Engineering Applications of Artificial Intelligence 164 (2026) 113400

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Research paper

A federated Artificial Intelligence testing and experimentation facility for the European energy sector

Elissaios Sarmas [a,*], Alexandre Lucas [b], Andrés Acosta [c], Ferdinanda Ponci [c], Pedro Rodriguez [d], Vangelis Marinakis [a]

[a] Decision Support Systems Laboratory, National Technical University of Athens, Athens, Greece
[b] INESC TEC, Porto, Portugal
[c] Institute for Automation of Complex Power Systems, E.ON Energy Research Center, RWTH Aachen University, Aachen, Germany
[d] Luxembourg Institute of Science and Technology, Luxembourg

ARTICLE INFO

Keywords:
Artificial Intelligence in energy
Artificial Intelligence-empowered software engineering
Energy systems
Federated learning
Digital twins

ABSTRACT

The application of Artificial Intelligence (AI) in the energy sector offers new opportunities for developing flexible, efficient, and sustainable infrastructures. Nevertheless, real-world deployment is still constrained by the lack of large-scale, integrated environments that can evaluate advanced algorithms under realistic operating conditions while ensuring regulatory compliance. This paper presents EnerTEF (which stands for Energy Testing and Experimentation Facility), a federated platform for testing and experimentation in the energy sector designed to address this gap. We introduce a unified TEF architecture that enables full-stack evaluation of intelligent systems, including predictive modeling, optimization, learning under data distribution shifts and federated learning across geographically distributed sites. The framework integrates high-fidelity digital twins, a privacy-preserving data exchange framework and regulatory sandboxing to support transparent, explainable and robust AI development. EnerTEF demonstrates how such a framework can be deployed in critical energy domains through three real-world scenarios including short-term hydropower generation forecasting, coordination between distribution network operators and distributed energy resources and real-time optimization of self-consumption for municipal buildings. Results show that EnerTEF effectively enables the development of novel AI models, improves cross-context generalizability and supports innovation for complex energy infrastructures, ultimately creating a practical, scalable path for addressing different energy-related problems and heterogeneous data.

1. Introduction

The convergence of digital technologies and energy systems has catalyzed a transformation in how electricity is generated, distributed and consumed (Wu et al., 2021). Among these technologies, Artificial Intelligence (AI) has emerged as a pivotal enabler in the shift toward more flexible, efficient and sustainable energy infrastructures. AI applications are now central to predictive maintenance, load forecasting, fault detection, distributed energy resource management, demand-side optimization and real-time grid operation (Ejiyi et al., 2025; Khan et al., 2022; Rajaperumal and Columbus, 2025). Despite this momentum, the transition from innovative AI concepts to tested, trusted and market-ready tools in the energy domain remains slow and fragmented.

A core barrier to this transition lies in the lack of large-scale, integrated environments that can rigorously evaluate AI-driven energy solutions under conditions approximating real-world operational complexity (Kyriakarakos, 2025). Existing testbeds often operate in isolation, focused on narrow technical scopes or single-domain use cases and lack the infrastructure needed to assess AI technologies' scalability, generalizability and compliance with regulatory standards. As a result, many AI solutions remain stuck at technology readiness levels (TRL) 4–6, unable to progress to deployment due to insufficient testing frameworks and validation pathways (Wijaya and Asif, 2025; Lavin et al., 2022).

At the European level, this gap is becoming increasingly critical. The European Commission's Digitalisation of Energy Action Plan (European Commission, 2022) emphasizes the strategic need for integrating digital technologies, and most importantly AI, into the energy sector to accelerate decarbonization, increase grid flexibility and empower consumers. Meanwhile, the proposed Artificial Intelligence Act (European Parliament and Council of the European Union, 2024) introduces

* Corresponding author.
E-mail address: esarmas@epu.ntua.gr (E. Sarmas).

https://doi.org/10.1016/j.engappai.2025.113400
Received 14 August 2025; Received in revised form 16 November 2025; Accepted 28 November 2025
Available online 30 November 2025

Nomenclature

AI	Artificial Intelligence
API	Application Programming Interface
CPU	Central Processing Unit
DER	Distributed Energy Resources
DPA	Data Processing Agreement
DSO	Distribution System Operator
DT	Digital Twin
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit
GUI	Graphical User Interface
HIL	Hardware-in-the-Loop
HPC	High-Performance Computing
HVAC	Heating, Ventilation, and Air Conditioning
IAM	Identity and Access Management
IDSA	International Data Spaces Association
IID	Independent and Identically Distributed
KPI	Key Performance Indicator
NTP	Network Time Protocol
PaaS	Platform as a Service
PTP	Precision Time Protocol
RES	Renewable Energy Sources
SCC	Standard Contractual Clauses
SLA	Service-Level Agreement
TEF	Testing and Experimentation Facility
TPU	Tensor Processing Unit
TSO	Transmission System Operator
UML	Unified Modeling Language

stringent requirements for transparency, accountability, robustness and ethical compliance for AI systems, particularly those deployed in high-impact domains such as energy. Without dedicated facilities to support these goals, Europe risks falling behind in the deployment of trustworthy AI solutions tailored to its complex, heterogeneous energy systems.

To address this structural deficiency, we introduce EnerTEF, a federated Testing and Experimentation Facility (TEF) specifically designed to validate AI-driven energy applications across multiple contexts, stakeholders and geographies. EnerTEF is conceived as a modular, distributed and extensible infrastructure that links physical and digital test environments under a common orchestration and governance framework. The initiative draws on recent research and development in AI model deployment, federated learning, digital twins, data sovereignty and human-in-the-loop experimentation. By doing so, it creates a scalable pathway from laboratory innovation to real-world impact.

EnerTEF distinguishes itself from previous initiatives by embracing a federated architecture composed of five primary nodes and multiple satellite facilities. Each node is dedicated to a specific domain within the energy value chain, ranging from renewable energy sources (RES) and building energy management to electric mobility and transmission system operations (TSOs). These nodes are integrated via a unified orchestration layer and a shared data infrastructure, facilitating secure data exchange, collaborative experimentation and cross-node interoperability. This federated design supports heterogeneous experimentation scenarios, including AI model training under diverse regulatory regimes, distributed algorithm validation and synthetic-real hybrid simulation. In addition to its technical architecture, EnerTEF introduces an ecosystemic approach to AI innovation in the energy sector. It operationalizes the principles of co-creation through "AI Living Labs", wherein technology developers, domain experts, regulators and end-users collaboratively define test scenarios, success criteria and ethical benchmarks. It also incorporates "regulatory sandboxes" to explore how AI technologies interact with current and emerging policy frameworks, enabling experimentation in a legally bounded, yet adaptive environment. By aligning with ongoing initiatives like the European Energy Data Space (Egorov et al., 2025) and the GAIA-X architecture (Tardieu, 2022), EnerTEF is designed to contribute not only to technological validation but also to the standardization and market readiness of AI solutions. EnerTEF was conceived, architected, and implemented by the authors and no prior public platform or external implementation predates this work. The main contributions of this study are presented below:

- We introduce an AI Testing and Experimentation platform that enables federated learning pipelines via an AI experimentation workbench and offers real-time access to physical assets through well-synchronized digital twins for a wide range of applications in the energy sector.
- We offer a sovereign framework for secure data exchange via an International Data Spaces Association (IDSA) compliant data space, removing any requirement to centralize raw data.
- We specify and implement protocols for orchestrated access to HPC resources across edge, on-premise, and cloud with bandwidth-aware scheduling, preemptive jobs, checkpointing and precise twin–asset time alignment, all under a unified security and access-control layer.
- We validate the EnerTEF framework on three operational real-world use cases, namely hydropower forecasting, distribution grid operations and predictive building control for optimization of self-consumption.

The remainder of this paper is structured as follows. Section 2 reviews related work across key thematic areas, including AI deployment in energy systems, TEF infrastructures, digital twin integration, federated learning and regulatory alignment. Section 3 presents the methodological framework underlying EnerTEF, detailing the structure of its nodes, federated architecture, orchestration platform and experimentation workflows. Section 4 introduces the application scenario and illustrates the system's operational capabilities through three representative use cases: hydropower forecasting (Section 4.2), distribution networks co-optimization (Section 4.3) and real-time predictive building control for optimizing buildings' self-consumption (Section 4.4) to illustrate practical benefits and policy alignment. Section 5 provides a critical discussion of the platform's methodological contributions, deployment insights and institutional implications. Finally, Section 6 concludes with a summary of findings and outlines directions for future research and policy integration.

2. Related work

The deployment of AI in energy systems has generated a surge of innovation across forecasting, control and optimization applications (Mazhar et al., 2025). However, as interest in AI-enabled energy systems matures, so does the recognition of a critical bottleneck: the absence of scalable, federated and regulation-compliant TEF infrastructures (Kallert et al., 2021; Wichert et al., 2001). While numerous efforts have explored simulation, data-driven modeling and localized testbeds in the past, very few efforts offer unified environments that support trustworthy AI deployment across the complex, multi-stakeholder and multi-layered structure of modern energy systems. This section reviews the state of the art across five thematic areas.

2.1. AI testing and experimentation infrastructures in energy systems

AI has become a cornerstone of digital transformation in the energy sector, particularly as energy systems transition toward decentralized, distributed and data-intensive architectures (Shaukat et al., 2023; Thombs, 2019). AI models, ranging from supervised learning algorithms for forecasting (Sarmas et al., 2022, 2024) to deep reinforcement learning for adaptive control (Zhang et al., 2021), are increasingly applied across various layers of the energy value chain. In generation, predictive models estimate solar irradiance or wind speeds to optimize renewable resource utilization (Ssekulima et al., 2016). In distribution, anomaly detection tools monitor voltage deviations or equipment faults in real-time (Sawas and Farag, 2023). On the consumer side, AI-powered energy management systems dynamically schedule load consumption (Papias et al., 2024), enable demand response and optimize storage dispatch (Toderean et al., 2025), particularly in smart buildings and energy communities (Hou et al., 2024).

Despite these advances, the deployment of AI in operational energy systems remains uneven. One fundamental challenge lies in the heterogeneity and fragmentation of data sources. Energy datasets are often siloed, domain-specific and of varying quality, complicating model training and validation. Moreover, most AI models are developed and tested under idealized conditions using synthetic or lab-scale data, which does not reflect the complexity, volatility and uncertainty inherent in real-world power systems. Another limitation is the "black-box" nature of many AI algorithms, especially deep learning models, which impedes transparency, explainability and operator trust. For stakeholders such as distribution system operators (DSOs), regulators and grid planners, the lack of interpretability and confidence in AI predictions is a major barrier to adoption.

