UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
DOUTORADO EM SISTEMAS DE GESTÃO SUSTENTÁVEIS

LEONARDO FONTOURA DO NASCIMENTO

GENERATIVE ENERGY FORECAST: A HYBRID FORECASTING MODEL COMBINING STATISTICAL AND GENERATIVE AI FOR SHORT-TERM ENERGY DEMAND

Niterói
2025

UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
DOUTORADO EM SISTEMAS DE GESTÃO SUSTENTÁVEIS

LEONARDO FONTOURA DO NASCIMENTO

GENERATIVE ENERGY FORECAST: A HYBRID FORECASTING MODEL COMBINING STATISTICAL AND GENERATIVE AI FOR SHORT-TERM ENERGY DEMAND

Tese apresentada ao Curso de Doutorado em Sistemas de Gestão Sustentáveis da Universidade Federal Fluminense como requisito parcial para obtenção do Grau de Doutor em Sistemas de Gestão Sustentáveis. Área de Concentração: Sistemas de Gestão da Sustentabilidade. Linha de Pesquisa: Gestão das Organizações Sustentáveis.

Orientador

Prof. Julio Vieira Neto, D.Sc.
Universidade Federal Fluminense – UFF

Co-orientador

Prof. Daniel Luiz de Mattos Nascimento, D.Sc.
University of Barcelona Business School – UB

Niterói
2025

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

F677g Fontoura, Leonardo
Generative Energy Forecast: A Hybrid Forecasting Model Combining Statistical and Generative AI for Short-term Energy Demand / Leonardo Fontoura. - 2025.
101 f.: il.

Orientador: Julio Vieira Neto.
Coorientador: Daniel Luiz De Mattos Nascimento.
Tese (doutorado)-Universidade Federal Fluminense, Escola de Engenharia, Niterói, 2025.

1. Hybrid Forecasting Model. 2. Generative Artificial Intelligence. 3. Neural Networks. 4. Energy Consumption. 5. Produção intelectual. I. Vieira Neto, Julio, orientador. II. De Mattos Nascimento, Daniel Luiz, coorientador. III. Universidade Federal Fluminense. Escola de Engenharia. IV. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

LEONARDO FONTOURA DO NASCIMENTO

GENERATIVE ENERGY FORECAST: A HYBRID FORECASTING MODEL COMBINING STATISTICAL AND GENERATIVE AI FOR SHORT-TERM ENERGY DEMAND

Tese apresentada ao Curso de Doutorado em Sistemas de Gestão Sustentáveis da Universidade Federal Fluminense como requisito parcial para obtenção do Grau de Doutor em Sistemas de Gestão Sustentáveis. Área de Concentração: Sistemas de Gestão da Sustentabilidade. Linha de Pesquisa: Gestão das Organizações Sustentáveis.

Aprovada em 11 de novembro de 2025.

BANCA EXAMINADORA:

Documento assinado digitalmente
gov.br JULIO VIEIRA NETO
Data: 24/11/2025 20:13:03-0300
Verifique em https://validar.iti.gov.br

Prof. Julio Vieira Neto, D.Sc. - Orientador
Universidade Federal Fluminense – UFF

DANIEL LUIZ DE MATTOS NASCIMENTO - DNI 60366679M
Date: 2025-11-23 19:28:16

Prof. Daniel Luiz de Mattos Nascimento, D.Sc. – Co-orientador
University of Barcelona Business School – UB

Prof. Jose Arturo Garza-Reyes, PhD.
The University of Derby

Prof. Guilherme Luz Tortorella, D.Sc.
The University of Melbourne

Rodrigo Goyannes Gusmão Caiado
Assinado de forma digital por Rodrigo Goyannes Gusmão Caiado
Dados: 2025.11.21 12:32:24 -03'00'

Prof. Rodrigo Goyannes Gusmão Caiado, D.Sc.
Pontifical Catholic University of Rio de Janeiro – PUC-Rio

Documento assinado digitalmente
gov.br ANDRE BARBOSA OLIVEIRA
Data: 18/11/2025 12:00:11-0300
Verifique em https://validar.iti.gov.br

Prof. André Barbosa Oliveira, D.Sc.
Universidade Federal Fluminense – UFF

Documento assinado digitalmente
gov.br GILSON BRITO ALVES LIMA
Data: 25/11/2025 09:20:33-0300
Verifique em https://validar.iti.gov.br

Prof. Gilson Brito Alves Lima, D.Sc.
Universidade Federal Fluminense – UFF

FIGURE LIST

TABLE LIST

EQUATION LIST

SUMMARY

ACKNOWLEDGEMENTS

First and foremost, I express my deepest gratitude to my beloved wife, Priscila Benfeita, and my son, Bernardo Orion, for their unwavering love, patience, and support throughout this journey. Your presence has been my most significant source of strength.

To my mother, Angela Fontoura, thank you for being my foundation. Your wisdom and kindness have shaped who I am, and I carry your values with me in every step.

To my sister-in-law, Amanda Benfeita, thank you for your kindness and encouragement in quiet and meaningful ways and for being my lawyer for absolutely everything.

To Flávio Werly, my best friend and more than a colleague who has been a constant presence throughout the years: Thank you for your unwavering support, loyalty, and sincere companionship along this journey.

To Claudio Carvalho, who retired this year, and with whom I worked from 2001 to the present. His leadership and example had a lasting influence on my professional journey.

To my dear friend and professor, André Barbosa, your guidance and belief in me have left a lasting impression, and I am deeply thankful for your mentorship.

To my supervisor and dear friend, Daniel Nascimento, I am especially grateful for your academic guidance, sharp perspective, and constant encouragement. Your support was instrumental throughout this work.

To my dear friends Valéria Almeida, Diego Paro, Toni Caldeira, and Felipe Fernandes, thank you for the countless hours we spent together on our online adventures. Your friendship and good humour helped me disconnect from daily worries and reminded me of the importance of laughter, even during the most intense moments of this journey.

To Júlio Vieira, my supervisor, thank you especially for your indispensable support in navigating the many bureaucratic hurdles.

Finally, I could not close this section without acknowledging my loyal companions, Rhaika, Bolt and Khiara, who are no longer with us but live on in my heart, and Iron. Each of you, in your own way, brought light, balance, and comfort, even on the most challenging days. Your presence, whether beside me or in memory, has been a constant source of strength and quiet joy throughout this journey.

ABSTRACT

In the face of rising global energy demands, climate imperatives, and the increasing digitalisation of infrastructure, short-term energy forecasting has become essential for operational efficiency and informed decision-making. While recent advances in Artificial Intelligence (AI), particularly Generative AI (Gen-AI), have expanded the forecasting toolkit, they also introduce new challenges related to interpretability, computational cost, and sustainability. This research aims to develop and validate a novel hybrid forecasting model, Generative Energy Forecast (GenEneCast), that integrates classical time-series decomposition, a Deep Learning Model (DLM), and Gen-AI to improve short-term energy consumption forecasts' accuracy, interpretability, and sustainability. The study employed a multi-method research design, beginning with a Scoping Review to map the main Gen-AI technologies in energy management. A questionnaire with domain experts was conducted to assess the relevance of these Gen-AI models for energy efficiency, followed by Friedman and Nemenyi statistical tests to rank the most suitable approaches. Based on these findings, the GenEneCast model was developed, integrating Holt-Winters decomposition, LSTM, and GPT-4-Turbo. Beyond narrative generation, the Large Language Model was employed to suggest optimal configurations for both the Holt-Winters model and the LSTM architecture, dynamically adjusting to the statistical profile of each input series. The model was evaluated using real-world energy data through an Action Research approach. The GenEneCast model was assessed through the Model Confidence Set alongside classical and neural forecasting methods to identify statistically superior models within a multiple comparisons framework. The GenEneCast model outperformed traditional standalone methods, enhancing interpretability and model adaptability. The Green Paradox-AI emerged as a key theoretical contribution, highlighting that excessive gains in predictive precision may lead to disproportionate environmental costs. This study underscores the value of integrating decomposition techniques, residual learning, and generative modelling within a unified and adaptive forecasting framework guided by expert input and robust statistical validation. GenEneCast provides practical flexibility, theoretical advancement in hybrid Gen-AI, and strategic value for decision-makers seeking explainable and resource-aware forecasting solutions.

Keywords: Hybrid Forecasting Model; Generative Artificial Intelligence; Energy Consumption Prediction; Green Paradox-AI; Neural Networks.

LIST OF ABBREVIATIONS AND ACRONYMS

Acronyms	**Name**
AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
CBB	Circular Block Bootstrap
CNNs	Convolutional Neural Networks
CV	Cross-Validation
DLMs	Deep Learning Models
ELBO	Evidence Lower Bound
FFN	Feed Forward Network
GANs	Generative Adversarial Networks
Gen-AI	Generative Artificial Intelligence
GLMs	Generative Language Models
GNNs	Graph Neural Networks
GPT	Generative Pre-trained Transformers
Green AI	Green Artificial Intelligence
LLMs	Large Language Models
LMs	Language Models
LSTMs	Short-Term Memory Networks
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MCS	Model Confidence Set
MHA	Multi-head self-attention
NLP	Natural Language Processing
RMSE	Root Mean Square Error
RNNs	Recurrent Neural Networks
RQs	Research Questions
TDP	Thermal Design Power
TNNs	Transformer Neural Networks
VAEs	Variational Autoencoders
Wh	Watt-hours

1 INTRODUCTION

The global energy landscape is undergoing a profound transformation marked by escalating demand, increasing resource volatility and mounting environmental concerns (WANG; LI; LI, 2025; XIAO et al., 2025). Rapid urbanisation, industrial expansion, and the transition to cleaner energy sources have introduced new layers of complexity to energy systems worldwide (FONTOURA et al., 2023; YU et al., 2025). These dynamics have heightened the risk of imbalances, price fluctuations (QIN et al., 2023; WANG et al., 2025), and supply insecurities (ALHAYDAR; MOHAMMED; GHAITHAN, 2024), particularly in regions where infrastructure remains underdeveloped (MANIKANDAN et al., 2024). As a result, accurate energy forecasting has become not only a technical necessity but a strategic imperative for ensuring operational stability, economic efficiency and environmental sustainability (RUAN et al., 2024).

Forecasting models now play a central role in supporting real-time decision-making, demand response programmes, grid optimisation and long-term planning (FONTOURA et al., 2025). However, traditional methods often fail to address contemporary energy systems' scale, complexity and uncertainty. In this scenario, artificial intelligence (AI) emerges as a critical enabler, offering advanced capabilities to model non-linear behaviours, learn from high-dimensional data, and support adaptive energy management in an increasingly decentralised and data-driven world (HUANG et al., 2023; ZHOU; LUND, 2023). In this sense, predictive modelling operates as a core element of operations management, since it enables more efficient planning, allocation, and control of energy resources, directly linking analytical accuracy to operational performance. Within this context, Deep Learning becomes essential not merely as an alternative but as a methodological necessity to capture the highly dynamic and multivariate dependencies of modern energy systems. Unlike classical statistical models, which assume linearity and stationarity, Deep Learning architectures can learn hierarchical representations directly from data, enabling them to identify non-linear interactions, abrupt regime shifts, and latent temporal dependencies that cannot be explicitly modelled through parametric approaches. This capability is particularly critical for energy consumption forecasting, where complex behavioural, climatic, and infrastructural factors interact simultaneously over time.

Recent years have witnessed a significant evolution in forecasting methodologies, particularly with the emergence of Generative Artificial Intelligence (Gen-AI) as a transformative technology across various domains (FOSSO WAMBA et al., 2023). Gen-AI has expanded the boundaries of what learning systems can achieve by enabling models to generate, adapt and interpret complex patterns within large volumes of data. In the context of energy management, this advancement is particularly relevant, as energy systems increasingly demand models capable of handling heterogeneous, non-linear and context-rich information (FONTOURA et al., 2025; MOUSAVI et al., 2025).

Among the most explored Deep Learning Models (DLMs), Recurrent Neural Networks (RNNs) have demonstrated notable effectiveness in time series forecasting due to their capacity to model sequential dependencies and evolving patterns over time (JOSEPH et al., 2023; KARAKAN, 2024). Their application in energy forecasting has enabled more accurate demand predictions (KHALESIAN; FURNO; LECLERCQ, 2024), anomaly detection (ZAVRAK; ISKEFIYELI, 2023), and load-balancing strategies (ASHWIN et al., 2023), all of which are crucial for improving efficiency and reliability in energy systems. Building upon these advances, the integration of Language Models (LMs), initially developed for natural language tasks, now extends beyond narrative generation to include intelligent parameter suggestion and model configuration. This expands their role in extracting insights from structured time series and unstructured contextual data (JIANG; DALE; LU, 2024). These technologies position DLMs and Gen-AI as a key enabler of the next generation of intelligent, adaptive and context-aware energy forecasting solutions.

1.1 Problem statement

The intensifying energy crisis, coupled with escalating environmental concerns and the global shift towards renewable sources, has elevated the strategic importance of energy forecasting to unprecedented levels (FONTOURA et al., 2023). Effective energy management now requires predictive models that are not only highly accurate but also context-aware and scalable, capable of adapting to volatile consumption patterns and increasingly complex infrastructural configurations. However, despite the significant advances in forecasting methodologies, there remain several unresolved research gaps that limit both theoretical development and practical application in the field.

First, existing approaches, whether grounded in traditional statistical methods (YOUSEFI; ARDEHALI; GHODUSINEJAD, 2023) or in deep learning architectures (JOSEPH et al., 2023), are frequently applied in isolation, revealing a methodological gap related to the fragmentation between interpretable statistical models and complex neural architectures. Statistical models offer transparency and low computational demand but are unable to capture nonlinear patterns, while deep learning models capture complexity at the cost of interpretability and energy efficiency. This separation prevents the exploitation of complementary strengths between these paradigms, representing a central gap in the literature on hybrid energy forecasting approaches.

Moreover, deep learning architecture-based forecasting typically requires significant computational resources during training (CARREON-ORTIZ et al., 2023). Their iterative learning mechanisms and reliance on long input sequences often result in elevated energy consumption and prolonged training times, highlighting a second research gap concerning the sustainability and computational efficiency of AI-driven forecasting systems. Few studies have examined the environmental cost of model complexity, leading to a paradox between predictive precision and energy consumption, a tension that this research conceptualises as the Green Paradox-AI. This theoretical void has not yet been systematically addressed in the energy forecasting domain.

In addition, despite the disruptive potential of Gen-AI, there remains a noticeable absence of empirical studies and practical frameworks that apply these technologies specifically to energy forecasting. The current academic and industrial landscape reveals a critical third gap in developing forecasting models that effectively integrate advanced AI capabilities, particularly those emerging from Gen-AI (FONTOURA et al., 2025). In particular, the literature lacks evidence on how generative models, such as Large Language Models (LLMs), can support both interpretability and dynamic parameter configuration in forecasting systems. Finally, the fourth gap shows that the literature lacks robust hybrid models that jointly optimise interpretability, predictive accuracy and computational efficiency, as well as validation procedures that statistically confirm their superiority over traditional approaches. This absence reinforces the need for rigorous evaluation using methodologies such as the Model Confidence Set (MCS), which has rarely been applied in the context of hybrid forecasting.

This limited exploration of Gen-AI within the energy domain highlights an urgent research challenge: to design and validate hybrid forecasting models that combine statistical decomposition, deep learning, and generative modelling in a single adaptive framework. Such a model should leverage Gen-AI not only for interpretability but also for dynamic model configuration, ensuring a balanced trade-off between accuracy, explainability, and sustainability. By integrating these methodological components, the model contributes to operations management by transforming predictive insights into actionable information for planning, optimisation, and decision-making in energy systems. Addressing these interrelated problems provides the foundation for developing a hybrid model proposed in this study.

In summary, the logical flow of this research follows a macro-level funnel that begins with the global energy transition and sustainability challenges, narrows through the identification of four key research gaps: methodological fragmentation, lack of computational efficiency, limited application of Gen-AI, and absence of statistically validated hybrid models. These gaps converge on the development of the proposed hybrid model. This funnel structure connects the contextual motivations and theoretical foundations to the methodological design and validation strategy, ensuring a coherent progression from problem identification to the formulation of research objectives and questions.

1.2 Research questions

This study seeks to explore the following research questions (RQs):

- RQ1: Which DLM technologies have the highest potential for supporting accurate and context-aware energy consumption forecasting?
- RQ2: Which statistical forecasting method provides the highest complementarity when integrated with selected technologies for enhancing energy consumption prediction?
- RQ3: Which characteristics of Gen-AI make them suitable for interpretability and assisting in configuring and adapting hybrid forecasting systems for monthly energy consumption?
- RQ4: What are the practical trade-offs between complexity, interpretability and computational efficiency in implementing the proposed hybrid model?
- RQ5: How can the proposed hybrid architecture, enhanced by Gen-AI-driven configuration and interpretation, support energy management in real-world scenarios?

1.3 General objective

The primary objective of this study is to design, develop, and evaluate the Generative Energy Forecast (GenEneCast). This hybrid forecasting model combines a statistical method, DLM, and Gen-AI technologies to improve the accuracy, contextual relevance, and operational applicability of monthly energy consumption predictions. The model is guided by the principles of Green AI and the Paradox theory, seeking to balance predictive performance with computational efficiency and environmental sustainability. Unlike prior approaches, GenEneCast also leverages a Generative Language Model (GLM) to assist in parameterising the statistical and neural components, enhancing adaptability across different forecasting scenarios.

1.4 Specific objectives

The specific objectives of this research are:

- To identify and analyse DLM technologies with the most significant potential for supporting accurate and context-aware energy forecasting.
- To evaluate the complementarity of statistical forecasting methods and the suitability of LLMs when integrated into hybrid Gen-AI architectures for energy forecasting applications, including their role in dynamic model configuration.
- To assess the practical trade-offs in implementing the GenEneCast model, with particular attention to complexity, interpretability and computational efficiency, while exploring its effectiveness in supporting strategic and operational decision-making in real-world energy management scenarios.
- To apply the MCS methodology to statistically validate the forecasting performance of GenEneCast compared with classical and neural network benchmarks.
- To conceptualise and propose the Green Paradox-AI framework as a novel theoretical foundation.

1.5 Thesis structure

This study is structured into seven sections. Section 1 introduces the research context, outlines the problem statement, and presents the study's objectives. Section 2 reviews the theoretical background, encompassing key forecasting methodologies, including statistical approaches, DLMs, and recent advances in Gen-AI. Section 3 describes the research design, combining a Scoping Review, expert questionnaire, statistical tests (Friedman and Nemenyi) and developing and evaluating the hybrid model within an action research framework. This section also details the use of the MCS methodology to statistically validate the forecasting performance of GenEneCast against classical and neural network benchmarks. Section 4 details the architecture and functioning of the proposed hybrid model, GenEneCast, which integrates Holt-Winters, LSTM, and GPT-4-Turbo for forecasting, model parameterisation, and narrative generation. Section 5 presents the empirical results, highlighting model performance across configurations and evaluating the practical use of the proposed approach, including statistical validation using the MCS methodology. Section 6 offers a critical discussion, interpreting the results through accuracy, interpretability, energy efficiency, and the Green Paradox-AI framework. Finally, Section 7 concludes the study by summarising the main contributions, practical and theoretical implications, and directions for future research.

1.6 Interdisciplinary nature of the research

The present research adopts an explicitly interdisciplinary approach, integrating theoretical and methodological foundations from operations management, AI, and statistical modelling. From the perspective of operations management, forecasting is understood as a strategic mechanism that supports planning, control, and decision-making processes in complex energy systems. It enables managers to align operational efficiency with long-term sustainability objectives, thereby linking analytical performance to tangible organisational outcomes. AI contributes to this domain by introducing adaptive and learning-based methods capable of addressing non-linear, multivariate, and dynamic challenges that characterise modern energy environments.

In parallel, statistical modelling provides the methodological rigour necessary for validation, interpretability, and reproducibility of results. It bridges the gap between empirical observation and computational representation, ensuring that the predictive models developed are both scientifically sound and operationally applicable. The inclusion of sustainability principles anchors this technical integration in a broader socio-environmental context, reinforcing the need for energy solutions that are not only efficient but also responsible and transparent. By combining these distinct yet complementary perspectives, this research constructs a coherent interdisciplinary framework where managerial, computational, and environmental dimensions converge to support better decision-making in energy operations.

2 THEORETICAL BACKGROUND

This section presents the theoretical foundations that support the development of a hybrid model for forecasting monthly energy consumption. The study adopts the Green AI paradigm (ALZOUBI; MISHRA, 2024) and the Paradox theory (CARTER; KAUFMANN; KETCHEN, 2020; CUNHA; PUTNAM, 2019) as its theoretical basis, focusing on models combining predictive accuracy and computational efficiency. The section reviews DLMs and Gen-AI technologies applied in energy management, including Generative Adversarial Networks (GANs) (GOODFELLOW et al., 2020), Transformer Neural Networks (TNNs) (SCHWALLER et al., 2021), Recurrent Neural Networks (RNNs) (ZHAN et al., 2021), Convolutional Neural Networks (CNNs) (KRIZHEVSKY; SUTSKEVER; HINTON, 2017), Variational Autoencoders (VAEs) (SHRESTHA; MAHMOOD, 2019), and Graph Neural Networks (GNNs) (WU et al., 2021). It also examines statistical and deep learning approaches identified in the literature, discussing their potential contributions and limitations in hybrid model development.

2.1 Applications of DLMs and Gen-AI in energy management

The initial analysis covered 88 documents that focused on applying DLMs and Gen-AI technologies to enhance energy efficiency. After a rigorous screening process based on defined inclusion and exclusion criteria, 36 papers were selected to form the final portfolio. These studies collectively involved contributions from 235 authors who examined the topic between 2016 and 2024. The field has demonstrated notable momentum, with an average annual growth rate of 22.28%, reflecting a growing scholarly interest. China is the leading contributor, accounting for 27.27% of all publications identified. Combined with India, South Korea, and the United States, these four countries account for 61.16% of the total global scientific output.

The literature reveals a clear dominance of Convolutional Neural Networks (CNNs) (HAIDER; KO, 2023; WANG et al., 2024b), which appear in 22 publications. Recurrent Neural Networks (RNNs) (AKBARZADEH et al., 2024; RAHMAN; SRIKUMAR; SMITH, 2018) follow, featuring in 10 studies. Other Gen-AI technologies are significantly less represented: Transformer Neural Networks (TNNs) (BARBIERATO; GATTI, 2024) appear in three papers, while Generative Adversarial Networks (GANs) (ESCORCIA-GUTIERREZ et al., 2024; MOZO et al., 2023) and Graph Neural Networks (GNNs) (IFTIKHAR et al., 2023) are each mentioned in only two. Variational Autoencoders (VAEs) (SUN, 2024) are cited in just one study. CNNs and RNNs account for approximately 80% of the Gen-AI technologies identified in energy management, reflecting an intense concentration in the literature on these more established neural architectures. This distribution also indicates a notable underrepresentation of more recent or specialised models. Table 1 summarises the AI-based approaches in the literature and their respective energy efficiency applications.

Table 1 – Gen-AI-based approaches to improving energy efficiency.

Authors	**Model utilised**	**Purpose**
Mozo et al. (2023)	Green AI	Integrates Machine Learning for security and employs Green AI to optimise models for reduced energy consumption while enhancing resilience against adversarial attacks through retraining with high-quality exemplars.
Zhao et al. (2023)	ARBiS	Developing ARBiS, a hardware-efficient CIM accelerator, addresses memory access pressure and hardware costs to enhance CNN acceleration efficiency and energy performance in AI edge devices.
Liu et al. (2022)	AIoT	Propose an AIoT-empowered ECCC system with FPGA-based CNN accelerators, reducing energy cost and execution time.
Sun (2024)	DMFF	Propose a Transformer-based multimodal, multi-temporal feature fusion method (DMFF) for superior occupancy detection, promising building energy savings.
Wang et al. (2024)	EDCompress	Propose EDCompress, an energy-aware model compression method for diverse edge device dataflows, optimises energy efficiency while maintaining accuracy and guiding CNN deployment on hardware.
Zawish et al. (2023)	Energy-Aware AI-Driven	Propose a deep reinforcement learning-based pruning scheme, adapting CNN complexity to energy constraints and enhancing energy efficiency without sacrificing accuracy.
Saptalakar and Latte (2023)	FPGA	Integrate IoT, data analytics, and energy-efficient CNN-based image optimisation for innovative city development, addressing reflection elimination and enhancing energy utilisation via FPGA deployment.
Akbarzadeh et al. (2024)	Green AI	Integrates Machine Learning for security and utilizes Green AI to optimise models for reduced energy usage. It safeguards against adversarial attacks through retraining with high-quality examples.
Iftikhar et al. (2023)	HunterPlus	Introduces HunterPlus, a novel approach to optimise cloud-fog task scheduling using Convolutional Neural Networks to reduce energy consumption per task and improve job completion rate.
Komala et al. (2023)	Multi-UAV	Integrate UAVs into IoT to collect sensor data efficiently using Q-learning and CNN-based route planning and reduce energy consumption.
Kaiser et al. (2023)	Neuromorphic-P2M	Propose and validate a novel energy-efficient, asynchronous, non-Von Neumann in-pixel processing paradigm for neuromorphic vision sensors, integrating analogue convolution operations to minimise computational energy while preserving high accuracy.
Sumathi et al. (2022)	NEWTR	Develop and evaluate an AI-based Tree Routing protocol using RNN and ZigBee in IoT environments, improving network lifetime and reducing energy consumption.

Deng et al. (2021)	PermCNN	Propose an efficient architecture for structured CNNs that enhances hardware energy efficiency.
Escorcia-Gutierrez et al. (2023)	AIBS-IoTH	Combining artificial intelligence and blockchain for energy-efficient and secure medical data transmission in IoT healthcare systems.
Barbierato and Gatti (2024)	Green AI	Discuss the ecological impact of Deep Learning, differentiating between Red AI, which consumes extensive energy, and Green AI, aimed at efficiency and sustainability.

Green AI stands out as the most frequently referenced paradigm in the literature. Introduced in 2020, the concept emerged in response to the growing concerns over the environmental footprint and rising computational demands associated with AI (SCHWARTZ et al., 2020). While conventional AI development has primarily focused on maximising predictive accuracy, often at the expense of energy efficiency, Green AI proposes a more balanced approach. It promotes the design and deployment of models that maintain high performance while significantly reducing computational resource consumption, thereby aligning technological advancement with environmental responsibility (ZHU; OTA; DONG, 2022).

The models identified in the literature are applicable across various domains, demonstrating the versatility of AI in supporting and enhancing energy management strategies. Key applications include telecommunication (MOZO et al., 2023; SUMATHI et al., 2022), Hardware (WANG et al., 2024b; ZHAO et al., 2023a), computer vision (DENG; LIAO; YUAN, 2021; LIU et al., 2022), smart city (AKBARZADEH et al., 2024; IFTIKHAR et al., 2023; SUN, 2024; ZAWISH et al., 2023), image processing (SAPTALAKAR; LATTE, 2023), environmental monitoring (KOMALA et al., 2023), autonomous vehicles (KAISER et al., 2023), healthcare systems (ESCORCIA-GUTIERREZ et al., 2024), and data analysis (BARBIERATO; GATTI, 2024).

