

ORIGINAL RESEARCH

Open access

A Data-Driven Energy Infrastructure Digital Twin Approach Using Retrieval-Augmented Generation

Ethan Wright¹, Chloe Bennett^{1*}, Jack Turner²

Abstract

Platforms based on digital twins are gaining widespread use in overseeing energy infrastructure, particularly within smart grid environments. Recent progress in artificial intelligence opens up new possibilities to build greater user confidence by strengthening predictive and prescriptive analysis capabilities and refining how users interact with these systems. This study introduces an innovative framework and architectural design for energy digital twins that combines data-driven techniques with knowledge-based elements. Machine learning-driven data integration and analysis are fused into a knowledge graph that captures asset condition information, forecasting results, and relevant domain expertise. This structure underpins a retrieval-augmented generation strategy, which powers an advanced conversational virtual assistant grounded in a large language model. The assistant delivers practical decision-making support for asset oversight and maintenance activities. Key parts of the suggested architecture have been aligned with readily available commercial tools to create a working prototype. This prototype was applied in a practical scenario involving a portion of the high-voltage energy network located in central Italy.

Keywords Digital twin, Energy management, Energy infrastructures, Retrieval-augmented generation, Natural user interface

*Correspondence:

Chloe Bennett

chloe.bennett@leeds.ac.uk

¹ Department of Smart Energy Networks, Faculty of Engineering, University of Leeds, Leeds, United Kingdom

² Department of Power Analytics, Faculty of Engineering, University of Sheffield, Sheffield, United Kingdom

Introduction

Over the last 10 years, the global energy sector has undergone profound changes driven by evolving technologies, regulatory shifts, and market forces. Key influences include rising concerns about climate change, breakthroughs in renewable energy technologies, and a stronger push for sustainable, robust energy networks [1]. Among the most prominent developments has been the swift expansion of distributed and renewable energy generation [2]. Solar and wind technologies have seen dramatic improvements in scale, performance, and cost-effectiveness, making them increasingly viable alternatives to conventional fossil-based sources. This move toward low-carbon options has been fueled by heightened environmental awareness and the urgent need to cut greenhouse gas emissions. National governments and global accords have accelerated the shift toward renewable energy, driven by financial incentives and supportive regulations. The 2015 Paris Agreement [3], in particular, galvanized international efforts to mitigate climate change, encouraging countries to rethink their energy approaches and pursue decarbonization pathways.

Motivated by these objectives, the incorporation of Information and Communication Technology (ICT) has played a central role in upgrading conventional power grids into intelligent smart grids. These advanced grids mark a major step forward in electricity delivery and control, forming a cornerstone of contemporary energy system modernization [4]. The global smart

grid sector is expanding rapidly, driven by demand for more efficient, eco-friendly energy solutions. According to Ref. [5], its market value neared USD 50 billion in 2022 and is projected to rise at a Compound Annual Growth Rate (CAGR) of 17.4%, approaching roughly USD 130 billion by 2028. Smart grids rely on sophisticated metering infrastructure and analytical tools to enhance system reliability while enabling improved demand management. They allow consumers to make better-informed choices about usage patterns. A defining characteristic is the use of digital communication networks that support continuous, real-time oversight and adjustment of grid elements. Smart sensors distributed across the system facilitate bidirectional information flow between users and operators, delivering detailed insights into consumption, production, and network status. This enables both utility managers and end users to adjust usage during off-peak times or favorable pricing windows, thereby improving overall efficiency and load balancing [6]. Furthermore, smart grids enhance system robustness by rapidly detecting faults, isolating faults, and rerouting power to limit service interruptions. They also incorporate mechanisms to identify cybersecurity risks and unauthorized intrusions, safeguarding the reliability of power delivery [7].

Despite these advances in digitizing energy systems, several obstacles remain. Issues around standardization and system compatibility continue to demand attention to ensure smooth incorporation of analytics into routine operations. The digital twin (DT) concept offers a powerful solution to these hurdles. A digital twin functions as a precise virtual mirror of a physical asset or operation, maintained through continuous, two-way data synchronization for applications such as modeling, diagnostics, supervision, and upkeep. First developed in aerospace and widely adopted in intelligent manufacturing, DT approaches hold strong potential to create more adaptable, durable, and environmentally friendly energy systems by delivering enhanced insight and control over physical components [8]. Energy-focused digital twins (EDTs) spanning extensive territories must manage diverse assets, simulation tools, data streams, processing workflows, and operational responsibilities. This inherent complexity creates substantial cognitive demands on users: traditional graphical interfaces, relying mainly on 3D renderings and standard input methods, often hinder efficient resource discovery, information access, and insight extraction, leading to suboptimal usability [9].

To push smart-grid digital twins toward greater maturity in user experience, new interaction strategies are essential. The present research targets the following objectives (RAs) concerning cutting-edge technologies in energy asset oversight:

- RA1: Advancing grid oversight via digital twins combined with data-centric methods: The work investigates how merging digital twins, novel visualization approaches, and analytical techniques can refine infrastructure monitoring, enabling precise real-time evaluation and proactive fault avoidance.
- RA2: Harnessing large language models (LLMs) to aid decision-making processes: It examines the advantages and challenges of applying LLMs in this domain, including data fusion, modeling accuracy, and domain-tailored reliability to support better-informed decisions and increased process automation.
- RA3: Combining retrieval-augmented generation (RAG) techniques with knowledge graphs (KGs): Particular emphasis is placed on using the Resource Description Framework (RDF) [10] to structure domain knowledge. A well-organized KG enhances RAG-based LLMs by providing precise, context-relevant, up-to-date details, thereby improving the quality of decision-support tools.

Building on these goals, the paper outlines a new data-oriented, knowledge-enriched architecture designed for energy digital twins. It delivers comprehensive awareness of network and asset conditions, collects sensor data streams, applies preprocessing and machine learning for issue identification and resolution, and—most importantly—features a virtual assistant powered by an RAG pipeline. This assistant merges a general-purpose large language model with specialized domain knowledge stored in a KG, offering users actionable guidance on performance anomalies and maintenance strategies. The primary contributions include:

- A foundational architecture for monitoring and managing energy infrastructure that unites digital twins with analytical methods;

- A natural user interface (NUI) that goes beyond standard displays and controls by enabling direct asset interaction via gestures and conversational dialogue;
- RDF-based annotation of digital twin elements and live data feeds within a knowledge graph, used to supply targeted, real-time context to the LLM driving the assistant;
- Mapping of architecture components to commercial off-the-shelf (COTS) solutions to improve practicality, cut costs, and shorten deployment timelines;
- Development of a cloud-hosted prototype using open-source components;
- Demonstration through a real-world application on a high-voltage segment in central Italy, illustrating the framework's key benefits.

The proposed approach will undergo thorough evaluation by benchmarking it against existing practices in energy infrastructure oversight. This comparison will emphasize gains in decision quality and operator awareness relative to earlier methods used in the same setting.

Background

This section briefly outlines core concepts related to advanced technologies now being integrated into energy systems, ensuring the paper remains accessible. It then surveys the most significant existing research in the area.

Emerging technologies and digital twins for energy infrastructures

The fusion of cutting-edge technologies with power grids marks an exciting area of progress in modern energy management. Ongoing innovations across multiple fields are expanding the possibilities for improved connectivity, automation, and smart functionality in smart grids. As noted in Ref. [11], the Internet of Things (IoT) has driven widespread deployment of sensors and connected devices, forming an extensive network of energy-related data accessible through numerous grid points. IoT-derived information supports superior monitoring, governance, and informed choices. The rollout of 5G networks has enabled ultra-fast, low-latency communication, vital for instantaneous data sharing in smart grid settings [12]. Such connectivity upgrades improve system responsiveness and dependability, facilitating applications such as self-managing grids, precise metering, and the smooth integration of electric vehicles.

Artificial intelligence (AI) and machine learning (ML) methods are now routinely applied to refine smart grid performance [13]. These approaches process enormous datasets from grid equipment to uncover trends and enhance operations. ML techniques excel at tasks such as anomaly detection, predictive maintenance, preventive strategies, and demand prediction, thereby improving overall reliability and efficiency [14]. Edge computing [15] stands out as a critical supporting technology in this space, relocating processing power nearer to data origins to minimize delays and enable immediate analysis. Rather than routing everything to distant centralized facilities, edge setups decentralize computation along network boundaries. This reduces vulnerability to single-point failures and bolsters system durability. It proves especially valuable when connectivity is constrained, costly, or restricted by data-privacy rules. In energy contexts, edge processing accelerates response to anomalies and enhances operational agility [16, 17]. It also accommodates scalable workloads through additional edge nodes and enables lightweight ML ("tiny ML") [18, 19] or serverless federated learning [20] at distributed points. These decentralized models necessitate innovative security frameworks to protect data integrity [21] and identify threats [22].

Over recent years, digital twin (DT) methodologies have gained traction as a highly effective strategy for supervising intricate systems, including energy networks. Numerous investigations and initiatives have explored DT applications in renewables, storage, and transmission/distribution [23]. At its core, a digital twin acts as a continuously updated virtual

replica of a physical item, system, or workflow [24]. It draws on sensor feeds and other inputs tied to the real entity to maintain an accurate, evolving model that reveals current state, efficiency, and dynamics. Incoming raw data keeps the twin in sync with its physical counterpart. Energy digital twins (EDTs) typically incorporate geospatial mapping, behavioral simulation, analytics engines, and robust connectivity layers. Geospatial components deliver location-aware 3D views, simulation layers model expected responses, analytics derive meaningful conclusions from data, and connectivity maintains live synchronization between physical and virtual realms. Digital twins often encompass the full lifespan of an asset—from initial design through active service and eventual decommissioning [25]. This end-to-end coverage empowers energy stakeholders to track, evaluate, and fine-tune performance across every phase. Additionally, DTs promote teamwork among designers, operators, engineers, and maintenance teams by allowing shared interaction with the virtual model [26]. In smart grid environments, such collaborative capabilities improve coordination and enhance decision processes throughout operations.

Related work

Digital twin (DT) approaches hold substantial promise for professionals overseeing industrial energy systems, enabling evidence-based choices, early detection of potential malfunctions, improved operational and network performance, streamlined processes, and stronger alignment with environmental goals. Nevertheless, broader implementation in the energy field faces barriers, including financial costs, regulatory compliance, and difficulties in assessing effectiveness. The following overview examines leading contemporary research on DT structures tailored to energy networks. The highlighted efforts combine technological integration, analytical processing, security measures, and advanced display methods to deliver a unified strategy for better managing and sustaining energy infrastructures.

In Ref. [27], researchers outlined a conceptual structure and supporting technology layers specifically for DTs in the energy domain. Their platform-independent D-Arc stack can be applied across various DT environments and grid configurations to enable anticipatory fault identification, boost overall performance, and refine operational workflows. Although it includes a process coordination model suitable for both local and remote (cloud) deployments, its general-purpose design may fall short in delivering tailored solutions for diverse application scenarios.

A different initiative in Ref. [28] introduced C3-FLOW, a hybrid cloud-edge DT system designed for overseeing low-emission electrical assets in intelligent industrial zones. This setup improves equipment timetabling, bandwidth assignment, and computing resource allocation, resulting in higher precision for distributed learning, better data transmission efficiency, and lower environmental impact through reduced carbon output. That said, the evaluation relied solely on simulated conditions, without validation involving actual hardware.

Work presented in Ref. [29] described two categories of DT representations to meet varying decision needs in Energy Cyber-Physical Systems (ECPSs): one optimized for lower data rates, suited to routine activities such as energy oversight and forward-looking upkeep, and another for high-frequency operations, such as disruption handling and recovery. The suggested design leverages serverless computation for observation and on-site responses within ECPS environments, utilizing IoT virtual representations as stand-ins. Testing occurred on linked embedded platforms simulating networked microgrids, but broader real-world expansion remains unexplored.

The advantages of applying DT principles to electricity control facilities were explored in Ref. [30]. Their concept relies on a modeling core, the “dynamic digital mirror,” to advance supervisory and regulatory capabilities in power networks. In a parallel effort [31], a comprehensive Power System Digital Twin (PSDT) was developed to enhance operational intelligence, grid governance, anomaly forecasting, live power evaluation, and troubleshooting. It merges intelligent sensing, 5G connectivity, cloud resources, and extensive data handling. Both concepts, however, remain at the conceptual stage, lacking implemented demonstrators.