The need for robust testing environments for energy systems has long been recognized, particularly for verifying hardware performance, ensuring grid stability and validating control strategies under edge-case scenarios. To this end, several major research projects and platforms have been established. ERIGrid 2.0 (Pellegrino et al., 2020), for example, integrates hardware-in-the-loop simulation (Ebe et al., 2018), power hardware testbenches and co-simulation tools to evaluate distributed energy resource integration and smart grid configurations (Heussen et al., 2019). Similarly, platforms such as FLEXGRID (Efthymiopoulos et al., 2023) and InterFlex (Khomami et al., 2020) provide environments for testing the interplay between grid automation technologies and market-based flexibility services.

However, these infrastructures primarily cater to power systems engineering and do not address the full requirements of AI-based service testing. Most focus on electrical or cyber–physical validation, such as protection schemes, communication protocols, or voltage regulation, and lack the architectural layers required for AI experimentation, such as data ingestion pipelines, machine learning libraries, or orchestration environments for training, tuning and testing models (González and Evans, 2019). Furthermore, few platforms support testing at scale, across different countries or network configurations, which limits their relevance in scenarios involving cross-border data flows, regulatory diversity, or federated architectures (Aaronson, 2019). EnerTEF fills this gap by providing interoperable access to energy data, coordinated via a common orchestration and governance framework. This enables not just the testing of individual algorithms or devices, but the assessment of full-stack AI services, including cloud–edge deployments, privacy mechanisms and regulatory compliance features, in a way that mirrors their eventual deployment environment.

2.2. Distributed digital twins & federated learning

On the one hand, digital twins (DTs) have emerged as one of the most promising tools for energy systems modeling, enabling high-resolution, bidirectional representations of physical infrastructure (Bazmohammadi et al., 2021). In energy contexts, DTs are employed to mirror the behavior of grid elements, buildings, generation assets and consumer systems, allowing operators and researchers to simulate future scenarios, test control strategies and monitor system health (Kabir et al., 2024; Djebali et al., 2024). Platforms such as SENTINEL (Martin et al., 2020) offer sophisticated DT capabilities that include integration of real-time sensor data, forecasting modules and optimization engines. Nevertheless, most DT applications in the energy sector are limited in scope and scale. They are often customized for specific environments or purposes, such as building-level energy optimization or wind turbine diagnostics, and lack standardization in terms of data formats, APIs and semantic models (Wolk and Reinhart, 2025). Moreover, their use is typically confined to software-only environments or simulations, with limited interoperability with physical assets or experimental infrastructures. This disconnect between simulation and field validation hampers the ability to use DTs as full-fledged testing environments for AI services, particularly in safety-critical or compliance-intensive applications (Nicholson et al., 2012). EnerTEF aims to integrate DTs into its federated architecture as operational assets. Each node of the EnerTEF network is paired with digital replicas of its physical infrastructure, enabling synchronized real-time data flows and model validation. DTs are active components of the experimentation process, supporting adaptive testing, hybrid simulation and the transfer of AI models across physical and virtual domains.

On the other hand, Federated learning (FL) is a novel approach to training machine learning models without centralizing sensitive or proprietary data. Instead, model training is distributed across multiple local nodes, with only model parameters, rather than raw data, shared for aggregation (Guerra et al., 2023). This paradigm is particularly attractive in the energy domain, where stakeholders often face legal, ethical, or logistical constraints in sharing data (Cheng et al., 2022). Applications of FL have been explored in areas such as distributed load forecasting, fault detection in microgrids, predictive maintenance of infrastructure and collaborative control of distributed energy resources (Michalakopoulos et al., 2024; Grataloup et al., 2024). While FL has theoretical promise, its implementation in energy contexts remains experimental. Most existing FL studies rely on simulated environments or homogeneous data sources that fail to reflect the operational diversity and noise typical of real-world energy systems. Moreover, technical challenges such as model drift, communication latency, data heterogeneity and system resilience under failure conditions are rarely addressed in practical deployments. Furthermore, FL solutions are seldom coupled with real-time control systems, physical devices, or domain-specific regulatory frameworks, limiting their applicability in critical infrastructure settings. EnerTEF incorporates federated learning as a core principle of its architecture. Through its orchestration layer, it enables privacy-preserving, compliant model training and testing across physically distributed nodes. Importantly, EnerTEF supports real-time federated workflows that can be integrated with both digital twins and physical testing infrastructure, offering an active environment for FL development, validation and benchmarking.

2.3. Regulatory and ethical governance of AI

The ethical and legal implications of AI in critical infrastructure domains are under increasing scrutiny, particularly with the impending adoption of the European Union's AI Act (Jørgensen et al., 2025). This regulatory framework introduces mandatory requirements for high-risk AI systems, including transparency, human oversight, robustness and accountability, as well as provisions for post-market monitoring and dynamic risk management (Machlev et al., 2022; Panagoulias et al., 2023). In the context of energy systems, which are classified as high-risk due to their societal and economic impact, these provisions are foundational (Srivastava et al., 2023).

Despite this, few existing AI development environments or testbeds incorporate mechanisms for regulatory testing, ethical review, or compliance auditing. Existing efforts are fragmented and largely limited to guidelines, checklists, or theoretical analyses. The concept of "regulatory sandboxes", while promising, has not yet been operationalized at scale for AI services in the energy sector. Moreover, current testing environments rarely engage regulators, standardization bodies, or civil society actors in the co-design or validation of AI systems. EnerTEF

Table 1
Related work review findings, indicating the gaps in prior related work and EnerTEF's contributions.

Area	Observed gap in prior work	Advantages of the EnerTEF framework
Data exchange (data space compliant)	Ad hoc or centralized sharing; weak provenance and usage control	IDSA-compliant connector; local data residency; usage-control contracts; auditable, policy-governed exchange
Real-time access via digital twins	Simulation-only twins; weak or no coupling to live assets; no timing discipline	Twins aligned to a common time base with latency budgets; hardware-/software-in-the-loop; twins as active test instruments alongside physical assets
Orchestration across HPC resources	Single-site tools; limited multi-resource scheduling; poor bandwidth awareness	HPC orchestrator coordinating edge/on-prem/cloud resources (CPU/GPU); bandwidth-aware scheduling; quotas and preemptible jobs; checkpointing
AI experiment workbench for federated learning at the edge	Centralized training assumptions; limited handling of non-IID data; no edge participation	Federated learning on edge/on-prem; asynchronous aggregation; partial participation; update compression; non-IID aware training with client/aggregate diagnostics
Regulatory sandboxes for data-owner rules (e.g., anonymity)	Guidance rather than operational enforcement; lack of verifiable policy application	Embedded regulatory sandbox; enforceable policies (masking/anonymization); optional differential privacy; automated evidence packs and compliance reports

addresses this shortcoming by integrating regulatory sandboxes into its node architecture and by embedding compliance metrics into the AI testing lifecycle. Each node supports scenario-based testing under different regulatory assumptions and provides interfaces for monitoring, audit and user feedback. Through this approach, EnerTEF not only supports technical validation but also fosters regulatory innovation, helping bridge the gap between AI capability and societal trust.

2.4. Related work review findings and EnerTEF advantages

In reviewing the existing body of literature across Section 2, existing platforms show value but leave some gaps across five different dimensions that EnerTEF addresses. First, data exchange is often ad hoc or centralized, whereas EnerTEF relies on data-space compliant mechanisms that provide provenance, usage control, and auditable exchange. Second, many platforms lack real-time access to physical assets through well-synchronized digital twins. EnerTEF couples twins to live telemetry with disciplined timing so they function as active test instruments. Third, orchestration across heterogeneous high-performance computing resources is limited. EnerTEF introduces an HPC orchestrator that coordinates edge, on-premise, and cloud resources with bandwidth-aware scheduling and checkpointing. Fourth, edge-capable federated learning is rarely supported. EnerTEF's AI experiment workbench enables federated learning at the edge with asynchronous aggregation, partial participation, and non-IID aware training. Finally, data-owner rules are often aspirational rather than enforceable; EnerTEF embeds regulatory sandboxes so owners can apply and prove policies such as anonymization and masking. The differentiating points among existing platforms and EnerTEF are summarized in Table 1

3. Methodological framework

The proposed methodological framework is motivated by the lack of an integrated, compliance-aligned setting in which AI for energy can be evaluated under real operating constraints, with lawful data use and verifiable governance. EnerTEF addresses this need by providing a pan-European, federated infrastructure that orchestrates and validates AI-based energy services in real, multi-stakeholder contexts. At its core, the platform enables federated learning pipelines through an AI experimentation workbench and provides real-time access to physical assets via well-synchronized digital twins, so that model behavior is examined against the dynamics of actual systems. Lawful and auditable data exchange is ensured by an International Data Spaces Association compliant data space that removes any requirement to centralize raw data. To support realistic scale, access to high-performance computing is orchestrated across edge, on-premise, and cloud resources with. The resulting training and evaluation pipelines account for non-identically distributed data across nodes, communication and coordination limits in cross-node learning, and strict timing requirements for twin-to-asset alignment, while a unified security and access-control layer and embedded regulatory sandbox enable policy-in-the-loop verification. This section outlines the architectural and procedural logic of the platform, including the structure of its testing environments and their federation.

3.1. Node structure

EnerTEF is structured as a network of testing facilities composed of several nodes, each potentially supported by smaller satellite facilities. A node serves as the primary unit of the infrastructure. It provides a coherent set of services and houses the necessary physical, digital and organizational assets for AI experimentation. A node can be operated by one or more legal entities working in close coordination and is responsible for delivering end-to-end testing capabilities in one or more energy domains. In cases where a single node cannot reach the necessary geographical or functional scale, satellites are introduced as smaller but fully operational testing environments. These satellites extend the capabilities of their associated nodes by providing complementary infrastructure and services, either through physical assets or remote access.

At the core of EnerTEF lies a distributed infrastructure composed of geographically dispersed, functionally specialized nodes. Each node is designed as a full-stack, interoperable environment that combines real physical assets (such as smart meters, building energy systems, EV charging infrastructure, battery energy storage systems and microgrids) with high-fidelity digital twins and software interfaces. These assets are representative of specific parts of the energy value chain and are operated by partners with domain-specific expertise. The structure of each node is presented in Fig. 1.

Each EnerTEF node is designed as a modular, multi-layered infrastructure that integrates physical experimentation, digital modeling and governance oversight into a single operational framework. At its foundation lies the Local Digital Platform, which serves as the central interface through which AI-based solutions interact with test environments. This platform enables developers to securely connect through client-side APIs, allowing their services to access relevant infrastructure components, datasets and decision support tools. Directly linked to the digital platform is the Local Data Space, a structured data management environment that consolidates operational, simulated and experimental information. The Local Data Space acts as a semantic and syntactic integration layer, enabling interoperability across heterogeneous sources and ensuring the traceability of data flows used for AI model training, validation and auditing. To preserve temporal coherence between digital and physical layers, nodes discipline timestamps to a common time base using Precision Time Protocol or Network Time Protocol, record sensor and actuation latencies, and expose these timings to experiments so that control and learning loops can be aligned with plant dynamics.

