Article:

Cali, Umit, Halden, Ugur, Andoni, Merlinda et al. (2026) Generative AI and LLM applications in renewable energy and smart grids: a systematic review for the sustainable energy transition. Artificial Intelligence Review. 153. ISSN: 0269-2821

https://doi.org/10.1007/s10462-026-11545-2

Artificial Intelligence Review (2026) 59:153
https://doi.org/10.1007/s10462-026-11545-2

Generative AI and LLM applications in renewable energy and smart grids: a systematic review for the sustainable energy transition

Umit Cali[1,10,11] · **Ugur Halden**[4] · **Merlinda Andoni**[3] · **Ferhat Ozgur Catak**[5] · **Si Chen**[3] · **Benoit Couraud**[3] · **Emre Kantar**[6] · **Samuel Knapper**[1] · **Ibrahim Kucukdemiral**[7] · **Huseyin Kusetogullari**[8] · **Murat Kuzlu**[9] · **Yashar Mousavi**[7] · **Sonam Norbu**[3] · **Taha Selim Ustun**[2] · **David Flynn**[3]

Received: 5 October 2025 / Accepted: 20 March 2026

Abstract

The global energy transition toward decarbonization and digitalization requires advanced methods to manage decentralized, data-intensive cyber-physical energy systems. This systematic review analyzes 106 research studies on Generative Artificial Intelligence (GenAI) and Large Language Models (LLMs) in renewable energy and smart grids, organized into seven application clusters covering forecasting, system design, operation, reliability, data and cybersecurity, and energy markets. The review situates these applications within a Cyber-Physical-Social Systems (CPSS) framework. Results show that GANs dominate current applications (47.2%), followed by LLMs (10.4%) and VAEs (9.4%), with growing adoption of diffusion and score-based models (7.5% each). Selected studies report improved probabilistic forecasting and uncertainty calibration using diffusion and score-based approaches, subject to dataset and evaluation setup. GenAI supports system planning through synthetic scenario generation, enhances operational decision support and demand response coordination, and contributes to reliability, cybersecurity, and market analysis. LLMs primarily function as language-driven decision support and knowledge integration components across multiple application domains. Despite computational and data-related constraints, GenAI represents an important enabler of the sustainable digital transition by supporting resilience, adaptability, and governance in renewable energy systems.

Keywords Cyber-physical-social systems · Energy digitalization · Energy transition · Generative artificial intelligence · Large language models · Sustainable energy systems · Smart grids · Renewable energy

---

Extended author information available on the last page of the article

1 Introduction

Over the last decade, the global energy sector has undergone a significant transformation driven by the duality of decarbonization and digitalization. In response to climate change, energy security challenges, and sustainability goals, countries across the world have accelerated their transition toward renewable energy-dominated systems. For example, Japan has committed to carbon neutrality by 2050 Ozawa et al. (2022), while the UK's legally binding "Net Zero" strategy outlines an evolving roadmap to eliminate net greenhouse gas emissions by the same year (Stewart and Burnett 2025). Meanwhile, Australia, despite its vast fossil fuel reserves, targets 82% renewable electricity generation by 2030 (Nelson et al. 2025), and China plans to peak emissions within the 2020 s and reach carbon neutrality by 2060, as outlined in its 14th Five-Year Plan (Zhan et al. 2024). These efforts reflect a growing global recognition that the shift to renewables is increasingly prioritized and supported by policy changes. Li and Han (2023).

Parallel to this green transition is a deepening wave of digitalization. Historically, power systems operated with limited sensing, control, and communication capabilities (Kalkal and Garg 2017) due to main grid operations being largely centralized, passive, and deterministic. However, the rise of Distributed Energy Resources (DERs), such as rooftop solar or wind energy, and low carbon technologies as electric vehicles EVs and battery systems, have introduced increased variability and complexity (Halden et al. 2021). Modern grids now demand enhanced observability, real-time data collection, intelligent control and optimization (Cali et al. 2021), and traditional tools face challanges for managing these increasingly decentralized, cyber-physical systems with high-frequency dynamics and inherited uncertainty due to the volatility of Renewable Energy Sources (RES).

In this context, energy systems are evolving into multi-layered ecosystems, involving not only physical infrastructure and cyber networks but also social, institutional, and human actors. To capture this integrated complexity, a Cyber-Physical-Social Systems (CPSS) framework is adopted (Cali et al. 2021), offering a holistic lens for studying interactions between hardware, digital platforms, and societal stakeholders.

However, the exponential growth in data volumes and the proliferation of controllable assets have made decision-making and optimization in these systems increasingly challenging (Mazhar et al. 2025). Consequently, Artificial Intelligence (AI), Machine Learning (ML) based methods have become more critical for managing this complexity. Traditional AI/ML approaches have improved forecasting, anomaly detection, and control across multiple domains (Mazhar et al. 2023; Balamurugan et al. 2025; Biswas et al. 2025). Yet, many of these techniques focus on narrow tasks such as short-term demand prediction and rely on point estimates, making them unsuitable for rare, or extreme scenarios in complex grid operations. Additionally, classical optimization methods used in conjunction with standard ML models such as population based algorithms can be computationally intensive and difficult to scale to large systems.

Recent developments in Generative AI (GenAI) provide new capabilities beyond those of predictive models. GenAI techniques such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Denoising Diffusion Probabilistic Models (DDPMs), Latent Diffusion Models (LDMs), Large Language Models (LLMs) and Vision Language Models (VLMs) are capable of producing synthetic scenarios, completing missing data, and simulating alternative outcomes (Sarmas et al. 2025). These capabilities have significant potential in scenario generation, predictive maintenance, energy policy modeling, and enhancing the resilience of decentralized grids.

To further clarify, traditional deep learning approaches such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks excel at tasks such as classification, regression, and time-series forecasting. These traditional AI methods predict specific outcomes based on historical patterns, typically producing point estimates or confidence intervals if desired by the user. However, GenAI models such as GANs, VAEs, and diffusion models are designed to learn the underlying data distribution and synthesize new, statistically plausible instances. For example, while an LSTM-based model can forecast a single renewable energy generation trajectory, a GAN or diffusion model can generate multiple diverse scenarios capturing a range of possible futures. LLMs extend this further by enabling natural language driven scenario exploration, policy analysis, and knowledge synthesis across unstructured text data. This fundamental distinction, which can be characterized as prediction versus generation, positions GenAI as complementary to, rather than a replacement for, existing deep learning methods in renewable energy applications.

Although many review studies have addressed AI/ML applications in the energy sector (Sarmas et al. 2025; Chen et al. 2025; Mongaillard et al. 2025; Biswas et al. 2025; Roman et al. 2021; Antonopoulos et al. 2020), few focus specifically on generative models, mainly because the application of GenAI remains in its early stages. As a result, power system engineers and energy experts often face the challenge of identifying which GenAI technologies are most relevant to their specific applications. In response to this challenge, this work provides a critical review of GenAI models and techniques in the smart grid and renewable energy domains, and combines a detailed analysis of the academic literature and up-to-date publications with a high-level conceptual mapping of these works using a CPSS framework for contextualizing model capabilities and limitations.

The main contributions of this work are:

- First, the research work provides a comprehensive synthesis of GenAI models by reviewing foundational techniques, learning mechanisms and theoretical properties of models relevant to the energy system domain.
- Second, it develops a structured taxonomy of GenAI models used by mapping model classes to use cases and applications across seven functional clusters in the smart grid and renewable energy domains, including forecasting and scenario generation, control and optimization, and markets and trading, among others. This aims to increase understanding of how GenAI models are currently used across different applications.
- Finally, the work introduces a high-level, energy-focused CPSS framework to map GenAI models within renewable energy systems and smart grids. The framework links physical grid assets, cyber and data infrastructures, operational control layers, and energy-specific market and regulatory contexts relevant to low-carbon power systems.

To achieve these objectives, the study adopts a two-stage literature review process, shown in Fig. 1. The first stage entails a comprehensive theoretical review of GenAI models, and is presented in Sect. 2. The second stage performs a systematic review of GenAI applications, as found in the academic literature, and is presented in Sect. 3. For both stages, the search strategy employed model-specific keywords, such as generic GenAI model terms

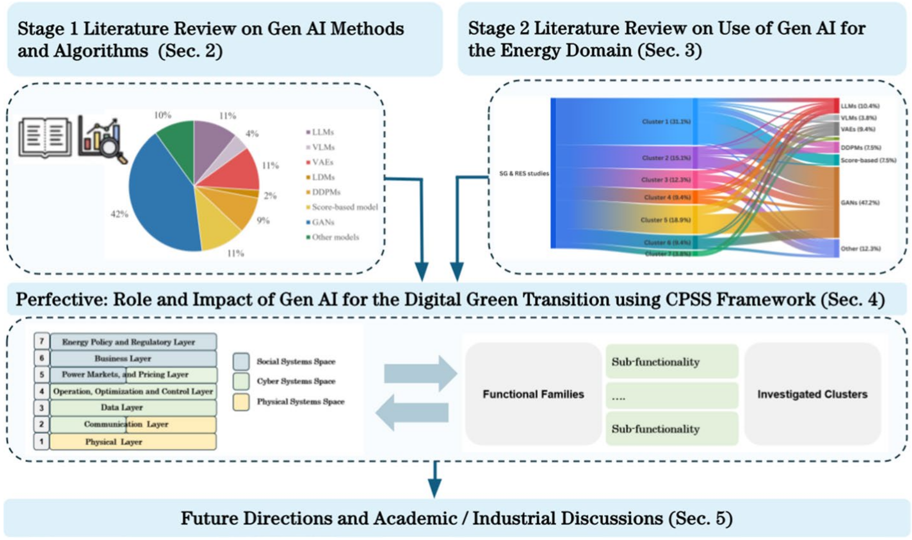

Fig. 1 Overview of the end-to-end scientific method of this study

(Sect. 2), and a combination of these terms with field-specific keywords from the renewable energy and smart grid domains. The findings of the two-staged literature review are fed into the CPSS framework, which provides a comprehensive lens for analyzing GenAI applications in energy systems. The review critically identifies promising areas, research gaps and limitations, and future research directions required for GenAI models to realize their full potential. Overall, the study aims to offer a balanced, cross-disciplinary perspective for researchers, policymakers, and industry practitioners navigating the digital-green transition. However, certain engineering limitations need acknowledgment. First, the review's scope captures published applications to date but cannot fully address emerging challenges in model deployment, such as computational requirements in resource constrained environments such as EdgeAI grid applications or real time latency requirements for critical grid operations, as they may change between various areas of applications. Similarly, as GenAI technologies evolve rapidly, some architectural advances or application domains may emerge beyond the temporal scope of this review. Moreover, the GenAI models found in the literature are deployed across different datasets, time horizons, temporal resolutions, and have fundamentally heterogeneous modeling assumptions. As a result, only limited comparative insights can be inferred, and direct comparison across models should be treated with caution. These limitations are inherent to any systematic review of a fast developing field and should be considered when interpreting the findings. Lastly, while national and regional decarbonization policies accelerate the deployment of renewable and digitally managed power systems, they implicitly impose engineering requirements for uncertainty-aware planning, stress testing, and adaptive operation that are not addressed by existing AI reviews, which largely focus on predictive models rather than generative AI and LLM-based scenario and decision support methods (Sarmas et al. 2025; Chen et al. 2025; Mongaillard et al. 2025; Biswas et al. 2025).

The remainder of this work is structured as follows: Sect. 2 presents the general overview of GenAI methods and algorithms, Sect. 3 the structured mapping of GenAI applications in smart grids and renewable energy, Sect. 4 presents the CPSS framework, Sect. 5 discusses limitations, open challenges and future research opportunities for GenAI models, whereas Sect. 6 concludes this work by providing key insights and recommendations for future research directions.

2 Overview of generative AI methods and algorithms

This section provides the technical background on key GenAI architectures that are transforming renewable energy and smart grid applications and that have been used in the literature analyzed in our study. We identified relevant studies by performing a literature search with model-specific keywords shown in Stage 1 Fig. 2. Overall, eight categories of generative models have been used, from established approaches such as GANs and Variational Autoencoders (VAEs) to recent advances in diffusion models and LLMs. Each subsection details the fundamental principles, architectural characteristics, training methodologies, and specific variants of these models. Understanding these technical foundations is essential for comprehending their applications in energy forecasting, system optimization, synthetic data generation, and intelligent grid management discussed in subsequent sections. Fig. 3 shows the distribution of papers included in this paper according to the generative AI model used. The diversity of these approaches reflects the rapid evolution of generative AI and its increasing relevance to addressing complex challenges in modern energy systems. Recent developments in Generative AI (GenAI) provide new capabilities beyond those of predictive models. GenAI techniques such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Denoising Diffusion Probabilistic Models (DDPMs), Latent Diffusion Models (LDMs), Large Language Models (LLMs) and Vision Language Models (VLMs) are capable of producing synthetic scenarios, completing missing data, and simulating alternative outcomes (Sarmas et al. 2025). Hereafter, stochastic differential equations (SDEs), noise-conditional score networks (NCSNs), conditional variational autoencoders (CVAEs), vector-quantized variational autoencoders (VQ-VAEs), and conditional generative adversarial networks (CGANs) are used as standard abbreviations. Table 1 illustrates the key characteristics and main applications of major GenAI models.

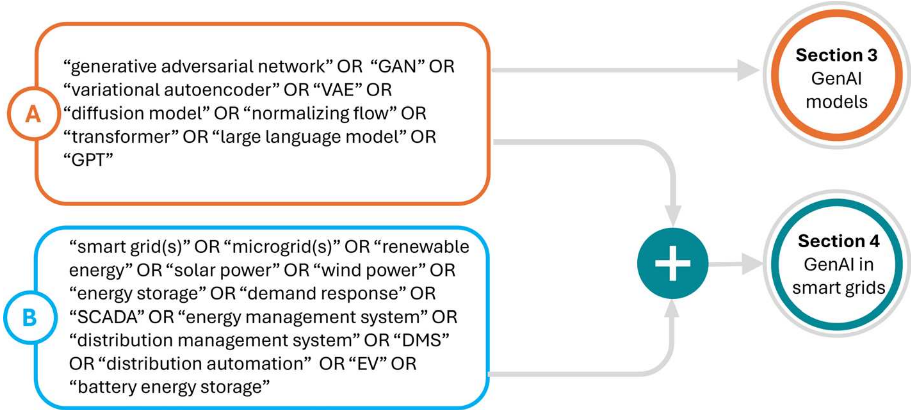

Fig. 2 Literature search workflow for the two literature review streams presented in this work

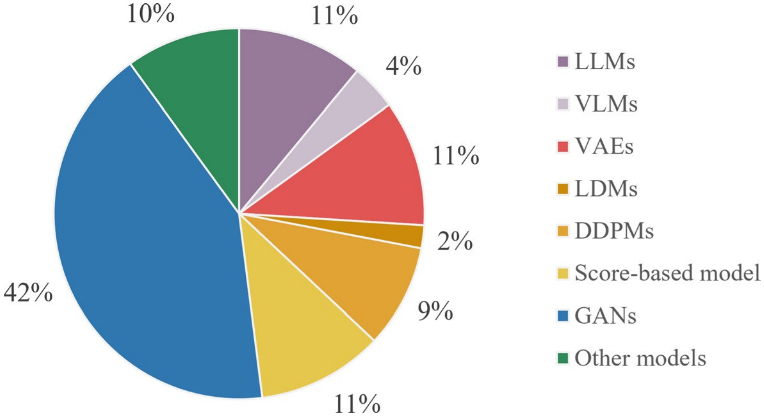

Fig. 3 Mapping and distribution of research works included in our analysis according to the generative AI model deployed

GenAI was first proposed in Goodfellow et al. (2020), which is a subset of machine learning (ML) that creates new data based on patterns learned from existing datasets (Zhang et al. 2025). Unlike traditional ML models that primarily focus on detection, classification, prediction, and decision-making, generative models are designed to create new instances that resemble the original data. This new data generation makes them potentially valuable across a wide range of applications, including image and video generation, speech synthesis, text creation, or the development of digital simulations (Sengar et al. 2024; Tedre and Vartiainen 2023). For instance, GenAI can create synthetic data by analyzing real-world data distributions and producing statistically similar examples. This feature provides benefits in fields where acquiring high-quality labeled data is challenging, time-consuming, or expensive.

In the field of renewable energy and grid systems, traditional ML models have been developed and used for resolving various problems, such as forecasting (Liao et al. 2024), anomaly detection (Burgos et al. 2024; Panthi 2020), and Optimal Power Flow (OPF) solution prediction (Khaloie et al. 2025; Li et al. 2024). In renewable energy, for instance, generative models help optimize energy storage solutions, and generate predictive maintenance insights. In smart grids, they can simulate power grid behaviors under different conditions, or enable the forecasting of energy demand and generation, helping to balance loads efficiently and integrate renewable sources more effectively. A detailed analysis of GenAI models applied in renewable generation applications and smart grids is presented in great detail in Sect. 3.

2.1 Large language models (LLMs)

LLMs have emerged as highly capable AI tools, demonstrating advanced capabilities across a wide range of domains (Fan et al. 2024). For instance, OpenAI's GPT-4/5, Anthropic's Claude, and Google's PaLM 2 exemplify these capabilities, shaping progress in Natural Language Processing (NLP) and beyond (Fan et al. 2024). At their core, these models are built on an autoregressive design and employ advanced training strategies in combination with Transformer-based architectures. The Transformer's attention mechanism enables

Table 1 Overview of generative AI models covered in this section

Model type	Key characteristics	Main applications	Notable variants
Large language models (LLMs) (Fan et al. 2024)	Transformer-based, autoregressive, attention mechanism	Text generation, summarization, classification, code generation	GPT-4/5, Claude, PaLM2
Visual language models (VLMs) (Zhang et al. 2024; Li et al. 2025)	Multimodal processing, image-text integration, contrastive learning	Image captioning, visual question answering, cross-modal retrieval	CLIP, ALIGN, BLIP, Flamingo, GIT
Variational autoencoders (VAEs) (Harrou et al. 2024)	Probabilistic encoding, latent space sampling, reconstruction	Data generation, anomaly detection, representation learning	CVAE, Denoising VAE, Hierarchical VAE, VQ-VAE
Latent diffusion models (LDMs) (Rombach et al. 2022)	Latent space diffusion, computational efficiency, high resolution	Image synthesis, conditional generation, inpainting	Stable Diffusion, Text-to-image models
Denoising diffusion probabilistic models (DDPMs) (Ho et al. 2020)	Iterative denoising, forward-reverse processes, stable training	High-quality generation, probabilistic modeling	Conditional DDPMs, Latent Diffusion Models
Score-based models (Song et al. 2021)	Score function learning, continuous-time processes, SDE framework	Uncertainty quantification, smooth generation, weather modeling	NCSN, Score SDE
Generative adversarial networks (GANs) (Goodfellow et al. 2020)	Adversarial training, generator-discriminator competition	Image generation, data augmentation, style transfer	CGAN, DCGAN, CycleGAN, GP-GAN, SRGAN
Other GenAI models (Durova et al. 2023; Wang and Nguyen 2025; Hayashi 2025)	Diverse architectures and approaches	Specialized applications	Autoregressive models, Transformer-based generative models, Hybrid models

these models to assess the relative importance of words within a passage, thereby capturing overall context and producing highly coherent sequences.

The training of LLMs relies on enormous text datasets and the adjustment of billions of internal parameters (Fan et al. 2024). Through this process, the models learn to minimize the discrepancy between predicted tokens and actual language use. One of the most advantageous uses of LLMs is their versatility, meaning they can perform a wide variety of tasks without the need for additional task-specific training. Extensive pre-training enables them to generate accurate, context-aware responses directly from user input. Another valuable application is question answering and information retrieval, where LLMs can extract relevant information from large corpora and generate contextually appropriate responses, making them particularly useful for research support and knowledge management.

Despite their capabilities, LLMs have certain limitations. Since they generate text sequentially, their processing cannot be easily parallelized. Additionally, they operate in a causal manner, relying only on past information to make predictions without access to future context. While the Transformer architecture improves handling of long-range dependencies compared to earlier models such as Recurrent Neural Networks (RNNs), challenges remain in retaining information over very long sequences due to context window constraints (Hussan et al. 2026; Sonata and Heryadi 2024).

In the energy sector, LLMs are finding diverse applications across multiple domains (Zhang et al. 2026). In energy forecasting and scenario analysis (where a scenario consists of different plausible futures underpinned by specific assumptions on demand, production, weather, or price forecasts), LLMs may assist with scenario-based planning, synthetic data generation for training forecasting models, and weather data interpretation (Mirshekali et al. 2025). For energy systems design and network planning, LLMs can be utilized to support decision-making by synthesizing technical documentation and design specifications (Hussan et al. 2026; Mirshekali et al. 2025). Within operation, control and automation, they can enable Natural Language Processing (NLP) interfaces for control systems and automated generation of operational procedures (Mirshekali et al. 2025; Zhang and Chen 2025). LLMs can also contribute to reliability, resilience and asset management through intelligent maintenance documentation, failure analysis, and predictive maintenance reporting. Within sensing, data, digitalization and cybersecurity domains, they can enhance cybersecurity threat intelligence, support smart energy IoT data interpretation, and assist in energy data disaggregation tasks (Zhang and Chen 2025). Meanwhile, for energy markets, trading and economics, LLMs can be used for analyzing market reports and regulatory documents to support trading strategies and economic assessments of various energy investments (Hussan et al. 2026; Zhang and Chen 2025). The mentioned use-cases for the applications of LLMs within the energy field are further analyzed in detail at Sect. 3

In practical power system operations, LLMs are increasingly positioned as decision-support and orchestration layers rather than direct controllers (Zhang et al. 2026). In Energy Management Systems (EMS) and Distribution Management Systems (DMS), LLMs can ingest heterogeneous inputs such as SCADA alarms, operator logs, maintenance reports, grid codes, and operational constraints, and synthesize this information into ranked remedial action suggestions during abnormal operating conditions or contingency events (Li et al. 2024). For example, LLM-based assistants can translate natural-language operator queries (e.g., “identify feeders with recurring voltage violations under high PV penetration”) into structured database queries or optimization requests executed by existing EMS tools, thereby improving operator situational awareness and reducing cognitive load.

In SCADA environments, LLMs have been explored for alarm rationalization, root-cause analysis, and automated generation of incident summaries, operating asynchronously and outside real-time protection and control loops to preserve deterministic safety guarantees. Recent studies also demonstrate the use of LLMs for operational knowledge management, where historical outage reports, switching procedures, and regulatory documents are jointly analyzed to support faster decision-making during grid disturbances. These use cases illustrate that LLMs complement, rather than replace, traditional control and optimization modules by enhancing interpretability, human–machine interaction, and decision support in complex cyber-physical energy systems.

2.2 Variational autoencoders (VAEs)

VAEs are a type of generative models used to generate new data, i.e., images, text, or audio, similar to the training data (Doersch 2016). VAEs are extended version of traditional autoencoders and have a similar architecture to them, consisting of an encoder and a decoder. In general, the encoder learns to extract meaningful latent variables from the input data, while the decoder uses these latent variables to reconstruct the input. However, unlike traditional autoencoders where the encoder outputs a fixed latent vector making it deterministic, the encoder in a VAE outputs two vectors: a mean vector representing the average or center of the data and a standard deviation vector indicating how much the values can vary. Those two vectors define a probability distribution in the latent space, making the process probabilistic.

To enable the random sampling process trainable with gradient-based methods, VAEs use a method called the reparameterization trick. This method introduces a parameter representing a random value selected from the normal distribution between zero and one. The reparameterization process allows the latent variable to be computed by combining the mean with the product of the standard deviation and the random parameter. This enables VAEs to select different points from the learned distribution to generate slightly varied but realistic outputs, as well as to produce diverse data samples while learning effectively during training.

The probabilistic nature of VAEs distinguishes them from traditional autoencoders by introducing controlled randomness in the encoding process. This randomness allows the model to generate new samples that are similar to but not identical to the training data. The encoder learns to map input data to parameters of probability distributions rather than fixed points, while the decoder learns to reconstruct realistic data from samples drawn from these distributions. This approach enables VAEs to capture the underlying variability and uncertainty present in the training data.

The training process of VAEs involves optimizing two competing objectives simultaneously. The first objective focuses on reconstruction accuracy, ensuring that the decoder can accurately recreate the original input from the latent representation. The second objective, known as the regularization term, encourages the learned latent distributions to conform to a prior distribution, typically a standard normal distribution. This dual optimization ensures that the latent space has desirable properties for generating new samples while maintaining reconstruction fidelity.

In the literature, there are several types of VAEs focusing on specific objectives. Conditional VAE (CVAE) enables controlled data generation by incorporating additional conditioning information (Pagnoni et al. 2018). Denoising VAE focuses on learning robust representations by training the model to reconstruct clean data from corrupted inputs (Creswell and Bharath 2018). Hierarchical VAE is designed for complex and structured data by employing multiple levels of latent variables (Vahdat and Kautz 2020). Vector Quantized VAE (VQ-VAE) uses discrete latent representations and has shown particular effectiveness in audio and image generation tasks (Wang et al. 2019).

