

Iurii Zhurov

IT specialist

SMB Cybersecurity Advisors, United States of America

ORCID: 0000-0002-8276-2230

E-mail: iurii.zhurov@smbsecurityadvisors.com

A METHOD FOR FORMING A SOCIAL ENGINEER PROFILE BASED ON A DETECTED PHISHING ATTACK ON AN INFORMATION SYSTEM

Phishing attacks remain one of the most widespread cyber threat vectors, as they combine technical mechanisms for compromising information systems with social engineering techniques. At the same time, existing phishing detection approaches are primarily focused on analyzing the technical parameters of an attack and do not provide a formalized transition from detection results to identifying the characteristics of the attacker, which complicates cyberattack attribution, cyber threat intelligence, and information security incident response. The aim of this study is to develop a method for constructing a social engineer profile based on the analysis of a detected phishing attack on an information system. The proposed method is based on extending a generalized tuple by integrating a set of attacker profile features formed according to the set-theoretic classification model of modern social engineering attack implementation approaches and the parameters of the detected phishing attack. The method consists of four main stages: constructing a formalized representation of the phishing attack, determining the characteristics of the social engineer profile, constructing an integrated social engineer profile, and visualizing the resulting profile. Unlike existing approaches, the proposed method provides a formalized transition from the parameters of a detected phishing attack to the construction of an integrated profile of its perpetrator. The proposed method was validated using a representative phishing attack, resulting in the construction of a social engineer profile, identification of its key characteristics, and generation of a diagram that provides a graphical representation of the resulting profile. The obtained results can be used to support cyberattack attribution, cyber threat intelligence, information security incident response, and the development of countermeasures against social engineering attacks.

Keywords: phishing, social engineer profile, attacker profile, social engineering, cyber threat intelligence, information security.

Надійшла до редакції (Received): 10.07.2026

Прийнята до друку (Accepted): 08.09.2026

Опубліковано онлайн (Available online): 15.09.2026

<http://creativecommons.org/licenses/by/4.0/>

This work is licensed under Creative Commons Attribution-noncommercial-sharealike 4.0 International License

UDC 004.896:620.9:697

DOI: 10.31673/2409-7292.2026.034410

Andrii Verlan

Doctor of Technical Sciences, Professor of the Department of IPZE NN IATE

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine,

Norwegian University of Natural Sciences and Engineering, Trondheim, Norway

ORCID: 0000-0002-6469-2638

E-mail: a.verlan@edu.kpi.ua

Andrii Dyachuk

Postgraduate student of the Department of IPZE NN IATE

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

ORCID: 0009-0003-6979-278X

E-mail: diachuk.a.o.-tv51f@edu.kpi.ua

GENERATIVE NEURAL NETWORK APPROACHES FOR ADAPTIVE BUILDING ENERGY FORECASTING AND CONTROL: A REVIEW OF METHODS AND MICROSERVICES ARCHITECTURES

This article presents a comprehensive review of the transition from static, rule-based Building Management Systems to data-driven artificial intelligence for optimizing energy consumption in the built environment. While discriminative neural networks have advanced energy forecasting, their practical deployment is frequently hindered by

regression model overfitting on constrained and noisy sensor datasets. The objective of this study is to evaluate the integration of generative neural networks into adaptive building energy forecasting and control, aiming to resolve the data and operational limitations of conventional machine learning. To achieve this goal, the specific tasks undertaken include synthesizing the methodological shift from classical sequence modelling to generative data augmentation, evaluating Model-Based Deep Reinforcement Learning (DRL) frameworks for autonomous HVAC control, and examining the decentralized microservices architectures necessary for edge-cloud deployment. The analysis indicates that generative models, such as TimeGAN, synthesize training data to overcome empirical data scarcity, thereby improving the sample efficiency of DRL agents and preventing equipment malfunction during exploratory learning. Additionally, hybrid control frameworks like SACEM and Generative Model Predictive Control (GMPC) maintain energy reduction targets alongside human thermal comfort in non-stationary environments. Ultimately, translating these computational advancements into physical building operations mandates a structural transition to containerized microservices. Specifically, the implementation of Temporal Domain Separation is essential to isolate the inference latency of generative algorithms from the sub-100 millisecond deterministic safety constraints of physical HVAC machinery, ensuring secure and adaptive facility management.

Keywords: generative neural networks, building energy management, smart buildings, deep reinforcement learning, HVAC, microservices architectures, LSTM, transformers.

Introduction

The built environment stands as a significant contributor to global energy consumption and anthropogenic greenhouse gas emissions. Historically, the management of electrical and thermal loads within commercial and residential facilities has relied heavily upon Building Management Systems (BMS) governed by static, rule-based controls (RBC) and first-principle thermodynamic models [1]. These conventional automation systems are relatively effective for nominal, steady-state operations within strictly defined parameters. However, they are challenged by the escalating complexity of the modern built environment. Modern smart buildings operate within highly non-stationary environments characterized by stochastic occupant behaviours, weather fluctuations, and the unpredictable integration of renewable energy resources [2]. The requirement to constantly recalibrate physical mathematical models using expensive domain expertise whenever these boundary conditions shift renders traditional RBC brittle, inefficient, and difficult to scale across diverse urban infrastructures.

To resolve these systemic limitations, the paradigm of building energy management has definitively shifted toward data-driven Artificial Intelligence (AI). Artificial Neural Networks (ANNs) have progressively established themselves as the computational core of modern smart building automation, functioning autonomously across the entire cyber-physical stack [3]. Early iterations of this transition utilized discriminative deep learning architectures such as Long Short-Term Memory (LSTM) networks and transformers to identify patterns in historical telemetry and forecast future energy loads [4]. However, as the demand for fully autonomous, adaptive control has intensified, the field has evolved beyond isolated forecasting.

The current frontier of building energy management is situated at the intersection of generative neural networks, model-based deep reinforcement learning (DRL), and distributed microservices architectures [5]. Modern smart buildings are adopting cloud-edge collaborative computing frameworks governed by decentralized microservices. This architectural transition ensures that safety-critical, deterministic control loops operate with sub-100 millisecond latency at the edge, while heavier generative AI tasks are processed asynchronously in the cloud [6].

This article proposes a structured evaluation of how generative neural networks and distributed microservices architectures can be integrated to advance adaptive building energy management. The specific tasks executed to achieve this include synthesizing the methodological progression from classical sequence modelling to generative data augmentation, assessing the application of model-based deep reinforcement learning for autonomous HVAC control, and examining the decentralized edge-cloud computing infrastructures necessary for practical deployment.

The main contribution of this review lies in bridging the theoretical developments of generative artificial intelligence with their physical implementation. By providing an exhaustive analysis of computing frameworks, this work demonstrates how modern architectures facilitate real-time

inference, ensuring that complex computational tasks do not compromise the deterministic safety constraints of the built environment.

Building energy consumption prediction models

1. Theoretical foundations of building energy forecasting

Before adaptive control can be achieved, a smart building must have the computational capacity to accurately predict its future energy states. Building energy consumption prediction models serve as the foundational tools for optimizing energy management across the entire lifecycle of a facility, covering the initial design, ongoing operation, and subsequent retrofitting phases [3].

A useful starting point is the review by Runge and Zmeureanu [7], which synthesized studies from 2000 to early 2019 on artificial neural networks for building energy forecasting. That review showed a field dominated by black-box ANN models, especially feed-forward neural networks, short-term forecasting horizons, whole-building load targets, and commercial-building applications. It also highlighted several gaps, including limited use of recurrent neural networks, grey-box models, ensemble methods, and richer control-oriented forecasting tasks.