Surrounding these digital layers is a suite of physical and virtual assets that make each node functionally complete. This includes Real-Life Facilities, which provide access to live operational infrastructure

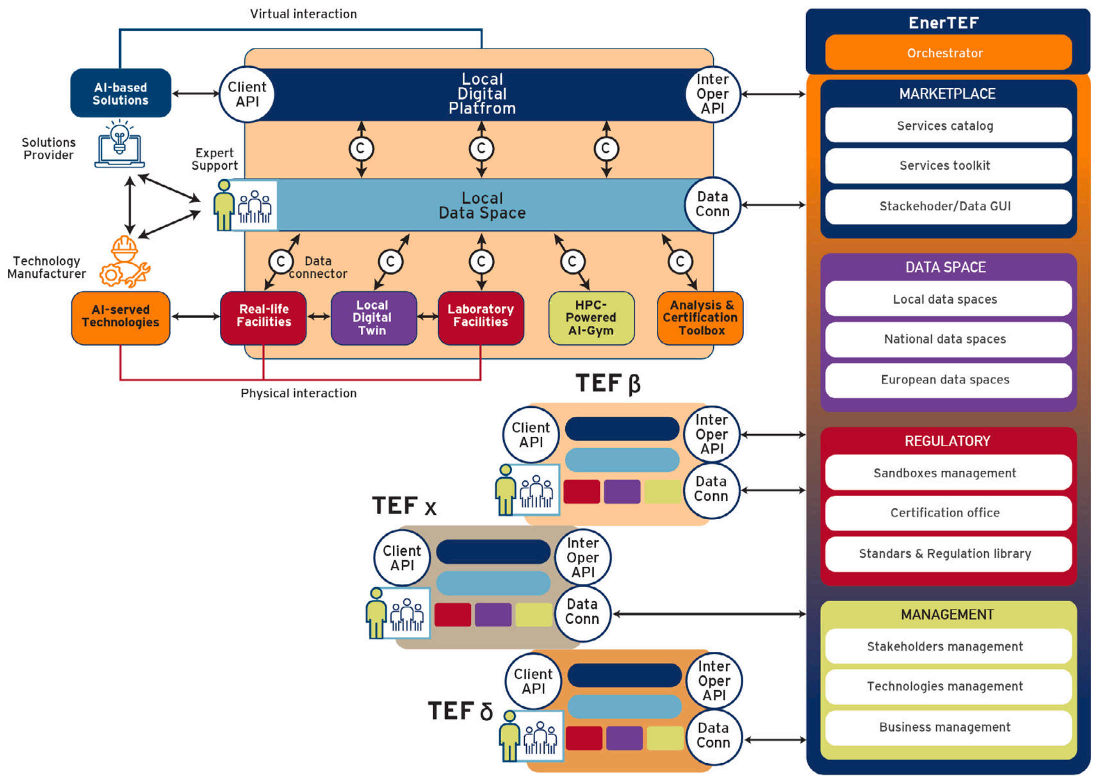


Fig. 1. The structure of each Testing & Experimentation node.

(e.g., grid-connected equipment, renewable installations, or building management systems); Local Digital Twins, which enable simulation-driven experimentation by replicating physical systems under varying boundary conditions; and Laboratory Facilities, which host high-resolution measurement, monitoring and testing equipment. Complementing these assets is an HPC-Powered AI Gym, offering compute-intensive environments for algorithmic training, particularly in reinforcement learning and federated AI scenarios. The Analysis and Certification Toolbox embedded in the node structure supports structured model evaluation, risk assessment and conformance testing, thereby operationalizing transparency, robustness and compliance as measurable properties. This system enables developers to not only validate performance but also to engage with institutional stakeholders around legal, ethical and operational requirements through the integration of sandboxing protocols and expert review panels. Twin update policies and error tolerances are encoded as experiment metadata: update cadence is tied to the relevant control period or state-estimation cycle, acceptable tracking error bands are specified per signal, and automatic fallbacks are triggered when error or latency budgets are exceeded.

Each node operates not in isolation, but as part of a federated architecture comprising multiple geographically and thematically differentiated instances denoted in the figure as TEF β, TEF x and TEF δ. These satellite nodes follow a shared reference architecture but can host domain-specific assets, stakeholder communities and regulatory scenarios tailored to local contexts. The interconnection between nodes is supported by Interoperability APIs and Data Connectors, which allow distributed AI experimentation, model portability and coordinated orchestration of test campaigns. At the federation level, the EnerTEF Orchestrator coordinates multi-node interactions, manages service discoverability through an integrated Marketplace and enforces access policies via standardized API gateways. To support experiments that span multiple nodes, the orchestrator exposes bandwidth-aware scheduling and message batching, and allows partial participation so that straggling nodes do not stall global progress. Finally, the orchestrator integrates with a tiered Data Space architecture, spanning local, national and European levels, and thus ensuring that experimentation can leverage context-relevant data while adhering to jurisdictional constraints.

3.2. EnerTEF federated architecture and functional stack

The EnerTEF architecture represents a federated, vertically integrated platform specifically designed to support the development, validation and regulatory alignment of AI solutions in the energy sector (Fig. 2). At its base are a series of geographically and thematically distributed TEF nodes, each linked to specific domains such as energy distribution, building energy management, hydropower forecasting, or grid stability services. These nodes host a combination of real-world infrastructures (e.g., municipal buildings, substations, microgrids), system-level hardware-in-the-loop (HIL) platforms, local digital twins that simulate operational dynamics, high-performance AI training environments and laboratory facilities for precision testing. The architecture supports experimentation that spans from edge-device control algorithms to cloud-deployed predictive analytics. By enabling AI developers to interact with both live assets and virtualized replicas, EnerTEF ensures that solutions are tested under the full complexity of environmental variability, user behavior, control interfaces and data quality conditions. This hybrid physical-digital capability is a distinctive strength of the platform, as it allows transition from purely algorithmic validation to system-level robustness assessment, which is a critical gap in conventional AI benchmarking. Federated training is configured to operate under non-identically distributed data: client sampling is stratified, updates are reweighted by sample quality and class imbalance, and proximal or momentum-based variants are used to stabilize convergence; convergence is monitored with drift and fairness diagnostics at both client and aggregate levels.

Above the infrastructural substrate lies the data and semantic interoperability framework, which underpins the federation's ability to operate as a unified system despite its geographical and institutional heterogeneity. This layer includes mechanisms for stream and batch data

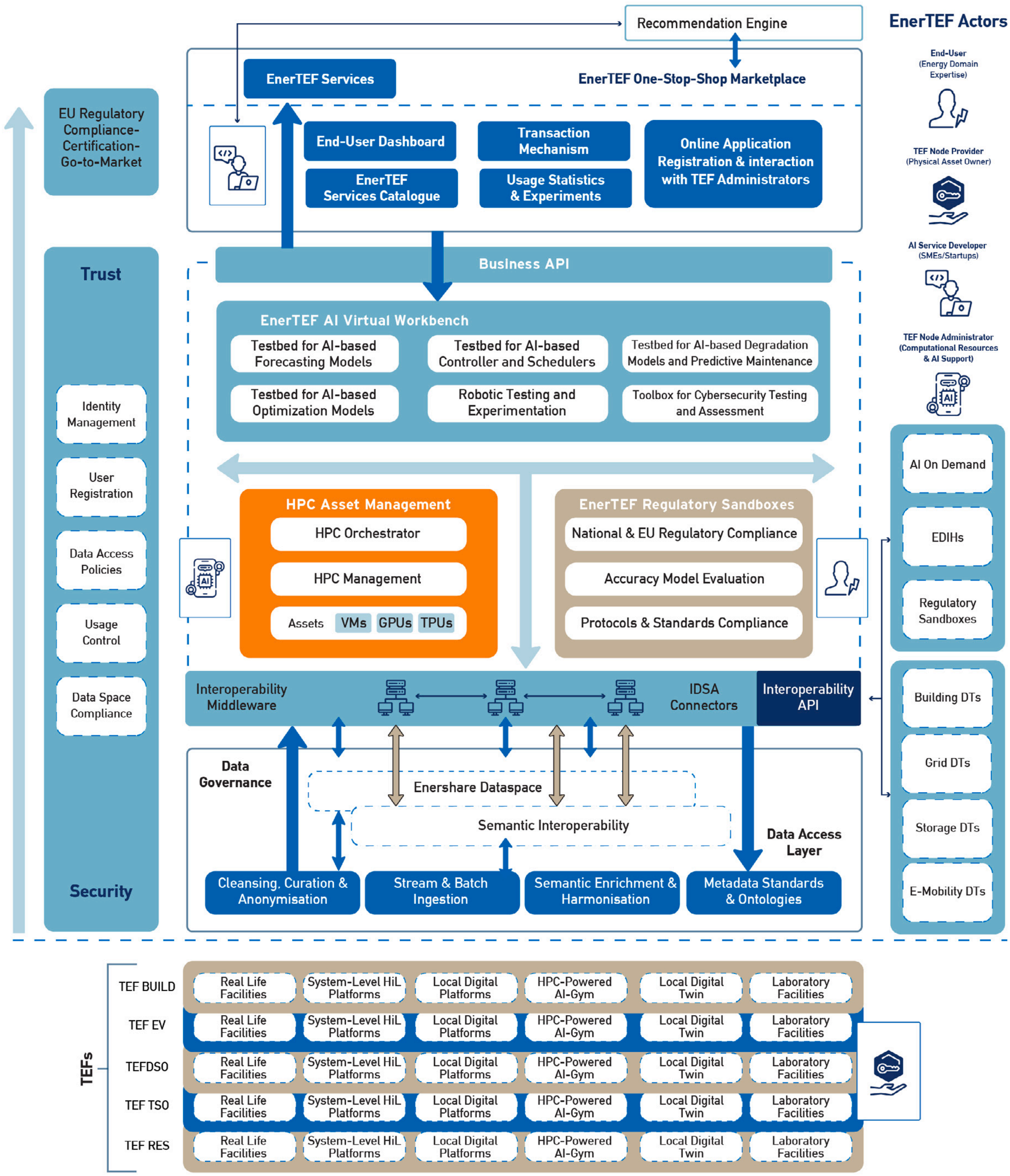


Fig. 2. The EnerTEF portal architecture.

ingestion, syntactic normalization, semantic enrichment, anonymization and metadata harmonization. EnerTEF employs a shared dataspace infrastructure referred to as the Enershare Dataspace, which ensures compliance with European data governance regulations, such as General Data Protection Regulation (GDPR) and the Data Governance Act, while maintaining usability for experimentation. Data exchange between nodes is facilitated through IDSA-compliant connectors and standardized Interoperability APIs, enabling AI developers to access datasets from different domains, regions and infrastructures without breaching localization or confidentiality agreements. To reduce communication overhead during cross-node learning, model updates are compressed using quantization or sparsification when appropriate, and aggregation cadence is adapted to network conditions; control-plane messages are separated from payloads to avoid head-of-line blocking. This foundation ensures that AI models can be trained and tested on diverse, representative data, while also supporting transfer learning, federated learning and cross-context benchmarking. Importantly, the semantic layer is not just a technical feature but a regulatory enabler: it creates the conditions for lawful experimentation, supports explainability by preserving data lineage and aligns with EU strategies for creating sector-specific data spaces and trustworthy digital infrastructures.

