Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Toward Resilient Explainable AI–Digital Twin Frameworks for Cyber-Secure Renewable Energy Systems: An Uncertainty-Aware Lifecycle-Integrated Architecture

MD NASIR Uddin

gccf.int@gmail.com

Multimedia University - Cyberjaya Campus https://orcid.org/0009-0000-7611-8150

Research Article

Keywords: Digital Twin; Explainable Artificial Intelligence (XAI); Agentic AI; Renewable Energy Systems; Cyber-Physical Resilience; Distributed Energy Resources (DER); Uncertainty Quantification; Multi-Agent Systems

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11659132/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Title: Toward Resilient Explainable AI–Digital Twin Frameworks for Cyber-Secure Renewable Energy Systems: An Uncertainty-Aware Lifecycle-Integrated Architecture

Abstract

The accelerating penetration of variable renewable energy sources into modern power systems demands digital infrastructures capable of supporting predictive, autonomous, and trustworthy grid operation. While digital twins (DT) and artificial intelligence (AI)—particularly multi-agent and agentic architectures—are key enablers, their integration remains fragmented across three critical dimensions: explainability, cybersecurity, and uncertainty management.

This paper proposes a conceptual, uncertainty-aware, lifecycle-integrated architecture that couples explainable AI (XAI) with digital twin modeling to support resilient, cyber-secure renewable energy operations. Based on a structured review spanning component-, building-, and urban-scale DT infrastructure, XAI-driven forecasting and fault diagnosis, distributed energy resource (DER) multi-agent control, and DER cybersecurity vulnerabilities, we identify a persistent scholarly disjunction. Few studies treat interpretability, adversarial resilience, and stochastic uncertainty as co-design requirements across the entire asset lifecycle.

Using a design-science methodology and thematic synthesis, we develop a five-layer architecture: sensing and data acquisition, digital twin simulation, XAI-mediated agentic decision-making, cybersecurity and resilience, and lifecycle governance. The core contribution lies in defining the structural logic through which these layers interoperate to achieve transparency, security, and adaptability. This framework provides a foundation for future simulation-based validation and advances debates on digitalized energy infrastructure governance.

Keywords: Digital Twin; Explainable Artificial Intelligence (XAI); Agentic AI; Renewable Energy Systems; Cyber-Physical Resilience; Distributed Energy Resources (DER); Uncertainty Quantification; Multi-Agent Systems

1. Introduction

1.1 Background and Context

Decarbonization commitments have driven a structural shift in electricity systems away from centralized, dispatchable generation toward distributed, variable, and weather-dependent renewable resources. This transition has been accompanied by a parallel digitalization of energy infrastructure, in which digital twin technology, understood as a virtual, continuously updated representation of a physical asset, has emerged as a central enabling technology (Aghazadeh Ardebili et al., 2024; Mansour et al., 2023). Digital twins have been applied across multiple spatial scales, from individual components such as batteries (Zhao et al., 2025) to entire buildings (Kaewunruen et al., 2018) and urban energy districts (Komninos, 2022; Mazzetto, 2024), and are

increasingly coupled with artificial intelligence to support forecasting, anomaly detection, and autonomous control (Wang et al., 2024).

Concurrently, the opacity of many high-performing AI models has motivated a substantial body of explainable AI (XAI) research within the energy domain, addressing applications ranging from wind-turbine condition monitoring (Astolfi et al., 2023) to grid-frequency dynamics (Drewnick et al., 2025) and building energy forecasting (Moon et al., 2024). A related and increasingly prominent thread concerns multi-agent systems (MAS) and, more recently, agentic AI architectures capable of autonomous, distributed decision-making across networks of distributed energy resources (González-Briones et al., 2018; Gill et al., 2026; Kiasari & Aly, 2026).

1.2 Problem Statement

Despite the individual maturity of these research streams, the simultaneous achievement of interpretability, cybersecurity, and uncertainty-aware operation across the full lifecycle of a renewable energy asset has received comparatively little structured attention. Three coupled challenges motivate this study. First, as digital twins assume an increasingly agentic character in DER coordination (Hasan & Nguyen, 2026; He & Jazizadeh, 2025), the interpretability of their decisions becomes a precondition for operator trust and regulatory acceptance, particularly where automated actions affect grid stability (Ridley, 2025; Q. V. Liao & Varshney, 2022). Second, the proliferation of networked DER and inverter-based resources materially expands the cyber-physical attack surface of the grid, a vulnerability long recognized in the DER cybersecurity literature (Qi et al., 2016; Sundararajan et al., 2018) yet rarely addressed jointly with digital twin or explainability design. Third, renewable generation is inherently stochastic, motivating uncertainty-aware forecasting and sensitivity-analysis approaches (Belik & Rubanenko, 2025; Tavoosi et al., 2021), a dimension that existing XAI–DT integration efforts have generally treated as secondary rather than constitutive.

1.3 Research Gap and Contribution

A reading of recent systematic reviews suggests that the digital twin literature (Aghazadeh Ardebili et al., 2024; Kim et al., 2025), the XAI-for-energy literature (Alsaigh et al., 2023; Machlev et al., 2022; Shadi et al., 2025), and the multi-agent and agentic AI literature (Izmirlioglu et al., 2024; Hosseini & Seilani, 2025) have developed largely in parallel, each surveying its own domain with limited cross-referencing of the others' design requirements. Where integration is attempted, it tends to address two of the three concerns rather than all three jointly, for instance explainable agentic control without an explicit cybersecurity model (Kiasari & Aly, 2026), or digital twin–agentic AI coupling without a corresponding interpretability layer (Hasan & Nguyen, 2026). This paper addresses that gap by proposing a conceptual architecture that explicitly co-locates explainability, cyber-resilience, and uncertainty quantification within a single lifecycle-integrated framework, with the aim of advancing the discussion beyond an additive combination of existing components toward a structurally coherent design logic for trustworthy, autonomous renewable energy management.

1.4 Structure of the Paper

The remainder of the paper proceeds as follows. Section 2 reviews the literature across four thematic domains, namely digital twin architectures, explainable AI in energy systems, multi-agent and agentic AI for DER management, and cybersecurity and governance, before synthesizing the cross-cutting research gap that motivates the proposed framework. Section 3 details the design-science-oriented methodology and systematic thematic synthesis process underlying the architecture's development.

2. Literature Review

2.1 Digital Twin Architectures for Renewable Energy and Urban-Scale Systems

Digital twin technology has been extensively reviewed as an enabling layer for the digitalization of energy systems. Aghazadeh Ardebili et al. (2024) conducted a systematic literature review identifying the enablers, design considerations, management practices, and computational challenges associated with digital twins of smart energy systems, providing one of the more comprehensive domain-level syntheses available. Complementary reviews have focused on narrower technological scopes: Li and He (2025) examined digital twin concepts for key components of renewable energy systems, while Kim et al. (2025) reviewed the applications, integrated technologies, and key challenges specific to digital twins for clean energy systems. At the component level, empirical implementations include Darbali-Zamora et al. (2021), who developed a real-time digital twin for state-estimation-based DER optimization in telemetry-sparse distribution networks; Belik and Rubanenko (2025), who conducted a sensitivity analysis of a digital twin model for an energy-community photovoltaic system; and Bayer and Pruckner (2023), who constructed a digital twin of a local energy system grounded in real smart-meter data. At a more granular asset level, Zhao et al. (2025) proposed a hierarchical, self-evolving digital twin method for multi-faceted battery situation awareness, illustrating the extension of digital twin logic toward adaptive, self-updating representations.

A parallel body of work situates digital twins within buildings and urban-scale energy districts. Kaewunruen et al. (2018) evaluated net-zero energy buildings through digital-twin modeling, while Agostinelli et al. (2021, 2022) examined cyber-physical systems combining digital twins and AI for building energy management and for a zero-energy port area in Italy. At the district and city scale, Savage et al. (2022) introduced a "universal digital twin" to examine the social-equity implications of heat-pump deployment; Lesnyak et al. (2023) applied digital twin concepts across building, campus, neighborhood, and urban scales in the context of heat transition; and Poshnath et al. (2025) proposed a policy-integrated digital twin approach for equitable renewable energy allocation in multi-owned buildings. Related contributions include Bocullo et al. (2023) on city-block renovation using renewable energy technologies, Villani et al. (2025) on urban sustainability and resilience in Venice, Malakhatka et al. (2025) on backcasting for positive energy districts, and Omar (2025) on digital twins for climate-responsive urban development integrating zero-energy buildings into smart-city strategies. Several reviews synthesize this urban digital twin literature

more broadly, including Mazzetto (2024), Ferré-Bigorra et al. (2022), Deng et al. (2021) on digital twin cities as a pattern of urban governance, Komninos (2022) on connected intelligence for carbon-neutral districts, and Omrany et al. (2025) on pathways toward resilient and sustainable cities through urban digital twins. Mansour et al. (2023) further surveyed the integration of IoT and digital twin technologies across electrical power systems more broadly, and Nneoma Ugwu et al. (2026) examined AI-driven digital twins for the real-time optimization of renewable energy grids, a contribution closely aligned with the operational concerns motivating the present framework.

Across this literature, digital twin design is treated predominantly as a modeling and data-integration problem: emphasis falls on simulation fidelity, sensor integration, and computational architecture, with comparatively limited attention to how the resulting models support transparent, explainable, or cyber-resilient decision-making once deployed operationally.

2.2 Explainable AI in Energy and Power Systems

The broader XAI literature has established conceptual foundations, surveying algorithmic approaches (Das & Rad, 2020), articulating human-centered design principles (Q. V. Liao & Varshney, 2022; Ehsan et al., 2021, 2022, 2023), and synthesizing evaluation methodologies (Kim et al., 2024; Ridley, 2025), that have since been adapted to the energy domain. Several reviews map this domain-specific uptake: Machlev et al. (2022) reviewed XAI techniques for energy and power systems, identifying recurring challenges and opportunities; Alsaigh et al. (2023) examined AI explainability and governance considerations specific to smart energy systems; and Shadi et al. (2025) reviewed XAI applications in energy-systems maintenance.

Applied studies span forecasting, fault diagnosis, and policy-relevant interpretation. In wind energy, Astolfi et al. (2023) applied XAI techniques to wind-turbine condition monitoring; W. Liao et al. (2024) interrogated the trustworthiness of XAI in wind power forecasting; Yassen et al. (2025) developed an explainable, LSTM-based wind power forecasting model; and Rengasamy and R (2025) proposed a hybrid random forest–XGBoost explainable framework for wind-turbine fault detection. In solar applications, Sriniva et al. (2024) applied XAI to energy prediction and anomaly detection in solar systems, and Garí-Galíndez et al. (2025) developed an XAI-driven methodology for accelerating research on thin-film kesterite photovoltaic technology. Building- and demand-side applications include Maarif et al. (2023), who combined LSTM forecasting with XAI for energy-usage prediction; Moon et al. (2024), who applied explainable ensemble learning to residential electricity-consumption forecasting; Iram et al. (2025), who used XAI to analyze weather-driven energy efficiency in residential buildings; and Konhäuser and Werner (2024), who used XAI to uncover the financial impact of energy-efficient building characteristics.

At the grid-operations level, Drewnick et al. (2025) applied XAI to deterministic and stochastic influences on power-grid frequency dynamics, Kruse et al. (2022) used XAI to analyze secondary control activation, and Pütz et al. (2023) examined regulatory changes in the German and Austrian power systems through an explainability lens. Several studies extend XAI toward policy and financial decision contexts, including Çankal and Ever (2025) on renewable energy consumption and financial performance, Adıgüzel Tüylü (2026) on interpreting primary energy-consumption

patterns across Europe and Türkiye, and Panagoulias et al. (2023), who proposed a tailored-explainability methodology for energy-management decision support. Notably, Perr-Sauer et al. (2025) offered a broad survey of XAI applications across renewable energy research, underscoring both the breadth of uptake and the continued absence of standardized design guidance linking explainability to operational deployment contexts such as digital twins.

Taken together, this literature demonstrates that XAI techniques have been productively adapted across nearly every renewable energy subdomain, yet predominantly as a post hoc analytical layer applied to an already-trained model, rather than as a structural requirement embedded within the broader digital twin or control architecture within which that model operates.

2.3 Multi-Agent Systems and Agentic AI for Distributed Energy Resource Management

The coordination of distributed energy resources has long drawn on multi-agent systems (MAS) research. Foundational and review-level contributions include González-Briones et al. (2018) on MAS applications in energy optimization, Izmirlioglu et al. (2024) on MAS for smart grids, Ma et al. (2019) on the use of ontologies within energy-sector MAS, Yao et al. (2023) on agent-based methods for multi-energy systems, and Rogers et al. (2021) on the challenges facing autonomous agents and MAS research in smart-grid delivery. Applied microgrid implementations extend from early architectures such as the CERTS microgrid concept (Lasseter et al., 2002) and Colson et al.'s (2011) MAS-based microgrid power management, through game-theoretic decentralized coordination (Karavas et al., 2017) and an Internet-of-Energy-based optimal MAS control scheme incorporating renewable resources (Alhasnawi et al., 2021), to more recent decentralized hierarchical MAS frameworks for sustainable energy management (Dragomir & Dragomir, 2025) and resilient multi-microgrid coordination (Aghmadi & Mohammed, 2024).

Distributed energy resource management systems (DERMS) constitute a related and increasingly important architectural concern. Ding et al. (2022) proposed a federated architecture for secure and transactive DERMS (FAST-DERMS), Strezoski (2023) reviewed the state of the art in DERMS and outlined directions for advancement, and Ilic et al. (2020) examined interactive coordination protocols within DERMS. Reinforcement-learning-based control has emerged as a prominent extension of this literature: Touzani et al. (2021) applied deep reinforcement learning to DER control for load flexibility, Bousnina and Guerassimoff (2024) coupled deep reinforcement learning with a digital twin case study for smart energy management, Zhang et al. (2023) developed a multi-agent deep-reinforcement-learning strategy for safe DER scheduling in energy hubs, and Y. Zhao et al. (2024) proposed safety-constrained multi-agent reinforcement learning for power-quality control in distributed renewable energy networks.

