

Article

A Retrieval-Augmented Generation Approach for Data-Driven Energy Infrastructure Digital Twins

Saverio Ieva ^{1,2}, Davide Loconte ¹, Giuseppe Loseto ^{2,3,*}, Michele Ruta ^{1,2}, Floriano Scioscia ^{1,2},
Davide Marche ⁴ and Marianna Notarnicola ⁴

- ¹ Department of Electrical and Information Engineering, Polytechnic University of Bari, Via E. Orabona 4, I-70125 Bari, Italy; davide.loconte@poliba.it (D.L.); michele.ruta@poliba.it (M.R.); floriano.scioscia@poliba.it (F.S.)
- ² donkeyPower S.r.l., Via E. Orabona 4, I-70125 Bari, Italy
- ³ Department of Engineering, LUM “Giuseppe Degennaro” University, S.S. 100 km 18, I-70010 Casamassima, Italy
- ⁴ Lutech S.p.A., Via M. Gorki 30/32C, I-20092 Cinisello Balsamo, Italy; davide.marche@lutech.it (D.M.); marianna.notarnicola@lutech.it (M.N.)
- * Correspondence: loseto@lum.it; Tel.: +39-080-5963515

Highlights:

What are the main findings?

- Novel data-driven and knowledge-based energy digital-twin framework integrating a Retrieval-Augmented Generation (RAG) approach.
- Prototype applied to a real-world scenario involving the management of high-voltage energy infrastructures, showcasing the framework feasibility and effectiveness in operational environments.

What is the implication of the main finding?

- Improved management of energy infrastructures enhancing the ability to predict future conditions and prescribe more informed and data-driven decisions in asset maintenance.
- Exploitation of a conversational virtual assistant to interact with users, improving the accessibility, interpretability, and usability of complex data for decision-makers.



Citation: Ieva, S.; Loconte, D.; Loseto, G.; Ruta, M.; Scioscia, F.; Marche, D.; Notarnicola, M. A Retrieval-Augmented Generation Approach for Data-Driven Energy Infrastructure Digital Twins. *Smart Cities* **2024**, *7*, 3095–3120. <https://doi.org/10.3390/smartcities7060121>

Academic Editor: Pierluigi Siano

Received: 9 September 2024

Revised: 17 October 2024

Accepted: 22 October 2024

Published: 24 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: digital-twin platforms are increasingly adopted in energy infrastructure management for smart grids. Novel opportunities arise from emerging artificial intelligence technologies to increase user trust by enhancing predictive and prescriptive analytics capabilities and by improving user interaction paradigms. This paper presents a novel data-driven and knowledge-based energy digital-twin framework and architecture. Data integration and mining based on machine learning are integrated into a knowledge graph annotating asset status data, prediction outcomes, and background domain knowledge in order to support a retrieval-augmented generation approach, which enhances a conversational virtual assistant based on a large language model to provide user decision support in asset management and maintenance. Components of the proposed architecture have been mapped to commercial-off-the-shelf tools to implement a prototype framework, exploited in a case study on the management of a section of the high-voltage energy infrastructure in central Italy.

Keywords: digital twin; energy infrastructures; energy management; retrieval-augmented generation; natural user interface

1. Introduction

Over the past decade, the global energy landscape has undergone a remarkable transformation, marked by significant shifts in technology, policy, and market dynamics. The evolution of the energy system during this period has been shaped by a combination of factors, including increasing concerns about climate change, advances in renewable energy technologies, and the growing recognition of the need for a more sustainable and resilient energy infrastructure [1]. In particular, the rapid growth of distributed and

renewable energy sources can be considered as one of the most notable trends in the last 10 years in the energy systems area [2]. Solar and wind power have experienced substantial increases in deployment and efficiency. Technological advancements, coupled with falling costs, have made these renewable sources more economically competitive with traditional fossil fuels. This shift towards cleaner energy has been accelerated by a greater awareness of the environmental impacts of fossil fuel consumption and the imperative to reduce greenhouse-gas emissions. Government policies and international agreements have driven the transition to a more sustainable energy system. In addition, many countries have adopted ambitious renewable energy targets, implemented incentive programs, and established regulatory frameworks to encourage the development and integration of clean energy technologies. Moreover, the Paris Agreement [3], signed in 2015, served as a catalyst for global cooperation on climate action, prompting nations to re-evaluate their energy strategies and commit to decarbonization.

Spurred by these goals, Information and Communication Technology (ICT) integration has been increasingly adopted to transform traditional energy grids into smart grids. They represent a groundbreaking enhancement in the management and distribution of electrical power, representing a key element in the modernization of energy systems and infrastructures [4]. The smart-grid market is experiencing rapid growth worldwide, driven by an increasing demand for efficient and sustainable energy solutions. As reported in [5], it was estimated at almost USD 50 billion in 2022 and is forecast to grow at a Compound Annual Growth Rate (CAGR) of 17.4%, reaching approximately USD 130 billion in 2028. Smart grids exploit advanced metering systems and data analytics to enhance the reliability of energy infrastructure and also enable better demand-side management, optimizing energy distribution and empowering consumers to make informed decisions about their energy consumption. One of the primary features of smart grids is the integration of digital communication systems that enable real-time monitoring and control of various components within the grid. Smart devices, deployed throughout the network, provide a two-way communication between consumers and utilities and enable a granular collection of data regarding electricity consumption, power generation, and grid conditions. At different levels, both system administrators and consumers can gain insights into energy usage patterns and make informed decisions to optimize distribution during periods of lower demand or lower pricing, contributing to demand-side management and overall grid efficiency [6]. Smart grids also enhance the infrastructure resilience against disruptions and outages. Automated monitoring and control systems can quickly identify and isolate faults, rerouting power to minimize downtime and improve overall system reliability. Additionally, smart grids can integrate procedures for the rapid detection of unauthorized access or cybersecurity threats, ensuring the security and integrity of the electricity supply [7].

While progress has been made in transitioning toward a more digital energy infrastructure, challenges persist. Standardization and interoperability concerns require ongoing attention to ensure the seamless integration of data-driven techniques within everyday grid operations. To overcome these limitations, the Digital Twin (DT) paradigm emerges as a key enabler; it consists of a representation of a real-world object or process within an information system, accurately reflecting its physical counterpart through real-time bidirectional data exchange for purposes including simulation, testing, monitoring, and maintenance. Originally introduced in the aerospace industry and popularized in smart manufacturing, DT architectures and technologies can be exploited for shaping more resilient, flexible, and sustainable energy infrastructures by providing a virtual counterpart of physical assets and processes to increase their visibility and controllability. As highlighted in [8], the implementation of Energy Digital Twins (EDTs) has the potential to transform the management of energy systems, resulting in enhanced energy efficiency, decreased downtime, and reduced maintenance costs.

Energy infrastructure DTs covering large areas integrate a wide number and range of elements, models, data sources, analytic processes, and management tasks. This complexity poses significant cognitive challenges for users: conventional Graphical User Interfaces

(GUIs), based only on Three-Dimensional (3D) visualizations and classical computer interaction paradigms, make it difficult to discover resources, retrieve information, and gain insights in the desired way, often resulting in a restricted user experience [9]. Novel approaches are required to enable smart-grid DTs to achieve maturity from a usability perspective. In particular, this work addresses the following Research Aims (RAs) related to the application of advanced technologies in energy infrastructure management:

- RA1. Enhancing grid management through DTs and data-driven approaches: By examining the integration of digital twins and innovative data visualization techniques with data-driven methods, the research highlights how this combination improves the monitoring of energy infrastructures. This vision aims to enable more accurate real-time assessments and failure prevention procedures.
- RA2. Leveraging Large Language Models (LLMs) for decision support: The study explores the challenges and benefits of utilizing LLMs in energy infrastructure. Key challenges include data integration and data modeling and ensuring accuracy in domain-specific tasks in order to obtain enhanced decision-making capabilities and improved automation.
- RA3. Integrating Retrieval-Augmented Generation (RAG) approaches with Knowledge Graphs (KGs): The exploitation of a Resource Description Framework (RDF) [10] for data modeling in RAG is also a focus of this study. By structuring domain-specific knowledge, a KG improves the performance of RAG-based LLMs in delivering real-time, accurate information, thereby increasing the effectiveness of decision support systems.

According to the previous research statements, this paper presents a novel data-driven, knowledge-based architecture for EDTs, aiming to provide complete visibility of the status of the energy network and its assets. The proposed framework supports gathering data streams from sensing devices, pre-processing them and performing Machine Learning (ML) procedures for detecting and addressing potential grid issues. Most notably, it includes a virtual assistant—based on an RAG approach integrating a foundation LLM with domain-specific information organized in a KG—to support end users and provide informed suggestions regarding energy infrastructure performance issues and maintenance procedures. The main contributions of the proposal are as follows:

- a reference architecture for energy infrastructure management and monitoring, integrating digital twins and data-driven methods;
- a Natural User Interface (NUI) extending conventional visualization and interaction paradigms with direct object retrieval and manipulation by means of a combination of gestures and a conversational virtual assistant;
- annotation of DT models and data streams into a RDF KG, exploited for RAG to provide real-time domain information to a general-purpose foundation LLM powering the conversational agent;
- characterization of each architectural module by means of a mapping with Commercial-Off-The-Shelf (COTS) components to enhance feasibility while simultaneously lowering development costs and time to market;
- a cloud-based prototype implementation integrating open-source software technologies and tools;
- a case study regarding a section of the high-voltage network in central Italy, showcasing key value propositions of the approach.

Finally, the research aims will be validated through a comprehensive analysis comparing the developed architecture against the current state-of-the-art in energy infrastructure management. This analysis will also highlight improvements in decision-making procedures and situational awareness resulting from the implementation of the proposed methodologies compared to the baseline approaches previously applied in the reference case study.

The remainder of the paper is organized as follows. Section 2 analyzes emerging technologies and relevant related work, while the proposal is described in detail in Section 3,

focusing on the overall architecture and its components. A real-world case study is presented in Section 4, also highlighting the available open-source tools selected to implement the prototype system. Section 5 provides a discussion about system peculiarities as well as the challenges and opportunities related to the proposed approach, and conclusions and future work directions are provided in Section 6.

2. Background

Here, the main notions regarding novel approaches and technologies integrated in energy systems are briefly recalled, in order to make the paper self-contained and easily understandable. Then, the most relevant related work is surveyed.

2.1. Emerging Technologies and Digital Twins for Energy Infrastructures

The integration of emerging technologies with energy grids represents a dynamic frontier in the evolution of modern energy systems. Rapid advancements in various technological domains have opened new possibilities for enhancing the capabilities and efficiency of smart grids from connectivity, automation, and intelligence standpoints. As reported in [11], the Internet of Things (IoT) has facilitated the proliferation of interconnected devices and sensors, creating a web of energy data where information can be easily retrieved by means of several endpoints located within the smart grid. The data collected from IoT devices enables better monitoring, control, and decision-making. The advent of 5G technology has also provided high-throughput, low-latency communication crucial for the real-time data exchange required in smart grids [12]. This connectivity evolution enhances the reliability and responsiveness of grid components, supporting applications like autonomous grid management, advanced metering, and seamless integration of electric vehicles into the grid.