Security and cybersecurity are enforced by the EnerTEF Identity and Access Management (IAM) layer and by data spaces principles. Identities are issued through the IAM with role-based permissions (e.g., developer, node administrator, end user, marketplace administrator) and least-privilege policies, with all actions logged for audit. Data remain at source in the Local Data Space and are not centrally stored in the cloud; exchange occurs only through the EnerTEF IDSA-compliant connector, which negotiates usage control, enforces access contracts, and mediates transfers. Connector-to-connector links use TLS 1.3 with mutual authentication; only metadata, model parameters, or approved aggregates are shared, according to node policies. Keys are managed locally, and compromise procedures are defined: isolate the affected node, revoke credentials, rotate keys, and restore from verified backups. This arrangement combines fine-grained access control with local data residency and auditable, policy-governed exchange.

While instantiated in energy, the architecture is domain-agnostic and transferable to other critical-infrastructure sectors (e.g., water, mobility, health). The architectural core is further reinforced by a set of interlinked virtual services and governance tools that turn the EnerTEF infrastructure into a full-stack platform for AI lifecycle management. Central to this layer is the AI Virtual Workbench, a suite of modular testbeds designed to support experimentation with AI-based forecasting, optimization, control, degradation prediction and cybersecurity under operationally realistic and stress-tested scenarios. Each testbed is powered by a high-performance computing (HPC) backend that manages computational elasticity across various processing units (e.g., VMs, GPUs, TPUs), coordinated by the EnerTEF orchestrator. These testbeds allow AI developers to replicate realistic control loops, stress-test models against dynamic inputs, compare algorithmic strategies and visualize intervention effects. Surrounding the workbench is a regulatory sandbox environment, enabling AI services to be tested under simulated policy conditions. The sandbox includes tools for post-hoc explainability, bias detection, human oversight modeling and conformance with the EU AI Act's high-risk AI requirements. On top of this sits the EnerTEF Marketplace and Services Interface, where stakeholders can discover testing assets, register use cases, access dashboards, manage experiments and interact with TEF administrators. A built-in recommendation engine provides intelligent matchmaking between AI solution profiles and available infrastructures or datasets, improving developer guidance and infrastructure utilization. Overall, this architecture transforms EnerTEF from a loose federation of facilities into a cohesive, institutional-grade platform that integrates experimentation, compliance and deployment readiness in a single digital continuum. Its layered design ensures alignment with European digital sovereignty, energy resilience and trustworthy AI goals, while remaining flexible enough to adapt to new technological and regulatory developments.

3.3. AI experimentation workbench

The AI experimentation workbench is a dynamic environment enabling TEF users to conduct advanced AI experiments across different problem categories. It provides a dynamic environment in which users can design, train, evaluate, and deploy models across a set of canonical problem categories (forecasting, optimization/control, anomaly detection, degradation prediction, reinforcement learning, and privacy-preserving/federated learning). It is built to operate at multiple data scales, from exploratory notebooks to production-grade pipelines, while keeping experiments efficient and reproducible.

The AI experimentation workbench consists out of a Kubernetes cluster. Kubernetes is an open-source orchestrator platform that automates the deployment, scaling and management of containerized applications. Containers which hold an application are lightweight and portable units which contain all needed dependencies. Services in Kubernetes define a set of pods and provide a stable endpoint (IP address) for access. The applications are either deployed with Helm charts or Docker images directly. Helm is a package manager for Kubernetes that simplifies the deployment of applications by providing a way to define, install, and manage Kubernetes applications using charts. Helm charts allow you to use templates for Kubernetes manifests (YAML files). The architecture of the AI experimentation workbench is demonstrated in Fig. 3 while its main components are summarized in Table 2.

The experimental process begins with a collaborative scenario definition phase involving AI developers, infrastructure operators, domain experts and regulators. Each experiment is grounded in a real-world use case, such as reinforcement learning for dynamic load control, federated forecasting of renewable generation, or anomaly detection in low-voltage networks. Based on these use cases, experimental scenarios are constructed to reflect both technical parameters (e.g., boundary conditions, operational constraints, failure modes) and contextual variables (e.g., regulatory obligations, user behavior, environmental targets). Models are then deployed either to real physical assets or to digital twins that reproduce asset dynamics with high fidelity. The orchestration platform manages the training modality in a local, centralized, or federated manner depending on data privacy, computational distribution and policy constraints. To guard against bias and instability with non-identically distributed data, client-level metrics are tracked alongside global metrics, gradient clipping and adaptive learning rates are employed when dispersion is high, and reweighting schemes limit dominance by large or atypical clients.

Execution within EnerTEF is characterized by rigorous monitoring, traceability and multi-perspective evaluation. Throughout the experiment lifecycle, all system interactions, control outputs, decision logic and performance logs are captured in standardized formats, enabling real-time observability and post-hoc auditability. Experiments may run synchronously in real environments or asynchronously in accelerated simulations. To further support compliance-aligned experimentation, EnerTEF incorporates a distributed regulatory sandbox framework that mirrors evolving European policy environments, particularly the EU AI Act. Each node can instantiate synthetic regulatory conditions to evaluate whether an AI system satisfies transparency, human oversight, traceability and risk mitigation obligations. These sandboxes include mechanisms for compliance report generation, automated documentation auditing and constraint enforcement, such as disabling autonomous control in high-uncertainty regimes. This regulatory overlay transforms EnerTEF into a pre-certification environment for AI deployment in critical energy contexts, allowing developers to anticipate and mitigate legal barriers prior to commercialization. Update frequency and error tolerances are specified per experiment as policies: for example, twin state-refresh periods are set as fractions of the control-loop period, prediction error bands trigger safe fallback or retraining, and communication intervals adjust to maintain end-to-end latency within declared budgets.

Finally, the platform supports systematic benchmarking across multiple axes of performance. Evaluation metrics span four categories: technical (e.g., accuracy, response latency, throughput), robustness (e.g., model drift, failure resilience), governance (e.g., explainability, fairness, auditability) and environmental (e.g., computational efficiency, carbon footprint). All indicators are collected using unified schemas and stored in centralized data repositories equipped with semantic indexing, enabling comparative analytics across nodes and domains. This evidence base not only informs individual AI service refinement but also contributes to the broader scientific understanding of trustworthy AI integration within complex energy systems. Benchmark suites include client-level as well as aggregate metrics to detect heterogeneous performance, and results are stored with full lineage so that conclusions are reproducible across nodes and releases.

4. Application scenario and use cases

4.1. Application scenario

The EnerTEF platform is designed to operationalize a federated, secure and policy-aligned digital experimentation ecosystem for the development and validation of AI services in the energy sector. At the core of this scenario lies the EnerTEF Portal, a Platform-as-a-Service (PaaS) environment that orchestrates the interaction between AI Service Developers, End-Users from the energy domain and Node

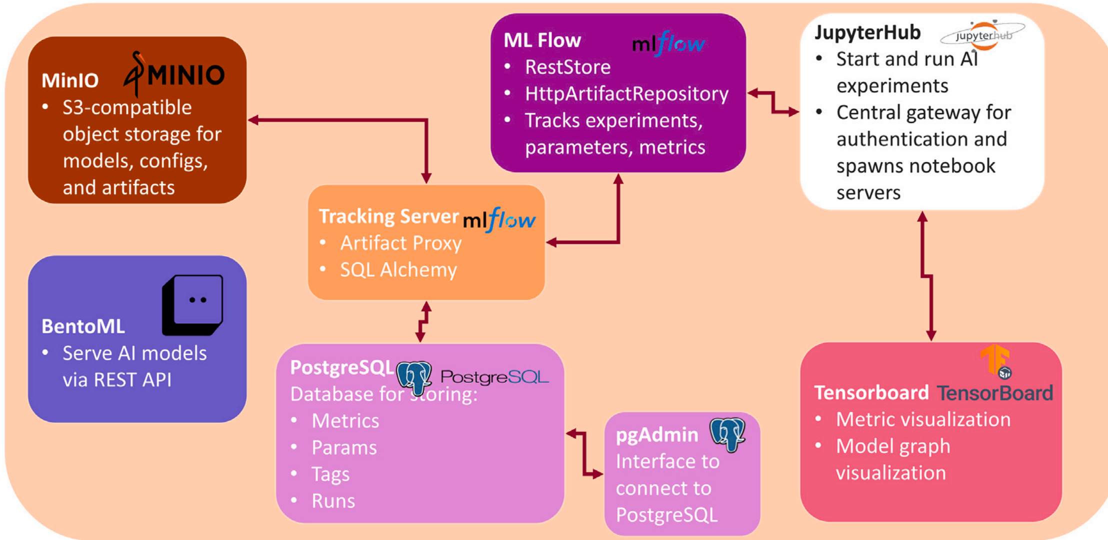


Fig. 3. The EnerTEF AI experimentation workbench.

Table 2
Core applications hosted in the AI Experimentation Workbench and their roles.

Component	Role (summary)	Primary use in workbench	License
MinIO	High-performance object storage for cloud-native/big-data workloads	Store models, configs, datasets, and run artifacts	GNU AGPL v3
MLflow	ML lifecycle tracking with web UI	Track parameters, metrics, artifacts; manage experiment runs	Apache 2.0
PostgreSQL	Robust relational database	Persist MLflow metadata (metrics, params, tags, runs)	PostgreSQL
pgAdmin	Admin tool for PostgreSQL	Inspect/manage databases; query and maintain MLflow data	PostgreSQL
JupyterHub	Multi-user notebook gateway	Spawn isolated user notebook servers with chosen environments	BSD 3-Clause
Jupyter Notebook	Interactive code and narrative environment	Develop/execute experiments; log with TensorBoard; tune with Optuna	BSD 3-Clause
TensorBoard	Training visualization toolkit	Visualize logs and metrics from experiments within notebooks	Apache 2.0

Administrators responsible for managing the underlying physical and digital infrastructures. This portal facilitates an end-to-end experimentation lifecycle by linking distributed datasets, HPC resources, physical assets and digital twins, all underpinned by strong data governance and interoperability standards. Through this architecture, EnerTEF offers a scalable and replicable model for conducting compliant, real-world AI experiments across heterogeneous infrastructure environments. The complete UML diagram is depicted in Fig. 4

Within this application scenario, AI developers are able to access a curated Services Catalogue, submit proposals in response to pre-defined challenges and configure experiments using the EnerTEF AI Experiment Workbench. Experiments are deployed through an integrated orchestration layer that automates resource provisioning and enforces usage agreements, ensuring GDPR compliance and semantic interoperability. End-Users (such as TSOs, DSOs and RES aggregators) can define service needs, monitor experiments via dedicated dashboards and validate solutions based on pre-defined key performance indicators (KPIs). Node Administrators ensure the availability, security and performance of federated assets while adhering to shared metadata, data access and reporting protocols. The Marketplace Transaction Management Module manages financial exchanges, licensing and service commercialization. The scenario culminates in the certification of AI services, enabling their listing within the Portal's "Validated Services" catalogue for wider adoption. This structured, modular and policy-aware scenario illustrates EnerTEF's capacity to bridge technological innovation with operational relevance and regulatory readiness.