The generative capabilities of VAEs make them particularly valuable in energy system applications where synthetic data generation is needed. Their ability to capture probabilistic relationships in data makes them well-suited for modeling the inherent uncertainties present in renewable energy systems, such as weather-dependent generation patterns and variable demand profiles. The controlled randomness introduced by the probabilistic framework allows for the generation of realistic scenarios (i.e. possible situation described by specific time series, such as demand, PV, wind production, weather, or price forecasts) that can be used for system testing, planning, and optimization under uncertainty.

2.3 Generative adversarial networks (GANs)

GANs are a type of generative model, which were introduced by Goodfellow et al. (2020) as one of the most exciting developments in AI world at the time. GANs are widely used to create new, realistic data, such as images, text, and music, by using two competing neural networks, specifically a Generator and a Discriminator. These networks are trained simultaneously with real data for different purposes, engaging in an adversarial training process that drives both networks to improve their respective capabilities.

The generator network aims to generate data samples similar to the real ones from the training dataset by taking random uniform or Gaussian noise as input (Goodfellow et al. 2020). The generator learns to transform this random noise into realistic data samples that mimic the patterns and characteristics of the training data (Wang and Nguyen 2025). Conversely, the discriminator network aims to classify data as real or fake, distinguishing between authentic data samples from the original training dataset and synthetic data samples produced by the generator (Goodfellow et al. 2020). The discriminator outputs a probability score indicating whether the input data is real or generated, with values closer to one indicating real data and values closer to zero indicating fake data.

The training process involves a competitive dynamic between these two networks (Goodfellow et al. 2020). As the generator becomes better at creating realistic samples, the discriminator must improve its ability to detect fake data. Simultaneously, as the discriminator becomes more skilled at identifying generated samples, the generator is forced to produce increasingly convincing outputs (Goodfellow et al. 2020). This adversarial relationship continues until the generator produces samples that are indistinguishable from real data, effectively reaching an equilibrium where the discriminator cannot reliably differentiate between real and generated samples (Choi et al. 2024).

The power of GANs lies in their ability to learn complex data distributions without requiring explicit modeling of the underlying probability distributions (Choi et al. 2024). The generator learns to capture the essential patterns and relationships present in the training data through the adversarial training process. This implicit learning approach allows GANs to generate high-quality samples across various data types and domains, making them particularly valuable for applications requiring realistic synthetic data generation (Rillig et al. 2023; Reddy 2024; Choi et al. 2024; Cali et al. 2024).

Training GANs presents several challenges, including mode collapse, where the generator produces limited varieties of samples, and training instability, where the two networks fail to reach a stable equilibrium. Researchers have developed various techniques to address these issues, including modified loss functions, architectural improvements, and regularization methods (Reddy 2024). Despite these challenges, GANs have demonstrated remarkable success in generating high-fidelity synthetic data across numerous applications (Rillig et al. 2023; Reddy 2024; Choi et al. 2024; Cali et al. 2024; Wang et al. 2018).

Beyond the original vanilla GAN, the literature presents multiple variants of the model designed to serve diverse objectives. Conditional GAN (CGAN) enables controlled data generation by incorporating additional conditioning information (Sarp et al. 2021). Deep Convolutional GAN (DCGAN) focuses specifically on image generation using convolutional neural network architectures (Patil et al. 2021). Cycle-consistent GAN (CycleGAN) specializes in image-to-image translation tasks (Liao et al. 2024). Gaussian–Poisson GAN (GP-GAN) is designed for image blending and image harmonization applications (Wu et al. 2019). Super-resolution GAN (SRGAN) targets medical imaging and video upscaling applications (Wang et al. 2018).

The versatility and effectiveness of GANs have made them particularly relevant for energy system applications, where synthetic data generation can address challenges related to data scarcity, privacy concerns, and scenario modeling. Their ability to generate realistic time series data, weather patterns, and operational scenarios makes them valuable tools for energy forecasting, system planning, and robust testing of energy management algorithms under diverse conditions.

2.4 Denoising diffusion probabilistic models (DDPMs)

DDPMs were first introduced by Ho et al. (2020). These models have emerged as highly effective generative models and are widely used in both computer vision and natural language processing applications (Ho et al. 2020; Alimanov and Islam 2023). DDPMs utilize a parameterized stochastic process, learned through variational inference, to reverse a noise-adding diffusion path (Ho et al. 2020). During training, data is progressively corrupted with noise in a forward process, and the model learns to reverse this trajectory to recover the original input.

DDPMs are particularly notable for their robust and stable training procedures, comprehensive modeling of data distributions, and scalability to large datasets (Ho et al. 2020; Alimanov and Islam 2023; Miraki et al. 2025). Unlike some other generative models that can suffer from training instabilities or mode collapse, DDPMs provide a more reliable training experience through their gradual denoising approach. The iterative nature of the diffusion process allows these models to capture fine-grained details and complex patterns in the data more effectively than single-step generation methods.

The fundamental principle underlying DDPMs involves two complementary processes: a forward diffusion process that gradually adds noise to clean data, and a reverse process that learns to remove this noise step by step. The forward process systematically corrupts the original data by adding Gaussian noise over multiple time-steps until the data becomes pure noise. This corruption process follows a predetermined schedule that controls how much noise is added at each step, ensuring a smooth transition from clean data to complete noise.

The reverse process represents the core learning component of DDPMs. During training, the model learns to predict and remove the noise that was added at each timestep of the forward process. This denoising capability is developed through a neural network that takes the noisy data and the current timestep as inputs and predicts the noise component that should be removed. The training framework is based on a reweighted variational lower bound, aiming to optimize the model's denoising capabilities (Choi et al. 2021).

Generation occurs by starting with a sample of pure noise and iteratively applying the learned denoising steps. At each timestep, the model removes a portion of the noise, gradually transforming the random noise into a coherent data sample that resembles the training distribution. This iterative refinement process allows DDPMs to generate high-quality samples with fine details and realistic characteristics.

The step-by-step generation process of DDPMs offers several advantages over other generative approaches. The gradual refinement allows for better control over the generation process and can produce more stable results. Additionally, the iterative nature enables the model to correct errors made in earlier steps, leading to improved final output quality. However, this multi-step approach also means that generation can be computationally more expensive compared to single-step methods.

DDPMs have demonstrated remarkable versatility across various domains and data types. Their ability to model complex distributions makes them suitable for generating high-resolution images, synthesizing realistic textures, and creating diverse samples from learned distributions. The robust training characteristics of DDPMs have made them increasingly popular for applications requiring reliable and high-quality generative capabilities.

In energy system applications, DDPMs have shown particular promise for energy forecasting and scenario generation (Capel and Dumas 2023; Miraki et al. 2025). Their ability to capture uncertainty and generate multiple plausible scenarios makes them valuable for probabilistic forecasting tasks. The gradual denoising process can effectively model the temporal dependencies and variability inherent in energy data, making DDPMs well-suited for generating realistic energy consumption patterns, renewable generation profiles, and other time-series data relevant to energy system planning and operation.

2.5 Latent diffusion models (LDMs)

Latent Diffusion Models (LDMs) are a class of generative models that combine the power of diffusion probabilistic processes with the efficiency of latent-space modelling (Chen et al. 2022). Introduced to address the scalability and resolution limitations of pixel-space diffusion models, LDMs operate by learning and denoising in a compressed latent representation space instead of directly in the high-dimensional image space (Alimanov and Islam 2023).

Operating in latent space provides significant advantages for generative modeling applications (Chen et al. 2022). LDMs reduce the memory and compute requirements by avoiding high-dimensional image space operations, making them more computationally efficient than their pixel-space counterparts (Alimanov and Islam 2023). The use of learned encoders and decoders allows for training on higher resolution images than typical pixel-space diffusion models, enabling high-resolution output generation. Additionally, the encoder-decoder pipeline enables flexible conditioning approaches, such as text-to-image generation and image inpainting, making LDMs highly versatile for various applications.

LDMs offer a promising tool for synthesizing high-quality visual data in energy-related domains (Alimanov and Islam 2023; Chen et al. 2022). They can generate plausible environmental conditions for model training when real data is scarce or incomplete, addressing common data availability challenges in renewable energy applications. These models excel at creating rare failure scenarios for grid components to augment predictive maintenance systems, helping to improve system reliability through better anomaly detection capabilities. Furthermore, LDMs can produce hypothetical weather, disaster, or load distribution maps for use in simulations and training, supporting comprehensive scenario planning and risk assessment in energy system operations.

LDMs demonstrate superior performance characteristics compared to other popular generative approaches. While GANs suffer from low training stability, LDMs offer high training stability similar to DDPMs. In terms of output quality, GANs produce high realistic results, but both DDPMs and LDMs achieve very high output quality that often surpasses GAN performance. Regarding training costs, GANs require moderate computational resources, DDPMs demand high computational investment, while LDMs achieve a favorable balance with moderate training costs, making them attractive for practical applications.

Ongoing research in LDMs is exploring several promising directions that could benefit energy applications. Hybrid architectures are being developed that integrate transformers for improved multi-modal synthesis, enabling more sophisticated content generation that combines different data types. Self-supervised pretraining approaches are being investigated to reduce reliance on large labeled datasets, which is particularly valuable in energy domains where labeled data can be expensive to obtain. Additionally, physics-informed LDMs are being developed that embed domain knowledge to improve realism and utility in simulation tasks, ensuring that generated content adheres to physical principles and constraints relevant to energy systems.

In conclusion, LDMs provide a powerful and flexible generative modeling paradigm that balances output quality with efficiency. Their adaptability makes them particularly suited for applications in energy infrastructure modeling, simulation, and synthetic data generation under the constraints of the digital-green transition.

2.6 Score-based models/score matching

Score-based generative models, also known as score matching models, represent a powerful class of generative models that have gained significant attention in recent years for their ability to generate high-quality samples while providing theoretical guarantees (Song and Ermon 2019). These models are particularly relevant for renewable energy applications due to their capacity to model complex, continuous distributions of energy-related variables such as wind speeds, solar irradiance patterns, and load profiles.

The fundamental principle underlying score-based models is the concept of score matching, which involves learning the score function (the gradient of the log probability density) of the data distribution rather than the density itself (Hyvärinen and Dayan 2005). This approach offers several advantages over traditional generative models, including avoiding the need for normalization constants and providing a more stable training procedure compared to GANs.

In more detail, score-based models work by learning to estimate the score function, which can be understood as the direction of steepest ascent in the probability landscape of the data. Rather than directly modeling the probability density function, these models learn how data points should move to increase their likelihood under the true data distribution (Vincent 2011). The training process typically employs denoising score matching, where the model learns to reverse the effect of adding noise to clean data samples. This denoising approach was pioneered by Vincent (2011) and later refined by Song and Ermon (2019) for generative modeling applications. The key insight is that by adding different levels of noise to the data and learning to predict the direction that removes this noise, the model implicitly learns the underlying data distribution.

A significant advancement came with the introduction of noise conditional score networks, where the model learns to handle multiple noise levels simultaneously (Song and Ermon 2020). This multi-scale approach allows the model to capture both fine-grained details and global structure in the data, making it particularly effective for complex datasets typical in renewable energy applications.

Overall, score-based models offer several unique advantages for renewable energy applications that distinguish them from other generative approaches. First, these models excel at capturing the smooth, continuous variations characteristic of meteorological data. Unlike discrete generative models that may produce artificial discontinuities, score-based models naturally generate smooth trajectories that better represent physical weather processes (Leinonen et al. 2020). This makes them particularly suitable for *weather pattern modeling*, such as generating realistic wind speed profiles, solar irradiance patterns, and temperature variations. In addition, the probabilistic nature of score-based models makes them valuable for *uncertainty quantification in renewable energy forecasting*. By generating multiple plausible future scenarios, these models can provide probabilistic forecasts that account for the inherent uncertainty in weather-dependent renewable generation (Chen et al. 2021). Score-based models can generate large volumes of synthetic renewable energy data that preserve the statistical properties of historical observations while exploring new regions of the data space (Rasul et al. 2021). This *synthetic data generation* is valuable for stress-testing grid operations, training other machine learning models, and conducting long-term planning studies where historical data may be insufficient. These capabilities are crucial for grid operators who need to plan for various possible outcomes. Finally, the score function learned by these models provides a natural *anomaly detection* mechanism. Unusual patterns in renewable energy systems, such as equipment malfunctions or extreme weather events, typically correspond to regions of low probability density that can be identified using the learned score function (Zhai et al. 2016). This capability enables predictive maintenance and early warning systems.

Recent developments have extended score-based models to continuous-time formulations using stochastic differential equations (SDEs) (Song and Ermon 2019). This framework treats the noise addition process as a continuous diffusion process, where data is gradually corrupted by noise over time. The reverse process, guided by the learned score function, can then generate new samples by starting from pure noise and gradually removing it. This SDE-based approach provides several theoretical advantages, including guaranteed convergence properties and the ability to trade off between sample quality and generation speed (Song et al. 2021a). SDE framework is a good foundation for modeling uncertainty, optimizing operations, as well as understanding the probabilistic behavior of smart grids.

Overall, score-based models offer several advantages over alternative generative approaches in renewable energy contexts. First, they provide stable training without the adversarial dynamics that can make GANs difficult to optimize (Song and Ermon 2019). Second, they naturally handle continuous data without requiring discretization, making them well suited for physical measurements. Third, they offer theoretical guarantees about convergence and sample quality that are often lacking in other generative models. However, these models also present certain challenges. The iterative sampling process can be computationally expensive, particularly for real-time applications (Song et al. 2021b). The choice of noise schedules and sampling parameters requires careful tuning and can significantly impact performance. Additionally, the models may require substantial computational resources during training, especially for high-dimensional or long sequence data typical in energy applications. Another important consideration for renewable energy applications is the integration of physical constraints into the generation process. Recent research has explored methods for incorporating domain knowledge into score-based models, such as ensuring non-negative power outputs or respecting physical limits on generation capacity (Liu et al. 2022). These constraints can be incorporated through specialized network architectures, constrained sampling procedures, or post-processing steps that project generated samples onto feasible regions.

2.7 Visual language models (VLMs)

VLMs represent a fascinating convergence of artificial intelligence capabilities, blending the power of visual perception with linguistic understanding to tackle complex tasks that require reasoning across both images and text. These models have emerged as a cornerstone of multi-modal AI, enabling systems to interpret and generate content that bridges the visual and textual domains. Unlike traditional AI models that focus solely on text or images, VLMs integrate these modalities, offering a versatile tool set for applications ranging from image captioning to visual question answering and beyond. In the context of renewable energy and smart grids, VLMs hold immense potential to transform how we analyze and manage energy systems, providing insights that are both visually informed and contextually rich (Sengar et al. 2024; Zhang et al. 2025).

The foundation of VLMs lies in the transformer architecture, a framework initially popularized in natural language processing for its ability to capture relationships within sequences of data. For VLMs, this architecture is adapted to process both text and images simultaneously. Images are typically broken down into smaller patches or features using pre-trained Convolutional Neural Networks (CNNs) or Vision Transformers (ViTs), which convert visual data into a sequence of embeddings. These embeddings are then processed alongside textual inputs, allowing the model to learn joint representations that capture the interplay between visual and textual information.

Training VLMs often involves a technique called contrastive learning, where the model learns to pair related images and texts while distinguishing them from unrelated pairs. Imagine teaching a model to recognize that a picture of a wind turbine belongs with a description of its power output, but not with a caption about a hydroelectric dam. This approach, used in models such as CLIP and ALIGN, allows VLMs to generalize across diverse tasks by learning from vast datasets of image-text pairs (Choi et al. 2024). By aligning visual and textual data, VLMs can perform tasks such as generating textual summaries from images or answering questions about visual scenes, which are critical for real-time decision-making in energy systems.

Several standout VLMs have pushed the boundaries of multimodal AI. For example, CLIP, introduced in 2021, leverages contrastive learning to align images and text, enabling robust performance across tasks like image classification and retrieval (Sengar et al. 2024). Similarly, ALIGN (2021) combines a Vision Transformer with a text encoder to achieve comparable results. BLIP (2022) enhances this by incorporating captioning alongside contrastive objectives, while Flamingo (2022) excels in few-shot learning, adapting quickly to new tasks with minimal data. GIT (2022), on the other hand, uses a Vision Transformer with a decoder for causal language modeling, enabling it to generate text directly from visual inputs (Zhang et al. 2025). These models showcase the diversity and adaptability of VLMs, making them powerful tools for energy applications.

In the realm of renewable energy and smart grids, VLMs offer practical solutions to real-world challenges. For instance, they can analyze satellite imagery to identify and classify renewable energy infrastructure, such as solar panels or wind turbines, with high accuracy (Wen et al. 2023). After natural disasters, VLMs can compare pre- and post-event images to assess damage to energy assets, streamlining recovery efforts. For maintenance, VLMs can process drone imagery alongside maintenance logs to detect anomalies, such as wear on turbine blades, prioritizing repairs efficiently. Additionally, VLMs can generate textual summaries of asset conditions from visual and sensor data, simplifying reporting for grid operators. These applications highlight how VLMs can enhance the digital-green transition by making energy systems smarter and more responsive.

Despite their promise, VLMs face significant challenges. Training these models requires enormous computational resources and large datasets, which can strain energy budgets and raise sustainability concerns (Reddy 2024). Biases in training data can also lead to unfair outcomes, such as misidentifying certain types of infrastructure or prioritizing certain regions over others (Mariani and Dwivedi 2024).

Looking ahead, the future of VLMs in energy systems is bright. Researchers are exploring lightweight architectures that can run on edge devices, enabling real-time decision-making in smart grids without relying on cloud computing (Weber et al. 2021). Fine-tuning VLMs for specific energy contexts, such as climate monitoring or grid automation, could further enhance their utility (Zhao et al. 2024). Additionally, improving the explainability of VLMs, i.e. making their decisions more transparent, will be critical for building trust among grid operators and regulators (Perera et al. 2024). By addressing these challenges, VLMs can play a pivotal role in creating intelligent, sustainable energy systems that align with the goals of the digital-green transition.

2.8 Other GenAI models

In this section, selected types of generative models such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs) and Denoising Diffusion Probabilistic Models (DDPMs), were briefly introduced. However, there are many other generative models that are also powerful and widely used within the broader AI landscape. These include Autoregressive Models, Transformer-Based Generative Models (Choi et al. 2023), and Hybrid Models (Larsen et al. 2016). Autoregressive models focus on predicting the likelihood of time-series events but can be computationally expensive to train, especially for long sequences. Transformer-based generative models are particularly effective for natural language processing tasks and complex sequence generation; however, similar to autoregressive models, they require significant computational resources. Fortunately, hybrid models can leverage the strengths of different generative model architectures,such as VAE-GANs, which combine the stable training of VAEs with the realistic image generation capabilities of GANs. It is evident that advancements in GenAI will continue, driven by ongoing support from both industry and academia, with a focus on improving the quality of generated samples and enhancing model stability.

In addition to well-established architectures such as GANs and VAEs, a number of advanced generative models have recently emerged. Transformer-based generative models leverage attention mechanisms to capture long-range dependencies, making them highly effective for sequential data, multimodal synthesis, and energy time-series forecasting, although their training demands remain computationally intensive (Vaswani et al. 2017). Building on this foundation, Diffusion Transformers (DiTs) combine the robustness of diffusion models with the scalability of transformers, offering state-of-the-art performance in high-resolution generation tasks and showing promise for spatiotemporal forecasting in complex energy systems (Peebles and Xie 2023; Capel and Dumas 2023). A distinct line of innovation is represented by Neural Radiance Fields (NeRFs), which model 3D scenes from 2D image collections by learning implicit volumetric representations. NeRFs are particularly relevant for creating immersive digital twins of renewable energy infrastructures, enabling photorealistic visualization and simulation of assets such as wind farms, solar installations, or urban energy systems (Mildenhall et al. 2021; Zhang et al. 2024). Collectively, these emerging models expand the generative AI landscape, extending applications from language and image synthesis toward physically grounded simulation, planning, and visualization across the renewable energy domain.

Given the diversity of Generative AI methods, it is interesting to compare them so that engineers and researchers can find the best fit for their applications and constraints. However, a quantitative comparison is difficult as it depends on the method implementation, datasets, and use case. As a result, Table 2 provides a qualitative comparison of these generative AI models in terms of training cost, training stability, and memory usage, based on trends commonly reported in the literature. According to Table 2, VAE models generally exhibit stable training and low computational and memory requirements, making them suitable for resource-constrained applications. GAN models often achieve efficient inference but can suffer from training instability and sensitivity to hyperparameter tuning. Diffusion-based and score-based models usually offer more stable optimization and strong

Table 2 Qualitative comparison of training cost, training stability, and memory usage across different GenAI models

Model	Training cost	Training stability	Memory usage
VAE (Doersch 2016; Pagnoni et al. 2018; Creswell and Bharath 2018)	Low–Medium	High	Low–Medium
GAN (Goodfellow et al. 2020; Sarp et al. 2021; Liao et al. 2024)	Medium	Low–Medium	Medium
Diffusion (DDPM/LDM) (Ho et al. 2020; Alimanov and Islam 2023; Miraki et al. 2025)	High	High	Medium–High
Score-based models (Song and Ermon 2019; Song et al. 2021b)	High	Medium–High	Medium–High
LLM (Fan et al. 2024)	Very High	High	Very High
VLM (Sengar et al. 2024; Zhang et al. 2025)	Very High	Medium–High	Very High

probabilistic modeling capabilities, at the expense of higher training and sampling costs due to their iterative generation processes. LLMs and VLMs incur substantially higher computational and memory demands, driven primarily by model size and, in inference settings, the growth of key–value caches with sequence length or multimodal context. Overall, the table highlights typical trade-offs among model families rather than absolute performance rankings, as actual computational requirements depend on model scale, data modality, and implementation details.

2.9 Scenario realism, validation, and trustworthiness considerations

The adoption of diffusion and score-based generative models in power and energy systems raises important questions regarding computational cost, scenario realism, and operational credibility. Unlike GAN-based approaches, diffusion models require iterative denoising processes during training and sampling, resulting in higher computational overhead (Ho et al. 2020; Karras et al. 2022). However, in power system applications these models are predominantly employed in offline settings, such as scenario generation for planning, stress testing, and resilience analysis, where training costs can be amortized and inference latency is not time-critical (Zhang et al. 2025). In this context, the improved training stability and superior mode coverage of diffusion-based models justify their higher computational cost, particularly for generating rare but operationally significant scenarios such as extreme weather events, correlated distributed energy resource (DER) failures, or demand spikes under high renewable penetration.

Ensuring the realism and credibility of synthetic scenarios is critical before they are used in downstream engineering analyses. Prior work emphasizes that generative models should not be treated as replacements for physical simulators or operational tools, but rather as scenario proposal mechanisms whose outputs must undergo multi-layered validation (Shukla and Deepa 2025). A common first layer involves statistical and distributional validation, where marginal distributions, temporal correlations, and cross-variable dependencies of synthetic data are compared against historical observations. This step ensures consistency with observed operating regimes while identifying distributional drift or mode collapse.

A second layer of validation relies on physics- and simulator-based assessment, in which synthetic scenarios are injected into established power system analysis tools, such as load flow solvers, contingency analysis frameworks, or dynamic stability simulators. Scenarios that violate physical constraints, operational limits, or protection requirements are filtered out, ensuring compliance with grid codes and safety margins. This simulation-in-the-loop approach preserves the role of domain knowledge and physics-based modeling in the validation process.

Beyond statistical and physical validation, recent studies highlight the importance of trustworthiness-oriented evaluation, particularly for safety-critical cyber-physical systems (Gawlikowski et al. 2023; Borji 2019). These approaches include uncertainty-aware generative modeling, out-of-distribution detection, and constraint-aware generation to prevent implausible or unsafe scenarios from influencing operational decision-making. Human-in-the-loop review and expert screening are often incorporated for high-impact use cases, reinforcing accountability and interpretability. Together, these practices establish a principled framework in which synthetic scenarios are assessed for credibility, robustness, and operational relevance, aligning generative modeling with emerging guidelines for trustworthy AI in critical infrastructure domains.

3 Generative AI in smart grids and renewable energy applications

Section 3 focuses on applications and use cases of generative AI in the smart grid and renewable energy context. Relevant literature studies were identified by conducting a systematic literature search on the Scopus and Web of Science databases, as illustrated in Fig. 4. To examine the applications of GenAI models in the energy domain, keywords reflecting areas and sub-domains related to smart grid operation and renewable generation were identified. Subsequently, we combined GenAI model-specific keywords with smart grid keywords and retrieved relevant studies. In the identification phase, filters were applied to include only peer-reviewed journal articles, conference papers, and scholarly books published between 2020 and 2025. The search included studies containing any combination of the query terms in the title, abstract, or keywords. This process yielded a total of 1,402 records. In the second phase, we reviewed the abstracts of the papers and ensured that the studies conformed to the predefined inclusion criteria. Duplicate records were also removed. This screening phase reduced the dataset to 349 studies. In the third phase, the screened studies were further refined through careful full-text assessment by the co-authors. Following this eligibility review, 106 studies were finally included in the systematic review (Sect. 3). A breakdown of studies according to the year of publication showcases a steady growth of research works deploying GenAI models in the smart grid and renewable energy domains illustrated in Fig. 5.