There is a more recent review that looks back across the ANN tradition from a broader perspective. El Alaoui and Rougui [8] describe ANN-based building energy prediction as a long-running line of research centered on the ability of neural networks to capture nonlinear relationships between historical consumption, environmental conditions, and operational variables. Their review spans studies published between 2001 and 2023 and classifies the main ANN families used in building energy research, including feedforward neural networks (FFNN), radial basis neural networks (RBNN), generalized regression neural networks (GRNN), and nonlinear autoregressive neural networks (NARNN). They argue that this diversity of ANN types reflects the range of building-energy tasks, from general regression and approximation to time-series forecasting with temporal dependencies.

2. Shift from classical ANN to broader sequence modelling

Later literature shows a clear broadening of methodological scope. It did not invalidate this earlier body of work. Instead, it expanded it through newer deep learning architectures, uncertainty-aware forecasting, transfer learning, and control-integrated approaches.

Liu et al.'s review of data-driven building energy prediction covered 116 studies and described a field that now includes a wider range of machine learning and deep learning methods, differentiated by data type, time scale, and application objective [9]. The review emphasizes that model choice depends on forecasting horizon, data quality, feature availability, and operational goals, rather than on a single universally best architecture.

As a result, the main methodological evolution is not merely the rise of a new architecture, but a shift toward sequence-aware, uncertainty-aware, and control-aware modelling. This includes recurrent networks, attention-augmented recurrent networks, temporal fusion transformers, patch-based transformers, adaptive MPC, and reinforcement-learning-enhanced control.

To evaluate the performance of these models, standard metrics such as Mean Absolute Error (MAE) and the coefficient of determination (R^2) are commonly used.

MAE measures the average magnitude of prediction errors:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (1)$$

where y_i and $\hat{y}_i$ denote the actual and predicted energy consumption, respectively.

R^2 quantifies the proportion of variance in the observed data explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (2)$$

with $\bar{y}$ representing the mean of the true values.

Together, these formulas provide a formal basis for understanding both the internal mechanics of the models and their effectiveness in predicting and controlling energy usage in buildings.

3. Continued relevance of recurrent models

Recurrent neural networks remain strongly represented in the energy literature, especially LSTM [9]. Their continued relevance is tied to the nature of building data: energy use, thermal states, occupancy effects, and weather-driven loads are fundamentally sequential processes with temporal dependencies, delayed effects, and recurring patterns. LSTM remains attractive because it was designed specifically to model such dependencies while mitigating the short-memory limitations of simpler recurrent networks.

Recent work confirms that LSTM is still a competitive model for building energy prediction. Mishra et al. (2024) developed an LSTM-based approach using historical consumption, environmental variables, and occupancy-related information for residential and commercial buildings [10]. Their results reported an average R^2 of 0.91 and average MAE of 0.013, with best results reaching $R^2 = 0.97$ and $MAE = 0.007$. The study found the LSTM model to outperform linear regression, decision trees, and random forest in the reported experiments. The authors also argue that the approach remains effective even when data are relatively limited.

4. Evolution of LSTM: attention, uncertainty, and optimization

While LSTM remains relevant, the contemporary literature [11] rarely treats plain LSTM as the methodological endpoint. Instead, LSTM is often improved through attention mechanisms, uncertainty modelling, and automated tuning.

Li et al. (2022) proposed an attention-based LSTM with fuzzy information granulation for interval building energy forecasting [12]. Their contribution is important because it moves beyond point prediction and explicitly addresses uncertainty, which is highly relevant for adaptive control. The study shows that building energy management benefits not only from mean forecasts, but also from interval forecasts that provide a representation of future uncertainty. It also reports that the attention-based LSTM outperformed conventional LSTM in interval forecasting.

For models that incorporate attention, such as Attention-LSTM, the network assigns different importance to previous timesteps to improve long-range dependency modelling.

The attention weight α_t for timestep t is calculated as:

$$\alpha_t = \frac{\exp(e_t)}{\sum_k \exp(e_k)}, \quad e_t = \text{score}(h_t, s), \quad (3)$$

where h_t is the LSTM hidden state and s is a context vector or query.

The attention mechanism enables the model to focus selectively on relevant past information, enhancing forecast accuracy for multi-step or irregular energy consumption patterns.

A later study on BO-STA-LSTM further illustrates this trajectory by combining LSTM with Bayesian optimization and spatial-temporal attention [13]. According to the article summary, the proposed model improved average prediction performance by 0.0885 in R^2 , with Bayesian optimization contributing 0.0717 and attention contributing 0.0560, and the final method outperformed CNN and TCN under the study conditions. This suggests that the recurrent family remains competitive when enhanced with contemporary mechanisms rather than used in its earlier plain form.

Taken together, these studies suggest that LSTM remains a robust sequence-modelling foundation, and modern variants strengthen it through attention and uncertainty handling.

5. Rise of transformers in building energy forecasting

The strongest challenge to recurrent models in recent years comes from transformer-based architectures. Originally developed for natural language processing, transformers have increasingly been adapted to time-series forecasting, including building energy applications. Their appeal lies in their ability to model long-range dependencies, process sequences in parallel, and support transfer learning across related datasets.

Zheng et al. (2023) applied Temporal Fusion Transformer to building energy consumption forecasting and emphasized both accuracy and interpretability [14]. This is notable because interpretability is often a concern in deep-learning-driven building operation. TFT and related architectures are therefore attractive not only as forecasters, but as models that can provide some insight into variable importance and temporal relevance.

Spencer et al. (2025) compared vanilla Transformer, Informer, and PatchTST across 16 Building Data Genome Project 2 datasets under transfer-learning settings [15]. The study reports that transfer learning was generally beneficial and that PatchTST outperformed the other transformer variants examined. This is substantial evidence that transformer-based models have become highly competitive, especially in settings involving multiple buildings, transfer across domains, and richer data environments. Therefore, in pure forecasting terms, the literature does show a clear movement toward transformers. However, that movement should be interpreted carefully. It demonstrates [16] that transformers are strong candidates for building energy forecasting, not that they are automatically the best choice for adaptive energy control as an integrated research problem. Hybrid models that combine the advantages of LSTM and transformer models are proposed as a step in solving this problem.

6 Discriminative deep learning frameworks and their limitations

The application of deep learning in this domain has historically focused on identifying the optimal algorithmic balance between prediction accuracy, computational speed, and mathematical complexity. Various neural network architectures demonstrate distinct strengths depending heavily on the temporal horizon of the forecast [4].

Table 1 presents a summary comparison of key sequence-model architectures used in building energy forecasting and their relevance for adaptive control. It highlights the core idea, main strengths and limitations, typical applications, and representative performance metrics for each model, including LSTM, attention-enhanced LSTM, BO-STA-LSTM, and transformer-based approaches such as TFT and PatchTST. The table allows for a quick overview of how these models differ in terms of accuracy, complexity, and suitability for integration into adaptive energy management systems.