2.2 DLMs and Gen-AI technologies used for energy management

Gen-AI refers to AI technologies capable of generating new content based on existing data and leveraging advanced models like neural networks (JACKSON et al., 2024; WAEL AL-KHATIB, 2023). Neural networks are parametric function approximators composed of layers of interconnected units that apply affine transformations followed by nonlinear activations and are trained via backpropagation to minimise a loss function; feed-forward networks process inputs independently, whereas recurrent architectures (e.g., RNNs/LSTMs) maintain a hidden state over time, which makes them suitable for modelling temporal dependencies typical of energy consumption series (HEWAMALAGE; BERGMEIR; BANDARA, 2021).

With technological advancements and progress in AI, Gen-AI has gained significant relevance, representing substantial breakthroughs in deep learning and unlocking numerous possibilities for intelligent design (LIAO et al., 2024). In this study, we leverage these properties by combining Holt–Winters for seasonality, an LSTM to learn non-linear residual dynamics, and a Gen-AI layer for configuration and interpretability. Thus, Gen-AI enables innovative applications, including developing and modelling energy forecasting models, as this research explores.

2.2.1 Generative Adversarial Networks (GANs)

The GAN is a generative model based on a game-theoretic interaction between two machine-learning models that facilitates the generation of samples from a target data distribution without the necessity of explicitly modelling the probability density function, comprising two neural networks trained simultaneously, the generator (G) and the discriminator (D) (GOODFELLOW et al., 2020). An illustration of the GAN concept is shown in Figure 1.

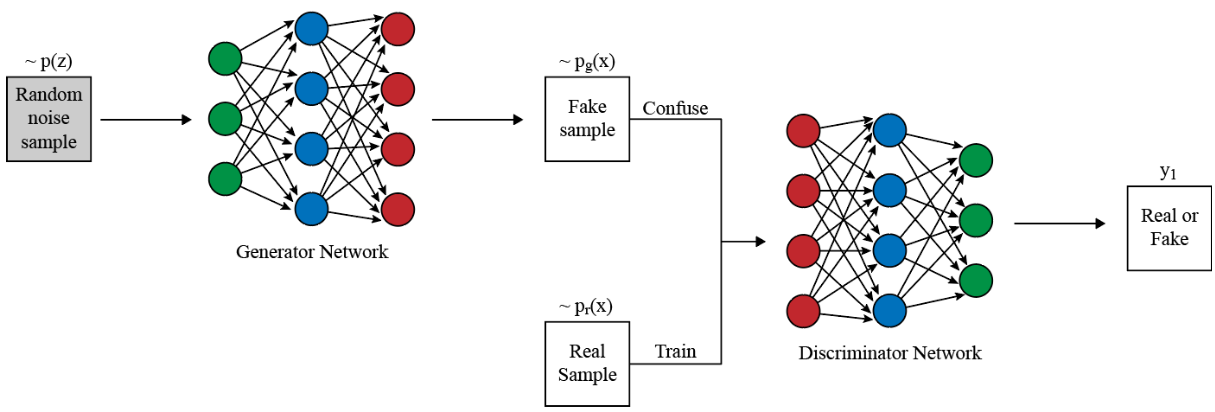


Figure 1 – Illustration of a GAN concept.

The generator, represented as $x_g = G\ (z;\ \theta_g)$, receives a random noise sample (z) from a prior distribution $p(z)$ and produces an output (x_g) intended to resemble real data samples (x_r), from the actual data distribution $p_r(x)$ (YI; WALIA; BABYN, 2019). The discriminator's function is expressed as $y_1 = D\ (x;\ \theta_d)$, and its role is to differentiate between real samples and those generated by the generator, wherein it accepts either real or generated samples as input and outputs a single value (y_1), indicating the likelihood that the input is real or fake

(GOODFELLOW et al., 2020). The ultimate goal is for the generated sample distribution, $p_g(x)$, to closely approximate the real data distribution, $p_r(x)$, after the training process (WOLTERINK et al., 2017). The training objectives for the discriminator and generator can be represented mathematically according to Equation 1:

$$\mathcal{L}_D^{GAN} = \max_D \mathbb{E}_{x_r \sim p_r(x)}[log\, D(x_r)] + \mathbb{E}_{x_g \sim p_g(x)}[log(1 - D(x_g))], \\ \mathcal{L}_G^{GAN} = \min_G \mathbb{E}_{x_g \sim p_g(x)}[log(1 - D(x_g)] \tag{1}$$

The interaction between the generator and the discriminator can be analogised to a forger and a detective, respectively, where the generator strives to create data that can deceive the discriminator, while the discriminator aims to identify the generated data as fake accurately (SHORTEN; KHOSHGOFTAAR, 2019). This adversarial process results in the generator receiving feedback from the discriminator in gradient information, which it uses to adjust its parameters and improve its output; consequently, the generator continually refines its ability to produce realistic data, achieving a state where the generated samples are indistinguishable from real data (YI; WALIA; BABYN, 2019).

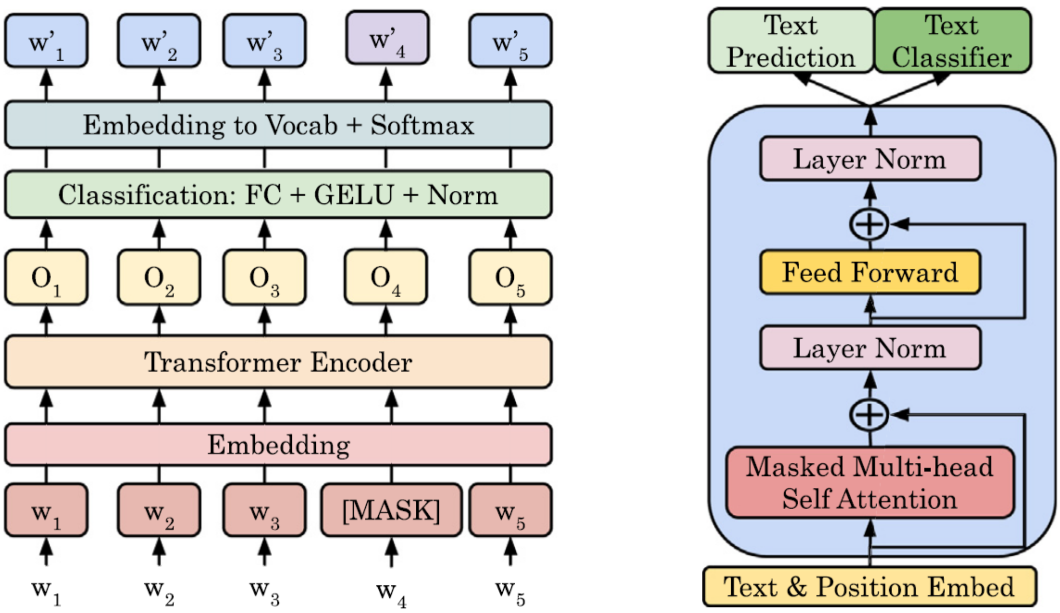
2.2.2 Transformer Neural Networks (TNNs)

TNNs are a class of deep neural network architectures that leverage multiple sequential self-attention layers to process and capture relationships within input sequences efficiently (SCHWALLER et al., 2021). LMs leverage Transformer architectures for their self-attention mechanisms, facilitating efficient text processing and generation. This enables them to capture long-range dependencies, train autoregressively by predicting the next token in a sequence, and scale effectively, resulting in emergent capabilities such as zero-shot learning (TRUHN et al., 2024).

LMs are not a new technology and have a long history in the AI domain, with ELIZA in 1966 being the first language created for natural language processing, which used scripts to simulate human conversations (SHAH et al., 2016; WEIZENBAUM, 1966). Traditionally, text generation is handled statistically by sampling each token, characterising it as a character, word, or part of a word, and based on the tokens that precede it; in other words, in this statistical perspective, the task of modelling or learning involves adjusting the parameters based on the training data, where the goal is to optimise these parameters to maximise the likelihood of each token in the training data (ROSENFELD, 2000).

In text generation, modern LMs leverage deep neural networks to learn patterns within data, where these networks process a sequence of tokens and predict the probabilities for each potential next token (REEDHA et al., 2022). Recent advancements in deep generative neural networks have sparked significant research into how humans might collaborate with AI systems driven by these networks, especially in creative endeavours (LIN et al., 2024). Extensive empirical evidence suggests that sufficiently LLMs can transcend simple memorisation, demonstrating impressive creativity and intelligence, such as effectively handling novel concepts not part of the training data but introduced in the prompt (CHEN et al., 2020; SCHWALLER et al., 2021).

The introduction of TNN has transformed the landscape of Natural Language Processing (NLP) in recent years. The TNN architecture, introduced in 2017, is defined by its use of encoder and decoder blocks and a self-attention mechanism, in which encoder blocks convert variable-length input data into fixed-size feature maps, while decoder blocks aim to reconstruct the original input from these feature maps (VÄRTINEN; HÄMÄLÄINEN; GUCKELSBERGER, 2024). Several large models based on Transformer architecture have been recently developed, with Generative Pre-trained Transformers (GPT) and Bidirectional Encoder Representations from Transformers (BERT) being among the most prominent (CHITTY-VENKATA et al., 2023). Figure 2 shows BERT and GPT architectures.



(a) BERT architecture (b) GPT architecture

Figure 2 – BERT and GPT architectures (CHITTY-VENKATA et al., 2023).

The BERT model, shown in Figure 2a, uses only the encoder component from the original Transform architecture and can predict missing words in a sentence by considering both the preceding and following context in the input sequence, which is why it is called bidirectional (MIOK et al., 2022). In contrast, the GPT model shown in Figure 2b is an LLM pre-trained on diverse text data unsupervised to perform prediction tasks, wherein only the decoder component is used, which includes positional encoding, Multi-head self-attention (MHA), a Pointwise Feed Forward Network (FFN), and normalisation (CHITTY-VENKATA et al., 2023). TNNs are generally designed to predict the next word in a sequence based on preceding words or prompts (SCHWALLER et al., 2021). Transformer models have demonstrated their effectiveness across various complex tasks, including generating music, images, and texts, which are highly realistic and similar to human creations (VÄRTINEN; HÄMÄLÄINEN; GUCKELSBERGER, 2024).

2.2.3 Recurrent Neural Networks (RNNs)

The advancements in deep learning led to the RNN model, which John Hopfield introduced in 1982 (ZHANG; WANG; LIU, 2014). The RNN is a versatile class of neural networks which have been utilised in research areas focused on sequential data, with the main characteristic being cyclic connections, allowing it to update its current state based on previous states and current input data, having the capability to learn and solve problems with potentially unlimited depth (ZHAN et al., 2021). Its models provide an effective mechanism for time series modelling because RNNs can be trained discriminatively (JAEGER; HAAS, 2004). Figure 3 shows the basic architectural structure of RNNs.

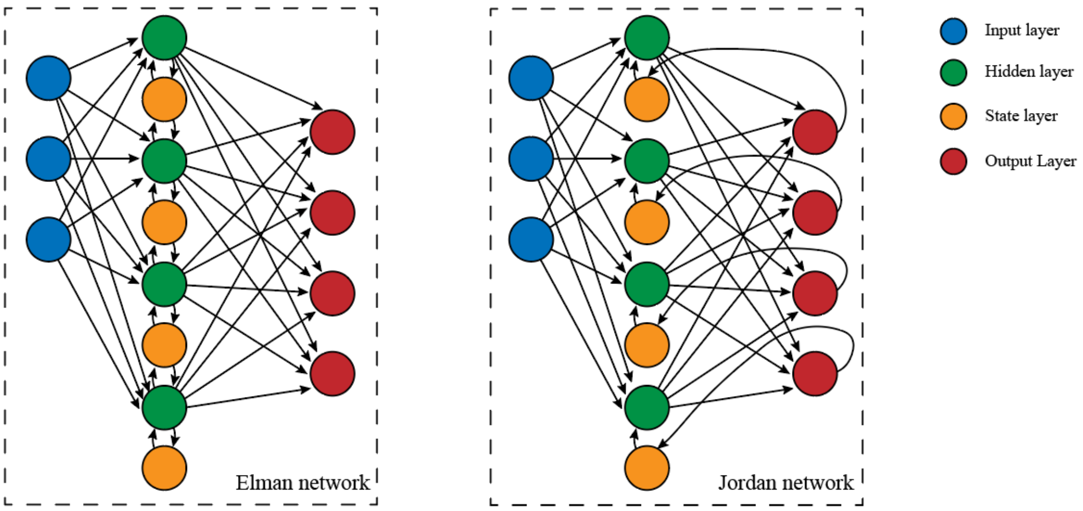


Figure 3 – The basic architectural structure of RNNs.

Figure 3 illustrates two models of RNNs, the first created by Jeffrey Elman in 1990 and the second created by Michael I. Jordan in 1986 (MACNAMARA; CUNNINGHAM; BYRNE, 1998). The Elman neural network maintains a memory of previous hidden layer states through its recurrent connections, allowing it to capture temporal dependencies in the data, while in the Jordan network, the recurrent connections receive input from the output layer's posterior probabilities (MESNIL et al., 2015). In RNNs, recurrent or hidden layers consist of cells influenced by previous states and current inputs through feedback connections, with the architecture of these layers varying and resulting in different types of RNNs distinguished by their recurrent cell design and network structure, enabling RNNs to possess diverse capabilities (HEWAMALAGE; BERGMEIR; BANDARA, 2021). Equation 2 presents the RNN mathematical representation of the output layer, while Equations 3 and 4 describe the hidden layer in the Elman and Jordan models, respectively.

$$y_t = \emptyset(W_{hy}h_t + b_y) \tag{2}$$

Where:

y_t is the output at time step *t*.

ϕ is the activation function.

W_{hy} is the weight matrix connecting the hidden state to the output.

h_t is the hidden state at time step *t*.

b_y is the bias term.

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \tag{3}$$

Where:

h_t is the hidden state at time step *t*.

σ is the activation function.

W_{hx} and W_{hh} are weight matrices.

x_t is the input at time step *t*.

b_h is the biased term.

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + W_{hy}y_{t-1} + b_h) \tag{4}$$

Where:

h_t is the hidden state at time step t.

σ is the activation function.

W_{hx}, W_{hh}, and W_{hy} are weight matrices.

y_{t-1} is the output from the previous time step.

x_t is the input at time step t.

y_t is the output at time step t.

b_h is the biased term.

Several RNN-based models have been developed to advance the capabilities of their predecessors. Among their variants, two widely adopted extensions are the Long Short-Term Memory (LSTM) and the Gated Recurrent Unit (GRU). The LSTM networks address the vanishing gradient problem of RNNs through gated cell structures, enabling the model to retain or discard information across time steps selectively (HOCHREITER; SCHMIDHUBER, 1997). This architecture allows LSTMs to effectively model long-term dependencies and complex consumption patterns, making them highly suitable for forecasting tasks involving delayed or cumulative effects (KONG et al., 2019). However, LSTM models demand considerable computational power and extended training times. Their architectural complexity and interpretability challenges necessitate advanced technical expertise for deployment. Consequently, while LSTMs offer superior accuracy, their adoption requires careful evaluation of trade-offs involving computational cost, implementation complexity, and operational timelines (DIVYANI SEN, 2025).

The GRUs are a recurrent neural network architecture developed to address the vanishing gradient problem commonly encountered in traditional RNNs. GRUs simplify the gating mechanism introduced in LSTM networks by combining the forget and input gates into a single update gate, resulting in a more streamlined architecture (XIA et al., 2021). Their capacity to model long-term temporal dependencies with reduced computational complexity makes GRUs suitable for forecasting energy consumption in dynamic and behaviourally variable environments (ZHAO et al., 2023b). Despite offering a computationally lighter alternative to LSTMS, GRUS may underperform in datasets with highly complex and long-range temporal patterns. Their simplified structure, while efficient, can sometimes limit the granularity with which intricate temporal relationships are captured (GAO et al., 2024). Furthermore, like other deep learning models, GRUs require careful tuning of hyperparameters and training conditions to achieve robust generalisation, which can be resource-intensive in real-world forecasting applications.

2.2.4 Convolutional Neural Networks (CNNs)

CNNs are the most well-known and widely used deep-learning algorithms (KRIZHEVSKY; SUTSKEVER; HINTON, 2017; LECUN; BENGIO; HINTON, 2015) and were modelled inspired by the structure of the visual system (HUBEL; WIESEL, 1962), which consists of three primary types of neural layers: convolutional layers, pooling layers, and fully connected layers, each serving a distinct function (VOULODIMOS et al., 2018; YAMASHITA et al., 2018). Their primary advantage is their ability to automatically detect and learn relevant features without the need for human intervention (ALZUBAIDI et al., 2021). In basic CNN architectures, as shown in Figure 4, generic object detection is achieved through bounding box regression, while salient object detection uses local contrast enhancement and pixel-level segmentation; that is, each layer is called a feature map, and the input layer is a 3D matrix representing pixel intensities across different colour channels (ZHAO et al., 2019).

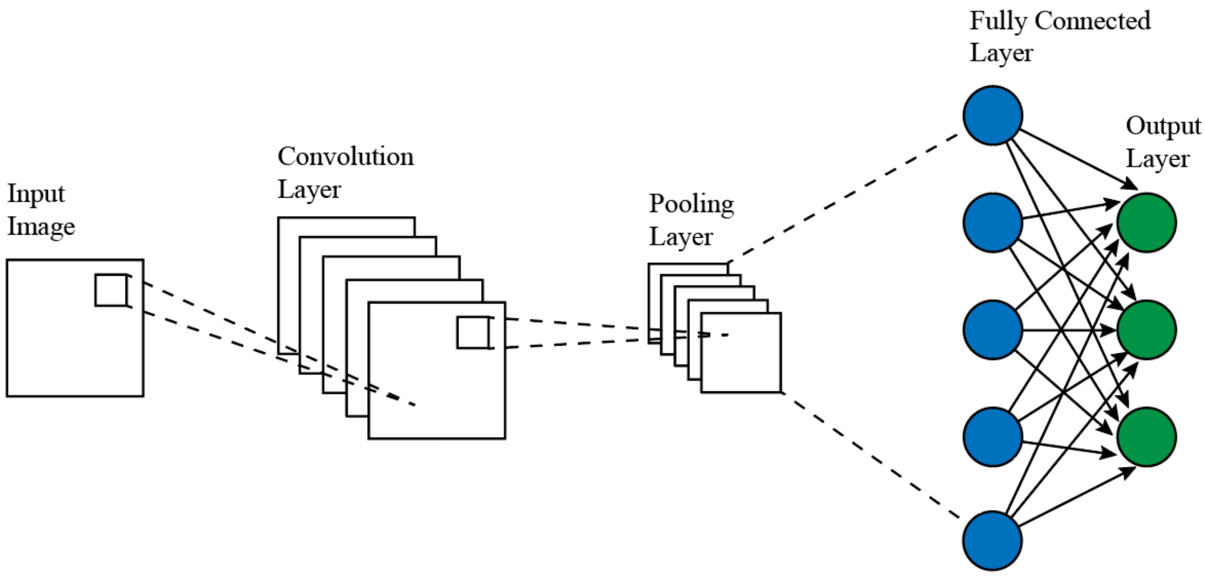


Figure 4 – The basic CNN architecture.

A convolutional layer is a critical element of CNN architecture responsible for feature extraction and typically involves a combination of linear and non-linear operations, namely the convolution operation and an activation function (YAMASHITA et al., 2018). Pooling layers are responsible for sub-sampling feature maps produced after convolutional operations, reducing the size of large feature maps, creating smaller, more compact feature maps

(ALZUBAIDI et al., 2021). In fully connected layers, each neuron is connected to all activations from the previous layer, allowing the layer to compute activations using matrix multiplication followed by a bias offset, allowing the transformation of the 2D feature maps into a 1D feature vector, which can then be classified into specific categories or as a feature vector for further processing (VOULODIMOS et al., 2018).

In each layer, the input image is convolved with a set of K kernels W = {W_1, W_2, …, W_K} and corresponding biases B = {b_1, …, b_K}, producing a set of new feature maps X_k (LITJENS et al., 2017). These feature maps are then passed through an element-wise non-linear transformation σ and this process is repeated for each convolutional layer l (VOULODIMOS et al., 2018), as shown in Equation 5.

$$X_k^l = \sigma(W_k^{l-1} * X^{l-1} + b_k^{l-1}) \tag{5}$$

Where:

X_k^l is the output feature map.

σ is the activation function.

W_k^{l-1} is the filter or kernel.

X^{l-1} is the input feature map.

$*$ is the convolution operation between the filter and the input feature map.

b_k^{l-1} is the bias term.

2.2.5 Variational Autoencoders (VAEs)

VAEs are described as a type of autoencoder comprising an encoder and a decoder used for unsupervised learning, which transforms high-dimensional data into a lower-dimensional latent representation that conforms to a Gaussian distribution (SHRESTHA; MAHMOOD, 2019), as shown in Figure 5. VAEs are a structured framework for learning deep latent variable models and their corresponding inference models, integrating directed probabilistic graphs and neural networks, where the generative model is a conditional Bayesian network that can be represented as a hierarchy of stochastic latent variables (KINGMA; WELLING, 2019).

By encoding input data into a latent space, the encoder learns to map inputs to this latent space, capturing the essential features in a compressed form, while the decoder reconstructs the input data from this latent space, allowing the model to generate new, similar instances by sampling from the latent distribution (MOHAMMADI et al., 2018; NISSEN et al., 2021). The

VAE aims to reduce reconstruction error and Kullback-Leibler divergence, facilitate the effective learning of latent variable models, and create new data samples that closely mirror the original dataset (ASPERTI; TRENTIN, 2020; HUSSAIN; MICHELUSI, 2022).

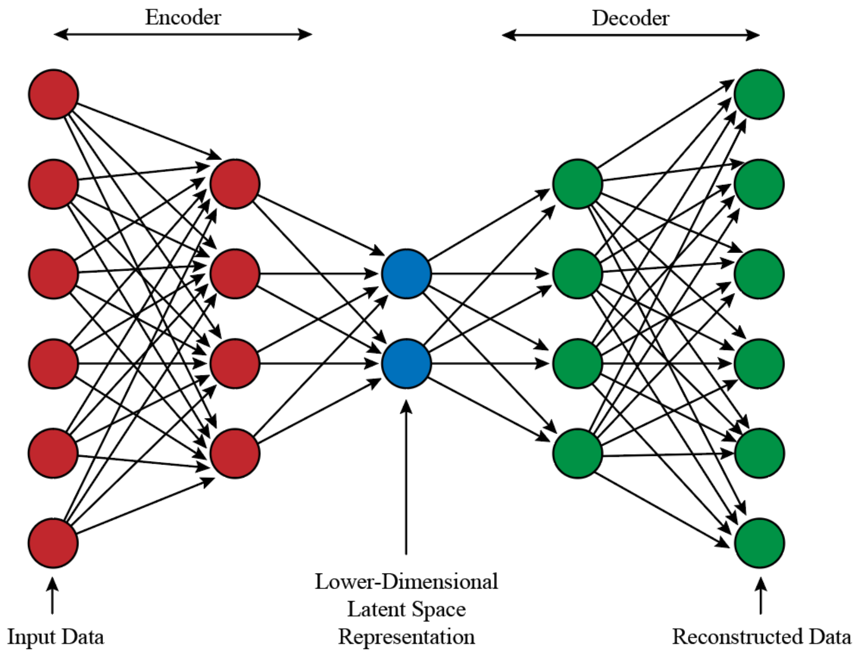


Figure 5 – The basic VAE architecture.

The VAE is regarded as a promising approach for modelling complex data, where efficient learning of latent distributions is essential (HOSNY et al., 2018). The mathematical representation of a VAE revolves around optimising the Evidence Lower Bound (ELBO) and can be expressed as Equation 6. The critical contribution of VAEs is their ability to perform efficient approximate inference using a reparameterisation trick, which reduces the variance in the gradients required for learning (KINGMA; WELLING, 2019).

$$\mathcal{L}_{\theta,\emptyset}(x) = \mathbb{E}_{q\emptyset(z|x)}[log\, p_\theta(x|z)] - D_{KL}(q_\emptyset(z|x)||p_\theta(z)) \tag{6}$$

Where:

$\mathcal{L}_{\theta,\emptyset}(x)$ represents the ELBO.

$p_\theta(x|z)$ represents the likelihood of the data given latent variables z.

$q_\emptyset(z|x)$ is the approximate posterior distribution over the latent variables z given the observed data x.

$p_\theta(z)$ denotes the prior distribution over the latent variables z.

D_{KL} denote the Kullback-Leibler divergence between the approximate posterior distribution and the prior distribution.

2.2.6 Graph Neural Networks (GNNs)

In recent years, there have been significant advancements in deep neural network techniques for processing graph data and have been extensively adopted in the field of computer vision (MA et al., 2024). A GNN is a deep learning model designed to process and analyse graph-structured data, representing entities as nodes and the relationships between them as edges (JIANG et al., 2021; WU et al., 2021). Unlike traditional neural networks that work with grid-like data such as images or sequences, GNNs are built to leverage the connectivity and structure of graphs (BATZNER et al., 2022; HOGAN et al., 2022). An example of the graphs is illustrated in the diagram shown in Figure 6.

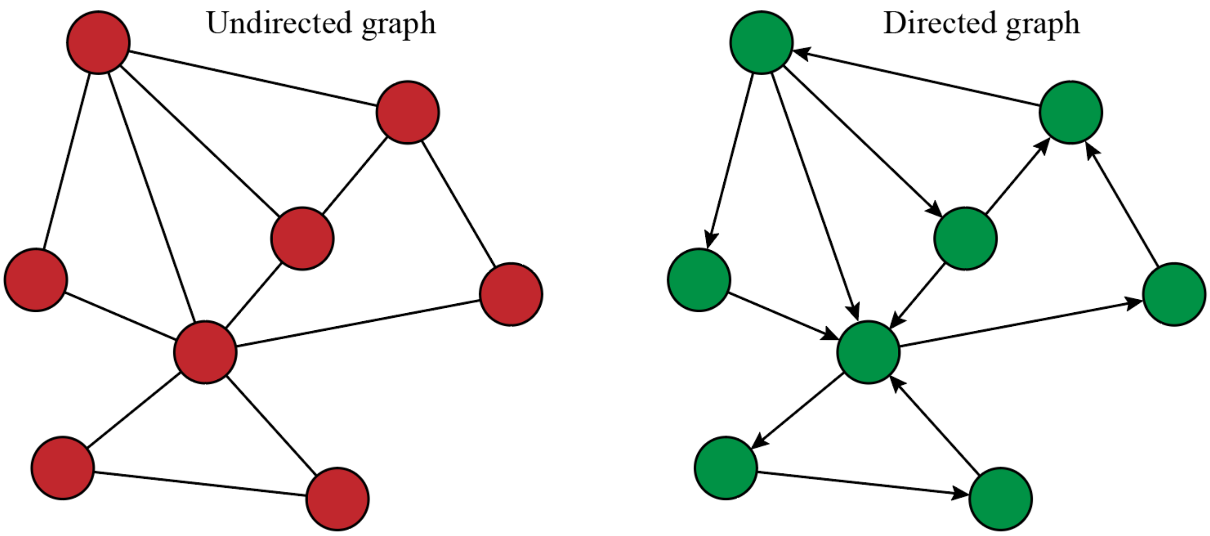


Figure 6 – Undirected and directed graph.