Artificial Intelligence (AI) and ML techniques have found application in optimizing the operation and management of smart grids [13]. These technologies can be used to analyze vast amounts of data generated by grid devices, identify patterns, and optimize grid performance. In particular, ML models can enhance fault detection, predictive and preventive maintenance, and load forecasting, contributing to the increased reliability and efficiency of the whole infrastructure [14]. In this area, edge computing [15] represents a key enabling technology, bringing computational resources closer to data generation sources in order to reduce latency and improve real-time processing capabilities. Instead of sending all data to a centralized cloud-based or on-premises datacenter for processing, edge computing distributes computational resources across the border of the local network. Decentralization reduces dependence on a single point of failure and enhances the overall resilience of the system. This approach is especially beneficial in scenarios where bandwidth is limited or expensive, as well as where transmitting large volumes of data to the cloud is impractical or prevented by privacy requirements. In the context of energy systems, edge computing allows for quicker decision-making, enabling faster responses to unexpected events and enhancing grid operations [16,17]. Moreover, it supports a scalable model adapted to varying workloads and requirements, where additional edge devices can be added to the network to distribute the processing load, granting a more flexible and dynamic infrastructure. Following this vision, intelligent systems deployed at the edge can perform advanced data processing and decision-making locally based on tiny ML [18,19] algorithms and server-less federated learning approaches [20]. These distributed approaches call for novel decentralized security architectures for data integrity [21] and detection of malicious behaviors [22].

In recent years, the exploitation of DT technologies has emerged as a further promising paradigm for optimizing and managing complex systems, particularly within the domain of energy infrastructures, with several studies and research projects being conducted across diverse areas, including renewable energy and energy storage and distribution [23]. Basically, a DT is a virtual representation of a physical object, system, or process [24]. This representation is created using data collected from sensors, devices, and other sources associated

with the physical entity. The DT serves as a dynamic and detailed model, offering insights into the real-world object's status, performance, and behavior. Raw data are continuously fed into the digital twin in order to obtain a comprehensive and accurate representation of the physical asset. EDTs usually comprise different modules, including geospatial and simulation models, data analytics, and connectivity. Geospatial models provide a georeferenced 3D visual representation of the physical entity, while simulation models produce a user-oriented representation of its behavior. Data analytics process the incoming data to derive insights, and connectivity ensures real-time updates between the physical and digital domains. Digital twins often cover the entire lifecycle of a physical entity, from design and development to operation and maintenance [25]. This comprehensive lifecycle management allows energy stakeholders to monitor, analyze, and optimize the performance of assets, systems, or processes throughout their activities. Finally, DTs facilitate the development of collaborative environments where various engineers, operators, and maintenance personnel can interact with and contribute to the digital representation [26]. In energy infrastructures, this collaboration enhances communication and decision-making across different processes of a smart grid.

2.2. Related Work

EDT technology offers significant potential in aiding the professionals involved in industrial energy management to make data-driven decisions, facilitating proactive identification of system failures, enhancing system and network efficiency, optimizing operations, and aligning with sustainability objectives. However, its practical adoption within the energy sector is affected by constraints such as cost implications, regulatory compliance, and challenges in evaluating performance. This section reviews state-of-the-art works on the DT frameworks applied to energy infrastructures. The selected studies integrate technology, data analytics, cybersecurity, and visualization in a holistic approach to enhance energy systems' management and sustainability.

A theoretical framework and technology stack for DTs within the energy sector have been introduced in [27]. The proposed *D-Arc* technology stack is platform-agnostic and applicable to any DT platform and grid systems, aiming to enable the proactive discovery of system failures, enhance efficiency, and optimize operations. While it defines a workflow orchestration model compatible with both on-premises and cloud infrastructures, the technology-agnostic approach may lack specificity for addressing different use cases. A cloud-edge EDT framework named *C³-FLOW* has been proposed in [28] to manage low-carbon electrical equipment in smart industrial parks. It optimizes device scheduling, channel allocation, and computational resource distribution, leading to increased accuracy in federated learning, communication efficiency, and reduction of carbon emissions. However, the proposal has only been analyzed in a simulation environment, and the architecture and algorithms have not been tested with real devices. In [29], two types of DT models have been proposed to fulfill decision-making requirements in Energy Cyber-Physical Systems (ECPSs): a low-bandwidth model for periodic tasks like energy management and predictive maintenance, and a high-bandwidth model for real-time functions such as outage management and system restoration. The proposed architecture employs server-less computing for monitoring and executing local actions in ECPS, exploiting IoT shadow states as digital replicas. The ECPS has been evaluated using interconnected embedded computers for networked microgrids; however, real-world scalability is untested.

The potential benefits of employing the DT concept in power system control centers have been introduced in [30]. The proposed architecture, based on a simulation engine called "dynamic digital mirror", offers a solution for further improvement in power system monitoring and control. Similarly, in [31], a Power System Digital Twin (PSDT) has been designed, aimed at enhancing grid operation intelligence, power system control, fault prediction and diagnosis, and power online analysis by integrating smart sensors, 5G communications, cloud platforms, and big-data processing. However, both proposals are described only as architectural models, without the development of a prototype. The

hierarchical EDT architecture in [32] uses real-time simulation to emulate the physical grid for enhanced planning and operation. The framework features a vendor-agnostic real-time simulation system and integrates hardware-in-the-loop physical system emulation. Nevertheless, significant challenges remain open, including improving parameter estimation accuracy, refining model granularity, and minimizing time delays. The EDT framework in [33] is aimed at robust smart-grid development. The platform facilitates comprehensive planning, construction, and operation of the power grid by integrating datasets including power, regulatory planning, and meteorological data. However, essential technologies such as DT representation and AI algorithms have not been fully integrated into the platform to offer production-grade services.

Finally, an architecture based on DT for smart power distribution systems has been proposed in [34], addressing the challenge of rapidly providing data models, analytics, and algorithms, incorporating AI and big-data technologies. However, insufficient sensor deployment is a key limitation, necessitating further exploration of DT technology applicability across the whole smart power distribution system, along with the need for additional DT model implementation.

3. Framework Architecture

The smart-grid concept emerges as a dynamic synergy between power infrastructures and ICT networks. Analogous to the human body, where the “circulatory system” ensures the flow of vital resources and the “nervous system” orchestrates complex monitoring and interactions, the smart grid harmonizes energy distribution and ICT systems, respectively. Following this vision, the Smart-Grid Architecture Model (SGAM) [35] has been defined to guide the development and implementation of smart-grid systems. As shown in Figure 1, it is a comprehensive framework conceptualizing five reference layers of smart energy networks that seamlessly integrate physical components, communication systems, and information technologies. The main goal of the proposed approach is the design of a reference data-driven framework that accommodates various DT applications adhering to SGAM guidelines and best practices of large-scale middleware development [36]: interoperability, self-adaptability, real-time functionality, scalability, reliability, and data security. As shown in Figure 2, the framework consists of different core layers, each with distinct responsibilities and functionalities. Individual components are detailed in the following subsections, along with technological choices for their reference implementation and integration.

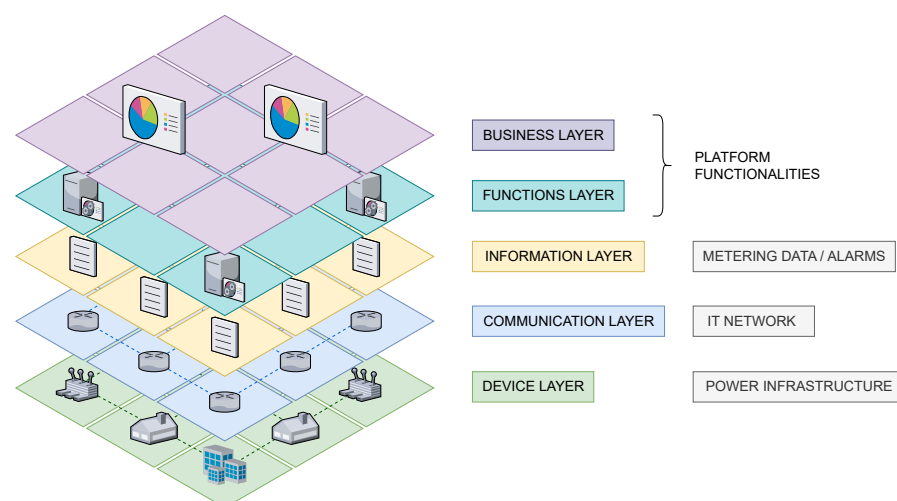


Figure 1. Smart-grid architecture model.

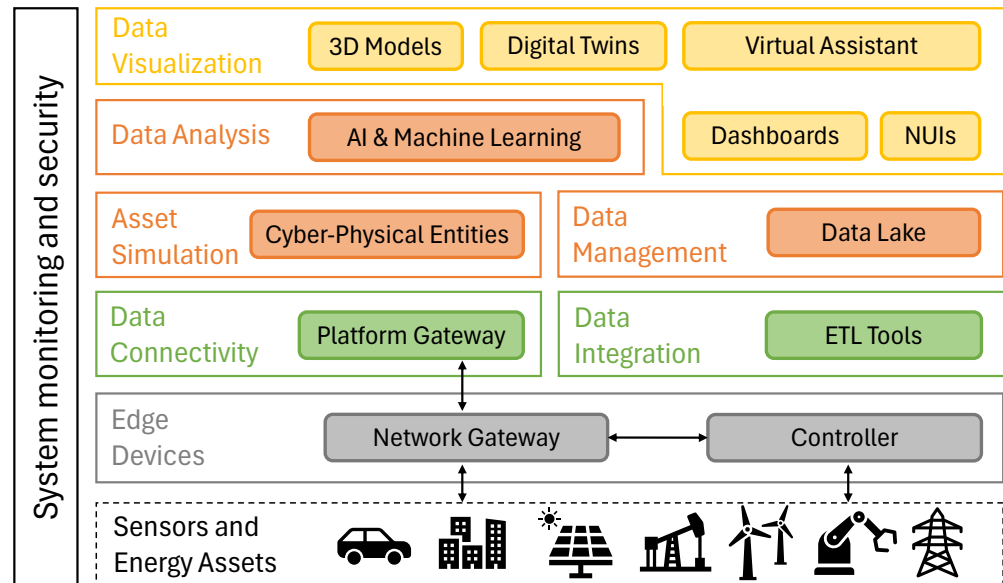


Figure 2. Reference architecture.