To evaluate the breadth and scalability of the EnerTEF architecture, three representative use cases were implemented across its federated testbeds. These scenarios were selected to interrogate distinct facets of the infrastructure: predictive modeling under natural resource constraints (TEF RES), regulatory co-optimization of network investments (TEF DSO) and real-time control of energy consumption to exploit PV-based energy production in complex urban environments (TEF BUILD).

4.2. Hydropower forecasting with time series AI (TEF RES)

The TEF RES node hosts an AI forecasting service that generates short-term production estimates for a hydropower facility operated by the Public Power Corporation (PPC) in Northern Greece. This application addresses the operational challenge of precise dispatch planning and market participation in the context of water-constrained renewable generation. Leveraging EnerTEF's standardized ingestion APIs, developers access historical operational data and weather forecasts to feed a time series modeling framework. Model development and hyperparameter optimization take place entirely within EnerTEF's secure experimentation environment, which provides integrated benchmarking against multiple statistical and machine learning baselines.

More specifically, we study the Asomata hydropower plant (108 MW, Western Macedonia, Greece), a reservoir-based facility serving multiple objectives (generation, upstream pump-back support, irrigation, downstream regulation). Reservoir management buffers short-term meteorological impacts, weakening direct weather–generation

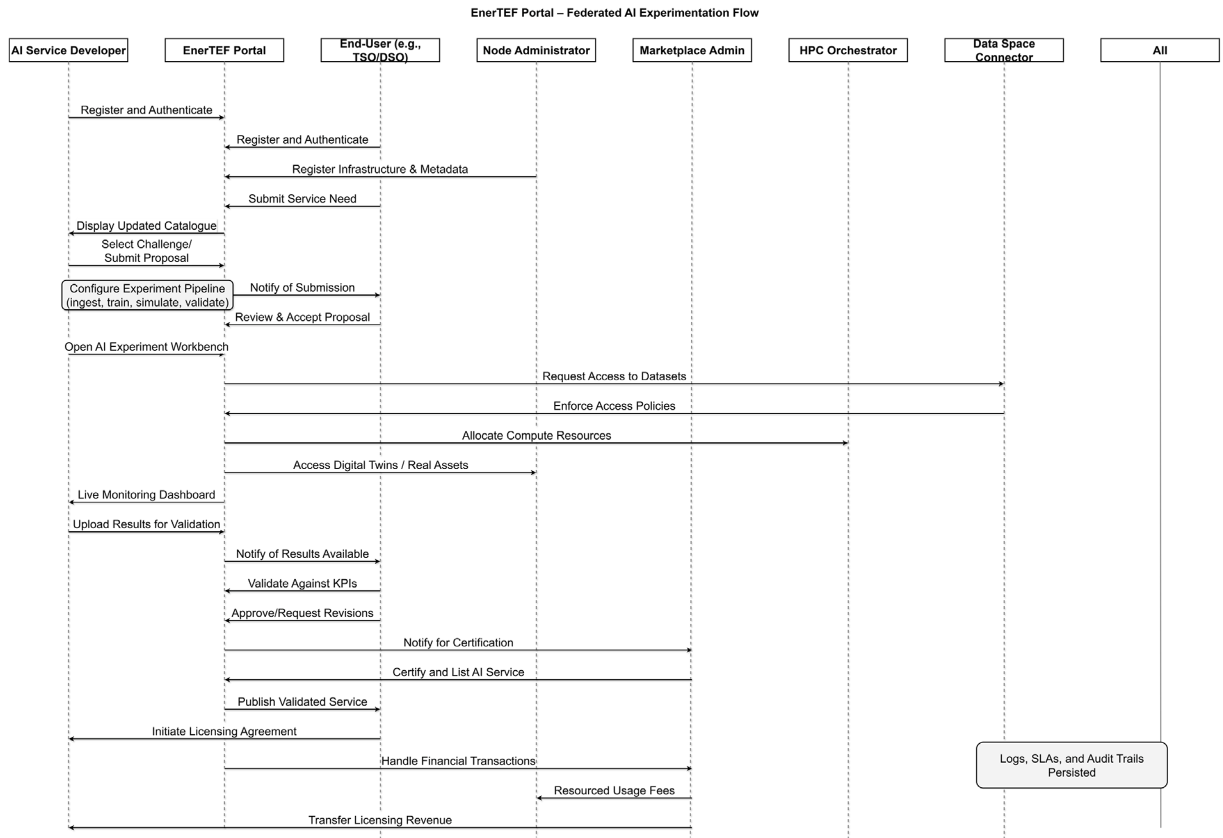


Fig. 4. UML diagram that describes the application scenario of the EnerTEF portal.

correlations. The dataset used in this study consists of daily-resolution time series covering hydropower generation and meteorological data from January 2014 to December 2023. The key variables are as follows:

- **Date**: Timestamp for each observation.
- **Total HPP**: Daily energy output of the Asomata HPP, measured in megawatt-hours (MWh).
- **temperature_2m_mean**: Mean air temperature at 2 m above ground level, in degrees Celsius.
- **precipitation_sum**: Total daily precipitation (rain and snow), in millimeters.
- **rain_sum**: Total daily rainfall, in millimeters.

Forecasts are produced with Prophet and compared to naïve and weekly seasonal-naïve baselines (Fragkiadaki et al., 2025). Mathematically, Prophet expresses the time series $y(t)$ as:

$$y(t) = g(t) + s(t) + h(t) + X\beta + \epsilon(t) \tag{1}$$

where $g(t)$ represents the non-periodic trend component, modeled as a piecewise linear or logistic growth curve with automatic change-point detection to capture structural breaks or shifts in the underlying process, $s(t)$ models periodic seasonal effects, such as daily, weekly, or yearly cycles, through Fourier series expansions that can flexibly approximate complex seasonal patterns, $h(t)$ accounts for the effects of known events or holidays; although not used in this study, this component can adjust forecasts around specific dates when deviations from normal behavior are expected, $X\beta$ corresponds to the influence of user-defined external regressors, allowing the integration of additional explanatory variables believed to impact the target variable and $\epsilon(t)$ is the error term, capturing noise and unexplained variability.

To assess models accuracy, three commonly used evaluation metrics were applied to the validation set:

Table 3
Model performance metrics on validation set.

Model	RMSE (MWh)	SMAPE (%)	R2 Score
Prophet (With regressors)	252.3356	53.30	0.1905
Prophet (Without regressors)	244.7474	45.23	0.2385
Naive (Persistence)	150.3782	83.16	−1.9843
Seasonal Naive (Weekly)	244.841	52.31	0.2379

Root Mean Squared Error (RMSE) — penalizes larger errors more heavily:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{2}$$

Symmetric Mean Absolute Percentage Error (SMAPE) — a scale-independent, symmetric variant:

$$\mathrm{SMAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \tag{3}$$

R-squared (R^2) Score — proportion of variance explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{4}$$

where $\bar{y}$ is the mean of the observed values.

Table 3 summarizes the performance of the Prophet models and the baseline models across the multiple evaluation metrics on the validation set.

Fig. 5 presents a comparison between actual hydropower generation and several forecasting methods over the validation period. The chart contrasts persistence and seasonal naïve models with Prophet-based forecasts, both with and without weather regressors. The results

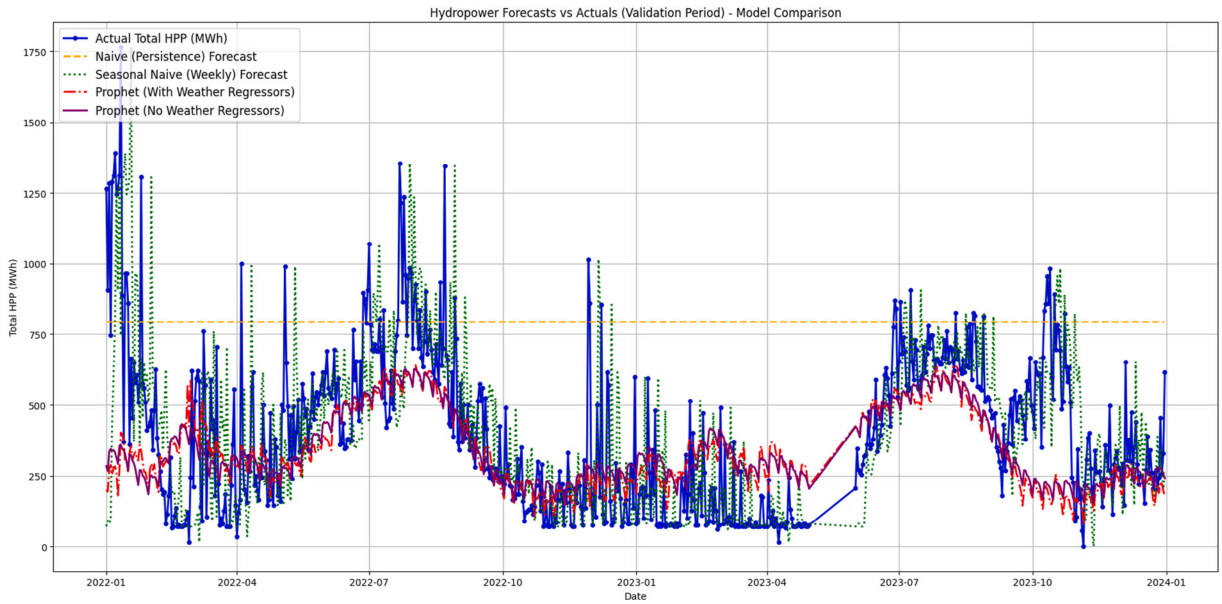


Fig. 5. Comparison of hydropower forecasts and actual generation for the PPC facility in Northern Greece over the validation period. The plot includes persistence, seasonal naïve, and Prophet models with and without weather regressors, demonstrating the accuracy gains achieved by incorporating meteorological data into AI-based forecasts.

illustrate the value of integrating weather data into AI models, with weather-aware Prophet forecasts tracking seasonal fluctuations and short-term variability more closely than the simpler baselines. This visual evidence supports the role of EnerTEF in enabling data-rich, comparative evaluation of forecasting strategies under operationally realistic conditions.