As shown in Figure 6, a directed graph is a type of graph in which all edges have a specific direction from one node to another, and an undirected graph can be viewed as a particular case of a directed graph, where each pair of connected nodes has two edges pointing in opposite directions (WU et al., 2021). A graph is undirected if and only if its adjacency matrix is symmetric (SCARSELLI et al., 2009). The GNNs are particularly powerful for tasks that involve understanding and leveraging the relationships and structures within graph data (SCARSELLI et al., 2009).

The core idea behind GNNs is to iteratively update the node representations by aggregating information from their neighbours, making them particularly effective for tasks such as node classification, graph classification, and link prediction (LI et al., 2022). GNNs can be categorised into four main types: Recurrent Graph Neural Networks, Convolutional Graph Neural Networks, Graph Autoencoders, and Spatial-Temporal Graph Neural Networks (WU et al., 2021). Assuming a graph $G = (V, E)$, the mathematical representation of a general GNN can be expressed by separating it into functions called AGGREGATE and COMBINE, as shown in Equation 7:

$$a_v^{(k)} = AGGREGATE^{(k)}\left(\left\{h_u^{(k-1)} : u \in N(v)\right\}\right), h_v^{(k)} = COMBINE^{(k)}\left(h_v^{(k-1)}, a_v^{(k)}\right) \quad (7)$$

Where:

$a_v^{(k)}$ is the aggregated information from the neighbours of node *v* at the *k-th* layer.

$h_v^{(k)}$ is the feature vector of node *v* at the *k-th* layer.

$h_u^{(k)}$ is the feature vector of the neighbouring node *u* at the *k-th* layer.

$N(v)$ is a set of nodes adjacent to *v*.

$AGGREGATE^{(k)}$ is the aggregation function at the *k-th* layer.

$COMBINE^{(k)}$ is the combination function at the *k-th* layer.

2.3 Classical statistical forecasting method

Energy consumption forecasting has traditionally employed classical statistical methods, notably the Holt-Winters exponential smoothing technique (DURMUS SENYAPAR; AKSOZ, 2024), the Seasonal Autoregressive Integrated Moving Average (SARIMA) model (NGO et al., 2022), and the Structural Time Series Model (STSM) (HARROU et al., 2021). More recently, the Prophet model has also gained attention for its modular formulation and ease of use (CHATURVEDI et al., 2022).

2.3.1 Holt-Winters exponential smoothing

Statistical models have been extensively applied in electricity consumption forecasting due to their simplicity, interpretability, and low computational demands (KIM; PARK; KIM, 2023; SERRANO et al., 2024; YOUSEFI; ARDEHALI; GHODUSINEJAD, 2023). Among these, the Holt-Winters exponential smoothing method is one of the most prominent, particularly in contexts involving seasonal time series (GAO et al., 2023; HUSSAIN; RAHMAN; MEMON, 2016; KHOT et al., 2023). Its capacity to capture level, trend and seasonal components makes it especially suitable for modelling monthly energy demand, which often exhibits regular consumption patterns.

The method was developed through the combined contributions of Charles C. Holt and Peter R. Winters. In 1957, Holt introduced a generalisation of exponential smoothing capable of modelling linear trends, laying the groundwork for more advanced time series techniques (HOLT, 2004). Three years later, in 1960, Winters extended this formulation by incorporating a seasonal component, resulting in a comprehensive model for forecasting data with both trend and seasonality (WINTERS, 1960).

What distinguishes Holt-Winters in operational contexts is its balance between simplicity and performance. It can be implemented quickly, delivers interpretable results, and often achieves forecasting accuracy comparable to that of more complex machine learning models, especially in short- to medium-term scenarios (DEKKER; VAN DONSELAAR; OUWEHAND, 2004; VEIGA et al., 2016). These characteristics make it a practical and widely used energy management tool where precision and efficiency are essential for real-time decision-making.

The Holt-Winters exponential smoothing method is available in two main formulations: additive and multiplicative. In the additive version, the seasonal component is assumed to have a constant absolute effect on the series over time, while the multiplicative version assumes that the seasonal influence varies in proportion to the level of the series (HYNDMAN et al., 2002). The choice between these formulations depends on the nature of the seasonal fluctuations. The multiplicative model is often more appropriate in energy consumption forecasting when seasonal variations scale proportionally with the level of demand.

This is particularly relevant in contexts where higher consumption periods are associated with amplified seasonal effects, such as during peak summer or winter. Unlike the additive model, where the seasonal component is assumed to have constant magnitude, the multiplicative formulation captures dynamic seasonal amplitudes relative to the overall consumption level. The mathematical representation of the multiplicative model is expressed through Equation 8 for forecasting, Equation 9 for level, Equation 10 for trend, Equation 11 for seasonality, and Equation 12 for the auxiliary term.

$$\hat{y}_{t+h|t} = (\ell_t + hb_t) \times s_{t+h-m(k+1)} \tag{8}$$

$$\ell_t = \alpha\left(\frac{y_t}{s_{t-m}}\right) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \tag{9}$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \tag{10}$$

$$s_t = \gamma\left(\frac{y_t}{\ell_t}\right) + (1 - \gamma)s_{t-m} \tag{11}$$

$$k = \left[\frac{h - 1}{m}\right] \tag{12}$$

Where:

$\hat{y}$ is the actual observed value at time t.

$\hat{y}_{t+h|t}$ is the forecast for time $t+h$ made at time t.

ℓ_t is the level component.

b_t is the trend component.

s_t is the seasonal component.

α, β, γ is the smoothing parameter.

m is the seasonal period.

k is the integer part of $\frac{h-1}{m}$.

2.3.2 Seasonal Autoregressive Integrated Moving Average (SARIMA)

The SARIMA model extends the classical ARIMA framework by explicitly incorporating seasonal dependencies into time series forecasting (KARAKAN, 2024). It captures both seasonal and non-seasonal dynamics by incorporating autoregressive terms, differencing operations, and moving averages, which makes it well-suited for modelling energy consumption data characterised by cyclical yearly behaviours (DURMUS SENYAPAR; AKSOZ, 2024). SARIMA provides a statistically rigorous and structured methodology, offering interpretability and formal parameter estimation (SISUTHOG; ATTANATHO; CHAIYA, 2022).

Despite its flexibility compared to non-seasonal ARIMA, SARIMA remains limited in capturing complex non-linear dynamics inherent in modern energy systems (AZIMI; HOOSHMAND; SOLEYMANI, 2021). Its dependence on linear relationships and the requirement for meticulous parameter calibration may limit its predictive accuracy and adaptability in environments marked by high volatility or human-influenced consumption patterns.

2.3.3 Prophet

The Prophet model, developed by Facebook's Core Data Science team, is a forecasting technique designed to handle time series data exhibiting multiple seasonality, trend changes, and holiday effects (BASHIR et al., 2022). Prophet decomposes time series into additive or multiplicative structures, allowing trend, seasonality, and external factors to be modelled separately. Its adaptability, resilience to missing values and anomalies, and low dependency on manual parameter tuning have contributed to its widespread adoption for short- and medium-term forecasting across various fields, including energy consumption analysis (SERRANO et al., 2024).

Despite its versatility, Prophet relies on predefined functional forms for trend and seasonality, which can constrain its capacity to capture highly irregular or purely stochastic behaviours in complex energy systems (WANG et al., 2024a). Furthermore, although Prophet can capture certain non-linear behaviours through changepoint detection, its performance may fall short when compared to deep learning approaches in highly dynamic or behaviour-driven energy usage scenarios. Even so, its high interpretability, straightforward implementation, and automated treatment of seasonality and holiday effects render it a practical option, especially in operational settings that demand fast and explainable forecasting outcomes.

2.3.4 The Structural Time Series Model (STSM)

The STSM separates a time series into distinct components, trend, seasonality, and irregular variation, while explicitly incorporating stochastic behaviour into each element. Its ability to accommodate deterministic and stochastic trends makes it a versatile framework for energy forecasting. Empirical research has shown that STSM performs effectively in modelling and forecasting electricity consumption across different sectors (DILAVER; HUNT, 2011). By uncovering the Underlying Energy Demand Trend (UEDT), a latent variable representing structural shifts in demand patterns, STSM facilitates more nuanced interpretations of energy consumption dynamics over time (FATIMA; XIA; AHAD, 2019).

The STSM enables the isolation of long-term behavioural changes from cyclical fluctuations, particularly relevant in settings affected by policy interventions, economic restructuring, or evolving consumer habits (CHITNIS; HUNT, 2012). Despite its strengths in providing transparency and interpretability through the explicit modelling of trend and seasonal components, the STSM may exhibit reduced performance in environments characterised by fast-evolving, non-linear behaviours or scenarios requiring the assimilation of high-dimensional data (HARROU et al., 2021).

2.4 Theoretical framework

This section explores the theoretical foundations of Green AI and Paradox Theory, which serve as complementary lenses to frame the development and evaluation of energy forecasting models. Green AI advocates for resource-efficient and environmentally conscious AI practices, emphasising the importance of reducing computational costs without compromising utility. Paradox Theory, in turn, provides a conceptual structure to navigate the tensions between competing demands, in this case, predictive performance versus sustainability. Together, these frameworks underpin the proposed approach, guiding both methodological choices and the interpretation of results.

2.4.1 Green AI theory

Green AI represents a paradigm shift in developing and deploying AI systems, placing environmental sustainability at the core of algorithmic design and implementation. The theory emerged in response to the exponential growth in computational demands associated with machine learning and generative models, particularly those relying on extensive data processing and energy-intensive infrastructures (ELMOUSALAMI; ELSHABOURY; ELYAMANY, 2024). Modern AI models, especially LLM, have seen their energy consumption grow faster than Moore’s Law, resulting in significant carbon emissions during training and inference phases (THEIS; WONG, 2017).

The central purpose of Green AI is to mitigate the environmental impacts of AI technologies without compromising their predictive capabilities (ALZOUBI; MISHRA, 2024). This is achieved through optimising algorithms, selecting energy-efficient models, and refining hardware architectures. Green AI aims to reduce the carbon footprint of AI by incorporating practices such as energy-aware training, computational resource management, and using renewable energy sources in data centres (RAMAN et al., 2024).

Among the objectives of Green AI is the promotion of energy efficiency in AI pipelines, which involves selecting algorithms with lower computational overhead, deploying lightweight model architectures, and applying efficient optimisation strategies (ELMOUSALAMI;

ELSHABOURY; ELYAMANY, 2024). The Green AI theory underscores the need to rethink traditional performance-centric approaches in AI development. By embedding sustainability principles into the core of algorithmic design and evaluation, it enables the creation of intelligent systems that are powerful, accurate, environmentally conscious, and operationally sustainable (ALZOUBI; MISHRA, 2024).

2.4.2 The Paradox theory

The Paradox theory offers a theoretical framework to comprehend and engage with persistent contradictions and tensions within organisational contexts. Rather than seeking linear or singular solutions, this perspective encourages organisations to manage complexity by embracing opposing forces as interdependent and coexisting elements (CARTER; KAUFMANN; KETCHEN, 2020; CUNHA; PUTNAM, 2019). From an organisational standpoint, paradoxes are understood as conflicting yet interconnected dimensions that remain present over time (SMITH; LEWIS, 2011). This theoretical approach diverges from traditional strategy paradigms that rely on binary logic, such as choosing between cost leadership or differentiation, prioritising either exploitation or exploration, or managing relationships through either trust or control mechanisms (KOHTAMÄKI; EINOLA; RABETINO, 2020; SMITH; LEWIS, 2011). The Paradox theory challenges this dichotomous reasoning by proposing that many strategic dilemmas cannot be resolved through exclusive choices but instead require the integration of competing demands to improve performance and adaptability (JAY, 2013; SCHREYÖGG; SYDOW, 2010).

Existing research has highlighted how paradoxical tensions influence decision-making across diverse areas, including social impact, financial performance, and innovation (SMITH; GONIN; BESHAROV, 2013). These tensions are particularly visible in hybrid organisations, where commercial objectives may conflict with social missions. In such cases, pursuing success in one domain may appear to compromise the other, and role ambiguity may also arise, such as when board members experience friction between participatory and directive responsibilities (JAY, 2013). With the increasing pace of technological advancement, the Paradox theory has gained prominence as a relevant lens for understanding operational dilemmas in contemporary organisations. Companies often face the challenge of remaining innovative to sustain competitiveness while striving for efficiency to maintain profitability. In this context, paradox theory encourages a mindset that acknowledges and manages these contradictions constructively. Rather than eliminating tensions, it seeks to harness them as drivers of innovation and strategic renewal (CUNHA; PUTNAM, 2019; KOHTAMÄKI; EINOLA; RABETINO, 2020).

3 METHODOLOGY

The methodological design of this study comprises six sequential and complementary stages, adopting a mixed-methods approach that combines qualitative and quantitative techniques. The research is characterised as exploratory (COOPER, 2023), descriptive (KUMAR; GUPTA; DAGAR, 2024), and explanatory (HAMIDU; BOACHIE-MENSAH; ISSAU, 2023) and follows a predominantly inductive logic (GRIMELL, 2023) with deductive elements (PETIT et al., 2023) in the model validation phase. The first stage involved a scoping review (DEIVANAYAGAM et al., 2023) to map the existing literature on DLMs and Gen-AI technologies applied to energy management, identifying research gaps and formulating research questions. In the second stage, a structured questionnaire (GARCIA-BUENDIA et al., 2022) was developed to gather expert opinions on the relevance of six DLM technologies for energy efficiency. The survey was conducted with thirty selected experts from academic and professional domains.

In the third stage, non-parametric statistical analysis was conducted using the Friedman test (REN et al., 2023) to identify significant differences in expert rankings, followed by the Nemenyi post hoc test (LU et al., 2023) to explore pairwise comparisons. These results supported the fourth stage, which focused on designing the GenEneCast hybrid forecasting model combining statistical decomposition, DLM, and Gen-AI techniques to improve the accuracy and efficiency of monthly electricity consumption predictions. In the fifth stage, an action research strategy was implemented (CORNISH et al., 2023). The model was applied in an organisational setting using real energy consumption data. Forecasts were evaluated for accuracy and practical relevance, allowing iterative model refinement. This stage reinforced the study’s contribution to bridging theory and practice in energy management. The sixth stage employed the MCS methodology to statistically compare the predictive performance of GenEneCast against classical and neural network benchmarks, identifying models that belong to the set of superior forecasters with statistical confidence. Figure 7 presents an overview of the research design.

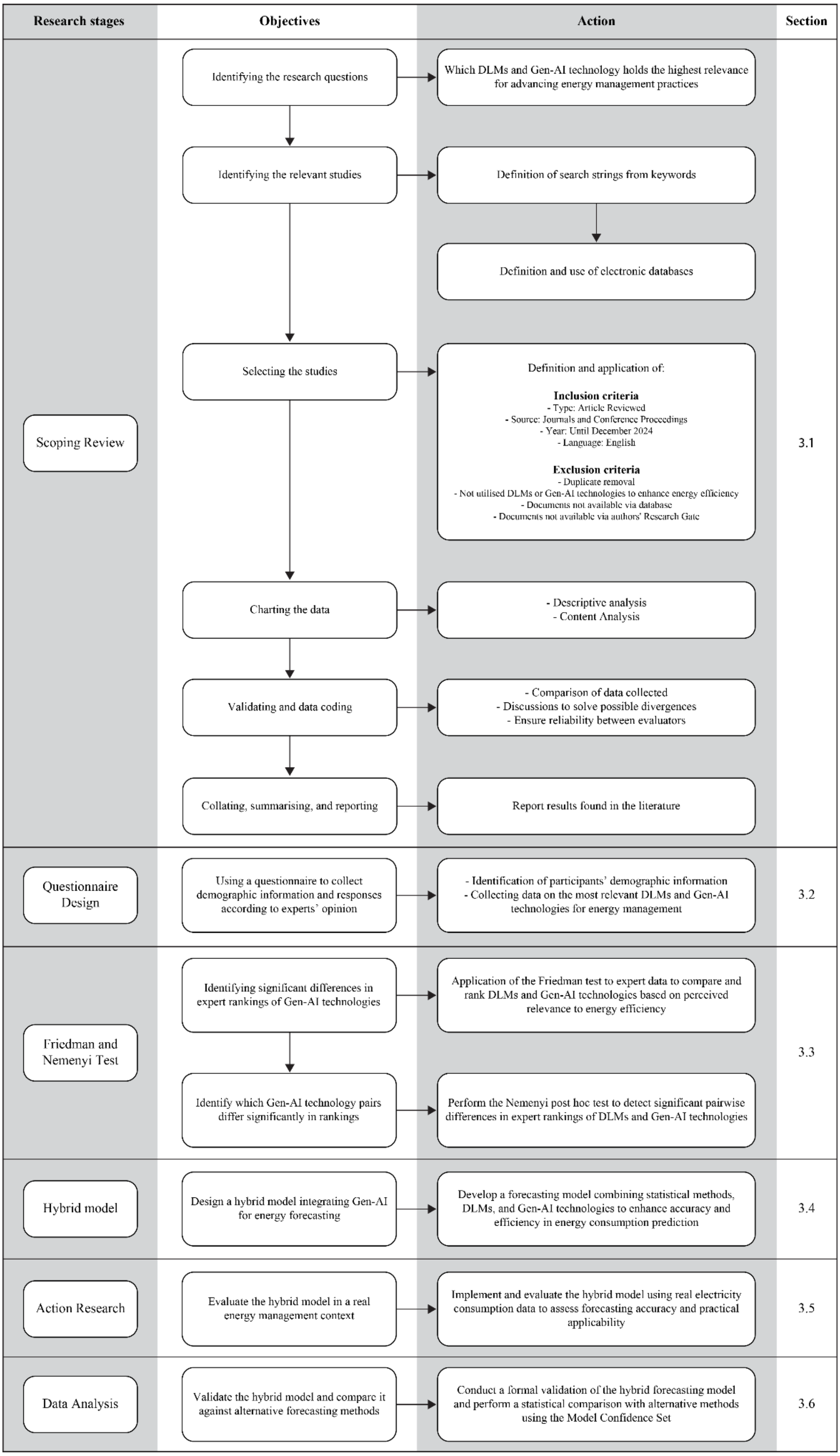


Figure 7 – Research design overview.

3.1 Scoping review

This study adopted a scoping review to identify and map key DLMs or Gen-AI technologies applied to energy efficiency to highlight their potential contributions to organisational energy management. Scoping reviews are particularly valuable for exploring the breadth and depth of available literature on a subject. They offer an overview of the volume, nature and range of existing studies, enabling broad mapping and more focused thematic exploration. This methodological approach also supports the identification of emerging trends and areas of concentrated research interest (ARMSTRONG et al., 2011; MUNN et al., 2018). Furthermore, scoping reviews are instrumental in uncovering gaps in the literature that may benefit from more rigorous investigation through systematic reviews or empirical studies (NASCIMENTO et al., 2022).

The primary purpose of a scoping review is to identify and synthesise existing evidence across a broad and often heterogeneous body of literature (ANDERSON et al., 2008), offering a structured overview of complex or emerging themes (DI PASQUALE; MIRANDA; NEUMANN, 2020). This methodological approach has been increasingly adopted in research involving AI, particularly in fields such as business (KANBACH et al., 2023), healthcare (MARTINS et al., 2024), and education (YAN et al., 2024), among others. The present study followed the framework proposed by Arksey and O'Malley (2005), and incorporating an additional step called validating and data coding (DANESE; MANFÈ; ROMANO, 2018). The procedure consisted of six stages: (i) identifying the research questions, (ii) identifying the relevant studies, (iii) selecting the studies, (iv) charting the data, (v) validating and data coding, and (vi) collating, summarising, and reporting. To ensure transparency and reproducibility, it is essential to document explicitly all methodological procedures employed throughout the review process (FONTOURA et al., 2023).

A combination of interdisciplinary discussions among the authors, structured brainstorming sessions, and a comprehensive literature analysis was conducted to address the first research question. These efforts sought to develop a well-rounded understanding of state-of-the-art DLMs or Gen-AI technologies within the context of energy efficiency. A literature review was conducted to establish the theoretical grounding for the study (SAIEG et al., 2018) and to identify gaps in current research (WEBSTER; WATSON, 2002). A structured search string was developed to support identifying relevant studies that aligned with the review's scope. This process was further enhanced by using a word tree, which visually summarises the main terms and concepts extracted from the literature, as illustrated in Figure 8.

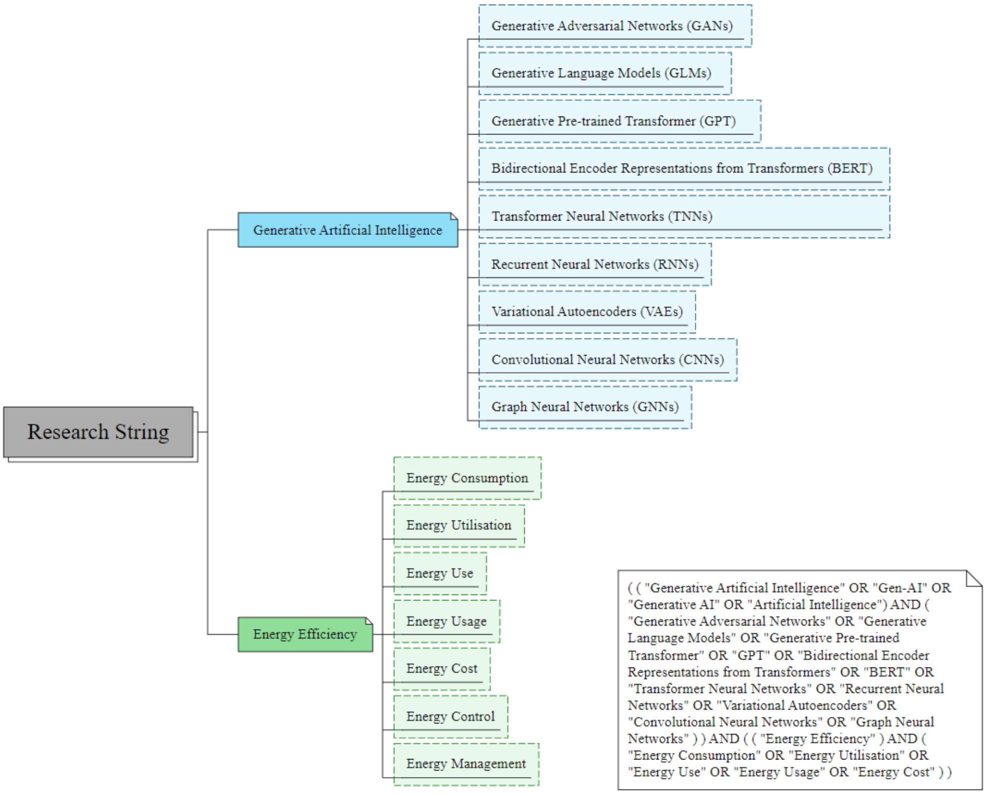


Figure 8 – Word tree and research string of the study.

A comprehensive portfolio of documents was assembled using the Scopus and Web of Science (WoS) electronic databases, configured to search titles, abstracts, and keywords. These platforms were selected due to their broad coverage of high-impact scholarly publications, ensuring the inclusion of influential and methodologically sound studies (FONTOURA et al., 2023). As illustrated in Figure 9, the initial search retrieved 70 documents from Scopus and 18 from WoS before applying the inclusion and exclusion criteria. The study selection process involved filtering these results through predefined eligibility parameters. The inclusion criteria comprised all publications available until December 2024, limited to peer-reviewed journal articles and conference papers published in English. These sources are widely recognised for their academic rigour and reliability in supporting literature reviews (POLLOCK et al., 2016). Exclusion criteria included removing duplicate records, studies that did not explicitly employ DLMs or Gen-AI technologies to improve energy efficiency, and documents that were not accessible through the databases or the authors' ResearchGate profiles. After applying these filters, 36 publications were retained for in-depth analysis.

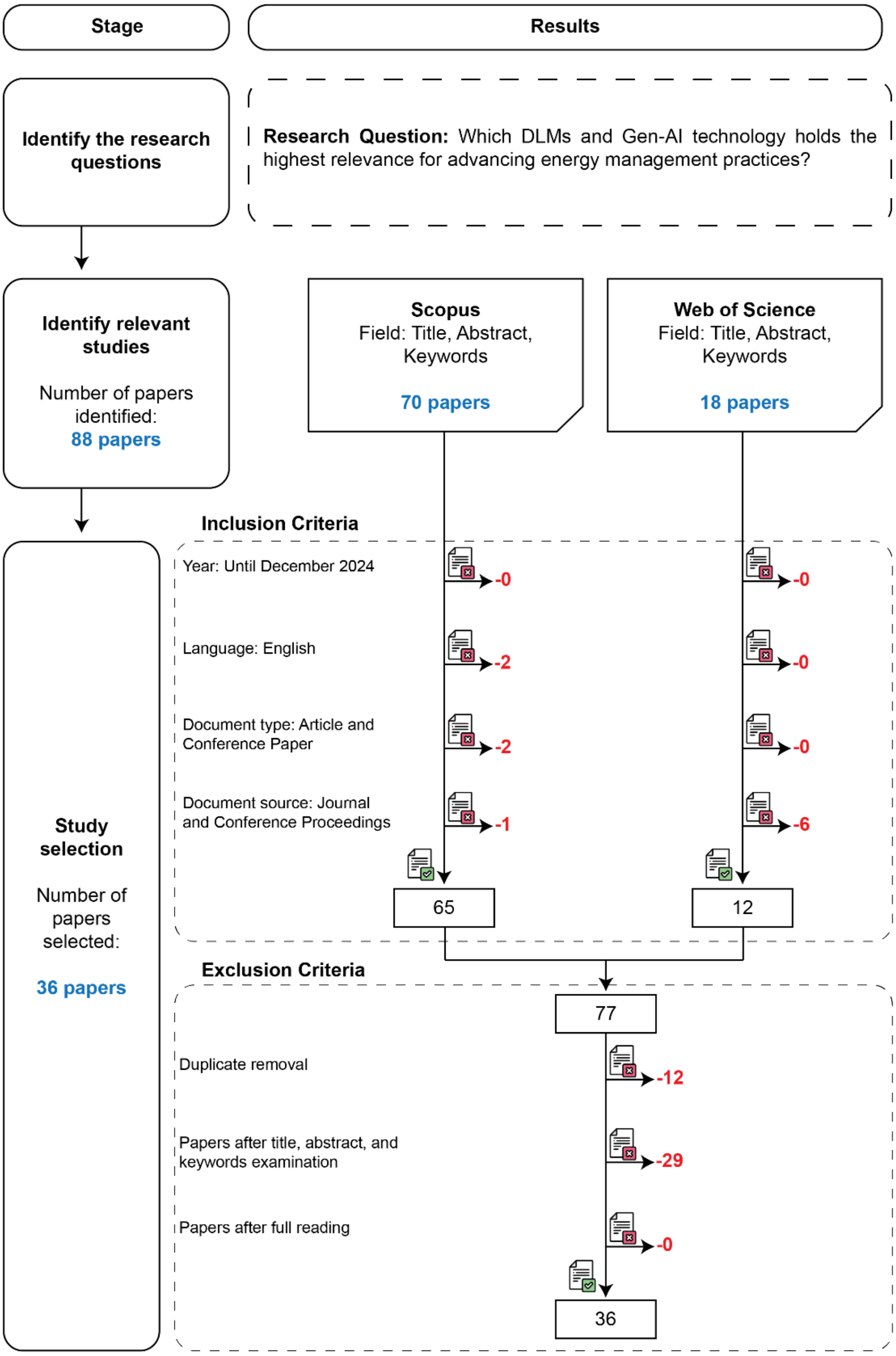


Figure 9 – Study selection.