The papers were further organized into 7 clusters according to the smart grid and renewable energy application area as in Fig. 6, following the main application proposed by the authors of these papers: energy forecasting and scenario analysis; energy system design and

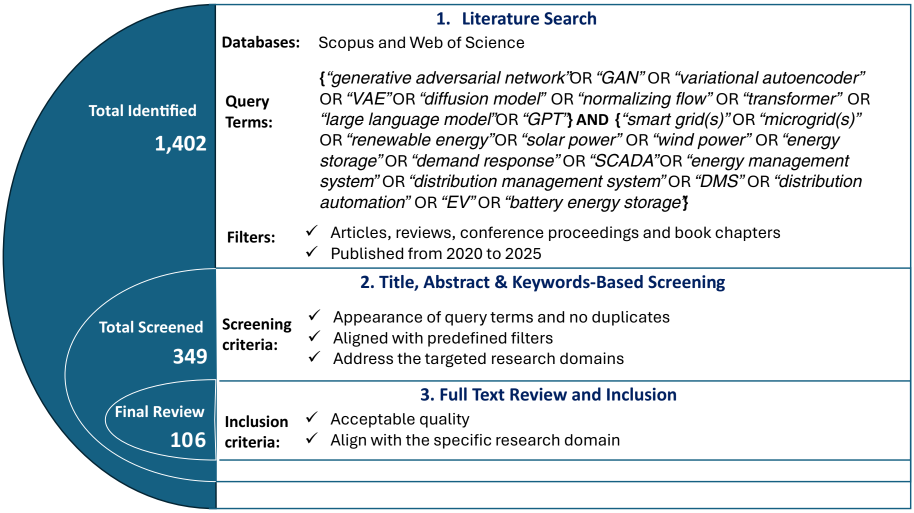

Fig. 4 Workflow showing the number of literature identified, screened, assessed for eligibility, and included in the final systematic review

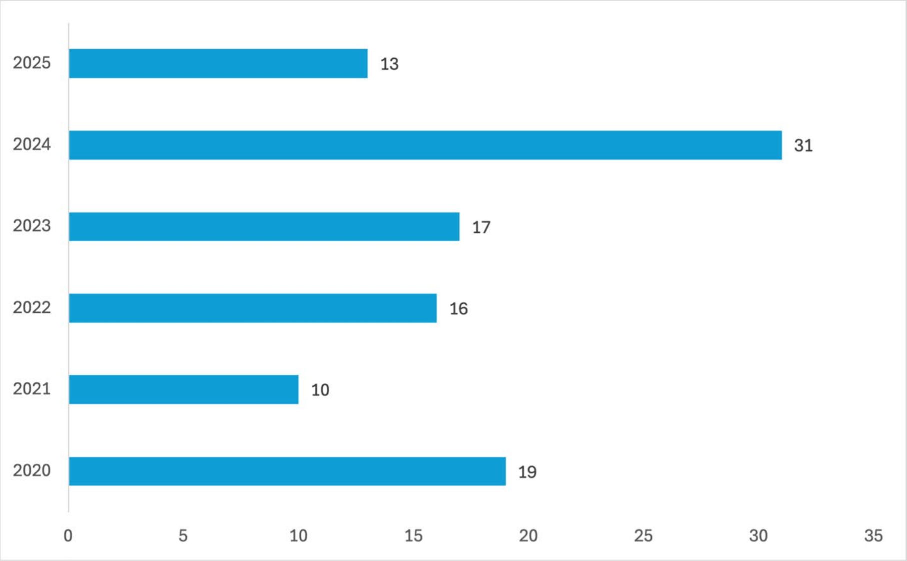

Fig. 5 Year-wise distribution of publication on GenAI models and appliaction in energy domain

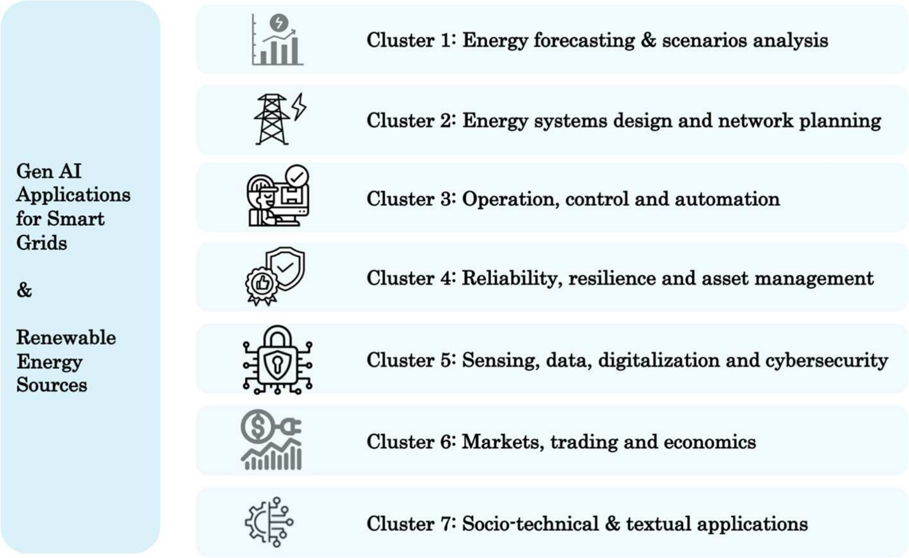

Fig. 6 Illustration of clusters for the performed literature review

network planning; operation, control, and automation; reliability, resilience, and asset management; sensing, data, digitalization, and cybersecurity; markets, trading, and economics; and socio technical and textual applications. In addition, within each cluster, we further classify the studies according to the main task where generative AI models are applied, such as data augmentation, synthetic time series generation, scenario generation, anomaly detection, feature extraction, and others.[1] Each cluster discusses key contributions, methodologies, and findings of recent works, along with comparative insights on their approaches and performance.

In the following section, we present the results of our literature review analysis. Fig. 7 illustrates how research papers included in Sect. 3 are mapped across different renewable generation and smart grid application fields represented by Clusters, and the family of generative AI models they employ. Other papers have also reviewed generative models within the energy system domain. Zhang et al. (2025) conducted a comprehensive review of deep generative models (DGM), including GANs, VAEs, NFs, diffusion-based and GPT models, highlighting the main usage of DGM, their potentials, challenges and future directions. The authors categorized research articles in various aspects, such as data generation, forecasting, situational awareness, modeling, and optimal decision making. Similar to this work, our analysis found that GANs was the most commonly used GenAI model in smart grids and renewable energy applications with applications across clusters (47.2%), followed by LLMs (10.4%) and VAEs (9.4%). Papers applying DDPMs (7.5%) and score-based models (7.5%) can also be found, especially related to Cluster 1 and 2. However, unlike other reviews of generative AI models in the energy sector, which primarily highlight the functions achieved by GenAI models (e.g., data synthesis, short-term forecasting, anomaly detection), our analysis emphasizes how these technologies are applied within energy systems, and more specifically in smart grids and renewable energy applications (e.g., future scenario analysis, system control, market participation, asset management). Moreover, our review highlights the final energy applications identified in the reviewed papers, which are categorized into the clusters shown in Fig. 6.

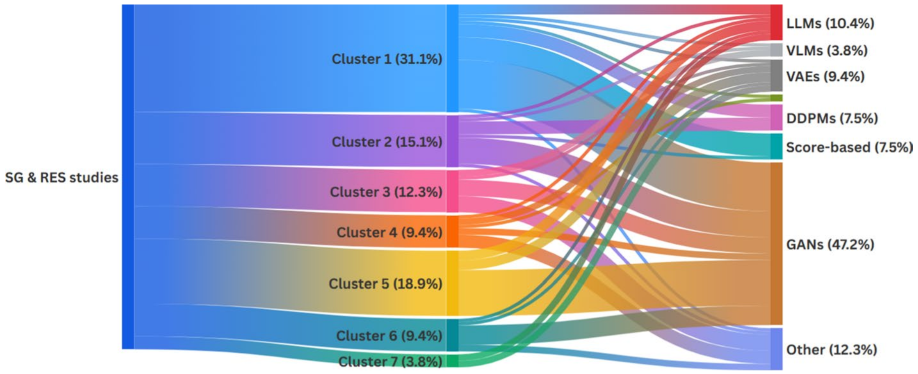

Fig. 7 Sankey diagram mapping research studies in Sect. 3 across two levels: renewable and smart grid application domains (clusters) and the generative AI models employed

---

[1] It should be noted that the classification of papers into different clusters is not always straightforward and clearly defined, due to overlaps that naturally arise among studies that might address multiple aspects of functionalities and tasks that cut across smart grid domains. In our work, we distinguish the studies according to their primary focus area as exemplified by their authors.

Table 3 Mapping of CPSS as a high-level conceptual framework extending the SGAM architecture, showing layers, functional families, sub-functionalities, clusters, and literature references

CPSS layer	Functional family	Sub-functionality	Cluster(s)	Selected References
Physical systems	System representation & simulation	Digital twins	2,4	Liao et al. (2024); Wang and Ou (2024)
		Anomaly detection & diagnostics	4	Zhang et al. (2022); Li et al. (2019)
		Grid studies & network planning	2	Li et al. (2025a, 2025b)
		Energy system design & sizing	2	Chen et al. (2022); Buster et al. (2024)
		Microgrid planning	2	Liu and Tang (2024)
		Climate-impact stress testing	4	[Future Work]^{a}
Communication	Resilience & stress testing	Cyberattack simulation	5	Omara and Kantarci (2024); Ding et al. (2022); Liu et al. (2024)
		Network traffic generation	5	Li et al. (2025); Efatinasab et al. (2024); Ahmadian et al. (2018)
		Synthetic cross-domain attack generation	5	[Future Work]^{a}
Data	Data-centric functions	Synthetic data generation	1	Zhang et al. (2023); He and Yuan (2023); Yilmaz and Korn (2022); Kumar et al. (2023)
		Scenario generation	1	Dumas et al. (2022); Bagheri et al. (2022); He and Yuan (2023); Yilmaz et al. (2024)
		Feature extraction	1	Zhang and Kong (2022); Zhang et al. (2023); Tao et al. (2023)
		Demand disaggregation	5	Çavdar and Feryad (2021); Badi et al. (2018)
		Energy forecasting	1	Huang et al. (2022); Xiong et al. (2025); Zhang et al. (2023); Tao et al. (2023)
		Data imputation & augmentation	1	Li et al. (2025); Zhang et al. (2023); He and Yuan (2023)
		Reliability & contingency analysis	4,5	Zhang and Yang (2023); Peng et al. (2021); Zhang et al. (2022); Li et al. (2019)
		Socio-energy demand generation	1	[Future Work]^{a}
Operation & control	Decision support & control	TSO/DSO network optimization	2,4	Li et al. (2025a, 2025b)
		Demand response & DSM simulation	3	Wen et al. (2023); Weber et al. (2021)
		Optimal power flow & dispatch	3	Khodayar et al. (2019); Wang et al. (2024); Zhang et al. (2024)

Table 3 (continued)

CPSS layer	Functional family	Sub-functionality	Cluster(s)	Selected References
		Grid control & management	3	Wen et al. (2023); Khodayar et al. (2019)
		Multi-energy system control	3	Weber et al. (2021); Zhang et al. (2024)
		AI-based and data driven control	3	Han et al. (2024)
		Blackout recovery pathway generation	4	[Future Work][a]
Power markets & pricing	Market & business functions	Price forecasting & bidding	6	Medina et al. (2024); Walter and Wagner (2024); Lu et al. (2022); Zhang and Wu (2021)
		Flexibility market simulation	4,6	[Future Work][a]
Business	Market & business functions	Techno-economic feasibility	6,7	Yilmaz et al. (2024); Walter and Wagner (2024)
		Investment risk assessment	6,7	Mukhamediev et al. (2020); Kim et al. (2021)
		Carbon trading & PPAs	6,7	[Future Work][a]
Policy & regulation	Policy & governance	AI-supported policy intelligence	7	Buster et al. (2024)
		Generative policy sandboxes	7	[Future Work][a]

[a] Future Work corresponds to functionalities that have not been encountered in the analyzed literature, but that are discussed as future research directions in Sect. 5

3.1 Cluster 1: Energy forecasting and scenario analysis

Accurate forecasting is a critical component of modern energy systems, enabling reliable operations, market efficiency, and effective integration of renewable resources. When forecasting or data generation is employed as a task to address specific smart grid applications (e.g., operational control, markets, or system design), the corresponding studies are listed under the most relevant focus cluster (2-7).

3.1.1 Short-term forecasting

Traditional forecasting models often struggle to capture the inherent uncertainty, complexity, non-linearity, and variability underlying modern energy systems that depend highly on renewables. Renewable generation, including solar, wind and hydro power generation, are often distributed and depend on weather patterns. High variability and sometimes data sparsity have made solar and wind forecasting a fertile ground for generative models.

On solar applications, generative AI models are used for *Photovoltaic (PV) production forecasting* at different time scales. Huang et al. (2022) developed a conditional GAN (CGAN) combined with a Bi-long short-term memory (LSTM) network for hour-ahead PV power forecasting that outperformed standard LSTM in prediction accuracy. In their architecture, the generator produces a predicted time series conditioned on recent observations, and the discriminator compares the fake forecast with the true sequence to improve

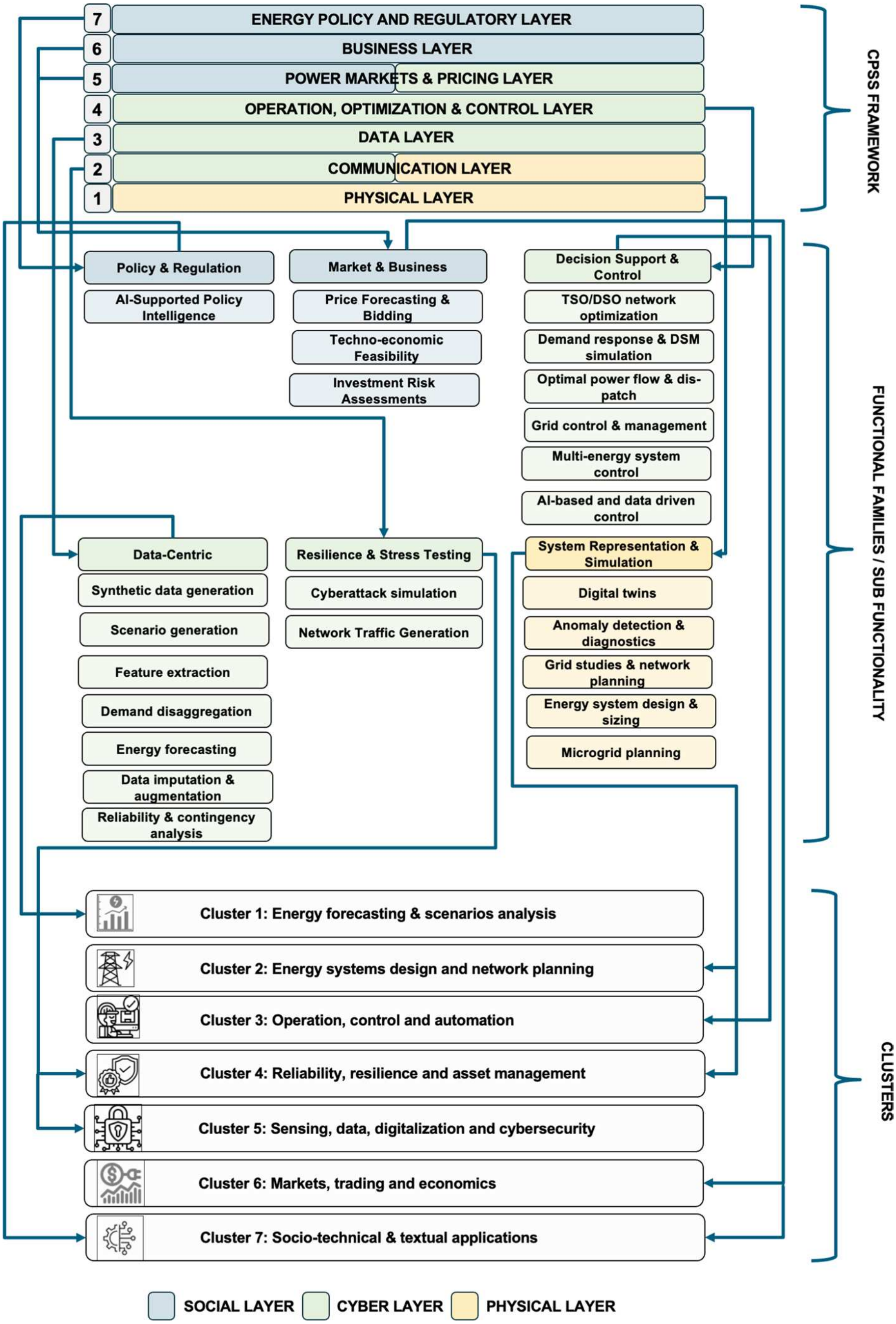

Fig. 8 Mapping of cyber physical and social system layers to clusters via GenAI use cases

realism. The combination of CGAN-BiLSTM outperformed other approaches, including standard LSTM, recurrent neural network (RNN), back-propagation neural network (BP), support vector machine (SVM), and Persistence models, regarding prediction accuracy. The comparison was achieved on a common real-world dataset and compared the values of five performance evaluation indicators, RMSE, MAE, nRMSE, R2, and R. Reduced values of error metrics were achieved by adding adversarial training to refine the temporal patterns. Li et al. (2022) took a similar approach for 1-6 h ahead PV forecasting by proposing a hybrid model integrating a conditional GAN for sequence generation with iterative forecasting. Notably, they found the GAN-based model better captured ramps and variability in PV output, thanks to its ability to learn complex data distributions. Other works focused on day-ahead PV forecasting based on diffusion-based models integrating transformer architectures Xiong et al. (2025) with the aim to enhance the accuracy and uncertainty quantification of predictions. Several research groups have demonstrated the effectiveness of score-based models in solar power forecasting. Lai et al. (2022) developed conditional score-based models that outperformed traditional time series methods by better capturing the multimodal nature of solar generation under different weather conditions. Their approach conditions the generation process on weather forecasts and time-of-day information to produce more accurate and calibrated predictions. The authors in Zhang and Kong (2022) proposed a hybrid photovoltaic output power forecasting model that combined a LSTM neural network with stochastic differential equation (SDE) method. According to the results, this hybrid approach effectively mitigates the uncertainty inherent in PV power generation and enhances decision support for the stable and safe integration of PV systems into the grid.

Generative AI models have also been used for short-term *wind forecasting*. For instance, Zhang et al. (2023) developed a model of few minutes-ahead wind forecasts capable of capturing wind gusts. The model integrated a CGAN with CNN and LSTM components, leveraging Lidar data as additional input. The CGAN was used to generate realistic wind speed fields from Lidar measurements of wind profiles. Such adversarial training allowed the model to anticipate rapid wind power changes better than conventional time-series models, illustrating the benefit of generative imaging of the wind inflow.

Several works have focused on *load forecasting*. The research performed in Tao et al. (2023) explored the applications of Bidirectional Encoder Representations from Transformers (BERT) for load forecasting. Specifically, BERT architecture was utilized for time-series forecasting by encoding load data from 52 transformers recorded at a 15-min resolution into high-dimensional representations. The BERT model outperformed a LSTM-forecasting model with respect to the forecasting error rate and mean absolute error (MAE) metrics. Specifically, the BERT model was benchmarked against an LSTM-based forecasting model. The results of the study showed that BERT model achieved a relative error of 48.64%, while LSTM had a higher error rate of 53.52%. Additionally, the BERT model's Mean Absolute Error (MAE) was 1.99 kW, compared to 2.91 kW for LSTM, indicating superior prediction accuracy. Likewise, Khan et al. (2024) proposed a GAN-augmented framework for hourly load forecasting in residential and commercial buildings as well as generation from renewable sources. After cleaning the raw data, a GAN generates synthetic time series data, which combined with actual data, created a hybrid dataset, and was used to feed and train a stacked gated recurrent unit (GRU) model. Trained and tested on six benchmark datasets from renewable generation and utility companies, across sixteen linear and nonlinear models for energy forecasting, the models trained on hybrid datasets presented improved performance across the MAE and RMSE metrics.

3.1.2 Weather forecasting

Closely related to energy forecasting and short-term forecasting, several works focus on forecasting weather parameters that are sequentially used for renewable generation forecasting. An early study that laid groundwork for image-based generative forecasting was performed by Andrianakos et al. (2019), where authors pioneered a GAN-based model to extrapolate all-sky camera images for nowcasting cloud cover. By treating sequential sky images as frames, their model generated the next few frames to predict cloud motion and resulting solar irradiance occlusion. A similar study from Nie et al. (2024) predicted cloud dynamic movements and their resulting effect on solar irradiance to provide short-term PV power forecasting. They proposed SkyGPT, a deep learning framework that uses stochastic video prediction by generating synthetic future sky images from past sequences via a physics-informed VideoGPT architecture that incorporates Vector Quantized Variational Autoencoders (VQ-VAE) and Transformers to capture realistic cloud dynamics. Generated images are are then processed by a U-Net-based predictor, trained on data from a 30kWp rooftop PV system, to estimate probabilistic solar power outputs. SkyGPT presented superior accuracy and robustness than other baseline techniques. Meng et al. (2021) introduced an hybrid model that coupled wavelet transforms with an adversarial network for solar power prediction based on global horizontal irradiance data. They decomposed the PV power time series into sub-frequency components using wavelet packet transform, trained a GAN-based predictor for each component, and then recombined the outputs. The model achieved higher accuracy than conventional ANN and persistence models on a solar farm dataset, particularly in capturing high-frequency fluctuations and was used to provide forecasts between 1-6 hr prediction window.

Generative AI models can prove promising in the development of spatio-temporal modeling that are able to capture dependencies across locations and time horizons in distributed energy systems, such as regional solar or wind farms. For example, Wen et al. (2023) introduced a novel GAN framework for regional solar forecasting based on the generation of regional solar irradiance maps. The authors used spatial Kriging interpolation of sensor data to train a model that predicts the spatiotemporal evolution of irradiance using a multi-scale GAN. This enabled estimation of non monitored PV outputs across a region with accuracy comparable to state-of-the-art deterministic methods, while offering broader geographic coverage.

Generally speaking, in solar forecasting (short-term or weather-based), GAN-based models consistently show lower error than baseline deep learning models when sufficient training data and carefully tuned architectures are in place (Huang et al. 2022; Wen et al. 2023), though improvements can be modest and dependent on weather patterns.

Although these models have shown promising accuracy in specific regional cases, broader evaluations indicate that model performance can vary substantially across different climatic and geographic contexts. For example, models trained on a single site often produce larger errors when applied offsite in distinct climate regimes, whereas global and transfer-learning strategies can mitigate some of this variability at the cost of increased training complexity and data requirements (Nie et al. 2024). Meta-analyses further demonstrate that forecast skill correlates with climate zone and dataset composition, underscoring the challenge of achieving uniform performance across heterogeneous regions (Nguyen and Müsgens 2023). These findings suggest that cross-regional inconsistencies remain an open challenge and motivate continued development of adaptation techniques, multi-location training protocols, and evaluation benchmarks that explicitly account for geographic and weather diversity.

3.1.3 Data imputation, data augmentation and synthetic data generation

Some works have used generative AI methods for creating synthetic data e.g. extending existing datasets, either by forecasting new samples to fill missing data (data imputation) or by extending existing datasets by creating new samples (data augmentation). Also, generative models are used to create synthetic time-series data to address data sparsity and improve generalization, particularly in regions or technologies with limited historical records. Research work (Li et al. 2025) integrated conditional GANs (CGANs) with self-attention layers for data imputation i.e. reconstruction of missing or corrupted data in time-series demand datasets. Paper (Zhang et al. 2023) aimed to deal with the data scarcity available for building-integrated PV data. They introduced SolarGAN, a deep generative model trained on a year of facade irradiance data to synthesize realistic hourly solar profiles for a synthetic year. The generated data retained key statistical features (e.g., sunny/cloudy periods) and proved effective for simulating building-integrated PV performance. Similarly, He and Yuan (2023) addressed the issue of limited data for new wind farms by developing a few-sample wind power prediction model using a Least-Squares GAN (LSGAN). Their approach generated additional synthetic wind power time series, effectively augmenting the small historical dataset of a new site. Based on a test wind farm in Spain, they demonstrated how a generative model can fill in informational gaps for improved wind forecasting. The LSGAN augmentation significantly reduced prediction error under data-scarce conditions, i.e. for a test wind farm in Spain, the hybrid model achieved an R^2 of 0.964 and much lower RMSE than a BiLSTM trained on real data only.