A layered DT structure in Ref. [32] employs instantaneous modeling to replicate the actual grid, supporting superior design and management activities. It incorporates a flexible, supplier-neutral simulation environment along with hardware-

integrated emulation of physical elements. Key unresolved issues persist, including achieving greater parameter-tuning precision, increasing model detail, and reducing latency.

In Ref. [33], an EDT platform targets resilient smart-grid advancement, supporting full-cycle grid development—from strategy and deployment to daily operations—by consolidating diverse inputs, including electrical metrics, policy directives, and weather information. Still, critical building blocks, including accurate DT modeling techniques and embedded AI routines, have yet to be fully integrated to enable enterprise-level functionality.

Lastly, Zhang et al. [34] put forward a DT-centric design for intelligent electricity distribution networks, addressing the need for rapid development of structured data representations, analytical tools, and processing routines while embedding AI and large-scale data methods. A major constraint noted is limited sensor coverage, calling for deeper investigation into DT suitability across the entire distribution setup and further development of the corresponding models.

Framework Architecture

The smart-grid paradigm arises from the close interplay between electrical distribution networks and digital communication infrastructures. Much like the human body—where the circulatory network transports essential substances and the neural network coordinates sophisticated observation and responses—the smart grid unites energy flow management with information technology coordination. In line with this analogy, the Smart-Grid Architecture Model (SGAM) [35] has been established as a guiding structure for designing and deploying intelligent grid solutions. As illustrated in Figure 1, SGAM provides a comprehensive conceptual framework comprising five interconnected tiers that effectively integrate tangible hardware, networking channels, and information systems within smart energy systems.

The primary objective of the current proposal is to create a standardized, analytics-focused framework capable of hosting a range of DT implementations while following SGAM principles and established guidelines for building expansive middleware platforms [36]. These guidelines emphasize seamless integration across systems, adaptive behavior, immediate responsiveness, capacity for growth, operational stability, and protection of information.

As depicted in Figure 2, the framework is organized into several fundamental strata, each assigned specific duties and capabilities. The subsequent subsections elaborate on the individual elements, including selected technologies for prototype development and their mutual connectivity.

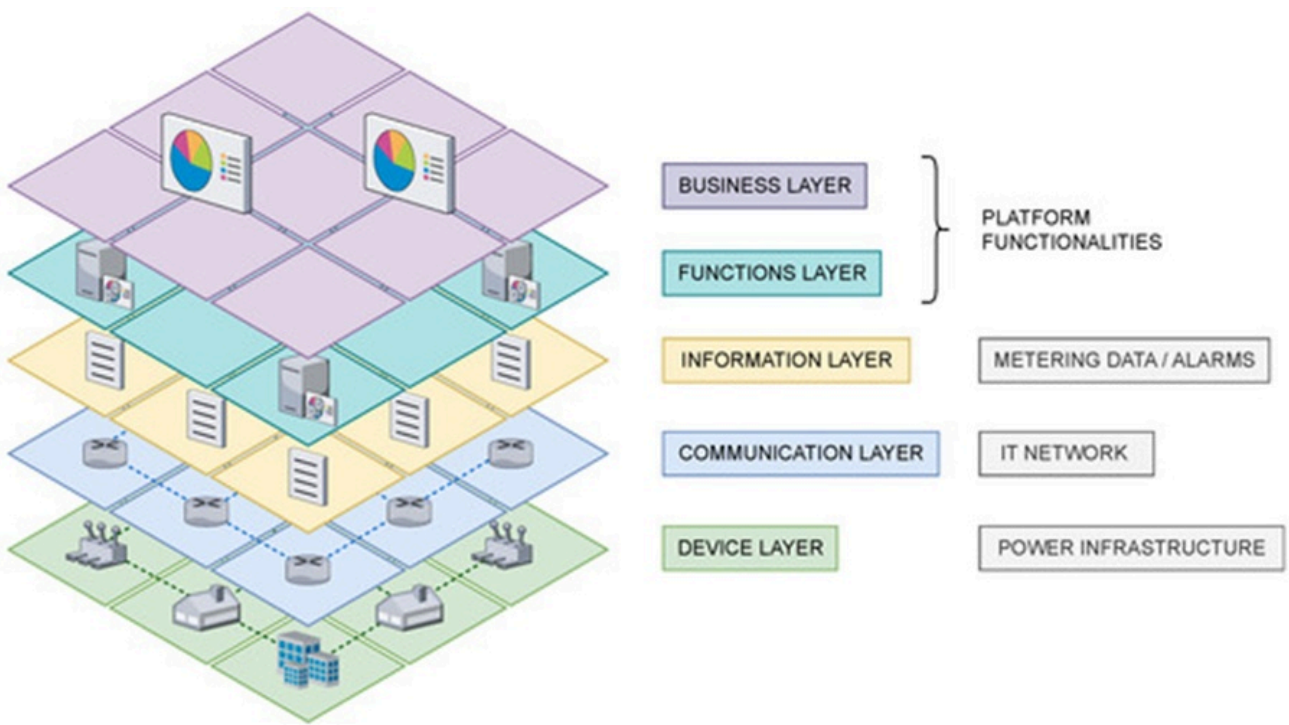


Figure 1. Smart-grid architecture model.

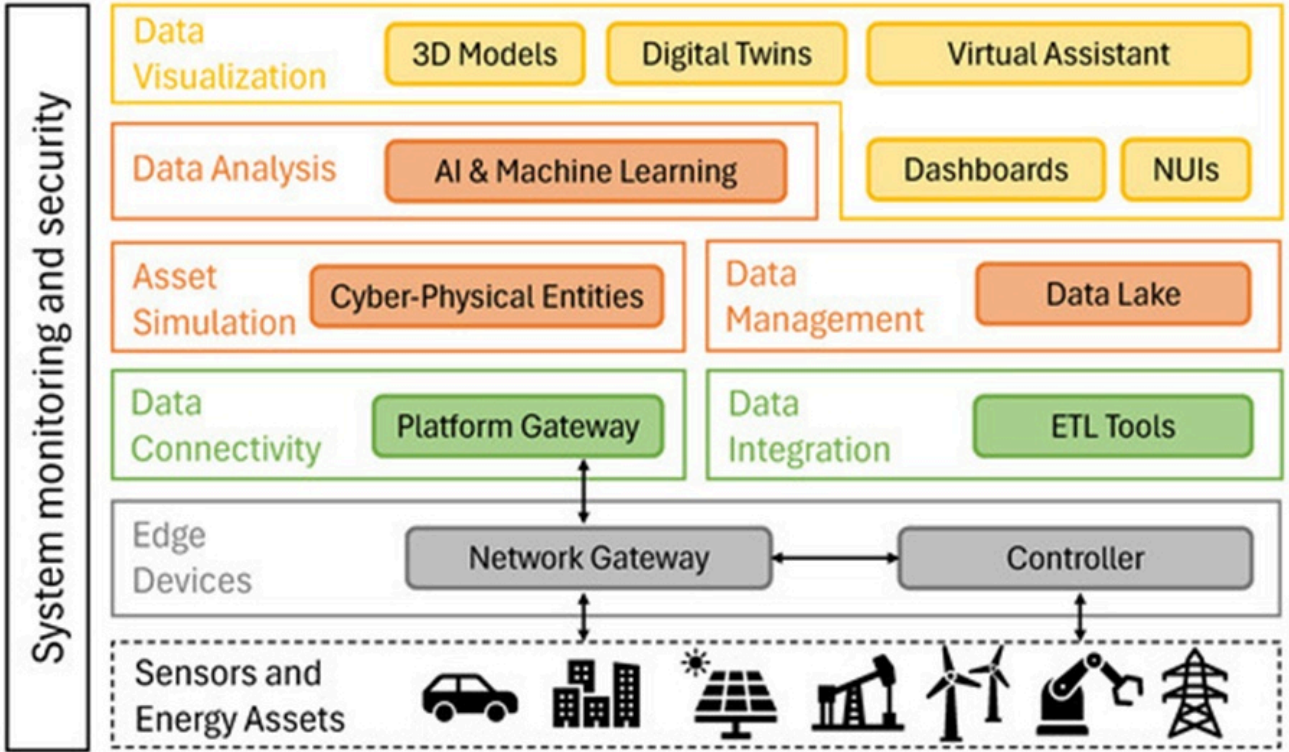


Figure 2. Reference architecture.

Sensor networks and communication

The architecture establishes connections at the most fundamental levels of sensing and networking, linking directly to on-site edge equipment while also interfacing with centralized cloud infrastructure. Positioned closest to the physical field devices, the entry-level layer deploys sensing instruments on actual assets and manages the continuous, two-way transfer of information from these devices to the digital platform.

Serving as the core communication infrastructure, the Data Connectivity layer ensures fluid interaction, real-time alignment, and effective merging of data between the physical domain and its virtual representation. To achieve this, the design supports multiple classes of equipment:

- **Standard sensors:** These instruments produce electrical signals that feed operational and surrounding-condition measurements directly into the digital twin. They supply immediate, interpretable values for process-related parameters, examples being thermal readings or electrical potential differences.
- **Process Analytical Technology (PAT) sensors** operate similarly to conventional sensors and capture information at predefined measurement locations. However, their raw outputs generally demand further analysis—often via statistical methods or dynamic monitoring models—before yielding actionable insights. A typical instance is a near-infrared spectroscopy unit.
- **Actuators:** These elements perform physical interventions on equipment or processes in response to instructions from higher-level control systems. They convert incoming digital or analog directives into corresponding mechanical motion, electrical switching, hydraulic pressure, or pneumatic force to modify real-world states.

Sensor-generated data streams are initially handled within the Edge Computing layer. Conducting computation at or near the point of origin delivers key improvements, notably shorter response times, lower network traffic, and stronger fault tolerance when connectivity is temporarily lost.

In alignment with principles of edge processing and distributed (swarm-based) intelligence [37, 38], the entire platform is structured around a microservices model [39]. This approach packages computational workloads and storage operations into independent, self-managing service units, enabling resource allocation to adapt fluidly and efficiently to changing network conditions and system demands.

Sensor networks in energy applications face complexity due to the broad array of existing communication standards and protocols. Although this variety enhances system robustness and adaptability, it creates significant obstacles to achieving unified, trouble-free cooperation among diverse devices.

As indicated in **Figure 2**, the infrastructure incorporates two main categories of communication hardware to bridge protocol differences and enable reliable exchanges throughout the system:

- **Gateways:** Installed locally, these units collect raw sensor outputs and execute protocol conversion, reformatting incoming messages from heterogeneous sources into a uniform internal structure. They commonly apply initial data refinement steps—including noise suppression, value scaling, and anomaly removal—to boost data reliability while lowering transmission overhead. Built-in security capabilities encompass user/device authentication, encrypted data channels, and fine-grained access rules to protect sensitive information flows.
- **Controllers:** Responsible for coordinating messaging between actuators and other parts of the network (such as gateways), these devices accept control directives from upstream components, relay them accurately to the target actuators linked to the physical infrastructure, and oversee the assignment of operations to ensure reliable, optimized communication performance.

Given the stringent demands of energy infrastructures—particularly around affordability, efficient data acquisition, and ultra-low latency for mission-critical messages—the platform adopts a messaging backbone built on the Message

Queueing Telemetry Transport (MQTT) protocol [40]. MQTT (<https://mqtt.org>, accessed on 8 September 2024) is a lightweight publish/subscribe messaging standard in which publisher clients send topic-tagged messages to a central broker, which then routes them to all subscribed clients monitoring those topics.

Gateways are also equipped to communicate with local or remote services via Representational State Transfer (REST) Application Programming Interfaces (APIs) [41]. REST APIs provide a consistent, adaptable, and low-overhead way to connect software components—especially in microservice architectures—using standard HTTP verbs to perform operations such as retrieving, modifying, and deleting resources.

Although multiple data encoding formats are supported, the platform adopts JavaScript Object Notation (JSON) (<https://www.json.org>, accessed on 8 September 2024) as its primary interchange format. JSON's broad compatibility, independence from programming languages, and readability for both people and machines make it an ideal choice.