The selected documents underwent a rigorous two-stage assessment process. In the first stage, titles, abstracts, and keywords were reviewed to confirm their relevance to the research scope. This was followed by a full-text analysis of the shortlisted studies, resulting in a final portfolio comprising 36 papers. The complete list of included studies is publicly available at https://zenodo.org/records/15257001. The analytical methods and data collection parameters were clearly defined for the data charting stage. Descriptive analysis (NÚÑEZ-MERINO et al., 2020) and content analysis (GARZA-REYES, 2015) were identified as the most appropriate methodological approaches for this investigation, allowing for quantitative and qualitative exploration of the selected material.

Subsequently, the validation and data coding stage involved systematically comparing extracted information, supported by collaborative discussions among the researchers to resolve discrepancies. This process enhanced the reliability of the findings and helped minimise potential errors and evaluator bias (FONTOURA et al., 2023). Finally, descriptive analysis was applied to synthesise the main results in the collation, summarising, and reporting stages. In contrast, content analysis was employed to categorise the core DLM technologies identified in the literature, with particular emphasis on those applied to improving energy-related practices.

3.2 Questionnaire design and expert panel

During the second stage, particular attention was dedicated to the design and validation of the questionnaire employed in the study. This process comprised multiple phases aimed at organising and clarifying all relevant aspects of the research for the participants (GARCIA-BUENDIA et al., 2022). In the initial phase, demographic information was collected, including academic qualifications, field of study, current professional role, total years of experience, and the country where the respondent was based. The second phase required participants to assess the perceived relevance of each DLM technology in promoting energy efficiency. This was measured using a five-point Likert scale ranging from 'very low relevance' to 'very high relevance' (DE LIMA; SEURING, 2023). All questionnaire items were thoroughly reviewed and discussed among the authors to minimise the potential for ambiguity or misinterpretation by respondents.

The questionnaire aimed to assess the perceived significance of each DLM technology in energy efficiency models. Before completing the instrument, participants were informed of the study's objectives and provided concise descriptions of each technology, including brief explanations of their respective functionalities. To ensure clarity and consistency, the questionnaire underwent a thorough validation process conducted by the authors to identify potential ambiguities or inconsistencies in wording and structure (FONTOURA et al., 2023; GARCIA-BUENDIA et al., 2022). Preliminary access to the questionnaire involved careful review and simulated responses, allowing the authors to verify that the content was unambiguous and comprehensible to all intended participants.

To ensure the precise selection of qualified participants, experts were chosen based on the criteria proposed by De Lima and Seuring (2023), prioritising individuals with demonstrable experience in technological domains and either conceptual or practical expertise in energy management. This expertise was identified through their areas of specialisation or academic background. To broaden the sample and ensure adequate representation, ninety-eight experts were invited to participate in the first round of the questionnaire. Of these, thirty responded, resulting in a response rate of 30.61%, as summarised in Table 2. Data collection was carried out throughout December 2024. Complete responses are available at https://zenodo.org/records/15257001.

Table 2 – Composition of the expert panel used in the research.

Characteristics	**n (%) Academics**	**n (%) Practitioner**	**n (%) Total**
Experience			
Up to 10 years	3 (20.00)	6 (40.00)	9 (30.00)
From 10 to 25 years	11 (73.33)	5 (33.33)	16 (53.33)
Over 25 years	1 (6.67)	4 (26.67)	5 (16.67)
Background			
Engineering	7 (46.67)	7 (46.67)	14 (46.67)
Computing and Information Technology	1 (6.67)	5 (33.33)	6 (20.00)
Business and Management Systems	4 (26.67)	0 (0.00)	4 (13.33)
Others	3 (20.00)	3 (20.00)	6 (20.00)
Qualification			
Bachelor	1 (6.67)	6 (40.00)	7 (23.33)
Master of Science (MSc)	5 (33.33)	8 (53.33)	13 (43.33)
Doctor (PhD or DSc)	9 (60.00)	1 (6.67)	10 (33.33)
Developing countries			
Brazil	9 (60.00)	6 (40.00)	15 (50.00)
Cuba	2 (13.33)	0 (00.00)	2 (6.67)
Saudi Arabia	0 (00.00)	1 (6.67)	1 (3.33)
Jordan	0 (00.00)	3 (20.00)	3 (10.00)
Developed countries			
Spain	3 (20.00)	0 (00.00)	3 (10.00)
United Kingdom	1 (6.67)	0 (00.00)	1 (3.33)
United Arab Emirates	0 (00.00)	5 (33.33)	5 (16.67)

Table 2 presents the thirty experts who participated in the study, categorised into two groups: academics and practitioners. The inclusion of both categories follows the recommendations of Cuhls et al. (2002), ensuring that participants possess domain-specific expertise aligned with their respective fields of knowledge. This segmentation was intended to support a more targeted analysis of each group's perspective, allowing for identifying the most relevant DLM technologies according to the professional and contextual realities in which these individuals operate.

The group of experts involved in the study demonstrated a diverse range of professional backgrounds and technical expertise. Several professionals with engineering qualifications reported substantial experience across fields such as Building Information Modelling (BIM), digital transformation, supply chain management, technology consultancy, artificial intelligence, and industrial automation. Experts from computer science and information technology backgrounds highlighted their skills in network infrastructure, server management, and cybersecurity. Participants trained in management systems emphasised their academic and practical proficiency in I4.0 technologies, interoperability, BIM integration, and energy efficiency models. Specialists from finance and business strategy disciplines also contributed valuable insights, mainly by applying quantitative methodologies, including statistical techniques and neural network models. Additionally, the study included professionals with specific expertise in blockchain and artificial intelligence, enriching the analysis with forward-looking technological perspectives.

The breadth of knowledge contributed by these experts significantly enhanced the study's findings on the applicability of DLM technologies to energy-related challenges. Notably, 76.66% of participants held a master's or doctoral degree, reinforcing the overall academic quality of the sample. Those with bachelor's degrees also brought substantial professional experience, with an average of 16.07 years in their respective domains. The sample was geographically diverse, comprising experts from developing and developed countries. Of the thirty participants, twenty-one were based in developing nations such as Brazil, Cuba, Jordan, and Saudi Arabia, while the remaining nine were from developed countries including Spain, the United Kingdom, and the United Arab Emirates. The authors intentionally sought a balanced distribution between academic and professional profiles, ultimately engaging fifteen experts from each group to ensure a consistent and well-rounded analysis of the results.

3.3 Friedman and Nemenyi tests

A non-parametric statistical procedure was employed to assess the relative importance attributed by domain experts to various DLM technologies in the context of energy management. This approach was selected due to the ordinal structure of the data, in which each expert ranked six technologies according to their perceived relevance. Since all participants evaluated the same set of technologies, the data reflected repeated measures under consistent conditions. In such scenarios, non-parametric methods are preferable as they do not require standard distribution assumptions or variance homogeneity (AHSAN et al., 2023). Accordingly, based on expert evaluations, the Friedman test was applied to identify statistically significant differences between the technologies.

The Friedman test is a non-parametric statistical procedure used to detect differences in the rankings of multiple treatments or conditions across repeated measures or matched groups (PITTA et al., 2006). It is particularly appropriate when the data consists of rankings or ordinal values across several treatments evaluated by the same subjects (BRANDWEIN et al., 2001). It tests the null hypothesis that the median rankings of the groups are equal, which, in this study, corresponds to verifying whether the experts perceived all six DLM technologies as equally relevant for energy efficiency (FARAMARZI et al., 2020; OSABA et al., 2016). The test calculates a chi-squared statistic based on the sum of the ranks for each condition, adjusted for the number of groups and blocks (OLADEJO; EKWE; MIRJALILI, 2024). In this context, each of the thirty specialists served as a block, and each Gen-AI technology represented a treatment. A p-value of less than 0.05 was considered statistically significant, indicating that at least one of the technologies differed in perceived relevance.

Mathematically, for a set of k conditions (DLM technologies) rated by N subjects (experts), the Friedman test operates as follows. First, each subject ranks the k conditions. The sum of ranks for each condition, denoted as R_j, is then calculated. The Friedman test statistic is given by Equation 13:

$$X_F^2 = \frac{12N}{k(k+1)} \sum_j R_j^2 - \frac{k(k+1)^2}{4} \tag{13}$$

Where:

N = Number of evaluators.

k = Number of items being compared.

R_j = Sum of the ranks assigned to item j.

Under the null hypothesis, the Friedman test statistic X_F^2 is assumed to follow approximately a chi-squared distribution with $k-1$ degrees of freedom, where k represents the number of technologies being compared. The null hypothesis is rejected if the computed test value exceeds the critical threshold defined by the chi-squared distribution at a given significance level. This outcome indicates that there are statistically significant differences in the rankings assigned to the technologies by the experts (FARAMARZI et al., 2020).

Upon obtaining a statistically significant result in the Friedman test, it became necessary to perform a post hoc analysis to identify where the differences occurred. For this purpose, the Nemenyi test was applied (LU et al., 2023). The Nemenyi test is a non-parametric multiple comparison procedure designed to follow the Friedman test when the null hypothesis is rejected (ZHAN et al., 2023). It performs pairwise comparisons between all treatments, identifying whether the differences in mean ranks between any two conditions are statistically significant (SHI; XIONG; LI, 2023; ZHAN et al., 2023). The test is based on a critical difference, which defines the minimum required difference between mean ranks for the result to be considered significant at a given confidence level. In this analysis, a significance threshold of 0.05 was used to assess which pairs of technologies were evaluated differently by the experts (JAUBERT et al., 2024). The Nemenyi test statistic is given by Equation 14:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{14}$$

Where:

CD = Critical Difference.

q_α = Critical value.

k = Number of groups.

N = Number of blocks.

The statistical analysis was conducted using Python, leveraging several specialised libraries. The *pandas* library was used to import and structure the data from an Excel spreadsheet (SUEUR et al., 2023). It provided tabular data manipulation functionalities, enabling the removal of identification columns and the selection of only the relevant variables. The *scipy.stats* module, part of the *SciPy* ecosystem (ASLAM et al., 2023) was used to apply the Friedman test through the *friedmanchisquare()* function. This function takes multiple aligned data arrays and returns the Friedman chi-squared statistic and corresponding p-value. To perform the post hoc analysis, the library *scikit-posthocs* was employed, which includes a function called *posthoc_nemenyi_friedman()* explicitly designed for this type of ranked data. This function returns a matrix of p-values resulting from pairwise comparisons, allowing the identification of statistically significant differences between technologies.

All tests were conducted under a significance level of 5%, following standard practice in empirical research. The methodology ensured statistical robustness by respecting the assumptions of non-parametric analysis and utilising tools well-suited to the data's ordinal nature. Together, the Friedman and Nemenyi tests provided a comprehensive view of the experts' perceptions, enabling the identification of Gen-AI technologies that were significantly prioritised or deprioritised in the context of energy efficiency.

3.4 Hybrid Gen-AI model construction approach

The construction of the hybrid forecasting model proposed in this study followed a structured and iterative methodological path. Rather than selecting components a priori, the model design was informed by combining theoretical foundations, literature review findings, and empirical evidence gathered through expert consultation and statistical analysis. Section 4 details the model architecture and outlines its development methodological rationale and stages. Initially, a set of forecasting techniques was considered, encompassing statistical, neural, and Gen-AI approaches. The integration of these technologies aimed to leverage the strengths of each: statistical models offer simplicity, interpretability and efficiency, while DLMs and Gen-AI models contribute advanced pattern recognition capabilities and adaptability to complex, non-linear data structures (FONTOURA et al., 2025; GAO et al., 2023; KHOT et al., 2023). In the selection process, we adopted a literature- and expert-driven logic: the statistical component was defined based on consolidated recommendations from forecasting studies; the neural component was chosen using both literature and an expert questionnaire analysed with Friedman and Nemenyi tests; and the use of Gen-AI was grounded in the literature for configuration and interpretability.

The model design was informed by the principles of Green AI, which emphasise the development of sustainable and computationally efficient models (ELMOUSALAMI; ELSHABOURY; ELYAMANY, 2024). In parallel, it was conceptually anchored in the Paradox theory, which promotes the constructive acceptance and management of seemingly contradictory demands. Rather than eliminating tensions between performance and efficiency, this perspective encourages their coexistence as a source of innovation and balance (CUNHA; PUTNAM, 2019; KOHTAMÄKI; EINOLA; RABETINO, 2020). Accordingly, each technological choice balanced predictive performance, transparency, and computational efficiency, preserving traceability between theoretical justification, expert evidence, and operational applicability.

The seasonal component of the time series was modelled using the Holt-Winters method, with its configuration, including the choice between additive and multiplicative seasonality, dynamically defined based on the statistical characteristics of the input series.

Residuals from this component were standardised and subsequently modelled using an LSTM trained with sequences of 12 months in an autoregressive structure. The GPT-4-Turbo assisted in parametrising both components, proposing optimal seasonal modes for the Holt-Winters model and key hyperparameters for the LSTM, such as the number of neurons, training epochs, and input window size. The integration of components occurred through direct summation of the LSTM-predicted residuals and the Holt-Winters forecast, with no additional weighting factor. Specifically, Holt–Winters was selected based on the forecasting literature, which indicates it as a transparent and effective technique for monthly energy consumption series; LSTM was chosen based on the literature and confirmed by the expert questionnaire, in which recurrent architectures were indicated as most suitable according to Friedman and Nemenyi statistical analysis; and GPT-4-Turbo was included following Gen-AI literature to support dynamic parameter configuration and produce managerially intelligible interpretations.

The construction process also incorporated preliminary experimentation and validation planning, ensuring module compatibility, data structure alignment, and operational feasibility. The final configuration was refined based on both technical criteria and the contextual requirements of the real-world organisational environment in which the model would later be tested under the action research phase (CORNWALL; JEWKES, 1995). This methodological foundation ensured that the model emerged not from isolated design decisions but from a reasoned synthesis of scientific literature, expert insight, and sustainability-oriented development practices.

3.5 Action research

The concept of action research was first introduced by Kurt Lewin in 1946 as an innovative approach to social inquiry that integrates theoretical development with practical intervention (SUSMAN; EVERED, 1978). It is a comparative investigation into the conditions and consequences of different forms of social action to promote meaningful and context-specific change. Through direct engagement with the social system under study, the researcher initiates transformation and produces critical knowledge grounded in practice (LEWIN, 1946). Action research thus combines knowledge generation with active intervention, enabling researchers to address immediate organisational challenges while contributing to theoretical advancement. This dual purpose allows for bridging theory and practice, aligning academic inquiry with real-world application (BASKERVILLE; WOOD-HARPER, 1998).

Action research is commonly classified into four main types, each distinguished by its methodological orientation and intended outcomes. Diagnostic action research systematically investigates a specific problem to gain a deep understanding before proposing or implementing potential solutions (COHEN; ALROI, 1981). Participatory action research prioritises stakeholders' active engagement throughout the inquiry and intervention phases, promoting shared ownership and collaborative problem-solving (CORNWALL; JEWKES, 1995). Empirical action research is characterised by collecting and analysing practical data, seeking to generate knowledge through real-world application and iterative learning cycles (COUGHLAN; COGHLAN, 2002). Lastly, experimental action research involves the implementation of structured interventions to evaluate alternative strategies and assess their impact, particularly within organisational and educational contexts (LIEDTKE et al., 2015).

3.5.1 Data Collection

This study employed a participatory action research approach using monthly electricity consumption data from Danxia Communication & Marketing, a company located in Brazil. The dataset spans a continuous period of 37 months, from September 2021 to September 2024, offering a solid empirical basis for model training, validation, and testing under conditions that closely resemble actual operations. Significantly, the most recent two years in the series exhibit heightened climatic variability, leading to more pronounced fluctuations and irregularities in consumption behaviour (WANG; ZHANG; YOUSEFI, 2024). These environmental changes introduce additional complexity to the forecasting task, providing a more stringent and realistic assessment of model performance.

The dataset was trained, validated, and tested forecasting models under time-aware conditions that reflect real-world applications. Partitioning strategies varied according to the modelling technique adopted. A chronological split was applied to separate training and testing subsets for statistical models. In contrast, neural network-based methods relied on walk-forward validation, which maintains the temporal order of observations by performing sequential, non-overlapping train-test cycles without random shuffling (GIANNETTI et al., 2025). Additionally, all models were subjected to external validation using real measurements from October 2024 until March 2025, providing a stringent test of forecasting capabilities beyond the original historical window and reinforcing the proposed framework's practical relevance.

3.5.2 Robustness Testing

To strengthen the dataset and enhance the statistical reliability of the model performance evaluation, additional synthetic data were generated using the Circular Block Bootstrap (CBB) approach (TURCO; LLASAT, 2011). This resampling technique maintains both the temporal dependencies and the seasonal dynamics of the original energy consumption series by sampling consecutive 12-month blocks with circular continuity. In total, 120 synthetic monthly observations were produced, corresponding to 10 years of simulated data that preserved the empirical characteristics of the real series, including mean, variance, and autocorrelation structure. These extended data were then used to perform 24-month-ahead forecasting simulations, enabling the assessment of model stability and accuracy under diverse stochastic scenarios.

The robustness and statistical significance of forecasting accuracy among competing models were further assessed through the MCS framework. The MCS was constructed at a 95% confidence level ($\alpha = 0.05$) using 10,000 bootstrap replications of loss differentials, with RMSE serving as the loss function across the 24 forecast horizons. The procedure iteratively removed statistically inferior models, retaining only those belonging to the superior predictive set at the chosen significance level. This method ensures that the model ranking reflects genuine performance differences rather than random variation. The integration of CBB-based synthetic generation with the MCS evaluation thus provides a comprehensive and reproducible framework for testing model generalisation and forecast reliability. The random seed and block length (12 months) were fixed to guarantee reproducibility and alignment with the annual consumption cycle observed in the empirical data.

3.5.3 Classical statistical models implementation

Holt-Winters, SARIMA, Prophet, and STSM models were implemented in Python using dedicated libraries. The *statsmodels* (MAGINGA et al., 2023) and *prophet* (EUFRASIO ESPINOSA; LENNY KOH, 2024) packages were employed for their respective model families. The *ExponentialSmoothing* class was configured for Holt-Winters with additive trend and multiplicative seasonality, optimising smoothing parameters to minimise forecast error. SARIMA was implemented using the *SARIMAX* class (AMIR RAZA et al., 2025), with optimal parameters *(p, d, q, P, D, Q, s)* selected via an extensive grid search procedure, minimising the Akaike Information Criterion (AIC).

Prophet was configured with automatic changepoint detection and yearly seasonality, allowing it to capture recurring patterns and structural shifts with minimal manual intervention. The STSM model was estimated using the *UnobservedComponents* class, configured with a local linear trend and multiplicative seasonal component, both treated as stochastic. This specification enabled the model to capture evolving long-term trends and dynamic seasonal behaviour, offering structural interpretability by decomposing the time series into latent components.

All models were trained exclusively on the training subset and evaluated on the testing data to preserve temporal consistency. Parameter tuning focused on minimising out-of-sample forecast error while avoiding overfitting. Although Prophet permits additional adjustments, such as seasonality priors and changepoint sensitivity, its default settings typically yielded robust and interpretable forecasts. In the case of STSM, its stochastic formulation facilitated adaptive learning of trends and seasonality while retaining the transparency of structural decomposition. All pre-processing, fitting, and evaluation procedures were executed using Python frameworks, ensuring methodological consistency across all traditional statistical models.

3.5.4 Neural Networks implementation

The SimpleRNNs, LSTM, and GRU models were implemented in Python using the *TensorFlow* (SUBBIAH; CHINNAPPAN, 2022). Each architecture consisted of a single recurrent hidden layer (*SimpleRNN, LSTM, or GRU*), followed by fully connected dense output layers. A consistent training pipeline was employed across all networks, using the *Adam* optimiser, and *mean_squared_error* as the loss function. Hyperparameter tuning explored 21 neuron configurations (ranging from 10 to 30), with fixed values for batch size 16, 200 training epochs, and learning rate 0.001. A total of 217 model executions were performed: 210 walk-forward training runs across 21 neuron configurations, 1 final training on the full dataset, and 6 recursive predictions for the forecast horizon.

The LSTM additionally employed a sliding window mechanism to structure input sequences and better capture long-term dependencies in energy consumption patterns (YAQUB et al., 2020), whereas the GRU aimed to reduce computational complexity while preserving forecasting capability (BOUBAKER et al., 2021). The SimpleRNN followed a similar minimalist structure (PAVLATOS et al., 2023), enabling consistent comparisons across all architectures.

3.6 Data analysis

The forecasting models were evaluated using a dual-criteria framework that combines predictive accuracy and computational efficiency, in alignment with Green AI principles (ALZOUBI; MISHRA, 2024). Forecast error metrics and runtime measurements were used to assess model performance and processing cost independently. In parallel, statistical robustness was ensured through the MCS methodology, which was applied exclusively to the squared forecast errors to identify models not significantly outperformed by others at a given confidence level (KHOO et al., 2024). This approach supports the selection of models that are not only accurate and statistically reliable but also operationally efficient and resource-conscious. To complement the statistical analysis, a Pareto frontier plot (SHEHADEH et al., 2024) was employed to explore trade-offs between predictive performance and computational effort visually (VIJAY et al., 2025), highlighting models that offer optimal accuracy and processing efficiency combinations.

Forecasting accuracy was assessed using two complementary metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) (LIU et al., 2024). RMSE served as the primary criterion for model comparison, as it penalises larger deviations more heavily and reflects global predictive fidelity (TURKI et al., 2024). MAE was computed as a secondary, descriptive metric to improve interpretability, offering a linear and intuitive measure of average forecast error (BOUTAHRI; TILIOUA, 2024). However, MAE was not used in the statistical testing procedure. The MCS procedure was applied to support formal model comparison. The comparison was based on monthly squared forecast errors (MSE), calculated for each model across the six forecast periods. From these values, the Average Squared Error (ASE) (DAINOTTI et al., 2024) was derived for each model as:

$$ASE_i = \frac{1}{T}\sum_{t=1}^{T}(\hat{y}_{i,t} - y_t)^2 \tag{15}$$

Where:

T: Number of forecast periods evaluated.

i: Index of the model being evaluated.

$\hat{y}_{i,t}$: Forecast value produced by model *i* at time *t*.

y_t: Actual observed value at time *t*.

To maintain comparability with RMSE in terms of scale and unit (kWh), the RMSE-like metric was defined as the square root of ASE:

$$\text{RMSE-like}_i = \sqrt{ASE_i} \tag{16}$$

These aggregated metrics (ASE and RMSE-like) were used to populate a loss matrix of dimension $T \times M$, where $T = 6$ forecast months, and M is the number of models. The MCS procedure proposed by Hansen et al. (2011) was then applied to determine the subset of models statistically indistinguishable from the best. Formally, the MCS tests the null hypothesis:

$$H_0\colon \mathbb{E}\left[d_{ij,t}\right] = 0 \quad \text{for all } i, j \in \mathcal{M}_0 \tag{17}$$

Where $d_{ij,t} = l_{i,t} - l_{j,t}$ denotes the difference in loss between models i and j at time t. The initial candidate set $\mathcal{M}_0$ contains all competing models. The algorithm proceeds iteratively, eliminating the model with the worst relative performance, based on the maximum t-statistic, until the null hypothesis of equal predictive ability can no longer be rejected. The remaining set, $\widehat{\mathcal{M}_{1-\alpha}^{*}}$, contains the models that are not significantly outperformed by any other at the 95% confidence level. This study implemented the MCS using a bootstrap-based procedure with 1,000 resamples and the squared error loss function (JIANG et al., 2024a).

Two distinct time-based metrics were used to evaluate computational efficiency comprehensively. Execution time, reflecting the actual CPU processing duration during model training and prediction, was employed to quantify the computational load and was instrumental in constructing Pareto frontiers that capture the balance between accuracy and efficiency. Concurrently, wall-clock time, which accounts for total elapsed time including I/O operations, waiting periods, and system-level overhead, was used to approximate energy consumption by multiplying it by the processor’s Thermal Design Power (TDP). This distinction between time metrics prevents overlap and enables a more precise analysis of each model's computational performance and environmental footprint.

All computational experiments were carried out on an Intel Core i7-9750H processor (TDP = 45 Watts), with GPU acceleration deliberately disabled to standardise the execution environment for both statistical and neural models. This setup ensures that all performance results are based solely on CPU computation, eliminating discrepancies that could arise from hardware-level acceleration. By jointly considering forecast accuracy, statistical reliability, and processing efficiency, the evaluation framework facilitates the identification of models that are not only technically sound but also aligned with sustainable and responsible AI implementation principles.

4 GENENECAST AS A HYBRID FORECASTING MODEL

The GenEneCast model is a hybrid forecasting framework that predicts short-term energy consumption based on monthly time series data. It captures seasonal patterns and non-linear residual behaviours by combining the statistical robustness of the Holt-Winters method (YANG et al., 2024) with the adaptive learning capabilities of an LSTM neural network (SHAO et al., 2025). In addition to the Generative Language Model (GLM) summaries of the results, GPT-4-Turbo is used to assist in configuring key model parameters, including the seasonal mode of Holt-Winters and the architecture of the LSTM. The model allows users to define the forecast horizon and incorporates contextual metadata, such as city, state, and country, to enhance location-aware analysis. This integrated architecture supports interpretable and accurate forecasting, particularly in contexts where historical regularities and subtle variations play a significant role.

4.1 GenEneCast architecture

GenEneCast introduces a hybrid framework for short- and medium-term energy forecasting, integrating statistical decomposition, neural modelling, and GLMs to improve predictive accuracy and interpretability. Unlike traditional hybrid approaches that rely on fixed configurations or manual tuning, GenEneCast leverages the generative capacity of GPT-4-Turbo to orchestrate and interpret the forecasting process dynamically. This integration enhances adaptability, reinforces explainability, and aligns with computational sustainability goals. Figure 10 outlines the technological architecture of the model.