Table 1

Comparison of sequence models for building energy forecasting and adaptive control

Model	Strengths	Weaknesses / Limitations	Typical applications	Reported metrics
LSTM	Good for time-series; stable for moderate datasets; compatible with control loops	Can struggle with long sequences; less accurate for multi-horizon	Short-/medium-term load forecasting; adaptive MPC; scenario generation	Avg $R^2 = 0.91$, Avg MAE = 0.013; best-case $R^2 = 0.97$, MAE = 0.007 [10]
Attention-LSTM	Captures long-range dependencies; interval forecasting; better multi-step predictions	Higher computational cost	Interval forecasting; uncertainty-aware energy prediction; scenario generation	MAPE improvement over plain LSTM ~3–5%; better interval coverage [12]
BO-STA-LSTM	Accounts for spatiotemporal correlations; outperforms CNN/TCN	Complex; tuning required	Multi-step forecasting; adaptive control integration	Avg R^2 improvement = +0.0885 [13]
Temporal Fusion Transformer (TFT)	Handles long sequences; multi-horizon; interpretable; transfer learning	Computationally heavy; larger data required	Multi-horizon forecasting; transfer learning; interpretability	Avg MAE reduction ~15–20% vs LSTM; multi-horizon R^2 often >0.9 [14]
PatchTST / Transformer variants	Strong forecasting; long-term and transfer learning	Very high computational cost; less direct for real-time control	Multi-building forecasting; long-term scenario generation; benchmark studies	MAE reduction ~16% vs vanilla Transformer; R^2 ~0.92–0.96 [15]

Accordingly, the literature [17] suggests that in adaptive control contexts the best model is not determined solely by forecasting benchmarks. A slightly less accurate predictor may still be preferable if it is easier to retrain online, more numerically stable, more transparent in deployment, or easier to couple with MPC or another supervisory control architecture.

7. The imperative for data augmentation

An important theoretical and practical limitation in discriminative forecasting is that a model's expected generalization error is inversely proportional to its training set size [18]. In the context of building energy modelling, capturing large, perfectly clean datasets is rarely achievable. Structural breaks in legacy HVAC operations, low sampling rates from aging IoT sensors, communication packet loss, and high data gathering costs severely restrict the size and quality of available training sets. When complex multivariate time-series models are trained on these limited, noisy datasets, they suffer from severe regression model overfitting.

To counteract the mathematical restrictions imposed by small sample sizes, researchers have pivoted to data augmentation methodologies. While traditional time-series augmentation methods predominantly utilized simple feature-space transformations (such as adding Gaussian noise or time-shifting) to artificially expand training set sizes, these methods yield only marginal improvements [18]. The introduction of generative neural networks into the forecasting pipeline represents a paradigm shift. By synthetically expanding the training set size using generative algorithms that understand the underlying physics and probability distributions of the building, the robustness of the forecasting models to out-of-sample data is vastly improved, significantly boosting out-of-sample prediction accuracies and mitigating the overfitting dilemma.

Adaptive HVAC control via Deep Reinforcement Learning

1 The transition from rule-based to data-driven control

In control-oriented building energy management, the forecaster is only one part of a larger decision system. The 2025 systematic review [1] by Rehman et al. on energy management models in smart homes shows that recent research does not concentrate only on prediction quality. It addresses renewable integration, demand-side management, user behaviour, data protection, automation, and grid stability. The authors explicitly note that consumption variability, occupancy, weather, and user preferences require real-time adaptability in energy management. Based on the investigation they identify a gap between theoretical advances and practical large-scale adoption of HEMS. They further argue that machine-learning-driven energy management remains underexplored in practice and that effective models must account for broader operational concerns rather than only algorithmic performance.

Another systematic review [15] by Michailidis et al. on machine learning for energy management in buildings focused specifically on real-world applications and found that practical systems are built around broader building energy management strategies rather than around a single forecasting model. The review highlights the operational importance of integration, deployment, and decision support in real environments.

Traditional building HVAC systems utilize predetermined, static logic. For example, a conventional rule-based system might activate a heating plant at a same time every weekday morning to ensure the facility reaches its target temperature by certain time, regardless of the external thermodynamic variables or actual occupancy levels. Conversely, AI-based control relies on data-driven adaptive logic. While the physical, mechanical elements of the HVAC system (fans, compressor valves, hydronic pumps) operate within strict physical limits, an AI layer acts as a manager that coordinates these components dynamically. By analysing localized weather forecasts, the building's thermal mass, and historical performance data, an AI system might calculate that on a mild day, heating can be delayed, whereas on a frigid day, it must start earlier. This intelligent preemption minimizes energy waste and improves equipment service life.

Deep Reinforcement Learning (DRL) algorithms are uniquely suited for this specific task. By formally framing the building control problem as a Markov Decision Process (MDP), the DRL agent autonomously learns an optimal control policy through continuous trial-and-error interaction with the

building environment [5]. Figure 1 illustrates the basic reinforcement learning loop as applied to building energy management.

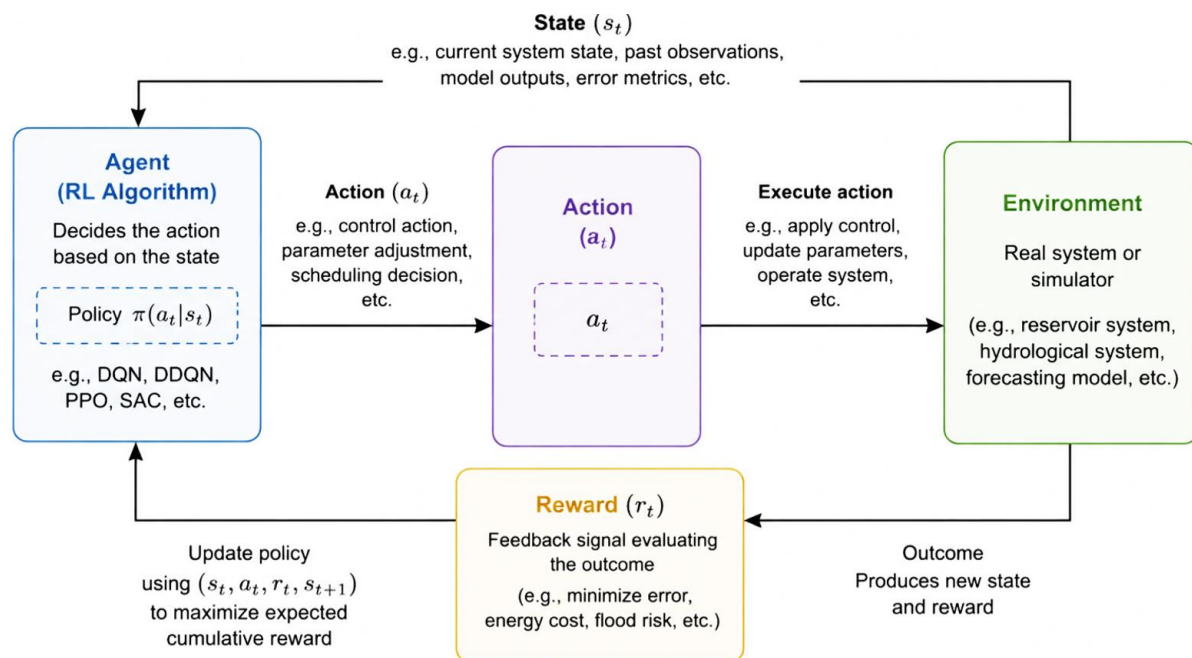


Fig. 1. Reinforcement Learning for Adaptive Energy Management

The RL agent observes the current state of the environment, including sensor readings such as indoor temperatures, energy consumption, occupancy, and weather conditions. Based on this state, the agent selects an action, such as adjusting HVAC setpoints or dispatching energy storage.