Data management and storage techniques

Higher within the layered design, raw incoming data undergo preprocessing—filtering and summarization—inside the Data Integration module. This module's core responsibility is to represent and harmonize information about the three lowest tiers of the SGAM reference model (Figure 1), ultimately constructing the integrated, coherent perspective illustrated in Figure 3.

Establishing well-defined logical data models yields several important gains: it facilitates clearer analytical interpretation and the uncovering of hidden correlations within datasets; it enforces consistent documentation standards and architectural patterns throughout the system; it helps prevent implementation mistakes when introducing new platform features; and it improves collaboration between technical data specialists and professionals focused on business analytics and strategic intelligence.

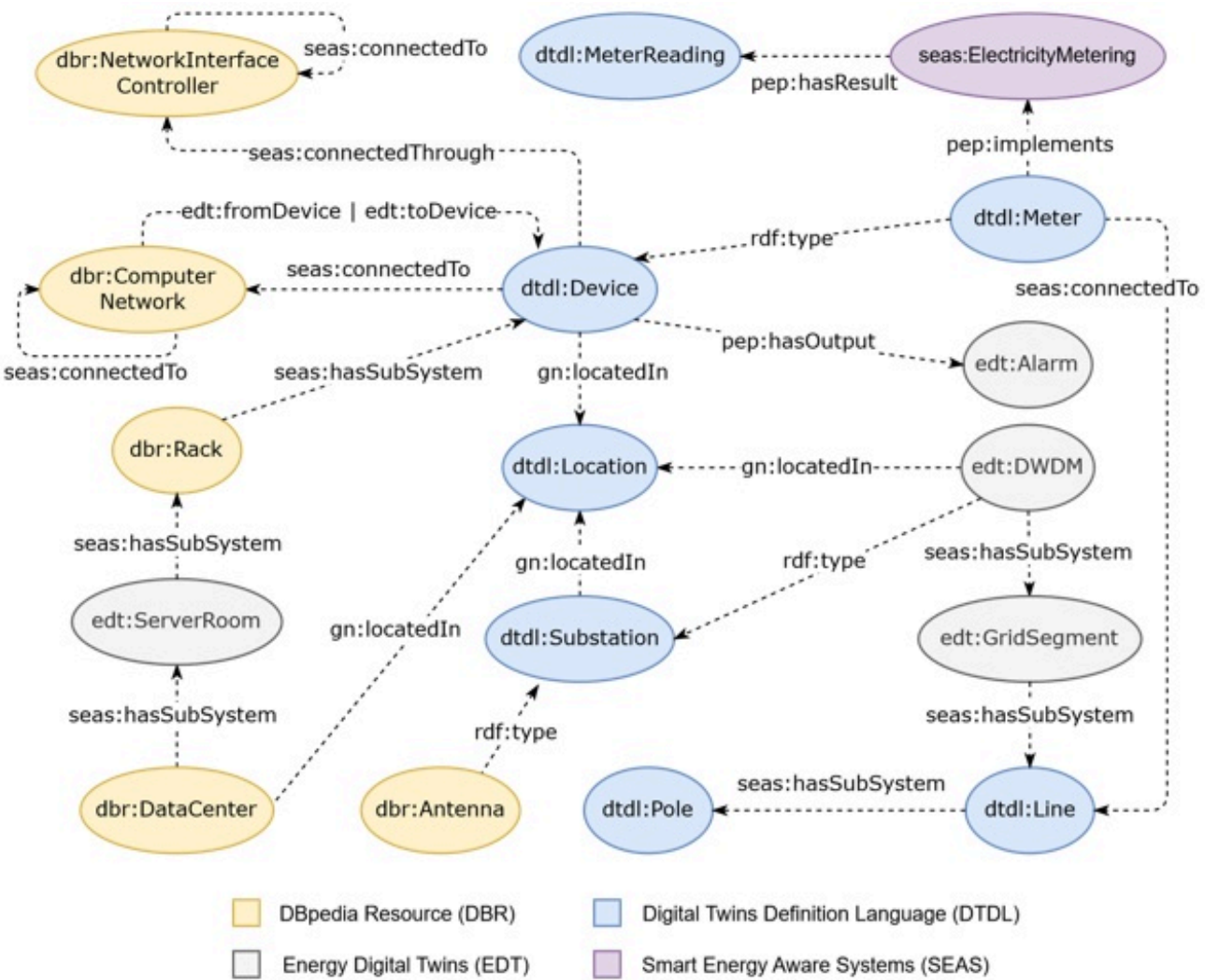


Figure 3. Reference graph-based data model.

All entities listed in [Table 1](#) are represented as RDF [10] resources, adhering to the principles of linked data [42]. This modeling choice ensures the method remains versatile and broadly applicable, while enabling semantic-driven annotation and analysis of the data. Specifically, a selection of widely recognized vocabularies serves as foundational ontologies. These capture core notions and relations relevant to energy infrastructure systems, communication networks, and metering processes.

Table 1. Entities defined in the reference data model.

Entity	Vocabulary	Description
Device	DTDL	An entity within the energy infrastructure that performs a specific function
Meter	DTDL	A physical asset utilized for measuring energy consumption and detecting events of interest
Electricity Metering	SEAS	The process of measuring and recording electrical energy usage
Meter Reading	DTDL	The recorded amount of electricity consumption via a meter

Alarm	EDT	A specific event triggered by a device indicating an issue
Location	DTDl	The physical location where a device or energy asset is situated
NIC	DBR	A Network Interface Controller (NIC) that represents a physical connection enabling interaction with external peripherals
Computer Network	DBR	A network of interconnected devices that can communicate with each other
Substation	DTDl	A part of the electrical generation, transmission, or distribution system
Antenna	DBR	A secondary substation designed for critical scenarios demanding high data availability in the smart grid
DWDM	EDT	A Dense Wavelength Division Multiplexing (DWDM) system supporting fiber-optic transmissions
Grid Segment	EDT	A segment of a power or communication network
Line	DTDl	A line that forms part of a telecommunication or power transmission network
Pole	DTDl	A utility pole that supports overhead power lines or other public utilities
DataCenter	DBR	A centralized facility for storing and processing large amounts of data related to the energy grid
Server Room	EDT	A temperature-controlled room for the continuous operation of servers in a data center
Rack	DBR	A physical container designed to house servers, networking devices, and other data center equipment

- Digital Twins Definition Language (DTDl) models were developed for energy grids (available at <https://github.com/Azure/pendigitaltwins-energygrid>, accessed on 8 September 2024) and for smart cities (available at <https://github.com/Azure/pendigitaltwins-smartcities>, accessed on 8 September 2024). These are specifically crafted to represent digital twin implementations, supporting the tracking and representation of assets in electrical networks and urban settings (Figure 3).
- The Smart Energy Aware Systems (SEAS) ontology (accessible at <https://w3id.org/seas/>, accessed on 8 September 2024) [43], which functions as a collection of straightforward foundational patterns that capture a range of engineering concepts and attributes within the broader energy domain (Figure 3).
- The Procedure Execution ontology (PEP) (found at <https://w3id.org/pep/>, accessed on 8 September 2024) supplies attributes for specifying processes, their outcomes, and outputs linked to measurement operations.
- Resources from DBpedia [44] (DBR) and the GeoNames (GN) ontology (located at <https://www.geonames.org/ontology/>, accessed on 8 September 2024) were used to represent components of the information technology network and spatial connections among asset positions (Figure 3).

In Figure 3, the various classes and associations from each vocabulary are marked using their corresponding RDF namespace abbreviations and distinguished by unique colors. Furthermore, a tailored vocabulary called EDT has been created to address specialized domain-specific classes (Figure 3) and characteristics absent from the listed ontologies.

The Asset Simulation component focuses on representing and simulating cyber-physical components of the infrastructure in a virtual form. This virtualization enables optimized resource assignment and oversight, ensuring the proper creation and ongoing support of digital counterparts throughout their lifecycles. It further promotes tight integration between real-world and simulated entities, creating a mutually reinforcing relationship that improves the architecture's overall performance.

The Data Management component serves as the central coordinator for the system, overseeing communications among diverse elements and functionalities. It additionally handles the extensive streams of information captured by on-site sensors, standardizes incoming data into uniform structures, and distributes it to downstream components. A primary difficulty in this environment is ensuring compatibility among the varying data structures used by heterogeneous sensor networks and information sources linked to the platform.

All information incorporated into the core knowledge repository is handled via a central data lake. This serves as a unified storage system capable of accommodating enormous quantities of raw, organized, and unorganized information required for examination or transformation. Consequently, sensor measurements can coexist alongside other relevant details, such as infrastructure, operational frameworks, or component inventories. In contrast to conventional data warehouses [45], a data lake accommodates inputs from multiple sources in diverse layouts and categories, enabling the ingestion and retention of heterogeneous collections without the need for mandatory preprocessing or standardization. Data lakes use expandable storage mechanisms such as cloud-based object stores or distributed file systems, providing cost-effective capacity compared with legacy data management systems. They also support flexible data investigation and insight generation through utilities and environments designed for on-the-fly querying, examination, and graphical representation of stored content.

Within the described framework, the data lake accommodates material originally residing in both Relational Database Management Systems (RDBMSs) and NoSQL systems. Notably, graph-oriented databases are recommended for capturing and interpreting intricate data linkages, particularly in evolving settings such as intelligent urban areas or electrical distribution networks. Information is persisted as knowledge graphs, where separate items are linked via precisely specified associations, yielding a coherent, meaning-focused perspective on the content. For operations that demand precise timing constraints—such as dynamic rankings or instantaneous analytics—in-memory storage solutions are also employed to cache frequently accessed data, thereby reducing disk operations and lowering response delays.

Data processing and analytics

The insights from the aforementioned module inform the Data Analysis tier. This tier delivers descriptive, predictive, and prescriptive analytical functions [46] targeted at representing and interpreting the condition and functional dynamics of energy infrastructures, as referenced in Ref. [47], drawing from sensor-gathered measurements and supplementary external inputs. When drawing on external sources, this tier can engage the Data Integration component to initiate retrieval operations, thereby enriching the knowledge repository with extra details sourced from cloud-hosted platforms.

Through techniques involving data mining and machine learning, the platform seeks to uncover practical insights and derive valuable understanding from the substantial data volumes produced by interconnected systems. As noted in Ref. [48], example use cases encompass:

- Fault identification and troubleshooting: detecting potential issues within the electrical network and tracking the condition and efficiency of components like transformers, generators, or substations.
- Proactive upkeep: utilizing structural and behavioral representations of grid equipment, examining time-series status information to forecast optimal maintenance timing, thereby avoiding breakdowns and reducing service interruptions.

- Demand prediction and equilibrium: estimating upcoming power requirements from records and distributing loads evenly across network segments to maintain reliability and avoid excess strain or bottlenecks.
- Demand-side management enhancement: refining programs that adjust consumption in response to high-demand intervals or variable pricing cues.

Based on required response immediacy and processing scale, the architecture supports two primary processing modes:

- Batch processing is suited to extensive datasets and aggregated operations. Information is typically accumulated, loaded, and handled to generate outputs on a scheduled basis, often spanning moderate to extended durations.
- Real-time processing is intended for scenarios that require swift responses and is executed by lightweight systems that handle high-frequency events. This is predominantly applied where numerous occurrences happen within brief windows.

Beyond these, the Data Analysis tier supports additional methodologies for replicating actual operational conditions via:

- What-if analysis [49]: modifications are applied to selected variables in a baseline case, followed by repeated simulation runs whose outcomes are retained for comparative evaluation and impact assessment.
- Goal-seeking analysis [50]: an intended outcome is specified through measurable indicators or Key Performance Indicators (KPIs), with the platform assisting in scenario construction to determine and assess viable paths toward achieving that target.

Data visualization and digital twins

Positioned at the highest level of the structure, the Data Visualization tier serves as the system's business intelligence front-end, laying the foundation for the fluid merging of real and virtual realms while advancing and optimizing the energy sector. Drawing on inputs from lower tiers, it constructs varied perspectives and simulated depictions, encompassing virtual replicas of key infrastructure parts (for instance, measurement instruments, grid connection points, or compact independent setups overseen by transmission towers).