Yilmaz and Korn (2022) generated synthetic load data of individual consumers resembling real historic data with the use of several GAN-based models. Santos et al. (2022) developed an innovative system for generating synthetic data designed to address the pervasive challenge of data scarcity in smart grid environments. Their methodology established a three-module architecture leveraging both linked data and GenAI capabilities through VAEs. The system automatically identified and extracted relevant datasets using linked data queries against open data repositories, and then employed VAEs as the core GenAI component to learn underlying data distributions. As demonstrated in several microgrid applications such as energy consumption forecasting, device-level usage analysis, and renewable energy production patterns, the generative approach enabled their system to created statistically meaningful synthetic data from limited samples. Through multiple case studies and comparative analysis, the authors demonstrated that the developed VAE-based approach maintained comparable accuracy in consumption forecasting. Recent work by Kumar et al. (2023) demonstrated how score-based models can generate synthetic load profiles that preserve both individual consumer patterns and aggregate system-level characteristics. This capability is particularly valuable for utilities that need to protect customer privacy while still generating realistic demand scenarios for planning purposes.

In the field of *hydro streamflow forecasting*, Perera et al. (2024) presented a modified GAN model in ungauged basins, incorporating explainable AI techniques that validated the model across known hydrological drivers. Their approach used adversarial training on inputs like climate indices and catchment characteristics, to enhance sparse data and trained the model to learn features critical for extreme events. The model achieved $R^2 \approx 0.93$ on monthly streamflow predictions and was capable of reproducing extreme flood events.

Finally, some works have used GAN models for data filtering and pre-processing such as the work by Kosana et al. (2022). They proposed a hybrid GAN + Temporal Convolutional Network (TCN) -GRU framework for wind speed forecasting, where a vanilla GAN was employed to preprocess and denoise raw wind speed data before feeding it into a TCN and GRU hybrid predictor. By adversarially training on historical wind sequences, the GAN learned to generate smoother and more representative input series, which helped the TCN-GRU model focus on meaningful patterns. Evaluation widely accepted metrics such as MAPE, MAE, RMSE and R^2 showed the model outperformed models without preprocessing, indicating improved predictive accuracy and stability. This suggests GANs can serve not only for direct prediction but also as data filters to enhance signal quality for other forecasting tools.

3.1.4 Scenario-based analysis

In this section, we analyze works that have used generative AI models for generating diverse and plausible future trajectories (e.g., solar or wind generation or demand) and probabilistic future scenarios, enabling robust decision-making under uncertainty.

Capel and Dumas (2023) explored the application of DDPMs for probabilistic energy forecasting, particularly for load, PV, and wind generation. Their study evaluated multiple deep generative models (DGMs), including GANs, VAEs, and normalizing flows (NFs), using the Global Energy Forecasting Competition 2014 (GEFcom2014) dataset (Hong et al. 2016), and assessed their impact on an electricity retailer's day-ahead market bidding strategy across four quality metrics. Within this experimental setup, the authors reported that DDPMs achieved better performance than the other evaluated DGM types in generating probabilistic PV and wind forecasts, while exhibiting a slight underestimation bias for PV and load forecasting. In contrast, NFs demonstrated stronger performance for load forecasting. The study also noted that DDPMs required longer training times and higher computational resources, whereas NFs were easier to calibrate, and GANs showed high sensitivity to hyperparameter tuning.

Paper (Dumas et al. 2022) likewise employed the GEFcom2014 dataset to compare NFs, VAEs, and GANs across seven forecast quality metrics. Their results indicated that NFs outperformed GANs and VAEs for PV and load forecasting across several metrics, while VAEs performed marginally better than NFs for wind forecasting. Based on these findings, the authors concluded that NFs can provide high-quality probabilistic scenarios for energy forecasting, particularly due to their ability to explicitly model complex probability distributions, making them suitable for power system operations and electricity market participation.

Miraki et al. (2025) focused on a specific subtype of DDPMs, namely Graph-based DDPMs (G-DDPMs), designed to capture spatial and temporal dependencies among energy variables. Their probabilistic time-series forecasting model was evaluated using two public hourly-resolution datasets containing renewable generation and electricity demand data from 370 different clients. The results of the study concluded that G-DDPM outperformed all baseline models in key error metrics, with 2.1% increase in continuous ranked probability score (CRPS), 4.4% increase in MAE and RMSE increase of 7.9%. Ortega et al. (2023); Trindade and Electricityloaddiagrams20112014 (2015).

Dumas et al. (2022) also presented a scenario-based probabilistic forecasting framework for wind, solar, and load, incorporating weather-based features such as temperature and humidity. Beyond GANs, their approach leveraged NFs for probabilistic solar power forecasting, generating large scenario ensembles that captured uncertainty more effectively than traditional quantile regression and enabled application in stochastic grid optimization.

Bagheri et al. (2022) employed GAN-based models to generate diverse wind power trajectories for day-ahead stochastic dispatch. Rather than producing a single forecast, the GAN generated multiple plausible realizations of wind power, improving the representation of non-Gaussian forecast error distributions compared to traditional approaches such as Gaussian noise or bootstrapping. Improvements were observed in metrics including prediction interval coverage probability and continuous ranked probability score, indicating a more accurate characterization of uncertainty.

In Yilmaz (2023), GAN-based models were used to generate realistic load profiles for bus clustering and end-user scenario generation, providing a synthetic alternative to costly smart metering data. Complementarily, Verdejo et al. (2019) modeled uncertainty in electric power systems using stochastic differential equations (SDEs), while Olsson et al. (2010) proposed logistic-type SDEs for probabilistic wind power forecasting, highlighting the flexibility of SDE-based frameworks for modeling state-dependent uncertainty and correlation structures.

Generative approaches have also been applied to water demand and usage patterns, which indirectly affect hydroelectric demand scheduling. For example, a GAN was used to generate realistic long-term water demand scenarios in a region (Yang et al. 2024), helping to plan combined water-energy systems. Such cross-domain scenario generation can feed into hydrothermal coordination models for optimizing power generation. Differences from wind/solar include more focus on monthly/annual scales (for water), and on combining data-driven models with hydrological simulations. Evaluation often requires assessing not just error metrics, but whether operational constraints (e.g., not spilling water, meeting flood control rules) are satisfied under generated scenarios, which introduces new metrics for the feasibility and reliability of generative outputs.

The studies above, illustrated that generative AI models can provide a promising avenue for scenario-based analysis. However, several works have shown that challenges remain. For example, one challenge noted is the stability of training when combining GANs with complex forecast models. To tackle this, researchers have tried techniques like the least-squares loss variant (LSGAN) for more stable gradients (He and Yuan 2023), or used Wasserstein GAN (WGAN) with gradient penalty in scenario generation to improve diversity without divergence. These technical refinements are important to ensure that the generator is able to learn the wind power distribution effectively.

Overall, generative AI has shown significant promise in energy forecasting and scenario analysis, addressing limitations in traditional forecasting models. In short-term forecasting applications, generative methods like conditional GANs combined with LSTM and CNN architectures have improved PV and wind power predictions by capturing variability and uncertainty better than baseline models. Transformer-based methods like BERT and diffusion models have also demonstrated advantages in load forecasting accuracy, outperforming traditional statistical and neural network models. Moreover, generative models enhanced by physical and meteorological constraints have contributed notably to accurate weather forecasting, specifically for solar irradiance prediction and cloud dynamics modeling, further supporting renewable generation forecasts. In scenario-based analyses, generative methods such as diffusion models and normalizing flows (NFs) have outperformed GANs and VAEs in generating high-quality probabilistic forecasts, as demonstrated in international forecasting competitions. These models enable better uncertainty quantification and robust decision-making in electricity market operations. Additionally, generative AI has proven effective in data imputation and augmentation tasks by synthesizing realistic time-series data to overcome historical data scarcity, particularly beneficial for smart grid and distributed renewable systems. Overall, integrating generative AI into forecasting workflows provides more reliable and interpretable energy scenarios, enhancing operational efficiency and risk management in modern energy systems. Key differences among approaches lie in the input features used (time-series vs. images, single-site vs. regional data), the type of generative model (standard GAN, CGAN, VAE, etc.), and the training strategy (e.g., adding auxiliary tasks or evolutionary training). Evaluation metrics are typically forecast errors (RMSE, MAE) and sometimes probabilistic scores for scenario-generating models.

Overall, Cluster 1 analysed studies related to forecasting and scenario-based analysis. According to our analysis, GANs emerge as a powerful model in the field of forecasting and scenario-based analysis, with almost half of the studies using GAN architectures for this purpose. Studies in Cluster 1 typically focus on the improvement of the prediction accuracy and typically report a comparison between the developed models to standard and established approaches in the literature, regarding well-established performance metrics and prediction accuracy indices, such as the mean absolute percentage error (MAPE), mean absolute error (MAE), root mean square error (RMSE), R^2. MAE offers the mean absolute error between samples (outlier-tolerant), while RMSE rates higher for deviating results. A lower MAPE indicates a better the forecasting model, while the higher R^2 score indicates how well the model fits the available data, i.e. superior performance of the predictive approach. Indicative performance metrics of reviewed studies are reported in this manuscript, when available. Direct quantitative comparison across studies however must be interpreted with caution, as the studies rely on heterogeneous datasets, different temporal resolutions, forecasting horizons, climatic conditions, and model assumptions. As a result, while the reported metrics provide valuable insight into individual model performance, general conclusions regarding the superiority of a specific forecasting approach or model across applications should be avoided.

3.2 Cluster 2: Energy systems design and network planning

This cluster includes all the works related to grid expansion, network planning and energy system siting and design, including multi-energy systems, microgrid design and DER/RES integration studies. In addition to these applications, we also indicate the specific tasks (e.g., data synthesis, text generation or long term forecasting) performed by the used GenAI models. Generative AI has started playing an essential role in energy system design and network planning by enabling realistic synthetic data generation for robust scenario analysis. GANs and diffusion models produce high-resolution synthetic load profiles and renewable generation patterns, which planners use to test system sizing, resilience, and performance under diverse future conditions without needing extensive historical data. Additionally, generative AI methods help capture rare but critical events, such as extreme weather conditions combined with low renewable output, thus enhancing system adequacy studies beyond traditional scenario methods.

Beyond control policies, generative models have been applied to synthetic data generation that support *grid studies*. For example, GANs have been used to generate synthetic load profiles or solar/wind production profiles for network planning studies, enabling testing of algorithms on “what-if” cases and scenario-based analysis overcoming the need for real data that span over a course of multiple years (He et al. 2024; Hu et al. 2023). Li et al. (2025a) applied Generative AI to active distribution network (ADN) planning by employing a Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) and K-means clustering to model uncertainty in PV generation and load profiles. These synthetic scenarios were used to inform a bi-level expansion planning framework that optimizes the siting and sizing of distributed energy resources including PVs, energy storage, and EV charging stations, alongside network reconfiguration and active management strategies. This work highlights the value of generative models in supporting robust, data-driven planning under high renewable penetration. Study (Li et al. 2025b) introduced an innovative distribution network planning method that combines GAN-generated uncertainty scenarios with Distributionally Robust Optimization (DRO) to determine optimal DER allocation. Applied to the IEEE 33-bus system, their framework achieved a 15% reduction in distribution losses and a 20% improvement in voltage stability compared to conventional methods. The GAN-generated scenarios capture a wide range of load and renewable generation variabilities, which are then integrated into a DRO formulation to ensure cost efficiency and resilience under adverse conditions. This work directly exemplifies how generative AI can enrich the uncertainty modeling phase in network design, leading to robust, adaptive planning outcomes for high-renewable power systems. In wind energy applications, Chen et al. (2022) applied diffusion-based score models to generate realistic wind speed sequences for wind farm planning. Their model successfully captured both the temporal dependencies and spatial correlations across multiple wind measurement locations, enabling more robust wind resource assessments.

Generative AI is also finding use in the design phase of hybrid systems. For *microgrid planning*, GANs have been used to generate high-resolution of years of synthetic demand and renewable production data reflecting climate change or other future conditions (Chen et al. 2018). Planners then optimize system sizing (PV, wind capacity, storage) on those synthetic profiles to ensure resilience. Some studies have begun exploring diffusion models to create realistic time-series, showing that they can capture seasonal and daily patterns effectively, which is crucial for design optimization (Lin et al. 2025; Xu et al. 2025; Zhao et al. 2025). This area is still nascent but growing. Liu and Tang (2024) employed WGAN-GP and K-medoids clustering in optimal microgrid planning to address the uncertainty and intermittency of renewable sources such as wind and PV generation. WGAN-GP was used to generate realistic scenarios for RES and demand without relying on predefined probability distributions, effectively capturing Knightian uncertainty. These scenarios were clustered using K-medoids to identify representative cases for planning. The generative approach was integrated into a multi-objective bi-level planning model, incorporating battery energy storage system (BESS) lifetime and info-gap decision theory under robustness and opportuneness strategies. Results show that this framework enhances planning resilience and cost-effectiveness compared to traditional deterministic methods, highlighting the potential of Generative AI in uncertainty-aware energy system design.

In complex *multi-energy systems* (electricity, heating, gas), generative models can help create integrated scenarios (e.g., a very cold, low-wind week leading to high heating demand and low wind power) to test system adequacy (Xu et al. 2023). Traditional scenario generation might miss such rare but critical combinations, whereas GANs trained on historical data augmented with physics (like weather extremes) could produce them.

A fascinating line of research is using generative design algorithms (like GAN-based topology generation) for laying out wind-solar co-located farms or distribution networks that accommodate renewables. By encoding design constraints into a generative model, researchers have shown the ability to automatically propose configurations (e.g., wind farm layout) that yield high efficiency and low wake losses (Shin et al. 2025), which an engineer can further refine. In a related design oriented application, Oh et al. (2019) proposed a deep generative design framework that integrates topology optimization with generative adversarial networks (GANs) to create engineering designs that balance aesthetics, structural performance, and diversity. While demonstrated on a 2D wheel design, the framework's iterative design exploration and evaluation methodology combining topology optimization for feasibility with GAN-based exploration for diversity offers a blueprint for renewable energy infrastructure layout, such as optimal wind-solar hybrid farm configurations or microgrid network topologies. By quantifying novelty, compliance, and other attributes, such an approach could allow planners to explore unconventional yet efficient layouts for distributed energy systems, moving beyond conventional parametric design limitations. Buster et al. (2024) have used LLM models in renewable energy siting to automate the extraction of zoning ordinances from legal documents, addressing the growing need for efficient access to siting data amid rapid renewable energy deployment in the United States. Their study introduces a hybrid methodology that integrates LLMs with a decision tree framework to populate a public wind and solar siting ordinance database developed by the National Renewable Energy Laboratory (NREL). Achieving an accuracy rate of 85–90%, this approach demonstrates strong potential for supporting downstream quantitative modeling and policy research. The work exemplifies how generative AI techniques can unlock efficiencies in managing complex regulatory information, thereby facilitating large-scale energy system planning and informing future siting strategies for grid expansion and integration of DERs.

Though not as mature as forecasting/control applications, planning applications indicate the breadth of generative AI's potential in renewable energy systems. Generative models support optimal planning of distribution networks and microgrids. By incorporating advanced uncertainty modeling, such as Wasserstein GANs and clustering techniques, researchers have developed frameworks for active distribution network planning, optimal siting of distributed resources, and microgrid resilience under renewable intermittency. These models effectively address complex design constraints and regulatory frameworks, automating processes like renewable energy zoning data extraction and topology optimization. Recent studies highlight substantial improvements in cost efficiency, voltage stability, and design novelty, demonstrating that generative AI significantly enhances adaptive planning strategies in renewable-integrated energy systems.

3.3 Cluster 3: Operation, control and automation

Here, we include all research works related to the operation and control of energy systems including grid optimisation and real-time control, microgrid operation and smart grid energy management. Alongside these applications, we also indicate the specific tasks performed by the used GenAI models (e.g., scenario generation, scheduling, etc). In recent years, generative AI techniques have emerged as powerful tools to address challenges in control, integration and optimization of renewables and power systems. Traditional machine learning techniques often struggle with limited data, non-stationary distributions, and uncertainty quantification. Generative models can mitigate such issues and enable novel optimization schemes, such as training controllers via simulated experiences or adversarial training or learning optimal control policies under uncertainty. While fewer studies apply generative AI directly in smart grid operation, control and automation (as opposed to forecasting), generative models are critical in producing outputs that can be leveraged to design robust control strategies, as illustrated by the works included in this section.

Accurate forecasts produced by generative models feed into energy management systems for PV farms and microgrids, improving decisions on battery usage and grid dispatch. For example, the scenarios generated in the work of Wen et al. (2023) can be used by utilities to estimate PV production in unmonitored parts of the distribution grid, enabling proactive voltage control or reserve allocation. Generative models can also simulate extreme solar scenarios (e.g., prolonged cloud cover or rapid ramps) to stress-test grid control strategies. Khodayar et al. (2019) developed a generative graph-based model for spatio-temporal solar forecasting that could produce a variety of PV production outcomes at different grid nodes. Such tools help grid operators design more robust inverter control settings and voltage regulation schemes, while quantifying the range of possible PV fluctuations.

Generative AI has been applied to multi-energy systems and microgrids, which integrate diverse energy technologies (e.g. solar, wind, storage) across multiple energy vectors (electricity, heat, gas). The operation of such systems presents a complex optimization problem resulting from stochastic energy sources and multiple uncertainties (RES generators, demand, prices). Generative models that can capture the joint distribution of these factors are valuable. Generative AI has been applied to these contexts mainly via scenario generation for stochastic optimization and via generative reinforcement learning/imitation learning to drive control strategies. Han et al. (2024) proposed a hybrid reinforcement learning (WGAN framework) to optimize distribution grid hosting capacity for renewables. Their method combines Proximal Policy Optimization (PPO), a reinforcement learning algorithm, with a Wasserstein GAN for generating high-quality scenarios of demand and renewable production, which the PPO uses to train on a wide range of conditions. This adversarial augmentation of the RL training improved the agent's robustness to out-of-sample events and achieved more than a two-fold improvement in hosting capacity, as showed by simulation analysis on a distribution network with high RES penetration. The work clearly demonstrated that an RL agent with generative scenario support outperforms one trained on limited historical scenarios, which underscores the value of GAN-driven diversity. In the work by Wang et al. (2024) a Generative Adversarial Imitation Learning (GAIL) agent learns the optimal energy scheduling policy for microgrid energy management by imitating an expert. In their setup, the discriminator tries to distinguish between the expert's actions (e.g., how much battery to charge, when to curtail solar) and the generator (policy) actions, providing feedback to make the policy more expert-like. The result was a control policy that nearly matched the expert dispatch cost without requiring an explicit model of the microgrid dynamics during deployment. Such approaches are promising for real-time economic dispatch in microgrids or virtual power plants, as they can learn from offline optimization solutions and then execute fast decisions online. A related work by Zhang et al. (2024) addressed dynamic economic dispatch in multi-vector energy systems (combining electricity, heat and gas) using generative imitation learning. Their framework used a GAN to generate a rich set of renewable fluctuation scenarios, and an imitation learning agent to learn a dispatch strategy that minimizes cost under those scenarios. Compared to deterministic optimization or a pure reinforcement approach, the generative element provided a broader training set, and the learned policy was less sensitive to single forecast errors. Weber et al. (2021) developed an open-source toolbox called OpenModelica microgrid gym (OMG) for simulation and data-driven control optimization of microgrids. Their methodology integrated generative probabilistic machine learning with power electronics control by leveraging Bayesian optimization algorithms to tune controller parameters safely in uncertain operating environments. According to the authors, the GenAI approach automatically discovered optimal controller parameters through iterative sampling and simulation, creating an adaptive system that learned from each experiment. The transfer from simulation to physical laboratory implementation validated the practical applicability of their generative approach. However, limitations included the extensive hyperparameter tuning required for the Bayesian optimization, and the challenges associated with defining appropriate safety thresholds and reward functions. Moreover, the developed generative model demonstrated limitations in providing fully reliable safety predictions due to inherent uncertainties in Gaussian process evaluations, indicating areas for future research improvement.

Generative AI in grid integration and control tends to focus on the robustness and scalability of optimization: The robustness is achieved by exposing optimization algorithms (or learning-based controllers) to a wide range of scenarios generated by generative models; while the scalability is managed by learning policies (via imitation or RL) that avoid heavy computation during real-time, effectively compressing the optimization solution into a neural network policy. Differences in approaches include whether the generative model is used purely for augmentation (e.g., scenario generation offline) or actively in the control loop (e.g., generating adversarial scenarios on-the-fly to challenge an RL agent). Evaluation often involves comparing costs, constraint violation counts, or reward metrics between approaches with and without the generative component. Our review analysis illustrates that generative AI models can significantly enhance the performance and reliability of smart grid operations. Although these GenAI techniques hold substantial promise for facilitating the digital-green transition, persistent challenges remain, particularly regarding generalizability across diverse grid topologies, consistency of synthetic data quality, and robustness of validation procedures. Future research should therefore prioritize designing adaptive architectures capable of operating reliably across heterogeneous microgrid environments, improving synthetic data generation methods, and rigorously addressing uncertainties associated with safety-critical grid controls. As these methodologies continue to evolve and mature, their broader integration will likely accelerate the development of more intelligent, resilient, and sustainable grid infrastructures essential for realizing our renewable energy objectives.

While the applications reviewed above demonstrate GenAI's potential for operational optimization and control, real-world deployment introduces critical challenges beyond algorithmic performance. Unlike offline applications such as planning and scenario generation where computational delays are acceptable, operational control in EMS, DMS, and distributed energy resource management system (DERMS) environments demands deterministic response times ranging from milliseconds for protection systems to minutes for economic dispatch. For context, protective relaying requires response times under 100 ms, while economic dispatch typically operates on 5-15 min cycles (Ghafari et al. 2025). Large diffusion-based and transformer architectures may exhibit inference latencies exceeding these requirements when deployed on centralized infrastructure, creating potential mismatches between model capabilities and operational constraints.

The injection of synthetic data or AI-generated control signals into live operational loops presents distinct cyber-physical risks. Model drift represents a primary concern, as GenAI models trained on historical data may degrade when grid conditions evolve due to increased renewable penetration or demand pattern changes (Zhang et al. 2025). Unlike conventional control systems with well-defined failure modes, generative models can produce plausible but physically infeasible scenarios—generation-demand imbalances violating power flow constraints or voltage setpoints outside equipment ratings—leading to potentially cascading failures if validation mechanisms are insufficient. For example, a diffusion model trained on pre-pandemic demand patterns might generate systematically biased forecasts under post-pandemic conditions, potentially triggering unnecessary load shedding. Adversarial vulnerabilities compound these risks, as malicious actors could craft inputs to trigger erroneous outputs, exploiting model opacity to bypass traditional anomaly detection (Li et al. 2025). Furthermore, the stochastic nature of sampling-based generative models introduces non-determinism conflicting with reproducibility and auditability requirements of critical infrastructure.

Current deployment strategies emphasize layered mitigation approaches preserving deterministic safety guarantees while leveraging GenAI capabilities. Physics-informed constraints are embedded into model architectures or post-processing layers, ensuring generated scenarios respect conservation laws, equipment limits, and operational constraints (Shukla and Deepa 2025). Hybrid frameworks combine GenAI scenario generation with traditional optimization engines or physics-based simulators, enabling validation loops where synthetic outputs are tested against power flow solvers or digital twins before influencing decisions (Choi et al. 2024). Human-in-the-loop architectures position GenAI as decision-support layers rather than autonomous controllers, with operators retaining authority over final commands during abnormal conditions (Choi et al. 2024). Moreover, edge computing deployments co-locating lightweight models with substation controllers reduce communication latency and enhance resilience (Mohammadabadi et al. 2024), though at the cost of reduced model complexity. As discussed in Sect. 2, current implementations position LLMs in advisory roles with latencies of 1-10 s, appropriate for decision support but outside real-time control loops (Ghafari et al. 2025). Ongoing research in verified AI may eventually enable certified safety guarantees for generative models in time-critical applications, though significant technical and regulatory barriers remain.

3.4 Cluster 4: Reliability, resilience and asset management

In this cluster, we include works related to energy system reliability and resilience, anomaly detection, condition monitoring and predictive maintenance. Beyond these applications, we also detail the specific tasks that are achieved by the GenAI models (e.g., data synthesis, anomaly detection, threat detection, etc).