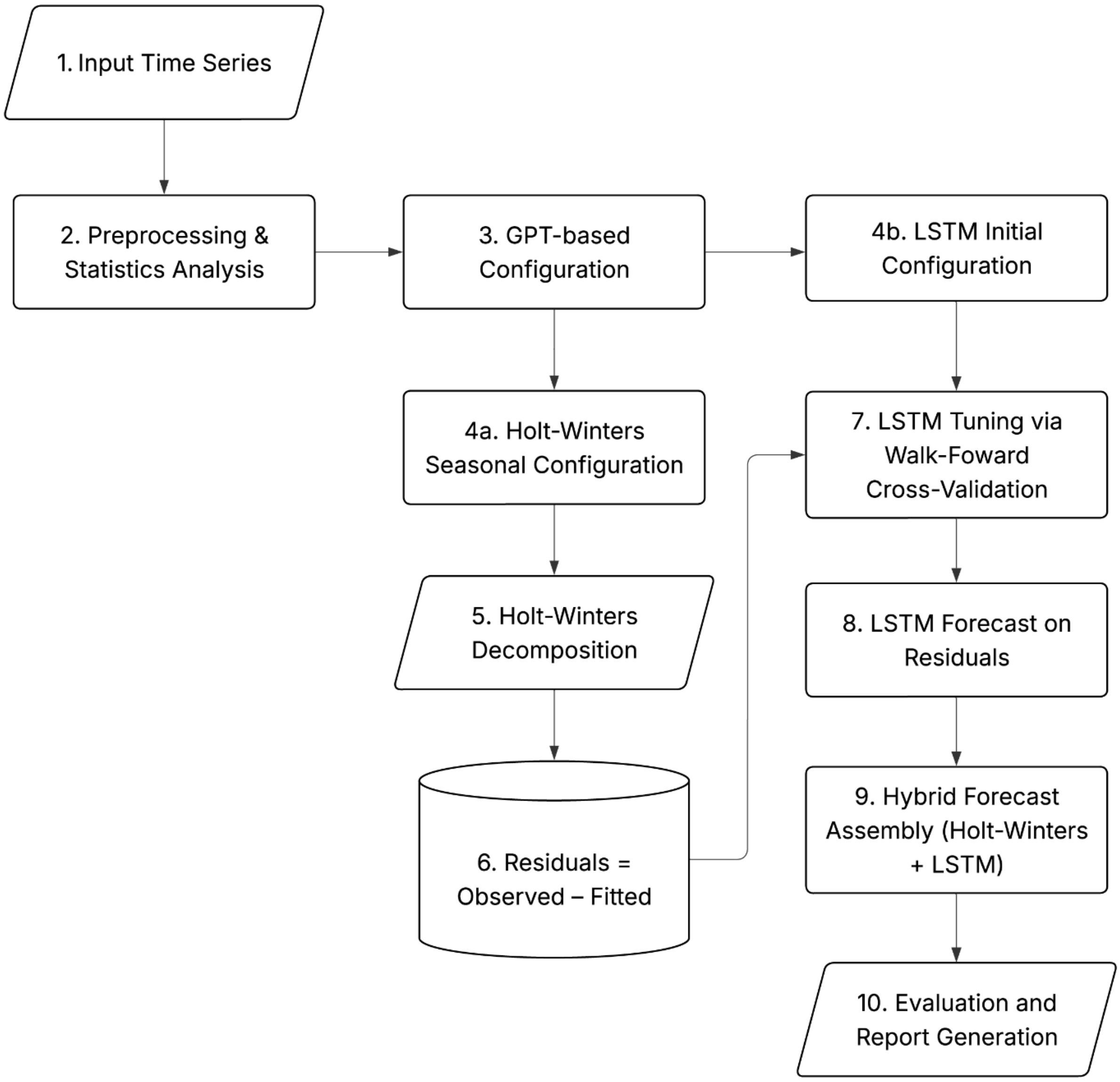


Figure 10 – Overview of the GenEneCast pipeline.

As shown in Figure 10, the forecasting process begins with decomposing the observed energy consumption series y_t using the Holt-Winters exponential smoothing method. The model is configured exclusively to capture seasonal patterns, with the trend component intentionally omitted to delegate the modelling of dynamic and nonlinear variations to the neural layer. This results in the additive and multiplicative form:

$$y = \begin{cases} S_t + \varepsilon_t, (additive) \\ S_t \times \varepsilon_t, (multiplicative) \end{cases} \tag{18}$$

Where S_t represents the seasonal component and ε_t denotes the residual series. The configuration of the Holt-Winters model, comprising the type of seasonality (additive or multiplicative) and the periodicity, is not set manually. Instead, it is automatically suggested by GPT-4-Turbo based on descriptive statistics extracted from the original series, such as mean, standard deviation, autocorrelation at lag 12, and the slope of the linear trend. These values are passed to the GPT model, which returns a valid JSON structure containing the optimal parameters θ^{HW}.

The residual component ε_t, representing the irregular and non-seasonal behaviours, is then modelled by an LSTM neural network. Before training, a second GPT-4-Turbo call is made using summarised residual statistics, including skewness, kurtosis, Autocorrelation Function (ACF), and Partial Autocorrelation Function (PACF) values (WARIS; TARIQ, 2024; ZHENG et al., 2024). This call dynamically configures the LSTM architecture, returning the number of time steps (input window size), dropout rate, and number of neurons, encoded as θ^{LSTM}. The final forecast on the horizon $t + h$ is the recombination of the seasonal forecast from Holt-Winters and the LSTM-based residual prediction:

$$\hat{y}_{t+h} = \hat{S}_{t+h} + \hat{\varepsilon}_{t+h} \tag{19}$$

This can be expressed in terms of the whole pipeline:

$$\hat{y}_{t+h} = HW_{\theta^{HW}}(y_{1:t}) + LSTM_{\theta^{LSTM}}(\epsilon_{1:t}) \tag{20}$$

With both sets of hyperparameters defined as:

$$\theta^{HW}, \theta^{LSTM} \sim \text{GPT-4-Turbo} \tag{21}$$

Training of the residual LSTM relied on backpropagation through time (BPTT). The network is unfolded over the selected input window, and gradients are propagated across time via the chain rule to minimise the RMSE-oriented loss, with *Adam* optimisation. In practice, the implementation uses truncated BPTT with a sequence length equal to the chosen input window. Training is performed on 3D tensors built from the residual series, which are passed to an LSTM layer compiled with MSE loss in *TensorFlow/Keras*, executing BPTT by default.

Hyperparameter tuning in GenEneCast is also adaptive. The neuron count suggested by GPT serves as a baseline for constructing a proportional search space, ranging from 60% to 140% of the recommended value. These candidate configurations are evaluated using repeated walk-forward cross-validation. The number of folds is dynamically adjusted according to the size of the input series. In contrast, the number of repetitions per fold depends on the selected forecasting precision mode (Low, Medium, or High). Under its most intensive setting, GenEneCast may execute up to 251 training cycles, allowing it to handle time series with diverse lengths and structures while ensuring transparency regarding computational expenditure.

The model is retrained using the complete residual dataset once the optimal configuration is selected, based on RMSE across validation segments. Forecasts are produced recursively, with each estimated residual incrementally added to its respective seasonal component. This approach ensures that both deterministic seasonality and dynamic behavioural patterns are captured. In addition to numeric prediction, GenEneCast incorporates a third modelling layer aimed at interpretability, powered by GLMs. Following the final forecast generation, the system transmits a structured data package, including predicted values, residual adjustments, and geographic context (city, state, country), to GPT-4-Turbo. The GLM produces three outputs: (i) a month-by-month narrative comparing forecasted consumption to the same period in the previous year; (ii) a contextual explanation grounded in potential climatic or behavioural influences; and (iii) an interpretation of the residual adjustments and associated error indicators (RMSE and MAE), all expressed in accessible language suitable for executive audiences.

These outputs are automatically assembled into a Word document that includes annotated charts and performance summary tables. The forecasting pipeline is structured through modular Python scripts, featuring dedicated functions for data preprocessing, statistical setup, neural modelling, validation routines, forecasting procedures, and interpretative reporting. By leveraging GPT-4-Turbo as both a configuration mechanism and an interpretability layer, GenEneCast effectively unites statistical precision with user-focused insights. This illustrates how Gen-AI can act simultaneously as a content generator and an adaptive co-modelling agent, advancing human-centred automation within complex forecasting scenarios. All source code and a report file example are publicly accessible in the repository and are available at: https://zenodo.org/records/15375974.

4.2 Libraries and implementation tools in GenEneCast

The GenEneCast forecasting pipeline was entirely implemented in Python, relying on a modular open-source library suite that enables statistical decomposition, deep learning, cross-validation, interpretability, and automated report generation. Together, these components constitute a cohesive and reproducible framework tailored for energy forecasting tasks, emphasising flexibility and transparency. The statistical component of the model was implemented using the *statsmodels* library (KOLOSOK et al., 2022), which provides a flexible interface for Holt-Winters exponential smoothing. Rather than manually configuring this method, GenEneCast uses GPT-4-Turbo to determine the optimal seasonal mode and periodicity based on series diagnostics. For neural network modelling, *TensorFlow* and its high-level API *Keras* were employed (LI et al., 2023; MASOOD et al., 2023; NIZAM et al., 2022). These libraries enabled the dynamic construction and training of LSTM architectures, with hyperparameters (e.g., number of neurons, window size, dropout rate) also being defined through GPT-driven analysis of residual statistics (WARIS; TARIQ, 2024; ZHENG et al., 2024).

The *scikit-learn* library was used to standardise residual inputs via *StandardScaler*, implement walk-forward cross-validation through *TimeSeriesSplit*, and compute evaluation metrics such as RMSE and MAE (ADUGNA; XU; FAN, 2022; HASSAN et al., 2022; SAEED; HAMA, 2023). Data manipulation and array-level operations were handled by `pandas` and `NumPy`, which ensured efficient time-indexing and vectorised computations (ABDENNUR et al., 2024; DAUTZENBERG et al., 2022; EL HACHIMI et al., 2022).

For result presentation and communication, *matplotlib* was used to generate publication-ready visualisations (MAY et al., 2022), while the *python-docx* library enabled the automatic generation of structured Word reports containing tables, annotated charts, and summaries (CHEN et al., 2024). To support interpretability, the *OpenAI* library facilitated integration with GPT-4-Turbo, which served both as a configuration engine and a generative interpreter (COOK, 2025; LIM et al., 2023b; WALKER et al., 2023). The generated narratives incorporated forecasted values, residual dynamics, and contextual metadata (city, state, country), producing domain-specific insights tailored for managerial audiences. Finally, supporting libraries such as *os*, *datetime*, and *psutil* were used for execution tracking, file management, and system-level monitoring during training and inference. Together, these tools form a robust and scalable infrastructure that underpins GenEneCast's goal of delivering accurate, explainable, and resource-aware energy forecasts.

4.3 Foundational technologies in GenEneCast

The GenEneCast model uses established forecasting techniques to deliver interpretable and efficient energy predictions through hybrid learning. At its core, the model leverages a combination of statistical decomposition and neural refinement, coordinated by a Gen-AI layer that dynamically configures model parameters and generates natural language interpretations. The Holt-Winters exponential smoothing method was selected as the statistical backbone due to its proven effectiveness in capturing strong seasonal patterns typical in electricity consumption data (PIRES; MARTINS, 2024). As a transparent and low-cost forecasting technique, Holt-Winters remains particularly attractive for operational settings where interpretability and computational efficiency are critical (AHMADI et al., 2023; ZÜGE; COELHO, 2024). Rather than fixing the seasonal mode a priori, as done in traditional models, GenEneCast uses GPT-4-Turbo to automatically determine whether additive or multiplicative seasonality is more appropriate, based on descriptive features extracted from the series.

GenEneCast incorporates an LSTM network trained on the residual series to address the nonlinear and dynamic components not captured by seasonal smoothing. As a variant of RNNs, LSTM is widely recognised for its ability to learn long-range dependencies and retain information across time steps, making it particularly well-suited for time series forecasting applications (EHRHART et al., 2022; ONAN, 2022; ZHEN et al., 2022). This neural layer enhances the statistical baseline by modelling residual trends and behavioural fluctuations, which are often influenced by irregular consumption patterns and external factors.

What differentiates GenEneCast from previous hybrid models is the integration of GPT-4-Turbo as a narrative generator and a dynamic configuration engine. It suggests the Holt-Winters parameters based on descriptive statistics (mean, standard deviation, autocorrelation, trend slope). It defines key LSTM hyperparameters, including the number of neurons, dropout rate, and window size, based on residual diagnostics such as skewness, kurtosis, and autocorrelation structures (WARIS; TARIQ, 2024; ZHENG et al., 2024). This adaptive approach reduces the need for manual tuning while improving responsiveness to dataset-specific characteristics. In addition, GPT-4-Turbo is used to produce human-readable summaries that contextualise forecast results using geographic metadata and temporal comparisons. As a result, GenEneCast bridges the gap between statistical rigour and managerial interpretability (LIM et al., 2023b; WALKER et al., 2023), supporting more accessible decision-making processes while preserving modelling transparency and reproducibility.

5 RESULTS

This section presents the empirical findings derived from two complementary analyses. The first involves a statistical assessment of expert rankings, providing insights into how different forecasting technologies are perceived in terms of relevance and suitability, particularly concerning the standing of the proposed hybrid model. The second analysis evaluates the predictive performance of the GenEneCast framework using quantitative indicators, highlighting its accuracy, robustness, and applicability in real-world energy consumption contexts. To ensure statistical rigour, the MCS procedure was applied at a 95% confidence level, allowing the identification of statistically superior models and enabling a rigorous comparison of forecasting approaches under uncertainty. The comparative evaluation of forecasting models emphasised the performance and stability of the GenEneCast framework. Accuracy metrics assessed over a six-month forecasting horizon using real energy consumption data demonstrated that all GenEneCast configurations consistently outperformed both statistical and neural network baselines, confirming their ability to balance forecasting precision and generalisation.

To verify the consistency and robustness of these findings, an additional analysis was conducted using 10,000 bootstrap replications of synthetic datasets generated through the CBB method, followed by evaluation via the MCS framework. As presented in Appendix A, the results obtained under simulated stochastic conditions closely mirrored those derived from real data, reinforcing the validity of the models' comparative behaviour. The GenEneCast-High configuration achieved the lowest RMSE (130.51), followed by Prophet (133.72) and GenEneCast-Medium (136.36). All four GenEneCast variants remained within the 95% confidence MCS, indicating that their predictive accuracies were statistically indistinguishable from the top-performing models. The GenEneCast-Low configuration also maintained competitive precision (RMSE = 136.58), outperforming all traditional baselines, including SARIMA (137.96), STSM (141.98), and Holt-Winters (154.66).

These findings demonstrate that the hybrid generative framework not only achieves superior performance in empirical forecasts but also sustains its reliability under stochastic perturbations in resampled datasets. The strong convergence between real-data and synthetic-data results underscores the robustness, adaptability, and generalisation strength of the GenEneCast approach in modelling complex temporal and seasonal energy consumption dynamics.

5.1 Statistical analysis of expert rankings

The Friedman test was employed to determine whether statistically significant differences existed in how practitioners and academics ranked six DLM technologies based on their perceived relevance to energy efficiency applications. As a non-parametric technique, it is well-suited for analysing repeated measures in which the same participants assess multiple items through ordinal rankings (BRANDWEIN et al., 2001; PITTA et al., 2006). Among practitioners, the test yielded a chi-square statistic of 3.3374 and a p-value of 0.6481, indicating no significant differences in their evaluations. This suggests that, within this group, the six Gen-AI technologies were viewed with relatively similar levels of relevance. Nonetheless, some subtle preferences emerged. RNNs achieved the highest mean rank (3.87), followed closely by GNNs (3.80) and GANs (3.73). VAEs (3.67), TNNs (3.60), and CNNs (3.53) followed in close succession, indicating a general equilibrium in practitioner perceptions.

In contrast, the results from academic experts revealed a more pronounced differentiation. The Friedman test returned a chi-square statistic of 25.1544 with a p-value of 0.0001, providing strong evidence of significant differences in the perceived relevance of the technologies. RNNs again ranked highest (4.40), followed by GANs (4.13) and VAEs (3.80). GNNs (3.67) and CNNs (3.47) were ranked slightly lower, while TNNs received the lowest mean score (2.80). The Nemenyi post hoc test was conducted to explore these distinctions further, as shown in Table 3. The analysis revealed significant differences between RNNs and TNNs (p-value = 0.0005) and between GANs and TNNs (p-value = 0.0099). These results confirm that academics considered RNNs significantly more relevant than other architectures, particularly when compared to transformer-based and less established Gen-AI techniques.

Table 3 – Nemenyi post hoc test (Academics)

Gen-AI	**GANs**	**TNNs**	**RNNs**	**CNNs**	**VAEs**	**GNNs**
GANs	1.000000	0.009894	0.970844	0.624125	0.925693	0.909831
TNNs	0.009894	1.000000	0.000481	0.462151	0.159273	0.177248
RNNs	0.970844	0.000481	1.000000	0.177248	0.494120	0.462151
CNNs	0.624125	0.462151	0.177248	1.000000	0.992016	0.994676
VAEs	0.925693	0.159273	0.494120	0.992016	1.000000	1.000000
GNNs	0.909831	0.177248	0.462151	0.994676	1.000000	1.000000

When all responses were combined, the overall Friedman test confirmed the existence of significant differences across expert rankings, with a chi-square statistic of 18.5671 and a p-value of 0.0023. RNNs continued to lead the ranking (4.13), followed by GANs (3.93), and both GNNs and VAEs tied (3.73). CNNs (3.50) and TNNs (3.20) remained in the lower positions. To identify which specific differences were statistically significant, the Nemenyi post hoc test was applied, as shown in Table 4.

Table 4 – Nemenyi post hoc test (Combined experts)

Gen-AI	GANs	TNNs	RNNs	CNNs	VAEs	GNNs
GANs	1.000000	0.109078	0.968702	0.537957	0.962412	0.999785
TNNs	0.109078	1.000000	0.010597	0.955299	0.514961	0.203335
RNNs	0.968702	0.010597	1.000000	0.128677	0.584097	0.893531
CNNs	0.537957	0.955299	0.128677	1.000000	0.955299	0.718178
VAEs	0.962412	0.514961	0.584097	0.955299	1.000000	0.993928
GNNs	0.999785	0.203335	0.893531	0.718178	0.993928	1.000000

The only significant difference observed in this combined dataset was between RNNs and TNNs (p-value = 0.0106), indicating that experts significantly preferred RNNs over TNNs. All other comparisons produced p-values well above the 0.05 threshold, suggesting that the remaining technologies were perceived with relatively similar relevance in the eyes of the experts. This outcome underscores experts' distinction between architectures traditionally used in time series forecasting, such as RNNs, and more recent transformer-based models that may be less familiar or applicable in the energy domain.

5.2 Internal results of the GenEneCast model

This section presents a comparative analysis of all forecasting models, particularly emphasising the proposed GenEneCast framework. Core accuracy metrics were evaluated across a six-month prediction window, and the MCS procedure was applied at a 95% confidence threshold to ensure statistical robustness. The findings provide a quantitative foundation for assessing the consistency and reliability of GenEneCast configurations relative to established benchmarks. GenEneCast was deployed in three precision modes, Low, Medium, and High, each calibrated to strike a specific balance between forecasting accuracy and computational overhead. Across the forecasting period from October 2024 to March 2025, the model architecture adapted dynamically according to the input series' statistical characteristics and the chosen precision level. The setup of each GenEneCast variant was guided by GPT-based recommendations, with the GLM evaluating key statistical attributes of the time series.

5.2.1 Architectural configuration and validation performance

Following its time series analysis using statistical descriptors, GPT-4-Turbo consistently recommended a multiplicative seasonal structure for the Holt-Winters component, with a periodicity 12, capturing the monthly cyclicality of energy consumption. This setting was constant across all executions, enabling architectural variation to be concentrated exclusively within the LSTM module. In its initial recommendation, GPT-4-Turbo proposed 32 neurons for the LSTM across all precision tiers, aiming to optimise the trade-off between model generalisation and training efficiency. However, this neuron count served as a reference point; the final configuration was determined dynamically via an internal tuning routine, which accounted for both the desired accuracy level and the depth of cross-validation applied.

The tuning process followed a walk-forward cross-validation strategy, preserving the temporal integrity of the dataset and closely reflecting real-world forecasting conditions. As repetitions increased, from one in the Low configuration to five in the High configuration, the model progressively adopted fewer neurons, ranging from 44 to 19. This adjustment highlights the inherent trade-off between network complexity and validation stringency when operating under realistic temporal constraints, as shown in Table 5.

Table 5 – Architecture and validation metrics across configurations

Model	Neurons selected	GPT suggestion	Folds	Repeats	Validation RMSE	Validation MAE
Low	44	32	3	1	0.81	0.64
Medium	25	32	3	3	0.96	0.77
High	19	32	3	5	0.92	0.71

Although the configuration space allowed for up to 251 training iterations, the system autonomously constrained this number by adapting to the statistical properties of the data and the required precision level. This self-regulating mechanism conserved computational resources and mitigated the risk of overparameterisation. Notably, despite undergoing the least intensive validation, the Low configuration yielded the best validation scores (RMSE: 0.81, MAE: 0.64), indicating a strong alignment between model simplicity and the underlying signal characteristics. The High configuration, although more conservative and subjected to more rigorous validation, also achieved reliable internal accuracy (RMSE: 0.92), reinforcing the stability of the residual learning structure across tuning intensities.

5.2.2 Forecast curves and residual diagnostics

The GenEneCast model's internal dynamics were further investigated by evaluating forecast trajectories and residual patterns across all precision configurations. Figure 11 displays the overlaid forecast curves of the three GenEneCast variants compared to the actual observed values. The solid black line represents measured energy consumption, while the dashed blue line depicts the in-sample fitted values produced during training. Forecasted points for the test period are distinguished by colour: green for GenEneCast – Low, orange for GenEneCast – Medium, and red for GenEneCast – High.

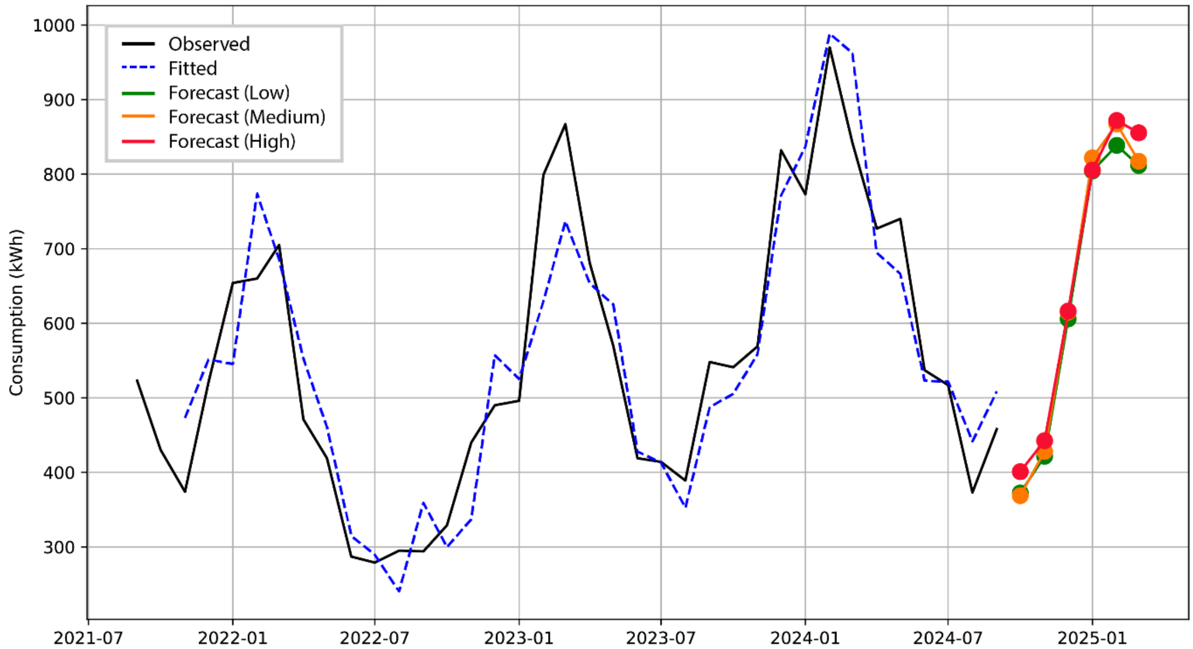


Figure 11 – GenEneCast final forecast

Figure 12 displays the GenEneCast model's residuals for all three configurations. The grey line illustrates recent residuals from the training phase. In contrast, the forecasted residuals during the testing period are colour-coded: green for GenEneCast – Low, orange for GenEneCast – Medium, and red for GenEneCast – High. These residuals capture the deviation between actual observations and the corresponding forecasts generated by each configuration. The absence of discernible trends, autocorrelation, or structural distortions across the residual series indicates that all configurations are statistically well-calibrated.

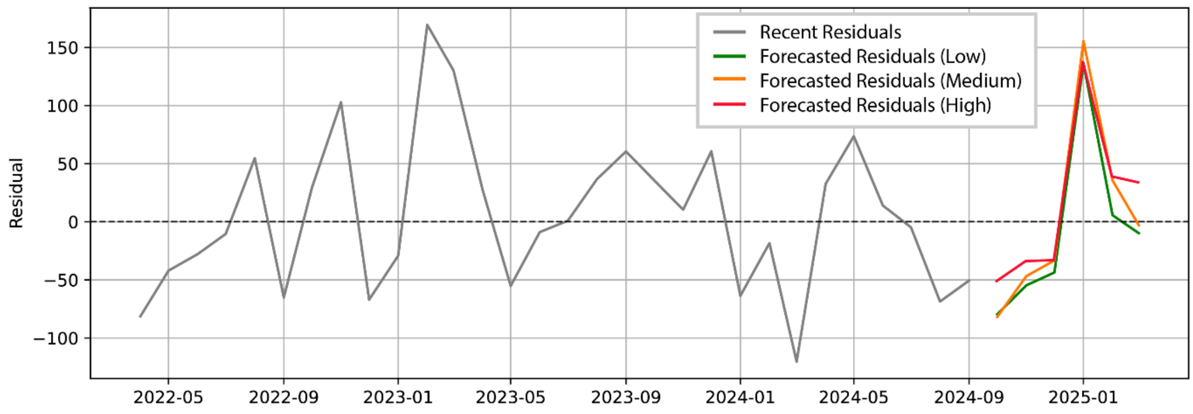


Figure 12 – Residuals of GenEneCast.

Under the Low configuration, the model delivered consistent and accurate forecasts, mirroring the actual energy consumption throughout the evaluation period. The residuals were predominantly symmetric and centred near zero, ranging from –80 to +136 units. The most pronounced positive residual occurred in January 2025, indicating a slight underestimation of peak summer demand. Nonetheless, no visual evidence of systematic bias or autocorrelation was observed. This residual pattern reflects a well-tuned hybrid decomposition, where the Holt-Winters component effectively captured the primary seasonal structure. At the same time, the LSTM correction layer accommodated nonlinear variations in the signal.

The Medium configuration produced a forecast trajectory broadly comparable to the Low mode, though with slightly larger deviations from actual values in January and February. Residuals exhibited a broader amplitude, ranging from –82 to +155 units. Despite this wider spread, the residuals remained statistically consistent, with no visible trend or structural dependency. These findings indicate that increasing the number of validation repetitions from 1 to 3 did not yield substantial gains in residual correction accuracy, though it preserved model robustness and generalisation capacity. The configuration maintained responsiveness to short-term fluctuations, particularly during transitional seasonal periods.

