizations. Cognitive biases and fair AI design analysis show how human bias interacts with AI outputs and design choices, shaping fairness and trust outcomes in socio-technical systems [14]. This transfers directly to building operations, where operators may anchor on historical schedules, over-trust automation during abnormal events, or attribute discomfort to arbitrary causes without any diagnostic evidence. Evidence on AI coding assistants highlights productivity effects alongside the risks of over-reliance and the need for validation, illustrating how users can accept model outputs beyond competence boundaries without explicit reliability signals [15]. The building parallel is operationally direct: if a building automation system does not signal competence boundaries and uncertainty, operators can become over-reliant during drift and then reverse into blanket rejection after visible failures.

3. Framework: Six Dimensions of Human-Centered AI for Intelligent Building Automation

3.1. Framework Overview

Human-centered AI in building automation is not a user interface concept. It is a deployment discipline that connects machine intelligence to physical constraints, safety bounds, and human oversight under nonstationary conditions. A six-dimensional framework is defined here to structure intelligent building automation and energy optimization under distribution shift and competence boundary management.

3.2. Dimensions 1–2: Objectives, Constraints, and Data Integrity

The first dimension is operational purpose alignment within explicit constraints. Building automation optimizes multiple objectives, including thermal comfort, ventilation adequacy, energy cost, demand response participation, and equipment longevity. AI for thermal comfort systems emphasizes that comfort must be treated as a primary objective with explicit requirements and constraints, and not as a residual after energy minimization [8]. The MPC literature shows that constraints are not peripheral but central to safe and efficient control, with HVAC control framed as optimization under comfort and equipment constraints [10]. Human-centered purpose alignment requires that objective priorities be explicit in the control design so that safe fallback modes remain coherent when uncertainty increases and the system must degrade.

The second dimension is the data observability and integrity across the IoT pipeline. IoT frameworks for smart buildings emphasize large-scale sensor data collection and analytics [2]. In deployment, observability means that the system can detect when sensors drift, data streams fail, and measurement distributions change. Occupancy inference from environmental sensors depends on stable relationships among CO₂, ventilation, and occupancy that drift over time [13]. Human-centered deployment requires treating data integrity as a monitored variable rather than an assumption, because unobserved sensor drift converts model errors into silent control errors.

3.3. Dimensions 3–4: Uncertainty and Competence Boundaries

The third dimension is uncertainty quantification, which produces calibrated reliability signals. Occupancy prediction reviews have shown diverse methods applied to occupancy forecasting for energy and comfort applications [3]. Impact studies have shown that errors in occupancy prediction translate into changes in HVAC control performance [4]. Comfort-focused analysis reinforces occupancy prediction as a key control input [6]. These studies justify uncertainty quantification as a control determinant. Human-centered deployment requires calibrated uncertainty estimates that indicate when predictions are unreliable, so that the controller can choose conservative actions and operators can interpret outcomes correctly under drift.

The fourth dimension is competence boundary enforcement through drift detection and reliability gating. A distribution shift is normal in buildings because occupancy patterns, environmental loads, and equipment behavior change continuously. RL control reviews explicitly recognize the challenges of generalization, safety, and deployment in buildings, which are intensified in non-stationary environments [11]. Demand response RL reviews similarly stress modeling and algorithmic considerations under variable external conditions [12]. Competence boundary enforcement requires drift detection at the input distribution, residual, and closed-loop performance levels, with policies that gate reliance on learned components when drift signals exceed thresholds.

3.4. Dimensions 5–6: Oversight and Resilient Degradation

The fifth dimension is human oversight and interaction under predictable circumstances. The human-in-the-loop HVAC literature establishes that oversight and interaction mechanisms are central to operational success because occupants and operators remain stakeholders with legitimate authority and feedback roles [7]. Occupant-centric multi-zone HVAC control aims to center occupant outcomes in the control policy design [9]. Human-centered oversight requires that when the system detects drift or high uncertainty, it can surface intelligible signals to operators and execute predictable

fallbacks rather than oscillating behaviors that create distrust. Cognitive bias analysis is relevant because operators and occupants interpret system behavior through biased heuristics; transparent reliability signaling and stable fallback behavior reduce the space for misattribution [14].

