Review

Federated Learning for Power Cyber-Physical Systems: Toward Secure, Resilient, and Explainable Intelligence

Zhiye Wang [*]

Posted Date: 17 September 2025

doi: 10.20944/preprints202509.1447.v1

Keywords: federated learning; power cyber-physical systems; security and resilience; explainable artificial intelligence; zero-trust architecture; digital twin validation; smart grid and ev integration

Review

Federated Learning for Power Cyber-Physical Systems: Toward Secure, Resilient, and Explainable Intelligence

Zhiye Wang

School of Electrical Engineering, Jiangxi University of Water Resources and Electric Power, Nanchang, China

Abstract

The digital transformation of power cyber-physical systems (CPSs) introduces unprecedented opportunities for optimization, forecasting, and real-time control, while simultaneously exposing critical vulnerabilities in data security, system resilience, and operator trust. Federated Learning (FL) provides a promising paradigm by enabling collaborative intelligence without raw data sharing, yet traditional approaches fall short in safety-critical energy infrastructures. This review advances the state of the art by presenting a holistic perspective on secure, resilient, and explainable FL for Power CPSs. We first analyze emerging threats—including model poisoning, backdoor insertion, and cross-layer false data injection—and map them to existing defenses such as robust aggregation, Byzantine resilience, differential privacy, and zero-trust authentication. We then synthesize architectural innovations, including personalized FL, digital twin–enhanced validation, and human-in-the-loop trust calibration, highlighting their potential to address system heterogeneity and operational risks. Real-world applications in load forecasting, intrusion detection, EV coordination, and microgrid control are surveyed to demonstrate feasibility. Finally, we outline future research directions linking adversarial robustness, explainability, scalable integration, and governance frameworks. This work positions federated learning as a cornerstone for trustworthy intelligence in next-generation power systems.

Keywords: federated learning; power cyber-physical systems; security and resilience; explainable artificial intelligence; zero-trust architecture; digital twin validation; smart grid and ev integration

1. Introduction

1.1. Background and Motivation

The digital transformation of electrical infrastructures is rapidly reshaping modern power systems into Power Cyber-Physical Systems (Power CPSs)—interconnected networks of physical assets, communication protocols, and intelligent control modules [1–3], as shown in Fig. 1 [4]. This transformation unlocks opportunities for optimization, fault detection, and real-time control, but also introduces new vulnerabilities in terms of data privacy, adversarial attack surfaces, and systemic trust [5].

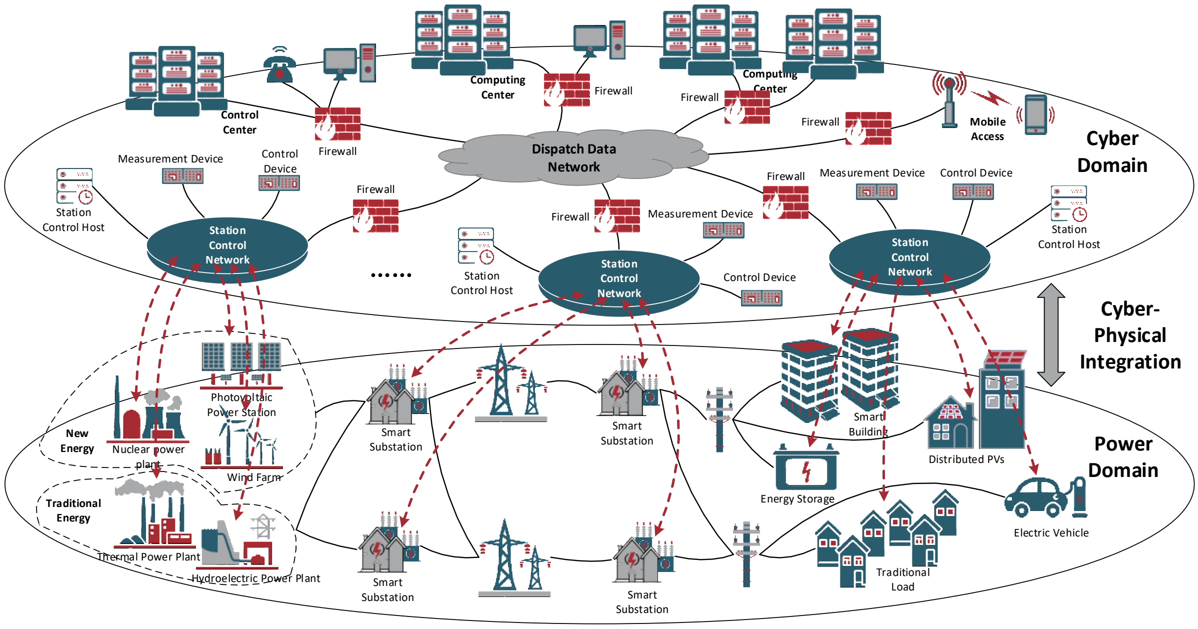

Figure 1. Cyber-Physical Architecture of Modern Power Systems.

Federated Learning (FL) has emerged as a promising paradigm to enable collaborative intelligence among power system stakeholders—such as transmission operators, distribution networks, microgrids [6,7], and electric vehicle (EV) aggregators [8,9]—without sharing raw data [10–12]. By training models locally and aggregating updates, FL complies with data sovereignty regulations and privacy mandates, which makes it particularly attractive for privacy-sensitive domains such as Supervisory Control and Data Acquisition (SCADA) [13] and EV charging infrastructures [14,15].

Yet, direct adoption of FL in Power CPSs faces critical **trust barriers**. Although privacy-preserving by design, FL is inherently vulnerable to malicious participants, model poisoning, free-rider behaviors, and the lack of interpretability. These weaknesses are unacceptable in safety-critical systems, where resilience, accountability, and ethical deployment are non-negotiable [16]. Traditional metrics such as accuracy and convergence fail to capture the full requirements of trustworthy operation in power grids.