Most recently, the literature has begun to articulate a distinct “agentic AI” paradigm, characterized by greater autonomy, goal-directed reasoning, and adaptability relative to conventional MAS. Gill et al. (2026) outlined the vision and challenges of agentic AI as an emerging field, Hosseini and Seilani (2025) conducted a systematic review of its role in shaping a smart future, and Tiwari (2025) theorized the conceptual transition from automation to genuine machine agency in urban AI

contexts. Within the energy domain specifically, Kiasari and Aly (2026) reviewed agentic AI frameworks for smart grids with explicit attention to autonomy, safety, and explainability, while Hasan and Nguyen (2026) examined the integration of agentic AI with digital twins for intelligent decision-making, a contribution directly relevant to, though narrower in scope than, the architecture proposed here. Related applications include He and Jazizadeh (2025) on context-aware, LLM-based agents for human-centered building energy management, Nong (2026) on agentic AI for autonomous cascade-failure detection and response in critical energy infrastructure, and Fernández-Miguel et al. (2025) on agentic AI for human-centric predictive maintenance in smart manufacturing.

While this literature increasingly acknowledges explainability as a desirable property of agentic control (Kiasari & Aly, 2026), the treatment of cybersecurity resilience and forecast uncertainty as co-equal, architecturally embedded constraints, rather than supplementary considerations, remains underdeveloped.

2.4 Cybersecurity, Resilience, and Governance in DER Ecosystems

The expansion of networked, inverter-based distributed energy resources has been recognized as a structural cybersecurity concern since early work by Qi et al. (2016) on cybersecurity for DER and smart inverters, and Sundararajan et al. (2018), who surveyed protocol-level vulnerabilities and mitigation strategies for DER cyber-physical security. Subsequent work has addressed access-control architecture, with Fragkos et al. (2022) comparing centralized and decentralized implementation considerations for DER access control, and aggregation-level governance, with Obi et al. (2020) reviewing standards and literature on customer-owned DER aggregation, and Cook et al. (2018) drawing lessons from utility-led DER aggregation programs in the United States. Valova and Brown (2022) further examined the technical and regulatory challenges of DER interconnection, while Kumar et al. (2020) situated cybersecurity considerations within the broader convergence of AI, IoT, and blockchain technologies in smart grids.

Notably, this cybersecurity literature has developed with limited direct engagement with either the digital twin or XAI streams reviewed above, despite digital twins constituting, by design, a high-fidelity digital representation of physical grid assets whose compromise would carry direct operational consequences, a structural vulnerability rarely addressed within the digital twin literature itself.

2.5 Synthesis and Identified Research Gap

Read together, four observations emerge. First, digital twin research has matured considerably in modeling fidelity and spatial scalability, extending from individual components to entire urban districts, but treats explainability and cybersecurity largely as external considerations rather than design-intrinsic requirements. Second, the XAI-for-energy literature has demonstrated broad applicability across forecasting, fault diagnosis, and policy contexts, yet remains predominantly model-centric, applied post hoc to a given algorithm rather than embedded within the operational architecture of a digital twin or control system. Third, multi-agent and agentic AI research has

progressed toward greater autonomy in DER coordination and has begun to acknowledge explainability as a design goal (Kiasari & Aly, 2026), but rarely integrates this with formal cybersecurity threat modeling or explicit uncertainty quantification. Fourth, the DER cybersecurity literature, while technically mature with respect to protocol- and access-control-level vulnerabilities, has developed largely independently of the digital twin and XAI streams.

This fragmentation constitutes the specific research gap addressed in this paper: no reviewed contribution articulates a single architecture in which explainability, cybersecurity resilience, and uncertainty quantification are treated as jointly binding constraints across the full lifecycle of a renewable energy digital twin, from design and commissioning through operation, maintenance, and eventual decommissioning. The framework developed in this paper is intended to occupy that conceptual space, rather than to incrementally extend any one of the four reviewed streams in isolation.

3. Research Methodology

3.1 Research Design and Philosophical Stance

This study adopts a conceptual, qualitative research design situated within the design science research (DSR) paradigm, in which the principal output is not an empirically tested hypothesis but a purposefully constructed artifact, here, a lifecycle-integrated architectural framework, intended to address a well-specified, practice-relevant problem (Section 1.3). This stance is appropriate where, as established in Section 2, the relevant phenomenon (the joint treatment of explainability, cybersecurity, and uncertainty in renewable energy digital twins) is not yet adequately theorized in the existing literature, which precludes a purely deductive, hypothesis-testing approach. The design-science orientation is combined with a systematic literature review (SLR) component, used to ground the proposed architecture in verified prior work rather than ad hoc specification.

3.2 Literature Search Strategy

The review underlying this paper was conducted across four thematic clusters corresponding to the structure of Section 2: (i) digital twin technologies for energy and urban systems; (ii) explainable artificial intelligence applied to power and energy systems; (iii) multi-agent systems and agentic AI for distributed energy resource management; and (iv) cybersecurity and governance of DER and smart-inverter infrastructure. Search terms were constructed by combining core concept keywords, including “digital twin,” “explainable AI,” “XAI,” “agentic AI,” “multi-agent system,” “distributed energy resource,” “cybersecurity,” “renewable energy,” and “smart grid,” using Boolean operators applied independently within each thematic cluster before cross-referencing results for thematic overlap. The resulting corpus spans publications from 2011 to 2026, reflecting both the foundational MAS and microgrid literature and the most recent agentic AI and XAI contributions, and is dominated by peer-reviewed journal articles, with a smaller proportion of conference proceedings, technical reports, and preprints retained where they represented otherwise underrepresented or emergent perspectives, agentic AI architectures being a notable example of a sub-literature that remains predominantly preprint- and proceedings-based at the time of writing.

3.3 Inclusion and Exclusion Criteria

Sources were retained where they (a) addressed at least one of the four thematic clusters in the context of energy, power, or closely related urban infrastructure systems; (b) were published in English; and (c) provided sufficient methodological or conceptual detail to support thematic synthesis, namely peer-reviewed articles, systematic reviews, technical reports from recognized research institutions, or clearly described preprints. Sources addressing multi-agent systems, digital twins, or explainable AI in domains unrelated to energy or critical infrastructure were excluded unless they contributed a methodological construct of direct relevance to the proposed architecture, as in the case of human-centered XAI principles drawn from clinical decision-support contexts (Panigutti et al., 2023). Where multiple reviews covered substantially overlapping ground, the most recent and most comprehensive review was prioritized for synthesis, with earlier or narrower reviews retained as supporting evidence.

3.4 Framework Development Approach

The proposed architecture was developed through an iterative thematic synthesis process rather than a single-pass coding exercise. Each retained source was first coded according to the lifecycle stage(s) of a renewable energy asset it addressed (design and planning, commissioning, operation, maintenance, decommissioning), the architectural layer to which it primarily contributed (data and sensing, simulation and modeling, decision and control, security, or governance), and whether it engaged explicitly with explainability, cybersecurity, or uncertainty as a stated design concern. This coding exercise made visible the cross-cutting gap formalized in Section 2.5: the great majority of sources contributed substantively to one or two architectural layers but rarely engaged more than two of the three cross-cutting concerns simultaneously. The resulting five-layer architecture, comprising sensing and data acquisition, digital twin simulation, XAI-mediated agentic decision-making, cybersecurity and resilience, and lifecycle governance, was constructed to formalize the points of interdependency between layers that the reviewed literature treats discretely, drawing in particular on the DERMS literature (Ding et al., 2022; Strezoski, 2023) for the data and control layers, the XAI-for-energy literature (Machlev et al., 2022; Perr-Sauer et al., 2025) for the decision layer's interpretability requirements, the DER cybersecurity literature (Qi et al., 2016; Sundararajan et al., 2018) for the resilience layer, and the urban and lifecycle-oriented digital twin literature (Lesnyak et al., 2023; Omrany et al., 2025) for the governance layer.

3.5 Validity, Reliability, and Limitations

As a conceptual contribution, this study does not generate primary empirical data, and the proposed architecture has not yet been validated through simulation, prototype deployment, or case-based application. This limitation is consistent with the design-science paradigm adopted, in which conceptual artifact development typically precedes empirical evaluation in subsequent research cycles. To mitigate subjectivity in the thematic coding process, source classification was cross-checked iteratively against each publication's stated objectives and conclusions; where ambiguity arose regarding a source's primary thematic contribution, the source was coded conservatively, that is, assigned to the layer or concern most explicitly stated in its title and abstract rather than inferred

from secondary implications. Future research should subject the proposed architecture to empirical validation through simulation-based testbeds or pilot implementation within an operational renewable energy distribution network, an avenue the authors intend to pursue in subsequent work.

4. Research Gap Analysis

The body of literature surrounding explainable artificial intelligence (XAI), digital twins (DT), multi-agent/agentic AI, and distributed energy resource (DER) cybersecurity has matured largely as four parallel rather than convergent streams, and the boundary conditions of each stream expose where the present study positions its contribution.

The first stream treats XAI as an analytical add-on to forecasting and fault-diagnosis pipelines in renewable energy assets — wind turbine condition monitoring (Astolfi et al., 2023; Rengasamy & R, 2025), wind power forecasting (Liao et al., 2024; Yassen et al., 2025), photovoltaic suitability and materials research (Shen et al., 2026; Garí-Galíndez et al., 2025), and building- or grid-level consumption forecasting (Maarif et al., 2023; Moon et al., 2024; Akyol et al., 2023; Konhäuser & Werner, 2024). Review-level syntheses (Machlev et al., 2022; Perr-Sauer et al., 2025; Alsaigh et al., 2023; Shadi et al., 2025) converge on a consistent observation: explainability is almost universally applied *post hoc*, as an interpretive layer attached after model training, rather than as a property designed into the sensing-to-decision pipeline. None of these works embeds explainability within a synchronized virtual representation of the physical asset across its operational lifecycle.

The second stream — digital twins of renewable and distributed energy systems — is comparatively mature on the side of state estimation, synchronization, and sensitivity analysis (Darbali-Zamora et al., 2021; Bayer & Pruckner, 2023; Belik & Rubanenko, 2025; Zhao et al., 2025), and on lifecycle-oriented conceptualization at the building, district, and urban scale (Aghazadeh Ardebili et al., 2024; Kim et al., 2025; Iossa et al., 2025; Mansour et al., 2023; Li & He, 2025). What this literature does not resolve is *interpretability of the twin's own reasoning*: sensitivity analyses quantify which parameters drive output variance (Belik & Rubanenko, 2025), but they do not produce stakeholder-facing explanations of why a twin recommends a given control or maintenance action, nor do they treat explainability as a lifecycle-stage-dependent requirement that differs between design, commissioning, operation, and decommissioning (cf. Executive Vice President & Imam, 2025, on utility infrastructure DT opportunities and challenges).

The third stream, multi-agent systems and the emerging "agentic AI" paradigm for grid and microgrid management (González-Briones et al., 2018; Rogers et al., 2021; Izmirlioglu et al., 2024; Hosseini & Seilani, 2025; Kiasari & Aly, 2026; Hasan & Nguyen, 2026), is where autonomy and decentralized decision-making are theorized, including explicit calls for "autonomous, safe, and explainable control frameworks" (Kiasari & Aly, 2026). However, this literature largely theorizes agentic decision logic independently of a calibrated, uncertainty-quantified digital twin substrate; explainability and digital twin fidelity are referenced as desiderata rather than architecturally integrated.

The fourth stream — DER and smart-inverter cybersecurity (Qi et al., 2016; Sundararajan et al., 2018; Fragkos et al., 2022; Strezoski, 2023; Ding et al., 2022) — addresses protocol-level and

access-control vulnerabilities in distributed energy resource communication but, as a body of work, is not cross-referenced against either the explainability or the digital-twin literatures reviewed above. Anomaly or intrusion detection in this stream is not coupled to an explanatory layer that would allow an operator to interrogate *why* a flagged event was classified as a cyber-physical anomaly rather than a benign operational disturbance.

The convergent gap, therefore, is not the absence of any one capability but the absence of an **architecture that treats explainability, digital-twin fidelity, cybersecurity-aware anomaly reasoning, and decision uncertainty as co-dependent properties distributed across the digital twin lifecycle**, rather than as four separately optimized subsystems. Resilience-oriented work that approaches this integration from an operational-continuity angle (Aghmadi & Mohammed, 2024; Feng et al., 2025, on self-healing fault diagnosis–isolation–recovery; Dragomir & Dragomir, 2025) gestures toward the requirement but does not formalize it as an explainability-by-design, security-by-design architecture for renewable energy systems specifically.

5. Research Questions and Objectives

Derived directly from the gap articulated above, the study is organized around four research questions, each paired with a corresponding objective.

RQ1. How can explainability be structurally embedded across the successive lifecycle stages of a digital twin of a renewable energy system, rather than applied as a post-hoc interpretive layer to a single model output? *Objective 1:* To identify, from the lifecycle-oriented digital twin literature (Aghazadeh Ardebili et al., 2024; Iossa et al., 2025; Mansour et al., 2023), the stage-specific explainability requirements that differ between design/planning, commissioning, operation, maintenance, and decommissioning/repowering.

RQ2. What architectural mechanisms allow cybersecurity-aware anomaly reasoning to be coupled with explainable decision support within a digital twin of a renewable energy asset or distributed energy resource network? *Objective 2:* To synthesize DER cybersecurity findings (Qi et al., 2016; Sundararajan et al., 2018; Fragkos et al., 2022) with XAI techniques reported for energy-domain anomaly and fault detection (Rengasamy & R, 2025; Sriniva et al., 2024) into a shared architectural layer.

RQ3. How can uncertainty quantification be coupled with explainability to support resilient, trustworthy decision-making by autonomous (multi-agent/agentic) controllers operating on a digital twin of a renewable energy system? *Objective 3:* To formalize, drawing on sensitivity-analysis and variability-aware forecasting studies (Belik & Rubanenko, 2025; Mathur et al., 2024; Fose et al., 2024; Tavoosi et al., 2021), the points in the agentic decision loop at which uncertainty estimates must be made explicit and explainable rather than absorbed silently into point predictions.

RQ4. What unifying architectural framework can integrate explainable AI, digital twin synchronization, multi-agent/agentic decision logic, and cybersecurity principles into a coherent, lifecycle-integrated, uncertainty-aware reference architecture for renewable energy systems? *Objective 4:* To construct and present such a reference architecture, specifying its layers, information flows, and the lifecycle stage at which each capability is activated.

These four objectives are cumulative: Objectives 1–3 each resolve one axis of the gap identified in Section 4, and Objective 4 integrates them into the architecture developed in Section 7.

6. Theoretical Foundations

The architecture developed in this study is not derived from a single theory but from the deliberate combination of four theoretical positions that, in the reviewed literature, have so far been advanced independently.

Digital twin as a lifecycle-synchronized representation. Across the systematic reviews consulted (Aghazadeh Ardebili et al., 2024; Mansour et al., 2023; Kim et al., 2025), the digital twin is consistently theorized not as a static simulation but as a representation that remains synchronized with its physical counterpart across the asset's lifecycle — design, deployment, operation, maintenance, and end-of-life. This study adopts that lifecycle premise as the organizing spine of the architecture, rather than treating the twin as a single-stage modeling exercise.