The agent receives positive rewards for achieving energy savings and severe penalties for comfort violations. Studies involving DRL agents deployed within physics-calibrated digital twins demonstrate that these systems can reduce annual HVAC energy demand by 10% to 35% [19]. They achieve these energy reductions while tightly maintaining operative temperatures within $\pm 0.5^\circ\text{C}$ of the setpoint and keeping indoor CO_2 concentrations near or below 800 ppm, adhering strictly to ASHRAE Standard 55 (thermal environmental conditions) and ASHRAE Standard 62.1 (ventilation for acceptable indoor air quality).

2. The RLDA framework

A significant barrier to the real-world deployment of DRL in building automation is sample efficiency. DRL agents traditionally require millions of interactions with an environment to converge on an optimal policy, which is impossible in a physical building where random exploration can cause equipment damage. To resolve this, Kurte, Kuldeep, et al. developed the Reinforcement Learning with Data Augmentation (RLDA) framework [20].

The RLDA framework leverages the TimeGAN architecture as an online data generator. The TimeGAN allows the DRL agent to pre-train extensively in a simulated environment before being deployed by generating vast amounts of synthetic, highly realistic thermodynamic transitions. Figure 2 illustrates how the RL agent interacts with the environment, observing the system state and generating control actions for flexible energy resources.

Experiences in the replay memory, stored as sequential tuples $\langle S_t, A_t, R_{t+1}, S_{t+1} \rangle$, are periodically sampled to create partial trajectories. These partial trajectories are used to train a TimeGAN model, which generates synthetic experience tuples.

The synthetic tuples are then combined with real experiences from the replay memory to form an augmented dataset for Deep Q-Network (DQN) training. This approach enables the agent to learn more efficiently from limited real-world data, improving policy performance for adaptive energy management across multiple flexible resources.

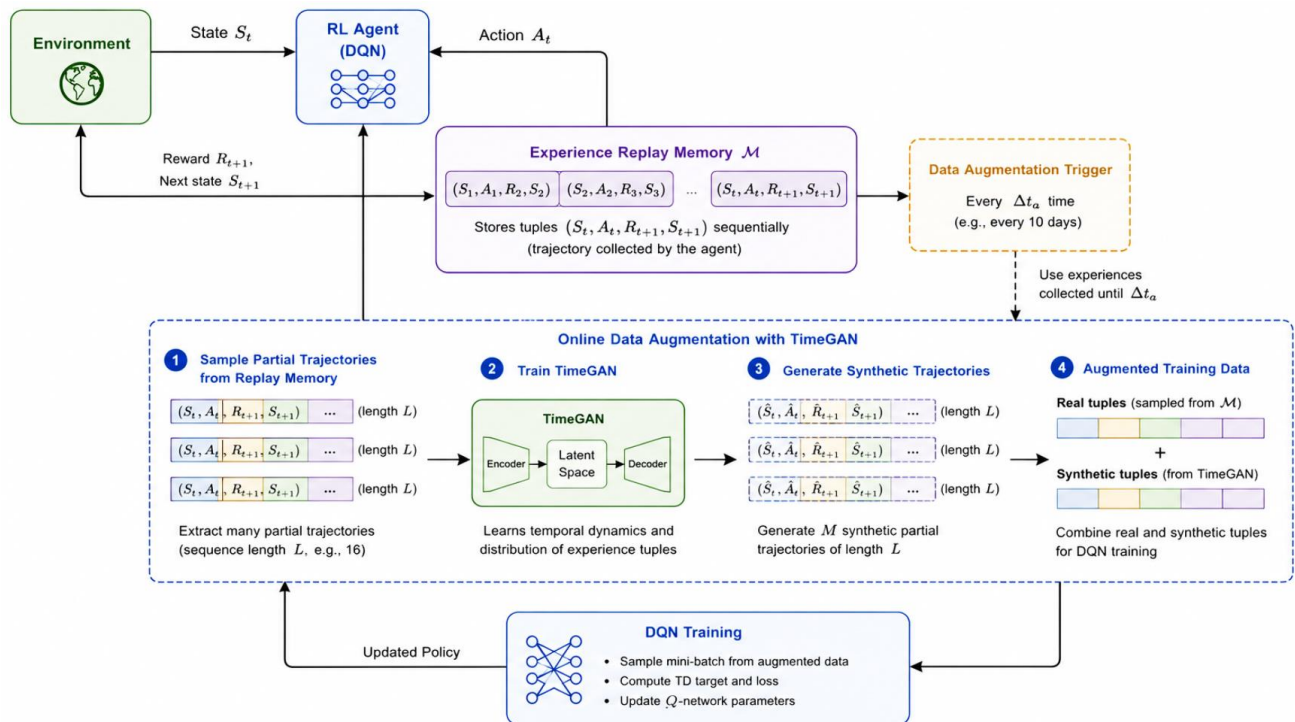


Fig. 2. The RLDA Framework Architecture

In case studies focusing on intelligent HVAC control, the implementation of the RLDA framework yielded a substantial 28% improvement in sample efficiency [20]. This reduction in required training time eliminates one of the primary bottlenecks preventing the widespread adoption of DRL-based intelligent control in real-world energy management systems.

3. The SACEM framework

Despite the promise of DRL, standard algorithms such as Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradient (DDPG), or vanilla Soft Actor-Critic (SAC) frequently exhibit critical flaws when deployed in live buildings. They often suffer from myopic decision-making and limited robustness to environmental uncertainty, frequently treating human comfort as a secondary objective to be sacrificed in pursuit of energy cost minimization.

To achieve environmentally sustainable management without compromising occupant well-being, advanced hybrid control frameworks have been developed [5]. The SACEM framework (Soft Actor-Critic with Cross-Entropy Method) represents a state-of-the-art solution designed specifically for commercial buildings in hot desert climates. A schematic representation of its architecture is provided in Figure 3.

SACEM bridges the gap between the predictive reasoning of Model Predictive Control (MPC) and the adaptability of DRL through three structural innovations:

1. *Two-layer decision architecture* couples a stochastic entropy-regularized policy (SAC) with a short-horizon action refinement optimizer (CEM). The SAC agent provides data-efficient off-policy learning, while the CEM layer explicitly evaluates candidate action sequences under a thermal transition model, allowing for vital decision-time look-ahead reasoning;

2. *Comfort-dominant reward structure* utilizes strongly asymmetric penalties that make comfort violations mathematically and economically irrational for the agent. This explicitly prioritizes thermal comfort even when the system operates under aggressive time-of-use (TOU) electricity pricing schemes;

3. *A safety-aware projection mechanism* acts as a hard boundary during the action refinement stage preventing the execution of unsafe candidate actions even during the exploratory phases of learning.