The visualization facilities further enable multi-platform engagement through digital twin capabilities, delivering high-level summaries suitable for building web- and mobile-based interfaces. Through this approach, users can observe and augment virtual replicas in augmented or virtual reality environments, incorporating community-contributed inputs and sophisticated display mechanisms that rely on responsive human–computer interaction.

The Data Visualization tier delivers an extensive array of digital capabilities that enable entities to fully capitalize on the advantages of digital twins across numerous contexts. In essence, the outlined structure furnishes the subsequent instruments for engaging with and altering information:

- Three-dimensional representations of cyber-physical items, permitting users to retrieve, inspect, and adjust assets via their virtual 3D forms;
- Dynamic control panels and business intelligence tools that facilitate ongoing system oversight, displaying outcomes from analytics, critical metrics, and alerts triggered under hazardous circumstances;
- Natural User Interfaces (NUIs) that permit interaction via instinctive motions and actions [51]. These interfaces aim to improve usability by reducing onboarding time and enabling smooth integration between people and machines.

As illustrated in **Figure 4**, NUIs mark an advancement beyond traditional interaction models. Command-Line Interfaces (CLIs) rely on abstract textual instructions, offering strong scriptability that suits automated batch operations more than live human engagement. Graphical User Interfaces (GUIs) narrow the divide through pictorial analogies, icons, menus, and dimensional renderings of reality, thereby boosting efficiency in interactive settings. NUIs advance this by allowing direct, multi-modal engagement—potentially incorporating gestures, voice, gaze tracking, and similar channels. They mirror everyday human communication patterns, such as spoken dialogue, touch, and physical gestures, rendering interaction more natural, allowing individuals to focus on objectives while reducing mental effort in handling content through technological means. Consequently, this approach aligns especially well with intricate systems such as digital twins.

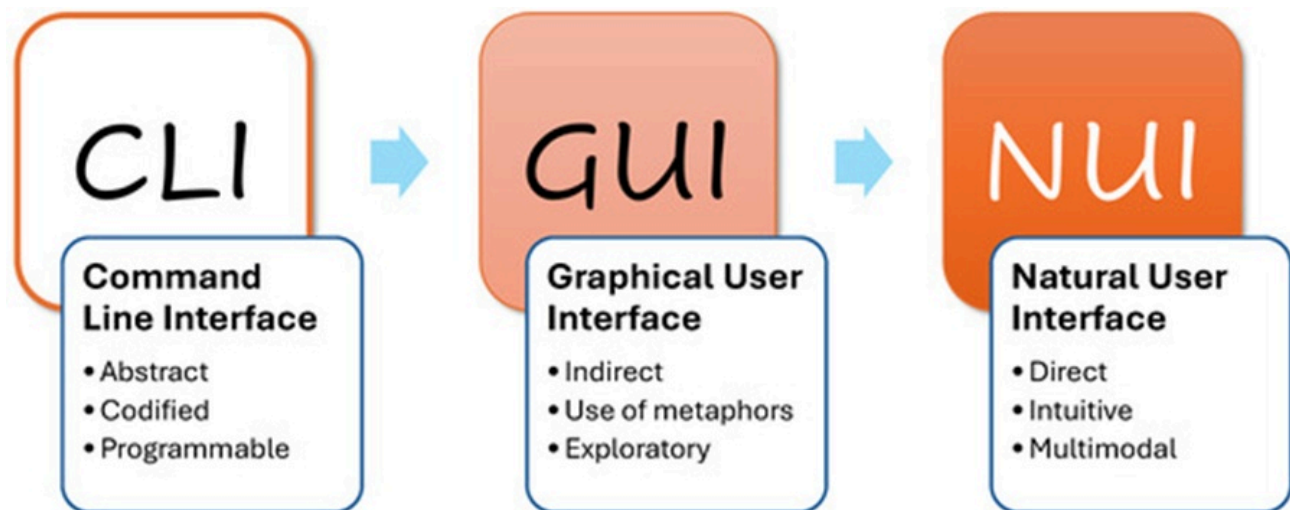


Figure 4. User interface paradigms.

The described digital twin platform incorporates targeted features that leverage generative artificial intelligence techniques to support decision-making and problem-solving, while also enabling natural user interfaces for interacting with the knowledge graph. A key element is the integration of a Retrieval-Augmented Generation (RAG) component [52], which merges the language understanding abilities of large language models with effective information retrieval mechanisms. This allows the system to pull pertinent details from extensive repositories on the fly during answer creation, resulting in responses that are better grounded and more precise. Although large language models excel across numerous domains, they often lack sufficient domain-specific awareness and can produce unreliable outputs when applied to real-world situations. The addition of retrieved external context through RAG significantly improves the model's grasp of relevant background, yielding more logical, contextually appropriate, and trustworthy results.

Furthermore, the system allows users to send targeted control instructions that trigger actions on the corresponding physical assets, enabling real-time management of infrastructure elements informed by insights from the digital counterpart.

The platform also offers a data export capability that produces datasets in RDF format. Through an interface to the central data lake, users can browse, search, refine, and extract selected portions of the stored information while respecting linked data guidelines [42].

System monitoring and security

Given the mission-critical role of electrical grid operations, safeguarding the platform's security and operational integrity is of utmost importance [53]. Real-time system monitoring focuses on identifying and addressing irregularities or potential malicious behavior as soon as they appear. At the same time, broader security measures aim to shield the grid from

cyberattacks, physical intrusions, and illicit access attempts. As illustrated in **Figure 2**, a dedicated cross-cutting layer has been incorporated to handle both operational and oversight functions, encompassing:

- **Access management:** implementation of multi-factor authentication combined with role-based permissions, utilizing the OpenID Connect (OIDC) protocol [54], to restrict access to confidential information and essential infrastructure elements—including substations, control centers, and data facilities—exclusively to verified individuals.
- **Ongoing surveillance:** a centralized orchestration service collects and evaluates logs along with network activity from diverse grid devices, empowering security personnel to spot degraded performance, unusual patterns, or emerging threats.
- **Vulnerability handling:** routine scanning for weaknesses, together with structured patch deployment workflows, to detect and correct flaws in hardware, firmware, and software across the grid ecosystem. This involves tracking vendor-issued security advisories and rapidly deploying necessary fixes to reduce exposure.
- **Incident handling strategy:** establishment of clearly defined protocols, assigned responsibilities, and escalation paths to contain and recover from security events efficiently.
- **Regulatory adherence:** sustained alignment with sector-specific norms and legal obligations, achieved through periodic security evaluations, enforcement of protective measures, and mandatory reporting of incidents to oversight bodies.

Case Study: Enhancing Grid Reliability through Digital Twins

To validate the feasibility and practical value of the proposed reference architecture (**Figure 2**), a working prototype has been constructed. This implementation connects to various manageable sensors and actuators, supporting multiple live data feeds that supply distributed decision nodes that operate either independently or under human oversight. The primary demonstration scenario, elaborated in the sections below, centers on oversight and control of a portion of the electrical infrastructure located in central Italy.

Prototype implementation

As shown in **Figure 5**, the prototype adopts a layered microservices design orchestrated via Kubernetes (<https://kubernetes.io>, accessed on 8 September 2024). Kubernetes is an open-source platform for automating the deployment, scaling, and operation of containerized workloads. Containers provide strong process isolation, allowing each service to execute independently without interference from others. This separation strengthens both security—by limiting the blast radius of faults or exploits—and overall system stability. Containerization also supports flexible, demand-driven scaling: additional instances can be launched or retired automatically in response to varying computational loads, optimizing resource efficiency and responsiveness.

The prototype employs a hybrid deployment model that combines on-site and cloud-hosted elements. This flexible arrangement enables organizations to customize the infrastructure footprint to meet specific constraints, including compliance obligations, data classification policies, and varying processing requirements.

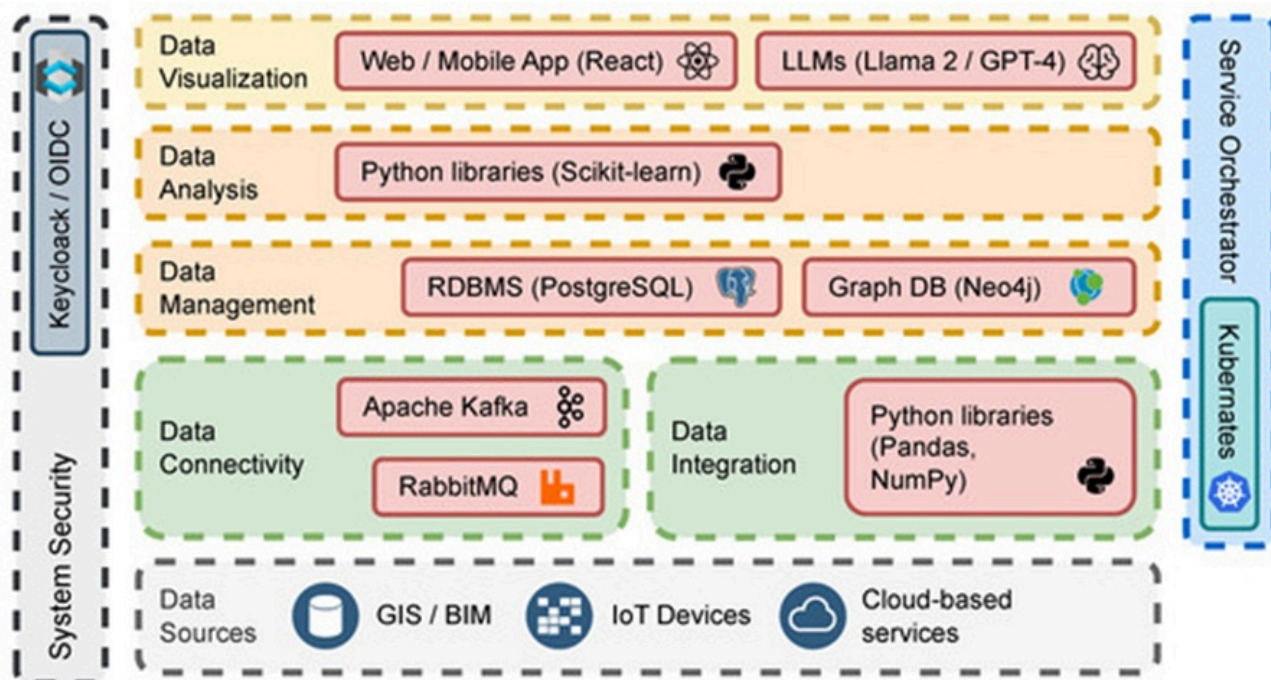


Figure 5. The structure of the implemented prototype.

Data on the electrical grid is gathered from multiple sources, including IoT sensors integrated into the monitoring network, Geographic Information Systems (GISs), Building Information Modeling (BIM) tools, and various cloud-hosted solutions dedicated to asset oversight and supervision. The EDT system incorporates specialized adapters to collect this information via two distinct interaction patterns: direct, real-time requests to REST API endpoints for immediate access; and indirect, event-driven delivery using message brokers or streaming frameworks (examples include Apache Kafka (<https://kafka.apache.org>, accessed on 8 September 2024) and RabbitMQ (<https://www.rabbitmq.com>, accessed on 8 September 2024)). Additionally, the platform embeds Keycloak (<https://www.keycloak.org>, accessed on 8 September 2024), a freely available Identity and Access Management (IAM) solution that handles user verification, identity federation, and detailed permission controls.

Data preparation steps—encompassing extraction, transformation, and loading (ETL)—are carried out via bespoke Python scripts that leverage the Pandas (<https://pandas.pydata.org>, accessed on 8 September 2024) and NumPy (<https://numpy.org>, accessed on 8 September 2024) packages for efficient handling and reshaping of information. These routines clean incoming raw inputs to ensure reliability, completeness, and compatibility with downstream tasks, addressing issues such as missing entries, extreme values, redundant entries, and other irregularities that might undermine data quality. The refined datasets are subsequently enriched with semantic annotations aligned with the core knowledge graph outlined in Figure 3, and the resulting data are persisted in Neo4j (<https://neo4j.com>, accessed on 8 September 2024), a specialized graph-oriented database optimized for storing and traversing highly connected data. Neo4j excels at representing complex interdependencies among items, enabling rapid relationship-based searches and sophisticated pattern matching across networks. Nodes and edges alike carry customizable attributes, supporting adaptable structures that suit varied information types while facilitating intuitive graphical exploration of the dataset. Supplementary descriptive details are maintained in PostgreSQL (<https://www.postgresql.org>, accessed on 8 September 2024), a robust, standards-compliant open-source relational database renowned for its stability and expandability. In this context, such metadata covers aspects such as asset geolocation, references to supporting materials or feeds, and application-specific labels (for instance, serial numbers, initiative references, or operator-assigned categories).