Explainability as a human-centered, audience-dependent property. The human-centered XAI (HCXAI) line of theorizing (Ehsan et al., 2021, 2022, 2023; Liao & Varshney, 2022; Ridley, 2025) frames explainability not as an intrinsic algorithmic feature but as a property that must be evaluated against the needs, expertise, and context of the human (or agent) consuming the explanation — a position reinforced by evaluation-focused reviews (Kim et al., 2024) and by co-design work in adjacent high-stakes domains (Panigutti et al., 2023). This study adopts the HCXAI position that explanation requirements differ by *who* is making a decision and *at which lifecycle stage*, which is the theoretical basis for treating explainability as lifecycle-dependent in Section 7 rather than as a single fixed module.

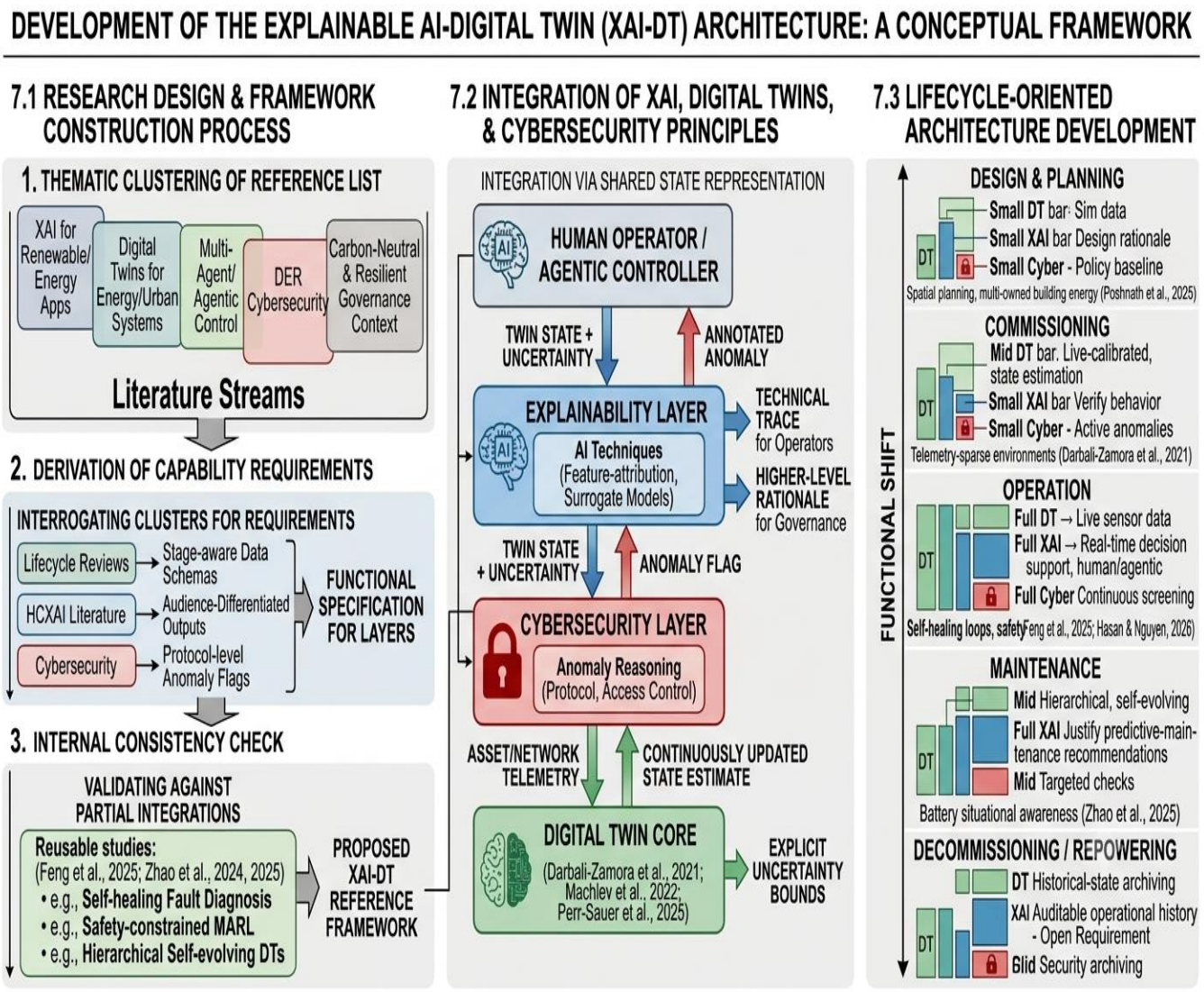
Cyber-physical security as an embedded, not bolted-on, system property. The DER cybersecurity literature (Qi et al., 2016; Sundararajan et al., 2018) theorizes vulnerability as arising at the interface between communication protocols and physical control actions in distributed energy resources, implying that security cannot be evaluated independently of the control logic it protects. This study extends that position by treating cybersecurity reasoning as a layer that must share representational state with the explainability and digital-twin layers, rather than as an isolated monitoring function.

Multi-agent and agentic decision-making as a decentralized control paradigm. The trajectory from classical multi-agent systems theory (González-Briones et al., 2018; Rogers et al., 2021; Farid, 2015) toward the more recent "agentic AI" framing (Hosseini & Seilani, 2025; Tiwari, 2025; Kiasari & Aly, 2026; Hasan & Nguyen, 2026) provides the decision-execution layer of the architecture: decentralized agents that act on the digital twin's state rather than a single centralized optimizer. Several of these sources explicitly call for safety- and explainability-constrained agentic control (Kiasari & Aly, 2026; Zhao et al., 2024, on safety-constrained multi-agent reinforcement learning), which this study treats as a theoretical requirement rather than an optional refinement.

Taken together, these four positions are combined — rather than each applied separately, as in the reviewed literature — into an integrative socio-technical resilience lens: a renewable energy digital twin is theorized here as a system whose *trustworthiness* depends jointly on lifecycle-stage-

appropriate explanation, security-aware state representation, and uncertainty-aware autonomous decision logic, none of which is sufficient on its own.

7. Development of the Explainable AI–Digital Twin Architecture



7.1 Research Design and Framework Construction Process

The architecture was constructed through a conceptual, literature-grounded design process rather than empirical data collection, consistent with the study's aim of producing a reference framework rather than testing a deployed system. The process proceeded in three stages.

First, the reference list compiled for this study was thematically clustered into the four streams identified in Section 4 — XAI for renewable/energy applications, digital twins for energy and urban systems, multi-agent/agentic control, and DER cybersecurity — together with a fifth cluster addressing carbon-neutral and resilient urban energy governance (Chen et al., 2022; Su & Urban,

2021; Lee & Jung, 2023), which supplies the policy and sustainability context within which the architecture must operate.

Second, each cluster was interrogated for the specific *capability requirement* it implies — for instance, the lifecycle reviews (Aghazadeh Ardebili et al., 2024) imply a requirement for stage-aware data schemas; the HCXAI literature (Ehsan et al., 2022) implies a requirement for audience-differentiated explanation outputs; the cybersecurity literature (Sundararajan et al., 2018) implies a requirement for protocol-level anomaly flags that can be passed upstream. These extracted requirements form the functional specification against which the architecture's layers, presented in Sections 7.2 and 7.3, were defined.

Third, the resulting layer specification was checked for internal consistency against the resilience-oriented studies that already combine two or more of the four streams in a narrower domain — self-healing power supply fault diagnosis and recovery (Feng et al., 2025), safety-constrained multi-agent reinforcement learning for power quality (Zhao et al., 2024), and hierarchical self-evolving digital twins for battery state awareness (Zhao et al., 2025) — to confirm that the proposed integration is an extension of, rather than a departure from, demonstrated partial integrations in the literature.

7.2 Integration of Explainable AI, Digital Twins, and Cybersecurity Principles

The architecture integrates the three technical pillars through a shared state representation rather than through sequential hand-offs between independent modules, which is the limitation identified in Section 4 for each pillar taken separately.

At the foundation, the **digital twin core** maintains a continuously updated, uncertainty-annotated state of the renewable energy asset or DER network, drawing on the synchronization and state-estimation methods reported for real-time digital twins (Darbali-Zamora et al., 2021) and on sensitivity-analysis methods for quantifying parameter-driven output variance (Belik & Rubanenko, 2025). Each state estimate carries an explicit uncertainty bound rather than being passed forward as a deterministic point value, addressing the gap noted in Section 4 between forecasting studies that handle variability (Mathur et al., 2024; Fose et al., 2024) and twin architectures that do not propagate that variability onward.

Layered on the twin core, the **explainability layer** does not operate on raw twin outputs in isolation but on the combined twin-state-plus-uncertainty representation, applying the categories of technique surveyed by Machlev et al. (2022) and Perr-Sauer et al. (2025) — feature-attribution and surrogate-model approaches of the kind reported for load and generation forecasting (Maarif et al., 2023; Akyol et al., 2023) and for fault/anomaly detection (Rengasamy & R, 2025; Sriniva et al., 2024). Consistent with the human-centered theoretical position adopted in Section 6, this layer generates differentiated explanation outputs: a technical trace for operators and a higher-level rationale for governance or oversight audiences, rather than a single undifferentiated output.

The **cybersecurity layer** receives the same twin-state representation and applies protocol- and access-level anomaly reasoning derived from the DER security literature (Qi et al., 2016;

Sundararajan et al., 2018; Fragkos et al., 2022), but — and this is the architectural contribution distinguishing it from that literature — every anomaly flag it raises is routed through the explainability layer before reaching a human operator or an agentic controller, so that a flagged event arrives annotated with the basis for the flag rather than as an unexplained binary alert.

7.3 Lifecycle-Oriented Architecture Development

Consistent with the lifecycle premise established in Section 6, the integrated layer stack described in 7.2 is not instantiated uniformly; its emphasis shifts across five lifecycle stages identified from the digital twin lifecycle literature (Aghazadeh Ardebili et al., 2024; Iossa et al., 2025; Executive Vice President & Imam, 2025). During **design and planning**, the digital twin core operates on design-stage simulation data rather than live telemetry, and the explainability layer is weighted toward design-rationale outputs — comparable in function to the spatial-planning digital twin approaches reported for multi-owned building energy allocation (Poshnath et al., 2025) — while the cybersecurity layer establishes baseline access-control policy rather than runtime anomaly detection.

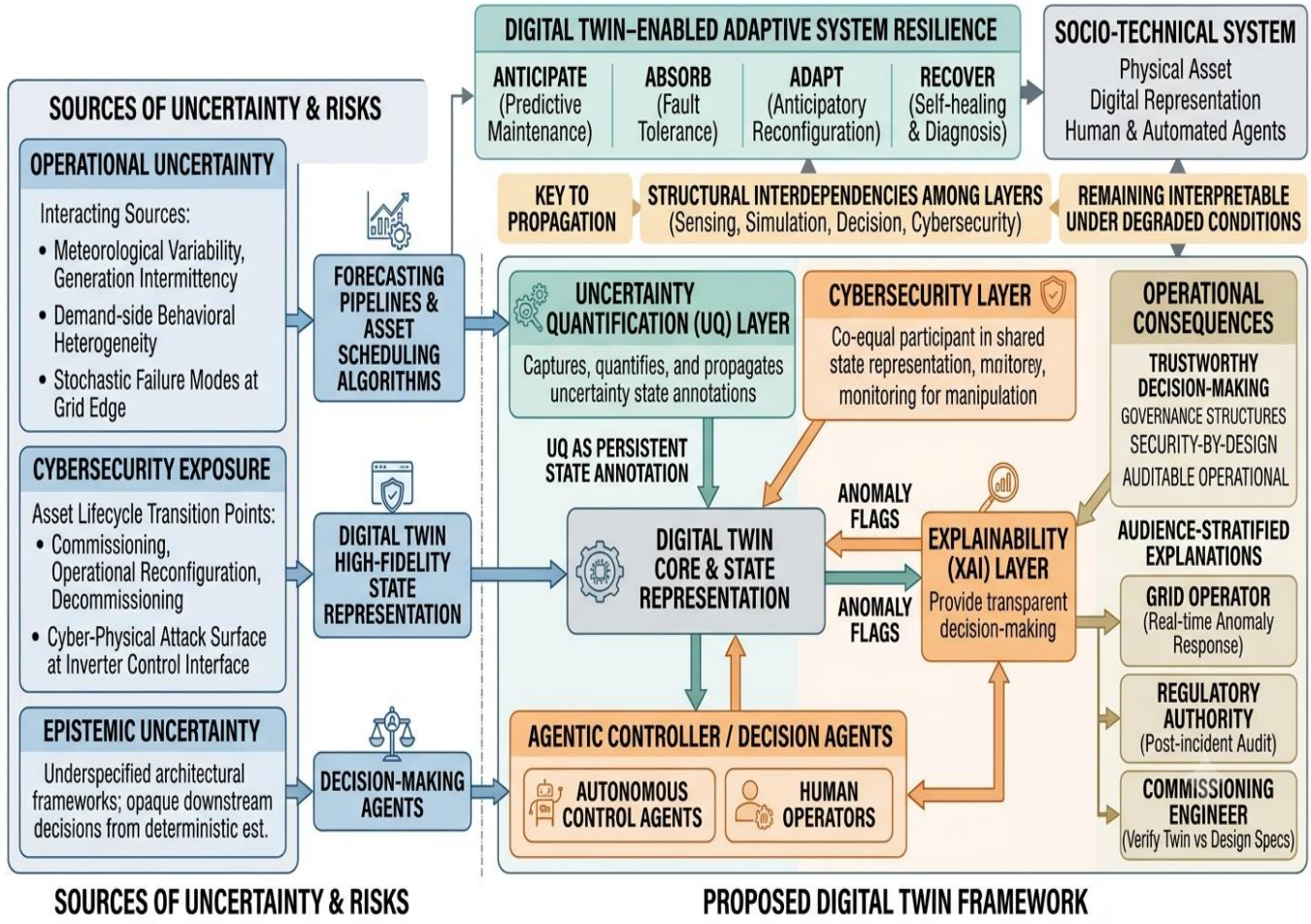
During **commissioning**, the twin transitions from simulated to live-calibrated state, drawing on the state-estimation approach used for telemetry-sparse environments (Darbali-Zamora et al., 2021); explainability outputs at this stage are oriented toward verifying that twin behavior matches commissioning-test expectations. During **operation**, all three layers run continuously: the twin core ingests live sensor data, the explainability layer supports real-time operator and agentic decision-making of the kind theorized for agentic AI–digital twin integration (Hasan & Nguyen, 2026; Kiasari & Aly, 2026), and the cybersecurity layer performs continuous anomaly screening, consistent with self-healing fault-diagnosis-and-recovery loops reported for subway power systems (Feng et al., 2025) and with safety-constrained multi-agent control for power quality (Zhao et al., 2024).