Generative AI has significantly advanced the fields of reliability, resilience, and asset management in energy systems, particularly through anomaly detection and predictive maintenance applications. In wind energy, generative methods like VAE-WGAN models and autoencoder-based architectures effectively detect subtle anomalies in turbine operations, enabling early maintenance scheduling and reducing downtime. These models leverage complex temporal dynamics and transfer learning across different turbine domains, improving fault detection accuracy even when labeled data is limited. Detecting early signs of turbine faults or performance degradation is crucial to schedule maintenance and for optimal control of wind turbines. Additionally, lower downtime also help with the integration of wind energy systems into overall energy generation systems as the investors can sell higher amounts of energy to the grid, maximizing their Return on Investments (ROIs). Zhang and Yang (2023) developed a deep LSTM-based VAE-WGAN model for wind turbine anomaly detection under semi-supervised training. In their approach, a VAE with LSTM layers serves as the generator, modeling the distribution of normal operational data, while a discriminator (using the WGAN framework) ensures the generated samples match real data distributions. The generative model proved adept at capturing complex temporal dynamics (using LSTM) and representing them in a lower-dimensional latent space. An interesting application of GANs in maintenance is domain adaptation across wind turbines. For instance, Peng et al. (2021) proposed an unsupervised transfer learning algorithm based on GANs to predict the health status of a wind turbine fleet. They introduced a specific *WT-GAN* (Generative Adversarial Networks for wind turbine health status prediction), where the generator learns to transform feature distributions from one turbine (source domain) to match another (target domain) without needing labeled failures in the target. By doing so, a fault detection model trained on one turbine can be applied to others, compensating for different operating conditions. This generative transfer approach is valuable because individual turbines often have few fault examples, but a GAN can leverage abundant normal data from all turbines to create a common feature space. Other studies have used autoencoder-based generative models for wind diagnostics. For example, Zhang et al. (2022) proposed a wind turbine anomaly detection and diagnosis method combining LSTM-based stacked denoising autoencoders and XGBoost, using advanced preprocessing, real-time detection, and feature contribution analysis. The effectiveness of their approach was validated on real SCADA data. On the other hand, a variation of MAD-GAN (Multiplicative Adversarial Deep GAN originally for video anomaly) was adapted for multivariate wind time-series anomaly detection (Li et al. 2019). These approaches typically report metrics like detection precision/recall or early fault detection lead time. Generative models often yield more robust detection of subtle anomalies than threshold-based methods, as they better capture the normal data manifold and can identify when data is off-manifold (indicative of a fault).

A different approach related to condition monitoring, predictive maintenance, and resilience is the integration of digital twins with generative models. Wang and Ou (2024) combine digital twins of hydro turbines with generative neural network models. The idea is to train a generative model on sensor data (vibration, pressure, etc.) to emulate the normal operating behavior of a hydro turbine. This surrogate model can then be used to simulate various operating scenarios or detect anomalies. Some initial studies have reported success in detecting subtle changes in vibration patterns using deep autoencoders before they trigger alarms in conventional systems. Generative AI-based digital twin solutions for predictive maintenance have been demonstrated in other industrial domains (Mikołajewska et al. 2025). In hydropower contexts, however, such generative applications are still emerging, and one noticeable trend is combining them with explainable AI and domain knowledge. Trust is a major factor for water resource managers, so methods like LIME (used by Perera et al. (2024)) or SHAP values can help interpret what a generative model is doing (e.g., which historical analogs or climate signals a GAN is using to generate a drought scenario). This is slightly less emphasized in wind/solar papers, but for hydro it appears early adopters are proactive in this regard. The integration of generative AI with digital twins is emerging as a promising approach for predictive maintenance in hydropower systems. By training generative neural networks on turbine sensor data, these virtual replicas help simulate operational scenarios and proactively identify maintenance needs. Early research combining generative models with explainable AI techniques has shown the potential to enhance trust among operators, ensuring transparency about how anomalies or failure scenarios are identified. Though still developing, generative AI-driven digital twins indicate substantial potential for improving resilience and operational reliability in hydroelectric and other critical infrastructure.

3.5 Cluster 5: Sensing, data, digitalization and cybersecurity

This cluster reviews studies around sensing and monitoring, data infrastructure and digitalization of power systems. We include works related to cybersecurity, grid analytics including big data analytics (BDA), IoT and smart energy applications and Non Intrusive Load Monitoring (NILM). We also report the particular tasks that achieved by the GenAI models (e.g., data augmentation, synthesis,...).

3.5.1 Security of cyber-physical systems & cybersecurity

As power systems become increasingly dependent on data and ICT equipment the threat of cyberattacks to vital critical infrastructure increases. This raises the need to invest in research around the topics of digital security and cybersecurity in the context of smart grids. Potential threats may manifest in different forms including malicious attacks, data and privacy violations, unauthorized access, data manipulation and communication network disturbance, which may lead to disruption of smart grid assets and overall system, causing poor performance, asset failure and system blackouts, as described in the work of Aghajari et al. (2025). Noura et al. (2025) illustrated relevant smart grid applications, and includes identification of cyberattacks through anomaly detection, unauthorized access detection such as identification of malware or security breaches in power system equipment, identification of unusual traffic patterns in communication networks, detection of fraudulent behaviour e.g. energy theft, and more generally any attacks on the cyber-physical infrastructure that might compromise safe operation and reliability, including attacks on metering and measurement equipment, control systems and communication networks. While ML and AI techniques have been utilised for detection and protection from such threats in many research works, the potential of generative AI methods specifically is still in its infancy. Generative AI methods have been used both for detecting and creating cyber-attacks that aim to explore the vulnerabilities of the system and inform better detection and protection methods.

In the field of *threat detection*, Wang et al. (2024) integrated GANs and VAEs to detect false data injection attacks (FDIA) in smart grid applications. FDIA aims to manipulate the state estimation (SE) process in smart grids by typically tampering measurement data at advanced metering infrastructures or (AMIs) or supervisory control and data acquisition systems (SCADA). SE refers to the projection of unknown system values such as voltage magnitudes and phase angles, which is required for optimal system management, optimal power flow, economic dispatch and emergency response. The authors claim that Generative AI methods improve the state of the art in FDIA, by not only detecting attacks, but also identifying their exact location. Likewise, Omara and Kantarci (2024) developed a framework to prevent and detect adversarial attacks on Vehicle-to-Microgrid applications. GAN-based methods were used in this work to detect cyber-attacks. Similarly, Li et al. (2025) employed GAN-based methods for voltage and load data augmentation to build a comprehensive dataset for cybersecurity studies. The dataset combined real data from network power flow simulations with synthetic data generated from simulations of denial-of-service attacks. This approach helps address the scarcity of attack-related data, which are often difficult to obtain. The resulting synthetic datasets can then be used to improve the detection of future cyber-attacks in power systems. Ding et al. (2022) combined several AI approaches with generative AI methods (deep auto-encoder, GAN and random forest) to detect intrusions in IoT 5 G networks and reported superior performance of the proposed method.

Generative AI methods can also be used to create and *launch attacks* on power systems. For instance, Liu et al. (2024) used GAN-based methods to launch FDIA attacks on power systems that increasingly rely on DERs and IoT equipment, hence they may exhibit underdeveloped protection schemes. The methods derived aim to perform data manipulation and treat the power system as a black box, in the sense that prior information on system parameters and topology is not required. Similarly, Dash and Khandeparkar (2023) created samples of disturbance variables using deep temporal convolutional GAN, which is combined with original phasor measurement unit (PMU) data to launch FDIA. In the work of Efatinasab et al. (2024), it is shown that GANs can be utilised to manipulate real data from sensing equipment and escape detection. In their work, a perpetrator only requires knowledge of the output of the system's stability model, to successfully launch an attack, without need of data interception or actual model knowledge. Similarly, Li et al. (2024) and Nawaz et al. (2025) formulated GAN methodologies for FDIA attacks and show that these approaches cannot be detected by state-of-the-art detection systems. Furthermore, Ahmadian et al. (2018) utilised GANs to replicate statistical distributions of normal operation data and lauch attacks on the communication layer of smart grid applications.

Generative AI methods can also be used by adversaries engaging in *energy theft* practices. Recently, Kim et al. (2024) provided a general overview of data-driven and ML methods for energy theft detection. The paper states that generative AI models, including diffusion-based LSTM, VAEs and GANs, are used for energy theft detection, especially in relation to data augmentation and expansion of limited datasets. Likewise, Elgarhy et al. (2024) focused on manipulation of smart meter data and AMIs for the purpose of electricity theft. Finally, Badr et al. (2023) demonstrated how malicious actors can exploit GANs to generate fraudulent electricity meter readings, enabling energy theft for illicit financial gain.

3.5.2 Smart energy through IoT

The adoption of Internet-of-Things (IoT) technologies in the energy sector has fundamentally transformed energy management practices. IoT integrates networks of interconnected sensors, actuators, and advanced analytics, enabling dynamic monitoring and informed decision-making. These smart systems significantly enhance energy distribution efficiency, support renewable energy integration, and promote effective demand-side management strategies through real-time user interaction (Al-Turjman and Abujubbeh 2019).

In recent years, GenAI chatbots are increasingly important in enhancing smart energy management systems within the IoT domain. Matharaarachchi et al. (2024) presented a structured approach to utilizing GenAI chatbots to tackle complexities associated with large-scale IoT energy systems aimed at net-zero emissions. Their approach involves six core modules: Intent Classifier, Knowledge Extractor, Database Retriever, Cached Hierarchical Vector Storage, Secure Prompting, and Conversational Interface with Language Generator. The proposed approach effectively streamlined decision-making processes by simplifying complex data analysis tasks and supporting optimized energy network operations, highlighted significant contributions towards reducing operational complexity and carbon emissions. Mohammadabadi et al. (2024) proposed a hybrid framework that combines GenAI with federated learning to address privacy, communication burden, and data scarcity in smart grid communication systems. This paper introduced an architecture in which intelligent edge agents locally process sensor data using fine-tuned GenAI models, avoiding the need to transmit raw data to a central server.

The study further details how diffusion models, unlike GANs which may suffer from mode collapse, use a process of iterative noise addition and denoising to generate high-fidelity, diverse synthetic datasets. The study concludes that this distributed GenAI framework not only safeguards privacy but also enhances forecasting accuracy, demonstrating clear advantages over traditional centralized or standard federated learning approaches in the context of smart grids. Charoensawat et al. (2024) introduced a novel integration framework combining IoT platforms with generative AI tools through an embedded ESP32-based proxy device. The study aimed to evaluate communication efficiency between the IoT system and the OpenAI platform, especially in latency-sensitive environments like smart energy monitoring. The experimental evaluation focused on round-trip time (RTT) measurements across two interfaces.

The study demonstrates the feasibility of lightweight AI-driven interactivity within IoT energy environments, particularly in generating real-time insights from sensor data. However, the authors acknowledged challenges such as fluctuating latency, lack of deterministic performance from AI backends, and scalability issues for broader deployments. These findings provide a foundation for further enhancements in AI-IoT integration, especially or responsive control system in energy management. Zhao et al. (2024) developed an intelligent web-based energy management system (iW-EMS) specifically designed for integrating and optimizing distributed energy resources using a combination of annealing and cone programming, it also leverages GenAI services using historical and real-time data to dynamically generate optimal energy management scenarios. By incorporating data from weather forecasts, market pricing, and consumption profiles, iW-EMS enhances energy efficiency and grid reliability. The web interface includes user-facing chatbot functionalities for interactive control and feedback. Their system was designed to optimize operational efficiency and improve responsiveness to variable renewable outputs. Simulation results indicate improvements in grid load leveling and operational efficiency. The inclusion of vehicle-to-grid (V2G) strategies highlights the added flexibility of using EVs as mobile storage assets.

This section demonstrated that Generative AI chatbbots, proxy-based device integration, and intelligent energy management interfaces are advancing the capabilities of IoT-enabled energy systems by improving operational clarity, automating user interaction, and optimising decision-making under various grid conditions. These innovations highlight the potential of GenAI technologies to bridge communication gaps in complex energy infrastructures, support flexible demand response, and facilitate integration of distributed resources such as EVs. Despite these advances, key challenges persist. The latency variability in AI-IoT communication, limitations in model scalability, and the difficulty of ensuring interoperability across diverse device environments emphasizes the need for further refinement. Additionally, integrating these systems into legacy infrastructure while maintaining system responsiveness and cybersecurity remains an open research concern.

3.5.3 Energy data dissaggregation

Çavdar and Feryad (2021) developed an energy disaggregation model combining bidirectional encoder representations from transformers (BERT) architecture with a novel optimization algorithm called AdaX for NILM. They applied BERT's bidirectional transformer technology, which was originally developed for natural language processing, to energy disaggregation tasks. The authors demonstrated superior performance of their BERT-NILM AdaX model across multiple metrics when applied to two public datasets. While their achievements showed promising results for accurate device-level energy consumption estimation without expensive sensors, shortcomings included performance variability across different appliance types. Badi et al. (2018) combined CNNs with VAEs for energy data disaggregation. The developed methodology leveraged the generative capabilities of VAEs, which employed a probabilistic encoder to learn parameter distributions from aggregated energy signals and a generative decoder to reconstruct appliance-specific consumption patterns. This generative approach enabled the system to model the latent distribution of energy consumption data more effectively than traditional methods, capturing the underlying stochastic variability of appliance usage patterns. The authors achieved significant performance improvements, demonstrating that the developed GenAI-based system outperformed state-of-the-art techniques by 44% on signal aggregate error (SAE) and 19% on mean absolute error (MAE) when tested on the UK-DALE dataset across five common household appliances. However, limitations remained, notably the necessity of appliance-specific network training (requiring separate networks per appliance type), the potential issue of latent space over-pruning, and the experimental scope limited to a single dataset with only five appliances. These limitations suggest a need for further studies to enhance generalization across diverse real-world scenarios.

3.6 Cluster 6: Markets, trading and economics

This cluster is about market operation and trading strategy analysis. In this cluster we also include studies related to forecasting of electricity prices, as this was the primary focus of these studies. In addition to these applications, we also indicate the specific tasks for which each GenAI model is employed.

Medina et al. (2024) explored the applications of GPTs, specifically OpenAI's ChatGPT for forecasting the electricity price trends within the Spanish electricity market. The performed study proposed two methodologies for integrating GPTs within the forecasting domain, namely in context example prompts, and in fine tuned GPT models. The first methodology involved providing GPT with specific examples within prompts to guide its analysis of specialized news feeds to Spanish electricity market. In the second methodology, GPT was fine-tuned using a dataset comprising specialized news and expert reports pertinent to the Spanish electricity market. The researchers aimed to extract valuable insights from energy news and expert reports to generate additional variables that could enhance electricity price trend predictions, extending the ChatGPT use as an Explainable Artificial Intelligence (XAI) for feature engineering. The findings of the research indicated that the insights derived from GPT's analysis closely aligned with subsequent price fluctuations, suggesting that GPT could effectively augment prediction models by introducing additional variables.

Recently, Yilmaz et al. (2024) developed an Electricity-GAN to create realistic electricity price scenarios for market simulation. They used GANs to generate synthetic electricity price data and demonstrated their results on a numerical study based on Turkey's intraday electricity market. In addition to effectively capturing price distribution, they also found that use of more complex GANs increased the computational power significantly without leading to a significant improvement in synthetic data quality.

Walter and Wagner (2024) followed a CGAN approach combined with a one-dimensional CNN and BLSTM, to generate 24-hour time-series data for short-term electricity price scenarios. Their model used historic electricity price data from the European EPEX SPOT market and input variables that typically affect electricity prices, such as production and demand data, and showed the effectiveness of CGANs in generating realistic price scenarios. Similarly, Lu et al. (2022) proposed a CTSGAN deep learning framework for probabilistic forecasting of day-ahead electricity prices in the Australian market. In subsequent work (Lu et al. 2023), the same authors expanded their approach by integrating weather variables, such as temperature, irradiance, and wind speed, to further improve the prediction accuracy. Zhang and Wu (2021) applied a GAN-based approach to forecast locational marginal prices (LMPs) from the perspective of the market participants in the US market. The resulting distributions were able to capture temporal and spatial correlations of LMPs and outperformed benchmark techniques around LMP forecasting. Research work (Hanif et al. 2023) compared GAN-based neural networks with standard machine learning techniques like SVMs, Random Forest and XG-Boost, and proved the effectivess of GANs in the application of electricity price forecasting. Demir et al. (2021) compared several deep generative methods including autoencoders and GANs for data augmentation of time-series electricity price datasets, which yielded to improved forecasting and regression accuracy. They found that VAEs and WGAN-GPs achieved the best-performing results, and in addition they reported that combining multiple techniques could result in further improvements, showcasing a promising future research direction.

3.7 Cluster 7: Socio-technical & textual applications

In this cluster, we include studies related to various applications that cannot be classified in the previous categories.

Generative AI has been utilized to explore *societal aspects* of the energy transition. For instance, generative AI models have been driving research efforts that try to understand public opinion, media narratives, and social responses related to renewable energy developments. Recently, (Mukhamediev et al. 2020) applied generative AI through topic modeling to classify negative news related to socially significant events. They trained a 200 topic BigARTM model, a generative probabilistic framework, on over 800,000 news articles from 40 Kazakhstani media sources (2018–2019), using regularization techniques to improve coherence and sparsity. The generated topic distributions were used within a Multiple Criteria Decision Making (MCDM) framework, where the Analytical Hierarchy Process (AHP) assigned weights to sentiment, objectivity, and social relevance. Their approach achieved an F1-score of 0.81 for negative news detection. Compared to BERT-based classification, the generative model showed similar accuracy but required far fewer labeled examples, highlighting its practicality for large-scale media analysis. Likewise, Kim et al. (2021) applied generative AI through a fine-tuned Robustly optimized Bidirectional Encoder Representations from Transformers (RoBERTa), an advanced transformer-based language model to analyze public sentiment on solar energy in the USA using Twitter data. They collected 1.58 million posts from 2020, filtering 266,686 by geolocation. Of these, 9,000 tweets were manually labeled as positive, neutral, or negative, and 6,300 labeled tweets were used to fine-tune RoBERTa, achieving 80.2% classification accuracy. The analysis revealed predominantly positive public sentiment toward solar energy, while highlighting regional and temporal variations that reflect shifting attitudes influenced by external factors. This demonstrates how generative AI models can effectively capture complex, dynamic trends in large-scale social data. For instance, the field of Generative AI has seen rapid advances across diverse scientific and engineering disciplines, mainly due to its potential to explore vast design spaces and optimize complex systems that were previously beyond the reach of traditional methods. Across domains such as materials science, mechanical engineering, and molecular design, researchers have begun to leverage generative models not only to accelerate discovery but also to embed domain specific physical, chemical, or thermodynamic constraints into the generative process itself.

In *materials discovery*, Xie et al. (2021) tackle a key challenge in solid-state design: generating periodic structures that are quantum-mechanically stable. Unlike small molecule generation, this task involves constructing complex 3D periodic atomic arrangements under bonding constraints, lattice symmetries, and energy minimization. Existing methods often oversimplify these structures or overlook critical physical invariances. To address this, the authors introduce the Crystal Diffusion Variational Autoencoder (CDVAE), a unified generative framework combining variational autoencoders with diffusion-based denoising guided by noise conditional score networks. By embedding physics-informed inductive biases, CDVAE advances generative AI for materials design while reducing reliance on costly quantum simulations in early-stage screening. Similarly, Hoffman et al. (2022) present Query-based Molecule Optimization (QMO), a generative AI framework addressing molecular optimization in drug discovery and materials science. QMO uses a pre-trained sequence-to-sequence autoencoder to map molecules into a latent space learned from unlabeled data, enabling generative exploration. A zeroth-order optimization strategy guides sampling in this space, using black-box property predictors to evaluate decoded molecules and estimate improvement directions, without relying on explicit gradients. This approach effectively generates molecules optimized for drug-likeness, solubility, and SARS-CoV-2 binding affinity, while maintaining structural similarity to leads. QMO also generalizes to tasks like minimizing peptide toxicity. By decoupling generative modeling from optimization, QMO provides a flexible and data-efficient tool for molecular design.

3.8 Large language model applications across smart grid and renewable energy clusters

LLMs constitute an emerging subset of generative AI applications in renewable energy and smart grids. However, our study shows limited current use of LLMs. As can be seen from the reviewed literature, LLMs are not deployed as primary numerical forecasting, optimization, or control tools. Instead, they operate as cross cutting, language driven support components integrated within application clusters for cases such as data generation for better forecasting, energy system planning, etc.

In forecasting and scenario analysis, LLMs are mainly used to assist scenario interpretation, generation of narrative explanations, and interaction with probabilistic outputs produced by other AI/ML models, rather than directly providing time series prediction (Zhang et al. 2026; Mirshekali et al. 2025). Thus, it can be stated that their role is to improve human understanding, provide more in depth data for modeling various assumptions during forecasting, energy system optimization, as well as generating contextual factors associated with renewable generation and demand scenarios.

Within energy system design and planning, generative AI supports a range of knowledge-intensive and optimization tasks. Liu and Tang (2024) employed WGAN-GP and K-medoids clustering for optimal microgrid planning, integrating the generative approach into a multi-objective bi-level planning model that incorporates battery energy storage system lifetime and info-gap decision theory, demonstrating enhanced planning resilience and cost-effectiveness compared to traditional deterministic methods. For regulatory and policy dimensions, Buster et al. (2024) used LLM models to automate the extraction of zoning ordinances from legal documents, achieving 85–90% accuracy in populating a public wind and solar siting ordinance database developed by NREL, thereby facilitating large-scale energy system planning and informing siting strategies for grid expansion and DER integration.

In operation, control, and automation of energy systems, generative AI has been applied primarily through scenario generation for stochastic optimization and generative reinforcement/imitation learning approaches for control strategies. Forecasts produced by generative models feed into energy management systems for PV farms and microgrids, improving decisions on battery usage and grid dispatch, while also enabling stress-testing of grid control strategies through simulation of extreme scenarios such as prolonged cloud cover or rapid ramps (Wang et al. 2019; Bagheri et al. 2022). Wang et al. (2024) employed GAIL for microgrid energy scheduling, where the policy learns to imitate expert dispatch decisions without requiring explicit system dynamics models during deployment. This approach has been extended to multi-vector energy systems combining electricity, heat, and gas, using generative models to produce renewable fluctuation scenarios while imitation learning agents optimize dispatch strategies (Zhang et al. 2024).

For reliability, resilience, and asset management in energy systems, generative AI methods have been applied primarily to anomaly detection and predictive maintenance, particularly in wind energy. Peng et al. (2021) proposed an unsupervised transfer learning algorithm based on GANs (WT-GAN) to predict health status across a wind turbine fleet, where the generator learns to transform feature distributions from one turbine to match another without requiring labeled failures in the target domain, enabling fault detection models trained on one turbine to be applied to others with different operating conditions. Additionally, Zhang et al. (2022) combined LSTM-based stacked denoising autoencoders with XGBoost for wind turbine anomaly detection and diagnosis, validating their approach on real SCADA data. These generative approaches yield more robust detection of subtle anomalies than threshold-based methods by better capturing the normal data manifold and identifying off-manifold deviations indicative of faults.

4 Cyber-physical and social systems (CPSS) for GenAI in energy systems

GenAI is increasingly being applied in energy systems to perform functions such as data augmentation, future scenario generation, and anomaly or threat detection. These functions support a wide range of applications, from system operation and control to market bidding. In fact, the potential of generative AI models cuts across multiple layers of modern power systems by integrating physical, cyber, data, business, policy and social aspects. In this section, building on Sect. 3, we map the use of GenAI onto the extended Smart Grid Architecture Model standard, which incorporates social aspects and thus yields a cyber-physical–social power system framework (Cali et al. 2024). As illustrated in Fig. and discussed in Sect. 4.2, this framework comprises the physical layer, the communication layer, the data layer, and the operational/functional layer, as well as the market, business, and policy/regulatory layers. Table summarises how the functionalities enabled by GenAI models (and highlighted in Sect. 3) contribute to each of these layers.

It is important to note that several sub-functionalities within the CPSS framework could be associated with multiple layers, reflecting the interconnected nature of cyber-physical-social systems. As an example, anomaly detection and diagnostics could theoretically be positioned within the Physical Systems layer (for critical infrastructure health monitoring), the Data layer (for statistical anomaly detection in datasets), or the Communication layer (for network intrusion detection, etc.). Similarly, functionalities such as reliability and contingency analysis could span both the Data layer (through data-driven reliability modeling) and the Communication layer (through network resilience assessment). However, for the purposes of this mapping, we have made deliberate decisions based on the primary application context within energy systems, coupled with the actual focus points of the cited articles within Table . Anomaly detection and diagnostics has been assigned to the Physical Systems layer because as within the energy domain, the primary focus is typically on the health and operational status of physical infrastructure such as wind turbines, power generators, and grid equipment. Likewise, reliability and contingency analysis has been positioned in the Data layer as it fundamentally relies on historical operational data and statistical modeling to assess system reliability and plan for contingencies. These assignments reflect the most prevalent use cases and research focus areas within the energy systems literature, while acknowledging that alternative categorizations could be justified based on different application contexts or research perspectives.