The sixth dimension is resilience and graceful degradation in a cyber-physical environment. Intelligent building visions emphasize cyber-aware and human-interacting systems, which implies resilience to both faults and adversarial conditions [1]. Resilience requires safe fallback control modes that maintain acceptable comfort and ventilation while reducing the risk of failure of AI components. MPC provides a structured fallback mechanism because it explicitly encodes constraints and can operate with simpler disturbance models when predictive intelligence is unreliable [10]. Human-centered resilience also requires that degradation pathways be designed and tested under distribution shifts rather than improvised after failures.

This framework treats human-centered AI as a competence boundary management system embedded in the building automation. It links sensing integrity, uncertainty quantification, drift detection, oversight, and graceful degradation into a single, cohesive deployment specification.

4. Research Gaps: Systematic Gap Identification Under Distribution Shift

4.1. Gap Mapping Approach

The gap analysis follows a deployment-mapping method. The cited literature is mapped across the building automation pipeline stages of sensing and data engineering, occupancy inference, control policy design, integration and service, and oversight. Each stage was evaluated using the six framework dimensions. The result is a coherent set of gaps centered on distribution shifts and competence boundary management.

4.2. Robustness Evaluation and Competence Boundary Indicators

The first gap is the absence of a systematic robustness evaluation under real-world distribution shifts. Occupancy detection and prediction studies often demonstrate strong performance on held-out sets from the same data regime, including environmental sensor occupancy detection results that assume stable sensor relationships within the training environment [13]. Occupancy prediction reviews document methods and applications but do not establish a standard evaluation protocol that stresses models under longitudinal shifts, such as hybrid work, seasonal transitions, or space repurposing [3]. HVAC impact studies have shown that occupancy prediction errors are important in control, yet most evaluations still focus on in-distribution performance rather than explicitly quantifying degradation under shifting occupancy regimes [4]. Comfort-focused analysis reinforces the operational importance of occupancy prediction but does not resolve how to validate the robustness under changing building usage patterns [6]. The implications of this deployment are direct. An in-distribution accuracy claim does not establish safe control of drift.

The second gap is the lack of competence boundary indicators that translate prediction uncertainties into operational decisions. Current building automation deployments frequently treat occupancy prediction as point estimates. Impact work has shown that HVAC performance depends on prediction quality and the controller's use of predictions [4]. Without calibrated uncertainty and reliability gating, control policies can amplify errors under drift by aggressively preconditioning spaces based on unreliable forecasts. AI for thermal comfort systems explicitly highlights the importance of handling uncertainty and operational constraints [8]; however, the literature still lacks a standardized mechanism that converts uncertainty into conservative control actions and operator-visible reliability signals.

4.3. Drift Detection and Graceful Degradation

The third gap is drift detection for building practical and actionable time-series data. IoT smart building data streams are noisy, seasonal, and subject to missing data and sensor drift [2]. Building time-series contain strong daily and weekly patterns that can mask shifts if drift detection methods

do not account for seasonality. RL reviews for building controls highlight challenges, including generalization and safety, which implies the need for drift detection and runtime monitoring; however, drift detection is not yet treated as a standard component of RL building control deployments [11]. Demand response RL reviews similarly emphasize modeling choices and sequential decision challenges; however, systematic drift detection and safe mode switching remain underdeveloped in operational deployments [12]. API integration frameworks address real-time integration, but they rarely specify drift detection and competence boundary monitoring as part of the integration layer, which is the natural place to implement them across services [5].

The fourth gap is the lack of graceful degradation strategies that are formally designed, evaluated, and integrated. Building automation requires safe performance in the event of partial failure. When occupancy prediction becomes unreliable, a system requires a controlled fallback, such as schedule-based control, rule-based ventilation minimums, or MPC with conservative disturbances and tighter safety margins. The MPC literature provides a structured foundation for constraint-aware fallback [10]; however, the integration of learned occupancy predictors with MPC under drift is not standardized. The RL literature highlights opportunities and challenges related to safety and deployment, which makes graceful degradation essential for real-world acceptance [11]. Human-in-the-loop HVAC perspectives emphasize that operators must remain capable of oversight and intervention, which requires predictable degradation pathways and clear status signals [7]. This gap does not imply the absence of candidate fallback strategies. This gap is due to the absence of systematic design patterns that define when to switch, what to switch to, and how to evaluate performance and stability during switching.