Analytical workflows rely on scikit-learn (<https://scikit-learn.org>, accessed on 8 September 2024), a widely used open-source machine learning toolkit offering straightforward yet powerful utilities for tasks spanning categorization, prediction,

grouping, and more. The library supports efficient computation on multi-core hardware and integration with distributed environments to accelerate model development and assessment. The outcomes of these computations enrich and update the underlying knowledge graph. The presentation tier includes all elements dedicated to improving usability through intuitive exploration and engagement with modeled objects. Notably, a browser-based interface was constructed with React (<https://react.dev>, accessed on 8 September 2024), a JavaScript library renowned for crafting dynamic, component-driven user experiences. Its modular design promotes component reuse throughout the application, simplifying the creation and upkeep of sophisticated front-ends by decomposing them into manageable, independently updatable pieces that enhance both productivity and long-term code quality.

For mobile accessibility across iOS and Android, React Native (<https://reactnative.dev>, accessed on 8 September 2024) was employed, reusing portions of web-oriented components to streamline cross-platform development and reduce effort compared to fully native implementations for each operating system.

Lastly, the architecture features an interactive conversational agent powered by a large core language model. This agent uses retrieval-enhanced generation techniques to facilitate troubleshooting, in-depth inquiries, and reasoned assistance, delivering precise, context-aware replies drawn directly from the stored knowledge graph. Two prominent large language models were assessed for powering this assistant: GPT-4 [55] and Llama 2 [56]. GPT-4 (<https://openai.com/research/gpt-4>, accessed on 8 September 2024), created by OpenAI, is a versatile multi-modal system that can interpret and generate content across text and images, yielding natural, user-friendly language. Various configurations of GPT-4 exist, differing in performance levels and associated costs. Llama 2 (<https://llama.meta.com/llama2>, accessed on 8 September 2024) comprises a family of openly released pre-trained and instruction-tuned models from Meta, spanning 7 billion to 70 billion parameters, offering a cost-free, customizable option compared to proprietary alternatives like GPT-4—ideal for scenarios prioritizing data privacy and tailoring. Both solutions deliver strong performance across language comprehension, content creation, programming support, query resolution, and workflow automation, addressing a broad spectrum of use cases. A side-by-side evaluation of GPT-4 and Llama 2 is presented in Table 2, summarizing their principal attributes to guide selection based on project-specific priorities and constraints.

Table 2. Comparative table of adopted LLMs.

Feature	Llama 2	GPT-4
License	Llama 2 community license	Commercial
Input	Text only	Text and images
Model size	7 B, 13 B, 70 B parameters	~1.76 T parameters
Model customization	Highly customizable via RAG or fine-tuning	Limited to selected organizations
Model updates	Periodic updates with potential new features	Updates are not immediately accessible to end users
Tokens per prompt	Up to 2048	Up to 8192
Integration with other systems	LLM backend required	Through RESTful API calls
Conversational skills	Fluent and natural-sounding responses for interactive chat	Immersive conversations integrating multiple forms of media
Privacy concerns	Data processing occurs on-premises, reducing exposure to third-party entities	Data processed on external servers, potentially subject to data breaches or misuse

Privacy represents a key priority in the proposed system. To ensure robust protection, encryption is applied to data at rest and in transit, safeguarding all confidential information throughout the operation. Crucially, no data is ever sent to external third-party providers: all processing related to machine learning, analytics, and visualization takes place entirely on local infrastructure.

Although the architecture supports integration with GPT-4, the team deliberately selected an on-premises deployment of Llama 2 to strengthen privacy guarantees. This local large language model serves as the foundation for the virtual assistant and has been fine-tuned specifically for the demands and characteristics of energy grid use cases.

The virtual assistant employs a Retrieval-Augmented Generation (RAG) strategy, as illustrated in [Figure 6](#). It combines user queries with information stored in the knowledge graph (KG) while preserving data confidentiality and avoiding dependence on remote servers.

In a basic setup without RAG, the pre-trained language model would answer solely based on its general training data, delivering only generic responses that lack awareness of local context. With RAG enabled, however, the user’s question is first transformed into an embedding vector. This vector enables similarity-based retrieval of pertinent facts from the Neo4j database — for instance, recorded incidents, details on energy assets, or metering readings.

The most relevant retrieved items are then incorporated to enrich the original query context, supplying domain-specific facts and additional viewpoints. The system also provides a variety of prompt templates to refine and optimize the language model’s conversational behavior.

Once the query context has been expanded, the augmented input is forwarded to the LLM, which generates coherent, natural-language answers deeply informed by the relevant grid-specific data.

Additionally, the platform remains adaptable: users can opt to interact with cloud-hosted language models whenever their privacy requirements permit. This flexible, hybrid model lets organizations benefit from the superior conversational capabilities, greater computational capacity, and broader access to knowledge offered by cloud-based LLMs, while still maintaining strict control over sensitive or proprietary information.

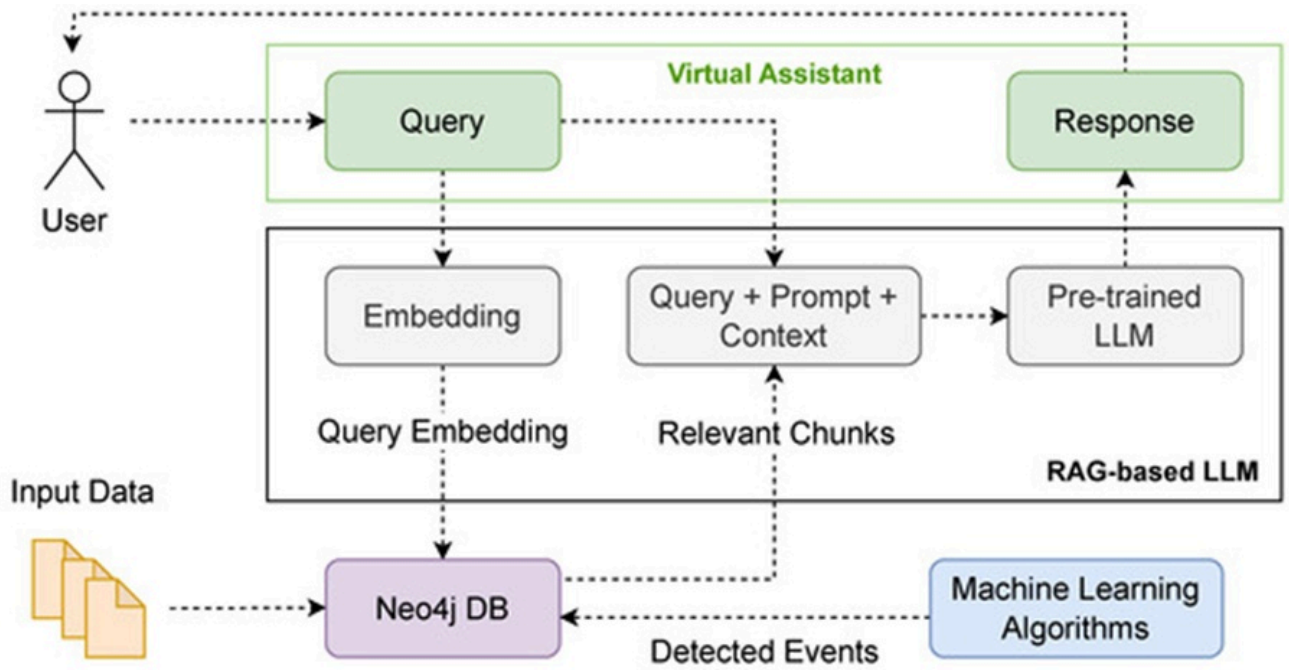


Figure 6. Retrieval augmented generation architecture.

Core performance metrics

To rigorously evaluate the advancements delivered by the suggested techniques, the real-world application also serves as a reference point for examining how the combination of digital twins (DTs), data-driven analytics, large language model (LLM)-driven advisory tools, and RDF-structured knowledge graphs improves essential operational elements. This is achieved by contrasting the introduced capabilities against conventional methods employed in the baseline setting.

In particular, the suggested solution has been examined across multiple assessment criteria aligned with the objectives :

- (RA1) Improving grid oversight via digital twins and data-centric techniques:
 - a. Seamless incorporation of live data streams to deliver ongoing status reports and evaluations of grid components, while rendering asset conditions in a clear, user-friendly manner;
 - b. Accurate simulation of actual physical assets and operational dynamics through digital twin modeling;
 - c. Proactive identification and mitigation of faults or deviations that might impair grid reliability;
 - d. Establishment of a synchronized virtual counterpart for physical grid elements, reflecting current operational metrics and supporting remote intervention capabilities.
- (RA2) Utilizing large language models to aid decision-making:
 - a. Application of foundation models adapted to specialized tasks involving the unique vocabulary, jargon, and ideas prevalent in the energy industry;
 - b. Comprehension and production of text concerning power systems, enabling responses to intricate, multi-dimensional inquiries about grid operations;
 - c. Delivery of clear, understandable rationales for outputs and recommendations, thereby boosting user confidence and comprehension;
 - d. Embedding of LLM capabilities within routine operational processes and automated systems in energy facilities to enable seamless collaboration with complementary data-handling solutions.
- (RA3) Combining RAG strategies with structured knowledge graphs:
 - a. Standardized ontology for labeling information in the knowledge graph, guaranteeing faithful depiction of real-world connections and facilitating pertinent retrieval;
 - b. Consolidation of inputs from diverse origins (such as IoT devices, relational databases, and event logs) into the knowledge graph, broadening the reach of the RAG mechanism;
 - c. Generation of highly pertinent, sector-tailored information that boosts response precision and lowers hallucination risks.

Application scenario evaluation

This portion outlines the deployment of the developed system to improve performance in the selected practical environment. The platform oversees a substantial portion of the electricity distribution grid in central Italy, encompassing roughly 75,000 km of power lines along with 900 substations. It specifically covers the municipal distribution network serving Rome, which features 70 primary substations and over 30,000 km of cabling.

In essence, the EDT platform serves as an all-encompassing solution for supervising and optimizing grid assets, incorporating predictive maintenance strategies driven by readings from intelligent metering devices.

The primary objectives of this deployment can be condensed as follows: “Smart metering equipment is installed throughout the grid to track energy usage continuously. Platform operators leverage the resulting data to anticipate possible device malfunctions or efficiency drops. Targeted notifications and guidance are transmitted to field technicians via a dedicated mobile interface, equipping them with essential details to resolve problems swiftly and effectively.”