During **maintenance**, the architecture's emphasis shifts toward the hierarchical, self-evolving twin behavior reported for battery situational awareness (Zhao et al., 2025), in which the explainability layer's primary function becomes justifying predictive-maintenance recommendations rather than real-time control actions. During **decommissioning or repowering**, the twin core's role narrows to historical-state archiving, and the explainability layer is repurposed to produce an auditable account of the asset's operational history — a function not addressed in the reviewed digital-twin lifecycle literature, which this architecture identifies as an open requirement rather than one resolved by existing studies.

This stage-by-stage redistribution of emphasis across an otherwise constant layer stack is the architectural device through which the present study operationalizes the lifecycle-integration claim advanced in Section 4 as the principal gap in the reviewed literature.

8. Conceptualizing Resilient and Cyber-Secure Renewable Energy Systems

8.1 Managing Uncertainty in Renewable Energy Operations



The structural integration of variable renewable energy sources into modern power systems introduces a class of operational uncertainty that is not merely statistical noise but a constitutive feature of system behavior — one that propagates through forecasting pipelines, asset scheduling algorithms, and autonomous control loops in ways that existing digital twin architectures have systematically underspecified. Uncertainty in renewable energy operations arises from at least three interacting sources: meteorological variability driving generation intermittency, demand-side behavioral heterogeneity, and the stochastic failure modes of distributed hardware operating at the grid edge (Fose et al., 2024; Tavoosi et al., 2021). While each source has been addressed individually within subdomain literature — variability-aware generation forecasting (Mathur et al., 2024), demand-response uncertainty in distributed energy resource management systems (Strezoski, 2023), and component-level sensitivity analysis for photovoltaic digital twins (Belik & Rubanenko, 2025) — the cross-layer propagation of uncertainty from raw sensor inputs through decision outputs remains architecturally underspecified in integrated frameworks.

This lacuna carries direct operational consequences. A digital twin that absorbs forecast uncertainty into a deterministic point estimate before passing that estimate to an agentic controller effectively renders the downstream decision opaque with respect to the confidence bounds on which it rests

— a structural deficiency that human-centered explainability frameworks identify as incompatible with meaningful operator oversight (Liao & Varshney, 2022; Ridley, 2025). The present framework addresses this by treating uncertainty quantification not as a preprocessing step to be resolved before the twin's decision logic engages, but as a persistent, explicitly propagated state annotation carried through each architectural layer. This design choice draws on the theoretical position advanced in Section 6: a renewable energy digital twin cannot be deemed trustworthy if the confidence bounds governing its recommendations are invisible to the operators and governance structures responsible for its outputs.

8.2 Cybersecurity Challenges Across Renewable Energy Lifecycles

The cybersecurity exposure of renewable energy infrastructure is not uniform across the asset lifecycle but intensifies at specific transition points — commissioning, operational reconfiguration, and decommissioning — at which the communication interfaces between physical hardware and its digital representation are most susceptible to manipulation. Foundational work by Qi et al. (2016) and Sundararajan et al. (2018) established that the cyber-physical attack surface of distributed energy resources resides principally at the interface between communication protocols and physical inverter control actions, a vulnerability that grows non-linearly with the proliferation of networked DER nodes. More recent governance-oriented analyses have extended this concern to the access-control architecture governing aggregator-level DER coordination (Fragkos et al., 2022; Obi et al., 2020), yet neither stream has formally addressed how a digital twin's high-fidelity state representation — by design an accurate reflection of the physical grid — constitutes a high-value target whose compromise would carry direct operational consequences.

This is a structural omission with compounding implications. An adversary who successfully manipulates the state representation within a digital twin does not merely corrupt a simulation; they corrupt the evidentiary basis on which autonomous control agents act and on which human operators form situational judgments. The framework developed here responds to this threat model by treating the cybersecurity layer not as an intrusion-detection appendage but as a co-equal participant in the shared state representation maintained by the digital twin core — one whose anomaly flags are architecturally required to pass through the explainability layer before reaching any decision-making agent or human operator. This routing requirement, absent from the DER cybersecurity literature surveyed in Section 2.4, operationalizes security-by-design as a property of the twin's architecture rather than a supplementary monitoring function.

8.3 Explainable AI for Transparent and Trustworthy Decision-Making

The deployment of XAI techniques within renewable energy systems has proceeded rapidly across forecasting, fault diagnosis, and policy-relevant interpretation (Machlev et al., 2022; Perr-Sauer et al., 2025), yet a persistent structural limitation characterizes nearly all reported implementations: explainability is applied post hoc to an already-trained model whose operational embedding — within a digital twin, a control loop, or a lifecycle-management system — is treated as architecturally separate from the explanation task. This post-hoc orientation, well-documented in systematic reviews (Alsaigh et al., 2023; Shadi et al., 2025), means that explanation outputs are generated for individual model predictions in isolation from the broader state context — uncertainty

bounds, cybersecurity status, lifecycle stage — within which those predictions carry meaning for the humans or agents consuming them.

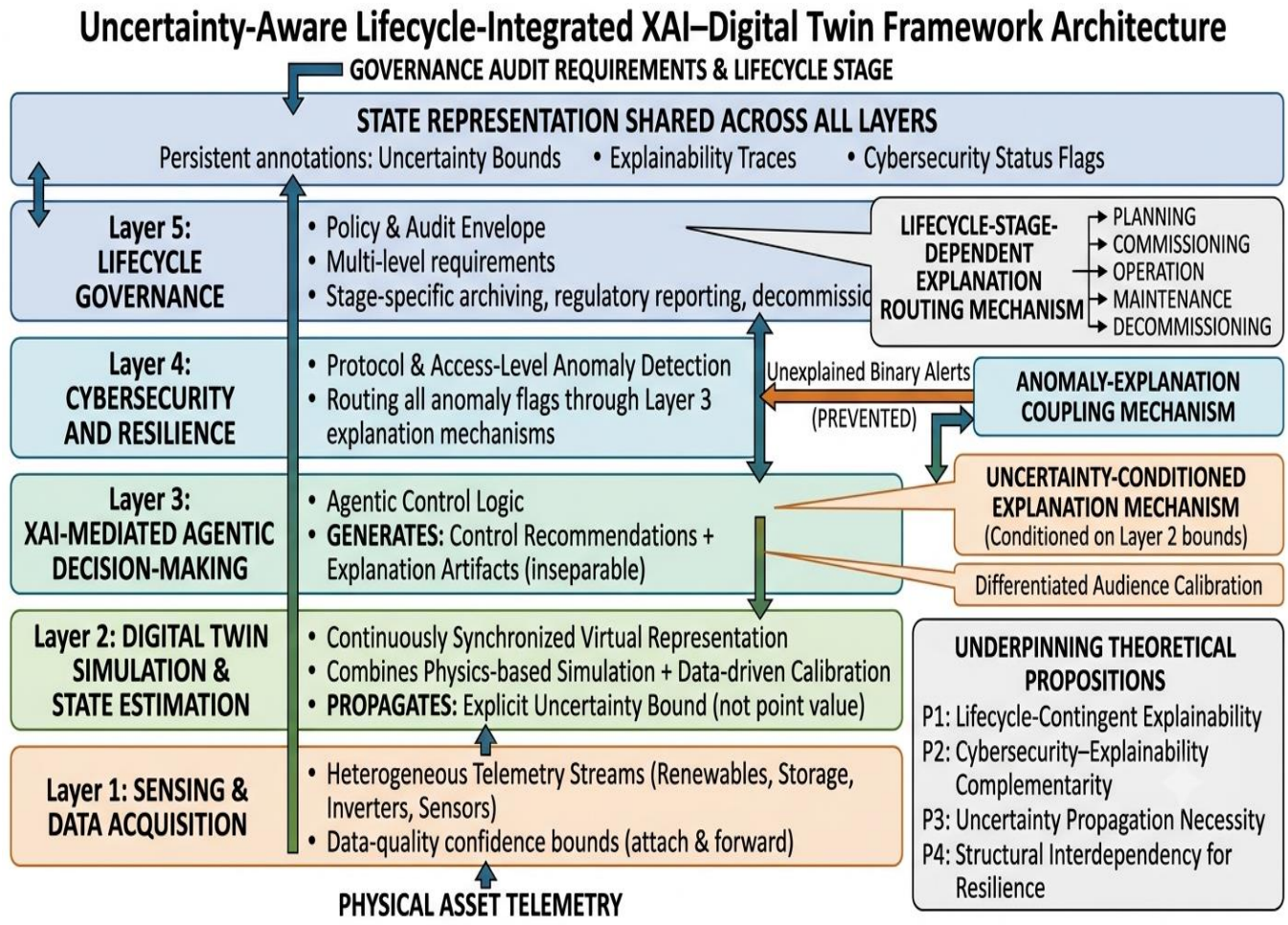
The human-centered XAI literature provides the theoretical corrective: explanation is not an intrinsic algorithmic property but an audience-dependent, context-sensitive function that must be evaluated against the needs, expertise, and decision context of its recipient (Ehsan et al., 2021, 2022, 2023; Liao & Varshney, 2022). A grid operator responding to a real-time anomaly requires a different explanatory artifact than a regulatory authority conducting a post-incident audit or a commissioning engineer verifying that twin behavior matches design specifications. This differentiation is not a presentational refinement; it is a functional requirement that existing model-centric XAI implementations do not satisfy because they generate a single undifferentiated output rather than audience-stratified explanations conditioned on lifecycle stage. The proposed framework formalizes this requirement by configuring the explainability layer to generate technically differentiated outputs across all identified lifecycle stages, from design rationale during planning through auditable operational history during decommissioning — the latter representing an open gap explicitly not addressed in the reviewed digital twin lifecycle literature.

8.4 Digital Twin–Enabled Adaptive System Resilience

Resilience in the context of renewable energy digital twins is frequently conflated with fault tolerance — the capacity to maintain operation under component failure — but the framework advanced here employs a more structurally precise definition: resilience is the capacity of the socio-technical system, comprising the physical asset, its digital representation, and the human and automated agents operating on that representation, to anticipate, absorb, adapt to, and recover from disruptions spanning operational, cybersecurity, and epistemic dimensions simultaneously. This definition draws on convergent work in multi-agent grid resilience (Aghmadi & Mohammed, 2024; Dragomir & Dragomir, 2025), self-healing fault diagnosis and recovery (Feng et al., 2025), and urban digital twin resilience (Omrany et al., 2025; Villani et al., 2025), synthesized here against the co-design requirements established in Sections 8.1–8.3.

Digital twin technology contributes to this conception of resilience not through simulation fidelity alone but through its capacity to maintain a continuously updated, uncertainty-annotated representation of physical asset state — enabling both predictive maintenance (Zhao et al., 2025) and the anticipatory reconfiguration of agentic control logic before disruption propagates across system layers. The architectural implication is that resilience cannot be designed into any single layer of the framework in isolation; it emerges from the structural interdependencies among the sensing, simulation, decision, and cybersecurity layers, each of which must remain interpretable under degraded conditions. This cross-layer resilience logic is operationalized in Section 9.

9. Uncertainty-Aware Lifecycle-Integrated XAI–Digital Twin Framework



9.1 Proposed Framework Architecture

The framework is constituted as a five-layer architecture in which each layer performs a distinct functional role while exchanging structured state representations with adjacent layers rather than producing encapsulated outputs. The layers, in order from physical infrastructure to governance, are: (1) Sensing and Data Acquisition; (2) Digital Twin Simulation and State Estimation; (3) XAI-Mediated Agentic Decision-Making; (4) Cybersecurity and Resilience; and (5) Lifecycle Governance. The organizing principle distinguishing this architecture from additive integration of its constituent technologies is the requirement that uncertainty bounds, explainability traces, and cybersecurity status flags be maintained as persistent, jointly visible annotations on the shared state representation passed between layers — not resolved, discarded, or encapsulated at any single layer boundary.

This architectural logic responds directly to the gap identified in Section 2.5: existing contributions treat explainability, cybersecurity, and uncertainty as independently optimized subsystems whose interfaces are unspecified. The five-layer stack specifies those interfaces explicitly, drawing on the DERMS literature (Ding et al., 2022; Strezoski, 2023) for the data and control plane, the human-centered XAI literature (Ehsan et al., 2022; Liao & Varshney, 2022) for explanation output

differentiation, the DER cybersecurity literature (Qi et al., 2016; Sundararajan et al., 2018; Fragkos et al., 2022) for the resilience layer's threat model, and the lifecycle-oriented digital twin literature (Aghazadeh Ardebili et al., 2024; Iossa et al., 2025) for the governance layer's stage-specific requirements.

9.2 Core Framework Layers and Functional Dimensions

Layer 1 — Sensing and Data Acquisition ingests heterogeneous telemetry streams from renewable generation assets, storage systems, inverter-based resources, and grid-edge sensors, applying quality-of-signal assessment and anomaly pre-screening before forwarding tagged data to Layer 2. Critically, this layer does not resolve sensor uncertainty into a clean data representation; it attaches data-quality confidence bounds that travel forward with each observation — a design choice informed by the telemetry-sparse state-estimation literature (Darbali-Zamora et al., 2021) and by the sensitivity analysis findings for photovoltaic digital twins (Belik & Rubanenko, 2025).

Layer 2 — Digital Twin Simulation and State Estimation maintains the continuously synchronized virtual representation of the physical asset or DER network, combining physics-based simulation with data-driven calibration in a manner consistent with hierarchical, self-evolving twin architectures (Zhao et al., 2025). Each state estimate produced at this layer carries an explicit uncertainty bound rather than a deterministic point value, drawing on variability-aware forecasting methods (Mathur et al., 2024; Fose et al., 2024) and propagating that uncertainty upward rather than absorbing it.

Layer 3 — XAI-Mediated Agentic Decision-Making receives the uncertainty-annotated twin state and applies agentic control logic of the kind theorized for smart-grid autonomy (Kiasari & Aly, 2026; Hasan & Nguyen, 2026), generating both control recommendations and, inseparably, the explanation artifacts required to justify those recommendations to differentiated audiences. Feature-attribution and surrogate-model techniques surveyed by Machlev et al. (2022) and Perr-Sauer et al. (2025) provide the core explainability mechanisms, calibrated here against the uncertainty bounds passed from Layer 2 — so that an explanation for a control recommendation also communicates the confidence level of the twin state on which it rests.