4.4. Monitoring Linkages and Bias-Aware Oversight

The fifth gap is monitoring, which connects prediction drift to control drift. Many systems monitor prediction accuracy offline and energy outcomes online, but they do not connect them. HVAC impact studies have demonstrated that prediction errors have closed-loop consequences, implying that drift detection should include both feature distribution drift and performance drift signals tied to comfort and energy outcomes [4]. Comfort requirements work emphasizes multi-objective trade-offs and the need for robust operation, which implies monitoring at the level of objective satisfaction rather than only prediction metrics [8]. Without this connection, a system can meet energy targets while failing comfort objectives for specific zones during drift, or it can appear stable in energy use while accumulating comfort complaints that erode trust and trigger workarounds.

The sixth gap is the oversight design, which treats human cognitive bias as an operational risk. Cognitive bias analysis of fair AI design shows that human interpretation and decision-making can be predictably distorted, especially when AI outputs are presented without competence boundaries and uncertainty contexts [14]. Evidence from AI coding assistants shows that productivity tools can induce over-reliance unless validation behaviors are supported and competence boundaries are clear [15]. In building operations, analogous risk is automation complacency under drift, followed by sudden rejection after failures. Human-in-the-loop HVAC work establishes that oversight and interaction mechanisms are important [7], but competence boundary communication and bias-aware interface design remain underdeveloped in building automation deployments.

These gaps converge on a single requirement. Intelligent building automation must treat distribution shifts as default operating conditions and operationalize competence boundaries through drift detection, uncertainty quantification, reliability gating, and graceful degradation.

5. Evaluation Metrics and Methodology

5.1. Metrics for Prediction Reliability, Shift, and Control Outcomes

Occupancy inference metrics must separate the classification performance from probabilistic reliability. For occupancy detection and prediction tasks, standard measures such as accuracy, F1,

and balanced accuracy are useful when class imbalance is present. However, they did not measure whether the predicted probabilities reflected reality. Probabilistic metrics, such as the Brier score and negative log-likelihood, quantify the quality of probabilistic outputs, and calibration measures quantify whether confidence aligns with empirical frequency. These probabilistic measures are critical because occupancy prediction is used as an input to controllers that decide whether to pre-condition, ventilate, or dim lighting, and HVAC impact evidence shows that errors translate into changes in control performance [4]. Occupancy prediction reviews justify the need to treat occupancy as a time-varying stochastic process rather than a deterministic signal, reinforcing the relevance of probabilistic evaluation [3]. Environmental sensor occupancy detection further motivates reliability evaluation because sensor drift can maintain apparent accuracy while degrading probability quality, especially when thresholds are tuned to historical conditions [13].

Distribution shift metrics must characterize how operating data diverge from the training data. Input drift can be quantified using divergence measures between feature distributions computed over sliding windows after seasonal adjustment. Drift in time-series data should be evaluated after accounting for expected periodicity, using residuals from seasonal baselines to avoid flagging normal weekly patterns as drift. Residual drift metrics quantify the change in model error distribution over time, which is more operationally relevant than raw feature drift when control depends on prediction error. Concept drift can be inferred when the mapping from features to occupancy changes, which can occur when ventilation settings change, occupancy becomes more meeting-driven, or sensor coverage changes. The drift detection performance is evaluated through the detection delay, false alarm rate, and missed detection rate under controlled shift injections.

Uncertainty quantification metrics must evaluate both the calibration and sharpness. Calibration measures whether the predicted probabilities match the observed frequencies, whereas sharpness measures whether the uncertainty intervals are informative rather than trivially wide. In occupancy-based HVAC control, uncertainty must be informative enough to trigger conservative control only when needed, because over-conservative gating reduces energy benefits and user acceptance, whereas under-conservative gating causes comfort failures during drift. AI for thermal comfort systems emphasizes that robust performance must respect constraints and uncertainty, reinforcing that uncertainty quantification must be evaluated as an operational signal rather than a modeling accessory [8]. RL control reviews highlight safety and generalization issues, reinforcing that uncertainty and competence boundaries must be measured and enforced, especially when policies are adapted over time [11].

The control performance metrics must combine energy, comfort, air quality, and equipment stress. Energy metrics include HVAC and lighting energy consumption normalized by weather and occupancy exposure, and demand metrics include peak demand and demand response compliance, where relevant. Comfort metrics include time outside defined comfort bands, temperature stability quantified by overshoot and rate of change, and complaint rates aligned to zones and operational modes, consistent with the human-in-the-loop HVAC emphasis on comfort outcomes as primary operational signals [7]. Air quality metrics include time above CO₂ thresholds and response latency after occupancy surges, which align with occupancy prediction applications in ventilation [3]. Equipment stress metrics include short cycling frequency, actuator saturation duration, and constraint-binding frequency, which are relevant under drift because degraded equipment changes the system response to control actions. The MPC literature motivates explicit constraint handling and provides a basis for evaluating constraint violations and robustness under model errors [10].