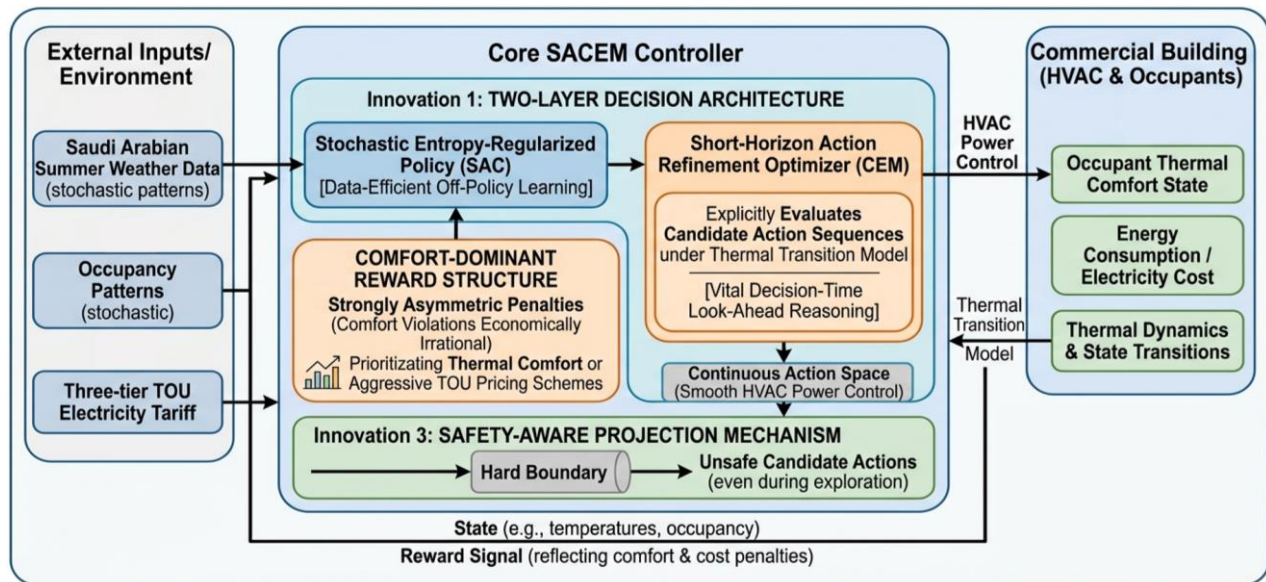


Fig. 3. The SACEM Framework Architecture

The technical implementation of SACEM relies on a continuous action space to enable smooth HVAC power control. In rigorous simulations utilizing weather conditions, stochastic occupancy patterns, and a three-tier TOU tariff, SACEM consistently demonstrated profound resilience intelligence. It successfully avoided brittle behaviours such as control oscillations or total system shutdowns during heat spikes. Performance metrics confirmed that SACEM achieved a comfort score of 95.8%, while simultaneously reducing energy consumption and operating costs by approximately 21% relative to the baseline DRL algorithms [5].

4. The integration of Generative AI in control

1. Generative modelling

While augmenting DRL with synthetic data via TimeGANs improves sample efficiency, the industry is moving toward a more holistic integration of generative AI into the control loop itself. In this context, generative modelling may include scenario generation, probabilistic forecasting, synthetic trajectory generation, multi-step sequence prediction, and uncertainty-aware forecasting [12]. Such methods are particularly relevant for adaptive building energy forecasting and control because they can represent multiple plausible futures rather than a single expected outcome, enabling control systems to evaluate risks, anticipate variability, and support robust or stochastic decision-making under uncertainty.

General-purpose sequential solvers like Monte Carlo Tree Search (MCTS) enable energy optimization capabilities. For example, in managing a residential heat pump, the system utilizes MCTS within the latent space to simulate hundreds of potential action sequences, identifying the precise trajectory that minimizes energy costs against dynamic grid pricing while perfectly maintaining thermal comfort. This paradigm empowers scalable demand response, allowing individual smart buildings to act as highly flexible assets that participate intelligently in grid-level market-clearing mechanisms, ultimately supporting the stability of the broader electrical grid.

2. Generative Model Predictive Control (GMPC)

Model Predictive Control (MPC) is an industrial paradigm that involves projecting a sequence of future actions and using an explicit mathematical model to evaluate that sequence by simulating system dynamics.

In practice, traditional MPC typically requires deriving linear models from first principles and solving a Mixed-Integer Linear Programming (MILP) problem at each decision step.

Figure 4 illustrates the basic architecture of Model Predictive Control (MPC) for a smart building energy management system.

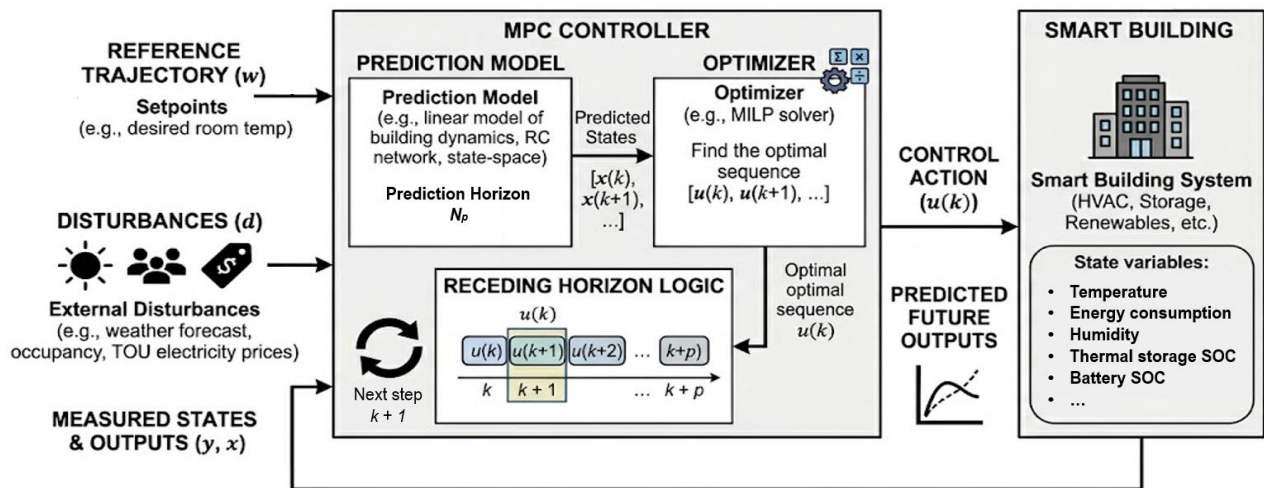


Fig. 4. Basic Architecture of Model Predictive Control for BEMS

At each control step, the MPC controller receives three main inputs: the reference or setpoint trajectory, such as desired indoor temperature or comfort limits; forecasted disturbances, such as weather, occupancy, and electricity prices; and the measured states and outputs of the building, such as zone temperature, energy use, and storage state of charge. These inputs are processed by the prediction model, which estimates the future behaviour of the building over a finite prediction horizon. The optimizer then uses these predicted states and outputs to determine an optimal sequence of control actions while satisfying comfort, actuator, energy, and operational constraints. Rather than implementing the entire planned sequence, MPC utilizes a receding horizon approach: it executes only the first calculated action, discarding the remaining trajectory. The building response is then measured and fed back to the controller, and the optimization is repeated at the next time step. This receding-horizon structure allows MPC to continuously adapt its decisions to updated measurements and changing external conditions.

A 2025 review [21] of Model Predictive Control for smart buildings shows that MPC remains a central framework in modern building energy management systems, applied to HVAC, storage, renewables, domestic hot water, EV charging, and lighting. The review emphasizes that MPC's value lies in optimizing control actions using predicted future behaviour under dynamic conditions.