4.1 The sustainable digital transition and GenAI

The global transition toward renewable energy and digitalization is driving the sustainable digital transition (SDT) and reshaping economies, industries, and societies. For instance, a recent comprehensive review by Shewit et al. highlights major advances in deep neural networks and digital twins for smart grids, showing their transformative impact on next-generation energy systems (Tsegaye et al. 2025).

Over the past two decades, the rapid uptake of renewable energy sources has disrupted traditional power systems, requiring more adaptive and intelligent energy management approaches. In parallel, digital technologies such as AI, Distributed Ledger Technology (DLT), Digital Twins, and next-generation ICT systems (5 G/6 G) are transforming the energy landscape, enabling more decentralized, efficient, and resilient infrastructures under the SDT. Among these, GenAI and LLMs stand out as transformative forces across smart grids and renewable energy domains. They enable scenario generation, synthetic data creation, and adaptive decision support that were not feasible with conventional approaches. To contextualize their role, we extend the CPSS framework by embedding GenAI into the Smart Grid Architecture Model (SGAM). This enhanced framework integrates GenAI, ML, and blockchain as digital enablers across physical, communication, data, operation, market, business, and policy layers. Existing research has already identified a broad set of sub-functionalities across these layers, clustered into domains such as digital twins for system representation, anomaly detection and diagnostics, grid studies and network planning, cyberattack simulation, synthetic data and scenario generation, energy forecasting, reliability and contingency analysis, and decision-support functions like demand response, DSM simulation, and multi-energy system control. At the market and business level, current clusters include price forecasting, techno-economic feasibility assessment, and investment risk analysis, while in the governance layer AI-supported policy intelligence has been highlighted. These use cases form the baseline of GenAI-enabled applications for energy systems. Within this structure, GenAI provides new functionalities such as climate-impact stress testing for physical systems, synthetic attack generation for communication networks, socio-energy demand simulation in the data layer, and blackout recovery planning in operational control. At the market and business level, it supports flexibility market simulation and carbon trading with PPAs, while at the policy level it enables generative policy sandboxes for regulatory foresight. Together, these extensions demonstrate how GenAI and LLMs can strengthen resilience, adaptability, and sustainability in the digital transition of energy systems.

4.2 Enhanced CPSS framework for GenAI in energy systems

4.2.1 Physical layer

The physical layer includes all tangible components of the energy system such as generators (both conventional and renewable), transmission and distribution lines, transformers, sensors, storage systems, and consumer side assets like rooftop PV, electric vehicles (EVs), etc.

GenAI enhances this layer by generating synthetic data for training predictive maintenance models, especially where real fault data is scarce. It can simulate failure scenarios for safety testing, assist in generative design for optimizing asset layout (wind turbine siting), and support digital twins for real time system simulation. These enable safe experimentation with control strategies, diagnostics, and behavior forecasting under varied conditions.

As mentioned, the physical layer encompasses system representation and simulation tasks that directly interact with tangible infrastructure and operational conditions of the energy system. Within this layer, digital twins emerge as a critical sub-functionality, enabling high fidelity representation of physical assets for testing, monitoring, and predictive maintenance. Similarly, anomaly detection and diagnostics provide tools for early fault identification and system health monitoring, ensuring operational reliability. Finally, TSO/DSO network optimization links planning with real-time decision making, enhancing system flexibility and resilience.

Therefore, when mapped within the CPSS framework, digital twins including anomaly detection and diagnostics are strongly associated with Cluster 4: Reliability and Asset Management, while network optimization links both to Cluster 2: Energy System Design and Planning and Cluster 4: Reliability and Asset Management. Collectively, these functionalities highlight how generative AI can augment the physical layer by improving resilience, predictive maintenance, and network-wide optimization.

4.2.2 Communication layer

The communication layer can be defined as the nervous system of the smart grid, providing the infrastructures for seamless and secure exchange of information. This includes both wired (e.g., fiber optics, Ethernet) and wireless (e.g., Wi-Fi, ZigBee, LoRa, 5 G) communication technologies. Effective communication is fundamental for enabling system-wide observability, control, coordination, and automation. As smart grids become more decentralized, reliable low-latency communication becomes increasingly important to support real-time demand response, peer-to-peer trading, and integration of distributed energy resources.

Within the CPSS mapping, this layer is represented by resilience and stress testing functionalities. These include cyberattack simulation, network traffic generation, and reliability and contingency analysis. Cyberattack simulation and traffic generation directly contribute to assessing the robustness of communication infrastructures, while reliability and contingency analysis bridges both operational resilience and cybersecurity preparedness.

Accordingly, when mapped to the CPSS framework, these sub-functionalities are primarily linked to Cluster 5: Cybersecurity and Digitalization, with reliability and contingency analysis also extending into Cluster 4: Reliability and Asset Management. Collectively, these highlight the role of generative AI in proactively stress-testing communication infrastructures, modeling attack or fault scenarios, and safeguarding the integrity of smart grid operations.

4.2.3 Data layer

The data layer is responsible for collecting, storing, organizing, and preprocessing data from various components of the energy system. This may includes real-time measurements from sensors and smart meters, historical operational data, weather forecasts, and market signals.

It acts as the foundation for all analytics, Machine Learning (ML) and Decision Making (DM). Ideally, data in this layer must be accurate, timely, and reliable to support effective grid operations and planning. A well-structured data layer ensures that higher-level systems have consistent access to actionable insights, whether it's for optimizing energy flow, forecasting demand, or detecting faults.

Similar to other layers, GenAI can also be utilized for further enhancing the data layer. It can be used to synthesize additional training data when real-world samples are limited or imbalanced, especially for rare events like equipment failures or cyberattacks. It can also fill in missing or corrupted data entries using imputation models that learn from existing patterns. Moreover, generative models can produce entirely new datasets for validating and benchmarking analytical tools under a wide range of hypothetical scenarios.

Within the CPSS mapping, the data layer is defined by data-centric functions such as synthetic data generation, scenario generation, and feature extraction. These sub-functionalities ensure that analytical and forecasting models are supported by diverse, high-quality datasets that capture both real and hypothetical operating conditions. Generative AI strengthens this layer by creating synthetic datasets to overcome data scarcity, generating alternative scenarios to test system robustness, and extracting meaningful features that improve the accuracy and interpretability of predictive models.

Accordingly, the data layer maps directly to Cluster 1: Forecasting and Scenario Analysis. By enhancing data availability, diversity, and reliability, generative AI ensures that forecasting, scenario planning, and decision-support models are more resilient and better equipped to handle uncertainty in modern energy systems.

4.2.4 Operation, optimization and control layer

The operation, optimization, and control layer represents the decision-making hub of the entire energy system due to the fact that it governs real-time operations such as Demand Side Management (DSM), load balancing, voltage and frequency control and optimal scheduling of distributed energy resources like batteries and electric vehicles. This layer ensures the efficient, stable, and secure functioning of the grid, particularly as variable renewable energy sources introduce new uncertainties. This layer consists of control algorithms, optimization engines, and system modeling to make decisions that are executed through the underlying communication and physical layers.

GenAI may enhance this layer by creating diverse operational scenarios that reflect real-world complexities, including sudden load changes, equipment malfunctions, or weather-driven fluctuations in renewable output. These scenarios can be used to train and validate AI-based control strategies, such as Reinforcement Learning (RL) agents or hybrid model-based approaches. GenAI can also contribute to the development of adaptive and explainable control logic by simulating a wide range of "what-if" conditions, helping domain experts refine algorithmic rules for enabling more intelligent grid response.

Similarly, generative AI has been applied to support power system operation through state estimation under partial, noisy, or missing measurement conditions by reconstructing latent system variables and estimating unknown control inputs, improving the completeness and reliability of estimated system states even in data-sparse scenarios (He et al. 2020; Pei et al. 2025). Generative models have also been used for anomaly detection in operational data by learning normal system behaviors and identifying deviations that signal faults or attacks, enabling proactive detection compared with conventional threshold-based methods (Badakhshan and Zhang 2025).

Accordingly, these sub-functionalities map directly to Cluster 3: Operation and Control. Collectively, they demonstrate how generative AI enhances this decision-making hub of the energy system by enabling robust, adaptive, and explainable control strategies that ensure reliable operations under uncertainty.

4.2.5 Power markets and pricing layer

This layer is responsible for the economic and financial governance of the entire energy system, which includes energy generation, consumption, distribution, and transmission. Additionally, mechanisms such as wholesale and retail energy markets, time-of-use pricing, real-time balancing markets, and demand response programs. Participants in this layer include utilities, prosumers, aggregators, and regulators, all of whom interact through market signals to optimize energy flows and achieve cost efficiency.

Generative AI enhances this layer by simulating realistic market environments, generating synthetic demand and price data, and creating unorthodox trading scenarios that allow for robust testing of bidding and risk assessment algorithms. It can also support long-term feasibility studies by modeling investment risks under different policy or regulatory conditions, and it provides new opportunities for simulating carbon markets and PPAs to evaluate their impacts on both system actors and overall market dynamics. Accordingly, these sub-functionalities map primarily to Cluster 6: Markets and Trading, with techno-economic and risk assessment cases also extending into Cluster 7: Socio Technical and Textual Applications. Collectively, they highlight how generative AI strengthens market simulation, risk forecasting, and design exploration, enabling more resilient and adaptive energy markets.

4.2.6 Business layer

AI-powered tools are reshaping energy market dynamics by improving energy price forecasting, Demand Side Management (DSM), and regulatory adaptability. Business viability in this evolving landscape is assessed using key techno-economic indicators such as Return on Investment (ROI), Levelized Cost of Electricity (LCOE), amortization periods, etc. as these metrics guide utilities, aggregators, and prosumers in navigating complex market conditions.

The integration of Generative AI (GenAI) introduces both opportunity and complexity. Its techno-economic feasibility must be evaluated through careful analysis of capital (CAPEX) and operational (OPEX) costs, ensuring that digital infrastructure investments are economically sound and strategically valuable. GenAI can enhance operational efficiency, reduce uncertainty, and support real time decision-making across market functions. As the DGS accelerates, combined assessments of digital and physical energy assets such as AI-based microgrids or GenAI powered digital twins become essential. These holistic evaluations capture the interconnected value and risk of co-dependent infrastructure.

GenAI can also enables advanced scenario modeling for carbon credit markets, Power Purchase Agreements (PPAs), and investment risk analysis, helping stakeholders anticipate regulatory shifts and market volatility. When combined with blockchain, GenAI supports decentralized energy markets by enabling Peer-to-Peer (P2P) trading and secure, tokenized transactions. This convergence fosters innovative, trusted, and inclusive business models within distributed energy ecosystems.

The business layer focuses on techno-economic feasibility, investment risk assessment, and carbon trading and PPAs. Generative AI supports this layer by enabling cost-benefit analysis, scenario modeling, and data-driven trading strategies. Accordingly, these sub-functionalities map to Cluster 6: Markets and Trading, while techno-economic assessments and risk evaluations also extend into Cluster 7: Socio Technical and Textual Applications.

4.2.7 Energy policy and regulatory layer

Policymakers and regulatory bodies are increasingly leveraging Large Language Models (LLMs) and AI-based simulation tools to draft policies, develop energy transition strategies, and ensure regulatory compliance. GenAI plays a crucial role in this domain by enabling scenario-based modeling, where synthetic datasets are generated to assess the potential impact of different policy decisions on energy markets, grid stability, and renewable integration. This capability allows for data-driven policymaking, reducing uncertainty and improving the effectiveness of regulatory interventions.

In addition to policy impact analysis, AI-driven risk assessment models provide critical insights into renewable energy integration, cybersecurity threats, and power grid modernization. These models assist in identifying vulnerabilities, optimizing energy market regulations, and ensuring long-term stability in energy systems. However, the increasing use of AI in regulatory processes also raises concerns about compliance with legal and ethical frameworks. Regulations such as the EU AI Act, GDPR, and other international legal standards must be carefully considered to ensure transparency, fairness, and accountability in AI-driven energy governance. Addressing these challenges will be essential to harness the full potential of GenAI while maintaining regulatory integrity and public trust.

The energy policy and regulatory layer includes policy impact simulation, LLM-assisted regulation drafting, and cyber-physical risk assessment. Generative AI supports this layer by enabling data-driven policymaking and risk analysis. Accordingly, these sub-functionalities map to Cluster 7: Socio Technical and Textual Applications, with cyber-physical risk assessment also linking to Cluster 5: Cybersecurity and Digitalization.

While this CPSS framework provides a comprehensive qualitative mapping of GenAI applications across energy system layers, further development is needed to strengthen both its quantitative rigor and operational utility for energy system stakeholders. Firstly, measurable metrics should be established for each layer, such as Technology Readiness Levels (TRLs) for physical systems, latency and bandwidth requirements for communication layers, data quality scores for synthetic generation, and cost-benefit ratios for operational control applications. Formal mapping rules could then specify which GenAI architectures are best suited for particular layer combinations based on performance benchmarks, computational constraints, and deployment contexts for instance, prioritizing edge compatible diffusion models for real time operation and control, while LLMs excel in policy and regulatory analysis.

Secondly, integration pathways with operational utility systems must be further denoted: the physical layer is mostly responsible for connecting to asset management platforms, whereas the operation layer interfaces with energy management systems and distribution management systems, while market and policy layers within overarching social layer supports trading platforms and regulatory compliance workflows. These operational points ensure that GenAI applications align with utility priorities including grid reliability, cost optimization, and regulatory adherence. Developing such quantitative frameworks and operational mappings would enable systematic assessment of GenAI maturity across CPSS layers, facilitating evidence based technology selection and investment prioritization for energy system participants and stakeholders.

5 Challenges, limitations, and future research directions

The integration of GenAI into renewable energy systems promises a significant paradigm shift in the design, optimization, and governance of energy infrastructure. While advanced techniques spanning diffusion models, generative adversarial networks, and large language models offer unprecedented capabilities in predictive modeling and autonomous control, their deployment introduces multifaceted challenges across technological, ethical, and socio-political domains. Key limitations include data quality dependencies, substantial computational footprints, and emergent risks from opaque decision-making processes in critical infrastructure. Addressing these requires interdisciplinary collaboration among computer science, energy engineering, and policy studies to ensure GenAI advances align with sustainability, equity, and resilience goals in the global energy transition. This alignment constitutes not merely a technical necessity but a societal obligation, as GenAI adoption will inevitably shape resource access, market dynamics, and environmental outcomes for decades.

5.1 Sustainability, compliance, and regulatory dimensions

The deployment of GenAI in power and energy systems is shaped not only by technical performance but also by sustainability considerations, data governance requirements, and regulatory obligations. As these models increasingly support planning, operation, and policy analysis, their environmental footprint, privacy and security implications, and legal accountability become central concerns. This sub-section reviewed these dimensions and their implications for responsible deployment in energy systems.

5.1.1 Energy intensity and carbon cost of model training

The rapid deployment of GenAI models raises critical environmental concerns, particularly regarding the carbon footprint of inference operations. Recent studies reveal that inference can account for up to 90% of a model's total lifecycle energy consumption, as training is a one-time event while inference occurs continuously at global scale (Jegham et al. 2025). For instance, GPT-4o's annual inference operations are projected to emit between 138,125 and 163,441 tons of CO2 equivalent and consume water equivalent to the annual drinking needs of approximately 1.2 million people (Jegham et al. 2025). Similarly, the 176-billion parameter BLOOM model deployed on Google Cloud Platform emitted approximately 19 kg of CO2 equivalent per day, with 75% of energy consumed merely to maintain the model in memory without actively processing requests (Luccioni et al. 2023). These substantial environmental impacts highlights the urgent need for development of carbon aware optimization strategies for GenAI deployment and utilization.

Generation directives present a promising future approach to reduce the carbon footprint of LLM inference without compromising output quality. For example, Li et al. (2024) introduced Sprout which is a framework that leverages generation directives, instructions that guide the autoregressive generation process to produce concise yet accurate responses. By strategically assigning directives to user prompts based on regional carbon intensity and grid conditions, Sprout demonstrated carbon emission reductions exceeding 40% in real-world evaluations using Llama2 models and global electricity grid data (Li et al. 2024). The key insight is that controlling token generation length through directives maintains model size and contextual understanding while significantly reducing computational overhead. For example, hosting larger models like Llama2-13B with generation directives can achieve both lower carbon emissions and higher correctness compared to smaller models like Llama2-7B without directives.

5.1.2 Data privacy and security risks

Generative AI introduces structural data privacy and security risks across the energy domain. These risks do not only stem from a single application but from the way GenAI models learn, generate, and recombine large-scale datasets originating from the nature pf the cyber-physical-social energy systems. GenAI models require extensive training data drawn from smart grids, renewable generation assets, markets, and user-facing systems. These datasets often include fine-grained consumption traces, geospatial information, operational logs, and social context data. When combined, they increase the likelihood of re-identification and unintended inference, even when individual datasets appear anonymized. Key risk vectors include the leakage of training data or AI models during centralized development and storage, where large foundation models aggregate sensitive energy, operational, and social datasets in shared environments. Inference attacks present another significant risk, as adversaries may exploit generative outputs to reconstruct or infer sensitive information embedded in the training data. Synthetic data generation also introduces risks when generated samples reproduce sensitive statistical signatures or rare patterns, undermining privacy and confidentiality guarantees. These risks are further amplified by cross-domain data fusion, where the combination of technical, operational, and social datasets increases both privacy exposure and cybersecurity vulnerabilities across interconnected energy systems. Unlike traditional predictive models, generative models aim to learn full data distributions. This increases the probability of memorization and reconstruction of sensitive patterns which can originate some risks in terms of digital privacy, trade secrets, cybersecurity and national security matters. In energy contexts, this includes household routines, industrial load profiles, and operational vulnerabilities of the critical infrastructure assets. From a security perspective, GenAI expands the attack surface. Models and datasets become high-value targets due to their dual operational and strategic relevance. Adversarial manipulation of training data, prompt injection in language-based interfaces, and poisoning of synthetic scenario generators threaten both analytical integrity and operational trust. However, GenAI approaches can also be used to enhance the cybersecurity of power systems by enabling more accurate state estimation, even in the presence of incomplete data (Ranjbar et al. 2025; He et al. 2020). State estimation can then be used to identify the occurrence of cyberattacks, such as false data injection. Privacy-preserving approaches appear frequently in the reviewed literature as enabling mechanisms rather than afterthoughts. Federated learning, differential privacy, secure aggregation, and validated synthetic data generation shall be considered as structural requirements for scalable GenAI deployment. These techniques reduce exposure but introduce trade-offs in accuracy, uncertainty calibration, and computational cost. Overall, data privacy and security risks in GenAI-enabled energy systems arise from scale, integration, and generative capacity. Addressing them requires design-time decisions rather than post-hoc controls.

5.1.3 Legal and regulatory concerns in grid governance and operations

The deployment of Generative AI in energy systems raises legal and regulatory challenges that extend beyond sector-specific compliance. These challenges emerge from the interaction between AI governance, data protection law, cybersecurity regulation, and critical infrastructure oversight. The reviewed literature highlights that GenAI applications increasingly operate across planning, operation, resilience assessment, and policy analysis. This breadth complicates regulatory alignment. A single model may process personal data, sensitive operational data, and synthetic futures within one analytical pipeline. In the European context, multiple regulatory regimes intersect. Data protection and digital privacy laws govern the collection, processing, and reuse of personal and social datasets employed during model training and inference. Cybersecurity laws regulate model integrity, access control mechanisms, and incident handling obligations to protect AI pipelines and associated digital infrastructure. Critical infrastructure laws frame energy systems as essential services with elevated protection duties, explicitly incorporating national security considerations and systemic risk prevention. In parallel, AI governance laws introduce risk-based obligations related to transparency, data quality management, and human oversight for AI systems deployed in safety- and security-relevant energy applications. These regimes impose cumulative, not alternative, responsibilities. Compliance therefore shifts from component-level checks to system-level governance. A central issue concerns accountability in AI-assisted decision-making. Generative models increasingly support scenario generation, stress testing, and policy analysis. While these functions are advisory, their outputs shape high-impact decisions. Legal responsibility remains with human actors, yet traceability across data sources, model versions, and generated outputs is often weak. Cross-jurisdictional and interoperability matters divergence intensifies this challenge. Binding, compliance-driven AI regulation in some regions contrasts with voluntary or sectoral approaches elsewhere. This creates uncertainty for transnational research projects, shared datasets, and open-source model reuse. Procurement and deployment practices emerge as regulatory leverage points. The current industrial and academic practices point toward embedding legal requirements into technical artifacts, including model documentation, dataset provenance records, and evaluation protocols. These artifacts support auditability and reduce downstream liability. More broadly, the regulation of GenAI in energy systems reflects a shift from static infrastructure governance toward dynamic model governance. Legal frameworks increasingly target processes rather than assets, focusing on how data are generated, transformed, and reused across cyber-physical-social systems. In this context, effective governance of Generative AI requires alignment between legal norms and technical architectures. Without this alignment, advances in modeling and scenario generation risk outpacing the institutional capacity to ensure lawful, secure, and accountable use.

5.2 Key challenges and limitations

5.2.1 Data-level challenges

GenAI deployment in renewable energy systems confronts fundamental data challenges stemming from scarcity and heterogeneity. High-quality, representative datasets are essential for training accurate and robust models; however, renewable infrastructure often generates sparse, noisy, and non-stationary data streams. This is particularly evident in emerging markets and off-grid deployments with limited monitoring capabilities. Weather-dependent renewable variability introduces temporal inconsistencies that violate standard machine learning assumptions of independent and identically distributed data, consequently creating generalization gaps across geographical contexts and evolving grid conditions. Researchers are pursuing physics-informed generative models, such as diffusion processes conditioned on meteorological data to synthesize realistic training samples. Nevertheless, significant validation challenges persist regarding physical plausibility, rare event capture, and equitable data governance frameworks. Ethical concerns emerge from GenAI's tendency to reproduce and amplify biases in training data, potentially leading to discriminatory outcomes in energy distribution and infrastructure siting.

5.2.2 Social acceptability of generative AI in the power sector

The social acceptability of generative AI in the power sector is increasingly recognized as a key factor influencing its adoption, alongside technical performance. It has been highlighted that stakeholders acceptance of AI in power systems is shaped by a set of socio-technical concerns, including trust, transparency, data governance, fairness, and regulatory accountability (Henao et al. 2025). Rather than being determined by accuracy alone, acceptance emerges from how AI systems align with institutional norms, ethical expectations, and stakeholder values.

A central barrier identified to GenAI solutions adoption in power systems is limited transparency in AI-driven decision-making (Henao et al. 2025). The opacity of complex models can reduce trust among operators, regulators, and consumers, particularly when AI systems are applied to safety-critical or high-impact operational tasks. In parallel, the extensive use of high-resolution operational and consumer data raises concerns around privacy, cybersecurity, and responsible data stewardship, which can further influence public and institutional confidence.

Broader research on the social acceptability of generative AI across domains reinforces these findings, emphasizing that cyber-security and surveillance are the main concerns about GenAI, followed by privacy risks, job destruction and system risks due to accuracy (Bialy et al. 2025), whereas perceived usefulness, explainability, and governance mechanisms are critical determinants of acceptance.

5.2.3 Model-level challenges

Computational complexity presents another critical limitation, as cutting-edge GenAI models require extensive training resources, with energy consumption potentially offsetting environmental benefits. Single training runs for advanced models can consume megawatt-hours of electricity, with carbon emissions comparable to small industrial facilities (Rillig et al. 2023). This paradox necessitates innovations in efficient architectures, optimized training protocols, and renewable-powered computing infrastructure. Furthermore, rigorous cost-benefit analyses must ensure operational footprints are justified by tangible energy savings. GenAI models deployed in power systems may experience performance drift over time as grid conditions evolve, highlighting the need for ongoing evaluation and validation to ensure reliable operation.

5.2.4 System integration challenges and reliability testing

Infrastructure integration barriers create substantial bottlenecks, as legacy grid components often lack digital architecture supporting advanced AI. Outdated monitoring systems with limited connectivity and proprietary protocols hinder real-time data exchange, particularly in developing economies where optimization needs are greatest. Overcoming these requires coordinated investment in modern sensors, secure edge computing, and industry-wide standardization to address interoperability issues and cybersecurity vulnerabilities.

These system-level limitations are compounded by regulatory gaps where policymakers struggle to establish governance frameworks for rapidly evolving AI in critical infrastructure. Dual-use threats present additional complications, as capabilities designed for grid optimization could be repurposed to fabricate demand forecasts or manipulate generation predictions. Mitigation requires technical safeguards such as blockchain-verified data provenance combined with adversarial robustness testing and cross-sector ethical guidelines.

5.3 Emerging opportunities and research directions

The rapid evolution of Generative AI, from GANs and VAEs to diffusion and large language models, has unlocked new possibilities across many domains, and renewable energy is no exception. Promising directions include accelerated sampling methods that reduce computational costs without sacrificing quality (Lu et al. 2022), and physics-informed score models that embed domain knowledge into learning processes, improving accuracy and physical consistency (Yang et al. 2023). Likewise, multi-modal extensions capable of jointly modeling solar, wind, and storage while preserving their interdependencies are gaining attention (Pandey et al. 2023).