Layer 4 — Cybersecurity and Resilience applies protocol- and access-level anomaly detection consistent with DER cybersecurity frameworks (Qi et al., 2016; Sundararajan et al., 2018) against the same shared twin state maintained in Layer 2, routing all anomaly flags through the explainability mechanisms of Layer 3 before those flags reach any human operator or agentic controller. This routing requirement — the architectural novelty that differentiates this layer from existing DER security implementations — ensures that no unexplained binary alert reaches a decision-making agent.

Layer 5 — Lifecycle Governance provides the policy and auditability envelope within which all lower layers operate, integrating multi-level governance requirements (Deng et al., 2021; Omrany et al., 2025) with stage-specific archiving, regulatory reporting, and decommissioning audit functions. This layer activates different explanation and documentation protocols depending on the lifecycle stage — a functional differentiation grounded in the urban digital twin governance literature (Aghazadeh Ardebili et al., 2024; Executive Vice President & Imam, 2025) and in the

equity-oriented digital twin frameworks addressing multi-owned building energy allocation (Poshnath et al., 2025) and heat-pump deployment equity (Savage et al., 2022).

9.3 Explainability, Resilience, and Cybersecurity Mechanisms

The co-location of explainability and cybersecurity reasoning on a shared state representation — rather than their sequential application to separate data streams — constitutes the primary mechanism through which the framework operationalizes trustworthy autonomy. Three specific mechanisms instantiate this co-location.

The **uncertainty-conditioned explanation mechanism** ensures that any explanation generated by Layer 3 is conditioned on the confidence bounds of the twin state it describes, so that a high-uncertainty forecast is accompanied by an explanation that communicates that uncertainty to the consuming audience rather than presenting a confident rationale for what is effectively a low-confidence recommendation. This mechanism responds to the concern raised by Liao et al. (2024) regarding the trustworthiness of XAI in wind power forecasting, extending it from a post-hoc evaluation concern to an architectural design requirement.

The **anomaly-explanation coupling mechanism** routes every cybersecurity anomaly flag from Layer 4 through the Layer 3 explanation pipeline before delivery to any downstream decision agent or human operator. This mechanism ensures that the basis for a flagged event — specifically, which features of the shared twin state triggered the anomaly classifier, and with what confidence — is transparent at the point of consumption, addressing the structural deficiency in existing DER intrusion-detection systems noted in Section 4.

The **lifecycle-stage-dependent explanation routing mechanism** configures the explainability layer to produce differentiated output artifacts calibrated to the audience and decision context of each lifecycle stage: design-rationale summaries during planning (cf. Poshnath et al., 2025), verification reports during commissioning (cf. Darbali-Zamora et al., 2021), real-time operator traces and agentic action justifications during operation (cf. Kiasari & Aly, 2026), predictive-maintenance rationales during scheduled maintenance (cf. Zhao et al., 2025), and auditable historical summaries during decommissioning — this last function representing an open architectural requirement not resolved by any reviewed contribution.

9.4 Theoretical Propositions and Research Assumptions

Four theoretical propositions, each derived from the foundations established in Section 6, underpin the framework's design logic and specify the empirically testable claims that subsequent validation studies should address.

Proposition 1 (Lifecycle-Contingent Explainability): The quality and operational utility of XAI outputs in a renewable energy digital twin are not uniform across the asset lifecycle but are functions of the lifecycle stage, the audience role, and the prevailing uncertainty level of the twin state — a proposition that follows from the HCXAI theoretical position (Ehsan et al., 2021, 2022;

Liao & Varshney, 2022) and that is operationalized in the lifecycle-stage-dependent explanation routing mechanism of Layer 3.

Proposition 2 (Cybersecurity–Explainability Complementarity): Effective cyber-physical anomaly response in DER ecosystems requires that anomaly detection and explanation generation be architecturally coupled on a shared state representation — specifically, that no anomaly flag reach a decision-making agent without an accompanying explanation of its evidential basis — a proposition grounded in the DER vulnerability literature (Qi et al., 2016; Sundararajan et al., 2018) and operationalized in the anomaly-explanation coupling mechanism of Layer 4.

Proposition 3 (Uncertainty Propagation Necessity): Trustworthy autonomous decision-making in agentic DER control requires that forecast uncertainty be propagated as an explicit annotation through all architectural layers rather than resolved into deterministic point estimates at any single layer boundary — a proposition grounded in variability-aware forecasting research (Mathur et al., 2024; Tavoosi et al., 2021) and operationalized in the uncertainty-conditioned explanation mechanism of Layer 3.

Proposition 4 (Structural Interdependency for Resilience): System-level resilience in a renewable energy digital twin is an emergent property of the structural interdependencies among the five architectural layers, not an attribute achievable within any single layer in isolation — a proposition grounded in the multi-agent resilience literature (Aghmadi & Mohammed, 2024; Dragomir & Dragomir, 2025) and in self-healing power systems research (Feng et al., 2025).

These propositions assume: (a) that the physical asset is equipped with sufficient telemetry density to support meaningful state estimation, consistent with the lower bound established by Darbali-Zamora et al. (2021) for telemetry-sparse environments; (b) that the governance environment within which the framework operates provides regulatory hooks for lifecycle-stage-differentiated audit requirements; and (c) that agentic control components possess sufficient computational resources to execute real-time explainability generation without latency incompatible with grid-stability requirements — an assumption identified as requiring empirical stress-testing in simulation-based validation.

9.5 Pathways for Future Empirical Validation

The framework's conceptual status, consistent with the design-science paradigm adopted in Section 3, does not constitute a limitation to be apologized for but a research stage whose successor is clearly specified. Three pathways for empirical validation are identified, ordered by methodological immediacy.

Simulation-based testbed validation represents the most proximate validation pathway, using established power systems simulation environments — augmented with XAI post-processing pipelines and synthetic cyber-attack injection — to test whether the uncertainty propagation, anomaly-explanation coupling, and lifecycle-stage routing mechanisms produce explanation artifacts that satisfy operational usefulness criteria. This pathway is technically feasible with currently available tools and should constitute the first empirical research cycle following this conceptual contribution.

Pilot deployment within an operational DER network constitutes the medium-term validation pathway, enabling evaluation of the framework's governance layer against real regulatory and operator-trust requirements rather than simulated proxies. The equitable energy allocation and multi-level governance literature (Poshnath et al., 2025; Savage et al., 2022; Omrany et al., 2025) provides the social-equity and stakeholder-engagement criteria against which such a pilot should be assessed alongside technical performance metrics.

Cross-lifecycle longitudinal evaluation represents the most demanding and most consequential validation pathway: following a single renewable energy asset or DER network through multiple lifecycle stages — from commissioning through operational steady-state through a maintenance intervention — to test whether the framework's lifecycle-stage-dependent mechanisms produce measurably superior operator trust, anomaly response times, and governance auditability relative to architectures in which explainability, cybersecurity, and uncertainty are managed as separate subsystems. This pathway directly tests Proposition 4 and is the validation design most capable of distinguishing the framework's contribution from incremental extension of any single prior stream.

10. Analysis and Contributions

10.1 Implications for Renewable Energy System Resilience

The framework's most consequential implication for renewable energy system resilience is the reconceptualization of resilience as a cross-layer emergent property rather than a single-layer engineering objective. Current practice, reflected across the multi-agent resilience literature (Aghmadi & Mohammed, 2024), the self-healing power systems literature (Feng et al., 2025), and the urban digital twin literature (Omrany et al., 2025), tends to optimize resilience within a defined functional scope — microgrid coordination, fault recovery, or city-scale infrastructure continuity — without specifying how resilience at one layer is conditioned on the interpretability and security properties of adjacent layers. The framework's Proposition 4 formalizes this conditioning: a digital twin that maintains operational continuity under a grid disturbance but whose anomaly reasoning is opaque to operators, or whose state representation has been adversarially compromised, is not resilient in any operationally meaningful sense of the term.

This reconceptualization has direct implications for how renewable energy resilience is measured and evaluated. Existing metrics — restoration time, load-shedding magnitude, frequency deviation thresholds — capture the physical dimension of resilience but are silent on its interpretability and security dimensions. The framework implies a need for composite resilience metrics that assess, jointly, the speed and accuracy of autonomous response, the quality and accessibility of the explanations accompanying that response, and the integrity of the twin state representation on which both are conditioned.

10.2 Advancing Trustworthy Explainable AI in Critical Energy Infrastructure

The framework advances the XAI-for-energy field beyond its current model-centric orientation by establishing explainability as a structural design requirement distributed across the full architectural stack — from the uncertainty annotation applied at the sensing layer through the audience-

differentiated explanation outputs generated at the decision layer to the auditable historical records produced at the governance layer. This advance responds directly to the systematic limitation identified by Perr-Sauer et al. (2025) and Machlev et al. (2022): that XAI has been productively adapted across nearly every renewable energy subdomain but almost universally as a post-hoc layer applied to an isolated model rather than as a constitutive property of the operational architecture.

The practical significance of this reorientation is greatest at the interface between autonomous agentic control and human operator oversight — the interface where, as Ridley (2025) and Liao and Varshney (2022) both argue, the absence of explanation is not merely an epistemic inconvenience but a barrier to regulatory acceptance and operator trust that, if unaddressed, will constrain the deployment of autonomous energy management systems regardless of their technical performance. By embedding explainability as a structural requirement in the agentic decision layer and routing all cybersecurity anomaly flags through that layer before delivery to operators, the framework creates the conditions under which regulatory trust can be grounded in architectural design rather than retrospective model audit.

10.3 Cybersecurity Governance Across the Renewable Energy Lifecycle

The framework's treatment of cybersecurity governance as a lifecycle-differentiated, architecturally embedded function — rather than a compliance layer applied uniformly across all operational states — constitutes a substantive advance over both the technical DER security literature (Qi et al., 2016; Sundararajan et al., 2018; Fragkos et al., 2022) and the governance-oriented reviews (Alsaigh et al., 2023; Kumar et al., 2020). The governance layer's stage-specific security protocols — baseline access-control policy during planning, telemetry-calibration integrity during commissioning, continuous anomaly screening during operation, predictive-maintenance data provenance during scheduled maintenance, and archival integrity during decommissioning — reflect the recognition that the attack surface of a renewable energy digital twin is not static but evolves with the asset's lifecycle state.

This lifecycle-differentiated cybersecurity governance model has direct implications for regulatory frameworks governing DER interconnection and aggregation. Existing standards and literature reviews (Obi et al., 2020; Valova & Brown, 2022; Cook et al., 2018) address access control and interconnection requirements in terms of operational steady-state rather than lifecycle stage, a framing that systematically underspecifies the security requirements at commissioning and decommissioning transitions — the precise moments at which the communication interfaces between physical hardware and digital representation are most susceptible to manipulation. The framework's governance layer provides the architectural basis for lifecycle-differentiated regulatory standards that existing frameworks do not yet specify.

10.4 Managing Uncertainty Through Digital Twin Intelligence

The framework's treatment of uncertainty as a persistent, propagated state annotation — rather than a preprocessing problem to be resolved before decision logic engages — establishes a principled architecture for what might be termed *digital twin intelligence under uncertainty*: the capacity of an autonomous, explanation-generating control system to make decisions whose confidence

bounds are transparent to all consuming audiences simultaneously. This capacity is not achievable within architectures that resolve uncertainty at a single layer boundary, because doing so severs the causal chain between the stochastic properties of physical generation and the confidence level of autonomous control recommendations — a severance that renders both the explanation and the cybersecurity anomaly detection epistemically incomplete.

The significance of this contribution is amplified by the increasing deployment of agentic AI architectures for DER coordination (Kiasari & Aly, 2026; Hasan & Nguyen, 2026; Nong, 2026), in which the volume and speed of autonomous decisions outpaces human review capacity. In this context, uncertainty-conditioned explanations serve not merely as decision-support artifacts but as the primary mechanism through which operator oversight is exercised over a largely autonomous system — making the quality of uncertainty propagation a governance requirement, not merely a technical preference.

10.5 Contributions to Sustainable and Secure Energy Transitions

Situated within the broader literature on carbon-neutral urban energy transitions (Chen et al., 2022; Su & Urban, 2021; Lee & Jung, 2023) and the multi-level governance of digitalized energy infrastructure (Deng et al., 2021; Omrany et al., 2025), the framework advances the field in a direction that existing technical contributions, however sophisticated individually, have not reached: the articulation of a single reference architecture in which the technical requirements of trustworthy autonomy — explainability, cyber-resilience, and uncertainty quantification — are treated as jointly binding design constraints distributed across the full lifecycle of a renewable energy asset.

This articulation matters not only for the immediate community of digital twin and XAI researchers but for the multi-level governance scholars and energy policy practitioners who must translate technical system design into regulatory mandates, procurement standards, and public accountability mechanisms. The framework's lifecycle governance layer provides a conceptual bridge between the engineering design choices embedded in Layers 1–4 and the policy and audit requirements that govern the deployment of autonomous systems in critical national infrastructure — a bridge whose construction is a prerequisite for the secure, equitable, and trustworthy energy transitions to which the global community of researchers and policymakers is committed.

11. Contributions

11.1 Theoretical Contributions

The theoretical contribution of this framework resides not in the refinement of any single constituent technology but in the articulation of a structurally coherent design logic that treats explainability, cybersecurity resilience, and uncertainty quantification as co-binding architectural constraints across the full lifecycle of a renewable energy digital twin — a logic absent from all four reviewed streams individually and from their partial integrations collectively. Three specific theoretical advances merit precise delineation.

The first advance is the formalization of **explainability as a lifecycle-contingent, audience-differentiated architectural property** rather than a model-level algorithmic feature. Existing theoretical treatments of XAI, including the human-centered frameworks developed by Ehsan et al. (2021, 2022, 2023) and Liao and Varshney (2022), establish that explanation quality is audience-dependent and context-sensitive, but they do not derive from that position a corresponding architectural requirement: that the systems within which AI models operate must be designed to generate differentiated explanation artifacts conditioned on lifecycle stage, operator role, and prevailing uncertainty level simultaneously. The framework instantiates this theoretical requirement as the lifecycle-stage-dependent explanation routing mechanism of Layer 3, providing the first architecture in the reviewed literature in which HCXAI principles are operationalized as structural design constraints rather than evaluation criteria applied post-deployment.