The standard operational sequence is outlined below, with each step or feature linked to the corresponding research goals:

- Platform administrators log in to the web-based interface and view the complete grid layout, including generation facilities, transformer stations, transmission lines, and support structures, as shown in **Figure 7**. Geographic positions for each substation are retrieved from OpenStreetMap (OSM) [57] using queries to the Overpass API (https://wiki.openstreetmap.org/wiki/Overpass_API) (RA1.a). Infrastructure-related information is persistently maintained in the core knowledge graph (RA3.a, RA3.b).
- When machine learning models flag irregularities, a prominent red marker appears on the map to indicate suspected problems at the affected site (RA1.c).
- Selecting any element reveals in-depth attributes, associated characteristics, and linked alerts for the equipment or facility. The interface also lists executable commands, allowing users to engage remotely with connected control devices. These commands facilitate direct manipulation of physical hardware, such as restarting units, modifying parameters, or launching repair sequences (RA1.a, RA1.d).
- The platform incorporates interactive 2D representations (**Figure 8a**), with clickable regions that unlock additional views or layers of the selected asset (RA1.b).
- Operators can manipulate 3D renderings (**Figure 8b**) via intuitive gesture controls in a natural user interface (NUI), enabling rotation, magnification, component isolation, and highlighting areas requiring attention. These models update in real time, offering an engaging, lifelike way to inspect asset health, structural details, and positional context (RA1.b).
- A green icon in the lower-right area of the screen activates the virtual assistant panel (**Figure 7**). Through straightforward natural-language inputs (for example, “Display full details on the latest alerts for substation X”), users obtain targeted insights into observed equipment and incidents, including alert categories, occurrence times, affected parts, and proposed corrective measures (RA2.a, RA2.b).
- Upon examining alert details, if an urgent or precautionary intervention is warranted, administrators trigger technician dispatch. The system auto-generates a structured alert summarizing the problem severity and required response timeline, then routes it to the assigned field personnel (RA2.c, RA2.d).
- Field technicians receive dispatches on their mobile devices, complete with precise asset positioning (**Figure 9a**) and step-by-step repair instructions (**Figure 9b**), thereby reducing downtime and maintaining optimal grid performance (RA1.a).
- After finishing the repair, technicians mark the task as resolved and attach observations or supplementary remarks. The platform instantly logs the closure, refreshes relevant records, and informs supervisory staff or coordinators (RA1.a, RA2.d).
- Insights gathered from completed interventions are captured and integrated into the RDF-based knowledge graph (RA3.a), enriching it with practical experience and situational knowledge valuable for subsequent evaluations (RA1.c).

Periodic reviews of accumulated data help pinpoint patterns, refine maintenance protocols, and enhance the quality of outputs from RAG-enhanced mechanisms (RA3.c).

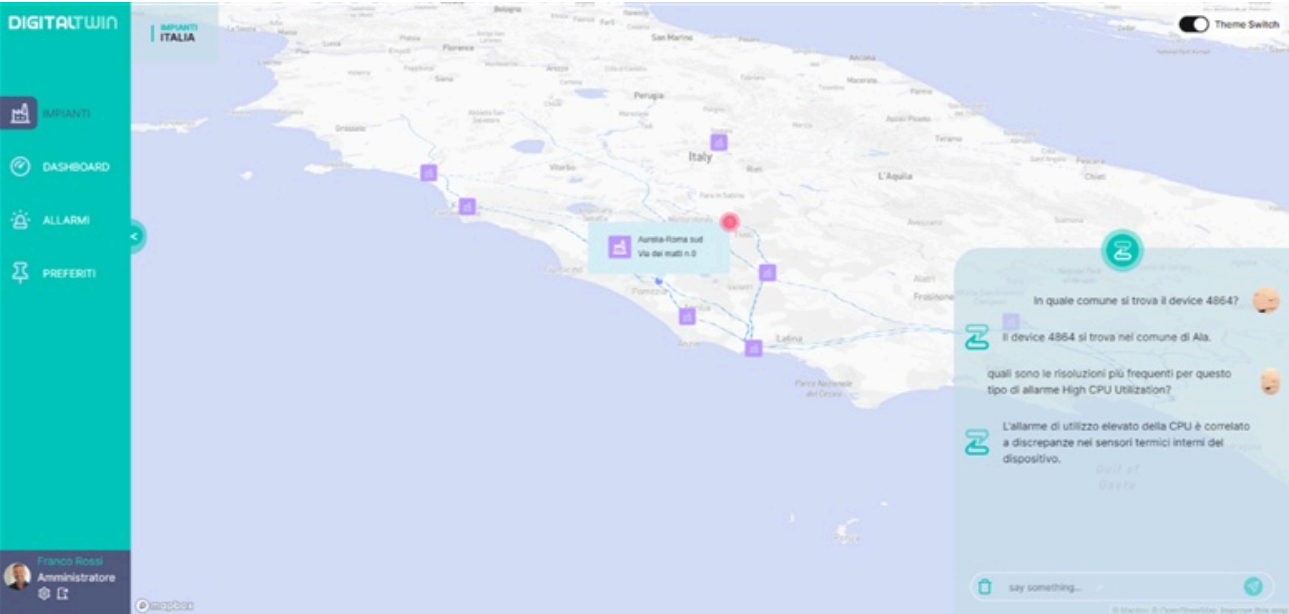


Figure 7. Digital representation of the energy grid.

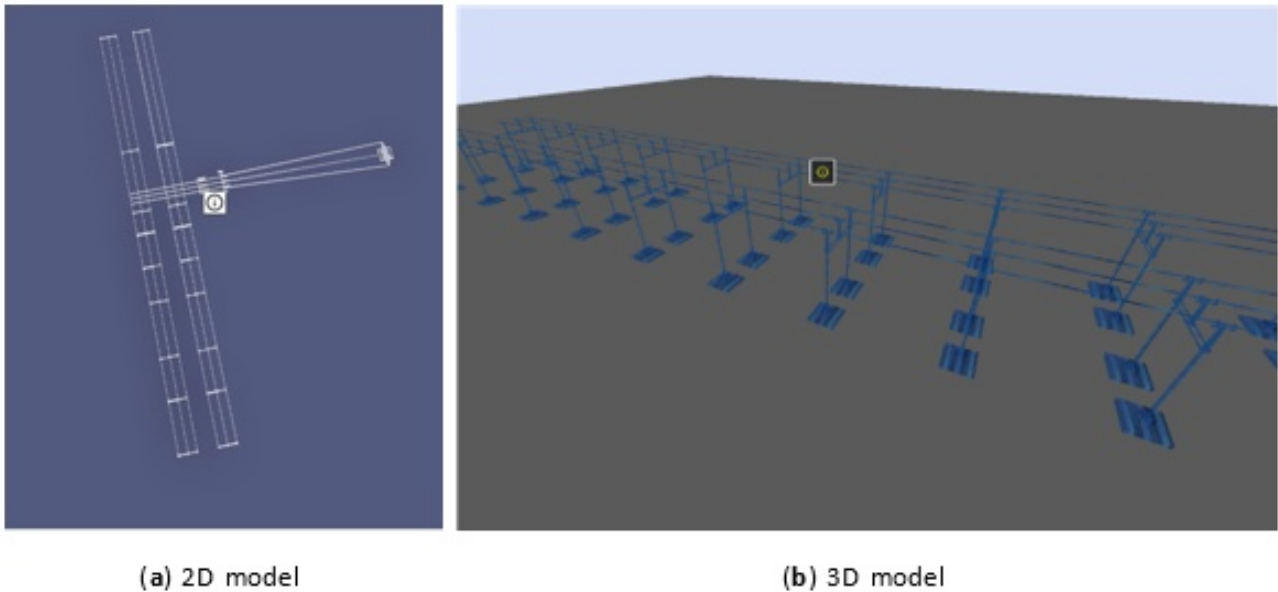


Figure 8. Digital-twin assets of the energy plant.

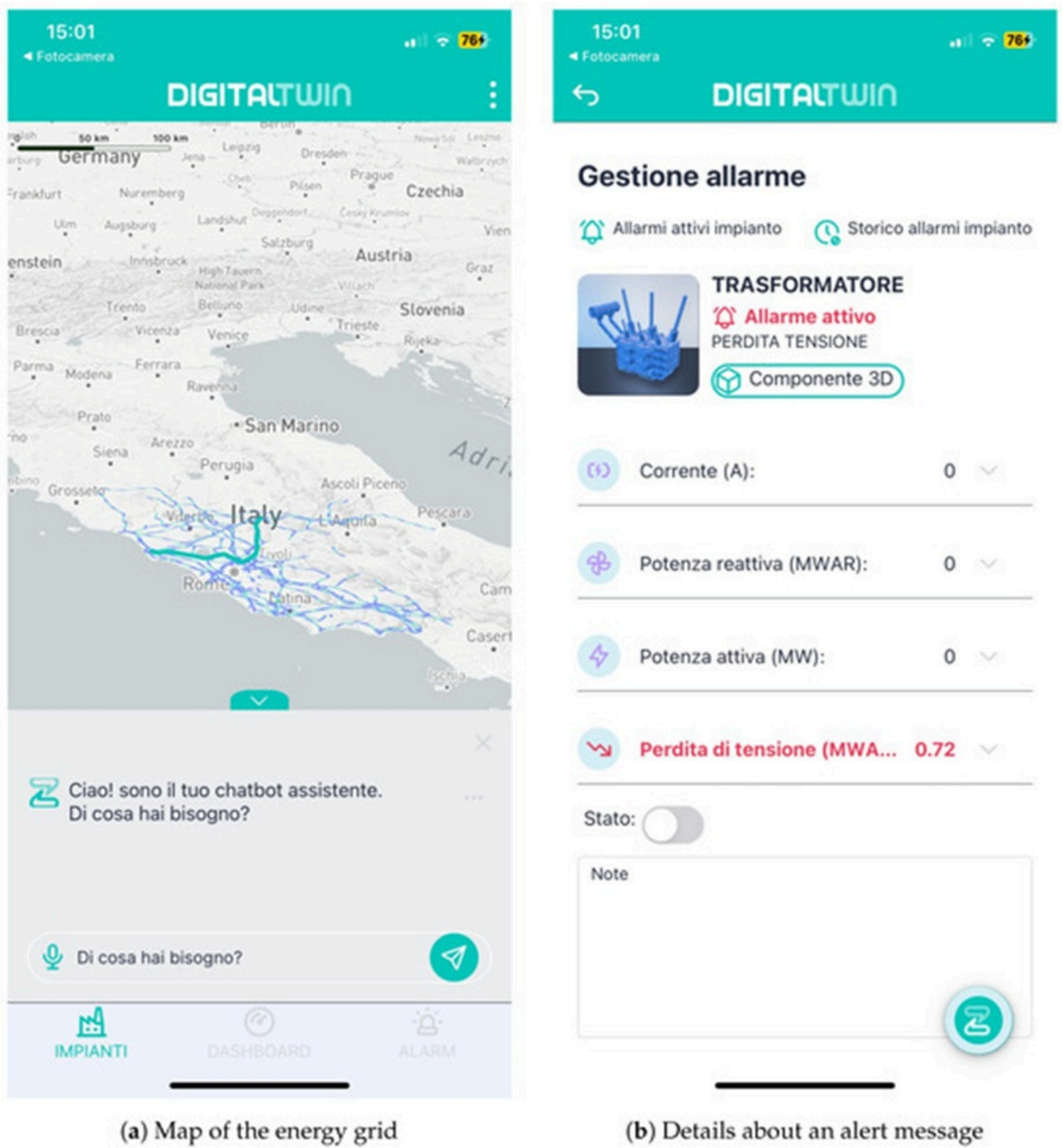


Figure 9. Screenshots of the mobile application.

This part evaluates the proposed approach by comparing it with existing state-of-the-art solutions. It does so through a detailed comparison with pertinent prior studies and by examining open issues and potential future directions.

Platform evaluation

Table 3 provides a side-by-side comparison of various Digital Twin (DT) frameworks, drawing on the research efforts outlined in the framework introduced in the present paper. The table is organized to assess the degree of support provided for an extensive range of capabilities, grouped into four primary categories:

Table 3. Comparative analysis of DT frameworks (✓: supported, ✗: not supported, *: partial support).

EDT platforms	[27]	[28]	[29]	[30]	[31]	[33]	[34]	This Work
Infrastructure								
Cloud Infrastructure	✓	✓	✓	✗	✓	✓	✓	✓
Edge Computing	✗	✓	✓	✗	*	✗	✓	✓
IoT Sensor Networks	✓	✓	✓	✓	✓	✓	✓	✓
Data Lake	*	*	✓	✓	✓	*	*	✓
Physical Asset Control	✓	*	*	*	✓	*	✓	*
Intelligence								
Real-time Data Processing	✓	✓	✓	✓	✓	✗	✓	✓
Data Analytics and ML	✓	✓	✓	*	✓	✓	✓	✓
Asset Simulation	*	✓	✓	✓	✓	✗	✓	*
Scenario Forecasting	*	*	✓	✓	✓	✓	✓	*
Knowledge Graphs	✗	✗	✗	✗	*	✗	✗	✓
LLM RAG	✗	✗	✗	✗	✗	✗	✗	✓
Cybersecurity								
System Auditing	*	✗	✓	✗	✗	✗	✗	✓
Access Control	✓	✗	✓	✗	✗	✗	✗	✓
Visualization								
Interactive GUI	✓	✗	✓	✓	✓	✓	✗	✓
GIS Maps	*	✗	✗	*	*	✓	*	✓
3D Models	✗	✗	✗	✗	*	✓	*	✓
Natural User Interface	✗	✗	✗	✗	✗	✗	✗	✓

- Infrastructure evaluates the extent to which each Digital Twin framework utilizes computing resources and coordinates real-world physical elements.
- Data Intelligence focuses on the platform's ability to perform data analysis and derive meaningful insights from data generated within the system.
- Cybersecurity examines the protective mechanisms each framework offers, including threat detection, monitoring, and safeguards to maintain data integrity and confidentiality in Digital Twin settings.