3.1. Sensor Networks and Communication

At the foundational layers of perception and connection, the architecture interfaces with physical devices at the edge of the network as well as with cloud-based resources. The lowest level, closest to the field devices for data acquisition, involves the installation of sensors on the physical assets and overseeing the bidirectional exchange of data from the devices to the platform. The Data Connectivity layer serves as the backbone for the system, facilitating seamless communication, synchronization, and integration between the physical environment and the digital realm. Specifically, the architecture incorporates different types of device:

- **Traditional sensors:** These generate signals enabling the digital twin to acquire operational and environmental data from real-world physical assets. They provide direct measurements of the variables related to the analyzed physical process, such as temperature and voltage.
- **Process Analytical Technology (PAT) sensors:** Similar to traditional sensors, PAT sensors record data from sampling points. Collected data cannot be directly interpreted as with traditional sensors. Typically, statistical and real-time observation models are employed to interpret the signal. A near-infrared spectrometer is an example of PAT sensor.
- **Actuators:** These control and manipulate physical elements or systems in response to signals or commands received from control units. They usually translate digital or analog signals into mechanical, electrical, hydraulic, or pneumatic actions to effect changes in the physical environment.

Data coming from sensors are processed by the Edge Computing layer. Processing data close to where they are generated brings considerable benefits in terms of processing latency, reduced data traffic, and increased resilience in the case of an interruption in the data connection. In order to take advantage of edge and swarm intelligence paradigms [37,38], the platform has been designed following a microservices architecture [39], where computation and data storage tasks are encapsulated in autonomous distributed application components in order to enable a more dynamic and scalable distribution according to network and system requirements. Challenges in sensor networks arise from the diversity of communication protocols and standards available within the energy ecosystem. This diversity, while enabling resilience and flexibility, presents several hurdles that must be addressed to ensure a seamless interoperability among heterogeneous devices. As shown in Figure 2,

the proposed infrastructure includes two types of communication device, aiming to support interactions among all infrastructure devices over multiple communication protocols:

- Gateways: These receive data locally from the sensors and perform protocol translation tasks, converting data from diverse protocols into a standardized format. Gateways may also perform preliminary data processing tasks, such as filtering, normalization, and data cleaning, to enhance the quality and integrity of the collected data. This processing helps mitigate noise and reduce bandwidth consumption. They implement authentication mechanisms and encryption protocols and access control policies to enforce security measures and data protection.
- Controllers: These handle communication tasks between actuator devices and other network components, such as gateways. They receive the control actions from the gateways, forward messages to the actuators connected to the physical assets, and manage task assignment to facilitate reliable and efficient communication within the energy network.

In order to respect the strict requirements of energy systems in terms of costs, efficiency of sensing and information-gathering operations, and reduction of transmission delays for critical data packets, the proposed platform exploits a communication architecture based on the Message Queuing Telemetry Transport (MQTT) protocol [40]. MQTT (<https://mqtt.org>, accessed on 8 September 2024) is an application protocol based on the publish/subscribe paradigm, where client nodes communicate through a broker node that distributes messages over the network. Client nodes, also called “publishers”, can send messages related to specific “topics”, which will be received by “subscriber” peers waiting for data related to the same target topic. Gateways can also interact with on-premises or cloud-based services exposing REpresentational State Transfer (REST) Application Programming Interfaces (APIs) [41]. REST APIs provide a uniform, flexible, and lightweight way to integrate application functionalities and to connect components in microservices architectures through simple Hypertext Transfer Protocol (HTTP) requests to perform standard operations like reading, updating, and deleting data. Information can be delivered to a client in different formats; however, JavaScript Object Notation (JSON) (<https://www.json.org>, accessed on 8 September 2024) has been selected as reference format in the proposed platform, as it is one of the most common representation standards, being programming language-agnostic and easily human- and machine-readable.

3.2. Data Management and Storage Techniques

Moving up the stack, collected data are filtered and aggregated by the Data Integration module. It is responsible for modeling information regarding the three lower layers of the SGAM framework (Figure 1) and creating the unified view depicted in Figure 3. The creation of logical data models introduces several advantages, including simplifying the reasoning and discovery of new relationships existing within collected data, creating uniformity in data documentation and system design, reducing errors during the development of novel functionalities of the platform, and simplifying communication between data engineers and business intelligence teams.

All of the entities reported in Table 1 are modeled as RDF [10] resources following linked-data best practices [42] in order to make the proposed approach general-purpose and support semantic-based data annotation and interpretation. In particular, the following well-known vocabularies are used as upper ontologies modeling basic concepts and properties concerning energy infrastructures, network communication, and metering procedures:

- *Digital Twins Definition Language (DTDl)* ontologies for energy grid (<https://github.com/Azure/opendigitaltwins-energygrid>, accessed on 8 September 2024) and smart cities (<https://github.com/Azure/opendigitaltwins-smartcities>, accessed on 8 September 2024), specifically proposed for modeling DT solutions, including the monitoring of grid and urban assets (concepts in blue in Figure 3);

- *Smart Energy Aware Systems (SEAS)* ontology (<https://w3id.org/seas/>, accessed on 8 September 2024) [43], designed as a set of simple core ontology patterns modeling multiple engineering-related concepts and properties of the energy ecosystem (concepts in violet in Figure 3);
- *Procedure Execution ontology (PEP)* (<https://w3id.org/pep/>, accessed on 8 September 2024), including the properties used to describe procedures, outputs, and results related to metering activities;
- *DBpedia* [44] resources (*DBR*) and *GeoNames (GN)* ontology (<https://www.geonames.org/ontology/>, accessed on 8 September 2024) have been exploited to model elements of the IT network and the relationships between asset locations (concepts in yellow in Figure 3).

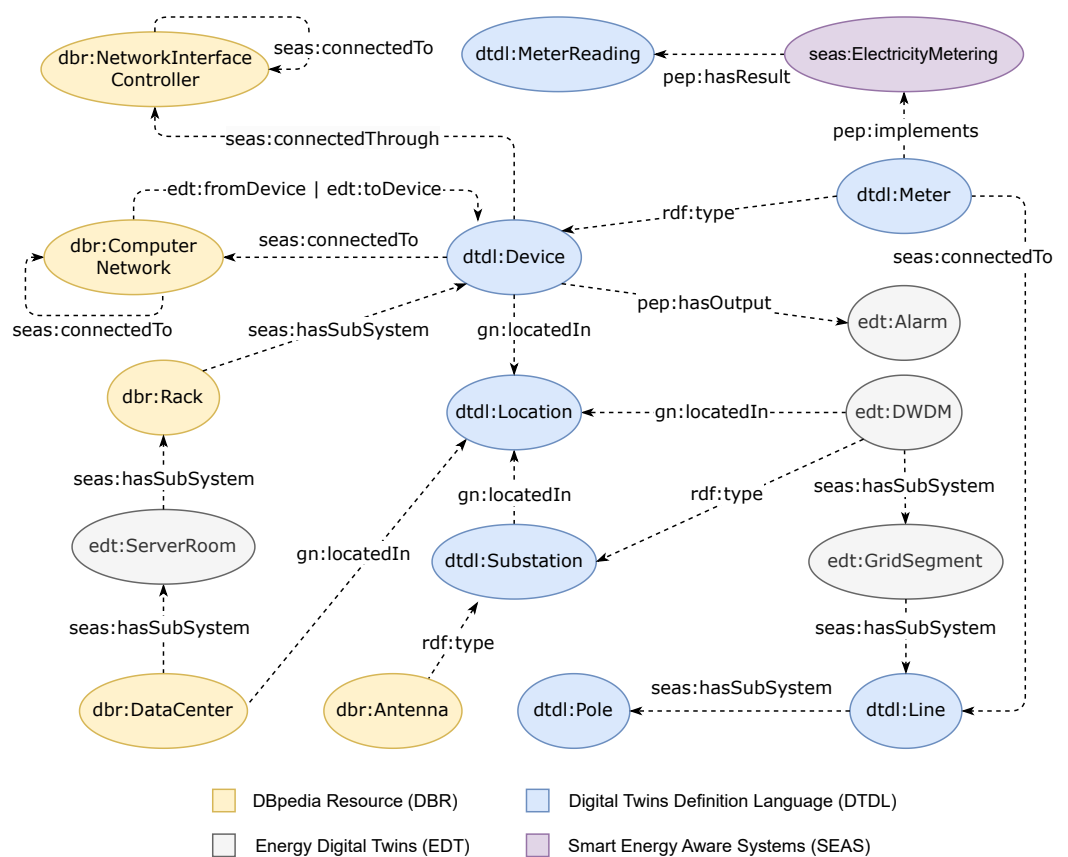


Figure 3. Reference graph-based data model.

In Figure 3, the entities and relationships belonging to each ontology are denoted by their respective RDF namespace prefix and are shown in a different color. In addition, a custom EDT vocabulary has been defined to model domain entities (reported in grey in Figure 3) and properties not included within the aforementioned ontologies.

The *Asset Simulation* layer is defined for abstracting and virtualizing cyber-physical entities within the infrastructure. This abstraction enables efficient resource allocation and management, ensuring that digital components are instantiated and maintained appropriately throughout their lifecycle. It also facilitates the entanglement between physical and virtual objects, fostering a symbiotic relationship that enhances the overall effectiveness of the architecture. Finally, the *Data Management* module acts as system orchestrator, coordinating interactions between various components and services. Additionally, it manages the large amount of data generated and collected by field sensors, normalizes the information received using standardized formats, and makes it available to other modules. A major challenge in such a scenario is to ensure the interoperability of the data models adopted by different device networks and data sources connected to the system.

Table 1. Entities defined in the reference data model.

Entity	Vocabulary	Description
Device	DTDL	A device included within the energy infrastructure
Meter	DTDL	Physical asset used for measuring energy consumption and detecting events of interest
Electricity Metering	SEAS	Process of measuring and recording electrical energy consumption
Meter Reading	DTDL	Amount of electricity consumption recorded through a meter
Alarm	EDT	Specific event related to an issue identified by a device
Location	DTDL	Physical location of a device or energy asset
NIC	DBR	Network Interface Controller (NIC) representing a physical connection through which the device interacts with external peripherals
Computer Network	DBR	Group of interconnected devices capable of communicating with each other
Substation	DTDL	Electrical station part of a generation, transmission, or distribution system
Antenna	DBR	Local secondary substation installed to support critical scenarios requiring high data availability in the smart grid
DWDM	EDT	Dense Wavelength Division Multiplexing (DWDM) system supporting fiber-optic transmissions
Grid Segment	EDT	Section of a power or communication network
Line	DTDL	A line composing a telecommunication or power transmission infrastructure
Pole	DTDL	A utility pole used to support overhead power lines or other public utilities
DataCenter	DBR	Centralized facility designed to store and process vast amounts of data related to the energy grid
Server Room	EDT	Air-conditioned room devoted to the continuous operation of computer servers in a datacenter
Rack	DBR	Physical container designed to house servers, networking devices, and other datacenter equipment

All data composing the reference knowledge base are managed through a reference data lake. It acts as a centralized repository that allows for the storage of a vast volume of raw, structured, and unstructured data needed for analysis or processing. In this way, sensor data can be managed along with further system information, including infrastructure specifications, business models, or bills of materials. Unlike traditional data warehouses [45], a data lake also accepts data from various sources, formats, and types, enabling organizations to ingest and store diverse datasets without prior transformation or normalization. Data lakes leverage scalable storage solutions, such as cloud object storage or distributed file systems, which offer cost-effective storage options compared to traditional data management tools. Moreover, they facilitate data exploration and discovery by providing tools and frameworks for ad hoc querying, analysis, and visualization of available data. In the proposed architecture, the data lake can manage data originally stored in both Relational Database Management Systems (RDBMSs) and *NoSQL* databases. In particular, graph databases are suggested for modeling and understanding the relationships inherent in data, especially in dynamic contexts such as a smart cities or power grids. Data are stored as knowledge graphs where distinct entities are connected through well-defined relationships, allowing a structured and semantic-oriented view of the data. In the presence of specific processes with strict time constraints, such as leaderboards or real-time data

analytics, in-memory databases are also exploited for caching data to reduce the number of disk accesses and minimize latency.