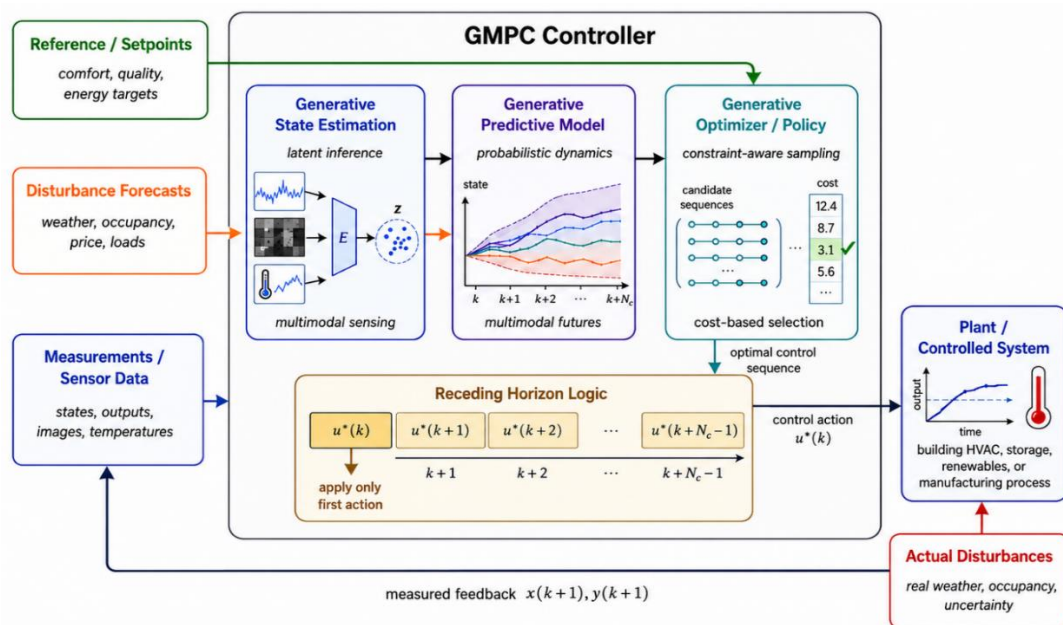


Fig. 5. Architecture of Generative Model Predictive Control for BEMS

Tohidi et al. (2024) further demonstrate the continued relevance of adaptive MPC [22] by proposing an adaptive controller for building thermal dynamics that accounts for parameter deviation and system constraints. It shows that the contemporary control problem is one of adaptation under changing conditions, not only prediction under static conditions.

However, HVAC systems exhibit inherently uncertain and highly nonlinear behaviors. When the thermodynamics are high-dimensional, establishing exact state-space equations becomes computationally intractable, resulting in significant performance degradation for traditional MPC under real-world uncertainties. Generative Model Predictive Control (GMPC) resolves these limitations by approximating the complex MPC control law using generative neural networks [23]. Figure 5 presents an architectural framework for GMPC applied to complex smart building HVAC systems.

Instead of relying on rigid engineering assumptions, GMPC utilizes diffusion-based generative models and adversarial networks that are capable of capturing the diverse, multimodal distributions of building behaviours entirely from data.

Microservices architectures

1. Transition to microservices architectures

The integration of Generative AI and real-time DRL into critical building infrastructure physically cannot be supported by legacy software systems. Historically, Building Energy Management Systems (BEMS) utilized monolithic application architectures or outdated Service-Oriented Architectures (SOA) [24]. Monolithic systems forced extensive, months-long testing cycles for even minor algorithmic adjustments, while SOA suffered from brittle orchestration logic and heavy XML-based communication overhead.

The transition to microservices architectures where the application is decomposed into small, independently deployable, and loosely coupled containerized logic units has proven mandatory for the realization of real-time management in smart buildings.

Table 2 compares monolithic, SOA-based, and microservices-based architectures in the context of an AI-enabled EMS.

Table 2

Comparison of application architectures in the context of an AI-enabled EMS

Aspect	Monolithic	SOA	Microservices
Structure	One large application	Several large services	Many small independent services
Typical size	1 deployable system	5-15 business services	10-100 small services
Example EMS modules	HVAC, lighting, storage, pricing, reports in one system	HVAC, billing, forecasting, reporting as separate services	Separate services for HVAC control, battery, EV charging, forecasting, occupancy, alerts
AI integration	Difficult to update AI models without affecting the whole EMS	AI can be added as shared services, but integration may be centralized	AI models can be deployed as independent prediction / control services
Prediction and control	Usually rule-based or centralized optimization	Can support forecasting and optimization services	Well suited for AI-based forecasting, anomaly detection, MPC/GMPC, and real-time control
Scalability	Whole EMS must be scaled	Major services can be scaled	Only AI-heavy services can be scaled, e.g., forecasting or control
Failure impact	One AI/control failure may affect the full EMS	Failure may affect connected services	Failure can be isolated, e.g., prediction service fails while HVAC fallback continues
Main drawback	Hard to modify as complexity grows	Can become middleware-heavy	Requires strong monitoring, orchestration, and cybersecurity

AI-enabled EMS introduces additional requirements, including continuous data ingestion, model training, model updating, uncertainty handling, and real-time inference. These requirements make the architecture more dynamic and computationally demanding. As a result, microservices become more suitable because prediction models, anomaly detection modules, MPC/GMPC

controllers, and optimization services can be deployed, updated, scaled, and monitored independently. For example, an AI forecasting service may require high computational resources during peak operation, while a reporting service may not. Separating these functions allows the EMS to remain flexible, fault-tolerant, and easier to upgrade. Therefore, the integration of AI shifts the preferred architecture from a tightly coupled monolithic system toward microservices architectures.

Simulations and real-world industrial deployments demonstrate that transitioning from SOA to containerized microservices EMS yields operational improvements. In comparative analyses, the annual failure time of energy management platforms decreased from 3.83 minutes under SOA to 0.26 minutes using microservices, pushing the overall system reliability to 99.99995% [25].

3. Temporal domain separation and agentic orchestration

A paramount architectural challenge in deploying Generative AI is that their inherent inference latency (typically 1 to 5 seconds) fundamentally conflicts with the safety-critical deterministic requirements (sub-100 milliseconds) of physical HVAC control loops [6].

To resolve this critical bottleneck, advanced microservices architectures implement an explicit design pattern known as Temporal Domain Separation. Rather than relying on a single monolithic AI component, the system deploys a highly distributed set of process-aware agents coordinated by an orchestrator.

Firstly, there is a deterministic supervisor that operates in a "hard real-time" domain (<100 ms). It continuously evaluates safety-critical states and retains absolute veto power over any recommendations generated by higher-level generative AI, guaranteeing operational safety.

Secondly, LLM assistant operates in a "soft real-time" domain (1–2 seconds) and handles complex, non-deterministic tasks such as holistic diagnostics, heuristic reasoning, and natural language interactions with building operators.

To ensure these LLMs can interpret numerical sensor data without hallucinations, the architecture utilizes bounded in-memory buffers to maintain a rolling dataset of logical windows (e.g., a 2-minute horizon for immediate state estimation and a 10-minute horizon for diagnostics). Incoming raw sensor records are actively enriched in real-time with semantic labels and fuzzy quality indices.

Managing the continuous lifecycle of generative models in a distributed environment requires specialized Machine Learning Operations (MLOps). Modern MLOps pipelines are designed to dynamically reconfigure computation graphs at runtime, automatically selecting the optimal neural network model based on dataset complexity and the computational constraints of the available edge resources [26]. At the hardware level, executing generative AI inference at the edge demands extreme energy efficiency to prevent the AI systems from negating the energy savings they create. Frameworks such as the Approximate Edge Inference System (AxIS) [27] adopt a full-system approach to approximate computing, selectively trading off microscopic degradations in application quality for significant energy savings.