The second advance concerns the **theoretical reconceptualization of cyber-physical security as a shared-state, explanation-coupled system property**. The DER cybersecurity literature (Qi et al., 2016; Sundararajan et al., 2018; Fragkos et al., 2022) theorizes vulnerability at the protocol-hardware interface but does not theorize the digital twin's state representation as a co-equal attack surface whose compromise propagates through explanation and decision layers. By positioning cybersecurity reasoning as a co-participant in the shared twin-state representation — and by requiring that all anomaly flags pass through the explainability layer before reaching any decision agent — the framework extends the theoretical boundary of security-by-design from the communication protocol level to the full socio-technical system within which that protocol operates. This extension has no direct precedent in the reviewed literature and constitutes a non-incremental theoretical contribution.

The third advance is the **formalization of uncertainty propagation as a governance requirement** in autonomous renewable energy control systems. Prior theoretical treatments position uncertainty quantification as a forecasting or sensitivity-analysis problem internal to a given model (Belik & Rubanenko, 2025; Mathur et al., 2024; Tavoosi et al., 2021). The framework repositions uncertainty propagation as an architectural requirement whose satisfaction is a precondition for meaningful operator oversight of agentic decision-making — a theoretical claim grounded in the convergence of the HCXAI literature's operator-trust arguments (Ridley, 2025; Liao & Varshney, 2022) with the governance literature's accountability demands (Deng et al., 2021; Omrany et al., 2025). Proposition 3 of Section 9.4 formalizes this claim in empirically testable terms, providing a theoretical foundation for subsequent simulation and validation studies.

Taken together, these three advances constitute a contribution to what the framework terms an **integrative socio-technical resilience theory** for digitalized energy infrastructure: a theoretical position in which the trustworthiness of an autonomous renewable energy system is understood as jointly produced by lifecycle-appropriate explanation, security-aware state representation, and uncertainty-transparent decision logic — none of which is theoretically sufficient in isolation, and none of which has previously been theorized as a co-design requirement with the other two.

11.2 Policy Contributions

The policy significance of the framework operates across three governance registers — regulatory standard-setting, procurement and deployment governance, and multi-level energy transition accountability — each of which is materially underserved by the current state of digital twin and XAI deployment practice.

At the **regulatory standard-setting level**, the framework's lifecycle-differentiated cybersecurity governance model provides an architectural basis for regulatory frameworks that existing DER interconnection standards do not yet specify. Current standards and their corresponding literature reviews (Obi et al., 2020; Valova & Brown, 2022; Cook et al., 2018) address operational steady-state security requirements but are structurally silent on the distinct vulnerabilities that arise at commissioning and decommissioning transitions — the lifecycle stages at which the communication interfaces between physical hardware and digital representation are most susceptible to adversarial manipulation. The framework's Layer 5 governance architecture, specifying stage-differentiated security protocols from design through decommissioning, provides a concrete structural reference for the development of lifecycle-integrated regulatory mandates governing DER aggregation, smart-inverter deployment, and digital twin operation in critical national energy infrastructure. Regulators in jurisdictions advancing digitalization roadmaps — including those operating under multi-level governance frameworks theorized by Deng et al. (2021) and Omrany et al. (2025) — will find in this framework a technically grounded basis for moving beyond operationally uniform compliance requirements toward stage-sensitive governance instruments.

At the **procurement and deployment governance level**, the framework's explicit specification of co-design requirements — that explainability, cybersecurity, and uncertainty quantification be treated as jointly binding constraints from the earliest architectural stage of a digital twin system rather than added as post-deployment refinements — provides procurement authorities and energy utilities with a structured set of technical criteria against which candidate digital twin platforms can be evaluated. The absence of such criteria in existing procurement guidance is a direct consequence of the fragmentation documented in Section 2.5: because the four reviewed streams have developed independently, no synthesis-level guidance specifying their joint requirements has been available to inform procurement decisions. The framework addresses this gap by providing, in the five-layer architecture and its associated four theoretical propositions, a reference specification against which vendors, integrators, and independent assessors can evaluate the trustworthiness of proposed digital twin deployments.

At the **multi-level energy transition accountability level**, the framework contributes to the governance of sustainable and secure energy transitions (Chen et al., 2022; Su & Urban, 2021; Lee & Jung, 2023) by providing the conceptual infrastructure for accountability mechanisms that span technical system design and public governance simultaneously. The framework's governance layer bridges the engineering choices embedded in Layers 1–4 — uncertainty propagation, anomaly-explanation coupling, lifecycle-stage routing — and the audit, reporting, and equity requirements that public accountability demands of autonomous systems operating in critical infrastructure. This bridge is a prerequisite for the operationalization of multi-level governance frameworks in

digitalized energy systems, including those addressing the social-equity dimensions of renewable energy deployment theorized by Savage et al. (2022) and Poshnath et al. (2025), where the opacity of automated allocation decisions constitutes a governance failure regardless of their technical optimality. Policymakers engaged in climate governance, just energy transition planning, and digital infrastructure regulation will find in this framework a principled basis for embedding transparency and security requirements into the institutional architecture of energy system digitalization — not as afterthoughts to technical deployment but as conditions of its legitimacy.

11.3 Practical Contributions

The framework's practical contributions operate at the level of system architecture, operational deployment, and institutional capacity — translating its theoretical and policy advances into actionable guidance for engineers, system operators, and energy infrastructure managers.

For **digital twin system architects and platform developers**, the five-layer architecture provides a reference design that resolves a persistent practical ambiguity: how to interface explainability, cybersecurity, and uncertainty management components within a single operational system without reducing any of the three to a supplementary module disconnected from the system's core state representation. The shared-state-representation principle — under which uncertainty bounds, explanation traces, and cybersecurity anomaly flags are maintained as persistent, jointly visible annotations rather than resolved or discarded at layer boundaries — provides a concrete integration protocol that existing platform architectures have not specified. This protocol is directly implementable in simulation-based testbed environments using established power systems tools and XAI post-processing pipelines, providing an immediate practical pathway from the conceptual framework to testable system instantiation.

For **grid operators and DER management system operators**, the framework's audience-differentiated explanation architecture provides the operational infrastructure for meaningful human oversight of autonomous control decisions at a scale and speed that exceeds unaided human cognitive capacity. The practical consequence of the lifecycle-stage-dependent explanation routing mechanism is that operators receive, at each lifecycle stage, precisely the explanation artifacts calibrated to their decision context — real-time control traces during operation, verification reports during commissioning, predictive-maintenance rationales during scheduled interventions — rather than undifferentiated model outputs that must be interpreted without contextual scaffolding. This differentiation, grounded in the HCXAI literature (Ehsan et al., 2022; Liao & Varshney, 2022) and operationalized in the framework's Layer 3, has direct implications for operator training, interface design, and the institutional protocols governing human-machine authority allocation in autonomous grid management systems.

For **energy asset owners and infrastructure managers**, the framework's decommissioning-stage audit function — the production of an auditable account of the asset's full operational history from the explainability and governance layers — addresses a practical gap that the reviewed digital twin lifecycle literature explicitly does not resolve. As renewable energy assets approach end-of-life or repowering decisions, the absence of structured, machine-readable operational histories limits the quality of decommissioning planning, residual-value assessment, and regulatory compliance documentation. The framework's governance layer specifies this audit function as an architectural

requirement rather than a best-practice recommendation, providing asset managers with both the technical rationale and the design reference needed to incorporate end-of-life auditability into digital twin procurement and deployment decisions from the outset.

12. Limitations

Intellectual honesty demands that the framework's limitations be stated with the same precision as its contributions, and four limitations warrant direct acknowledgment.

The most structurally significant limitation is the framework's **exclusively conceptual status at the time of submission**. The five-layer architecture, its integration mechanisms, and the four theoretical propositions of Section 9.4 have not yet been subjected to simulation-based, prototype, or operational validation. This limitation is consistent with, and explicitly anticipated by, the design-science research paradigm adopted in Section 3, in which conceptual artifact development precedes empirical evaluation in subsequent research cycles (cf. the methodology discussion in Section 3.5). However, it means that the claims advanced in Sections 9 and 10 remain theoretical propositions rather than empirically supported findings — a distinction that must be preserved in any subsequent citation or policy application of this work. The computational feasibility of real-time explainability generation within the latency constraints of grid-stability-relevant agentic control, in particular, is an assumption identified in Section 9.4 as requiring empirical stress-testing rather than a design property whose satisfaction has been demonstrated.

A second limitation concerns the framework's **implicit assumption of sufficient telemetry density**. The architecture, particularly its Layer 2 state-estimation and uncertainty-propagation mechanisms, assumes instrumentation density sufficient to support meaningful real-time synchronization between physical asset state and digital twin representation. For legacy renewable energy infrastructure, community-scale distributed generation, and assets deployed in resource-constrained operational environments, this assumption may not hold, and the framework does not currently specify degraded-mode operating protocols for telemetry-sparse conditions beyond the lower bound established by Darbali-Zamora et al. (2021) for distribution-network state estimation.

Third, the framework **does not address the computational and energy overhead** of simultaneously running real-time digital twin synchronization, multi-agent agentic control, XAI explanation generation, and continuous cybersecurity anomaly screening at the edge — a co-deployment scenario that imposes non-trivial computational demands that the framework's current specification does not quantify. For small-scale DER deployments and resource-constrained edge computing environments, this overhead may constitute a practical barrier to implementation that architectural refinement alone cannot resolve.

Fourth, the framework's **governance layer, while theoretically grounded in multi-level governance literature** (Deng et al., 2021; Omrany et al., 2025), does not engage with the institutional heterogeneity of specific regulatory jurisdictions. Lifecycle-differentiated cybersecurity and audit requirements will interact differently with the governance architectures of different national energy regulatory systems, and the framework's current specification is at a level of abstraction that does not map directly onto any particular jurisdiction's regulatory instruments.

Empirical validation through pilot deployment (Section 9.5) will require engagement with jurisdiction-specific regulatory environments that this conceptual contribution cannot anticipate in full.

13. Future Research Directions

Five research directions, each directly derived from the framework's propositions and limitations, define the agenda for subsequent empirical and theoretical work.

The most proximate priority is **simulation-based validation of the five-layer architecture** using established power systems testbed environments augmented with synthetic cyber-attack injection, XAI post-processing pipelines, and multi-agent control frameworks. The specific validation targets are the three integration mechanisms described in Section 9.3 — uncertainty-conditioned explanation, anomaly-explanation coupling, and lifecycle-stage-dependent explanation routing — assessed against operationally defined criteria for explanation quality, anomaly response latency, and control decision confidence-bound transparency. This research direction directly addresses the framework's status as an unvalidated conceptual artifact and should constitute the immediate successor study.

The second direction concerns the **development of composite resilience metrics** that jointly assess the physical, interpretability, and security dimensions of renewable energy digital twin performance, as implied by the framework's reconceptualization of resilience in Section 8.4 and 10.1. Existing resilience metrics within the multi-agent and self-healing systems literature (Aghmadi & Mohammed, 2024; Feng et al., 2025) capture physical recovery performance but are silent on explanation quality and state-representation integrity under adversarial conditions. Constructing and empirically testing composite metrics that span all three dimensions would provide the field with evaluation instruments capable of assessing the framework's full contribution rather than its physical performance dimension alone.

The third direction addresses the framework's telemetry-density assumption through **investigation of degraded-mode operating protocols** for telemetry-sparse and resource-constrained deployment environments. Building on the sparse-environment state-estimation literature (Darbali-Zamora et al., 2021) and the sensitivity-analysis methods for low-instrumentation photovoltaic systems (Belik & Rubanenko, 2025), this research direction should develop architectural extensions that specify how the framework's uncertainty propagation and explanation mechanisms degrade gracefully — rather than fail catastrophically — when telemetry coverage falls below the density assumptions of the full five-layer architecture.

The fourth direction concerns the **extension of the framework to urban-scale and multi-asset DER networks**, including positive energy districts (Malakhatka et al., 2025), campus-to-city scale heat transition systems (Lesnyak et al., 2023), and equity-sensitive multi-owned building energy allocation systems (Poshnath et al., 2025; Savage et al., 2022). The current framework is specified at a level of abstraction applicable to both single-asset and network-scale deployments, but the governance layer's equity and multi-stakeholder dimensions are most consequential — and most empirically tractable — at the urban district scale where social-equity outcomes are directly

measurable alongside technical performance. Collaboration between technical systems researchers and multi-level governance scholars will be essential for this extension.

The fifth direction addresses what remains the most demanding and most consequential open question in the framework's empirical agenda: **longitudinal cross-lifecycle evaluation** of a single renewable energy asset or DER network progressing through multiple successive lifecycle stages. Only a longitudinal study design can test Proposition 4 — that system-level resilience is an emergent property of cross-layer structural interdependencies rather than a single-layer attribute — and only such a design can assess whether the framework's lifecycle-stage-dependent mechanisms produce measurably superior governance auditability, operator trust, and anomaly response performance relative to architectures that treat explainability, cybersecurity, and uncertainty as separately optimized subsystems. This research direction will require sustained institutional partnerships with grid operators, regulatory bodies, and renewable energy asset managers, and represents the most significant medium-term investment the research community can make toward empirically grounding the trustworthy autonomous energy management systems to which both the academic literature and global energy transition policy aspire.

14. Conclusion

The accelerating penetration of variable renewable energy into modern power systems has produced a design imperative that the field has not yet fully confronted: autonomous energy management systems must be simultaneously trustworthy, secure, and uncertainty-aware across the full arc of an asset's operational life — not merely within the operational steady-state that most existing architectures address. The framework developed in this paper responds to that imperative not by incrementally extending any single prior stream of scholarship but by articulating the structural design logic through which explainability, cybersecurity resilience, and uncertainty quantification must interoperate as co-binding architectural constraints if the trustworthiness of renewable energy digital twins is to be more than an aspirational claim.

The central argument developed across Sections 8–13 is that the fragmentation of the four reviewed streams — XAI for energy applications, digital twin architectures, multi-agent and agentic AI control, and DER cybersecurity — has produced a landscape rich in partial integrations but devoid of the structural coherence required for trustworthy autonomous operation. Each stream has advanced its own design requirements with considerable sophistication; what none has addressed is how those requirements constrain one another when co-located within a single operational system whose outputs carry consequences for grid stability, operator safety, and public accountability simultaneously. The five-layer architecture proposed here addresses precisely that co-location problem, specifying not only what each layer must do but how layers must exchange uncertainty bounds, explanation traces, and security status flags to satisfy the demands of all three dimensions jointly.