The High configuration generated the most conservative set of forecasts, with residual adjustments ranging from approximately -51 to +138 units. The residual profile shows a narrower error band compared to the Medium configuration, particularly during the final months of the testing period. Positive deviations in January and February again point to a slight underestimation of peak demand, though with reduced magnitude. The smoother residual trajectory and lower variance suggest the influence of more intensive validation (five repetitions), which likely mitigated overfitting risks. Although the validation RMSE was marginally higher than the Low configuration, residual diagnostics confirm that the High mode delivered stable and reliable forecasts under more rigorous cross-validation constraints.

Overall, the residual behaviour observed across all configurations attests to the internal consistency of the GenEneCast model. Forecast errors remained near zero, with no evidence of autocorrelation or systematic bias. The stability of residual patterns, especially during seasonal demand peaks, further validates the effectiveness of the hybrid decomposition and correction approach. Moreover, walk-forward cross-validation enhanced the realism and robustness of the internal evaluation process, closely simulating actual forecasting conditions.

5.3 MCS-based comparative analysis

This section reports the numerical outcomes of the forecasting experiment conducted from October 2024 to March 2025. The models under evaluation comprise statistical approaches (STSM, Holt-Winters, SARIMA, and Prophet), neural networks (SimpleRNN, LSTM, and GRU), and the proposed hybrid architecture, GenEneCast, tested under three precision settings: Low, Medium, and High. A row containing the actual observed values is also included as a reference baseline for comparative analysis.

Monthly forecast errors (in kWh) are presented alongside the ASE and the corresponding RMSE-like value, summarising overall predictive performance. Execution time (ET, in seconds) reflects each model's actual CPU processing duration. Estimated energy consumption (EC, in watt-hours) was computed by multiplying the processor's TDP by the total wall-clock time, representing the whole period during which the system remained active. The final column indicates whether the model was retained within the 95% Model Confidence Set (MCS), using RMSE as the evaluation loss function.

As shown in Table 6, all models remained within the 95% MCS, indicating no statistically significant inferiority relative to the best-performing alternative. Among them, RMSE-like values ranged from 82.73 (GenEneCast – Medium) to 238.27 (Prophet), evidencing a performance gap of nearly 2.9 times. GenEneCast – Low, despite having an RMSE-like of 89.23, approximately 7.9% higher than Medium and 6.0% higher than High, achieved virtually equivalent predictive accuracy. Crucially, it did so with markedly greater computational efficiency: it required only 2.69 Wh and 283.73 seconds to complete training and inference, significantly outperforming standalone neural models such as GRU (178.36 RMSE-like, 3132.80 seconds, 29.09 Wh), LSTM (174.58 RMSE-like, 2968.73 seconds, 27.87 Wh), and SimpleRNN (145.87 RMSE-like, 2736.58 seconds, 26.084 Wh). These results highlight the value of combining statistical decomposition with lightweight residual learning, as implemented in the GenEneCast – Low configuration, which delivered superior efficiency without compromising accuracy.

Classical models such as SARIMA and Holt-Winters also completed execution in under three seconds. They were included in the MCS, indicating that, under certain conditions, simpler models may yield statistically comparable performance. The individual monthly forecasts revealed interesting variability; nevertheless, GenEneCast – Low consistently maintained lower error values across all months, contributing to its overall advantage in ASE and RMSE-like metrics.

Table 6 – Comparative performance of forecasting models

Model	**Oct 2024**	**Nov 2024**	**Dec 2024**	**Jan 2025**	**Feb 2025**	**Mar 2025**	**ASE**	**RMSE-like**	**ET**	**EC**	**In MCS (95%)**
GenEneCast - Medium	370.00	429.46	616.06	823.16	868.54	818.54	6844.2240	82.7298	831.08	7.713	Yes
GenEneCast - High	380.00	442.46	616.35	805.39	871.90	855.41	7087.2541	84.1858	1360.89	12.514	Yes
GenEneCast - Low	372.35	421.53	605.73	804.26	838.59	811.69	7962.6780	89.2338	283.73	2.689	Yes
STSM	468.48	478.71	614.04	622.65	772.85	749.40	18707.4502	136.7752	0.89	0.011	Yes
SimpleRNN	467.68	528.30	530.15	730.10	722.77	920.27	21268.3985	145.8369	2736.58	26.084	Yes
LSTM	412.97	450.73	511.74	586.70	654.12	713.04	30477.1946	174.5772	2968.73	27.871	Yes
GRU	408.86	445.11	505.96	578.58	649.13	704.14	31811.6544	178.3582	3132.80	29.087	Yes
SARIMA	532.51	656.43	807.61	726.48	1069.95	1021.65	37168.0621	192.7902	2.81	0.034	Yes
Holt-Winters	577.31	604.63	816.70	843.38	1050.01	1029.95	37881.4817	194.6317	2.36	0.030	Yes
Prophet	674.04	645.09	741.27	1057.66	1133.91	1053.50	56771.0761	238.2668	2.36	0.033	Yes
Real Data	385.00	445.00	500.00	785.00	1025.00	784.00	–	–	–	–	–

6 DISCUSSION

The following discussion analyses the performance of the evaluated models based on three main criteria: predictive accuracy, execution time, and energy consumption. Key findings derived from the Pareto frontier are highlighted, emphasising the strategic positioning of the GenEneCast model compared to the others. Furthermore, in light of the Green Paradox-AI, the discussion addresses the practical implications of these results, considering model applicability and the trade-offs between complexity and efficiency.

6.1 Reflections on Gen-AI applicability for energy management

The comparative analysis of expert profiles provides valuable insights into how different professional backgrounds influence the perceived relevance of Gen-AI technologies in energy efficiency. Across all three analytical perspectives, practitioners, academics, and the combined expert group, RNNs consistently emerged as the most highly regarded technology, underscoring their consolidated position in time series forecasting (PUSHPAVALLI et al., 2024). Among practitioners, although the Friedman test did not indicate statistically significant differences (p-value = 0.6481), RNNs still achieved the highest mean rank (3.87). This result suggests that, even without a strong analytical distinction among models, RNNs are intuitively recognised by professionals for their suitability in modelling sequential data, likely due to their established track record and accessibility in operational contexts (FONTOURA et al., 2025). Their ability to learn temporal dependencies makes them highly practical in forecasting energy consumption, where seasonal and cyclical patterns are common.

The view among academic experts was even more emphatic. In this group, the Friedman test confirmed the presence of significant differences in ranking (p-value = 0.0001), and RNNs once again led with the highest mean score (4.40). The Nemenyi post hoc test validated this distinction, revealing that RNNs were rated significantly higher than TNNs (p-value = 0.0005). This result reinforces the strong academic confidence in RNNs' predictive capabilities and alignment with theoretical expectations in time series modelling (DUTTA BARUAH; MUÑOZ ORGANERO, 2024; YANG; LIU, 2024). The preference for RNNs over newer architectures, such as transformer-based models, may also reflect academic caution regarding such techniques' complexity, resource demands and domain-specific maturity.

In the combined analysis of all thirty experts, RNNs maintained their lead (mean rank 4.13), and the overall Friedman test once again identified significant differences among technologies (p-value = 0.0023). The Nemenyi test confirmed a statistically significant difference between RNNs and TNNs (p-value = 0.0106), highlighting that the preference for RNNs is prominent in isolated subgroups and consistent when both practitioner and academic perspectives are merged. This convergence suggests a shared understanding of RNNs as a reliable and context-appropriate choice for energy forecasting (YANG; LIU, 2024).

The consistent prioritisation of RNNs across all groups likely stems from their balance between performance, interpretability, and computational feasibility (FONTOURA et al., 2025). Unlike more complex architectures, RNNs are relatively lightweight, widely implemented, and backed by a large body of empirical studies demonstrating their effectiveness in forecasting tasks. This practicality and theoretical robustness position RNNs as the most suitable foundation for hybrid Gen-AI forecasting models to improve energy efficiency.

While other technologies such as GANs, GNNs and VAEs also received moderate to high evaluations, their scores varied more significantly between expert groups. GANs, for instance, were consistently ranked in second place but did not exhibit significant statistical separation from the other models in most comparisons. VAEs and GNNs were perceived with cautious optimism, perhaps due to their growing but still limited application in the energy domain. CNNs and TNNs generally ranked lower, particularly among academics, where TNNs showed the weakest performance, reinforcing concerns about their complexity, interpretability, and current level of adoption in energy-related scenarios (JIANG et al., 2024b; XUE et al., 2024). In summary, the unanimous appreciation of RNNs across all expert profiles validates their role as the preferred architecture for energy management. Their consistent ranking reflects theoretical credibility and practical applicability, making them a solid cornerstone in developing computationally efficient and domain-relevant hybrid Gen-AI models.

6.2 Green Paradox-AI as a guiding perspective in the development of AI

In response to the growing tension between model complexity, energy consumption, and the pursuit of innovation and operational efficiency (FONTOURA et al., 2025), this study introduces the Green Paradox-AI framework, as shown in Figure 13. This theoretical contribution draws from Green AI (ZHU; OTA; DONG, 2022) and Paradox theory (LIM et al., 2023a) to propose an integrated perspective for designing AI systems in contexts that demand both performance and responsibility. Green AI promotes the development of models that prioritise accuracy and sustainability. It challenges the prevailing emphasis on maximum performance at all costs and calls for reduced computational load, energy use, and environmental impact (RAMAN et al., 2024). Paradox theory, in turn, offers a lens through which conflicting objectives, such as innovation and efficiency or interpretability and complexity, are viewed not as dilemmas to be resolved but as tensions to be managed (JARZABKOWSKI et al., 2022). These tensions are seen as productive forces that can drive strategic innovation when held in balance.

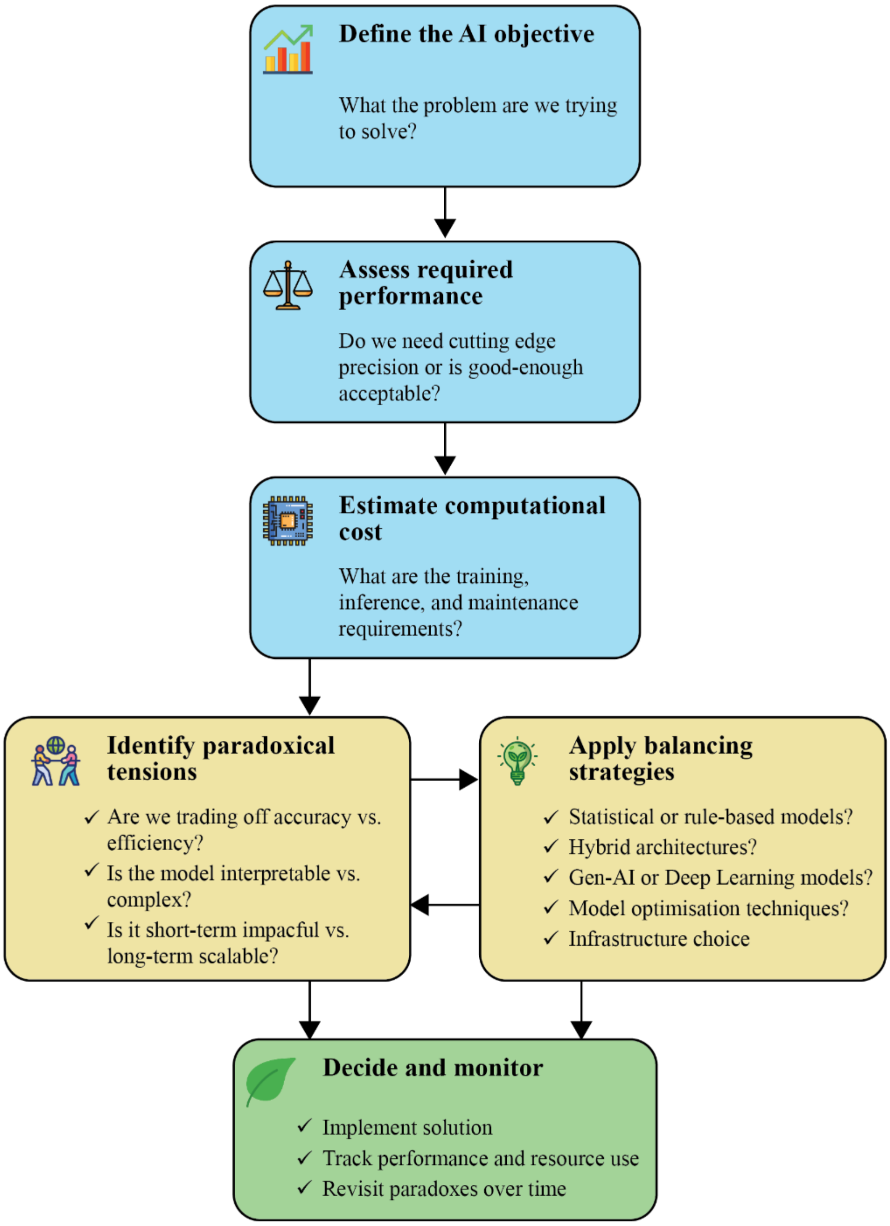


Figure 13 – The Green Paradox-AI framework.

The Green Paradox-AI framework brings together these two perspectives by treating the design of AI systems as a space of intentional contradiction. It recognises that modern decision-support systems must respond to increasing data availability and rising demands for precision while contending with real-world constraints such as limited infrastructure and sustainability targets. Rather than advocating for overly simplified or excessively complex solutions, the framework encourages hybrid approaches that strategically combine traditional and advanced methods. In the context of the GenEneCast model, for example, the integration of Holt-Winters with LSTM reflects a deliberate effort to balance interpretability with adaptive learning. Similarly, the selective use of LMs to automate reporting illustrates how innovation can serve operational goals without unnecessary computational overhead.

By highlighting the coexistence of ecological and organisational pressures, Green Paradox-AI provides a foundation for developing intelligent systems that are technically effective and socially and environmentally aligned. This approach is especially relevant for AI practitioners and decision-makers committed to embedding sustainability into the core of technological innovation. The accompanying decision flowchart translates this framework into a practical tool for AI project management. It outlines six sequential steps to support conscious, strategic choices throughout the model development process. It begins with clearly defining the AI objective, aligning technical efforts with organisational needs. Managers then evaluate the level of predictive performance required, reflecting on whether state-of-the-art precision is necessary or whether a simpler model may suffice. The following step involves estimating the computational demands of the potential solution, including training and inference costs, maintenance, and infrastructure.

Once these foundational elements are in place, the framework invites managers to identify potential paradoxical tensions. These include performance versus efficiency, complexity versus clarity, and short-term gains versus long-term sustainability (BEDI; TOSHNIWAL, 2019; CHARIZANOS; DEMIRHAN; İÇEN, 2024; ZHOU et al., 2025). Rather than treating these tensions as obstacles, the framework encourages managers to embrace them as opportunities to make more informed, balanced decisions. This leads to applying balancing strategies, such as combining statistical and machine learning models, adopting hybrid architectures, or selecting Gen-AI tools where appropriate.

Decisions at this stage are made with attention to strategic impact and environmental footprint. Finally, the framework calls for continuous monitoring of the implemented solution regarding accuracy, resource usage, maintainability, and evolving organisational goals. The cyclical nature of this process reflects a commitment to adaptive management and long-term alignment. The Green Paradox-AI framework thus reframes AI development as a dynamic and iterative process that empowers organisations to navigate complexity without compromising on responsibility. By following the structured decision flow, managers are better equipped to make deliberate, balanced choices that harmonise innovation, performance, and sustainability.

6.3 Implications for engineering management

The empirical findings from the GenEneCast model yield important managerial insights for energy forecasting and wider production operations. Specifically, they add meaningful nuance to ongoing discussions in the literature regarding evaluating and implementing AI-driven tools in operational environments. While a substantial portion of the engineering and operations management discourse remains anchored in performance-centric adoption models that emphasise predictive accuracy above all, the results presented here suggest a broader, more balanced perspective (FILDES et al., 2009; MAKRIDAKIS; SPILIOTIS; ASSIMAKOPOULOS, 2018), the findings presented here align more closely with an emerging view that AI deployment must balance multiple objectives, including computational efficiency, sustainability, and organisational fit (SCHWARTZ et al., 2020).

The trade-offs identified in the GenEneCast model, especially those involving accuracy, execution time, and energy consumption, echo concerns highlighted in recent Green AI research, which emphasises that gains in predictive precision frequently incur disproportionately high energy costs (ZUO et al., 2020). While GenEneCast – Medium achieved the lowest RMSE-like, the cost of this gain, over three times the energy consumption and time of the Low configuration, reflects precisely the diminishing returns problem (ZHANG; GE; CHAI, 2019). Yet, such cost implications are often underrepresented in the literature on forecasting in engineering systems, which tends to evaluate models based almost exclusively on accuracy metrics (CONTI et al., 2023).

The performance achieved by the GenEneCast – Low configuration offers a practical rebuttal to performance-maximalist approaches, supporting the view that models delivering adequate accuracy, aligned with operational requirements and constrained resource availability, may be preferable to those pursuing marginal accuracy gains at substantial computational cost. This perspective aligns with the concept of bounded rationality, which recognises decision-making under limited resources and information (BERGAN; FITZPATRICK, 2024), but also updates it for the era of AI-based decision systems, where computational and environmental costs are increasingly salient managerial constraints.

Furthermore, the configurability of GenEneCast reinforces recent calls for adaptive, modular AI architectures in production systems (GOMES et al., 2023). Its capacity to dynamically tailor model complexity to the specific forecasting context directly addresses concerns raised in the literature, highlighting flexibility as a key requirement for closing the gap between theoretical innovation and practical implementation in operational settings

(REHMAN et al., 2024). Integrating a GLM to support configuration decisions, based on statistical characteristics of the input series, further enhances the model's adaptability. This design enables non-expert users to engage meaningfully with the system while reducing dependence on specialised data science personnel. Such an approach aligns with growing demands for AI systems that are both interpretable and participatory (TURNER; GARN, 2022).

The Green Paradox-AI, formalised in this study, provides a conceptual structure largely absent in current literature. While many studies acknowledge energy and computational costs (LOPEZ-PEREZ et al., 2022), few offer a framework that explicitly models these as strategic constraints in model selection. GenEneCast's empirical demonstration that a less complex model (Low) can remain statistically robust (95% MCS) while reducing energy usage by approximately 78% compared to the High configuration and over 90% compared to the most energy-intensive model (GRU) provides compelling support for this new paradigm.

In summary, the findings of this research both align with and critically extend current literature. They reinforce the growing advocacy for eco-efficient, scalable, and flexible AI solutions in operational contexts, while challenging the prevailing emphasis on accuracy as the primary model selection criterion. The Green Paradox-AI framework, therefore, offers a meaningful theoretical and managerial perspective, reframing AI not merely as a technical instrument for maximising predictive performance, but as a decision-making asset constrained by real-world resource limitations and embedded within complex production environments.

6.4 Managerial insights

The outcomes of this study provide meaningful guidance for decision-makers responsible for deploying forecasting systems in operational settings. A central insight is that model selection should move beyond solely relying on predictive accuracy and instead rely on a balanced assessment that considers forecast quality, computational efficiency, and contextual limitations, an approach increasingly acknowledged in contemporary AI deployment literature (SCHWARTZ et al., 2020). The GenEneCast model, through its configurable architecture, enables scenario-specific strategies by allowing managers to choose the most suitable configuration, Low, Medium, or High, according to their operational priorities. This perspective reinforces the idea that selecting an AI model is not purely a technical exercise, but rather a resource-aware managerial decision, particularly in environments constrained by budget, energy usage, or time (JIN et al., 2025; SARKAR; KHANAPURI; TIWARI, 2025).

In operational contexts where speed, energy efficiency, and scalability are essential, such as embedded systems, low-power infrastructures, or high-frequency forecasting environments, the Low configuration stands out as the preferred choice. It delivers dependable accuracy while substantially lowering execution time and energy usage, thereby reinforcing the rationale behind sustainable AI practices (WHEELDON et al., 2020). Conversely, opting for Medium or High configurations may be justified in high-stakes or mission-critical scenarios, such as fault detection, anomaly identification, or safety monitoring, where even marginal improvements in accuracy can warrant greater computational expenditure. However, such decisions should be guided by a cost-benefit rationale, in line with the principles of the Green AI movement, which cautions against accuracy maximalism when it comes at the expense of energy sustainability (LI et al., 2024).

Selecting a forecasting model constitutes a constrained managerial decision rather than a purely technical optimisation task. The Green Paradox-AI framework developed in this study reinforces this view by demonstrating that pursuing maximum accuracy is not always advantageous, particularly when incremental gains entail disproportionately high computational costs. This interpretation is consistent with the principle of bounded rationality, which emphasises decision-making within real-world limitations (BERGAN; FITZPATRICK, 2024), whereby managers pursue reasonable solutions within the limits of available resources.

From a financial perspective, lowering computational and energy expenditures without sacrificing forecast reliability directly impacts operational profit margins. In budget-constrained environments, especially those reliant on large-scale digital infrastructures or embedded AI systems, choosing lighter models like GenEneCast – Low helps minimise overhead and enhances the return on forecasting investment. This reinforces the role of forecasting as a technical asset and a strategic driver for sustainable and cost-effective operations.

6.5 Practical selection for forecasting models

In practical applications, selecting forecasting models requires balancing multiple competing objectives, most notably, predictive accuracy, processing speed, and energy efficiency. To support this decision-making process, the present study introduces a Pareto frontier graph as a visual tool that helps identify models offering optimal trade-offs across these three dimensions. The graph, presented in Figure 14, simultaneously plots each model according to its RMSE-like value (x-axis), energy consumption in watt-hours (Wh) (y-axis), and execution time in seconds (z-axis).

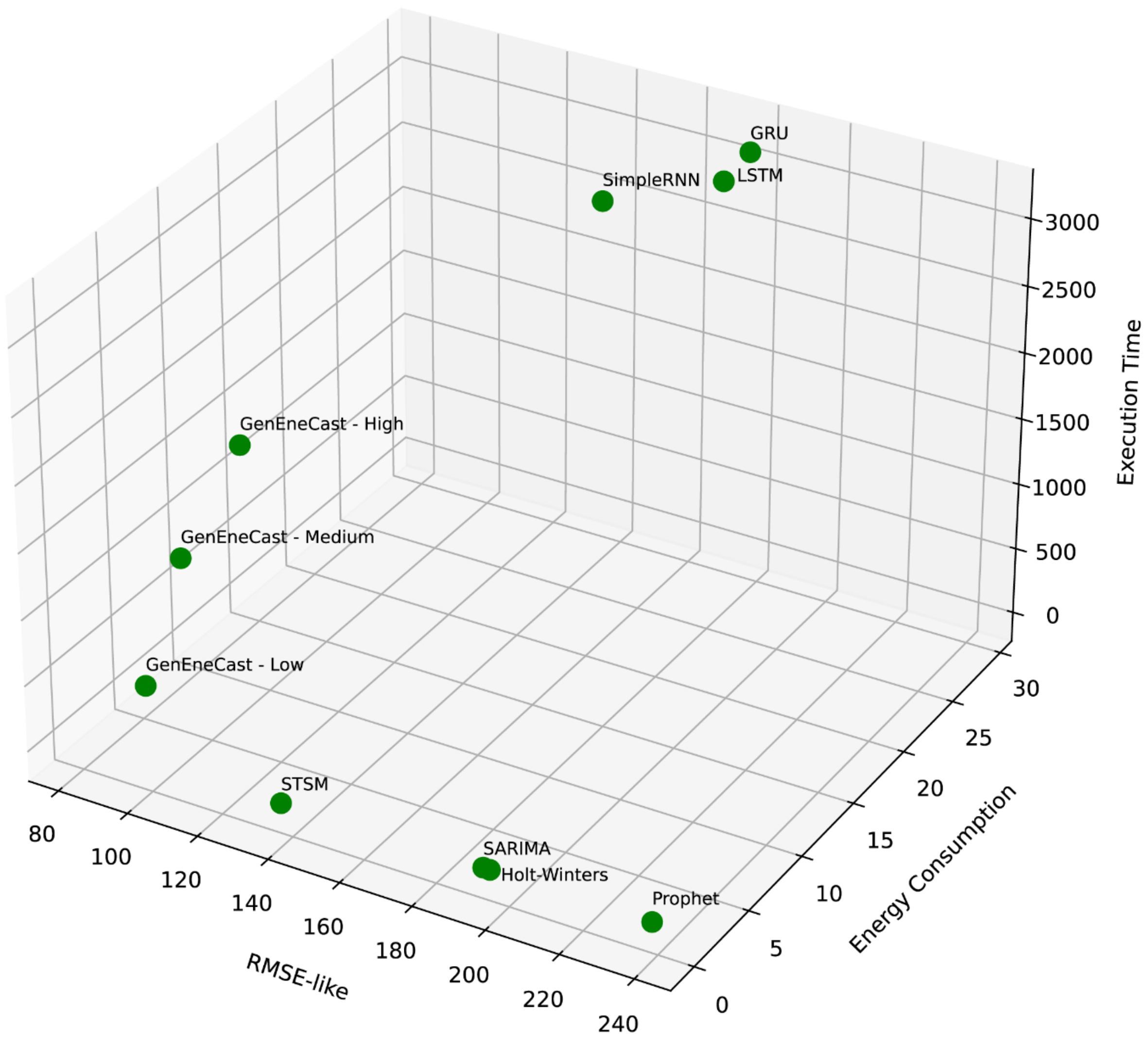


Figure 14 – Pareto Efficiency Frontier of forecasting models.

This approach is grounded in Pareto optimality (CHEN; TIAN, 2025), where a solution is considered efficient if no criterion can be improved without deteriorating at least one of the others. In forecasting, this visualisation provides practitioners and decision-makers with a valuable means of evaluating different models' statistical performance and computational efficiency. This consideration is increasingly critical in sustainability-driven settings. However, it is essential to acknowledge that the specific values represented in this Pareto frontier, namely, energy consumption and execution time, are context-dependent, as they reflect the characteristics of the particular time series dataset and the hardware–software environment employed in this study. Nevertheless, the structure of the Pareto frontier graph is fully replicable across other forecasting contexts. While numerical results will vary depending on the data and implementation, the method remains valid as a comparative framework for evaluating trade-offs across multiple forecasting objectives.

Within this context, the three configurations of the GenEneCast model dominate the frontier. The GenEneCast – Medium configuration represents the most balanced solution, achieving low RMSE-like (82.73), moderate energy use (7.713 Wh), and acceptable execution time (831.08 seconds). The GenEneCast – Low configuration, while slightly less accurate (RMSE-like = 89.23), offers the lowest energy consumption (2.689 Wh) and fastest execution (283.73 seconds), making it ideal for embedded or constrained environments. The GenEneCast – High variant provides a marginal gain in accuracy (RMSE-like = 84.19) at a significantly higher energy cost (12.514 Wh), illustrating a key dynamic in model selection: greater complexity does not always translate into proportional performance gains.

Other models, such as LSTM, GRU, and SimpleRNN, although included within the 95% MCS, remain distant from the efficiency frontier. Their comparatively high energy consumption and execution times indicate reduced suitability for deployment in production settings, especially when more efficient alternatives are available. This reinforces insights from recent literature that caution against overuse of generic deep learning models when simpler or hybrid architectures yield similar or better results at lower cost (AGGA et al., 2021; SAJJAD et al., 2020). Conversely, traditional statistical models such as SARIMA, Holt-Winters, and Prophet exhibit low computational demand but poor forecasting accuracy, limiting their relevance for applications that require precise predictions.