3.3. Data Processing and Analytics

Information provided through the module is exploited by the Data Analysis layer. It provides descriptive, predictive, and prescriptive analytics [46] capabilities aimed at modeling and understanding the state and operational behaviors of energy systems, ref. [47] starting from data collected by sensing devices and external resources. In the latter case, the layer may interact with the Data Integration module to activate data crawling capabilities in order to extend the knowledge base with additional information obtained from cloud-based services. By means of data mining and machine-learning algorithms, the system aims to identify actionable insights and extract useful knowledge from the vast volumes of data generated by the connected infrastructures. As highlighted in [48], potential applications include the following:

- Fault detection and diagnosis: identify possible malfunctions of the energy grid and monitor the health and performance of grid assets such as transformers, generators, or substations;
- Preventive maintenance: based on grid equipment structural and behavioral models, analyze status data series and predict when maintenance should be performed to prevent failures and minimize downtime;
- Load forecasting and balancing: predict future electricity demand based on historical data and balance the load across different parts of the grid to ensure stable operation and prevent overloading or congestion;
- Demand response optimization: optimize demand response programs to manage electricity consumption during peak periods or in response to pricing signals.

Depending on the timing of the response and the amount of data to be processed, the framework enables two main types of processing task:

- Batch processing for large volumes of data and group of transactions. Usually, data is collected, inserted, and processed to produce results automatically. Tasks are scheduled over even medium to long periods of time;
- Real-time processing for tasks requiring a rapid response and performed by simple systems managing rapid transactions. Mainly used in environments where many events occur in a short time.

The Data Analysis layer enables the application of further techniques to simulate real-world operational scenarios according to the following analytical approaches:

- What-if analysis [49]: one or more parameters of the reference scenario are expected to be changed and, after the execution of multiple simulation sessions, results are stored to allow comparisons and sensitivity analysis;
- Goal-seeking analysis [50]: a simulation objective is defined in terms of observable variables or Key Performance Indicators (KPIs), and the system will provide support in defining scenarios aimed to identify and evaluate suitable ways to obtain the specific result.

3.4. Data Visualization and Digital Twins

At the top of the architecture, the *Data Visualization* layer operates as a business intelligence interface of the system, laying the groundwork for the seamless integration of physical and digital environments, driving innovation and efficiency in the energy domain. It uses data from the underlying layers to build different views and virtual representations, including digital models associated with the constituent elements of the energy infrastructure (e.g., metering devices, power-grid nodes, small and autonomous system environments monitored by a transmission/reception tower). The visualization modules also enable a multi-channel interaction with DT functionalities, providing available high-level information for the development of web and mobile applications. In this way, end

users are capable of visualizing and enriching digital models through augmented or virtual reality tools, exploiting crowdsourcing approaches and advanced visualization solutions based on interactive human–computer interfaces.

The Data Visualization layer exposes a comprehensive set of digital solutions empowering organizations to harness the full potential of digital twins in several application scenarios. Basically, the proposed architecture provides the following tools in order to interact with and manipulate data:

- 3D models representing cyber-physical objects. Users can access, view, and modify assets through their 3D virtual representations;
- *Interactive dashboards* and *business intelligence* applications used to monitor in real-time the whole system and show the results of data processing, relevant KPIs, and warnings received in the presence of dangerous conditions;
- *NUIs* that enable users to interact with the platform using intuitive movements and gestures [51]. Natural User Interfaces are designed to enhance user experience by reducing the learning curve and enabling a seamless human–machine integration.

As Figure 4 shows, NUIs represent an evolution of established user interface paradigms. *Command-Line Interfaces* (CLIs) are abstract in nature, as actions are encoded in text commands; this, however, grants them the advantage of easy programmability by means of scripts. For this reason, they are more productive for batch systems rather than for real-time human–machine interaction. In *Graphical User Interfaces* (GUIs), the adoption of visual metaphors and symbols reduces the human–machine gap by making use of icons, menus, and 2D/3D representations of the real world. Interactive applications benefit from GUIs to increase user productivity. NUIs push interactivity even further by enabling direct and multimodal input possibly, including hand and body gestures, speech, eye tracking, and more. They adopt patterns that humans use daily to interact with each other, such as conversational language, touch, and gestures. This makes interaction more intuitive, enabling users to focus on their goals and reducing the cognitive burden of content manipulation and information processing via artificial systems. For this reason, this paradigm is particularly fit for complex platforms such as DTs.

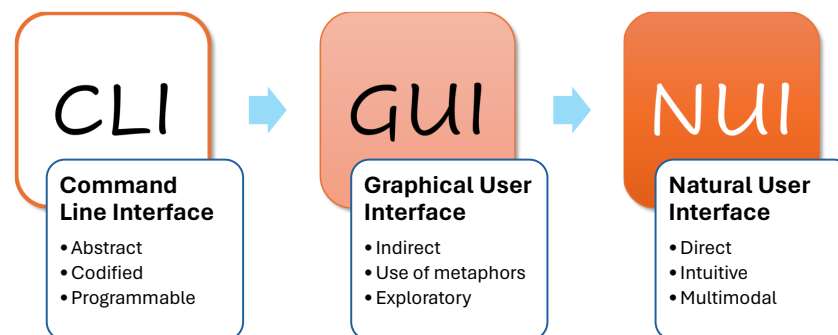


Figure 4. User Interface paradigms.

The proposed DT platform provides specific functionalities leveraging generative AI methods to support problem-solving and NUIs for querying the KG. The framework integrates an RAG module combining the strengths of LLMs with information retrieval and search capabilities, enabling relevant information from vast datasets to be dynamically fed into the model during the response generation process [52]. LLMs have demonstrated remarkable proficiency in various tasks, but they often lack context specificity and struggle with generating accurate responses in real-world scenarios. By incorporating external knowledge, RAG equips LLMs with a better understanding of context, leading to more informed, coherent, and accurate outputs for generating high-quality responses. Users are also able to interact with the system to issue specific actuation commands that are executed

on the physical twin, allowing direct control over physical assets based on real-time insights derived from the digital twin.

Finally, the platform includes a *data export* functionality for generating datasets in the RDF. Users navigate the reference data lake, query and filter available information, and export different subsets of the data according to linked-data principles [42].

3.5. System Monitoring and Security

Ensuring the security and integrity of the system is paramount in power-grid scenarios due to the critical nature of its operations [53]. System monitoring aims to detect and respond to any anomalies or suspicious activities in real-time, whereas system security involves protecting the power-grid infrastructure from cyber threats, physical attacks, and unauthorized access. As shown in Figure 2, a vertical layer has been designed to support both technical and governance activities, including:

- *Access control*: multi-factor authentication and role-based access control functionalities, based on the OpenID Connect (OIDC) [54] protocol, are implemented to ensure that only authorized personnel can access sensitive data and critical infrastructure components such as substations, control systems, and datacenters;
- *Continuous monitoring*: a service orchestrator aggregates and analyzes network traffic and log data collected from various grid components, enabling security teams to identify performance issues, suspicious activities, or potential security breaches;
- *Vulnerability management*: regular vulnerability assessments and patch management procedures are conducted to identify and remediate security issues in grid components and software applications. This includes monitoring for security updates and patches released by vendors and promptly applying them to mitigate potential risks;
- *Incident response planning*: defining strict procedures, roles, and responsibilities for effectively managing and mitigating security incidents when they occur;
- *Compliance and regulations*: ensuring compliance with industry standards is essential for maintaining the security and reliability of the power grid. Compliance requirements include conducting regular security audits, implementing security controls, and reporting security incidents to regulatory authorities.

4. Case Study: Enhancing Grid Reliability Through Digital Twins

Following the reference architecture shown in Figure 2, a platform prototype has been developed to prove the feasibility of the proposal and to evaluate its capabilities and effectiveness. The platform is connected to several controllable sensing and actuation components, with multiple continuous data streams feeding various distributed decision points, either autonomously or under human supervision. The reference case study, described in details in the following sections, refers to monitoring and managing a section of the energy infrastructure in central Italy.

4.1. Prototype Implementation

As depicted in Figure 5, the prototype consists of a multi-layer microservice architecture deployed using *Kubernetes* (<https://kubernetes.io>, accessed on 8 September 2024) as the service orchestrator. It is an open-source container orchestration engine for automating management and scaling of containerized applications. Containers offer process-level isolation, ensuring that services run independently of each other. This isolation enhances security and reliability by reducing the risk of software conflicts and minimizing the impact of failures or vulnerabilities in one container on other services within the environment. Moreover, containerized services enable one to scale the platform easily and dynamically, based on processing demand. Containers can be spun up or down to handle fluctuations in workload, ensuring optimal resource utilization and performance. The prototype supports a hybrid deployment architecture, blending both on-premises and cloud components. This approach allows organizations to tailor their customized setup according to specific needs, such as regulatory requirements, data sensitivity, or workload demands.

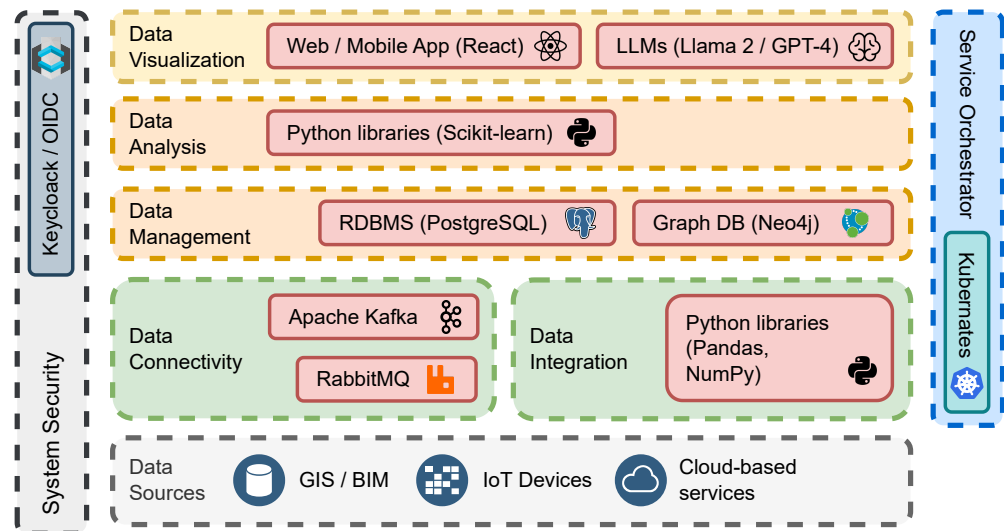


Figure 5. Prototype architecture.