Looking ahead, research must also focus on efficient, edge-compatible implementations for real-time deployment, enabling use in smart meters, microgrids, and distributed assets. In parallel, diffusion models already show promise in probabilistic electricity market forecasting (Capel and Dumas 2023), and their integration with physical grid constraints could lead to hybrid optimization frameworks spanning electricity, heat, and hydrogen systems. Together, these efforts position GenAI as a key enabler of uncertainty quantification, scenario generation, and robust decision-making for resilient, AI-driven energy management.

Beyond academic research, commercial scale deployment of GenAI in the energy sector has accelerated significantly, with major industry players developing domain specific LLMs for operational applications. Saudi Aramco has pioneered industry first initiatives including Aramco Metabrain AI, a 250 billion parameter GenAI model trained on seven trillion data points encompassing over 90 years of company history, which is capable of analyzing drilling plans, geological data, and provide precise forecasts for refined products including pricing trends and market dynamics (Nadig 2026). Additionally, in a collaborative effort, Aramco Americas, the Society of Petroleum Engineers (SPE), and i2k Connect developed EnergyLLM (ELLM) which is an energy specific LLM fine tuned from Meta's Llama 3.3 with training data from SPE's OnePetro database containing over 314,000 technical items. The developed model currently is in testing phase and demonstrates significantly preferred performance compared to general purpose models when evaluated by domain experts (Warren 2025; Eckroth et al. 2025). Similarly, Repsol has integrated Microsoft Copilot and established the first Generative AI Competence Center in the European energy sector, reporting productivity gains exceeding 2 h per employee per week through automated document analysis, summarization, and task optimization across their operations (Repsol 2026).

GenAI demonstrates transformative potential for energy management through enhanced predictive capabilities and adaptive control systems. Diffusion models show superior performance in probabilistic forecasting for electricity markets (Capel and Dumas 2023), enabling robust bidding strategies that account for renewable uncertainty. Future research should consequently integrate these with physical grid constraints to develop hybrid optimization frameworks for multi-energy systems spanning electricity, heat, and hydrogen vectors.

The convergence of GenAI with IoT enables autonomous energy systems that dynamically reconfigure based on environmental and market signals. Generative adversarial imitation learning frameworks show promise for deriving optimal microgrid dispatch policies from limited operational data (Wang et al. 2024). Extending these to incorporate real-time satellite imagery and weather forecasts could therefore enable fully self-configuring community energy systems that balance local generation, storage, and demand without centralized control.

Decentralized intelligence through edge-deployable GenAI addresses latency and privacy concerns while enabling real-time grid responses. Federated learning architectures with generative models significantly reduce communication overhead (Mohammadabadi et al. 2024), though critical needs remain for hardware-efficient diffusion networks and secure collaborative protocols. Knowledge distillation techniques can accordingly make advanced AI accessible for resource-constrained edge devices.

Policy frameworks must evolve to ensure transparent and accountable AI in critical infrastructure. Explainable GenAI techniques validate predictions against physical principles (Perera et al. 2024), demonstrating how interpretability supports regulatory compliance. Future work should therefore develop human-AI interfaces for participatory governance, incorporating risk-aware policymaking and algorithmic auditing mechanisms tailored to energy systems.

Generative design accelerates sustainable technology development by exploring solution spaces beyond conventional approaches. Physics-informed models create novel energy materials while optimizing multiple constraints (Xie et al. 2021). Similar methodologies could consequently revolutionize photovoltaic nanomaterials and energy storage systems through multi-scale simulation integrating manufacturing limitations.

Socio-technical co-design requires inclusive innovation, aligning technical solutions with community needs. Sentiment-aware models reveal regional disparities in energy adoption attitudes (Kim et al. 2021), informing equitable engagement strategies. Next-stage research should accordingly establish citizen-owned generative models for community energy planning and verifiable data cooperatives that democratize innovation while preserving privacy.

Moreover, beyond these directions, Table . highlights additional future opportunities where GenAI and LLM applications remain underexplored. One such area is carbon trading and power purchase agreements (PPAs), where generative models could simulate contract structures, policy-driven price trajectories, and risk profiles across jurisdictions. This would support investors and utilities in designing resilient financing and hedging strategies. In the physical systems layer, climate-impact stress testing represents another open direction. GenAI could generate long-term synthetic weather and climate scenarios, such as prolonged droughts or multi-year wind anomalies, to evaluate grid resilience under extreme conditions. At the communication layer, synthetic cross-domain attack generation could extend cybersecurity testing by creating hybrid IT-OT-IoT threat patterns not found in historical data, strengthening proactive defense of smart energy infrastructures. For the data layer, socio-energy demand generation offers a pathway to overcome privacy restrictions and data scarcity. GenAI could generate realistic behavioral datasets for prosumers, EV adoption, and demand-side flexibility under varying cultural and policy contexts. In operation and control, future work could explore blackout recovery pathway generation, where generative models propose synthetic black-start and restoration sequences to support contingency planning under high uncertainty. At the market level, flexibility market simulation could leverage GenAI to create synthetic bidding curves, aggregator behaviors, and congestion events, enabling stress tests of emerging local flexibility markets under evolving regulatory frameworks. In policy and regulation, generative policy sandboxes could simulate the outcomes of alternative governance frameworks, such as hydrogen certification schemes or carbon border adjustments. This would provide decision-makers with forward-looking tools to anticipate system-wide impacts before regulatory adoption. Finally, LLMs represent a cross-cutting opportunity across layers. They could support automated analysis of PPAs and contracts, enable compliance monitoring with evolving legal frameworks such as the AI Act and GDPR, and serve as natural-language assistants for system operators and prosumers. LLMs also offer a pathway to integrate technical and textual knowledge, for example, linking sensor data with maintenance logs or policy documents, thereby extending the scope of GenAI into governance, business intelligence, and user-centric innovation.

These opportunities extend the current research agenda by linking physical, cyber, social, market, business, and policy layers. They position GenAI and LLMs not only as drivers of efficiency and prediction, but also as tools for resilience, foresight, and governance in future energy systems.

6 Conclusion

This systematic review examined the relevance and potential role of generative AI models in renewable energy and smart grid applications, analyzing more than 100 research works across seven functional clusters within an extended and integrated CPSS framework. Our findings show that GenAI models, including GANs, VAEs, diffusion models, score-based approaches, and LLMs offer enhanced capabilities, particularly in synthetic data generation, scenario generation capturing uncertainty, and supporting decision-making processes in complex, decentralized energy systems. GANs represent the most widely used class of GenAI models across all domains, while LLMs emerge as a promising future trend, especially when applied to facilitate policy intelligence, regulatory compliance, and stakeholder engagement. Diffusion-based and score-based models show particular promise for probabilistic forecasting, while physics-informed and multi-modal architectures represent critical emerging frontiers.

Our work aimed to provide informative insights to energy practitioners and researchers regarding the use of generative AI models. First, this is achieved by providing the theoretical background, strengths and limitations of the most prominent GenAI model types, including engineering trade-offs such as training cost, stability and memory usage. Second, we provide a structured mapping of different model classes to seven application domains, so practitioners can identify which models are most relevant to specific problem settings. Third, use a CPSS framework to situate the models across different functional layers, such as physical, communication, data, operational, markets, business and policy layers, representing an integrated socio-techno-economic approach.

Furthermore, the work critically discusses limitations and provides practical guidance e.g. model type selection should be aligned to the research task and problem setting (structured taxonomy), combination of GenAI models with physics-based models is recommended in the cases where safety is critical, synthetic data generation addresses critical challenges of data scarcity and privacy, providing utilities and regulators with realistic testbeds for planning and market design, but their accuracy should be benchmarked against existing datasets. Moreover, explainability of solutions and human-in-the loop controls are essential for validation before real-world uptake.

The study also identified emerging opportunities, including climate-impact stress testing, blackout recovery planning, flexibility market simulation, carbon trading with PPAs, and generative policy sandboxes demonstrate GenAI's expanding scope across physical, cyber, and social system layers as future work. However, deployment must address computational costs, data quality dependencies, and ethical considerations. Regulatory frameworks such as the EU AI Act should underscore requirements for transparency, explainability, and trustworthiness. In addition, there are sustainability concerns related to the training of the models, security and legal requirements, while edge-deployment constraints emerge as key practical barriers.

Limitations of our analysis relate to the temporal cut-off of research works under review being set as studies published from 2020-2025. This means that our review cannot capture unpublished pilots or architectural advances that emerge after our cut-off in a rapidly developing field. Moreover, direct model comparison is limited due to the heterogeneity in datasets and modeling assumptions across studies, hence comparative analysis in this paper is mainly qualitative and intends to guide an initial assessment of the potential solutions, which should be followed up with rigorous benchmarking. A promising future direction of this work is to extend this review to include industry pilots, empirical case studies, and standardized benchmarking across common datasets and use cases. Overall, we position GenAI and LLMs as important catalysts for the sustainable digital transition, offering unprecedented capabilities to address uncertainty, decentralization, and complexity, while advancing decarbonization goals and ensuring resilient, adaptive, and equitable energy infrastructures for future generations. However, realizing GenAI's potential calls for interdisciplinary collaboration spanning engineering, data science, policy, and social domains.

Author contributions Umit Cali: Conceived and designed the analysis; collected the data; contributed data or analysis tools; performed the analysis; wrote the paper; supervision of whole research. Ugur Halden: Collected the data; contributed data or analysis tools; performed the analysis; wrote the paper. Merlinda Andoni: Collected the data; contributed data or analysis tools; performed the analysis; wrote the paper. Ferhat Ozgur Catak: Contributed data or analysis tools; performed the analysis; wrote the paper. Si Chen: Collected the data; contributed data or analysis tools; performed the analysis; wrote the paper. Benoit Couraud: Wrote the paper; language wash and final proofreading. Emre Kantar: Wrote the paper; prepared Figure 4. Samuel Knapper: Performed the analysis; wrote the paper; language wash, final proofreading, and initial Figure 6 version. Ibrahim Kucukdemiral: Wrote the paper. Huseyin Kusetogullari: Collected the data; contributed data or analysis tools; performed the analysis; wrote the paper. Murat Kuzlu: Performed the analysis; wrote the paper. Yashar Mousavi: Collected the data; contributed data or analysis tools; performed the analysis; wrote the paper. Sonam Norbu: Collected the data; contributed data or analysis tools; performed the analysis; wrote the paper. Taha Selim Ustun: Wrote the paper. David Flynn: Supervision of research team from University of Glasgow; wrote the paper; final proofreading.

Funding Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital). This work was supported in part by the UK EPSRC through the "DecarbonISation PAThways for Cooling and Heating (DISPATCH)" project [EP/V042955/1], the EPSRC "Hydrogen Integration for Accelerated Energy Transitions Hub (HI-ACT)" project [EP/X038823/2], and by the EPSRC and UK Department for Transport in the "Decarbonised Adaptable and Resilient Transport Infrastructures (DARe)" project [EP/Y024257/1].

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Aghajari HA, Niknam T, Shasadeghi M, Sharifhosseini S, Taabodi M, Sheybani E, Javidi G, Pourbehzadi M (2025) Analyzing complexities of integrating renewable energy sources into smart grid: a comprehensive review. Appl Energy 383:125317

Ahmadian S, Malki H, Han Z (2018) Cyber attacks on smart energy grids using generative adversarial networks. In: 2018 IEEE global conference on signal and information processing (GlobalSIP). IEEE, pp. 942–946

Al Badi H, Boland J, Bruce D, Albadi M (2018) Dust event impact on photovoltaic systems: role of humidity in soiling and self-cleaning. In: 2018 IEEE international conference on smart energy grid engineering (SEGE). IEEE, pp. 342–345

Alimanov A, Islam MB (2023) Denoising diffusion probabilistic model for retinal image generation and segmentation. In: 2023 IEEE international conference on computational photography (ICCP). IEEE, pp. 1–12

Al-Turjman F, Abujubbeh M (2019) IoT-enabled smart grid via SM: an overview. Future Gener Comput Syst 96:579–590

Andrianakos G, Tsourounis D, Oikonomou S, Kastaniotis D, Economou G, Kazantzidis A (2019) Sky image forecasting with generative adversarial networks for cloud coverage prediction. In: 2019 10th International conference on information, intelligence, systems and applications (IISA). IEEE, pp. 1–7

Antonopoulos I, Robu V, Couraud B, Kirli D, Norbu S, Kiprakis A, Flynn D, Elizondo-Gonzalez S, Wattam S (2020) Artificial intelligence and machine learning approaches to energy demand-side response: a systematic review. Renew Sustain Energy Rev 130:109899

Badakhshan S, Zhang J (2025) Generative Ai-enhanced real-time anomaly detection in integrated energy systems. IEEE Trans Smart Grid. https://doi.org/10.1109/TSG.2025.3630562

Badr MM, Mahmoud MM, Abdulaal M, Aljohani AJ, Alsolami F, Balamsh A (2023) A novel evasion attack against global electricity theft detectors and a countermeasure. IEEE Internet Things J 10(12):11038–11053

Bagheri F, Dagdougui H, Gendreau M (2022) Stochastic optimization and scenario generation for peak load shaving in smart district microgrid: sizing and operation. Energy Build 275:112426

Balamurugan M, Narayanan K, Raghu N, Arjun Kumar G, Trupti V (2025) Role of artificial intelligence in smart grid-a mini review. Front Artif Intell 8:1551661

Bialy F, Elliot M, Meckin R (2025) Perceptions of AI across sectors: a comparative review of public attitudes. https://arxiv.org/abs/2509.18233

Biswas P, Rashid A, Al Masum A, Al Nasim MA, Ferdous AA, Gupta KD, Biswas A (2025) An extensive and methodical review of smart grids for sustainable energy management-addressing challenges with AI, renewable energy integration and leading-edge technologies. IEEE Access

Borji A (2019) Pros and cons of gan evaluation measures. Comput Vis Image Underst 179:41–65. https://doi.org/10.1016/j.cviu.2018.10.009

Buster G, Pinchuk P, Barrons J, McKeever R, Levine A, Lopez A (2024) Supporting energy policy research with large language models: a case study in wind energy siting ordinances. Energy AI 18:100431

Cali U, Kuzlu M, Pipattanasomporn M, Kempf J, Bai L (2021) Digitalization of power markets and systems using energy informatics. Springer

Cali U, Catak FO, Halden U (2024) Trustworthy cyber-physical power systems using AI: dueling algorithms for PMU anomaly detection and cybersecurity. Artif Intell Rev 57(7):183

Capel EH, Dumas J (2023) Denoising diffusion probabilistic models for probabilistic energy forecasting. In: 2023 IEEE Belgrade PowerTech. IEEE, pp. 1–6

Çavdar İH, Feryad V (2021) Efficient design of energy disaggregation model with BERT-NILM trained by ADAX optimization method for smart grid. Energies 14(15):4649

Charoensawat P, Inthasuth T, Almahjoub A, Saadi M (2024) Enhancing iot system efficiency through integration with openai: an esp32-based proxy device approach. In: 2024 International technical conference on circuits/systems, computers, and communications (ITC-CSCC). IEEE, pp. 1–5

Chen Y, Wang Y, Kirschen D, Zhang B (2018) Model-free renewable scenario generation using generative adversarial networks. IEEE Trans Power Syst 33(3):3265–3275

Chen C, Buhmann E, Kasieczka G et al (2021) Score-based generative models for calorimeter shower simulation. Phys Rev D 106(9):092009

Chen X, Wang J, Li M, Zhang H (2022) Diffusion-based generative models for wind speed simulation in renewable energy applications. Appl Energy 328:120156

Chen D, Lin X, Qiao Y (2025) Perspectives for artificial intelligence in sustainable energy systems. Energy 318:134711

Choi J, Kim S, Jeong Y, Gwon Y, Yoon S (2021). ILVR: Conditioning method for denoising diffusion probabilistic models. arXiv preprint arXiv:2108.02938

Choi J, Park J, Kyung K, Kim NS, Ahn JH (2023) Unleashing the potential of pim: accelerating large batched inference of transformer-based generative models. IEEE Comput Archit Lett 22(2):113–116

Choi SL, Jain R, Feng C, Emami P, Zhang H, Hong J, Kim T, Park S, Ding F, Baggu M et al (2024) Generative AI for power grid operations. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO (United States)

Creswell A, Bharath AA (2018) Denoising adversarial autoencoders. IEEE Trans Neural Netw Learn Syst 30(4):968–984

Dash SP, Khandeparkar KV (2023) A false data injection attack on data-driven strategies in smart grid using gan. In: International conference on industrial, engineering and other applications of applied intelligent systems. Springer, pp. 313–324

Demir S, Mincev K, Kok K, Paterakis NG (2021) Data augmentation for time series regression: applying transformations, autoencoders and adversarial networks to electricity price forecasting. Appl Energy 304:117695

Ding S, Kou L, Wu T (2022) A GAN-based intrusion detection model for 5G enabled future metaverse. Mob Netw Appl 27(6):2596–2610

Doersch C (2016) Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908

Dos Santos R, Aguilar J, R-Moreno MD (2022) A synthetic data generator for smart grids based on the variational-autoencoder technique and linked data paradigm. In: 2022 XVLIII Latin American computer conference (CLEI). IEEE, pp. 1–7

Dumas J, Wehenkel A, Lanaspeze D, Cornélusse B, Sutera A (2022) A deep generative model for probabilistic energy forecasting in power systems: normalizing flows. Appl Energy 305:117871

Durova MA, Zein AN, Borisova SV, Mishin AA, Kan D, Osipov SK (2023) Autoregressive models for solving the problem of forecasting active energy complexes. In: 2023 5th International youth conference on radio electronics, electrical and power engineering (REEPE), vol. 5, pp. 1–6. https://doi.org/10.1109/REEPE57272.2023.10086906

Eckroth J, et al (2025) Building EnergyLLM: A domain-specific large language model trained on SPE content. In: SPE annual technical conference and exhibition, Houston, Texas, USA. https://doi.org/10.2118/D021S020R009. Society of Petroleum Engineers. https://onepetro.org/SPEATCE/proceedings-abstract/25ATCE/25ATCE/D021S020R009/792077

Efatinasab E, Brighente A, Rampazzo M, Azadi N, Conti M (2024) GAN-GRID: A novel generative attack on smart grid stability prediction. In: European symposium on research in computer security. Springer, pp. 374–393

Elgarhy I, El-Toukhy AT, Badr MM, Mahmoud M, Fouda MM, Alsabaan M, Kholidy HA (2024) Secured cluster-based electricity theft detectors against blackbox evasion attacks. In: 2024 IEEE 21st Consumer communications & networking conference (CCNC). IEEE, pp. 333–338

Fan L, Li L, Ma Z, Lee S, Yu H, Hemphill L (2024) A bibliometric review of large language models research from 2017 to 2023. ACM Trans Intell Syst Technol 15(5):91–19125. https://doi.org/10.1145/3664930

Gawlikowski J, Tassi CRN, Ali M, Lee J, Humt M, Feng J, Kruspe A, Triebel R, Jung P, Roscher R et al (2023) A survey of uncertainty in deep neural networks. Artif Intell Rev 56(Suppl 1):1513–1589

Ghafari M, Sami A, Rezapour H, Ghayour SS, Lin F, Basaran K, Lazaroiu GC, Siano P (2025) A comprehensive review on the application of large language models (llms) in power systems. IEEE Access 13:209450–209486

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial networks. Commun ACM 63(11):139–144. https://doi.org/10.1145/3422622

Guato Burgos MF, Morato J, Vizcaino Imacaña FP (2024) A review of smart grid anomaly detection approaches pertaining to artificial intelligence. Appl Sci 14(3):1194

Halden U, Cali U, Dynge MF, Stekli J, Bai L (2021) DLT-based equity crowdfunding on the techno-economic feasibility of solar energy investments. Sol Energy 227:137–150

Han J, Cai C, Pan W, Liu H, Xu Z (2024) Hybrid proximal policy optimization–Wasserstein generative adversarial network framework for hosting capacity optimization in renewable-integrated power systems. Energies 17(24):6234

Hanif M, Shahzad MK, Mehmood V, Saleem I (2023) Epfg: electricity price forecasting with enhanced gans neural network. IETE J Res 69(9):6473–6482

Harrou F, Dairi A, Dorbane A, Sun Y (2024) Enhancing wind power prediction with self-attentive variational autoencoders: a comparative study. Res Eng 23:102504

Hayashi H (2025) A hybrid of generative and discriminative models based on the Gaussian-coupled softmax layer. IEEE Trans Neural Netw Learn Syst 36(2):2894–2904. https://doi.org/10.1109/TNNLS.2024.3358113

He H, Yuan M (2023) A novel few-sample wind power prediction model based on generative adversarial network and quadratic mode decomposition. Front Energy Res 11:1211360

He Y, Chai S, Xu Z, Lai CS, Xu X (2020) Power system state estimation using conditional generative adversarial network. IET Gener Transm Distrib 14(24):5823–5833. https://doi.org/10.1049/iet-gtd.2020.0836

He Z, Huang R, Wang X, Fei F, Zhou H, Xu C, Ai X (2024) Renewable energy scenario generation based on improved generative adversarial network. In: 2024 4th international conference on energy, power and electrical engineering (EPEE). IEEE, pp. 1175–1180

Henao F, Edgell R, Sharma A, Olney J (2025) Ai in power systems: a systematic review of key matters of concern. Energy Inform 8(1):76. https://doi.org/10.1186/s42162-025-00529-1

Ho J, Jain A, Abbeel P (2020a) Denoising diffusion probabilistic models. Adv Neural Inf Process Syst 33:6840–6851

Ho J, Jain A, Abbeel P (2020b) Denoising diffusion probabilistic models. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H (eds.) Advances in neural information processing systems, vol. 33, pp. 6840–6851. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf

Hoffman SC, Chenthamarakshan V, Wadhawan K, Chen P-Y, Das P (2022) Optimizing molecules using efficient queries from property evaluations. Nat Mach Intell 4(1):21–31

Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ (2016) Probabilistic energy forecasting: global energy forecasting competition 2014 and beyond. Elsevier

Hu Y, Li Y, Song L, Lee HP, Rehm P, Makdad M, Miller E, Lu N (2023) MultiLoad-GAN: A GAN-based synthetic load group generation method considering spatial-temporal correlations. IEEE Trans Smart Grid 15(2):2309–2320

Huang X, Li Q, Tai Y, Chen Z, Liu J, Shi J, Liu W (2022) Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. Energy 246:123403

Hussan U, Wang H, Peng J, Jiang H, Rasheed H (2026) Transformer-based renewable energy forecasting: a comprehensive review. Renew Sustain Energy Rev 226:116356

Hyvärinen A, Dayan P (2005) Estimation of non-normalized statistical models by score matching. J Mach Learn Res 6(4):1

Jegham N, Abdelatti M, Koh CY, Elmoubarki L, Hendawi A (2025) How hungry is Ai? Benchmarking energy, water, and carbon footprint of llm inference. arXiv preprint arXiv:2505.09598

Kalkal P, Garg VK (2017) Transition from conventional to modern grids: modern grid include microgrid and smartgrid. In: 2017 4th International conference on signal processing, computing and control (ISPCC). IEEE, pp. 223–228

Karras T, Aittala M, Aila T, Laine S (2022) Elucidating the design space of diffusion-based generative models. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A. (eds.) Advances in neural information processing systems, vol. 35, pp. 26565–26577. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/a98846e9d9cc01cfb87eb694d946ce6b-Paper-Conference.pdf

Khaloie H, Dolanyi M, Toubeau J-F, Vallée F (2025) Review of machine learning techniques for optimal power flow. Appl Energy 388:125637

Khan N, Khan SU, Farouk A, Baik SW (2024) Generative adversarial network-assisted framework for power management. Cogn Comput 16(5):2596–2610

Khodayar M, Mohammadi S, Khodayar ME, Wang J, Liu G (2019) Convolutional graph autoencoder: a generative deep neural network for probabilistic Spatio-temporal solar irradiance forecasting. IEEE Trans Sustain Energy 11(2):571–583

Kim SY, Ganesan K, Dickens P, Panda S (2021) Public sentiment toward solar energy–opinion mining of twitter using a transformer-based language model. Sustainability 13(5):2673

Kim S, Sun Y, Lee S, Seon J, Hwang B, Kim J, Kim J, Kim K, Kim J (2024) Data-driven approaches for energy theft detection: a comprehensive review. Energies 17(12):3057

Kosana V, Teeparthi K, Madasthu S (2022) A novel and hybrid framework based on generative adversarial network and temporal convolutional approach for wind speed prediction. Sustain Energy Technol Assess 53:102467

Kumar R, Patel A, Singh M, Thompson S (2023) Score-based generative models for synthetic load profile generation in smart grids. IEEE Trans Smart Grid 14(3):2145–2157