In this study, the Pareto frontier graph is a practical and visually intuitive instrument for guiding multicriteria forecasting decisions. It enables managers, engineers, and data scientists to select models based on technical performance and alignment with energy consumption, response time, and scalability requirements. The consistent dominance of GenEneCast configurations across all three performance axes underscores the model's adaptability and reinforces the core theoretical contribution of this work. Aligned with the Green Paradox-AI framework, these findings demonstrate that optimal forecasting is not achieved through maximal complexity, but through deliberate, strategic balance.

7 CONCLUSION

This study proposed and validated the GenEneCast model, a novel hybrid forecasting framework that integrates Holt-Winters decomposition, LSTM residual modelling, and GLM through GPT-4-Turbo. Across all configurations, GenEneCast achieved competitive forecast accuracy while offering differentiated trade-offs in execution time and energy consumption. Notably, the Low configuration delivered near-optimal predictive performance with a substantial reduction in computational cost, showcasing the model's capacity to balance performance and efficiency under resource constraints. This empirical evidence underpins the Green Paradox-AI, a conceptual framework developed to illuminate the tensions between forecast precision and sustainable AI deployment.

In response to RQ1, expert-based evaluations revealed that RNNs were consistently perceived as the most relevant Gen-AI technology for energy forecasting, outperforming Transformer-based and convolutional alternatives. This informed the architectural foundation of GenEneCast and confirmed alignment between practitioner perception and model design. Regarding RQ2, experimental results demonstrated that Holt-Winters decomposition provided strong complementarity when integrated with LSTM residual learning, outperforming traditional baselines in predictive accuracy and robustness. The residual decomposition approach enhanced signal clarity, enabling the neural component to model trend dynamics more effectively.

In addressing RQ3, the integration of GPT-4-Turbo enabled automated configuration of model hyperparameters and generation of interpretative summaries, reducing technical entry barriers and increasing transparency. This confirmed that GLMs can act as meta-intelligence agents, facilitating usability and interpretability within hybrid AI pipelines. RQ4 was addressed through Pareto frontier analysis and MCS evaluation, which revealed clear trade-offs between model complexity, interpretability, and computational efficiency. The Low configuration, for example, remained within the 95% confidence set while consuming over 90% less energy than more complex deep learning baselines. In response to RQ5, applying GenEneCast to a real-world dataset demonstrated its practical viability. Its modular, adaptive design allowed configuration to local constraints and interpretability needs, suggesting that the framework can support energy management decisions in dynamic, resource-limited environments.

From a theoretical perspective, this study contributes to the energy forecasting literature by introducing Green Paradox-AI as a new lens through which trade-offs in model selection can be evaluated. Rooted in Green AI and Paradox Theory, this framework moves beyond the traditional accuracy-centric paradigm, arguing for models that achieve satisfying performance levels while minimising resource consumption. It also repositions AI-based forecasting as a bounded rationality challenge in engineering management, requiring holistic, context-aware decisions rather than technical optimisation alone.

The findings offer valuable guidance for engineering managers and decision-makers in operations and energy management. Moreover, the model's dynamic configurability and interpretability, facilitated by GPT-4-Turbo, enable broader adoption by users with limited technical expertise, promoting democratised and participatory AI use in industrial settings. The Pareto frontier analysis further supports strategic model selection by revealing GenEneCast's multidimensional efficiency across predictive, computational, and energetic axes.

However, the research also presents certain limitations. The results are based on a single real-world dataset from a Brazilian company, which, although representative, may not capture all the variability present in other sectors or geographies. Future research should explore the generalisability of GenEneCast across broader datasets, integrate additional features such as weather or calendar effects, and assess performance under parallel or cloud-based execution frameworks. Comparative studies incorporating transformer-based architectures or physics-informed neural models could enrich the analysis.

In closing, GenEneCast offers a robust and adaptive approach to forecasting short- and medium-term energy consumption, uniting statistical rigour, neural adaptability, and generative interpretability. By explicitly addressing the performance–efficiency paradox through the Green Paradox-AI framework, the model advances a more sustainable, flexible, and contextually grounded path for forecasting in the era of intelligent operations. Its methodological architecture, interpretative capabilities, and cost-effective deployment make it valuable for engineering research and practical decision-making in data-driven energy systems.

REFERENCES

ABDENNUR, N. et al. Bioframe: operations on genomic intervals in Pandas dataframes. **Bioinformatics**, v. 40, n. 2, 1 fev. 2024.

ADUGNA, T.; XU, W.; FAN, J. Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. **Remote Sensing**, v. 14, n. 3, p. 574, 25 jan. 2022.

AGGA, A. et al. Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models. **Renewable Energy**, v. 177, p. 101–112, nov. 2021.

AHMADI, A. et al. Statistical and deep learning models for reference evapotranspiration time series forecasting: A comparison of accuracy, complexity, and data efficiency. **Computers and Electronics in Agriculture**, v. 215, p. 108424, dez. 2023.

AHSAN, M. et al. Support vector data description with kernel density estimation (SVDD-KDE) control chart for network intrusion monitoring. **Scientific Reports**, v. 13, n. 1, p. 19149, 6 nov. 2023.

AKBARZADEH, O. et al. Heating-Cooling Monitoring and Power Consumption Forecasting Using LSTM for Energy-Efficient Smart Management of Buildings: A Computational Intelligence Solution for Smart Homes. **Tsinghua Science and Technology**, v. 29, n. 1, p. 143–157, 2024.

ALHAYDAR, H.; MOHAMMED, A.; GHAITHAN, A. Feasibility Analysis of Integration Renewable Energy for Industrial Center in Neom City. **Process Integration and Optimization for Sustainability**, v. 8, n. 1, p. 241–258, 7 mar. 2024.

ALZOUBI, Y. I.; MISHRA, A. Green artificial intelligence initiatives: Potentials and challenges. **Journal of Cleaner Production**, v. 468, p. 143090, ago. 2024.

ALZUBAIDI, L. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. **Journal of Big Data**, v. 8, n. 1, p. 53, 31 mar. 2021.

AMIR RAZA, M. et al. Global progress towards the Coal: Tracking coal reserves, coal prices, electricity from coal, carbon emissions and coal phase-out. **Gondwana Research**, v. 139, p. 43–72, mar. 2025.

ANDERSON, S. et al. Asking the right questions: Scoping studies in the commissioning of research on the organisation and delivery of health services. **Health Research Policy and Systems**, v. 6, n. 1, p. 7, 9 dez. 2008.

ARKSEY, H.; O'MALLEY, L. Scoping studies: towards a methodological framework.

International Journal of Social Research Methodology, v. 8, n. 1, p. 19–32, fev. 2005.

ARMSTRONG, R. et al. “Scoping the scope” of a cochrane review. **Journal of Public Health**, v. 33, n. 1, p. 147–150, 1 mar. 2011.

ASHWIN, M. et al. Efficient resource management in 6G communication networks using hybrid quantum deep learning model. **Computers and Electrical Engineering**, v. 106, p. 108565, mar. 2023.

ASLAM, F. N. et al. Trends in participant race and sex reporting in lung cancer phase <scp>III</scp> clinical trials. **Cancer Reports**, v. 6, n. S1, 8 set. 2023.

ASPERTI, A.; TRENTIN, M. Balancing Reconstruction Error and Kullback-Leibler Divergence in Variational Autoencoders. **IEEE Access**, v. 8, p. 199440–199448, 2020.

AZIMI, Z.; HOOSHMAND, R.-A.; SOLEYMANI, S. Energy management considering simultaneous presence of demand responses and electric vehicles in smart industrial grids. **Sustainable Energy Technologies and Assessments**, v. 45, p. 101127, jun. 2021.

BARBIERATO, E.; GATTI, A. Toward Green AI: A Methodological Survey of the Scientific Literature. **IEEE Access**, v. 12, p. 23989–24013, 2024.

BASHIR, T. et al. Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. **Energy Reports**, v. 8, p. 1678–1686, nov. 2022.

BASKERVILLE, R.; WOOD-HARPER, A. T. Diversity in information systems action research methods. **European Journal of Information Systems**, v. 7, n. 2, p. 90–107, 19 jun. 1998.

BATZNER, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. **Nature Communications**, v. 13, n. 1, p. 2453, 4 maio 2022.

BEDI, J.; TOSHNIWAL, D. Deep learning framework to forecast electricity demand. **Applied Energy**, v. 238, p. 1312–1326, mar. 2019.

BERGAN, D. E.; FITZPATRICK, N. Policymaker perceptions of citizen needs: heuristics, accuracy, and partisan differences. **Behavioural Public Policy**, v. 8, n. 3, p. 519–547, 19 jul. 2024.

BOUBAKER, S. et al. Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia. **IEEE Access**, v. 9, p. 36719–36729, 2021.

BOUTAHRI, Y.; TILIOUA, A. Machine learning-based predictive model for thermal comfort and energy optimization in smart buildings. **Results in Engineering**, v. 22, p. 102148, jun. 2024.

BRANDWEIN, M. S. et al. Mucoepidermoid Carcinoma. **The American Journal of Surgical Pathology**, v. 25, n. 7, p. 835–845, jul. 2001.

CARREON-ORTIZ, H. et al. Architecture Optimization of a Non-Linear Autoregressive Neural Networks for Mackey-Glass Time Series Prediction Using Discrete Mycorrhiza Optimization Algorithm. **Micromachines**, v. 14, n. 1, p. 149, 6 jan. 2023.

CARTER, C. R.; KAUFMANN, L.; KETCHEN, D. J. Expect the unexpected: toward a theory of the unintended consequences of sustainable supply chain management. **International Journal of Operations & Production Management**, v. 40, n. 12, p. 1857–1871, 1 set. 2020.

CHARIZANOS, G.; DEMIRHAN, H.; İÇEN, D. An online fuzzy fraud detection framework for credit card transactions. **Expert Systems with Applications**, v. 252, p. 124127, out. 2024.

CHATURVEDI, S. et al. A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. **Energy Policy**, v. 168, p. 113097, set. 2022.

CHEN, L. et al. TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. **Bioinformatics**, v. 36, n. 16, p. 4406–4414, 15 ago. 2020.

CHEN, Q. et al. A service-oriented framework for large-scale documents processing and application via 3D models and feature extraction. **Simulation Modelling Practice and Theory**, v. 133, p. 102903, maio 2024.

CHEN, Y.; TIAN, Z. A crude oil price forecasting framework based on Constraint Guarantee and Pareto Fronts Shrinking Strategy. **Applied Soft Computing**, v. 174, p. 112996, abr. 2025.

CHITNIS, M.; HUNT, L. C. What drives the change in UK household energy expenditure and associated CO2 emissions? Implication and forecast to 2020. **Applied Energy**, v. 94, p. 202–214, jun. 2012.

CHITTY-VENKATA, K. T. et al. A survey of techniques for optimizing transformer inference. **Journal of Systems Architecture**, v. 144, p. 102990, nov. 2023.

COHEN, A.; ALROI, N. Diagnostic action research as an instrument in teacher education. **Journal of Education for Teaching**, v. 7, n. 2, p. 176–186, 7 maio 1981.

CONTI, P. et al. Multi-fidelity surrogate modeling using long short-term memory networks. **Computer Methods in Applied Mechanics and Engineering**, v. 404, p. 115811, fev. 2023.

COOK, D. A. Creating virtual patients using large language models: scalable, global, and low cost. **Medical Teacher**, v. 47, n. 1, p. 40–42, 2 jan. 2025.

COOPER, G. Examining Science Education in ChatGPT: An Exploratory Study of

Generative Artificial Intelligence. **Journal of Science Education and Technology**, v. 32, n. 3, p. 444–452, 22 jun. 2023.

CORNISH, F. et al. Participatory action research. **Nature Reviews Methods Primers**, v. 3, n. 1, p. 34, 27 abr. 2023.

CORNWALL, A.; JEWKES, R. What is participatory research? **Social Science & Medicine**, v. 41, n. 12, p. 1667–1676, dez. 1995.

COUGHLAN, P.; COGHLAN, D. Action research for operations management. **International Journal of Operations & Production Management**, v. 22, n. 2, p. 220–240, 1 fev. 2002.

CUHLS, K.; BLIND, K.; GRUPP, H. **Innovations for our Future: Delphi '98: New Foresight on Science and Technology**. [s.l.] Physica-Verlag HD, 2002.

CUNHA, M. P. E; PUTNAM, L. L. Paradox theory and the paradox of success. **Strategic Organization**, v. 17, n. 1, p. 95–106, fev. 2019.

DAINOTTI, M. G. et al. Inferring the Redshift of More than 150 GRBs with a Machine-learning Ensemble Model. **The Astrophysical Journal Supplement Series**, v. 271, n. 1, p. 22, 1 mar. 2024.

DANESE, P.; MANFÈ, V.; ROMANO, P. A Systematic Literature Review on Recent Lean Research: State-of-the-art and Future Directions. **International Journal of Management Reviews**, v. 20, n. 2, p. 579–605, abr. 2018.

DAUTZENBERG, E. et al. GraphIAST: A graphical user interface software for Ideal Adsorption Solution Theory (IAST) calculations. **Computer Physics Communications**, v. 280, p. 108494, nov. 2022.

DE LIMA, F. A.; SEURING, S. A Delphi study examining risk and uncertainty management in circular supply chains. **International Journal of Production Economics**, v. 258, p. 108810, abr. 2023.

DEIVANAYAGAM, T. A. et al. Envisioning environmental equity: climate change, health, and racial justice. **The Lancet**, v. 402, n. 10395, p. 64–78, jul. 2023.

DEKKER, M.; VAN DONSELAAR, K.; OUWEHAND, P. How to use aggregation and combined forecasting to improve seasonal demand forecasts. **International Journal of Production Economics**, v. 90, n. 2, p. 151–167, jul. 2004.

DENG, C.; LIAO, S.; YUAN, B. PermCNN: Energy-Efficient Convolutional Neural Network Hardware Architecture with Permuted Diagonal Structure. **IEEE Transactions on Computers**, v. 70, n. 2, p. 163–173, 2021.

DI PASQUALE, V.; MIRANDA, S.; NEUMANN, W. P. Ageing and human-system errors in manufacturing: a scoping review. **International Journal of Production Research**, v. 58, n. 15, p. 4716–4740, 2 ago. 2020.

DILAVER, Z.; HUNT, L. C. Industrial electricity demand for Turkey: A structural time series analysis. **Energy Economics**, v. 33, n. 3, p. 426–436, maio 2011.

DIVYANI SEN. Explainable Deep Learning for Time Series Analysis: Integrating SHAP and LIME in LSTM-Based Models. **Journal of Information Systems Engineering and Management**, v. 10, n. 16s, p. 412–423, 6 mar. 2025.

DURMUS SENYAPAR, H. N.; AKSOZ, A. Empowering Sustainability: A Consumer-Centric Analysis Based on Advanced Electricity Consumption Predictions. **Sustainability**, v. 16, n. 7, p. 2958, 2 abr. 2024.

DUTTA BARUAH, R.; MUÑOZ ORGANERO, M. Explicit Context Integrated Recurrent Neural Network for applications in smart environments. **Expert Systems with Applications**, v. 255, p. 124752, dez. 2024.

EHRHART, M. et al. A Conditional GAN for Generating Time Series Data for Stress Detection in Wearable Physiological Sensor Data. **Sensors**, v. 22, n. 16, p. 5969, 10 ago. 2022.

EL HACHIMI, C. et al. Data Science Toolkit: An all-in-one python library to help researchers and practitioners in implementing data science-related algorithms with less effort. **Software Impacts**, v. 12, p. 100240, maio 2022.

ELMOUSALAMI, H.; ELSHABOURY, N.; ELYAMANY, A. H. Green artificial intelligence for cost-duration variance prediction (CDVP) for irrigation canals rehabilitation projects. **Expert Systems with Applications**, v. 249, p. 123789, set. 2024.

ESCORCIA-GUTIERREZ, J. et al. Privacy Preserving Blockchain with Energy Aware Clustering Scheme for IoT Healthcare Systems. **Mobile Networks and Applications**, v. 29, n. 1, p. 1–12, 2024.

EUFRASIO ESPINOSA, R. M.; LENNY KOH, S. C. Forecasting the ecological footprint of G20 countries in the next 30 years. **Scientific Reports**, v. 14, n. 1, p. 8298, 9 abr. 2024.

FARAMARZI, A. et al. Equilibrium optimizer: A novel optimization algorithm. **Knowledge-Based Systems**, v. 191, p. 105190, mar. 2020.

FATIMA, T.; XIA, E.; AHAD, M. Oil demand forecasting for China: a fresh evidence from structural time series analysis. **Environment, Development and Sustainability**, v. 21, n. 3, p. 1205–1224, 13 jun. 2019.

FILDES, R. et al. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. **International Journal of**

Forecasting, v. 25, n. 1, p. 3–23, jan. 2009.

FONTOURA, L. et al. Real-time energy flow mapping: A VSM-based proposal for energy efficiency. **Journal of Cleaner Production**, v. 419, p. 137871, set. 2023.

FONTOURA, L. et al. Energy Gen-AI technology framework: A perspective of energy efficiency and business ethics in operation management. **Technology in Society**, v. 81, p. 102847, jun. 2025.

FOSSO WAMBA, S. et al. Are both generative AI and ChatGPT game changers for 21st-Century operations and supply chain excellence? **International Journal of Production Economics**, v. 265, p. 109015, nov. 2023.

GAO, X. et al. Model selection with decision support model for US natural gas consumption forecasting. **Expert Systems with Applications**, v. 217, p. 119505, maio 2023.

GAO, Y. et al. Successful application of predictive information in deep reinforcement learning control: A case study based on an office building HVAC system. **Energy**, v. 291, p. 130344, mar. 2024.

GARCIA-BUENDIA, N. et al. Performance measurement of lean supply chain management: a balanced scorecard proposal. **Production Planning & Control**, p. 1–21, 28 set. 2022.

GARZA-REYES, J. A. Lean and green – a systematic review of the state of the art literature. **Journal of Cleaner Production**, v. 102, p. 18–29, set. 2015.

GIANNETTI, C. et al. Deep learning for robust forecasting of hot metal silicon content in a blast furnace. **The International Journal of Advanced Manufacturing Technology**, v. 136, n. 9, p. 4045–4054, 1 fev. 2025.

GOMES, J. et al. A scientific software ecosystem architecture for the livestock domain. **Information and Software Technology**, v. 160, p. 107240, ago. 2023.

GOODFELLOW, I. et al. Generative adversarial networks. **Communications of the ACM**, v. 63, n. 11, p. 139–144, 22 out. 2020.

GRIMELL, J. Moral injury: understanding Swedish veterans who are assessed but not diagnosed with PTSD. **Frontiers in Psychiatry**, v. 14, 4 dez. 2023.

HAIDER, M. H.; KO, S.-B. Booth Encoding-Based Energy Efficient Multipliers for Deep Learning Systems. **IEEE Transactions on Circuits and Systems II: Express Briefs**, v. 70, n. 6, p. 2241–2245, 2023.

HAMIDU, Z.; BOACHIE-MENSAH, F. O.; ISSAU, K. Supply chain resilience and performance of manufacturing firms: role of supply chain disruption. **Journal of Manufacturing Technology Management**, v. 34, n. 3, p. 361–382, 28 abr. 2023.

HANSEN, P. R.; LUNDE, A.; NASON, J. M. The Model Confidence Set. **Econometrica**, v. 79, n. 2, p. 453–497, maio 2011.

HARROU, F. et al. A Data-Driven Soft Sensor to Forecast Energy Consumption in Wastewater Treatment Plants: A Case Study. **IEEE Sensors Journal**, v. 21, n. 4, p. 4908–4917, 15 fev. 2021.

HASSAN, W. H. et al. Evaluation of gene expression programming and artificial neural networks in PyTorch for the prediction of local scour depth around a bridge pier. **Results in Engineering**, v. 13, p. 100353, mar. 2022.

HEWAMALAGE, H.; BERGMEIR, C.; BANDARA, K. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. **International Journal of Forecasting**, v. 37, n. 1, p. 388–427, jan. 2021.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1 nov. 1997.

HOGAN, A. et al. Knowledge Graphs. **ACM Computing Surveys**, v. 54, n. 4, p. 1–37, 31 maio 2022.

HOLT, C. C. Forecasting seasonals and trends by exponentially weighted moving averages. **International Journal of Forecasting**, v. 20, n. 1, p. 5–10, jan. 2004.

HOSNY, A. et al. Artificial intelligence in radiology. **Nature Reviews Cancer**, v. 18, n. 8, p. 500–510, 17 ago. 2018.

HUANG, R. et al. Longevity-aware energy management for fuel cell hybrid electric bus based on a novel proximal policy optimization deep reinforcement learning framework. **Journal of Power Sources**, v. 561, p. 232717, mar. 2023.

HUBEL, D. H.; WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. **The Journal of Physiology**, v. 160, n. 1, p. 106–154, jan. 1962.

HUSSAIN, A.; RAHMAN, M.; MEMON, J. A. Forecasting electricity consumption in Pakistan: the way forward. **Energy Policy**, v. 90, p. 73–80, mar. 2016.

HUSSAIN, M.; MICHELUSI, N. Learning and Adaptation for Millimeter-Wave Beam Tracking and Training: A Dual Timescale Variational Framework. **IEEE Journal on Selected Areas in Communications**, v. 40, n. 1, p. 37–53, jan. 2022.

HYNDMAN, R. J. et al. A state space framework for automatic forecasting using exponential smoothing methods. **International Journal of Forecasting**, v. 18, n. 3, p. 439–454, jul. 2002.

IFTIKHAR, S. et al. HunterPlus: AI based energy-efficient task scheduling for cloud–

fog computing environments. **Internet of Things (Netherlands)**, v. 21, 2023.

JACKSON, I. et al. Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation. **International Journal of Production Research**, p. 1–26, 31 jan. 2024.

JAEGER, H.; HAAS, H. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. **Science**, v. 304, n. 5667, p. 78–80, 2 abr. 2004.

JARZABKOWSKI, P. et al. Enabling Rapid Financial Response to Disasters: Knotting and Reknotting Multiple Paradoxes in Interorganizational Systems. **Academy of Management Journal**, v. 65, n. 5, p. 1477–1506, out. 2022.

JAUBERT, O. et al. <scp>HyperSLICE</scp> : <scp>HyperBand</scp> optimized spiral for low-latency interactive cardiac examination. **Magnetic Resonance in Medicine**, v. 91, n. 1, p. 266–279, 6 jan. 2024.

JAY, J. Navigating Paradox as a Mechanism of Change and Innovation in Hybrid Organizations. **Academy of Management Journal**, v. 56, n. 1, p. 137–159, fev. 2013.

JIANG, D. et al. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. **Journal of Cheminformatics**, v. 13, n. 1, p. 12, 17 fev. 2021.

JIANG, H. et al. Fragility assessment of sea-crossing cable-stayed bridge subjected to multi-hazard action via TKC and R-vine copula. **Engineering Structures**, v. 307, p. 117874, maio 2024a.

JIANG, Y. et al. A Flow Rate Estimation Method for Gas–Liquid Two-Phase Flow Based on Transformer Neural Network. **IEEE Sensors Journal**, v. 24, n. 16, p. 26902–26913, 15 ago. 2024b.

JIANG, Y.; DALE, R.; LU, H. Transformability, generalizability, but limited diffusibility: Comparing global vs. task-specific language representations in deep neural networks. **Cognitive Systems Research**, v. 83, p. 101184, jan. 2024.

JIN, H. et al. Vesper: A Versatile Sparse Linear Algebra Accelerator With Configurable Compute Patterns. **IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems**, v. 44, n. 5, p. 1731–1744, maio 2025.

JOSEPH, L. P. et al. Near real-time wind speed forecast model with bidirectional LSTM networks. **Renewable Energy**, v. 204, p. 39–58, mar. 2023.

KAISER, M. A.-A. et al. Neuromorphic-P2M: processing-in-pixel-in-memory paradigm for neuromorphic image sensors. **Frontiers in Neuroinformatics**, v. 17, 2023.

KANBACH, D. K. et al. The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective. **Review of Managerial Science**, 13 set. 2023.

KARAKAN, A. Predicting Energy Production in Renewable Energy Power Plants Using Deep Learning. **Energies**, v. 17, n. 16, p. 4031, 14 ago. 2024.

KHALESIAN, M.; FURNO, A.; LECLERCQ, L. Improving deep-learning methods for area-based traffic demand prediction via hierarchical reconciliation. **Transportation Research Part C: Emerging Technologies**, v. 159, p. 104410, fev. 2024.

KHOO, Z. DE et al. Forecasting volatility of stock indices: Improved GARCH-type models through combined weighted volatility measure and weighted volatility indicators. **The North American Journal of Economics and Finance**, v. 71, p. 102112, mar. 2024.

KHOT, A. C. et al. 2D Ti3C2T MXene-derived self-assembled 3D TiO2nanoflowers for nonvolatile memory and synaptic learning applications. **Journal of Materials Science & Technology**, v. 150, p. 1–10, jul. 2023.

KIM, H.; PARK, S.; KIM, S. Time-series clustering and forecasting household electricity demand using smart meter data. **Energy Reports**, v. 9, p. 4111–4121, dez. 2023.

KINGMA, D. P.; WELLING, M. An Introduction to Variational Autoencoders. **Foundations and Trends® in Machine Learning**, v. 12, n. 4, p. 307–392, 2019.

KOHTAMÄKI, M.; EINOLA, S.; RABETINO, R. Exploring servitization through the paradox lens: Coping practices in servitization. **International Journal of Production Economics**, v. 226, p. 107619, ago. 2020.

KOLOSOK, S. et al. Renewable Energy and Energy Innovations: Examining Relationships Using Markov Switching Regression Model. **Marketing and Management of Innovations**, v. 2, n. 1, p. 151–160, 2022.

KOMALA, C. R. et al. Multi-UAV computing enabling efficient clustering-based IoT for energy reduction and data transmission. **Journal of Intelligent and Fuzzy Systems**, v. 45, n. 1, p. 1717–1730, 2023.

KONG, W. et al. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. **IEEE Transactions on Smart Grid**, v. 10, n. 1, p. 841–851, jan. 2019.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. **Communications of the ACM**, v. 60, n. 6, p. 84–90, 24 maio 2017.

KUMAR, P.; GUPTA, S.; DAGAR, V. Sustainable energy development through non-residential rooftop solar photovoltaic adoption: Empirical evidence from India. **Sustainable Development**, v. 32, n. 1, p. 795–814, fev. 2024.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p.

436–444, 28 maio 2015.

LEWIN, K. Action Research and Minority Problems. **Journal of Social Issues**, v. 2, n. 4, p. 34–46, 14 nov. 1946.

LI, H. et al. Slim-neck by GSConv: a lightweight-design for real-time detector architectures. **Journal of Real-Time Image Processing**, v. 21, n. 3, p. 62, 29 jun. 2024.

LI, Q. et al. A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection. **IEEE Transactions on Knowledge and Data Engineering**, v. 35, n. 4, p. 3347–3366, 1 abr. 2023.

LI, T. et al. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. **Mechanical Systems and Signal Processing**, v. 168, p. 108653, abr. 2022.