Graceful degradation metrics must quantify safe performance in the presence of AI unreliability. The key measure is the performance retention under drift, defined as the ratio between the degraded-mode objective satisfaction and nominal-mode objective satisfaction under the same disturbance regime. Degradation should preserve safety and acceptability, measured as bounded comfort violations, maintained minimum ventilation, and reduced equipment stress, while accepting some loss of energy. The human-in-the-loop HVAC literature implies that acceptable degradation must be

predictable and interpretable to operators, which makes the stability of mode switching and clarity of the system status part of the evaluation [7].

5.2. Shift-Aware Evaluation Methodology and Gating Calibration

The methodology is staged to connect the modeling evidence to the operational evidence. Offline evaluation establishes the baseline predictive performance and probabilistic reliability under in-distribution conditions and synthetic or historical shift scenarios. This includes time-aware validation rather than random splits, using rolling-origin evaluation that respects temporal causality and produces performance trajectories across seasons and occupancy periods. Occupancy prediction reviews support this shift-aware framing by emphasizing the temporal structure and application contexts in which occupancy patterns vary [3]. HVAC impact studies justify the coupling of predictive evaluation to control outcomes because prediction errors affect HVAC performance in closed-loop operations [4]. Environmental sensor occupancy detection motivates explicit shift scenarios involving sensor drift and ventilation changes because these factors directly alter feature distributions [13].

Closed-loop evaluation uses simulation or digital twin environments to test controllers under controlled distribution shifts, sensor drift, and equipment degradation scenarios. The objective is not to prove a single controller optimal but to establish robustness envelopes and validate competence boundary policies. Controllers were evaluated with and without uncertainty gating and drift-triggered fallback to quantify the robustness of runtime monitoring versus baseline control design. MPC provides a structured baseline and fallback mode because it explicitly encodes constraints and can operate under conservative disturbance assumptions when predictive intelligence is unreliable [10]. RL controllers are evaluated under shifts with explicit safety constraints and mode-switching policies, consistent with the RL deployment challenges identified in building controls and demand response reviews [11,12].

Field evaluation was executed as a monitored pilot in limited zones with strict logging of sensor inputs, model outputs, uncertainty estimates, drift indicators, control actions, and mode-switch events. The field design treats distribution shifts as expected events, using real operational changes such as schedule updates, seasonal transitions, and partial sensor outages as natural experiments. The field evaluation focused on the detection performance, stability of mode switching, objective satisfaction under drift, and operator intervention rates. The integration layer is instrumented to preserve metadata and enable competence boundary enforcement, aligning with the requirements of a real-time occupancy-based HVAC integration architecture [5]. Human-in-the-loop HVAC principles constrain field operations by requiring predictable recourse and maintainable oversight, especially during conservative fallback operations [7].

The evaluation logic was direct. A system that demonstrates high in-distribution accuracy but fails to detect drift and degrade safely is not deployable. A system that detects drift but over-triggers the safe mode is also not deployable because it eliminates energy benefits and encourages operator disablement. The target is calibrated detection and calibrated uncertainty, paired with controlled degradation that maintains safe, acceptable comfort, and ventilation.

6. Case Analysis: Illustrative Applications

6.1. HVAC Automation and Occupancy Inference Deployments

The case analysis demonstrates how competence boundary management becomes concrete in intelligent building automation. Each case is presented as a deployment pattern, the shift and drift risks it faces, and the reliability and degradation mechanisms required to maintain an acceptable performance.