Three theoretical propositions formalize the framework's claims in empirically testable terms — that explainability is lifecycle-contingent and audience-differentiated, that cybersecurity reasoning and explanation generation must be architecturally coupled on a shared state representation, and that uncertainty propagation is a governance requirement rather than a preprocessing convenience

— and a fourth identifies system-level resilience as the emergent property of their structural integration rather than the aggregate of their individual optimization. Together, these propositions define a research agenda whose completion will require simulation-based validation, composite resilience metric development, degraded-mode protocol specification, urban-scale extension, and longitudinal cross-lifecycle evaluation — a program of work that this conceptual contribution is designed to enable rather than to substitute.

The framework's policy and practical implications are commensurate with its theoretical ambition. For regulators advancing lifecycle-differentiated governance of DER and digital twin deployments, it provides an architectural reference for security and audit standard-setting that existing frameworks have not specified. For procurement authorities evaluating digital twin platforms, it provides joint technical criteria that fragmented vendor claims cannot currently satisfy. For grid operators navigating the boundary between autonomous control and meaningful human oversight, it provides the architectural infrastructure for explanation artifacts calibrated to their decision context rather than to the model developer's convenience. And for the broader community of researchers, practitioners, and policymakers committed to secure, equitable, and trustworthy energy transitions, it provides a principled demonstration that trustworthiness is not a property that can be added to an autonomous energy system after its architecture is fixed — it must be designed in from the beginning, distributed across every layer, and sustained across every stage of the asset's life.

15. Declaration of the Use of Artificial Intelligence

(AI) The authors used Claude AI during the preparation of this conceptual article to support language editing, drafting, and manuscript structuring. The AI tool was used exclusively to improve the presentation and clarity of the text and did not generate the scholarly concepts, theoretical arguments, critical interpretations, or conclusions of the study. All AI-assisted outputs were critically reviewed, substantially revised where necessary, and approved by the authors, who accept full responsibility for the content and integrity of the manuscript.

16. Declarations

Conflict of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

17. References

Adıgüzel Tüylü, A. N. (2026). Interpreting Primary Energy Consumption in Europe and Türkiye Using Explainable Artificial Intelligence. *ELECTRICA*, *26*, 1–12. https://doi.org/10.5152/electrica.2026.25405

Aghazadeh Ardebili, A., Zappatore, M., Ramadan, A. I. H. A., Longo, A., & Ficarella, A. (2024). Digital Twins of smart energy systems: A systematic literature review on enablers, design, management and computational challenges. *Energy Informatics*, *7*(1), 94. https://doi.org/10.1186/s42162-024-00385-5

Aghmadi, A., & Mohammed, O. A. (2024). Operation and Coordinated Energy Management in Multi-Microgrids for Improved and Resilient Distributed Energy

Resource Integration in Power Systems. *Electronics*, *13*(2), 358. https://doi.org/10.3390/electronics13020358

Agostinelli, S., Cumo, F., Guidi, G., & Tomazzoli, C. (2021). Cyber-Physical Systems Improving Building Energy Management: Digital Twin and Artificial Intelligence. *Energies*, *14*(8), 2338. https://doi.org/10.3390/en14082338

Agostinelli, S., Cumo, F., Nezhad, M. M., Orsini, G., & Piras, G. (2022). Renewable Energy System Controlled by Open-Source Tools and Digital Twin Model: Zero Energy Port Area in Italy. *Energies*, *15*(5), 1817. https://doi.org/10.3390/en15051817

Akyol, S., Das, M., & Alatas, B. (2023). Modeling the Energy Consumption of R600a Gas in a Refrigeration System with New Explainable Artificial Intelligence Methods Based on Hybrid Optimization. *Biomimetics*, *8*(5), 397. https://doi.org/10.3390/biomimetics8050397

Alhasnawi, B. N., Jasim, B. H., Rahman, Z.-A. S. A., Guerrero, J. M., & Esteban, M. D. (2021). A Novel Internet of Energy Based Optimal Multi-Agent Control Scheme for Microgrid including Renewable Energy Resources. *International Journal of Environmental Research and Public Health*, *18*(15), 8146. https://doi.org/10.3390/ijerph18158146

Alsaigh, R., Mehmood, R., & Katib, I. (2023). AI explainability and governance in smart energy systems: A review. *Frontiers in Energy Research*, *11*, 1071291. https://doi.org/10.3389/fenrg.2023.1071291

Astolfi, D., De Caro, F., & Vaccaro, A. (2023). Condition Monitoring of Wind Turbine Systems by Explainable Artificial Intelligence Techniques. *Sensors*, *23*(12), 5376. https://doi.org/10.3390/s23125376

Bayer, D., & Pruckner, M. (2023). A digital twin of a local energy system based on real smart meter data. *Energy Informatics*, *6*(1), 8. https://doi.org/10.1186/s42162-023-00263-6

Belik, M., & Rubanenko, O. (2025). Sensitivity analysis of digital twin model for energy community PV system. *Scientific Reports*, *15*(1), 29097. https://doi.org/10.1038/s41598-025-13707-8

Bocullo, V., Martišauskas, L., Gatautis, R., Vonžudaitė, O., Bakas, R., Milčius, D., Venčaitis, R., & Pupeikis, D. (2023). A Digital Twin Approach to City Block Renovation Using RES Technologies. *Sustainability*, *15*(12), 9307. https://doi.org/10.3390/su15129307

Bousnina, D., & Guerassimoff, G. (2024). Optimal energy management in smart energy systems: A deep reinforcement learning approach and a digital twin case-study. *Smart Energy*, *16*, 100163. https://doi.org/10.1016/j.segy.2024.100163

Çankal, A., & Ever, D. (2025). The Effects of Renewable Energy Consumption on Financial Performance: An Explainable Artificial Intelligence (XAI)-Based Research on the BIST Sustainability Index. *International Journal of Energy Economics and Policy*, *15*(4), 204–213. https://doi.org/10.32479/ijeep.19602

Chen, L., Msigwa, G., Yang, M., Osman, A. I., Fawzy, S., Rooney, D. W., & Yap, P.-S. (2022). Strategies to achieve a carbon neutral society: A review. *Environmental Chemistry Letters*, *20*(4), 2277–2310. https://doi.org/10.1007/s10311-022-01435-8

Colson, C. M., Nehrir, M. H., & Gunderson, R. W. (2011). Multi-agent Microgrid Power Management. *IFAC Proceedings Volumes*, *44*(1), 3678–3683. https://doi.org/10.3182/20110828-6-IT-1002.01188

Cook, J. J., Ardani, K. B., O'Shaughnessy, E. J., Margolis, R. M., & Smith, B. (2018). *Expanding PV Value: Lessons Learned from Utility-led Distributed Energy Resource Aggregation in the United States* (NREL/TP--6A20-71984, 1483067; p. NREL/TP--6A20-71984, 1483067). https://doi.org/10.2172/1483067

Darbali-Zamora, R., Johnson, J., Summers, A., Jones, C. B., Hansen, C., & Showalter, C. (2021). State Estimation-Based Distributed Energy Resource Optimization for Distribution Voltage Regulation in Telemetry-Sparse Environments Using a Real-Time Digital Twin. *Energies*, *14*(3), 774. https://doi.org/10.3390/en14030774

Das, A., & Rad, P. (2020). *Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey* (arXiv:2006.11371). arXiv. https://doi.org/10.48550/arXiv.2006.11371

Deng, T., Zhang, K., & Shen, Z.-J. (Max). (2021). A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *Journal of Management Science and Engineering*, *6*(2), 125–134. https://doi.org/10.1016/j.jmse.2021.03.003

Ding, F., Liu, W., MacDonald, J., Ogle, J., Pratt, A., Saha, A., Hagerman, J., & Baggu, M. (2022). *Federated Architecture for Secure and Transactive Distributed Energy Resource Management Solutions (FAST-DERMS)* (NREL/TP-5D00-81566, 1839591, MainId:82339; p. NREL/TP-5D00-81566, 1839591, MainId:82339). https://doi.org/10.2172/1839591

Dragomir, O. E., & Dragomir, F. (2025). A Decentralized Hierarchical Multi-Agent Framework for Smart Grid Sustainable Energy Management. *Sustainability*, *17*(12), 5423. https://doi.org/10.3390/su17125423

Drewnick, T., Wen, X., Oberhofer, U., Rydin Gorjão, L., Beck, C., Hagenmeyer, V., & Schäfer, B. (2025). Analyzing deterministic and stochastic influences on the power grid frequency dynamics with explainable artificial intelligence. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *35*(3), 033153. https://doi.org/10.1063/5.0239371

Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M., Streit, M., Wachter, S., Riener, A., & Riedl, M. O. (2021). Operationalizing Human-Centered Perspectives in Explainable AI. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3411763.3441342

Ehsan, U., Wintersberger, P., Liao, Q. V., Watkins, E. A., Manger, C., Daumé Iii, H., Riener, A., & Riedl, M. O. (2022). Human-Centered Explainable AI (HCXAI): Beyond Opening the Black-Box of AI. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–7. https://doi.org/10.1145/3491101.3503727

Ehsan, U., Wintersberger, P., Watkins, E. A., Manger, C., Ramos, G., Weisz, J. D., Daumé Iii, H., Riener, A., & Riedl, M. O. (2023). Human-Centered Explainable AI (HCXAI): Coming of Age. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–7. https://doi.org/10.1145/3544549.3573832

Farid, A. M. (2015). Multi-Agent System Design Principles for Resilient Coordination & Control of Future Power Systems. *Intelligent Industrial Systems*, *1*(3), 255–269. https://doi.org/10.1007/s40903-015-0013-x

Feng, J., Yu, T., Zhang, K., & Cheng, L. (2025). Integration of Multi-Agent Systems and Artificial Intelligence in Self-Healing Subway Power Supply Systems: Advancements in Fault Diagnosis, Isolation, and Recovery. *Processes*, *13*(4), 1144. https://doi.org/10.3390/pr13041144

Fernández-Miguel, A., Ortíz-Marcos, S., Jiménez-Calzado, M., Fernández Del Hoyo, A. P., García-Muiña, F. E., & Settembre-Blundo, D. (2025). Agentic AI in Smart Manufacturing: Enabling Human-Centric Predictive Maintenance Ecosystems. *Applied Sciences*, *15*(21), 11414. https://doi.org/10.3390/app152111414

Ferré-Bigorra, J., Casals, M., & Gangolells, M. (2022). The adoption of urban digital twins. *Cities*, *131*, 103905. https://doi.org/10.1016/j.cities.2022.103905

Fose, N., Singh, A. R., Krishnamurthy, S., Ratshitanga, M., & Moodley, P. (2024). Empowering distribution system operators: A review of distributed energy resource forecasting techniques. *Heliyon*, *10*(15), e34800. https://doi.org/10.1016/j.heliyon.2024.e34800

Fragkos, G., Johnson, J., & Tsiropoulou, E. E. (2022). Centralized and Decentralized Distributed Energy Resource Access Control Implementation Considerations. *Energies*, *15*(17), 6375. https://doi.org/10.3390/en15176375

Garí-Galíndez, J., Fonoll-Rubio, R., Andrade-Arvizu, J., Witte, W., Perez-Rodriguez, A., Paetel, S., Vidal-Fuentes, P., Guc, M., & Izquierdo-Roca, V. (2025). Explainable Artificial Intelligence Driven Methodology for Accelerated Research of Complex Systems: Case Study of Thin-Film Photovoltaic Kesterite-Based Technology. *Advanced Energy Materials*, *15*(35), 2502420. https://doi.org/10.1002/aenm.202502420

Gill, S. S., Murugesan, S. S., Anurag, K. A., Verma, P., Kaur, H., Kumar, S., & Kumar, M. (2026). *Agentic AI: Vision and Challenges*. https://doi.org/10.36227/techrxiv.176779782.20689880/v1

González-Briones, A., De La Prieta, F., Mohamad, M. S., Omatu, S., & Corchado, J. M. (2018). Multi-Agent Systems Applications in Energy Optimization Problems: A State-of-the-Art Review. *Energies*, *11*(8), 1928. https://doi.org/10.3390/en11081928

Hasan, A., & Nguyen, D. T. (2026). Integrating agentic AI and digital twins for intelligent decision-making systems. *Array*, *29*, 100721. https://doi.org/10.1016/j.array.2026.100721

He, T., & Jazizadeh, F. (2025). *Context-aware LLM-based AI Agents for Human-centered Energy Management Systems in Smart Buildings* (arXiv:2512.25055). arXiv. https://doi.org/10.48550/arXiv.2512.25055

Hosseini, S., & Seilani, H. (2025). The role of agentic AI in shaping a smart future: A systematic review. *Array*, *26*, 100399. https://doi.org/10.1016/j.array.2025.100399

Ilic, M. D., Jaddivada, R., & Korpas, M. (2020). Interactive protocols for distributed energy resource management systems (DERMS). *IET Generation, Transmission & Distribution*, *14*(11), 2065–2081. https://doi.org/10.1049/iet-gtd.2019.1022

Iossa, R., Domenighini, P., & Cotana, F. (2025). Digital Twins from Building to Urban Areas: An Open Opportunity to Energy, Environmental, Economic and Social Benefits. *Applied Sciences*, *15*(19), 10795. https://doi.org/10.3390/app151910795

Iram, S., Shahzad, A. R., Farid, H. M. A., & Shakeel, H. M. (2025). Analyzing the impact of weather conditions on energy efficiency in residential buildings using machine learning techniques with explainable artificial intelligence. *Advances in Building Energy Research*, *19*(5), 625–659. https://doi.org/10.1080/17512549.2025.2527675

Izmirlioglu, Y., Pham, L., Son, T. C., & Pontelli, E. (2024). A Survey of Multi-Agent Systems for Smartgrids. *Energies*, *17*(15), 3620. https://doi.org/10.3390/en17153620

Kaewunruen, S., Rungskunroch, P., & Welsh, J. (2018). A Digital-Twin Evaluation of Net Zero Energy Building for Existing Buildings. *Sustainability*, *11*(1), 159. https://doi.org/10.3390/su11010159

Karavas, C.-S., Arvanitis, K., & Papadakis, G. (2017). A Game Theory Approach to Multi-Agent Decentralized Energy Management of Autonomous Polygeneration Microgrids. *Energies*, *10*(11), 1756. https://doi.org/10.3390/en10111756

Kiasari, M., & Aly, H. (2026). Agentic Artificial Intelligence for Smart Grids: A Comprehensive Review of Autonomous, Safe, and Explainable Control Frameworks. *Energies*, *19*(3), 617. https://doi.org/10.3390/en19030617

Kim, J., Maathuis, H., & Sent, D. (2024). Human-centered evaluation of explainable AI applications: A systematic review. *Frontiers in Artificial Intelligence*, *7*, 1456486. https://doi.org/10.3389/frai.2024.1456486