LIAO, W. et al. Generative AI design for building structures. **Automation in Construction**, v. 157, p. 105187, jan. 2024.

LIEDTKE, C. et al. User-integrated innovation in Sustainable LivingLabs: an experimental infrastructure for researching and developing sustainable product service systems. **Journal of Cleaner Production**, v. 97, p. 106–116, jun. 2015.

LIM, W. M. et al. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. **The International Journal of Management Education**, v. 21, n. 2, p. 100790, jul. 2023a.

LIM, Z. W. et al. Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. **eBioMedicine**, v. 95, p. 104770, set. 2023b.

LIN, J. et al. Decision-Oriented Dialogue for Human-AI Collaboration. **Transactions of the Association for Computational Linguistics**, v. 12, p. 892–911, 1 ago. 2024.

LITJENS, G. et al. A survey on deep learning in medical image analysis. **Medical Image Analysis**, v. 42, p. 60–88, dez. 2017.

LIU, J. et al. Machine learning-based techniques for land subsidence simulation in an urban area. **Journal of Environmental Management**, v. 352, p. 120078, fev. 2024.

LIU, X. et al. Collaborative Edge Computing with FPGA-Based CNN Accelerators for Energy-Efficient and Time-Aware Face Tracking System. **IEEE Transactions on Computational Social Systems**, v. 9, n. 1, p. 252–266, 2022.

LOPEZ-PEREZ, D. et al. A Survey on 5G Radio Access Network Energy Efficiency: Massive MIMO, Lean Carrier Design, Sleep Modes, and Machine Learning. **IEEE Communications Surveys & Tutorials**, v. 24, n. 1, p. 653–697, 2022.

LU, M. et al. A Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting. **Water**, v. 15, n. 7, p. 1265, 23 mar. 2023.

MA, J. et al. A review of graph neural networks and pretrained language models for knowledge graph reasoning. **Neurocomputing**, v. 609, p. 128490, dez. 2024.

MACNAMARA, S.; CUNNINGHAM, P.; BYRNE, J. Neural networks for language identification: a comparative study. **Information Processing & Management**, v. 34, n. 4, p. 395–403, jul. 1998.

MAGINGA, T. J. et al. Design and implementation of IoT sensors for nonvisual symptoms detection on maize inoculated with Exserohilum turcicum. **Smart Agricultural Technology**, v. 5, p. 100260, out. 2023.

MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. **PLOS ONE**, v. 13, n. 3, p. e0194889, 27 mar. 2018.

MANIKANDAN, S. et al. Efficient energy consumption in hybrid cloud environment using adaptive backtracking virtual machine consolidation. **Scientific Reports**, v. 14, n. 1, p. 22869, 2 out. 2024.

MARTINS, A. et al. Unlocking human-like conversations: Scoping review of automation techniques for personalized healthcare interventions using conversational agents. **International Journal of Medical Informatics**, v. 185, p. 105385, maio 2024.

MASOOD, M. et al. Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. **Applied Intelligence**, v. 53, n. 4, p. 3974–4026, 4 fev. 2023.

MAY, R. M. et al. MetPy: A Meteorological Python Library for Data Analysis and Visualization. **Bulletin of the American Meteorological Society**, v. 103, n. 10, p. E2273–E2284, out. 2022.

MESNIL, G. et al. Using Recurrent Neural Networks for Slot Filling in Spoken Language Understanding. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 23, n. 3, p. 530–539, mar. 2015.

MIOK, K. et al. To BAN or Not to BAN: Bayesian Attention Networks for Reliable Hate Speech Detection. **Cognitive Computation**, v. 14, n. 1, p. 353–371, 26 jan. 2022.

MOHAMMADI, M. et al. Semisupervised Deep Reinforcement Learning in Support of IoT and Smart City Services. **IEEE Internet of Things Journal**, v. 5, n. 2, p. 624–635, abr. 2018.

MOUSAVI, R. et al. Revolutionizing solar energy resources: The central role of generative AI in elevating system sustainability and efficiency. **Applied Energy**, v. 382, p. 125296, mar. 2025.

MOZO, A. et al. A Machine-Learning-Based Cyberattack Detector for a Cloud-Based SDN Controller. **Applied Sciences (Switzerland)**, v. 13, n. 8, 2023.

MUNN, Z. et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. **BMC Medical Research Methodology**, v. 18, n. 1, p. 143, 19 dez. 2018.

NASCIMENTO, D. L. DE M. et al. Circular value stream mapping 4.0: Proposed general model and application to a digital 3D printing recycling factory. **Sustainable Production and Consumption**, v. 34, p. 600–612, nov. 2022.

NGO, N.-T. et al. Developing a hybrid time-series artificial intelligence model to forecast energy use in buildings. **Scientific Reports**, v. 12, n. 1, p. 15775, 21 set. 2022.

NISSEN, J. N. et al. Improved metagenome binning and assembly using deep variational autoencoders. **Nature Biotechnology**, v. 39, n. 5, p. 555–560, 4 maio 2021.

NIZAM, H. et al. Real-Time Deep Anomaly Detection Framework for Multivariate Time-Series Data in Industrial IoT. **IEEE Sensors Journal**, v. 22, n. 23, p. 22836–22849, 1 dez. 2022.

NÚÑEZ-MERINO, M. et al. Information and digital technologies of Industry 4.0 and Lean supply chain management: a systematic literature review. **International Journal of Production Research**, v. 58, n. 16, p. 5034–5061, 17 ago. 2020.

OLADEJO, S. O.; EKWE, S. O.; MIRJALILI, S. The Hiking Optimization Algorithm: A novel human-based metaheuristic approach. **Knowledge-Based Systems**, v. 296, p. 111880, jul. 2024.

ONAN, A. Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. **Journal of King Saud University - Computer and Information Sciences**, v. 34, n. 5, p. 2098–2117, maio 2022.

OSABA, E. et al. An improved discrete bat algorithm for symmetric and asymmetric Traveling Salesman Problems. **Engineering Applications of Artificial Intelligence**, v. 48, p. 59–71, fev. 2016.

PAVLATOS, C. et al. Utilization of Artificial Neural Networks for Precise Electrical Load Prediction. **Technologies**, v. 11, n. 3, p. 70, 26 maio 2023.

PETIT, H. J. et al. Exploring Barriers and Facilitators to Reducing the Environmental Impact of the Operating Room. **Journal of Surgical Research**, v. 292, p. 197–205, dez. 2023.

PIRES, C.; MARTINS, M. V. Enhancing Water Management: A Comparative Analysis of Time Series Prediction Models for Distributed Water Flow in Supply Networks. **Water**, v. 16, n. 13, p. 1827, 27 jun. 2024.

PITTA, F. et al. Physical Activity and Hospitalization for Exacerbation of COPD. **Chest**, v. 129, n. 3, p. 536–544, mar. 2006.

POLLOCK, M. et al. What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative metasummary. **Systematic Reviews**, v. 5, n. 1, p. 190, 14 dez. 2016.

PUSHPAVALLI, M. et al. Enhancing Electrical Power Demand Prediction Using LSTM-Based Deep Learning Models for Local Energy Communities. **Electric Power Components and Systems**, p. 1–18, 4 abr. 2024.

QIN, M. et al. Low-carbon economic multi-objective dispatch of integrated energy system considering the price fluctuation of natural gas and carbon emission accounting. **Protection and Control of Modern Power Systems**, v. 8, n. 1, p. 61, 23 dez. 2023.

RAHMAN, A.; SRIKUMAR, V.; SMITH, A. D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. **Applied Energy**, v. 212, p. 372–385, fev. 2018.

RAMAN, R. et al. Green and sustainable AI research: an integrated thematic and topic modeling analysis. **Journal of Big Data**, v. 11, n. 1, p. 55, 22 abr. 2024.

REEDHA, R. et al. Transformer Neural Network for Weed and Crop Classification of High Resolution UAV Images. **Remote Sensing**, v. 14, n. 3, p. 592, 26 jan. 2022.

REHMAN, H. U. et al. A modular artificial intelligence and asset administration shell approach to streamline testing processes in manufacturing services. **Journal of Manufacturing Systems**, v. 72, p. 424–436, fev. 2024.

REN, Y. et al. A novel Q-learning based variable neighborhood iterative search algorithm for solving disassembly line scheduling problems. **Swarm and Evolutionary Computation**, v. 80, p. 101338, jul. 2023.

ROSENFELD, R. Two decades of statistical language modeling: where do we go from here? **Proceedings of the IEEE**, v. 88, n. 8, p. 1270–1278, ago. 2000.

RUAN, J. et al. On Vulnerability of Renewable Energy Forecasting: Adversarial Learning Attacks. **IEEE Transactions on Industrial Informatics**, v. 20, n. 3, p. 3650–3663, mar. 2024.

SAEED, M. H.; HAMA, J. I. Cardiac disease prediction using AI algorithms with SelectKBest. **Medical & Biological Engineering & Computing**, v. 61, n. 12, p. 3397–3408, 8 dez. 2023.

SAIEG, P. et al. Interactions of Building Information Modeling, Lean and Sustainability on the Architectural, Engineering and Construction industry: A systematic review. **Journal of Cleaner Production**, v. 174, p. 788–806, fev. 2018.

SAJJAD, M. et al. A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting. **IEEE Access**, v. 8, p. 143759–143768, 2020.

SAPTALAKAR, B. K.; LATTE, M. V. FPGA-based reflection image removal using cognitive neural networks. **Applied Nanoscience (Switzerland)**, v. 13, n. 3, p. 2539–2553, 2023.

SARKAR, P.; KHANAPURI, V. B.; TIWARI, M. K. Accelerating the stabilized column generation using machine learning. **Computers & Industrial Engineering**, v. 200, p. 110837, fev. 2025.

SCARSELLI, F. et al. The Graph Neural Network Model. **IEEE Transactions on Neural Networks**, v. 20, n. 1, p. 61–80, jan. 2009.

SCHREYÖGG, G.; SYDOW, J. CROSSROADS—Organizing for Fluidity? Dilemmas of New Organizational Forms. **Organization Science**, v. 21, n. 6, p. 1251–1262, dez. 2010.

SCHWALLER, P. et al. Extraction of organic chemistry grammar from unsupervised learning of chemical reactions. **Science Advances**, v. 7, n. 15, 9 abr. 2021.

SCHWARTZ, R. et al. Green AI. **Communications of the ACM**, v. 63, n. 12, p. 54–63, 17 nov. 2020.

SERRANO, A. L. M. et al. Statistical Comparison of Time Series Models for Forecasting Brazilian Monthly Energy Demand Using Economic, Industrial, and Climatic Exogenous Variables. **Applied Sciences**, v. 14, n. 13, p. 5846, 4 jul. 2024.

SHAH, H. et al. Can machines talk? Comparison of Eliza with modern dialogue systems. **Computers in Human Behavior**, v. 58, p. 278–295, maio 2016.

SHAO, Z. et al. Region-Object Relation-Aware Dense Captioning via Transformer. **IEEE Transactions on Neural Networks and Learning Systems**, v. 36, n. 3, p. 4184–4195, mar. 2025.

SHEHADEH, A. et al. An expert system for highway construction: Multi-objective optimization using enhanced particle swarm for optimal equipment management. **Expert Systems with Applications**, v. 249, p. 123621, set. 2024.

SHI, S.; XIONG, H.; LI, G. A no-tardiness job shop scheduling problem with overtime consideration and the solution approaches. **Computers & Industrial Engineering**, v. 178, p. 109115, abr. 2023.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on Image Data Augmentation for

Deep Learning. **Journal of Big Data**, v. 6, n. 1, p. 60, 6 dez. 2019.

SHRESTHA, A.; MAHMOOD, A. Review of Deep Learning Algorithms and Architectures. **IEEE Access**, v. 7, p. 53040–53065, 2019.

SISUTHOG, W.; ATTANATHO, L.; CHAIYA, C. Conversion of empty fruit bunches (EFBs) by hydrothermal carbonization towards hydrochar production. **Energy Reports**, v. 8, p. 242–248, dez. 2022.

SMITH, W. K.; GONIN, M.; BESHAROV, M. L. Managing Social-Business Tensions: A Review and Research Agenda for Social Enterprise. **Business Ethics Quarterly**, v. 23, n. 3, p. 407–442, 23 jul. 2013.

SMITH, W. K.; LEWIS, M. W. Toward a Theory of Paradox: A Dynamic equilibrium Model of Organizing. **Academy of Management Review**, v. 36, n. 2, p. 381–403, abr. 2011.

SUBBIAH, S. S.; CHINNAPPAN, J. Deep learning based short term load forecasting with hybrid feature selection. **Electric Power Systems Research**, v. 210, p. 108065, set. 2022.

SUEUR, M. et al. PyC2MC: An Open-Source Software Solution for Visualization and Treatment of High-Resolution Mass Spectrometry Data. **Journal of the American Society for Mass Spectrometry**, v. 34, n. 4, p. 617–626, 5 abr. 2023.

SUMATHI, A. C. et al. NEWTR: a multipath routing for next hop destination in internet of things with artificial recurrent neural network (RNN). **International Journal of Machine Learning and Cybernetics**, v. 13, n. 10, p. 2869–2889, 2022.

SUN, K. DMFF: Deep multimodel feature fusion for building occupancy detection. **Building and Environment**, v. 253, 2024.

SUSMAN, G. I.; EVERED, R. D. An Assessment of the Scientific Merits of Action Research. **Administrative Science Quarterly**, v. 23, n. 4, p. 582, dez. 1978.

THEIS, T. N.; WONG, H.-S. P. The End of Moore's Law: A New Beginning for Information Technology. **Computing in Science & Engineering**, v. 19, n. 2, p. 41–50, mar. 2017.

TRUHN, D. et al. Large language models and multimodal foundation models for precision oncology. **npj Precision Oncology**, v. 8, n. 1, p. 72, 22 mar. 2024.

TURCO, M.; LLASAT, M. C. Trends in indices of daily precipitation extremes in Catalonia (NE Spain), 1951–2003. **Natural Hazards and Earth System Sciences**, v. 11, n. 12, p. 3213–3226, 9 dez. 2011.

TURKI, E. et al. Forecasting spare part extractions from returned systems in a closed-loop supply chain. **International Journal of Production Research**, v. 62, n. 21, p. 7860–7876, 30 nov. 2024.

TURNER, C. J.; GARN, W. Next generation DES simulation: A research agenda for human centric manufacturing systems. **Journal of Industrial Information Integration**, v. 28, p. 100354, jul. 2022.

VÄRTINEN, S.; HÄMÄLÄINEN, P.; GUCKELSBERGER, C. Generating Role-Playing Game Quests With GPT Language Models. **IEEE Transactions on Games**, v. 16, n. 1, p. 127–139, mar. 2024.

VEIGA, C. P. DA et al. Demand forecasting based on natural computing approaches applied to the foodstuff retail segment. **Journal of Retailing and Consumer Services**, v. 31, p. 174–181, jul. 2016.

VIJAY, U. et al. A hybrid modelling approach coupling physics-based simulation and deep learning for battery electrode manufacturing simulations. **Energy Storage Materials**, v. 75, p. 103883, fev. 2025.

VOULODIMOS, A. et al. Deep Learning for Computer Vision: A Brief Review. **Computational Intelligence and Neuroscience**, v. 2018, p. 1–13, 2018.

WAEL AL-KHATIB, A. Drivers of generative artificial intelligence to fostering exploitative and exploratory innovation: A TOE framework. **Technology in Society**, v. 75, p. 102403, nov. 2023.

WALKER, H. L. et al. Reliability of Medical Information Provided by ChatGPT: Assessment Against Clinical Guidelines and Patient Information Quality Instrument. **Journal of Medical Internet Research**, v. 25, p. e47479, 30 jun. 2023.

WANG, C. et al. Minute-level ultra-short-term power load forecasting based on time series data features. **Applied Energy**, v. 372, p. 123801, out. 2024a.

WANG, D.; ZHANG, Y.; YOUSEFI, N. Urban Water-Energy consumption Prediction Influenced by Climate Change utilizing an innovative deep learning method. **Scientific Reports**, v. 14, n. 1, p. 30931, 28 dez. 2024.

WANG, Q.; LI, Y.; LI, R. Integrating artificial intelligence in energy transition: A comprehensive review. **Energy Strategy Reviews**, v. 57, p. 101600, jan. 2025.

WANG, W. et al. A novel forecasting framework leveraging large language model and machine learning for methanol price. **Energy**, v. 320, p. 135123, abr. 2025.

WANG, Z. et al. EDCompress: Energy-Aware Model Compression for Dataflows. **IEEE Transactions on Neural Networks and Learning Systems**, v. 35, n. 1, p. 208–220, 2024b.

WARIS, U.; TARIQ, S. Temporal dynamics and forecasting of aerosol optical depth in megacities Lahore and Karachi: Insights from the Indo-Gangetic Basin and southern Pakistan, and implications for Sustainable development. **Atmospheric Pollution Research**, v. 15, n. 7, p. 102146, jul. 2024.

WEBSTER, J.; WATSON, R. T. Analyzing the Past to Prepare for the Future: Writing a Literature Review. **MIS Quarterly**, v. 26, n. 2, p. xiii–xxiii, 24 dez. 2002.

WEIZENBAUM, J. ELIZA—a computer program for the study of natural language communication between man and machine. **Communications of the ACM**, v. 9, n. 1, p. 36–45, jan. 1966.

WHEELDON, A. et al. Learning automata based energy-efficient AI hardware design for IoT applications. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 378, n. 2182, p. 20190593, 16 out. 2020.

WINTERS, P. R. Forecasting Sales by Exponentially Weighted Moving Averages. **Management Science**, v. 6, n. 3, p. 324–342, abr. 1960.

WOLTERINK, J. M. et al. Generative Adversarial Networks for Noise Reduction in Low-Dose CT. **IEEE Transactions on Medical Imaging**, v. 36, n. 12, p. 2536–2545, dez. 2017.

WU, Z. et al. A Comprehensive Survey on Graph Neural Networks. **IEEE Transactions on Neural Networks and Learning Systems**, v. 32, n. 1, p. 4–24, jan. 2021.

XIA, M. et al. A Stacked GRU-RNN-Based Approach for Predicting Renewable Energy and Electricity Load for Smart Grid Operation. **IEEE Transactions on Industrial Informatics**, v. 17, n. 10, p. 7050–7059, out. 2021.

XIAO, M. et al. High-entropy materials for solid oxide cells: Synthesis, applications, and prospects. **Journal of Energy Chemistry**, v. 104, p. 268–296, maio 2025.

XUE, J. et al. Cooperative Deep Reinforcement Learning Enabled Power Allocation for Packet Duplication URLLC in Multi-Connectivity Vehicular Networks. **IEEE Transactions on Mobile Computing**, v. 23, n. 8, p. 8143–8157, ago. 2024.

YAMASHITA, R. et al. Convolutional neural networks: an overview and application in radiology. **Insights into Imaging**, v. 9, n. 4, p. 611–629, 22 ago. 2018.

YAN, L. et al. Practical and ethical challenges of large language models in education: A systematic scoping review. **British Journal of Educational Technology**, v. 55, n. 1, p. 90–112, 6 jan. 2024.

YANG, C.-H. et al. Deep learning-based air pollution analysis on carbon monoxide in Taiwan. **Ecological Informatics**, v. 80, p. 102477, maio 2024.

YANG, Y.; LIU, Z. A novel probabilistic approach for driving the sustainable energy circular economy: Innovating efficiency in renewable resource markets. **Sustainable Cities**

and Society, v. 108, p. 105461, ago. 2024.

YAQUB, M. et al. Modeling of a full-scale sewage treatment plant to predict the nutrient removal efficiency using a long short-term memory (LSTM) neural network. **Journal of Water Process Engineering**, v. 37, p. 101388, out. 2020.

YI, X.; WALIA, E.; BABYN, P. Generative adversarial network in medical imaging: A review. **Medical Image Analysis**, v. 58, p. 101552, dez. 2019.

YOUSEFI, H.; ARDEHALI, A.; GHODUSINEJAD, M. H. BRICS or G7? Current and future assessment of energy and environment performance using multi-criteria and time series analyzes. **Energy Strategy Reviews**, v. 49, p. 101164, set. 2023.

YU, Z. et al. Spatiotemporal dynamics and key drivers of carbon emissions in regional construction sectors: Insights from a Random Forest Model. **Cleaner Environmental Systems**, v. 16, p. 100257, mar. 2025.

ZAVRAK, S.; ISKEFIYELI, M. Flow-based intrusion detection on software-defined networks: a multivariate time series anomaly detection approach. **Neural Computing and Applications**, v. 35, n. 16, p. 12175–12193, 5 jun. 2023.

ZAWISH, M. et al. Energy-Aware AI-Driven Framework for Edge-Computing-Based IoT Applications. **IEEE Internet of Things Journal**, v. 10, n. 6, p. 5013–5023, 2023.

ZHAN, C. et al. A hybrid machine learning framework for forecasting house price. **Expert Systems with Applications**, v. 233, p. 120981, dez. 2023.

ZHAN, H. et al. DenseNet-CTC: An end-to-end RNN-free architecture for context-free string recognition. **Computer Vision and Image Understanding**, v. 204, p. 103168, mar. 2021.

ZHANG, H.; WANG, Z.; LIU, D. A Comprehensive Review of Stability Analysis of Continuous-Time Recurrent Neural Networks. **IEEE Transactions on Neural Networks and Learning Systems**, v. 25, n. 7, p. 1229–1262, jul. 2014.

ZHANG, L.; GE, R.; CHAI, J. Prediction of China's Energy Consumption Based on Robust Principal Component Analysis and PSO-LSSVM Optimized by the Tabu Search Algorithm. **Energies**, v. 12, n. 1, p. 196, 8 jan. 2019.

ZHAO, C. et al. ARBiS: A Hardware-Efficient SRAM CIM CNN Accelerator with Cyclic-Shift Weight Duplication and Parasitic-Capacitance Charge Sharing for AI Edge Application. **IEEE Transactions on Circuits and Systems I: Regular Papers**, v. 70, n. 1, p. 364–377, 2023a.

ZHAO, Z.-Q. et al. Object Detection With Deep Learning: A Review. **IEEE Transactions on Neural Networks and Learning Systems**, v. 30, n. 11, p. 3212–3232, nov.

2019.

ZHAO, Z. et al. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. **Engineering Applications of Artificial Intelligence**, v. 121, p. 105982, maio 2023b.

ZHEN, Y. et al. Temporal convolution network based on attention mechanism for well production prediction. **Journal of Petroleum Science and Engineering**, v. 218, p. 111043, nov. 2022.

ZHENG, G. et al. A multifactor hybrid model for carbon price interval prediction based on decomposition-integration framework. **Journal of Environmental Management**, v. 363, p. 121273, jul. 2024.

ZHOU, Y.; LUND, P. D. Peer-to-peer energy sharing and trading of renewable energy in smart communities ─ trading pricing models, decision-making and agent-based collaboration. **Renewable Energy**, v. 207, p. 177–193, maio 2023.

ZHOU, Z. et al. Government adoption of generative artificial intelligence and ambidextrous innovation. **International Review of Economics & Finance**, v. 98, p. 103953, mar. 2025.

ZHU, S.; OTA, K.; DONG, M. Green AI for IIoT: Energy Efficient Intelligent Edge Computing for Industrial Internet of Things. **IEEE Transactions on Green Communications and Networking**, v. 6, n. 1, p. 79 – 88, 2022.

ZÜGE, C. V.; COELHO, L. DOS S. Granular Weighted Fuzzy Approach Applied to Short-Term Load Demand Forecasting. **Technologies**, v. 12, n. 10, p. 182, 1 out. 2024.

ZUO, Y. et al. Performance and Cost Assessment of Machine Learning Interatomic Potentials. **The Journal of Physical Chemistry A**, v. 124, n. 4, p. 731–745, 30 jan. 2020.

Appendix A – Results of Robustness Testing via CBB + MCS

Data	GenEneCast High	Prophet	GenEneCast Medium	GenEneCast Low	SARIMA	STSM	Holt Winters	Simple RNN	GRU	LSTM	Synthetic Real Data
Forecast 1	491.93	377.24	509.05	506.65	506.69	476.43	463.95	489.46	401.97	379.45	541
Forecast 2	530.04	418.90	531.37	526.32	538.45	520.61	515.99	390.48	417.75	382.57	569
Forecast 3	688.81	533.69	681.93	671.08	677.19	642.21	670.59	426.06	465.49	432.39	832
Forecast 4	660.44	562.23	656.30	636.94	712.21	673.23	704.35	566.44	542.74	511.20	773
Forecast 5	850.93	742.08	865.90	876.64	881.82	865.83	909.57	563.50	653.64	617.31	970
Forecast 6	768.94	762.10	826.50	795.20	891.77	887.72	931.16	734.69	792.55	746.03	842
Forecast 7	705.39	586.47	714.37	719.67	706.72	701.60	727.75	795.80	854.23	791.32	727
Forecast 8	669.82	523.36	689.10	678.98	648.33	633.18	653.03	649.00	693.67	711.65	740
Forecast 9	498.07	377.88	509.34	498.94	490.99	477.60	476.76	552.12	554.46	605.11	537
Forecast 10	480.21	369.22	472.62	474.21	481.06	468.76	468.58	432.89	463.33	508.90	517
Forecast 11	408.84	334.00	388.48	398.64	440.56	432.31	416.20	418.86	429.03	443.48	373
Forecast 12	500.27	421.96	469.53	485.50	523.14	528.51	540.62	364.54	420.36	399.08	458
Forecast 13	441.46	369.76	387.36	417.56	480.03	484.63	475.94	462.40	423.54	379.38	430
Forecast 14	514.68	411.34	475.73	509.36	528.06	528.81	529.30	417.26	446.31	387.58	374
Forecast 15	686.29	524.65	657.82	689.37	646.95	650.41	687.85	420.43	497.69	428.32	521
Forecast 16	736.21	553.59	715.79	739.76	675.46	681.42	722.43	565.62	582.22	501.88	654
Forecast 17	928.32	729.95	925.81	937.44	874.99	874.02	932.88	589.11	699.83	602.41	660
Forecast 18	938.39	747.90	952.90	953.63	897.96	895.91	954.97	668.42	827.21	707.56	705
Forecast 19	653.60	577.70	708.41	661.15	712.70	709.80	746.32	743.10	832.09	764.45	471
Forecast 20	519.91	513.84	601.17	548.21	642.51	641.38	669.66	622.69	668.40	730.41	419
Forecast 21	389.43	369.51	432.58	404.83	486.59	485.79	488.88	521.25	542.32	640.98	287
Forecast 22	444.74	360.79	454.99	452.75	477.77	476.96	480.46	416.41	464.92	547.15	279
Forecast 23	417.31	326.27	422.06	434.66	440.74	440.51	426.73	399.95	437.76	471.68	295
Forecast 24	546.80	415.17	551.04	561.41	540.46	536.71	554.27	352.12	429.97	419.35	294
RMSE	130.50697	133.71929	136.35964	136.57937	137.9576	141.97904	154.66485	168.42138	172.82657	194.27114	-
MCS (95%)	True	True	True	True	True	True	True	True	True	True	-
Rank RMSE	1	2	3	4	5	6	7	8	9	10	-