Conclusions

This review demonstrates that the transition toward highly sustainable, energy-efficient smart buildings demands a fundamental reorientation away from deterministic, physics-led engineering conventions and toward adaptive, data-driven intelligence. The results confirm that integrating generative neural networks provides a mathematically grounded response to the empirical data limitations of traditional discriminative machine learning. By synthetically augmenting sparse sensor data, these advanced architectures enable the development of robust predictive models that resist statistical overfitting while preserving occupant privacy. The results demonstrate that deep reinforcement learning frameworks successfully reconcile ambitious energy reduction targets with stringent thermal comfort requirements, allowing smart buildings to operate as dynamic, grid-interactive assets. The results show that translating these theoretical control models into physical operations necessitates highly specialized software infrastructures. Specifically, the adoption of decentralized microservices and temporal domain separation ensures that the soft real-time inference latency of generative architectures does not violate the hard real-time safety constraints of physical heating, ventilation, and air-conditioning machinery.

Compared to previous reviews that primarily emphasized isolated, short-term load forecasting using classical sequence models, this synthesis highlights a broader operational paradigm. While earlier literature established discriminative networks as the baseline for prediction, the current study illustrates a necessary shift toward systems where predictive foresight is actively coupled with generative data augmentation and autonomous, real-time control loops.

Despite these comprehensive findings, a notable limitation of this review is its reliance on simulation-heavy literature. Although theoretical and simulated frameworks demonstrate significant potential, large-scale, longitudinal data from physical generative AI deployments in commercial buildings remains relatively scarce, which constrains the ability to fully evaluate long-term hardware degradation and real-world maintenance costs.

Consequently, this article proposes that future research should prioritize empirical validation through physical pilot deployments. Subsequent studies must focus specifically on minimizing the computational overhead of edge-based generative inference, developing highly energy-efficient hardware solutions, and establishing standardized protocols for retrofitting legacy building management systems with decentralized, fault-tolerant microservices.

References

1. Rehman, U. u., Faria, P., Gomes, L., & Vale, Z. (2025). Future of Energy Management Models in Smart Homes: A Systematic Literature Review of Research Trends, Gaps, and Future Directions. *Process Integration and Optimization for Sustainability*. <https://doi.org/10.1007/s41660-025-00506-x>.
2. Alanne, K., & Sierla, S. (2022). An overview of machine learning applications for smart buildings. *Sustainable Cities and Society*, 76, 103445. <https://doi.org/10.1016/j.scs.2021.103445>.
3. Yin, Q., Han, C., Li, A., Liu, X., & Liu, Y. (2024). A review of research on building energy consumption prediction models based on artificial neural networks. *Sustainability*, 16(17), 7805. <https://doi.org/10.3390/su16177805>.
4. Qiu L. Deep learning approaches for building energy consumption prediction. (2024). *Frontiers in Environmental Research*, 2(3). <https://doi.org/10.61784/fer3012>.
5. Alshammari, A., Elhadi, A. A. E., & Ibrahim, A. O. (2026). Environmentally sustainable HVAC management in smart buildings using a reinforcement learning framework SACEM. *Sustainability*, 18(2), 1036. <https://doi.org/10.3390/su18021036>.
6. González-Potes, A., Martínez-Castro, D., Paredes, C. M., Ochoa-Brust, A., Mena, L. J., Martínez-Peláez, R., Félix, V. G., & Félix-Cuadras, R. A. (2026). Hybrid AI and LLM-Enabled Agent-Based Real-Time Decision Support Architecture for Industrial Batch Processes: A Clean-in-Place Case Study. *AI*, 7(2), 51. <https://doi.org/10.3390/ai7020051>.
7. Runge, J., & Zmeureanu, R. (2019). Forecasting energy use in buildings using artificial neural networks: A review. *Energies*, 12(17), 3254. <https://doi.org/10.3390/en12173254>.
8. El Alaoui M., & Rougui M. (2024). Examining the application of artificial neural networks (ANNs) for advancing energy efficiency in building: A comprehensive reviews. *Journal of Sustainability Research*, 6(1). <https://doi.org/10.20900/jsr20240001>
9. Liu, H., Liang, J., Liu, Y., & Wu, H. (2023). A review of data-driven building energy prediction. *Buildings*, 13(2), 532. <https://doi.org/10.3390/buildings13020532>.
10. Mishra, A., Lone, H. R., & Mishra, A. (2024). DECODE: Data-driven energy consumption prediction leveraging historical data and environmental factors in buildings. *Energy and Buildings*, 307, 113950. <https://doi.org/10.1016/j.enbuild.2024.113950>.
11. Yang, F., & Mao, Q. (2023). Auto-Evaluation model for the prediction of building energy consumption that combines modified Kalman filtering and Long Short-Term Memory. *Sustainability*, 15(22), 15749. <https://doi.org/10.3390/su152215749>.
12. Li, Y., Tong, Z., Tong, S., & Westerdahl, D. (2022). A data-driven interval forecasting model for building energy prediction using attention-based LSTM and fuzzy information granulation. *Sustainable Cities and Society*, 76, 103481. <https://doi.org/10.1016/j.scs.2021.103481>.
13. Li, G., Wang, Y., Xu, C., Wang, J., Fang, X., & Xiong, C. (2024). BO-STA-LSTM: Building energy prediction based on a Bayesian Optimized Spatial-Temporal Attention enhanced LSTM method. *Developments in the Built Environment*, 100465. <https://doi.org/10.1016/j.dibe.2024.100465>.
14. Zheng, P., Zhou, H., Liu, J., & Nakanishi, Y. (2023). Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture. *Applied Energy*, 349, 121607. <https://doi.org/10.1016/j.apenergy.2023.121607>.
15. Spencer, R., Ranathunga, S., Boulic, M., van Heerden, A., & Susnjak, T. (2025). Transfer learning on transformers for building energy consumption forecasting - A comparative study. *Energy and Buildings*, 336, 115632. <https://doi.org/10.1016/j.enbuild.2025.115632>.

16. Bastakoti, B., Parajuli, S., & Regmi, B. (2026). LSTM vs. Transformer models in power forecasting: A comprehensive survey. Preprints.org. <https://doi.org/10.20944/preprints202603.0624.v1>.
17. Michailidis, P., Minelli, F., Michailidis, I., Kurucan, M., Coban, H. H., & Kosmatopoulos, E. (2025). Machine learning for energy management in buildings: A systematic review on real-world applications. *Energies*, 19(1), 219. <https://doi.org/10.3390/en19010219>.
18. Demir, S., Mincev, K., Kok, K., & Paterakis, N. G. (2021). Data augmentation for time series regression: Applying transformations, autoencoders and adversarial networks to electricity price forecasting. *Applied Energy*, 304, 117695. <https://doi.org/10.1016/j.apenergy.2021.117695>.
19. Amangeldy, B., Tasmurzayev, N., Imankulov, T., Baigarayeva, Z., Izmailov, N., Riza, T., Abdugarimov, A., Mukazhan, M., & Zhumagulov, B. (2025). AI-Powered Building Ecosystems: A Narrative Mapping Review on the Integration of Digital Twins and LLMs for Proactive Comfort, IEQ, and Energy Management. *Sensors*, 25(17), 5265. <https://doi.org/10.3390/s25175265>.
20. Kurte, K., Amasyali, K., Munk, J., & Zandi, H. (2022). Deep reinforcement learning with online data augmentation to improve sample efficiency for intelligent HVAC control. *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 479–483. <https://doi.org/10.1145/3563357.3566168>.
21. Michailidis, P., Michailidis, I., Minelli, F., Coban, H. H., & Kosmatopoulos, E. (2025). Model predictive control for smart buildings: Applications and innovations in energy management. *Buildings*, 15(18), 3298. <https://doi.org/10.3390/buildings15183298>.
22. Tohidi, S. S., Cali, D., & Madsen, H. (2024). Adaptive model predictive controller for building thermal dynamics. *IEEE Control Systems Letters*, 8, 1325–1330. <https://doi.org/10.1109/LCSYS.2024.3410170>.
23. Lee, S. K., Stone, R. F. P., Gao, M., Zhang, W., Sha, Z., & Ko, H. (2025). Generative model predictive control in manufacturing processes: A review. *arXiv*. <https://doi.org/10.48550/arXiv.2511.17865>.
24. Nuñez, I., Rovetto, C., Cruz, E., Smolarz, A., Concepcion, D., & Cano, E. E. (2024). Design of a microservices-based architecture for residential energy efficiency monitoring. *International Journal of Electronics and Telecommunications*, 70, 1089. <https://doi.org/10.24425/ijet.2024.152511>.
25. Lyu, Z., Wei, H., Bai, X., & Lian, C. (2020). Microservice-based architecture for an energy management system. *IEEE Systems Journal*, 14(4), 5061–5072. <https://doi.org/10.1109/jsyst.2020.2981095>.
26. Pitsun, O., & Shymchuk, M. (2026). Scalable MLOps pipeline with complexity-driven model selection using microservices. *Technologies*, 14(1), 45. <https://doi.org/10.3390/technologies14010045>.
27. Ghosh, S. K. (2023). Energy efficient edge inference systems [Doctoral dissertation, Purdue University]. Purdue University Graduate School. <https://doi.org/10.25394/PGS.23737488>.