Grid information is collected from several data sources, including IoT devices connected to the sensor network, Geographic Information Systems (GISs), Building Information Modeling (BIM) systems, and cloud-based platforms for asset management and monitoring. The EDT platform integrates different connectors in order to retrieve data using two communication modes: *synchronous*, accessing REST API-based services; *asynchronous*, through publish-subscribe systems or event streaming platforms (e.g., *Apache Kafka* (<https://kafka.apache.org>, accessed on 8 September 2024) or *RabbitMQ* (<https://www.rabbitmq.com>, accessed on 8 September 2024)). The platform also integrates *Keycloak* (<https://www.keycloak.org>, accessed on 8 September 2024), an open-source Identity and Access Management (IAM) framework providing user authentication and federation, as well as fine-grained authorization functionalities.

Extract, Transform, and Load (ETL) procedures have been implemented through custom scripts in the *Python* programming language and exploit *Pandas* (<https://pandas.pydata.org>, accessed on 8 September 2024) and *NumPy* (<https://numpy.org>, accessed on 8 September 2024) libraries for data manipulation. They are used to pre-process raw data in order to ensure that the information used for the subsequent analysis is accurate, complete, and is in suitable formats. This involves detecting and handling missing values, outliers, duplicate records, and other anomalies that could compromise the integrity of the dataset. Polished data are then annotated according to the reference KG, described in Figure 3, and stored in *Neo4j* (<https://neo4j.com>, accessed on 8 September 2024), a graph database management system. It is designed for efficiently managing and querying graph-structured data characterized by different relationships between entities. *Neo4j* enables a natural modeling of interconnected data and supports powerful queries based on a fast traversal of relationships and complex graph patterns. Both nodes and relationships are associated with data properties, enabling a flexible schema that accommodates diverse data structures and also provide a rich and clear data visualization. Additional metadata are stored using *PostgreSQL* (<https://www.postgresql.org>, accessed on 8 September 2024), an open-source RDBMS widely adopted for its robustness, extensibility, and adherence to SQL standards. In the proposed scenario, useful metadata include contextual information about energy assets (e.g., location, links to external documentation or data sources) and custom attributes related to specific system requirements or applications (e.g., unique identifiers, project codes, user-defined tags).

Analytics procedures have been implemented using *scikit-learn* (<https://scikit-learn.org>, accessed on 8 September 2024), a popular open-source ML library. It provides a simple and efficient toolkit for data analysis, covering a wide range of activities, including classification, regression, and clustering. It also supports parallel and distributed computing, allowing users to leverage multi-core processors and distributed computing frameworks for accel-

erated model training and evaluation. Information obtained as a result of the processing procedures are used to further enhance the KG. As described in Section 3.3, the visualization layer encompasses all services that enhance the user experience by facilitating navigation and interaction with represented elements. In particular, a web application has been developed using *React* (<https://react.dev>, accessed on 8 September 2024), an open-source library for building interactive web-based User Interfaces (UIs) in the *JavaScript* programming language. It provides a component-based architecture, which allows developers to reuse components across different parts of an application. This modular approach makes it easier to build complex user interfaces by breaking them down into smaller components. The re-usability of components not only saves development time but also improves the maintainability of the codebase. Each component can be easily updated or modified without affecting other parts of the application. *React Native* (<https://reactnative.dev>, accessed on 8 September 2024) has been adopted to develop a cross-platform mobile application for iOS and Android platforms. It integrates a subset of React components, implemented for the web application, reducing the overall development cost compared to building separate native apps for each platform.

Finally, the proposed architecture integrates a conversational virtual assistant based on a foundation Large Language Model. It leverages retrieval-augmented generative AI functionalities to support problem-solving and advanced search activities, providing detailed and contextually accurate responses starting from data stored within the KG. Two distinct LLMs have been evaluated to implement the virtual assistant: *GPT-4* [55] and *Llama 2* [56]. *GPT-4* (<https://openai.com/research/gpt-4>, accessed on 8 September 2024) is a Large Multimodal Model (LMM) developed by OpenAI, capable of processing both image and text inputs and producing human-oriented outputs in natural language. Multiple models based on *GPT-4* are available, each with different capabilities and price points. *Llama 2* (<https://llama.meta.com/llama2>, accessed on 8 September 2024) proposes a collection of pre-trained and fine-tuned LLMs developed by Meta, ranging in scale from 7 billion to 70 billion parameters. It serves as a freely available alternative to closed models like *GPT-4*, enabling the development of customized and privacy-oriented solutions. Both platforms offer useful capabilities, including Natural Language Understanding (NLU), Text Generation (TG), Code Generation (CG), Question Answering (QA), and Task Automation (TA), satisfying diverse user needs and requirements. A comparative analysis between *GPT-4* and *Llama 2* is reported in Table 2, highlighting the main features of each platform, usually assisting designers in making informed decisions based on their specific expectations and preferences.

In the proposed implementation, privacy is an important concern. To address this, both at-rest and in-transit data encryption are exploited, ensuring that all sensitive information remains protected throughout the system's operation. Importantly, no data are uploaded to third-party services: even for ML, analytics, and visualization, all processing is carried out locally. Analogously, although the framework is compatible with *GPT-4*, an on-premises installation of *Llama 2* has been preferred to enhance the privacy of the solution. Therefore, *Llama 2* has been used to develop the virtual assistant, fine-tuned to meet the specific needs and requirements of energy grid scenarios. By utilizing an on-premises LLM backend, the virtual assistant operates by adopting a RAG approach, as reported in Figure 6, integrating input data stored in the KG while ensuring data privacy and minimizing reliance on external servers. Basically, the virtual assistant receives a user query; without RAG, the pre-trained LLM could be directly used to respond to the question, providing only basic information without knowledge of contextual data. On the contrary, by integrating a RAG approach, the query is converted into an embedding vector in order to retrieve relevant factual information available within the Neo4j database, e.g., detected events, information about energy infrastructures, and metering data. Relevant resources, obtained by comparing the query against the indexed documents, are used to expand the context of the input query, providing additional information and perspectives. The framework also supports different prompt templates applied to improve the outputs of LLMs conversations. The augmented query is then passed to the LLM for response generation,

which produces contextual outputs in natural language, leveraging domain-specific data. Furthermore, the platform offers flexibility by enabling interaction with cloud-based LLMs in accordance with specific privacy policies. This hybrid approach allows users to exploit the benefits of cloud-based models, such as advanced conversational skills, enhanced processing power, and access to a broader range of data, while still maintaining control over system-sensitive information.

Table 2. Comparative table of adopted LLMs.

Feature	GPT-4	Llama 2
License	commercial	Llama 2 community license
Input	text and images	text only
Model size	~1.76 T parameters	7 B, 13 B, 70 B parameters
Model customization	limited to selected organizations	highly customizable through RAG or fine-tuning procedures
Model updates	updates not immediately accessible to end users	periodic updates with potential new features
Tokens per prompt	up to 8192	up to 2048
Integration with other systems	through RESTful API calls	LLM backend required
Conversational skills	immersive conversations integrating multiple forms of media	fluent and natural-sounding responses for interactive chat experiences
Privacy concerns	data processed on external servers, potentially subject to data breaches or misuse	data processing occurs on-premises, reducing exposure to third-party entities

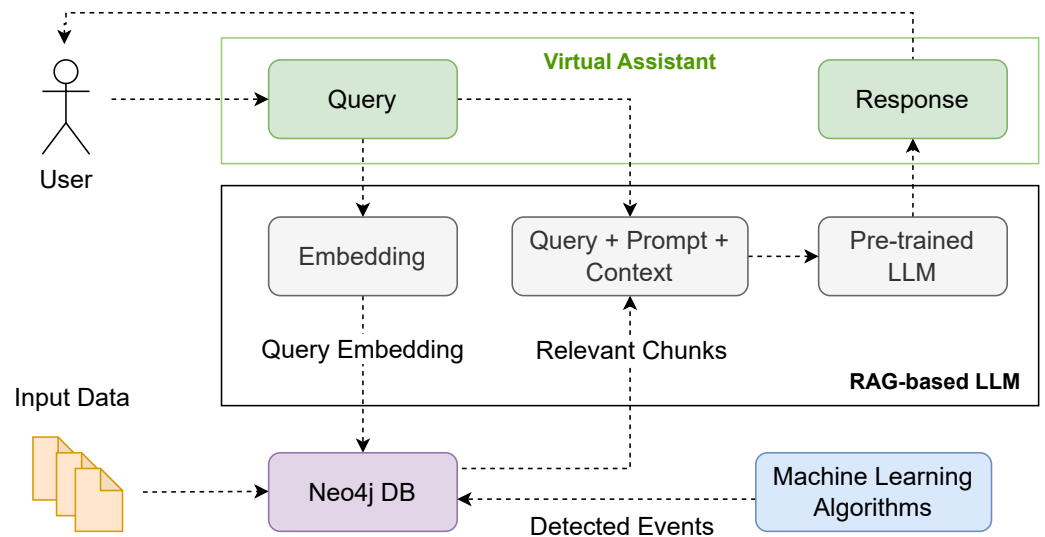


Figure 6. Retrieval augmented generation architecture.

4.2. Key Functionality Indicators

In order to thoroughly assess the improvements introduced by the proposed methodologies, the case study also serves as a benchmark to analyze how the integration of DTs, data-driven methods, LLM-based decision support, and RDF-based KGs enhances key operational aspects by comparing the proposed functionalities to baseline approaches used in the reference scenario. Specifically, the proposed approach has been analyzed with respect to several evaluation dimensions derived from the research aims in Section 1:

- (RA1) Enhancing grid management through DTs and data-driven approaches:
 - a. real-time data integration to provide updates and assessments about the grid infrastructure and to represent the asset status in a intuitive and easily interpretable format;
 - b. DT replication of real-world infrastructure in terms of asset behaviors and system processes;
 - c. failure and anomaly prevention to detect irregularities within the grid infrastructure that could lead to performance issues;
 - d. creation of a virtual representation of physical grid components that mirrors real-time operational data and enables remote control functionalities.
- (RA2) Leveraging LLMs for decision support:
 - a. exploitation of pre-trained models on domain-specific tasks requiring specialized language, terminology, and concepts of the energy sector;
 - b. understanding and generating textual content related to energy infrastructures for responding to complex, multi-faceted queries related to grid management;
 - c. provide comprehensible explanations related to the generated content and insights, increasing operator trust and understanding of its outputs;
 - d. LLM integration into existing operational workflows and automation systems within the energy infrastructure, in order to interact with other data management tools and platforms.
- (RA3) Integrating RAG approaches with knowledge graphs:
 - a. reference data model for annotating data in the knowledge graph, ensuring it accurately represents real-world relationships and supports relevant queries;
 - b. integration of data from various sources (e.g., IoT sensors, databases, log files) into the KG, expanding the scope of the RAG framework;
 - c. provide contextually relevant and domain-specific results, improving the accuracy of generated responses and reducing errors.