Visualization describes the human–machine interaction approaches adopted by each proposal, encompassing features such as interactive graphical user interfaces, geospatial (GIS) mapping, three-dimensional modeling, and natural user interfaces (NUIs).

The table uses a set of symbols to indicate the degree of support provided by each compared framework: a check mark (✓) signifies complete support, a cross (✗) indicates no support, and a star (*) denotes partial implementation or features planned for future inclusion.

Notably, none of the existing frameworks included in the comparison offer Retrieval-Augmented Generation (RAG) functionality that combines knowledge graphs (KG) with large language models (LLM). Such integration would allow conversational natural user interfaces to support the manipulation of energy digital twin (EDT) assets and deliver knowledge-enriched analytical responses. Moreover, most of the reviewed platforms pay limited attention to cybersecurity measures—an area of critical importance when overseeing essential societal infrastructure such as electrical grids.

The D-Arc framework [27] provides strong coverage across numerous essential EDT characteristics, particularly in cloud-based infrastructure, IoT sensor integration, real-time data handling, advanced analytics, machine learning, asset modeling, predictive scenario simulation, access management, and interactive graphical interfaces. It shows partial capabilities in edge computing and system auditing, though these are not thoroughly elaborated. On the other hand, it lacks explicit provisions for direct physical asset actuation, knowledge graph- or LLM-driven reasoning, geospatial mapping, and natural user interfaces—areas that remain open for improvement.

C3-Flow [28] concentrates primarily on formal mathematical modeling and distributed computation of EDT representations. As a result, several practical features expected in mature production systems receive little or no explicit attention, including cybersecurity provisions, data storage design, and user interface solutions. While physical asset control seems feasible in principle, it is not directly demonstrated in that work [28].

The architecture presented in Ref. [29] for electric cyber-physical systems (ECPSSs) relies on Amazon Web Services to implement a secure cloud-edge architecture that ensures encrypted, authenticated communication. While this addresses basic security needs, more advanced elements such as knowledge graphs, large language models, and sophisticated interactive interfaces fall outside the scope of the experiments, which prioritize core operational reliability and performance optimization over broad feature coverage.

In Ref. [30], the focus is on integrating EDT concepts with IoT deployments, such as Phasor Measurement Units (PMUs) and Intelligent Electronic Devices (IEDs), to improve real-time data handling, asset simulation, and forecasting in power system control environments. The study mentions graphical interfaces to support operator decision-making. It implies some level of physical actuation via smart substations and control logic. Yet, it provides no detailed discussion of data mining techniques, RAG-based approaches, security safeguards, or geospatial visualization incorporating 3D elements.

The Digital Twin proposal for the power sector in Ref. [31] stands out as one of the most comprehensive among the reviewed works, incorporating smart sensors, 5G connectivity, and large-scale data analytics. It briefly references the use of a knowledge graph within the infrastructure and employs GIS maps and 3D models derived from laser scanning to visualize underground pipe systems. Regrettably, cybersecurity considerations are not addressed in that publication.

The EDT system described in [34] aims to enhance predictive maintenance and fault identification in electricity distribution networks. It effectively leverages cloud, edge, and IoT layers for data acquisition and processing, integrates machine-learning-driven prognostics, and supports physical device control. While interactive interfaces are suggested, along with references to GIS mapping and 3D visualization, these elements remain underdeveloped in the presented solution.

As summarized in **Table 3**, the framework introduced in this paper delivers a comprehensive Digital Twin architecture tailored to the energy infrastructure sector. It distributes computational workloads effectively across the full cloud-to-edge

spectrum and manages large-scale, geographically dispersed sensor deployments. A scalable data lake architecture supports both structured and unstructured data for sophisticated analytics. In contrast, direct physical asset control enables real-time monitoring and remote actuation of connected equipment. Core operational features—real-time processing, analytics, and machine learning—are fully incorporated to improve energy system performance. These are further strengthened by innovative RAG mechanisms that leverage ontologies and knowledge graphs to augment foundation large language models, delivering precise, context-aware decision support. The prototype also embeds strong cybersecurity practices, such as multi-factor authentication, role-based permissions, and ongoing anomaly detection with real-time response capabilities. Finally, visualization and user interaction are cornerstones of the platform, with full support for GIS-based mapping, 3D modeling, and a natural user interface powered by LLMs. This enables intuitive, chat-style interaction for system management and information access, as validated through the prototype implementation.

Challenges and opportunities

The combination of information and communication technologies (ICT) with digital twins in energy management applications presents a wide array of challenges and opportunities, spanning both technical and human dimensions [58]. Drawing on insights from the framework's design and the implementation of the presented use case, the following sections outline the primary elements that most strongly affect the deployment and uptake of digital twins in energy infrastructures.

Compatibility with existing infrastructure: A large portion of current energy systems still rely on Supervisory Control and Data Acquisition (SCADA) systems for real-time oversight and control of operational processes. These systems often rely on vendor-specific protocols and technologies, creating significant barriers to integration with contemporary software solutions. Efforts to develop uniform communication standards and intermediary layers, such as the Open Platform Communications Unified Architecture (OPC-UA) [59], aim to facilitate smoother connectivity between legacy equipment and digital twin environments. Nevertheless, older systems produce enormous volumes of heterogeneous data that vary in frequency, structure, and format. Harmonizing information from such varied sources demands considerable effort, including mapping, standardization, and conversion steps to ensure uniformity and precision. Feeding machine learning algorithms with erroneous or inconsistent data can severely undermine the reliability of forecasts and analytical outputs, potentially leading to misguided recommendations and diminished confidence in data-supported decisions.

Furthermore, numerous legacy installations run on outdated hardware and software, leaving them vulnerable to significant cybersecurity risks, including unauthorized intrusions, data leaks, and malicious attacks. Systems running discontinued or unsupported software become prime targets for exploitation, as newly identified flaws remain unaddressed, enabling attackers to breach defenses. Many of these older platforms also employ antiquated verification methods and permission structures that fail to incorporate current protective measures—such as robust encryption standards, multi-factor verification, and role-based permission systems—thereby increasing the risk of data exposure during transmission or storage. Safeguarding the confidentiality, accuracy, and availability of information originating from legacy sources is vital for protecting essential infrastructure and organizational assets. The shift toward modern security paradigms, including defense-in-depth [60] and zero-trust [61] models, offers promising avenues for mitigating longstanding weaknesses and strengthening the overall robustness of energy networks.

Human-centric considerations in deploying digital twins: Achieving effective adoption of digital twin (DT) solutions in energy operations depends not solely on engineering advancements but also on a thorough understanding of human factors, including psychological, interpersonal, and structural factors that shape how people engage with and rely on technological tools [62]. Recognizing how innovative methods can be aligned with everyday human tasks is fundamental to realizing the full benefits of DTs. As highlighted in [63], digital twins extend beyond tools designed exclusively for specialists; they can actively support judgment calls, task execution, and adaptation to changing circumstances, thereby fundamentally altering the duties performed by personnel. In this shifting environment, operators are expected to work in close partnership with DT systems, monitoring their outputs and intervening as needed for particular activities. Particular

attention should therefore be given to the following human-related aspects when applying DTs to energy facilities, as they are essential for promoting acceptance, optimal usage, and long-term viability:

The quality of the user interface significantly influences how effectively individuals can engage with the DT platform. Interfaces should be straightforward and approachable, delivering pertinent data, interpretations, and forecasts in formats that align with users' natural thought patterns, thereby enabling faster, more precise responses to fluctuating energy conditions. Customizing these interfaces to match the knowledge levels and requirements of energy sector professionals helps streamline adoption and reduces the time needed to become proficient [64].

Building confidence in energy digital twin (EDT) systems is critical to widespread acceptance. Operators need assurance that the digital representations produced are dependable and precise. Greater emphasis should be placed on research into verification techniques, clear disclosure of model limitations and uncertainties, and mechanisms for incorporating user input, all of which help demystify sophisticated models and foster greater reliance, even among less technical users [65].

The intricate nature of EDT platforms typically requires thorough training and ongoing skill development for operators and managers. Users must develop a solid understanding of the system's features and potential. Sustained educational efforts enable personnel to fully exploit the technology, reach well-informed conclusions, and resolve operational challenges efficiently [66].

Given that DTs handle extensive datasets, robust cybersecurity practices are indispensable to ensure the ethical handling of information, safeguard personal privacy, and comply with applicable legal standards. In particular, energy-oriented digital twins should be classified as vital systems, requiring comprehensive protection of confidentiality, data integrity, and system availability, as well as privacy considerations for individuals and the whereabouts and condition of critical components [67].

Conclusion

The present work proposes an innovative framework for energy digital twins that merges data-driven techniques with structured domain knowledge. It incorporates: (i) continuous collection of data streams from monitoring sensors; (ii) real-time data analysis and machine learning techniques for predictive upkeep and performance evaluation; (iii) a knowledge graph grounded in ontology that integrates labeled observations, established domain expertise, and outputs from machine learning processes; (iv) a retrieval-augmented generation mechanism that leverages the knowledge graph to augment a foundational large language model, enabling a conversational AI assistant to support users in decisions concerning energy asset efficiency and maintenance strategies; (v) an intuitive interface that blends dialogue-based interaction with visualization of geospatial data through 3D representations of infrastructure elements within a geographic information system. These capabilities are underpinned by a unified, scalable architecture built on cloud and edge microservices, featuring distinct functional tiers and dedicated cross-cutting layers for device connectivity, data harmonization, and security.

A working prototype of this framework has been developed using readily available commercial components and applied to a practical scenario involving a portion of the electricity distribution network in central Italy. At present, the system operates exclusively in Italian. It has been developed primarily to serve energy companies operating within Italy, where numerous supporting documents and regulatory references are available only in Italian. Using the native language improved text processing accuracy and made the platform's features more accessible and comprehensible to its intended audience. That said, plans are underway to introduce multilingual capabilities as part of ongoing development, broadening the system's reach to users across different language regions and improving its overall inclusivity.

Looking ahead, efforts will focus on maturing the prototype into a fully realized framework and conducting field trials in operational settings to gather practical insights. These evaluations will help examine key implementation factors, such as:

(i) system performance and capacity to scale; (ii) compatibility and ease of connection with diverse legacy infrastructures; (iii) quality and effectiveness of user interactions; (iv) levels of acceptance and trust among operators and stakeholders. Additional research directions include advancing machine learning techniques—particularly by integrating explainable artificial intelligence (XAI) methods to bolster credibility—refining semantic structures to enhance retrieval-augmented generation performance in conversational agents, and expanding functionalities related to asset management, control, and simulation of operational scenarios within the overall framework.

Acknowledgements

None

Conflict of interest

Davide Marche and Marianna Notarnicola are employed by Lutech S.p.A.; Saverio Ieva, Giuseppe Loseto, Michele Ruta, and Floriano Scioscia are partners in donkeyPower S.r.l. The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Financial support

The research was supported by the Digital Enterprise grant (grant number NIL6S28), co-funded by Lutech S.p.A. and the European Regional Development Fund for the Apulia Region 2014/2020 Operating Program.

Ethics statement

None

Received: 15 Mar 2024

Revised: 26 Apr 2024

Accepted: 08 Jun 2024

Published online: 20 September 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

Joseph A, Balachandra P. Smart grid to energy internet: A systematic review of transitioning electricity systems. *IEEE Access*. 2020;8:215787-805.