4.3. DER-DSO co-optimization for grid planning (TEF DSO)

The TEF DSO node enables AI-supported coordination between distribution system operators and third-party service providers using a digital-twin workflow built around DPsim (Mirz et al., 2019) and VILLASframework (Bach and Monti, 2025). The Wunsiedel–Schönbrunn distribution grid is imported into DPsim to form the backbone of the twin. Electrical measurements from multiple substations are mapped to the virtual grid to support services such as state estimation, short-term demand/generation forecasting and what-if studies for flexibility activation. EnerTEF's orchestration provides authenticated access, versioned scenarios and audit-ready logs, while hybrid execution (hardware-in-the-loop and co-simulation) allows tests at different fidelities.

Two local platforms operationalize the twin. The DT *builder* provides a JupyterHub-based environment exposing DPsim's Python API, where users construct scenarios from Common Information Model (CIM) or MATPOWER descriptions and load/generation profiles. Input artefacts can be pulled through the dataspace connector, and the output is a Docker image that encapsulates the twin. The DT *manager*, built on VILLAframework, deploys the twin image and offers a web frontend with configurable dashboards (plots, gauges, topology images) and run-time controls (buttons, sliders, numeric inputs) to trigger contingencies or parameter changes. The VILLASweb backend bridges the twin to external components via VILLASnode protocols, enabling connections to databases, physical assets and AI modules and supporting distributed co-simulation and HIL.

Moreover, TEF-DSO integrates a part of the SOGNO platform, namely pyVolt (Ghoreishi et al., 2025), a microservice which provides a weighted least squares (WLS) calculation for state estimation. pyVolt uses CIMpy to read network data based on the Common Information Model (CIM IEC61970). State estimation is formulated in the standard WLS form with measurement vector z, state x, measurement function $h(\cdot)$, and covariance R:

$$\hat{x} \;=\; \arg\min_{x} \left(z - h(x)\right)^{\top} R^{-1} \left(z - h(x)\right), \tag{5}$$

whose normal-equation solution is

$$\hat{x} \;=\; \left(H^{\top} R^{-1} H\right)^{-1} H^{\top} R^{-1} z, \tag{6}$$

where $H = \partial h/\partial x$ is the Jacobian evaluated at the current iterate (or at $\hat{x}$ for the linear case). Residuals $r = z - h(\hat{x})$ and their normalized forms support bad-data detection. Topology or parameter changes can be injected at runtime via the DT manager to test estimator resilience. For load/solar forecasting on the same feeders, models are trained within the workbench and fed to the twin to close the planning–operation loop. Accuracy is reported using MAE and RMSE on voltages and flows for state estimation and on power series for forecasting, with stress tests that include fault injections, large load ramps, and inverter-rich scenarios. Extreme cases are exercised by scripted events in the manager (e.g., line outages, islanding commands), while the VILLASframework ensures deterministic message timing and protocol interoperability so that end-to-end latency budgets can be verified.

Fig. 6 shows the DT manager interface. This combination of DPsim-based modeling, VILLAS-mediated I/O and EnerTEF orchestration yields a reproducible pipeline in which AI services are deployed against a controllable yet realistic grid twin, evaluated under nominal and extreme conditions and packaged as versioned artefacts for cross-site reuse.

In addition, a live view of simulators, controllers and gateways, their status and uptime and active scenarios is provided. Through its structured infrastructure templates, scenario definitions, and co-creative development process, TEF DSO demonstrates how digital twins and AI microservices can converge within EnerTEF to produce scalable, explainable, and operationally relevant grid solutions. The combination of live infrastructure management (as depicted in Fig. 6) and advanced simulation capabilities supports DSOs in improving observability in low-voltage grids, enhancing network state awareness, and enabling proactive operational planning.

4.4. PV self-consumption optimization for municipal buildings (TEF BUILD)

At TEF BUILD, the Serafeio municipal complex in Athens is used to demonstrate AI-assisted control that maximizes on-site use of rooftop

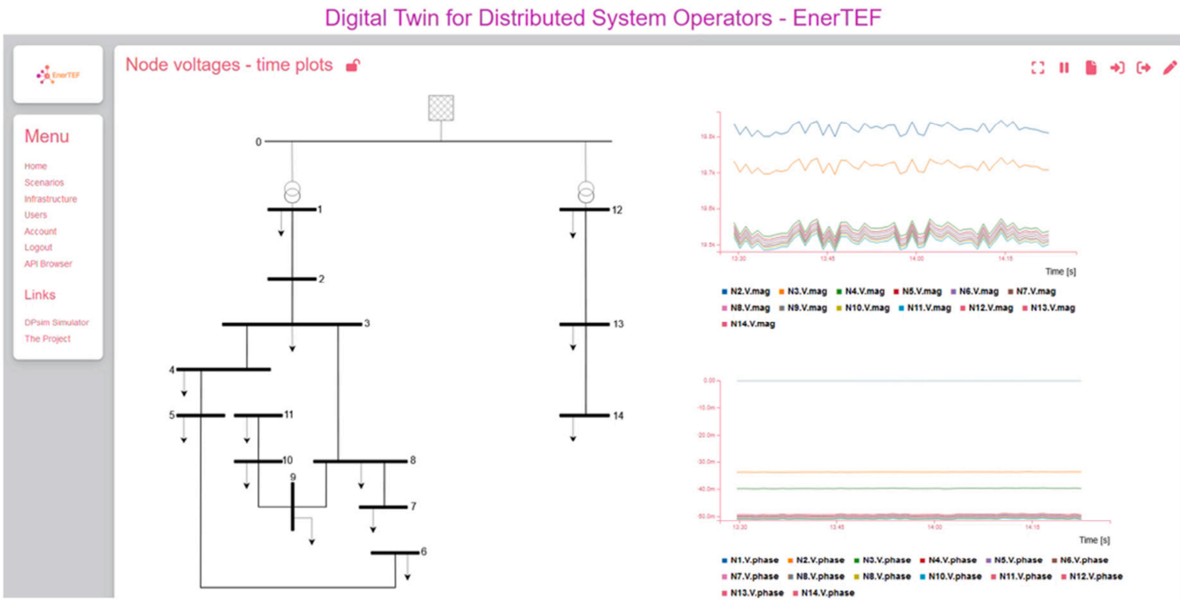


Fig. 6. TEF DSO digital twin infrastructure interface. The dashboard provides real-time visibility of infrastructure controllers, simulators, and gateways, enabling operators and developers to orchestrate distributed simulations, manage communications, and monitor system health for AI-enabled grid planning and operation.

photovoltaics (PV) while preserving comfort. Real-time data streams from the city platform provide electric demand at 15–60 min resolution, PV generation, and HVAC telemetry; a digital twin encapsulates a first-order thermal model of key zones. EnerTEF orchestrates training and deployment so that the same policy is validated in the twin and then executed on live assets with audit logs and operator override.

Control is posed as a finite-horizon predictive optimization. Let $P_{\mathrm{pv}}(t)$ be PV output, $P_{\mathrm{load}}(t)$ the intrinsic (uncontrolled) demand, and u_t the HVAC power set by the controller (positive when drawing power). Grid import/export are $P_{\mathrm{imp}}(t) \geq 0$ and $P_{\mathrm{exp}}(t) \geq 0$ with power balance

$$P_{\mathrm{load}}(t) + u_t \;=\; P_{\mathrm{pv}}(t) + P_{\mathrm{imp}}(t) - P_{\mathrm{exp}}(t). \tag{7}$$

The indoor temperature T_z evolves as

$$T_z(t{+}1) = a\,T_z(t) + b\,u_t + e\,T_{\mathrm{out}}(t). \tag{8}$$

Over horizon H, the controller minimizes energy cost while promoting self-consumption and penalizing comfort violations and excessive switching:

$$\min_{\{u_k, P_{\mathrm{imp},k}, P_{\mathrm{exp},k}\}} \sum_{k=t}^{t+H-1} \Big(c_e(k)\,P_{\mathrm{imp}}(k) + \lambda_{\mathrm{exp}}\,P_{\mathrm{exp}}(k) + \lambda_c\,|T_z(k) - T_{\mathrm{set}}(k)|_+ + \lambda_s\,\|\Delta u_k\|_1 \Big), \tag{9}$$

$$\text{s.t.}\quad T_{\min} \leq T_z(k) \leq T_{\max}, \qquad 0 \leq u_k \leq u_{\max}, \quad P_{\mathrm{imp}}(k), P_{\mathrm{exp}}(k) \geq 0, \tag{10}$$

together with the power balance and thermal dynamics. The λ_{exp} term discourages exporting PV when flexible demand can be shifted; $|x|_+ = \max\{x, 0\}$ counts only comfort violations outside $[T_{\min}, T_{\max}]$. The same formulation can be realized as reinforcement learning by defining the reward

$$r_k = -c_e(k)\,P_{\mathrm{imp}}(k) - \lambda_{\mathrm{exp}}\,P_{\mathrm{exp}}(k) - \lambda_c\,|T_z(k) - T_{\mathrm{set}}(k)|_+ - \lambda_s\,\|\Delta u_k\|_1, \tag{11}$$

and training a policy $\pi_\theta(a_k \,|\, s_k)$ in the twin, then deploying it with safety interlocks. Performance is reported with quantitative, reproducible indicators computed online: the self-consumption ratio

$$\mathrm{SCR} = \frac{\sum_k \min\{P_{\mathrm{pv}}(k),\ P_{\mathrm{load}}(k) + u_k\}}{\sum_k P_{\mathrm{pv}}(k)}, \tag{12}$$

the self-sufficiency ratio

$$\mathrm{SSR} = \frac{\sum_k \min\{P_{\mathrm{pv}}(k),\ P_{\mathrm{load}}(k) + u_k\}}{\sum_k \big(P_{\mathrm{load}}(k) + u_k\big)}, \tag{13}$$