4.3. Case Study Analysis

This section describes how the proposed implementation has been adopted to enhance operations in the reference case study. The platform manages a wide section of the central Italy electric power distribution network, consisting of about 75,000 km of energy lines and 900 substations. It includes the distribution network of the municipality of Rome, composed of 70 primary cabins and more than 30,000 km of lines. In detail, the EDT platform serves as a comprehensive tool for managing and optimizing the grid infrastructure, exploiting predictive maintenance techniques leveraging data collected from smart meters.

Requirements of the case study can be summarized as: “Metering devices are deployed in the energy grid to monitor consumption in real-time. By analyzing the data generated by these meters, platform users are able to proactively identify potential equipment failures or performance degradation. Actionable alerts and recommendations will be forwarded to maintenance workers through the mobile application, providing useful information and guiding them to address identified issues effectively”.

A reference workflow is described in what follows, and each task or functionality is mapped to the specific research statements outlined in Section 4.2:

- System administrators access the web platform and display the whole energy grid, composed of power plants, substations, lines and poles, as depicted in Figure 7. For each power station, geographic coordinates are extracted from *OpenStreetMap* (OSM) [57] and queried via the *Overpass* API (https://wiki.openstreetmap.org/wiki/Overpass_API) (RA1.a). All data related to the grid infrastructure are stored in the reference KG (RA3.a, RA3.b).
- In the presence of warnings detected through ML algorithms, a red icon is shown on the map to highlight potential anomalies on the specific node (RA1.c).

- Each item can be selected in order to display detailed information and relevant properties, as well as alarm messages associated with the devices, hosted by a plant or asset. A list of available actions enabled on the platform is also shown, enabling the user to remotely interact with network-connected actuators. These actions provide direct remote control over the physical devices in the system, enabling tasks such as resetting equipment, adjusting set points, or initiating maintenance procedures (RA1.a, RA1.d).
- The system also integrates 2D models (Figure 8a), presenting selectable zones allowing users to open further representations of the same asset (RA1.b).
- Users interact with 3D models (Figure 8b) through a gesture-based NUI to perform actions like rotating the model, zooming in/out on specific components, or highlighting potential maintenance areas. Models respond dynamically, providing a realistic and immersive experience for exploring data regarding equipment conditions, structural components, and spatial relationships (RA1.b).
- The green button placed in the bottom right corner of the user interface opens a virtual assistant box (Figure 7). By means of simple requests in natural language (e.g., “Show me detailed information about recent alarms at plant X”), administrators receive useful information about monitored devices and detected events, including alarm type, reference timestamp, affected components, and suggested actions for resolution (RA2.a, RA2.b).
- After reviewing the alarm information, in case of a critical issue requiring immediate or preventive maintenance, system administrators notify technicians. A warning message is automatically composed, indicating the nature of the issue and the urgency for maintenance action, and sent to the maintenance operators responsible for the plant (RA2.c, RA2.d).
- Operators receive the notification through the mobile app and are shown information about device location (Figure 9a) and maintenance actions (Figure 9b), thereby minimizing maintenance time and ensuring the plant’s operational efficiency (RA1.a).
- Once the maintenance task is successfully completed, operators close the issue and provide notes or additional insights gained during the process. The platform receives the confirmation, updates the maintenance record in real-time, and notifies administrators or maintenance coordinators (RA1.a, RA2.d).
- Information collected from maintenance completions are annotated to enrich the RDF-based KG (RA3.a) with valuable insights and contextual data useful for future analysis (RA1.c). Data are analyzed periodically to identify recurring issues, improve maintenance procedures, and optimize responses provided by the RAG-based solutions (RA3.c).

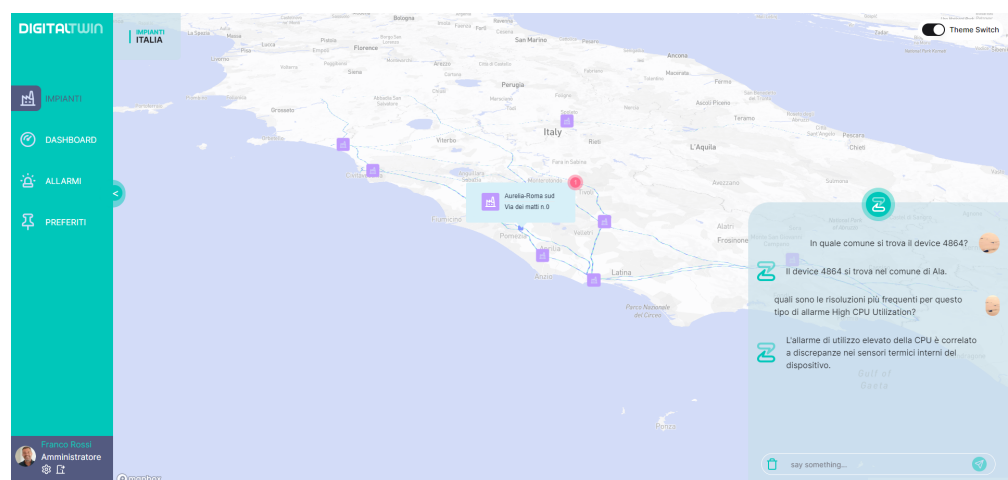
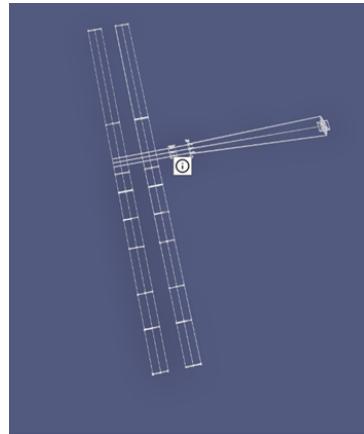
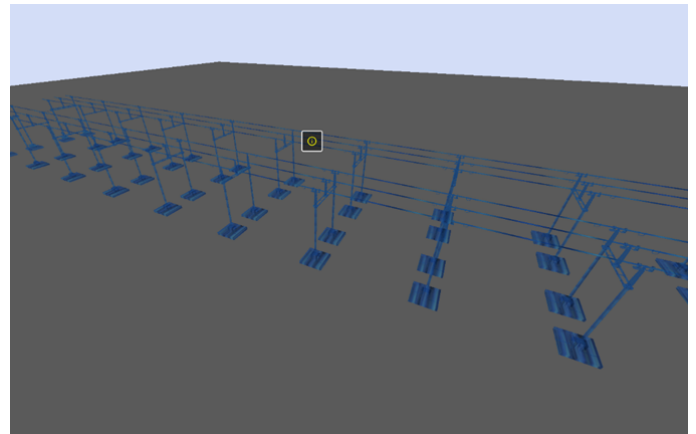


Figure 7. Digital representation of the energy grid.

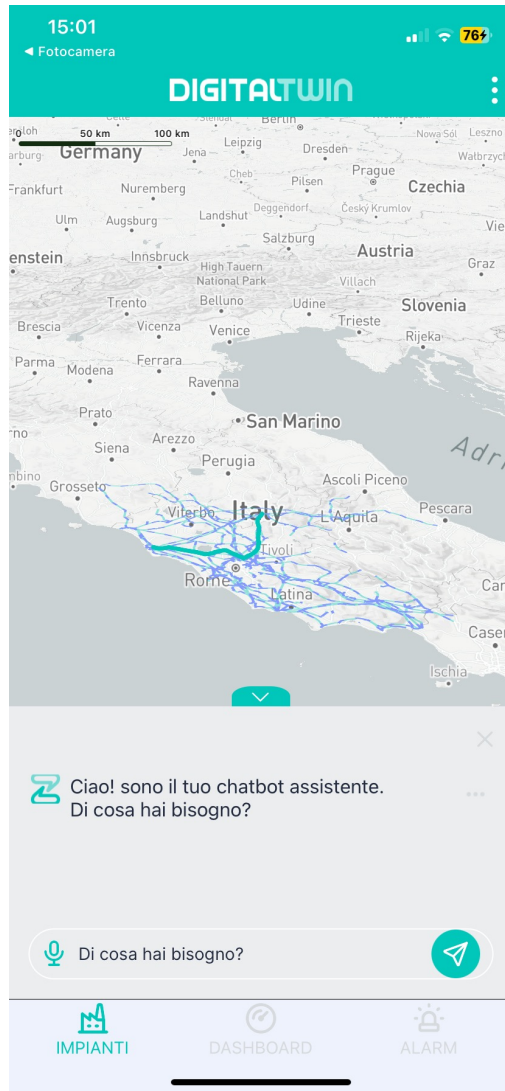


(a) 2D model



(b) 3D model

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



(a) Map of the energy grid



(b) Details about an alert message

Figure 9. Screenshots of the mobile application.

5. Discussions

This section assesses the proposal within the context of the state-of-the-art by means of a comparison with relevant related work and a discussion of open challenges and opportunities.

5.1. Platform Evaluation

Table 3 provides a comparative analysis of Digital-Twin (DT) frameworks across the research works described in Section 2.2 and this paper. It is structured to evaluate the support for a broad spectrum of features, categorized into four main areas:

- Infrastructure assesses how each DT framework leverages computational resources and manages physical components within the infrastructure.
- Data Intelligence examines the platform capabilities in terms of data analysis to extract knowledge from data generated within the framework.
- Cybersecurity assesses the capabilities of each framework for monitoring and protecting against threats, to ensure data integrity and confidentiality within the DT environments.
- Visualization details the human-machine interface solutions of each proposal, including support for interactive GUIs, GIS maps, 3D models, and NUIs.

The table employs a symbolic notation to indicate the level of support each referenced framework provides for the listed features: a check mark (✓) represents full support, a cross mark (✗) denotes lack of support, and a star (★) indicates partial support or instances where future updates are planned.

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	✗	✗	✗	✗	✗	✗	✗	✓

It can be noted that all the frameworks found to compare with the one described in this paper lack support for RAG capabilities based on KG and LLM integration, which enable one to expose EDT asset manipulation and knowledge-enhanced analytics capabilities through conversational NUIs. Furthermore, the majority of the proposed platforms lack focus on cybersecurity features, which are essential when managing critical infrastructures for human society, like energy grids.