Верлань Андрій Анатолійович

доктор технічних наук, професор кафедри ПЗЕ НН ІАТЕ
Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна,
Норвезький університет природничих та технічних наук, Тронгейм, Норвегія
ORCID: 0000-0002-6469-2638
E-mail: a.verlan@edu.kpi.ua

Дячук Андрій Олегович

аспірант кафедри ПЗЕ НН ІАТЕ
Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна
ORCID: 0009-0003-6979-278X
E-mail: diachuk.a.o.-tv51f@edu.kpi.ua

ПІДХОДИ НА ОСНОВІ ГЕНЕРАТИВНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ АДАПТИВНОГО ПРОГНОЗУВАННЯ ЕНЕРГОСПОЖИВАННЯ БУДІВЕЛЬ І КЕРУВАННЯ: ОГЛЯД МЕТОДІВ ТА МІКРОСЕРВІСНИХ АРХІТЕКТУР

Стаття представляє всебічний огляд переходу від статичних заснованих на правилах систем керування будівлями до керованого даними штучного інтелекту для оптимізації енергоспоживання в забудованому середовищі. Хоча дискримінативні нейронні мережі покращили прогнозування енергоспоживання, їх практичному розгортанню часто перешкоджає перенавчання регресійних моделей на обмежених та зашумлених наборах даних із датчиків. Метою цього дослідження є аналіз інтеграції генеративних нейронних мереж в адаптивне прогнозування та керування енергоспоживанням будівель для подолання інформаційних та експлуатаційних обмежень традиційного машинного навчання. Для досягнення цієї мети було виконано такі завдання: синтез методологічного переходу від класичного моделювання послідовностей до генеративного розширення даних, оцінювання модельних фреймворків глибокого навчання з підкріпленням для автономного управління системами HVAC, а також дослідження децентралізованих мікросервісних архітектур, необхідних для

периферійно-хмарного розгортання. Аналіз показує, що генеративні моделі, такі як TimeGAN, синтезують навчальні дані для подолання емпіричної нестачі інформації, тим самим покращуючи ефективність вибірки DRL-агентів і запобігаючи збоєм обладнання під час дослідницького навчання. Крім того, гібридні фреймворки керування, такі як SACEM та генеративне модельне прогнозувальне керування, забезпечують досягнення цілей зі зниження енергоспоживання разом із підтримкою теплового комфорту людини в нестационарних середовищах. Зрештою, перенесення цих обчислювальних досягнень у фізичну експлуатацію будівель вимагає структурного переходу до контейнеризованих мікросервісів. Зокрема, впровадження часового розділення доменів є важливим для ізоляції затримки інференсу генеративних алгоритмів від детермінованих обмежень безпеки фізичного обладнання HVAC, що гарантує безпечне та адаптивне керування.

Ключові слова: генеративні нейронні мережі, керування енергоспоживанням будівель, розумні будівлі, глибоке навчання з підкріпленням, HVAC, мікросервісні архітектури, LSTM, трансформери.

Надійшла до редакції (Received): 14.07.2026

Прийнята до друку (Accepted): 08.09.2026

Опубліковано онлайн (Available online): 15.09.2026

<http://creativecommons.org/licenses/by/4.0/>

This work is licensed under Creative Commons Attribution-noncommercial-sharealike 4.0 International License

УДК 004.056:327.88:316.776.4

DOI: 10.31673/2409-7292.2026.037203

Дудатєв Андрій Веніамінович

кандидат технічних наук, доцент, доцент кафедри захисту інформації

Вінницький національний технічний університет, Вінниця, Україна

ORCID: 0000-0002-7944-2404

E-mail: dydatyev.av@gmail.com

Гарнага Володимир Анатолійович

кандидат технічних наук, доцент, доцент кафедри захисту інформації

Вінницький національний технічний університет, Вінниця, Україна

ORCID: 0009-0000-9180-6445

E-mail: garnaga.volodymyr@vntu.edu.ua

ТРАНСФОРМАЦІЯ ВИДІВ ІНФОРМАЦІЙНОГО ПРОТИБОРСТВА: КОНЦЕПЦІЯ МЕТАІНФОРМУВАННЯ

У контексті сучасних гібридних війн інформаційне протиборство, зокрема інформаційні війни, набуває ключового значення та стає ефективним інструментом досягнення стратегічних і тактичних цілей. Інформаційний простір перетворюється на окреме поле бою, де вплив спрямований передусім на свідомість, поведінку та процеси прийняття рішень людини, формуючи необхідні наративи й установки без прямого застосування сили. Реалізація інформаційно-психологічних спеціальних операцій забезпечує доступ до інформаційних та управлінських ресурсів через цілеспрямований вплив на окремих осіб, персонал організацій і населення держави загалом. Такий вплив здійснюється через мас-медіа, соціальні мережі та інші канали комунікації, що значно підвищує його ефективність і масштаб. У статті проведено аналіз відомих комунікативних моделей, що дозволяють описати процеси передачі та сприйняття інформації. На цій основі запропоновано поняття тансінформування як форми впливу, що поєднує трансляцію, трансформацію та адаптацію повідомлень до цільової аудиторії. Також запропоновано концептуальну модель метаінформування, яка враховує процеси наслідування, накопичення та набуття знань упродовж життєвого циклу людини, зокрема під впливом цілеспрямовано поширюваної інформації. Запропонований підхід сприяє кращому розумінню механізмів інформаційно-психологічного впливу та підвищенню ефективності протидії в умовах інформаційних конфліктів. На завершення запропоновано систему аналітичних виразів, що дозволяють кількісно оцінювати ефективність проведення інформаційно-психологічних операцій, а також алгоритм процесу проведення такого оцінювання.

Ключові слова: комунікативні моделі, інформаційно-психологічних спеціальних операції, мем, трансінформування, мем-програмування, ІПСО.

Вступ

Доступ до сучасних глобальних інформаційних ресурсів, зокрема наукових, освітніх, розважальних тощо – суттєво розширив можливості людини щодо отримання інформації