Lai G, Chang W-C, Yang Y, Liu H (2022) Conditional score-based diffusion models for solar power forecasting. IEEE Trans Sustain Energy 14(2):1031–1041

Larsen ABL, Sønderby SK, Larochelle H, Winther O (2016) Autoencoding beyond pixels using a learned similarity metric. In: International conference on machine learning. PMLR, pp. 1558–1566

Leinonen J, Nerini D, Berne A (2020) Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. IEEE Trans Geosci Remote Sens 59(9):7211–7223

Li W, Han M (2023) Mapping renewable energy transition worldwide: gravity trajectory, contribution decomposition and income levels. Renew Energy 206:1265–1274

Li F, Zheng H, Li X (2022) A novel hybrid model for multi-step ahead photovoltaic power prediction based on conditional time series generative adversarial networks. Renew Energy 199:560–586

Li D, Chen D, Jin B, Shi L, Goh J, Ng S-K (2019) Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks. In: International conference on artificial neural networks. Springer, pp. 703–716

Li C, Kies A, Zhou K, Schlott M, El Sayed O, Bilousova M, Stöcker H (2024a) Optimal power flow in a highly renewable power system based on attention neural networks. Appl Energy 359:122779

Li B, Jiang Y, Gadepally V, Tiwari D (2024b) Sprout: Green generative ai with carbon-efficient llm inference. In: Proceedings of the 2024 conference on empirical methods in natural language processing, pp. 21799–21813

Li J, Yang Y, Sun J (2024c) Risks of practicing large language models in smart grid: threat modeling and validation. arXiv preprint arXiv:2405.06237

Li F, Shen W, Bi Z, Su X (2024d) Sparse adversarial learning for fdia attack sample generation in distributed smart grids. Comput Model Eng Sci (CMES) 139(2)

Li P, Shen Y, Shang Y, Alhazmi M (2025a) Innovative distribution network design using gan-based distributionally robust optimization for dg planning. IET Gener Transm Distrib 19(1):13350

Li Y, Zhang S, Li Y (2025b) Ai-enhanced resilience in power systems: adversarial deep learning for robust short-term voltage stability assessment under cyber-attacks. Chaos Solitons Fract 196:116406. https://doi.org/10.1016/j.chaos.2025.116406

Li H, Li HF, Shi Y (2025c) Vision language models: methods, datasets and training, and bibliometric analysis. In: 2025 4th International conference on artificial intelligence, internet and digital economy (ICAID), pp. 242–245. https://doi.org/10.1109/ICAID65275.2025.11034473

Li J, Li F, Sun K, Sun B (2025d) Bi-level coordinated expansion planning of high-PV-penetration active distribution networks incorporating multiple flexible resources considering WGAN-GP-based uncertainty generation. Energy, 136685

Li J, Zhao Z, Jin T (2025e) Short-term load forecasting in smart grids: a CGAN-self data reconstruction and BiTCN-BiGRU-self attention model with demand response optimization. Expert Syst Appl, 128553

Liao Z, Kally J, Ru S (2024a) Probabilistic modeling of renewable energy sources in smart grids: a stochastic optimization perspective. Sustain Cities Soc 109:105522

Liao W, Wu L, Xu S, Fujimura S (2024b) A novel approach for intelligent fault diagnosis in bearing with imbalanced data based on cycle-consistent gan. IEEE Trans Instrument Meas

Lin N, Palensky P, Vergara PP (2025) Energydiff: Universal time-series energy data generation using diffusion models. IEEE Trans Smart Grid

Liu H, Tang W (2024) Multi-objective bi-level programs for optimal microgrid planning considering actual bess lifetime based on WGAN-GP and info-gap decision theory. J Energy Storage 89:111510

Liu Y, Martinez C, Johnson M, Brown E (2022) Constrained score-based generative models for energy system applications. Energy AI 9:100182

Liu Z, Liu M, Wang Q, Tang Y (2024) False data injection attacks on data-driven algorithms in smart grids utilizing distributed power supplies. Engineering

Lu X, Qiu J, Lei G, Zhu J (2022a) Scenarios modelling for forecasting day-ahead electricity prices: case studies in Australia. Appl Energy 308:118296

Lu C, Zhou Y, Bao F, Chen J, Li C, Zhu J (2022b) Dpm-solver: a fast ode solver for diffusion probabilistic model sampling in around 10 steps. Adv Neural Inf Process Syst 35:5775–5787

Lu X, Qiu J, Lei G, Zhu J (2023) An interval prediction method for day-ahead electricity price in wholesale market considering weather factors. IEEE Trans Power Syst 39(2):2558–2569

Luccioni AS, Viguier S, Ligozat A-L (2023) Estimating the carbon footprint of bloom, a 176b parameter language model. J Mach Learn Res 24(253):1–15

Mariani M, Dwivedi YK (2024) Generative artificial intelligence in innovation management: a preview of future research developments. J Bus Res 175:114542

Matharaarachchi A, Mendis W, Randunu K, De Silva D, Gamage G, Moraliyage H, Mills N, Jennings A (2024) Optimizing generative AI chatbots for net-zero emissions energy internet-of-things infrastructure. Energies 17(8):1935

Mazhar T, Irfan HM, Haq I, Ullah I, Ashraf M, Shloul TA, Ghadi YY, Imran EDH (2023) Analysis of challenges and solutions of IoT in smart grids using AI and machine learning techniques: a review. Electronics 12(1):242

Mazhar T, Shahzad T, Rehman AU, Hamam H et al (2025) Integration of smart grid with industry 5.0: applications, challenges and solutions. Meas Energy 5:100031

Menéndez Medina A, Heredia Álvaro JA (2024) Using generative pre-trained transformers (GPT) for electricity price trend forecasting in the Spanish market. Energies 17(10):2338

Meng F, Zou Q, Zhang Z, Wang B, Ma H, Abdullah HM, Almalaq A, Mohamed MA (2021) An intelligent hybrid wavelet-adversarial deep model for accurate prediction of solar power generation. Energy Rep 7:2155–2164

Mikołajewska E, Mikołajewski D, Mikołajczyk T, Paczkowski T (2025) Generative AI in AI-based digital twins for fault diagnosis for predictive maintenance in Industry 4.0/5.0. Appl Sci 15(6):3166

Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R (2021) Nerf: Representing scenes as neural radiance fields for view synthesis. Commun ACM 65(1):99–106

Miraki A, Parviainen P, Arghandeh R (2025) Probabilistic forecasting of renewable energy and electricity demand using graph-based denoising diffusion probabilistic model. Energy AI 19:100459

Mirshekali H, Reza Shadi M, Ghanadi Ladani F, Reza SH (2025) A review of large language models for energy systems: applications, challenges, and future prospects. IEEE Access 13:163162–163188. https://doi.org/10.1109/ACCESS.2025.3610994

Mohammadabadi SMS, Entezami M, Moghaddam AK, Orangian M, Nejadshamsi S (2024) Generative artificial intelligence for distributed learning to enhance smart grid communication. Int J Intell Netw 5:267–274

Mongaillard T, Lasaulce S, Varma VS (2025) Smart energy management: from conventional optimization to Generative AI techniques. Internet of vehicles and computer vision solutions for smart city transformations. Springer, Cham, pp 337–361

Mukhamediev RI, Yakunin K, Mussabayev R, Buldybayev T, Kuchin Y, Murzakhmetov S, Yelis M (2020) Classification of negative information on socially significant topics in mass media. Symmetry 12(12):1945

Nadig S (2026). Saudi Aramco Unveils industry's first generative AI model. https://www.offshore-technology.com/news/saudi-aramco-unveils-industry-first-generative-ai-model/ Accessed 2026-01-26

Nawaz R, Akhtar R, Khan SU, Bu S, Mahmood MH (2025) Deep learning-driven false data injection attack in renewable integrated smart grids. Eng Appl Artif Intell 156:110953

Nelson T, Nolan T, Gilmore J (2025) Effective policy to achieve the Australian government's commitment to 82 per cent renewable energy by 2030. Australasian J Environ Manag, 1–24

Nguyen TN, Müsgens F (2023) A meta-analysis of solar forecasting based on skill score. https://arxiv.org/abs/2208.10536

Nie Y, Paletta Q, Scott A, Pomares LM, Arbod G, Sgouridis S, Lasenby J, Brandt A (2024a) Sky image-based solar forecasting using deep learning with heterogeneous multi-location data: dataset fusion versus transfer learning. Appl Energy 369:123467. https://doi.org/10.1016/j.apenergy.2024.123467

Nie Y, Zelikman E, Scott A, Paletta Q, Brandt A (2024b) Skygpt: Probabilistic ultra-short-term solar forecasting using synthetic sky images from physics-constrained videogpt. Adv Appl Energy 14:100172

Noura HN, Yaacoub JPA, Salman O, Chehab A (2025) Advanced machine learning in smart grids: an overview. Internet Things Cyber-Phys Syst

Oh S, Jung Y, Kim S, Lee I, Kang N (2019) Deep generative design: integration of topology optimization and generative models. J Mech Des 141(11):111405

Olsson M, Perninge M, Söder L (2010) Modeling real-time balancing power demands in wind power systems using stochastic differential equations. Electric Power Syst Res 80(8):966–974

Omara A, Kantarci B (2024) An ai-driven solution to prevent adversarial attacks on mobile vehicle-to-microgrid services. Simul Model Pract Theory 137:103016

Ozawa A, Tsani T, Kudoh Y (2022) Japan's pathways to achieve carbon neutrality by 2050-scenario analysis using an energy modeling methodology. Renew Sustain Energy Rev 169:112943

Pagnoni A, Liu K, Li S (2018) Conditional variational autoencoder for neural machine translation. arXiv preprint arXiv:1812.04405

Pandey A, Weber K, Nakamura Y, O'Brien P (2023) Multivariate score-based diffusion models for renewable energy portfolio optimization. Renew Sustain Energy Rev 178:113245

Panthi M (2020). Anomaly detection in smart grids using machine learning techniques. In: 2020 First international conference on power, control and computing technologies (ICPC2T). IEEE, pp. 220–222

Patil A et al (2021) DCGAN: Deep convolutional GAN with attention module for remote view classification. In: 2021 International conference on forensics, analytics, big data, security (FABS). IEEE, vol. 1, pp. 1–10

Peebles W, Xie S (2023) Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 4195–4205

Pei J, Wang P, Wang J, Shi D (2025) Deep generative model-aided power system dynamic state estimation and reconstruction with unknown control inputs or data distributions. https://arxiv.org/abs/2501.02928

Peng D, Liu C, Desmet W, Gryllias K (2021) Deep unsupervised transfer learning for health status prediction of a fleet of wind turbines with unbalanced data. In: Proceedings of the annual conference of the PHM society 2021

Perera U, Coralage D, Ekanayake I, Alawatugoda J, Meddage D (2024) A new frontier in streamflow modeling in ungauged basins with sparse data: a modified generative adversarial network with explainable AI. Res Eng 21:101920

Ranjbar MA, Azad S, Nazari-Heris M, Mohammadpourfard M (2025) Overview of key power system state estimation methods with a focus on artificial intelligence-based approaches. e-Prime - Adv Electric Eng Electron Energy 14:101125. https://doi.org/10.1016/j.prime.2025.101125

Rasul K, Seward C, Schuster I, Vollgraf R (2021) Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. Int Conf Mach Learn. PMLR, 8857–8868

Reddy S (2024) Generative ai in healthcare: an implementation science informed translational path on application, integration and governance. Implement Sci 19(1):27

Repsol: LLM (Large Language Models): what are they and examples? Repsol. https://www.repsol.com/en/energy-move-forward/innovation/llm/index.cshtml Accessed 2026-01-26

Rillig MC, Ågerstrand M, Bi M, Gould KA, Sauerland U (2023) Risks and benefits of large language models for the environment. Environ Sci Technol 57(9):3464–3466

Rojas Ortega S, Castro-Correa P, Sepúlveda-Mora S, Castro-Correa J (2023) Renewable energy and electricity demand time series dataset with exogenous variables at 5-minute interval. Mendeley Data. https://doi.org/10.17632/fdfftr3tc2.1 . https://doi.org/10.17632/fdfftr3tc2.1

Roman D, Saxena S, Robu V, Pecht M, Flynn D (2021) Machine learning pipeline for battery state-of-health estimation. Nat Mach Intell 3(5):447–456

Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B (2022) High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042 . IEEE

Sarmas E, Marinakis V, Doukas H (2025) Future directions of intelligent energy management and the role of Generative AI. Artificial intelligence for energy systems: driving intelligent. Flexible and optimal energy management. Springer, Cham, pp 257–266

Sarp S, Kuzlu M, Wilson E, Guler O (2021) WG 2 AN: Synthetic wound image generation using generative adversarial network. J Eng 2021(5)

Sengar SS, Hasan AB, Kumar S, Carroll F (2024) Generative artificial intelligence: a systematic review and applications. Multimedia Tools Appl, 1–40

Shin Y, Yun T, Yun S, Woo S, Lim S, Park J (2025) Wind farm layout optimization with diffusion models. In: Proceedings of the 31st ACM SIGKDD conference on knowledge discovery and data mining, Volume 2, pp. 2574–2585

Shukla PK, Deepa K (2025) Ai-based synthetic data generation techniques for improved fault classification in power systems. Ain Shams Eng J 16(8):103485. https://doi.org/10.1016/j.asej.2025.103485

Sonata I, Heryadi Y (2024) Comparison of lstm and transformer for time series data forecasting. In: Proceedings of the 7th international conference on informatics and computational sciences (ICICoS). IEEE, pp. 491–495

Song Y, Ermon S (2019) Generative modeling by estimating gradients of the data distribution. Adv Neural Inform Process Syst 32

Song Y, Ermon S (2020) Improved techniques for training score-based generative models. Adv Neural Inf Process Syst 33:12438–12448

Song Y, Durkan C, Murray I, Ermon S (2021a) Maximum likelihood training of score-based diffusion models. Adv Neural Inf Process Syst 34:1415–1428

Song J, Meng C, Ermon S (2021b) Denoising diffusion implicit models. Int Conf Learn Represent

Song Y, Sohl-Dickstein JN, Kingma DP, Kumar A, Ermon S, Poole B (2021c) Score-based generative modeling through stochastic differential equations. In: Proceedings of the 35th international conference on neural information processing systems. Curran Associates, Inc., pp. 12487–12502

Stewart I, Burnett N (2025) The UK's plans and progress to reach net zero by 2050. House of Commons Library, 1–5

Tao P, Ma H, Li C, Liu L (2023) Intelligent grid load forecasting based on BERT network model in low-carbon economy. Front Energy Res 11:1197024

Tedre M, Vartiainen H (2023) How text-to-image generative AI is transforming mediated action? Authorea Preprints

Trindade A (2015) ElectricityLoadDiagrams20112014. https://doi.org/10.24432/C58C86. UCI machine learning repository, [Online]

Tsegaye S, Heyi KG, Endaylalu MT, Melaku ZA, Turufi KT (2025) Deep neural networks in smart grid digital twins: evolution, challenges, and future outlooks. IEEE Access 13:114845–114864. https://doi.org/10.1109/ACCESS.2025.3585967

Vahdat A, Kautz J (2020) NVAE: A deep hierarchical variational autoencoder. Adv Neural Inf Process Syst 33:19667–19679

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inform Process Syst 30

Verdejo H, Awerkin A, Kliemann W, Becker C (2019) Modelling uncertainties in electrical power systems with stochastic differential equations. Int J Electric Power Energy Syst 113:322–332

Vincent P (2011) A connection between score matching and denoising autoencoders. Neural Comput 23(7):1661–1674. https://doi.org/10.1162/NECO_a_00142

Walter V, Wagner A (2024) Probabilistic simulation of electricity price scenarios using conditional generative adversarial networks. Energy AI 18:100422

Wang H, Ou S (2024) Structured neural network modeling for developing digital twins models of hydropower generation units. In: 2024 IEEE 19th conference on industrial electronics and applications (ICIEA). IEEE, pp. 1–6

Wang AX, Nguyen BP (2025) Ttvae: Transformer-based generative modeling for tabular data generation. Artif Intell 340:104292

Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y, Change Loy C (2018) Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV) workshops

Wang X, Takaki S, Yamagishi J, King S, Tokuda K (2019a) A vector quantized variational autoencoder (VQ-VAE) autoregressive neural f_0 model for statistical parametric speech synthesis. IEEE/ACM Trans Audio Speech Lang Process 28:157–170

Wang F, Zhang Z, Liu C, Yu Y, Pang S, Duić N, Shafie-Khah M, Catalao JP (2019b) Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. Energy Convers Manag 181:443–462

Wang Y, Zhou Y, Ma J et al (2024a) A locational false data injection attack detection method in smart grid based on adversarial variational autoencoders. Appl Soft Comput 151:111169

Wang K, Zhang Z, Meng K, Lei P, Wang R, Yang W, Lin Z (2024b) Optimal energy scheduling for microgrid based on GAIL with Wasserstein distance. AIP Adv 14(8)

Warren EF (2025). Curiosity meets comprehension: how energy LLM is bridging knowledge gaps. J Petrol Technol. Accessed 2026-01-26

Weber D, Heid S, Bode H, Lange JH, Hüllermeier E, Wallscheid O (2021) Safe Bayesian optimization for data-driven power electronics control design in microgrids: from simulations to real-world experiments. IEEE Access 9:35654–35669

Wen H, Du Y, Chen X, Lim EG, Wen H, Yan K (2023) A regional solar forecasting approach using generative adversarial networks with solar irradiance maps. Renew Energy 216:119043

Wu H, Zheng S, Zhang J, Huang K (2019) GP-GAN: Towards realistic high-resolution image blending. In: Proceedings of the 27th ACM international conference on multimedia, pp 2487–2495

Xie T, Fu X, Ganea O-E, Barzilay R, Jaakkola T (2021) Crystal diffusion variational autoencoder for periodic material generation. arXiv preprint arXiv:2110.06197

Xiong B, Chen Y, Chen D, Fu J, Zhang D (2025) Deep probabilistic solar power forecasting with transformer and gaussian process approximation. Appl Energy 382:125294

Xu J, Ji Z, Liu X, Bao Y, Zhang S, Wang W, Pang Z (2023) Two-stage scheduling of integrated energy systems based on a two-step DCGAN-based scenario prediction approach. Front Energy Res 10:1012367

Xu C, Xu P, Dai Y, Su S, Zhang L, Zhang J, Bai Y, Gao T, Xie Q, Shang L et al (2025) A two-stage generative architecture for renewable scenario generation based on temporal scenario representation and diffusion models. Energies 18(5):1275

Yang C, Rodriguez M, Kim S-H, Anderson R (2023) Physics-informed score-based models for renewable energy forecasting. Nat Energy 8(4):285–294

Yang C, Meng J, Liu B, Wang Z, Wang K (2024) A water demand forecasting model based on generative adversarial networks and multivariate feature fusion. Water 16(12):1731

Yilmaz B (2023) A scenario framework for electricity grid using generative adversarial networks. Sustain Energy Grids Netw 36:101157

Yilmaz B, Korn R (2022) Synthetic demand data generation for individual electricity consumers: generative adversarial networks (GANs). Energy AI 9:100161

Yilmaz B, Laudagé C, Korn R, Desmettre S (2024) Electricity GANs: generative adversarial networks for electricity price scenario generation. Commodities 3(3):254–280

Zhai S, Cheng Y, Lu W, Zhang Z (2016) Deep structured energy based models for anomaly detection. Int Conf Mach Learn. PMLR, 1100–1109

Zhan J, Wang C, Wang H, Zhang F, Li Z (2024) Pathways to achieve carbon emission peak and carbon neutrality by 2060: a case study in the Beijing-Tianjin-Hebei region. China Renew Sustain Energy Rev 189:113955

Zhang L, Chen Z (2025) Opportunities of applying large language models in building energy sector. Renew Sustain Energy Rev 214:115558

Zhang Y, Kong L (2022) Photovoltaic power prediction based on hybrid modeling of neural network and stochastic differential equation. ISA Trans 128:181–206

Zhang Z, Wu M (2021) Locational marginal price forecasting using convolutional long-short term memory-based generative adversarial network. In: 2021 IEEE power & energy society general meeting (PESGM). IEEE, pp. 1–5

Zhang C, Yang T (2023) Anomaly detection for wind turbines using long short-term memory-based variational autoencoder Wasserstein generation adversarial network under semi-supervised training. Energies 16(19):7008

Zhang C, Hu D, Yang T (2022) Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost. Reliab Eng Syst Saf 222:108445

Zhang Y, Schlueter A, Waibel C (2023a) Solargan: Synthetic annual solar irradiance time series on urban building facades via deep generative networks. Energy AI 12:100223

Zhang J, Zhao Z, Yan J, Cheng P (2023b) Ultra-short-term wind power forecasting based on CGAN-CNN-LSTM model supported by lidar. Sensors 23(9):4369

Zhang W, Shi J, Wang J, Jiang Y (2024a) Dynamic economic dispatch of integrated energy system based on generative adversarial imitation learning. Energy Rep 11:5733–5743

Zhang J, Huang J, Jin S, Lu S (2024b) Vision-language models for vision tasks: a survey. IEEE Trans Pattern Anal Mach Intell 46(8):5625–5644. https://doi.org/10.1109/TPAMI.2024.3369699

Zhang J, Zhang A, Lv J, Liu D, Sun X, Wu B (2024c) Research on a three-dimensional reconstruction algorithm for power grid tower poles based on a foreground-excluded neural radiance field. Sens Mater 36(12):5491–5505. https://doi.org/10.18494/SAM5348

Zhang X, Glaws A, Cortiella A, Emami P, King RN (2025) Deep generative models in energy system applications: review, challenges, and future directions. Appl Energy 380:125059

Zhang C, Zhang J, Lu J, Zhao Y (2026) Large language models meet energy systems: opportunities, challenges, and future perspectives. Appl Energy 403:127076. https://doi.org/10.1016/j.apenergy.2025.127076

Zhao L, Li Q, Ding G (2024) An intelligent web-based energy management system for distributed energy resources integration and optimization. J Web Eng 23(1):165–195

Zhao W, Shao Z, Yang S, Lu X (2025) A novel conditional diffusion model for joint source-load scenario generation considering both diversity and controllability. Appl Energy 377:124555

Authors and Affiliations

Umit Cali[1,10,11] · Ugur Halden[4] · Merlinda Andoni[3] · Ferhat Ozgur Catak[5] · Si Chen[3] · Benoit Couraud[3] · Emre Kantar[6] · Samuel Knapper[1] · Ibrahim Kucukdemiral[7] · Huseyin Kusetogullari[8] · Murat Kuzlu[9] · Yashar Mousavi[7] · Sonam Norbu[3] · Taha Selim Ustun[2] · David Flynn[3]

✉ Umit Cali
umit.cali@ntnu.no

Ugur Halden
ugur.halden@digitalgreen.energy

Merlinda Andoni
merlinda.andoni@glasgow.ac.uk

Ferhat Ozgur Catak
f.ozgur.catak@uis.no

Si Chen
si.chen@glasgow.ac.uk

Benoit Couraud
benoit.couraud@glasgow.ac.uk

Emre Kantar
emre.kantar@sintef.no

Samuel Knapper
samuel.knapper@york.ac.uk

Ibrahim Kucukdemiral
ibrahim.kucukdemiral@gcu.ac.uk

Huseyin Kusetogullari
huseyin.kusetogullari@bth.se

Murat Kuzlu
mkuzlu@odu.edu

Yashar Mousavi
seyedyashar.mousavi@gcu.ac.uk

Sonam Norbu
sonam.norbu@glasgow.ac.uk

Taha Selim Ustun
taha.ustun@fukushima-u.ac.jp

David Flynn
david.flynn@glasgow.ac.uk

[1] School of Physics, Engineering and Technology, University of York, Heslington, York YO10 5DD, UK

[2] Fukushima Renewable Energy Institute, AIST, 2-2-9 Machiikedai, Koriyama 963-0298, Japan

[3] Smart System Group, Autonomous Systems and Connectivity, James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK

[4] Digital Green Solutions Oü, Skala tn 7-2, 10141 Tallinn, Estonia

[5] Department of Computer Science, University of Stavanger, Kjell Arholms gate 41, 4021 Stavanger, Norway

[6] SINTEF Energy Research, Sem Sælands vei 11, 7034 Trondheim, Norway

[7] Department of Engineering, School of Science and Engineering, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, UK

[8] Department of Computer Science, Blekinge Institute of Technology, 37141 Karlskrona, Sweden

[9] Batten College of Engineering and Technology, Old Dominion University, 5115 Hampton Blvd, Norfolk 23529, VA, USA

[10] Department of Electric Energy, Norwegian University of Science and Technology, 7491 Trondheim, Norway

[11] Artificial Intelligence, Smart Infrastructure and Robotics Research Center, American University of Sharjah, Sharjah 26666, United Arab Emirates