The D-Arc [27] framework demonstrates support for many key features for an EDT framework, in particular for cloud infrastructure, IoT sensor networks, real-time data processing, data analytics, machine learning, asset simulation, scenario forecasting, access control, and interactive GUIs. It offers partial support for edge computing and system

auditing, alluding to some capabilities in these areas, albeit not explicitly detailed. However, the framework does not explicitly support physical asset control, knowledge-based capabilities via KG or LLM integration, GIS maps, and natural user interfaces, highlighting potential areas for future development. C³-Flow [28] prioritizes mathematical system modeling and computational distribution of the EDT models. Given this focus, certain features required for production-grade EDT platforms are not explicitly covered. Cybersecurity, data storage architecture, and GUIs are not explicitly detailed. Physical asset control appears to be possible, even if not explicitly performed in [28]. The framework designed in [29] for ECPSs leverages Amazon Web Services (AWSs) to outline a secure and integrated cloud-edge architecture. This infrastructure covers cybersecurity requirements by means of encrypted and authenticated data exchanges. Advanced characteristics like KGs, LLMs, and interactive GUIs are out of the scope of the performed experimentation, which focuses on essential operational functionalities and system optimization rather than comprehensive feature integration. The proposal in [30] primarily emphasizes the integration of EDT technology and IoT sensor networks, such as Phasor Measurement Units (PMUs) and Intelligent Electronic Devices (IEDs), to enhance real-time data processing, asset simulation, and scenario forecasting within power system control centers. Although the work highlights the GUIs for human intervention and suggests a degree of physical asset actuation through intelligent substations and control mechanisms, it lacks explicit discussion on data mining or RAG methods, security mechanisms, or geospatial components incorporating 3D models into maps. The DT for the power industry in [31] is one of the most feature-complete proposals, emphasizing aspects such as smart sensors, 5G communications, and big-data analytics. Moreover, the work briefly mentions KG integration in the infrastructure. GIS maps and 3D models are used in laser scanning technology for underground pipe network visualization. Nevertheless, the work does not describe any cybersecurity-related aspects. The EDT in [34] highlights how to improve predictive maintenance and fault detection in power distribution. The platform exploits cloud, edge, and IoT infrastructures for data collection and management and includes ML-based predictive maintenance and physical control. Although the work suggests the use of interactive interfaces and mentions GIS maps and 3D models, these aspects are not fully developed in the proposal.

As highlighted in Table 3, the proposed approach presents a feature-complete DT architecture for the energy infrastructure domain. The framework leverages the distribution of computational resources across the cloud-to-edge continuum, as explained in Section 3.1, and it is able to control geographically-distributed sensor networks. The data lake enables scalable storage and efficient management of both structured and unstructured data for advanced analytics and processing (Section 3.3), while physical asset control allows for real-time interaction with and management of network-connected devices, supporting remote actuation and monitoring (Section 3.4). The framework also integrates essential functions such as real-time data processing, analytics, and machine learning to optimize energy infrastructure management, as explained in Section 3.3. These capabilities are enhanced with innovative features like RAG, which exploits ontologies and KGs to contextualize and enrich the outputs of a foundation LLMs, enabling more accurate and relevant responses for decision support, as described in Section 3.4. The prototype integrates robust cybersecurity measures, including multi-factor authentication, role-based access control, and continuous system monitoring to detect and respond to anomalies in real time, as outlined in Section 3.5. Finally, data visualization and the GUI are central to the platform, offering support for GIS maps, 3D models, and a NUI. Powered by LLMs, the platform enables chat-like interactions for intuitive system control and information retrieval, as demonstrated in the prototype outlined in Section 4.3.

5.2. Challenges and Opportunities

The integration of ICT technologies and digital twins in energy management scenarios introduces a spectrum of challenges and opportunities, ranging from technical to human aspects [58]. Based on the experiences that emerged during the design of the framework and

the development of the proposed case study, the key factors mostly influencing digital-twin implementation and adoption in energy systems are discussed in what follows.

Integration with legacy systems: existing energy infrastructures frequently include Supervisory Control and Data Acquisition (SCADA) systems for monitoring and controlling industrial processes in real-time. These devices often use proprietary protocols and technologies that may not easily interface with modern applications. The development of standardized communication protocols and middleware solutions, such as the Open Platform Communications Unified Architecture (OPC-UA) [59], aims to enable a seamless integration with digital twins. However, legacy systems generate vast amounts of data with different rates, formats, and structures. Integrating diverse data sources can be complex and time-consuming, as it requires data mapping, normalization, and transformation processes to ensure data consistency and accuracy. Processing inaccurate data can have deep implications for the accuracy and effectiveness of predictive models and analytics. When noisy data is fed into ML models, the output becomes unreliable, leading to flawed conclusions and eroding trust in data-driven, decision-making processes.

In addition, many legacy systems are built on outdated hardware and software platforms, introducing several security risks, such as unauthorized access, data breaches, and cyberattacks. In the presence of discontinued or unmaintained software, systems are no longer protected against emerging threats. Hackers actively target unmaintained software as they know that any discovered vulnerabilities will remain unpatched, providing them with opportunities to exploit weaknesses and compromise systems. Legacy systems also use outdated authentication mechanisms and access controls that lack modern security features such as strong encryption keys and protocols, multi-factor authentication, and role-based access control, leaving sensitive data vulnerable to interception. Ensuring the security, privacy, and integrity of data collected and transmitted by legacy systems is essential to protect critical assets and business information. Along with the adoption of novel technologies, migration toward *defense-in-depth* strategies [60] and *zero-trust architectures* [61] offers compelling opportunities for addressing existing issues and improving the resilience of current energy systems.

Human factors in digital-twin implementation: the successful implementation of DT technology in energy management scenarios relies not only on technical aspects but also on a deep understanding and integration of human factors encompassing the cognitive, social, and organizational elements that influence how individuals interact with technology and make decisions [62]. Understanding how emerging approaches can be integrated with human activities is a prerequisite to unlock the full potential of DTs. As observed in [63], digital twins are not only expert-centric tools. They can play an active role in decision-making, task completion, and adaptation to evolving conditions, significantly reshaping the responsibilities of human operators. In this evolving landscape, end-users are anticipated to collaborate closely with DT applications, overseeing their performance and results in specific tasks. In particular, considering the following human factors in the context of energy infrastructures is crucial for ensuring user acceptance, effective utilization, and overall success:

- The design of the UI plays an important role in facilitating *user interaction* with the DT platform in order to create intuitive and user-friendly interfaces, providing relevant information, insights, and predictions. UIs should show information in a way that aligns with the users' cognitive processes, facilitating quicker and more accurate decision-making in response to dynamic energy scenarios. Tailoring the interface to the specific needs and expertise of energy management professionals ensures efficient use and minimizes the learning curve associated with adopting new technologies [64]
- Increasing *trust* in EDT systems is paramount for their successful adoption. Human operators must have confidence in the accuracy and reliability of the DT's representations. Research must increase focus on developing validation mechanisms, transparent communication of uncertainties, and incorporating user feedback to make these com-

plex models more understandable to (non-expert) users and enhance trust in the system [65].

- The complexity of EDT platforms often requires comprehensive *training* and *education* programs for end-users. Energy management operators and decision-makers need to be familiar with the capabilities and functionalities of the DT. Ongoing training programs help users harness the full potential of the technology, make informed decisions, and troubleshoot issues effectively [66].
- As DTs collect and process large amounts of data, cybersecurity considerations become crucial, aiming to ensure responsible data use, privacy protection, and compliance with relevant regulations. In particular, EDTs must be treated as critical systems in which security issues need to be considered in terms of confidentiality, integrity, and availability of both data and resources, along with privacy issues with respect to entities as well as location and status of assets [67].

6. Conclusions

This paper has introduced a novel data-driven and knowledge-based energy digital-twin framework, integrating (i) data stream gathering from sensing devices; (ii) data stream mining and machine learning for predictive maintenance and analytics; (iii) an ontology-based knowledge graph combining annotated data, background domain knowledge, and ML outcomes; (iv) retrieval-augmented generation exploiting the KG to enrich a foundation large language model, leveraged to provide a conversational virtual agent for assisting users in making decisions on energy infrastructure performance and maintenance management; (v) a natural user interface combining conversational interaction with geospatial information presentation based on 3D models of infrastructure assets in a geographic information system. All these functions are supported by a cohesive scalable cloud-edge microservice architecture, characterized by functional layers and by cross-layers dedicated to device support, data integration, and cybersecurity. A prototype of the proposed framework has been developed, leveraging commercial-off-the-shelf components for a case study regarding the management of a section of the energy distribution infrastructure in central Italy. Currently, the platform only supports Italian as the reference language. The system is mainly conceived for Italian companies of the energy sector, and several documents and regulations referenced in the platform are only available in Italian. Using the local language facilitated easier text interpretation, also enhancing clarity and understanding of the platform functionalities. However, as part of future work, multi-language support is being considered and will be developed as an ongoing activity. This will enable the platform to accommodate users from diverse linguistic backgrounds, further expanding its accessibility and usability.

Moreover, future work regards expanding the prototype into a complete implementation of the proposed framework and using it in pilot trials in production environments in order to gather feedback, enabling the assessment of critical aspects of DT implementation, including the following: (i) performance and scalability; (ii) ease of integration with a broad range of legacy systems; (iii) user interaction effectiveness; (iv) acceptability and trust of the system from end users. Further investigation will concern improving and expanding ML methods and applications, with a focus on eXplainable Artificial Intelligence (XAI) approaches to further increase system trust, semantic representation refinement for improving RAG effectiveness for the conversational virtual assistant, and enhancing asset control and scenario simulation use cases within the framework.

Author Contributions: Conceptualization, S.I., G.L., M.R., F.S., D.M. and M.N.; methodology, S.I., D.L., G.L. and F.S.; software, D.M. and M.N.; validation, S.I., D.L. and D.M.; formal analysis, G.L., M.R. and F.S.; investigation, S.I., D.L., G.L., M.R. and F.S.; resources, D.L., D.M. and M.N.; data curation, D.M. and M.N.; writing—original draft preparation, all authors; writing—review and editing, G.L., M.R. and F.S.; visualization, D.L., D.M. and M.N.; supervision, M.R. and F.S.; project administration, S.I., D.M. and M.N.; funding acquisition, M.R., D.M. and M.N. All authors have read and agreed to the published version of the manuscript.

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

Data Availability Statement: Data available on request due to restrictions (e.g., privacy, legal, or ethical reasons).

Conflicts 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 appeared to influence the work reported in this paper.