The first case is API-integrated occupancy-based HVAC automation. Real-time occupancy-based HVAC integration frameworks demonstrate how occupancy intelligence can be served through APIs and connected to building management systems, enabling dynamic HVAC

adjustments based on inferred occupancy [5]. This architecture is a natural insertion point for competence boundary enforcement because it centralizes the inference outputs and control commands. Under hybrid work patterns and organizational restructuring, occupancy periodicity changes and meeting-driven occupancy have become less predictable. Under sensor drift and data outages, the feature distributions and missing data patterns change. A robust implementation treats occupancy outputs as probabilistic, propagates uncertainty metadata through the API, and couples drift indicators to the mode-switch logic. When drift is detected or uncertainty rises beyond a calibrated threshold, the system switches from forecast-driven pre-conditioning to conservative schedule-based control or MPC with conservative disturbances, preserving minimum ventilation and bounded comfort outcomes. HVAC impact evidence supports this gating logic by showing that occupancy prediction errors affect control outcomes, making prediction reliability a control variable rather than a modeling detail [4]. Comfort-focused analysis supports the same mechanism by framing occupancy prediction as a comfort lever that must be used responsibly in the face of uncertainty [6]. The evaluation objective was to demonstrate that drift detection triggers the safe mode before comfort violations accumulate and that the safe mode avoids excessive energy waste while preserving acceptable indoor conditions.

The second case is occupancy detection and prediction using environmental sensors in a single zone or small office. Environmental sensor occupancy detection demonstrates strong inference using light, temperature, humidity, and CO₂ measurements [13]. In deployment, this pattern is attractive because it uses existing sensors and reduces privacy risks compared to camera-based sensing. The robustness hazard is the systematic drift in CO₂ sensors, changes in ventilation rates, and changes in daylight and lighting schedules. These shifts alter the mapping between environmental features and occupancy, producing a concept drift even when the sensor streams appear intact. Competence boundary management relies on monitoring feature distributions after seasonal adjustment, monitoring residual drift using periodic ground truth checks where feasible, and monitoring physical consistency constraints, such as CO₂ rise rates relative to ventilation settings. When drift signals rise, the system reduces its reliance on environmental occupancy inference for control and defaults to conservative ventilation minimums and temperature setpoints. This case illustrates that drift detection must be tailored to building time-series periodicity because daily cycles can mask drift if naive thresholds are used.

6.2. Multi-Zone and Reinforcement-Learning Control Under Non-Stationarity

The third case is the multi-zone occupant-centric HVAC control in commercial buildings. Occupant-centric multi-zone control aims to optimize HVAC decisions based on occupant outcomes while coordinating zones under shared constraints [9]. The distribution shift includes floor-level reassignments, changes in meeting space usage, and changes in equipment response as systems age. The risk is uneven degradation, where some zones become systematically misconditioned owing to sensing gaps or drift in zone-level models, and the controller continues to optimize aggregate energy metrics while comfort failures concentrate in specific areas. Human-in-the-loop HVAC framing requires that comfort and energy trade-offs be monitored as coupled outcomes and that operators retain meaningful oversight [7]. Competence boundary management for multi-zone systems requires zone-specific drift indicators rather than building-level aggregates because drift can be localized to sensors or equipment in specific zones. Under detected drift, the controller degrades locally by tightening comfort constraints, widening safety margins, and reverting to simpler control logic for affected zones while maintaining coordination through constraint-aware MPC at the system level. MPC's explicit constraint framework of MPC supports safe local degradation while preserving system-level feasibility [10]. This case also highlights the need to treat operator intervention rates and complaint trajectories as drift-related signals rather than purely human behavioral noise.

The fourth case is reinforcement learning for building control and demand response under non-stationary conditions. RL reviews for building controls emphasize opportunities for adaptive policies but also identify challenges, including safety constraints, generalization, and operational integration

[11]. Demand response RL reviews emphasize the complexities of sequential decisions under external signals and changing conditions [12]. In this context, distribution shift includes changes in occupancy regimes, tariff structures, equipment efficiency, and sensor drift. The core hazard is policy brittleness masked by high confidence in the learned value estimates, leading to unsafe or inefficient actions during drift. Competence boundary management for RL requires runtime monitoring of state distribution drift, reward and constraint violation statistics, and conservative action constraints that prevent unsafe exploration. When drift indicators rise, the RL policy execution is gated, and the control falls back to the MPC or rule-based demand response strategies that encode hard constraints explicitly [10]. This hybrid pattern treats the RL as an optimization layer when confidence is warranted and treats the MPC as a safety layer when uncertainty increases. The evaluation focuses on constraint satisfaction under shifts, stability of mode switching, and the ability to maintain demand response obligations without inducing comfort violations during abnormal occupancy conditions.