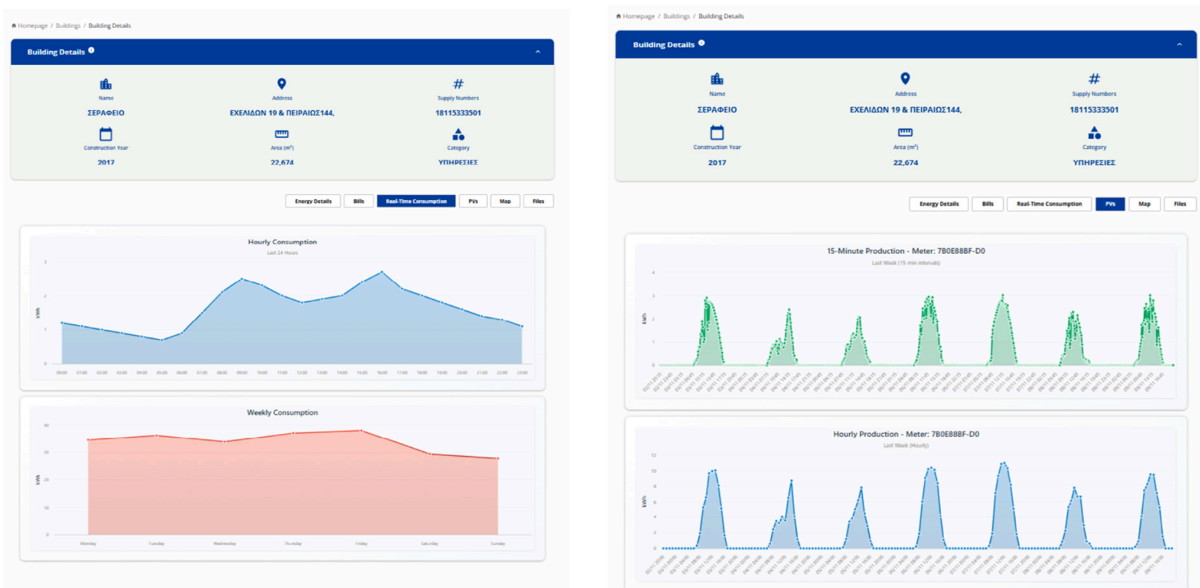
peak-demand reduction, comfort-violation rate (fraction of minutes with $T_z \notin [T_{\min}, T_{\max}]$), and end-to-end latency. Stess tests are exercised first at accelerated time in the twin and then in shadow mode on the live system. Fig. 7 (Serafeio consumption and production screenshots) illustrates the city platform views used during evaluation — hourly/weekly building demand and PV production — which are linked in EnerTEF to the controller's logs and compliance reports to verify that higher self-consumption is achieved without degrading comfort.

The controller was benchmarked against the city's rule–based HVAC schedule (baseline) using minute-level telemetry from Serafeio. Over the three summer months the EnerTEF policy increased on-site PV use and reduced imports without degrading comfort. SCR rose by 22%, SSR by 11%, and peak demand was cut by 14%. Grid exports fell as flexible load was shifted to PV hours. Comfort-violation rate remained below 1% and end-to-end median latency stayed well under one second. During a one-week heatwave (mean T_{out}>35 °C), SCR remained 58% versus 37% for the baseline while comfort constraints were respected (see Table 4).

5. Discussion

The EnerTEF infrastructure reflects a deliberate response to an increasingly urgent challenge: the absence of controlled yet contextually meaningful environments in which to assess the systemic implications of AI deployment in energy systems. What sets EnerTEF apart is its grounding in infrastructure realism, its capacity to link AI evaluation with the operational characteristics of actual physical systems, and its attention to the institutional and regulatory frictions that accompany such deployment. The use cases described earlier were not selected for their novelty alone but for their ability to stress the infrastructural, organizational and regulatory boundaries within which AI must operate.

The core methodological insight from these experiments lies in EnerTEF's capacity to identify model fragility that is typically concealed under controlled conditions. In contrast to simulation-heavy testing environments, where boundary conditions can be carefully contained,



(a) Consumption of the Serafeio Building (b) Rooftop-PV production of the Serafeio Building

Fig. 7. Sample interface from the City of Athens municipal building digitalization platform, integrated within TEF BUILD. The interface displays building details and real-time energy consumption and production data at hourly and weekly resolutions, supporting predictive control strategies and enabling real-time demand response and self-consumption optimization across the city's public building portfolio.

Table 4
Serafeio building, June-August 2025: self-consumption (SCR), self-sufficiency (SSR), peak demand, imported/exported energy, and comfort-violation rate. "pp" denotes percentage points.

Method	SCR (%)	SSR (%)	Peak (kW)	Imports (MWh)	Exports (MWh)	Comfort viol. (% time)
Baseline (rule-based)	41.2	28.1	310	242.6	96.4	0.9
EnerTEF controller	**63.4**	**39.5**	**267**	**199.0**	**62.7**	**0.7**
Relative change	+22.2 pp	+11.4 pp	−14.0%	−18.0%	−35.0%	−0.2 pp

EnerTEF's testbeds embed AI systems in live, heterogeneous contexts characterized by asynchronous data, partial observability and infrastructure constraints. These contexts expose latent assumptions within AI models regarding data availability, actuation latencies, or environmental predictability that would otherwise be obscured. The capacity to detect such brittleness in early development stages represents a material gain in both safety and deployment viability.

The federated design of the infrastructure further extends its analytical capacity. By decentralizing experimentation across nodes with divergent infrastructure characteristics and regulatory obligations, EnerTEF introduces a vector of systemic diversity into the testing process. This enables comparative performance analysis across settings, which is particularly relevant in the fragmented landscape of European energy systems. The orchestration layer facilitates this cross-node experimentation without flattening heterogeneity, supporting simultaneous local adaptation and controlled generalization. The result is a testing protocol capable of approximating deployment complexity without sacrificing experimental control.

Institutionally, EnerTEF constitutes a shift in the epistemology of AI testing: from one premised on technical sufficiency to one oriented toward regulatory and ethical coherence. Embedding regulatory sandboxes and stakeholder review directly within the test workflow is not merely a procedural adjustment; it repositions experimentation as a site of governance. This model presumes that algorithmic behavior cannot be evaluated independently of its socio-institutional context. It also foregrounds tensions: between model opacity and the legal imperative for explainability, between optimization performance and distributional fairness and between technical flexibility and institutional standardization. These tensions are neither incidental nor easily resolvable; they are constitutive of AI deployment in high-stakes domains and must be engaged structurally.

Despite these advances, the scalability of EnerTEF across domains, governance models, and technical protocols remains an open question. Operationally, two constraints are prioritized: affordability for small and medium-sized enterprises (SMEs) and startups, and clarity over governance and ownership. EnerTEF will be based on a three-tier, usage-based access model with trial credits, transparent per-experiment budget caps, off-peak and spot pricing, and optional priority support under a service-level agreement (SLA); discounts for smaller organizations will follow applicable EU state-aid rules. Shared evaluation metrics, standardized experiment documentation, and alignment with European certification processes will be formalized. At the technical level, scale costs will be reduced by moving models to the data to limit egress, employing cached reference datasets, packaging experiments as reproducible containers, and deploying a budget-aware scheduler with explicit service objectives to manage contention across nodes. For governance, user-retained intellectual property with clear access-rights and joint-results provisions will be the default; cross-node operation will be supported through federated identity, standard Data Processing Agreements (DPAs) and Standard Contractual Clauses (SCCs) for cross-border data, and immutable audit logs. A focused set of key performance indicators (KPIs), including cost to serve, time to approval, reproducibility rate, utilization, and conversion from trial to paid access, will be reported and benchmarked against peer testing facilities.

If these extensions are realized, EnerTEF can evolve from a testbed to a regulatory and methodological reference infrastructure.

6. Conclusion and future work

This paper has presented EnerTEF as a practical and forward-looking response to the growing need for rigorous, transparent and multi-contextual testing of AI systems in the energy domain. The facility does more than host technical evaluations; it provides a structured space in which complex systems, human actors and legal frameworks intersect. The fusion of federated infrastructures with real-time experimentation and embedded governance mechanisms sets EnerTEF apart from both simulation-only platforms and isolated testbeds. It offers a template for how digital innovation in high-stakes sectors can be developed with integrity from the outset.

The outcomes of the use cases presented, spanning predictive building control, distributed solar forecasting and market-based flexibility, highlight not only the versatility of the EnerTEF setup but its ability to uncover how AI systems behave under actual operational constraints. These experiments reaffirm that performance metrics alone are insufficient. What matters equally is how these systems respond to unpredictable inputs, how their decisions are interpreted by users and regulators and whether their operation can withstand the pressures of scale, variability and scrutiny. In this regard, EnerTEF succeeds not just in technical validation, but in creating the conditions for social and institutional learning around AI.

Looking beyond the energy sector, the architecture and methods pioneered by EnerTEF have relevance wherever digital systems meet physical infrastructure and public accountability. The idea that testing must be federated, co-creative and legally informed is not unique to energy. It is a condition of responsible innovation in any critical domain. As new European frameworks for AI governance take shape, infrastructures like EnerTEF can play a dual role: supporting developers in meeting compliance requirements and informing policymakers about the practicalities of implementation.

Future work will aim to expand the EnerTEF network to include additional domains and geographies, mainly oriented towards the modernization of transmission and distribution grids and collaboration between TSOs and DSOs, following the strategic plan of the European Commission for a new era of digitalized energy networks. Priority will be given to joint TSO–DSO use cases, including situational awareness across voltage levels, flexibility activation and verification, coordination of protection and stability in inverter-rich systems, so that grid operations can be stress-tested under realistic cross-boundary conditions. As a unifying research line, a foundation model for grid operators will be pursued, trained on multimodal operational data from TSO and DSO contexts (events, telemetry, topology, weather, and markets) and designed for safe adaptation to local conditions. To support this agenda, EnerTEF will expand to new nodes built for grid collaboration allowing for data access, interoperability, and operational safeguards. Subsequent expansion to adjacent domains, such as renewable energy, e-mobility, buildings, hydrogen, and district heating will be undertaken in concert with the grid end-users, ensuring that services in these domains are validated against TSO-DSO operational constraints. Geographic roll-out will prioritize jurisdictions with open data and data sharing–friendly legislation, clear interoperability mandates, and mature regulatory sandbox processes, in order to enable lawful access, reproducibility, and efficient federation formation before broader deployment.

CRediT authorship contribution statement

Elissaios Sarmas: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Alexandre Lucas:** Writing – original draft, Validation, Resources, Project administration, Investigation, Formal analysis, Conceptualization. **Andrés Acosta:** Writing – original draft, Visualization, Supervision, Software, Methodology, Data curation. **Ferdinanda Ponci:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Conceptualization. **Pedro Rodriguez:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Vangelis Marinakis:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

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.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

Acknowledgment

The work presented is based on research conducted within the framework of the Horizon Europe research and innovation programme EnerTEF (Grant Agreement No. 101172887). The content of the paper is the sole responsibility of its author and does not necessarily reflect the views of the EC.

Data availability

Data will be made available on request.

References

Aaronson, S.A., 2019. Data is different, and that's why the world needs a new approach to governing cross-border data flows. Digit. Policy, Regul. Gov. 21 (5), 441–460.