Cronin J, Anandarajah G, Dessens O. Climate change impacts on the energy system: A review of trends and gaps. *Clim Change*. 2018;151(1):79-93.

Schleussner CF, Rogelj J, Schaeffer M, Lissner T, Licker R, Fischer EM, et al. Science and policy characteristics of the Paris Agreement temperature goal. *Nat Clim Change*. 2016;6(9):827-35.

Dileep G. A survey on smart grid technologies and applications. *Renew Energy*. 2020;146:2589-625.

Statista. Smart Grid Market Value Worldwide 2022–2028. 2023. Available at: <https://www.statista.com/study/111848/smart-grids-worldwide/>. Accessed September 8, 2024.

Alotaibi I, Abido MA, Khalid M, Savkin AV. A comprehensive review of recent advances in smart grids: A sustainable future with renewable energy resources. *Energies*. 2020;13(23):6269.

Alasali F, Itradat A, Abu Ghalyon S, Abudayyeh M, El-Naily N, Hayajneh AM, et al. Smart Grid Resilience for Grid-Connected PV and Protection Systems under Cyber Threats. *Smart Cities*. 2023;7(1):51-77.

Do Amaral J, Dos Santos C, Montevechi J, De Queiroz A. Energy Digital Twin Applications: A Review. *Renew Sustain Energy Rev*. 2023;188:113891.

Manickam R, Vollmar J, Prabhakar GM. User Experience–Digital Twin Maturity Model (UX-DTMM). In: *Proceedings of the International Conference on Research into Design*. Bangalore, India; 2023:877-89.

Cyganiak R, Wood D, Lanthaler M. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation. 2014. Available at: <https://www.w3.org/TR/rdf11-concepts/>. Accessed September 8, 2024.

Hossein Motlagh N, Mohammadrezaei M, Hunt J, Zakeri B. Internet of Things (IoT) and the energy sector. *Energies*. 2020;13(2):494.

Ahmadzadeh S, Parr G, Zhao W. A review on communication aspects of demand response management for future 5G IoT-based smart grids. *IEEE Access*. 2021;9:77555-71.

Entezari A, Aslani A, Zahedi R, Noorollahi Y. Artificial intelligence and machine learning in energy systems: A bibliographic perspective. *Energy Strategy Rev*. 2023;45:101017.

Omitaomu OA, Niu H. Artificial intelligence techniques in smart grid: A survey. *Smart Cities*. 2021;4(2):548-68.

Cao K, Liu Y, Meng G, Sun Q. An overview on edge computing research. *IEEE Access*. 2020;8:85714-85728.

Minh QN, Nguyen VH, Quy VK, Ngoc LA, Chehri A, Jeon G. Edge Computing for IoT-Enabled Smart Grid: The Future of Energy. *Energies*. 2022;15(17):6140.

Arcas GI, Cioara T, Anghel I, Lazea D, Hangan A. Edge Offloading in Smart Grid. *Smart Cities*. 2024;7(2):680-711.

Ruta M, Scioscia F, Loseto G, Pinto A, Di Sciascio E. Machine learning in the Internet of Things: A semantic-enhanced approach. *Semant Web*. 2019;10(2):183-204.

Dhaou IB. Design and Implementation of an Internet-of-Things-Enabled Smart Meter and Smart Plug for Home-Energy-Management System. *Electronics*. 2023;12(19):4041.

Loconte D, Ieva S, Pinto A, Loseto G, Scioscia F, Ruta M. Expanding the cloud-to-edge continuum to the IoT in serverless federated learning. *Future Gener Comput Syst*. 2024;155:447-462.

Aljadani N, Gazdar T. A novel security architecture for WSN-based applications in smart grid. *Smart Cities*. 2022;5(2):633-49.

Xu Z, Salehi Shahraki A, Rudolph C. Blockchain-Based Malicious Behaviour Management Scheme for Smart Grids. *Smart Cities*. 2023;6(6):3005-31.

Yu W, Patros P, Young B, Klinac E, Walmsley TG. Energy digital twin technology for industrial energy management: Classification, challenges and future. *Renew Sustain Energy Rev*. 2022;161:112407.

Singh M, Fuenmayor E, Hinchy EP, Qiao Y, Murray N, Devine D. Digital twin: Origin to future. *Appl Syst Innov*. 2021;4(2):36.

Dietz M, Pernul G. Digital twin: Empowering enterprises towards a system-of-systems approach. *Bus Inf Syst Eng*. 2020;62(2):179-84.

Liu S, Lu Y, Shen X, Bao J. A digital thread-driven distributed collaboration mechanism between digital twin manufacturing units. *J Manuf Syst*. 2023;68:145-159.

Gouriseti SNG, Bhadra S, Sebastian-Cardenas DJ, Touhiduzzaman M, Ahmed O. A Theoretical Open Architecture Framework and Technology Stack for Digital Twins in Energy Sector Applications. *Energies*. 2023;16(12):4853.

Liao H, Zhou Z, Liu N, Zhang Y, Xu G, Wang Z, et al. Cloud-Edge-Device Collaborative Reliable and Communication-Efficient Digital Twin for Low-Carbon Electrical Equipment Management. *IEEE Trans Ind Inform*. 2023;19(2):1715-24.

Saad A, Faddel S, Mohammed O. IoT-Based Digital Twin for Energy Cyber-Physical Systems: Design and Implementation. *Energies*. 2020;13(18):4762.

Brosinsky C, Westermann D, Krebs R. Recent and prospective developments in power system control centers: Adapting the digital twin technology for application in power system control centers. In: *Proceedings of the 2018 IEEE International Energy Conference (ENERGYCON)*. Limassol, Cyprus; 2018:1-6.

Huang J, Zhao L, Wei F, Cao B. The Application of Digital Twin on Power Industry. *IOP Conf Ser Earth Environ Sci*. 2021;647(1):012015.

Ruhe S, Schaefer K, Branz S, Nicolai S, Bretschneider P, Westermann D. Design and Implementation of a Hierarchical Digital Twin for Power Systems Using Real-Time Simulation. *Electronics*. 2023;12(13):2747.

Liu T, Yu H, Yin H, Zhang Z, Sui Z, Zhu D, et al. Research and Application of Digital Twin Technology in Power Grid Development Business. In: *Proceedings of the 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE)*. Chongqing, China; 2021:383-7.

Zhang G, Huo C, Zheng L, Li X. An Architecture Based on Digital Twins for Smart Power Distribution System. In: *Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*. Chengdu, China; 2020:29-33.

Uslar M, Rohjans S, Neureiter C, Prössl Andrén F, Velasquez J, Steinbrink C, et al. Applying the smart grid architecture model for designing and validating system-of-systems in the power and energy domain: A European perspective. *Energies*. 2019;12(2):258.

Zhang J, Ma M, Wang P, Sun XD. Middleware for the Internet of Things: A survey on requirements, enabling technologies, and solutions. *J Syst Archit*. 2021;117:102098.

Deng S, Zhao H, Fang W, Yin J, Dustdar S, Zomaya AY. Edge intelligence: The confluence of edge computing and artificial intelligence. *IEEE Internet Things J*. 2020;7(9):7457-69.

Loseto G, Scioscia F, Ruta M, Gramegna F, Ieva S, Fasciano C, et al. Osmotic cloud-edge intelligence for IoT-based cyber-physical systems. *Sensors*. 2022;22(6):2166.

Velepucha V, Flores P. A survey on microservices architecture: Principles, patterns and migration challenges. *IEEE Access*. 2023;11:88339-58.

Mishra B, Kertesz A. The use of MQTT in M2M and IoT systems: A survey. *IEEE Access*. 2020;8:201071-201086.

Li L, Chou W, Zhou W, Luo M. Design patterns and extensibility of REST API for networking applications. *IEEE Trans Netw Serv Manag*. 2016;13(1):154-67.

Bizer C, Heath T, Berners-Lee T. Linked data-the story so far. In: *Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web*. ACM; 2023:115-43.

Lefrançois M. Planned ETSI SAREF Extensions based on the W3C&OGC SOSA/SSN-compatible SEAS Ontology Patterns. In: *Proceedings of the Workshop on Semantic Interoperability and Standardization in the IoT (SIS-IoT)*. Amsterdam, Netherlands; 2017.

Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, et al. DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. *Semant Web*. 2015;6(2):167-95.

Sawadogo P, Darmont J. On data lake architectures and metadata management. *J Intell Inf Syst*. 2021;56(1):97-120.

Roy D, Srivastava R, Jat M, Karaca MS. A complete overview of analytics techniques: Descriptive, predictive, and prescriptive. In: *Decision Intelligence Analytics and the Implementation of Strategic Business Management*. Springer; 2022:15-30.

Ahmad T, Madonski R, Zhang D, Huang C, Mujeeb A. Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm. *Renew Sustain Energy Rev*. 2022;160:112128.

Alimi OA, Ouahada K, Abu-Mahfouz AM. A review of machine learning approaches to power system security and stability. *IEEE Access*. 2020;8:113512-31.

Kegel L, Hahmann M, Lehner W. Generating what-if scenarios for time series data. In: *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*. Chicago, IL, USA; 2017:1-12.

Nguyen TN, Gonzalez C. Effects of decision complexity in goal-seeking gridworlds: A comparison of instance-based learning and reinforcement learning agents. In: *Proceedings of the 18th International Conference on Cognitive Modelling*. Online; 2020.

Jin Y, Ma M, Zhu Y. A comparison of natural user interface and graphical user interface for narrative in HMD-based augmented reality. *Multimed Tools Appl*. 2022;81(4):5795-826.

Pan S, Luo L, Wang Y, Chen C, Wang J, Wu X. Unifying large language models and knowledge graphs: A roadmap. *IEEE Trans Knowl Data Eng*. 2024;36(7):3580-99.

Gunduz MZ, Das R. Cyber-security on smart grid: Threats and potential solutions. *Comput Netw*. 2020;169:107094.

Sakimura N, Bradley J, Jones M, De Medeiros B, Mortimore C. OpenID Connect Core 1.0 incorporating errata set 2. OpenID Foundation Specification. 2023. Available at: https://openid.net/specs/openid-connect-core-1_0.html. Accessed October 21, 2024.

Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, et al. GPT-4 Technical Report. *arXiv*. 2023. arXiv:2303.08774.

Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv*. 2023. arXiv:2307.09288.

Haklay M, Weber P. Openstreetmap: User-generated street maps. *IEEE Pervasive Comput*. 2008;7(4):12-8.

Adel A. Unlocking the future: Fostering human-machine collaboration and driving intelligent automation through Industry 5.0 in smart cities. *Smart Cities*. 2023;6(5):2742-82.

Veichtlbauer A, Ortmayr M, Heistracher T. OPC UA integration for field devices. In: *Proceedings of the 2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. Emden, Germany; 2017:419-24.

Mosteiro-Sanchez A, Barcelo M, Astorga J, Urbieto A. Securing IIoT using defence-in-depth: Towards an end-to-end secure industry 4.0. *J Manuf Syst*. 2020;57:367-78.

Teerakanok S, Uehara T, Inomata A. Migrating to zero trust architecture: Reviews and challenges. *Secur Commun Netw*. 2021;2021:1-10.

Neumann WP, Winkelhaus S, Grosse EH, Glock CH. Industry 4.0 and the human factor—A systems framework and analysis methodology for successful development. *Int J Prod Econ*. 2021;233:107992.

Agrawal A, Thiel R, Jain P, Singh V, Fischer M. Digital Twin: Where do humans fit in? *Autom Constr*. 2023;148:104749.

Nardo M, Forino D, Murino T. The evolution of man–machine interaction: The role of human in Industry 4.0 paradigm. *Prod Manuf Res*. 2020;8(1):20-34.

Bonney MS, de Angelis M, Dal Borgo M, Wagg DJ. Contextualisation of information in digital twin processes. *Mech Syst Signal Process*. 2023;184:109657.

Molino M, Cortese CG, Ghislieri C. The promotion of technology acceptance and work engagement in Industry 4.0: From personal resources to information and training. *Int J Environ Res Public Health*. 2020;17(7):2438.

Alcaraz C, Lopez J. Digital twin: A comprehensive survey of security threats. *IEEE Commun Surv Tutor*. 2022;24(3):1475-503.