Kim, M., Ghobadi, F., Tayerani Charmchi, A. S., Lee, M., & Lee, J. (2025). Digital Twins for Clean Energy Systems: A State-of-the-Art Review of Applications, Integrated Technologies, and Key Challenges. *Sustainability*, *18*(1), 43. https://doi.org/10.3390/su18010043

Komninos, N. (2022). Net Zero Energy Districts: Connected Intelligence for Carbon-Neutral Cities. *Land*, *11*(2), 210. https://doi.org/10.3390/land11020210

Konhäuser, K., & Werner, T. (2024). Uncovering the financial impact of energy-efficient building characteristics with eXplainable artificial intelligence. *Applied Energy*, *374*, 123960. https://doi.org/10.1016/j.apenergy.2024.123960

Kruse, J., Schäfer, B., & Witthaut, D. (2022). Secondary control activation analysed and predicted with explainable AI. *Electric Power Systems Research*, *212*, 108489. https://doi.org/10.1016/j.epsr.2022.108489

Kumar, N. M., Chand, A. A., Malvoni, M., Prasad, K. A., Mamun, K. A., Islam, F. R., & Chopra, S. S. (2020). Distributed Energy Resources and the Application of AI, IoT, and Blockchain in Smart Grids. *Energies*, *13*(21), 5739. https://doi.org/10.3390/en13215739

Lasseter, R., Akhil, A., Marnay, C., Stephens, J., Dagle, J., Guttromsom, R., Meliopoulous, A. S., Yinger, R., & Eto, J. (2002). *Integration of distributed energy resources. The CERTS Microgrid Concept* (LBNL--50829, 799644; p. LBNL--50829, 799644). https://doi.org/10.2172/799644

Lee, J., & Jung, S. (2023). Towards Carbon-Neutral Cities: Urban Classification Based on Physical Environment and Carbon Emission Characteristics. *Land*, *12*(5), 968. https://doi.org/10.3390/land12050968

Lesnyak, E., Belkot, T., Hurka, J., Hörding, J. P., Kuhlmann, L., Paulau, P., Schnabel, M., Schönfeldt, P., & Middelberg, J. (2023). Applied Digital Twin Concepts Contributing to Heat Transition in Building, Campus, Neighborhood, and Urban Scale. *Big Data and Cognitive Computing*, *7*(3), 145. https://doi.org/10.3390/bdcc7030145

Li, Q., & He, Y. (2025). An Overview of Digital Twin Concept for Key Components of Renewable Energy Systems. *International Journal of Robotics and Automation Technology*, *8*, 29–47. https://doi.org/10.31875/2409-9694.2021.08.4

Liao, Q. V., & Varshney, K. R. (2022). *Human-Centered Explainable AI (XAI): From Algorithms to User Experiences* (arXiv:2110.10790). arXiv. https://doi.org/10.48550/arXiv.2110.10790

Liao, W., Fang, J., Ye, L., Bak-Jensen, B., Yang, Z., & Porte-Agel, F. (2024). Can we trust explainable artificial intelligence in wind power forecasting? *Applied Energy*, *376*, 124273. https://doi.org/10.1016/j.apenergy.2024.124273

Ma, Z., Schultz, M. J., Christensen, K., Værbak, M., Demazeau, Y., & Jørgensen, B. N. (2019). The Application of Ontologies in Multi-Agent Systems in the Energy Sector: A Scoping Review. *Energies*, *12*(16), 3200. https://doi.org/10.3390/en12163200

Maarif, M. R., Saleh, A. R., Habibi, M., Fitriyani, N. L., & Syafrudin, M. (2023). Energy Usage Forecasting Model Based on Long Short-Term Memory (LSTM) and eXplainable Artificial Intelligence (XAI). *Information*, *14*(5), 265. https://doi.org/10.3390/info14050265

Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI*, *9*, 100169. https://doi.org/10.1016/j.egyai.2022.100169

Malakhatka, E., Wallbaum, H., Abouebeid, S., Hofer, G., Pooyanfar, P., Dursun, İ., Weber, G., Geçer, H. S., & Thuvander, L. (2025). Fostering Sustainable Urban Energy Transitions: Backcasting for Positive Energy Districts and Digital Twin strategies in a European context. *CIB Conferences*, *1*(1). https://doi.org/10.7771/3067-4883.1959

Mansour, D. A., Numair, M., Zalhaf, A. S., Ramadan, R., Darwish, M. M. F., Huang, Q., Hussien, M. G., & Abdel-Rahim, O. (2023). Applications of IoT and digital twin in electrical power systems: A comprehensive survey. *IET Generation, Transmission & Distribution*, *17*(20), 4457–4479. https://doi.org/10.1049/gtd2.12940

Mathur, A., Kumari, R., Meena, V. P., Singh, V. P., Azar, A. T., & Hameed, I. A. (2024). Data-driven optimization for microgrid control under distributed energy resource variability. *Scientific Reports*, *14*(1), 10806. https://doi.org/10.1038/s41598-024-58767-4

Mazzetto, S. (2024). A Review of Urban Digital Twins Integration, Challenges, and Future Directions in Smart City Development. *Sustainability*, *16*(19), 8337. https://doi.org/10.3390/su16198337

Moon, J., Maqsood, M., So, D., Baik, S. W., Rho, S., & Nam, Y. (2024). Advancing ensemble learning techniques for residential building electricity consumption forecasting: Insight from explainable artificial intelligence. *PLOS ONE*, *19*(11), e0307654. https://doi.org/10.1371/journal.pone.0307654

Nneoma Ugwu, C., Chukwudi Ogenyi, F., Nnenna Ugwu, J., Okechukwu Ugwu, P.-C., & Ben Okon, M. (2026). Integration of AI-driven digital twins for real-time optimization of renewable energy grids. *Frontiers in Energy Research*, *14*, 1748233. https://doi.org/10.3389/fenrg.2026.1748233

Nong, Y. (2026). *Agentic AI for Autonomous Cascade Failure Detection and Response in Critical Energy Infrastructure—Proof-of-Concept Implementation and Validation Framework for Multi-Agent Decision Systems*. https://doi.org/10.36227/techrxiv.176784457.77367434/v1

Obi, M., Slay, T., & Bass, R. (2020). Distributed energy resource aggregation using customer-owned equipment: A review of literature and standards. *Energy Reports*, *6*, 2358–2369. https://doi.org/10.1016/j.egyr.2020.08.035

Omar, O. (2025). Digital Twins for Climate-Responsive Urban Development: Integrating Zero-Energy Buildings into Smart City Strategies. *Sustainability*, *17*(15), 6670. https://doi.org/10.3390/su17156670

Omrany, H., Mehdipour, A., Oteng, D., & Al-Obaidi, K. M. (2025). The uptake of urban digital twins in the built environment: A pathway to resilient and sustainable cities. *Computational Urban Science*, *5*(1), 20. https://doi.org/10.1007/s43762-025-00177-x

Panagoulias, D. P., Sarmas, E., Marinakis, V., Virvou, M., Tsihrintzis, G. A., & Doukas, H. (2023). Intelligent Decision Support for Energy Management: A Methodology for Tailored Explainability of Artificial Intelligence Analytics. *Electronics*, *12*(21), 4430. https://doi.org/10.3390/electronics12214430

Panigutti, C., Beretta, A., Fadda, D., Giannotti, F., Pedreschi, D., Perotti, A., & Rinzivillo, S. (2023). Co-design of Human-centered, Explainable AI for Clinical Decision Support. *ACM Transactions on Interactive Intelligent Systems*, *13*(4), 1–35. https://doi.org/10.1145/3587271

Perr-Sauer, J., Ugirumurera, J., Gafur, J., Bensen, E. A., Nguyen, T., Paul, S., Severino, J., Nag, A., Vijayshankar, S., Gasper, P., Finegan, D. P., Holden, J., Mueller, J., Graf, P., Tripp, C., & Egan, H. (2025). Applications of explainable artificial intelligence in renewable energy research. *Energy Reports*, *14*, 2217–2235. https://doi.org/10.1016/j.egyr.2025.08.046

Poshnath, A., Chen, Y., Rismanchi, B., & Rajabifard, A. (2025). Adaptive spatial planning for equitable renewable energy allocation in multi-owned buildings: A policy-integrated digital twin approach. *Energy and Buildings*, *344*, 115988. https://doi.org/10.1016/j.enbuild.2025.115988

Pütz, S., Kruse, J., Witthaut, D., Hagenmeyer, V., & Schäfer, B. (2023). Regulatory Changes in German and Austrian Power Systems Explored with Explainable Artificial Intelligence. *Companion Proceedings of the 14th ACM International Conference on Future Energy Systems*, 26–31. https://doi.org/10.1145/3599733.3600247

Qi, J., Hahn, A., Lu, X., Wang, J., & Liu, C. (2016). Cybersecurity for distributed energy resources and smart inverters. *IET Cyber-Physical Systems: Theory & Applications*, *1*(1), 28–39. https://doi.org/10.1049/iet-cps.2016.0018

Rengasamy, P., & R, R. (2025). Explainable artificial intelligence framework for wind turbine fault detection using random forest – Extreme gradient boosting hybrid model. *Results in Engineering*, *28*, 107446. https://doi.org/10.1016/j.rineng.2025.107446

Ridley, M. (2025). HUMAN-CENTERED explainable artificial intelligence: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology*, *76*(1), 98–120. https://doi.org/10.1002/asi.24889

Rogers, A., Ramchurn, S., & Jennings, N. (2021). Delivering the Smart Grid: Challenges for Autonomous Agents and Multi-Agent Systems Research. *Proceedings of the AAAI Conference on Artificial Intelligence*, *26*(1), 2166–2172. https://doi.org/10.1609/aaai.v26i1.8445

Savage, T., Akroyd, J., Mosbach, S., Hillman, M., Sielker, F., & Kraft, M. (2022). Universal Digital Twin – the impact of heat pumps on social inequality. *Advances in Applied Energy*, *5*, 100079. https://doi.org/10.1016/j.adapen.2021.100079

Shadi, M. R., Mirshekali, H., & Shaker, H. R. (2025). Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects. *Renewable and Sustainable Energy Reviews*, *216*, 115668. https://doi.org/10.1016/j.rser.2025.115668

Shen, J., Zheng, F., Tesema, F. B., Deng, W., Bellotti, A. G., & Xie, J. (2026). Deep learning and explainable Artificial Intelligence for large-scale photovoltaic suitability analysis in Australian cities. *Renewable Energy*, *259*, 125025. https://doi.org/10.1016/j.renene.2025.125025

Sriniva, K., Shinde, S. C., Acharya, P., & Roy, S. (2024). *Explainable AI for Energy Prediction and Anomaly Detection in Solar Energy Systems*. In Review. https://doi.org/10.21203/rs.3.rs-4922729/v1

Strezoski, L. (2023). Distributed energy resource management systems—DERMS: State of the art and how to move forward. *WIREs Energy and Environment*, *12*(1), e460. https://doi.org/10.1002/wene.460

Su, C., & Urban, F. (2021). Carbon Neutral China by 2060: The Role of Clean Heating Systems. *Energies*, *14*(22), 7461. https://doi.org/10.3390/en14227461

Sundararajan, A., Chavan, A., Saleem, D., & Sarwat, A. I. (2018). A Survey of Protocol-Level Challenges and Solutions for Distributed Energy Resource Cyber-Physical Security. *Energies*, *11*(9), 2360. https://doi.org/10.3390/en11092360

Tavoosi, J., Suratgar, A. A., Menhaj, M. B., Mosavi, A., Mohammadzadeh, A., & Ranjbar, E. (2021). Modeling Renewable Energy Systems by a Self-Evolving Nonlinear Consequent Part Recurrent Type-2 Fuzzy System for Power Prediction. *Sustainability*, *13*(6), 3301. https://doi.org/10.3390/su13063301

Tiwari, A. (2025). Conceptualising the emergence of Agentic Urban AI: From automation to agency. *Urban Informatics*, *4*(1), 13. https://doi.org/10.1007/s44212-025-00079-7

Touzani, S., Prakash, A. K., Wang, Z., Agarwal, S., Pritoni, M., Kiran, M., Brown, R., & Granderson, J. (2021). Controlling distributed energy resources via deep reinforcement learning for load flexibility and energy efficiency. *Applied Energy*, *304*, 117733. https://doi.org/10.1016/j.apenergy.2021.117733

Valova, R., & Brown, G. (2022). Distributed energy resource interconnection: An overview of challenges and opportunities in the United States. *Solar Compass*, *2*, 100021. https://doi.org/10.1016/j.solcom.2022.100021

Villani, L., Gugliermetti, L., Barucco, M. A., & Cinquepalmi, F. (2025). A Digital Twin Framework to Improve Urban Sustainability and Resiliency: The Case Study of Venice. *Land*, *14*(1), 83. https://doi.org/10.3390/land14010083

Wang, X., Wang, H., Bhandari, B., & Cheng, L. (2024). AI-Empowered Methods for Smart Energy Consumption: A Review of Load Forecasting, Anomaly Detection and Demand Response. *International Journal of Precision Engineering and Manufacturing-Green Technology*, *11*(3), 963–993. https://doi.org/10.1007/s40684-023-00537-0

Yao, R., Hu, Y., & Varga, L. (2023). Applications of Agent-Based Methods in Multi-Energy Systems—A Systematic Literature Review. *Energies*, *16*(5), 2456. https://doi.org/10.3390/en16052456

Yassen, M. A., El-Kenawy, E.-S. M., Abdel-Fattah, M. G., Ismail, I., & Mostafa, H. E.-D. S. (2025). Explainable artificial intelligence for wind power forecasting model based on long short-term memory. *Neural Computing and Applications*, *37*(19), 14589–14611. https://doi.org/10.1007/s00521-025-11230-5

Zhang, X., Wang, Q., Yu, J., Sun, Q., Hu, H., & Liu, X. (2023). A Multi-Agent Deep-Reinforcement-Learning-Based Strategy for Safe Distributed Energy Resource Scheduling in Energy Hubs. *Electronics*, *12*(23), 4763. https://doi.org/10.3390/electronics12234763

Zhao, K., Liu, Y., Zhou, Y., Ming, W., & Wu, J. (2025). A Hierarchical and Self-Evolving Digital Twin (HSE-DT) Method for Multi-Faceted Battery Situation Awareness Realisation. *Machines*, *13*(3), 175. https://doi.org/10.3390/machines13030175

Zhao, Y., Zhong, H., & Lim, C. C. (2024). Safety-Constrained Multi-Agent Reinforcement Learning for Power Quality Control in Distributed Renewable Energy Networks. *Computers, Materials & Continua*, *79*(1), 449–471. https://doi.org/10.32604/cmc.2024.048771