6.3. Pipeline View of Comfort Optimization

The fifth case is real-time occupancy-based comfort optimization, framed as a deployment pipeline rather than a single model. Occupancy-based comfort analysis emphasizes that occupancy prediction influences comfort outcomes when integrated into HVAC control [6]. API-based integration shows how occupancy intelligence can be connected to building management systems [5]. Therefore, the deployment pipeline must include explicit reliability outputs, drift detection, and graceful degradation; otherwise, comfort optimization becomes fragile. This case demonstrates that robust automation is a system behavior property. This is not achieved by selecting a single high-performing occupancy model from a held-out test set. This is achieved by instrumenting the pipeline so that the system can detect when the world has changed and respond with controlled degradation that keeps occupants safe and reasonably comfortable.

7. Discussion

7.1. Implications for Competence-Aware Building Automation

Intelligent building automation and energy optimization require a shift in the definition of evidence. In-distribution prediction accuracy is necessary but insufficient because buildings are non-stationary environments, and distribution shifts are routine rather than exceptional. The correct deployment target is competence-aware automation that continuously evaluates whether the models remain valid and enforces safe behavior when they do not. IoT infrastructures increase sensing coverage and data volume, but they also increase drift surfaces and missing data modes, making observability and data integrity monitoring a first-order control requirement rather than a data engineering preference [2]. Occupancy modeling literature and HVAC impact evidence show that prediction errors propagate into closed-loop outcomes, which means that uncertainty and drift signals must be part of the control design, not an analytics add-on [3,4]. MPC provides a structured baseline and fallback because it explicitly encodes constraints and can preserve safety and feasibility even when disturbance predictors degrade [10]. RL offers adaptive opportunities but faces generalization and safety constraints, which makes competence boundary enforcement and fallback to constraint-based control deployment requirements [11,12]. Human-in-the-loop HVAC perspectives set a hard constraint on adoption: operators and occupants require predictable behavior under degradation and intelligible indicators of system reliability; otherwise, automation produces over-reliance during drift, followed by rejection after failures [7]. Cognitive bias analysis strengthens this point by treating the misinterpretation of AI outputs as a predictable failure mode that must be mitigated by explicit competence boundary signaling and stable degradation behavior [14]. Evidence from AI coding assistants reinforces that competence boundary ambiguity causes over-reliance unless validation behaviors are structurally supported, which transfers to building operations as a requirement for drift-aware oversight and verification mechanisms [15].

7.2. Limitations and Operational Constraints

However, the limitations of this study are that ground truth data for occupancy and comfort are sparse in many buildings, which constrains calibration and concept drift validation, even when proxy sensing is available [3,13]. Sensor drift and data outages are difficult to separate from genuine occupancy changes without independent measurements, which increases the risk of both false alarms and missed drift detection. Buildings contain strong periodic patterns that complicate drift detection in time-series because seasonality can mimic drift if not explicitly modeled. Integration constraints and legacy building management systems can limit data retention, metadata propagation, and runtime computation, which restrict the sophistication of monitoring and uncertainty models that can be deployed within operational constraints [5]. Multi-objective trade-offs remain unavoidable. Conservative degradation reduces the risk but can reduce energy savings and demand response performance; therefore, gating thresholds must be calibrated with explicit acceptance criteria rather than tuned for maximum detection sensitivity. Organizational context also remains a confounding factor. Complaint rates and operator interventions reflect both building conditions and organizational norms; therefore, human-centered monitoring must interpret these signals within the operational context rather than treating them as purely objective ground truths. These limitations do not negate the deployment requirements. They define the engineering constraints within which competent boundary management must operate.

8. Conclusion

8.1. Summary

Intelligent building automation and energy optimization depend on IoT sensing and AI control; however, real buildings operate under a continuous distribution shift from hybrid work, seasonal dynamics, sensor drift, and equipment degradation. The resulting competence boundary problem is the dominant deployment risk: models that perform well in distribution can become confidently incorrect in operation, and controllers can amplify prediction errors into comfort, air quality, and equipment stress failures. A six-dimensional human-centered AI framework structures the solution by binding objective alignment, data observability, uncertainty quantification, drift detection, human oversight, and resilience into a deployable specification. The systematic gap analysis identifies the missing robustness evaluation under shift, missing competence boundary indicators that gate control behavior, underdeveloped drift detection methods tailored to building time-series periodicity, and weak graceful degradation patterns that preserve acceptable performance when AI components become unreliable. A shift-aware evaluation methodology and metric set connect prediction reliability to closed-loop outcomes and establish evidence requirements for real deployment. Case analyses show that API-integrated occupancy-based HVAC, environmental-sensor occupancy inference, multi-zone occupant-centric control, and RL-based control become deployable only when uncertainty and drift signals are monitored, acted upon, and paired with safe fallback to constraint-based MPC and conservative operation modes.