References

- Joseph, A.; Balachandra, P. Smart grid to energy internet: A systematic review of transitioning electricity systems. *IEEE Access* **2020**, *8*, 215787–215805. [CrossRef]
- Cronin, J.; Anandarajah, G.; Dessens, O. Climate change impacts on the energy system: A review of trends and gaps. *Clim. Change* **2018**, *151*, 79–93. [CrossRef]
- Schleussner, C.F.; Rogelj, J.; Schaeffer, M.; Lissner, T.; Licker, R.; Fischer, E.M.; Knutti, R.; Levermann, A.; Frieler, K.; Hare, W. Science and policy characteristics of the Paris Agreement temperature goal. *Nat. Clim. Change* **2016**, *6*, 827–835. [CrossRef]
- Dileep, G. A survey on smart grid technologies and applications. *Renew. Energy* **2020**, *146*, 2589–2625. [CrossRef]
- Statista. Smart Grid Market Value Worldwide 2022–2028. 2023. Available online: <https://www.statista.com/study/111848/smart-grids-worldwide/> (accessed on 8 September 2024).
- Alotaibi, I.; Abido, M.A.; Khalid, M.; Savkin, A.V. A comprehensive review of recent advances in smart grids: A sustainable future with renewable energy resources. *Energies* **2020**, *13*, 6269. [CrossRef]
- Alasali, F.; Itradat, A.; Abu Ghalyon, S.; Abudayyeh, M.; El-Naily, N.; Hayajneh, A.M.; AlMajali, A. Smart Grid Resilience for Grid-Connected PV and Protection Systems under Cyber Threats. *Smart Cities* **2023**, *7*, 51–77. [CrossRef]
- Do Amaral, J.; Dos Santos, C.; Montevechi, J.; De Queiroz, A. Energy Digital Twin Applications: A Review. *Renew. Sustain. Energy Rev.* **2023**, *188*, 113891. [CrossRef]
- Manickam, R.; Vollmar, J.; Prabhakar, G.M. User Experience–Digital Twin Maturity Model (UX-DTMM). In Proceedings of the International Conference on Research into Design, Bangalore, India, 9–11 January 2023; pp. 877–889.
- Cyganiak, R.; Wood, D.; Lanthaler, M. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation, W3C. 2014. Available online: <https://www.w3.org/TR/rdf11-concepts/> (accessed on 8 September 2024).
- Hossein Motlagh, N.; Mohammadrezaei, M.; Hunt, J.; Zakeri, B. Internet of Things (IoT) and the energy sector. *Energies* **2020**, *13*, 494. [CrossRef]
- 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*, 77555–77571. [CrossRef]
- 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. [CrossRef]
- Omitaomu, O.A.; Niu, H. Artificial intelligence techniques in smart grid: A survey. *Smart Cities* **2021**, *4*, 548–568. [CrossRef]
- Cao, K.; Liu, Y.; Meng, G.; Sun, Q. An overview on edge computing research. *IEEE Access* **2020**, *8*, 85714–85728. [CrossRef]
- Minh, Q.N.; Nguyen, V.H.; Quy, V.K.; Ngoc, L.A.; Chehri, A.; Jeon, G. Edge Computing for IoT-Enabled Smart Grid: The Future of Energy. *Energies* **2022**, *15*, 6140. [CrossRef]
- Arcas, G.I.; Cioara, T.; Anghel, I.; Lazea, D.; Hangan, A. Edge Offloading in Smart Grid. *Smart Cities* **2024**, *7*, 680–711. [CrossRef]
- 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*, 183–204. [CrossRef]
- Dhaou, I.B. Design and Implementation of an Internet-of-Things-Enabled Smart Meter and Smart Plug for Home-Energy-Management System. *Electronics* **2023**, *12*, 4041. [CrossRef]
- 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. [CrossRef]
- Aljadani, N.; Gazdar, T. A novel security architecture for WSN-based applications in smart grid. *Smart Cities* **2022**, *5*, 633–649. [CrossRef]
- Xu, Z.; Salehi Shahraki, A.; Rudolph, C. Blockchain-Based Malicious Behaviour Management Scheme for Smart Grids. *Smart Cities* **2023**, *6*, 3005–3031. [CrossRef]
- Yu, W.; Patros, P.; Young, B.; Klinac, E.; Walmsley, T.G. Energy digital twin technology for industrial energy management: Classification, challenges and future. *Renew. Sustain. Energy Rev.* **2022**, *161*, 112407. [CrossRef]
- Singh, M.; Fuenmayor, E.; Hinchy, E.P.; Qiao, Y.; Murray, N.; Devine, D. Digital twin: Origin to future. *Appl. Syst. Innov.* **2021**, *4*, 36. [CrossRef]
- Dietz, M.; Pernul, G. Digital twin: Empowering enterprises towards a system-of-systems approach. *Bus. Inf. Syst. Eng.* **2020**, *62*, 179–184. [CrossRef]

26. 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. [\[CrossRef\]](#)
27. Gourisetti, S.N.G.; Bhadra, S.; Sebastian-Cardenas, D.J.; Touhiduzzaman, M.; Ahmed, O. A Theoretical Open Architecture Framework and Technology Stack for Digital Twins in Energy Sector Applications. *Energies* **2023**, *16*, 4853. [\[CrossRef\]](#)
28. Liao, H.; Zhou, Z.; Liu, N.; Zhang, Y.; Xu, G.; Wang, Z.; Mumtaz, S. Cloud-Edge-Device Collaborative Reliable and Communication-Efficient Digital Twin for Low-Carbon Electrical Equipment Management. *IEEE Trans. Ind. Inform.* **2023**, *19*, 1715–1724. [\[CrossRef\]](#)
29. Saad, A.; Faddel, S.; Mohammed, O. IoT-Based Digital Twin for Energy Cyber-Physical Systems: Design and Implementation. *Energies* **2020**, *13*, 4762. [\[CrossRef\]](#)
30. 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, 3–7 June 2018; pp. 1–6.
31. Huang, J.; Zhao, L.; Wei, F.; Cao, B. The Application of Digital Twin on Power Industry. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *647*, 012015. [\[CrossRef\]](#)
32. 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*, 2747. [\[CrossRef\]](#)
33. Liu, T.; Yu, H.; Yin, H.; Zhang, Z.; Sui, Z.; Zhu, D.; Gao, L.; Li, Z. 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, 8–11 April 2021; pp. 383–387.
34. 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, 28–31 May 2020; pp. 29–33.
35. Uslar, M.; Rohjans, S.; Neureiter, C.; Pröbstl Andrén, F.; Velasquez, J.; Steinbrink, C.; Efthymiou, V.; Migliavacca, G.; Horsmanheimo, S.; Brunner, H.; 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*, 258. [\[CrossRef\]](#)
36. Zhang, J.; Ma, M.; Wang, P.; Sun, X.d. Middleware for the Internet of Things: A survey on requirements, enabling technologies, and solutions. *J. Syst. Archit.* **2021**, *117*, 102098. [\[CrossRef\]](#)
37. Deng, S.; Zhao, H.; Fang, W.; Yin, J.; Dustdar, S.; Zomaya, A.Y. Edge intelligence: The confluence of edge computing and artificial intelligence. *IEEE Internet Things J.* **2020**, *7*, 7457–7469. [\[CrossRef\]](#)
38. Loseto, G.; Scioscia, F.; Ruta, M.; Gramegna, F.; Ieva, S.; Fasciano, C.; Bilenchi, I.; Loconte, D. Osmotic cloud-edge intelligence for IoT-based cyber-physical systems. *Sensors* **2022**, *22*, 2166. [\[CrossRef\]](#)
39. Velepucha, V.; Flores, P. A survey on microservices architecture: Principles, patterns and migration challenges. *IEEE Access* **2023**, *11*, 88339–88358. [\[CrossRef\]](#)
40. Mishra, B.; Kertesz, A. The use of MQTT in M2M and IoT systems: A survey. *IEEE Access* **2020**, *8*, 201071–201086. [\[CrossRef\]](#)
41. 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*, 154–167. [\[CrossRef\]](#)
42. 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: New York, NY, USA, 2023; pp. 115–143.
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, The Netherlands, 11–14 September 2017.
44. Lehmann, J.; Isele, R.; Jakob, M.; Jentzsch, A.; Kontokostas, D.; Mendes, P.N.; Hellmann, S.; Morsey, M.; Van Kleef, P.; Auer, S.; et al. Dbpedia—A large-scale, multilingual knowledge base extracted from wikipedia. *Semant. Web* **2015**, *6*, 167–195. [\[CrossRef\]](#)
45. Sawadogo, P.; Darmont, J. On data lake architectures and metadata management. *J. Intell. Inf. Syst.* **2021**, *56*, 97–120. [\[CrossRef\]](#)
46. Roy, D.; Srivastava, R.; Jat, M.; Karaca, M.S. A complete overview of analytics techniques: Descriptive, predictive, and prescriptive. In *Decision Intelligence Analytics and the Implementation of Strategic Business Management*; Springer: Cham, Switzerland, 2022; pp. 15–30.
47. 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. [\[CrossRef\]](#)
48. Alimi, O.A.; Ouahada, K.; Abu-Mahfouz, A.M. A review of machine learning approaches to power system security and stability. *IEEE Access* **2020**, *8*, 113512–113531. [\[CrossRef\]](#)
49. 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, 27–29 June 2017; pp. 1–12.
50. Nguyen, T.N.; 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, 20 July–1 August 2020.
51. 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*, 5795–5826. [\[CrossRef\]](#)

52. 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*, 3580–3599. [[CrossRef](#)]
53. Gunduz, M.Z.; Das, R. Cyber-security on smart grid: Threats and potential solutions. *Comput. Netw.* **2020**, *169*, 107094. [[CrossRef](#)]
54. Sakimura, N.; Bradley, J.; Jones, M.; De Medeiros, B.; Mortimore, C. OpenID Connect Core 1.0 incorporating errata set 2. *OpenID Found. Specif.* **2023**. Available online: https://openid.net/specs/openid-connect-core-1_0.html (accessed on 21 October 2024).
55. Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. Gpt-4 technical report. *arXiv* **2023**, arXiv:2303.08774.
56. Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv* **2023**, arXiv:2307.09288.
57. Haklay, M.; Weber, P. Openstreetmap: User-generated street maps. *IEEE Pervasive Comput.* **2008**, *7*, 12–18. [[CrossRef](#)]
58. Adel, A. Unlocking the future: Fostering human–machine collaboration and driving intelligent automation through industry 5.0 in smart cities. *Smart Cities* **2023**, *6*, 2742–2782. [[CrossRef](#)]
59. Veichtlbauer, A.; Ortmayer, M.; Heistracher, T. OPC UA integration for field devices. In Proceedings of the 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), Emden, Germany, 24–26 July 2017; pp. 419–424.
60. 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*, 367–378. [[CrossRef](#)]
61. Teerakanok, S.; Uehara, T.; Inomata, A. Migrating to zero trust architecture: Reviews and challenges. *Secur. Commun. Netw.* **2021**, *2021*, 1–10. [[CrossRef](#)]
62. Neumann, W.P.; Winkelhaus, S.; Grosse, E.H.; Glock, C.H. Industry 4.0 and the human factor—A systems framework and analysis methodology for successful development. *Int. J. Prod. Econ.* **2021**, *233*, 107992. [[CrossRef](#)]
63. Agrawal, A.; Thiel, R.; Jain, P.; Singh, V.; Fischer, M. Digital Twin: Where do humans fit in? *Autom. Constr.* **2023**, *148*, 104749. [[CrossRef](#)]
64. 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*, 20–34. [[CrossRef](#)]
65. Bonney, M.S.; de Angelis, M.; Dal Borgo, M.; Wagg, D.J. Contextualisation of information in digital twin processes. *Mech. Syst. Signal Process.* **2023**, *184*, 109657. [[CrossRef](#)]
66. Molino, M.; Cortese, C.G.; 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*, 2438. [[CrossRef](#)] [[PubMed](#)]
67. Alcaraz, C.; Lopez, J. Digital twin: A comprehensive survey of security threats. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 1475–1503. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.