Bach, A., Monti, A., 2025. Remote real-time testing of physical components using communication setup automation. IEEE Access.

Bazmohammadi, N., Madary, A., Vasquez, J.C., Mohammadi, H.B., Khan, B., Wu, Y., Guerrero, J.M., 2021. Microgrid digital twins: Concepts, applications, and future trends. IEEE Access 10, 2284–2302.

Cheng, X., Li, C., Liu, X., 2022. A review of federated learning in energy systems. In: 2022 IEEE/IAS Industrial and Commercial Power System Asia. I&CpS Asia, IEEE, pp. 2089–2095.

Djebali, S., Guerard, G., Taleb, I., 2024. Survey and insights on digital twins design and smart grid's applications. Future Gener. Comput. Syst. 153, 234–248.

Ebe, F., Idlbi, B., Stakic, D.E., Chen, S., Kondzialka, C., Casel, M., Heilscher, G., Seitl, C., Bruendlinger, R., Strasser, T.I., 2018. Comparison of power hardware-in-the-loop approaches for the testing of smart grid controls. Energies 11 (12), 3381.

Efthymiopoulos, N., Makris, P., Tsaousoglou, G., Steriotis, K., Vergados, D.J., Khaksari, A., Herre, L., Lacort, V., Martinez, G., Lorente, E.L., et al., 2023. FLEXGRID–A novel smart grid architecture that facilitates high-RES penetration through innovative flexibility markets towards efficient stakeholder interaction. Open Res. Eur. 1, 128.

Egorov, A., Ganesan, K., Glória, G., Silva, R.S., e Silva, N.S., Krisper, U., 2025. Comprehensive governance models for energy data spaces: A path to secure and interoperable data sharing in Europe. In: 2025 21st International Conference on the European Energy Market. EEM, IEEE, pp. 1–6.

Ejiyi, C.J., Cai, D., Thomas, D., Obiora, S., Osei-Mensah, E., Acen, C., Eze, F.O., Sam, F., Zhang, Q., Bamisile, O.O., 2025. Comprehensive review of artificial intelligence applications in renewable energy systems: current implementations and emerging trends. J. Big Data 12 (1), 169.

European Commission, 2022. Digitalising the Energy System – EU Action Plan. Communication COM(2022) 552 Final, European Commission, URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022DC0552.

European Parliament and Council of the European Union, 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence. Regulation 2024/1689, Official Journal of the European Union, URL http://data.europa.eu/eli/reg/2024/1689/oj.

Fragkiadaki, A., Sarmas, E., Anastasiadis, A., Papadimitriou, P., Marinakis, V., 2025. Short-term forecasting of hydropower plant generation using the prophet model. In: 2025 16th International Conference on Information, Intelligence, Systems & Applications. IISA, IEEE, pp. 1–8, (in press).

Ghoreishi, M., Carreras, L., Nakti, G., Santori, F., Monti, A., Geri, A., 2025. Advancements in digital twin: Distributed energy resources simulation for supporting system level scenario evaluation. In: 2025 IEEE Kiel PowerTech. IEEE, pp. 1–6.

González, G., Evans, C.L., 2019. Biomedical Image Processing with Containers and Deep Learning: An Automated Analysis Pipeline: Data architecture, artificial intelligence, automated processing, containerization, and clusters orchestration ease the transition from data acquisition to insights in medium-to-large datasets. BioEssays 41 (6), 1900004.

Grataloup, A., Jonas, S., Meyer, A., 2024. A review of federated learning in renewable energy applications: Potential, challenges, and future directions. Energy AI 17, 100375.

Guerra, E., Wilhelmi, F., Miozzo, M., Dini, P., 2023. The cost of training machine learning models over distributed data sources. IEEE Open J. Commun. Soc. 4, 1111–1126.

Heussen, K., Steinbrink, C., Abdulhadi, I.F., Nguyen, V.H., Degefa, M.Z., Merino, J., Jensen, T.V., Guo, H., Gehrke, O., Bondy, D.E.M., et al., 2019. Erigrid holistic test description for validating cyber-physical energy systems. Energies 12 (14), 2722.

Hou, L., Tong, X., Chen, H., Fan, L., Liu, T., Liu, W., Liu, T., 2024. Optimized scheduling of smart community energy systems considering demand response and shared energy storage. Energy 295, 131066.

Jørgensen, B.N., Gunasekaran, S.S., Ma, Z.G., 2025. Impact of EU laws on AI adoption in smart grids: A review of regulatory barriers, technological challenges, and stakeholder benefits. Energies 18 (12), 3002.

Kabir, M.R., Halder, D., Ray, S., 2024. Digital twins for iot-driven energy systems: A survey. IEEE Access.

Kallert, A., Lottis, D., Shan, M., Schmidt, D., 2021. New experimental facility for innovative district heating systems—District LAB. Energy Rep. 7, 62–69.

Khan, M.A., Saleh, A.M., Waseem, M., Sajjad, I.A., 2022. Artificial intelligence enabled demand response: Prospects and challenges in smart grid environment. Ieee Access 11, 1477–1505.

Khomami, H.P., Fonteijn, R., Geelen, D., 2020. Flexibility market design for congestion management in smart distribution grids: The dutch demonstration of the interflex project. In: 2020 IEEE PES Innovative Smart Grid Technologies Europe. ISGT-Europe, IEEE, pp. 1191–1195.

Kyriakarakos, G., 2025. Artificial intelligence and the energy transition. Sustainability 17 (3), 1140.

Lavin, A., Gilligan-Lee, C.M., Visnjic, A., Ganju, S., Newman, D., Ganguly, S., Lange, D., Baydin, A.G., Sharma, A., Gibson, A., et al., 2022. Technology readiness levels for machine learning systems. Nat. Commun. 13 (1), 6039.

Machlev, R., Heistrene, L., Perl, M., Levy, K.Y., Belikov, J., Mannor, S., Levron, Y., 2022. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. Energy AI 9, 100169.

Martin, N., Madrid-López, C., Talens-Peiró, L., Süsser, D., Gaschnig, H., Lilliestam, J., 2020. Observed Trends and Modelling Paradigms on the Social and Environmental Aspects of the Energy Transition. Deliverable 2.1. Sustainable Energy Transitions Laboratory (SENTINEL) Project. Universitat de Barcelona, Institute for Advanced Sustainability Studies (IASS).

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

Michalakopoulos, V., Sarantinopoulos, E., Sarmas, E., Marinakis, V., 2024. Empowering federated learning techniques for privacy-preserving pv forecasting. Energy Rep. 12, 2244–2256.

Mirz, M., Vogel, S., Reinke, G., Monti, A., 2019. DPsim—A dynamic phasor real-time simulator for power systems. SoftwareX 10, 100253.

Nicholson, A., Webber, S., Dyer, S., Patel, T., Janicke, H., 2012. SCADA security in the light of Cyber-Warfare. Comput. Secur. 31 (4), 418–436.

Panagoulias, D.P., Sarmas, E., Marinakis, V., Virvou, M., Tsihrintzis, G.A., Doukas, H., 2023. Intelligent decision support for energy management: A methodology for tailored explainability of artificial intelligence analytics. Electronics 12 (21), 4430.

Papias, I., Michalakopoulos, V., Sarmas, E., Marinakis, V., 2024. Aggregated flexibility dynamic forecasting of residential energy prosumers for DR programs. In: 2024 15th International Conference on Information, Intelligence, Systems & Applications. IISA, IEEE, pp. 1–9.

Pellegrino, L., Lazzari, R., Verga, M., Sandroni, C., 2020. Research infrastructures integration to foster smart grid testing. In: 2020 AEIT International Annual Conference. AEIT, IEEE, pp. 1–5.

Rajaperumal, T., Columbus, C.C., 2025. Transforming the electrical grid: the role of AI in advancing smart, sustainable, and secure energy systems. Energy Inform. 8 (1), 51.

Sarmas, E., Dimitropoulos, N., Marinakis, V., Mylona, Z., Doukas, H., 2022. Transfer learning strategies for solar power forecasting under data scarcity. Sci. Rep. 12 (1), 14643.

Sarmas, E., Forouli, A., Marinakis, V., Doukas, H., 2024. Baseline energy modeling for improved measurement and verification through the use of ensemble artificial intelligence models. Inform. Sci. 654, 119879.

Sawas, A., Farag, H.E., 2023. Real-time detection of stealthy IoT-based cyber-attacks on power distribution systems: A novel anomaly prediction approach. Electr. Power Syst. Res. 223, 109496.

Shaukat, N., Islam, M.R., Rahman, M.M., Khan, B., Ullah, B., Ali, S.M., Fekih, A., 2023. Decentralized, democratized, and decarbonized future electric power distribution grids: a survey on the paradigm shift from the conventional power system to micro grid structures. IEEE Access 11, 60957–60987.

Srivastava, P.R., Mangla, S.K., Eachempati, P., Tiwari, A.K., 2023. An explainable artificial intelligence approach to understanding drivers of economic energy consumption and sustainability. Energy Econ. 125, 106868.

Ssekulima, E.B., Anwar, M.B., Al Hinai, A., El Moursi, M.S., 2016. Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review. IET Renew. Power Gener. 10 (7), 885–989.

Tardieu, H., 2022. Role of Gaia-X in the European data space ecosystem. In: Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Springer International Publishing Cham, pp. 41–59.

Thombs, R.P., 2019. When democracy meets energy transitions: A typology of social power and energy system scale. Energy Res. Soc. Sci. 52, 159–168.

Toderean, L., Cioara, T., Anghel, I., Sarmas, E., Michalakopoulos, V., Marinakis, V., 2025. Demand response optimization for smart grid integrated buildings: review of technology enablers landscape and innovation challenges. Energy Build. 326, 115067.

Wichert, B., Dymond, M., Lawrance, W., Friese, T., 2001. Development of a test facility for photovoltaic-diesel hybrid energy systems. Renew. Energy 22 (1–3), 311–319.

Wijaya, E.P., Asif, M., 2025. Technology readiness level assessment on digital technologies for energy efficiency. Transp. Res. Procedia 84, 512–519.

Wolk, S., Reinhart, C., 2025. Semantic building energy modeling: Analysis across geospatial scales. Build. Environ. 276, 112883.

Wu, Y., Wu, Y., Guerrero, J.M., Vasquez, J.C., 2021. Digitalization and decentralization driving transactive energy Internet: Key technologies and infrastructures. Int. J. Electr. Power Energy Syst. 126, 106593.

Zhang, G., Hu, W., Cao, D., Huang, Q., Chen, Z., Blaabjerg, F., 2021. A novel deep reinforcement learning enabled sparsity promoting adaptive control method to improve the stability of power systems with wind energy penetration. Renew. Energy 178, 363–376.