References

1. M. Manic, K. Amarasinghe, J. Rodriguez-Andina, and C. G. Rieger, "Intelligent Buildings of the Future: Cyberaware, Deep Learning Powered, and Human Interacting," *IEEE Industrial Electronics Magazine*, vol. 10, no. 4, pp. 32–49, 2016, doi: 10.1109/MIE.2016.2615575.
2. A. P. Plageras, K. E. Psannis, C. Stergiou, H. Wang, and B. B. Gupta, "Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings," *Future Generation Computer Systems*, vol. 82, pp. 349–357, 2018, doi: 10.1016/j.future.2017.09.082.
3. Z. Chen, C. Jiang, and L. Xie, "A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment," *Renewable and Sustainable Energy Reviews*, vol. 167, Art. no. 112704, 2022, doi: 10.1016/j.rser.2022.112704.

4. M. Esrafilian-Najafabadi and F. Haghighat, "Impact of occupancy prediction models on building HVAC control system performance: Application of machine learning techniques," *Energy and Buildings*, vol. 257, Art. no. 111808, 2022, doi: 10.1016/j.enbuild.2021.111808.
5. S. Adepoju and S. David, "An Intelligent API Framework for Real-time Occupancy-Based HVAC Integration in Smart Building Management Systems," *Journal of Knowledge Learning and Science Technology*, vol. 4, no. 1, pp. 63–69, 2025, doi: 10.60087/jklst.v4.n1.007.
6. S. A. Adepoju, "How machine learning can revolutionize building comfort: Accessing the impact of occupancy prediction models on HVAC control system," *World Journal of Advanced Research and Reviews*, vol. 25, no. 1, pp. 2315–2327, 2025, doi: 10.30574/wjarr.2025.25.1.0161.
7. W. Jung and F. Jazizadeh, "Human-in-the-loop HVAC operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions," *Applied Energy*, vol. 239, pp. 1471–1508, 2019, doi: 10.1016/j.apenergy.2019.01.070.
8. A. Ghahramani, P. Galicia, D. Lehrer, Z. Varghese, Z. Wang, and Y. Pandit, "Artificial Intelligence for Efficient Thermal Comfort Systems: Requirements, Current Applications and Future Directions," *Frontiers in Built Environment*, vol. 6, Art. no. 49, 2020, doi: 10.3389/fbuil.2020.00049.
9. A. Dhumane, S. Chiwhane, S. Thakur, P. Khanna, S. Raghuvanshi, and S. Kautish, "Artificial Intelligence (AI)-Based Occupant-Centric Heating Ventilation and Air Conditioning (HVAC) Control System for Multi-Zone Commercial Buildings," *Sustainability*, vol. 14, no. 23, Art. no. 16107, 2022, doi: 10.3390/su142316107.
10. G. Serale, M. Fiorentini, A. Capozzoli, D. Bernardini, and A. Bemporad, "Model Predictive Control (MPC) for Enhancing Building and HVAC System Energy Efficiency: Problem Formulation, Applications and Opportunities," *Energies*, vol. 11, no. 3, Art. no. 631, 2018, doi: 10.3390/en11030631.
11. Z. Wang and T. Hong, "Reinforcement learning for building controls: The opportunities and challenges," *Applied Energy*, vol. 269, Art. no. 115036, 2020, doi: 10.1016/j.apenergy.2020.115036.
12. J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Applied Energy*, vol. 235, pp. 1072–1089, 2019, doi: 10.1016/j.apenergy.2018.11.002.
13. L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models," *Energy and Buildings*, vol. 112, pp. 28–39, 2016, doi: 10.1016/j.enbuild.2015.11.071.
14. S. Adepoju and M. Adepoju, "Cognitive Biases: Understanding and Designing Fair AI Systems for Software Development," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 9, p. 6, 2025, doi: 10.14569/IJACSA.2025.0160901.
15. S. Adepoju, "GitHub Copilot's Impact on Developer Productivity: A Review of Early Evidence," *International Journal of Scientific Research in Science and Technology*, vol. 10, no. 4, pp. 814–822, 2023, doi: 10.32628/IJSRST2221192.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.