To bridge this gap, the paradigm of Trustworthy Federated Learning (TFL) has emerged [17,18]. Unlike conventional FL approaches that focus primarily on privacy, TFL expands the scope to encompass five interdependent pillars: robustness against adversarial manipulation [19]; fairness in performance across heterogeneous clients; explainability to ensure transparent and interpretable decisions [20]; accountability via auditable and secure learning workflows; and resilience to maintain operation under disruptions such as communication delays or client dropouts [21,22]. Together, these dimensions form the foundation of a trustworthy federated framework tailored to the high-stakes environment of power system operations.

Despite increasing attention, the literature remains fragmented. Most works address isolated technical aspects—such as secure aggregation or differential privacy—while lacking a unified view of trust as a multi-dimensional system requirement in Power CPSs. This review responds to that gap by integrating existing research, introducing a holistic trust framework, and mapping architectural paradigms to the specific challenges of Power CPSs.

1.2. Contributions of this Review

This review advances the understanding of Trustworthy Federated Learning (TFL) in Power CPSs through three key contributions:

Multi-Dimensional Trust Framework: We propose a holistic trust perspective that extends beyond privacy to encompass security, resilience, and explainability, while also addressing fairness and accountability. Unlike conventional FL surveys, this framework positions trust not as a single technical requirement but as a cross-cutting operational principle essential for heterogeneous and safety-critical CPS environments.

Architecture–Trust–Application Mapping: We establish a structured mapping that links federated learning architectures—horizontal, vertical, cross-silo, cross-device, zero-trust, personalized, and digital twin–enhanced—to their corresponding trust characteristics and practical applications in power systems, such as load forecasting, intrusion detection, EV coordination, and microgrid management. This integration clarifies deployment trade-offs and highlights research gaps.

Threat–Defense–Gap Synthesis: We deliver the first Power CPS–specific synthesis that systematically maps adversarial threats (e.g., poisoning, backdoors, Sybil, Byzantine, and false data injection) to evolving defenses (robust aggregation, differential privacy, secure multi-party computation, blockchain, and zero-trust validation). By identifying both strengths and open gaps, we provide a critical agenda for developing adaptive, hybrid, and resilient defense frameworks.

2. Foundations of TFL in Power CPS

2.1. Overview of Federated Learning in Power CPSs Context

A power CPS typically comprises three tightly coupled layers—cyber, integration, and physical—highlighting the interplay between data, control, and energy flows. This layered architecture provides the foundation for understanding how FL can be embedded within power system operations. FL is a decentralized paradigm in which multiple clients—such as substations, renewable energy assets [23,24], or EV charging clusters—collaboratively train a global model while keeping their raw data local [25–27]. This inherently aligns with data locality, privacy mandates, and the heterogeneous nature of power infrastructure systems [28].

In FL for Power CPSs, the overarching goal is to collaboratively train a global model by leveraging distributed data from multiple entities such as substations, renewable plants, and EV clusters, without centralizing sensitive measurements. Physically, this formulation reflects the idea that each client contributes to the system-wide intelligence based on its own operational data, while preserving local privacy and autonomy. The global optimization objective can be expressed as:

$$\min_{\mathbf{w}\subset\mathbb{R}^d} F(\mathbf{w}) = \sum_{i=1}^{N} p_i F_i(\mathbf{w}), \quad F_i(\mathbf{w}) = \mathbb{E}_{\xi\sim\mathcal{D}_i}[\ell(\mathbf{w};\xi)] \tag{1}$$

where $\mathbf{w}$ denotes the global model parameter vector of dimension d, which is jointly optimized across all participating clients. The total number of clients is N, representing distributed entities such as power plants, substations, or EV aggregators. Each client i is assigned a weight p_i, typically normalized according to the size of its local dataset $\mathcal{D}_i$. The function $F_i(\mathbf{w})$ corresponds to the local objective defined on client i, expressed as the expected loss $\ell(\mathbf{w};\xi)$ over a training sample ξ drawn from $\mathcal{D}_i$. Collectively, these definitions capture how the global model integrates heterogeneous data contributions from diverse CPS components while preserving local autonomy.

To operationalize the global optimization in federated learning, each client performs local stochastic gradient descent (SGD) updates before communicating with the central server. This process reflects the physical setting of Power CPSs, where substations, renewable assets, or EV fleets independently process their data and only share model updates rather than raw measurements. The global model is then aggregated from these local updates, ensuring both efficiency and privacy. Formally, the procedure is described as:

$$\mathbf{w}_i^{t,k+1} = \mathbf{w}_i^{t,k} - \eta \nabla \ell\left(\mathbf{w}_i^{t,k}; \xi_i^{t,k}\right), \quad \mathbf{w}_i^{t,0} = \mathbf{w}^t, \mathbf{w}^{t+1} = \sum_{i \in \mathcal{S}_t} \frac{n_i}{\sum_{j \in \mathcal{S}_t} n_j} \mathbf{w}_i^{t,K}. \tag{2}$$

Here, $\mathbf{w}_i^{t,k}$ denotes the local model parameters of client *i* at communication round *t* and local iteration *k*, while $\mathbf{w}^t$ represents the global model at the beginning of round *t*. Each local update is computed using a learning rate η and a sampled data point $\xi_i^{t,k}$ from the local dataset. After *K* local steps, the final local model $\mathbf{w}_i^{t,k}$ is sent to the server. The central server aggregates the participating clients s_t by a weighted average, where the weights are proportional to the local dataset sizes n_i. This formulation embodies the FedAvg algorithm, which balances global consistency with local autonomy, making it well-suited for heterogeneous and distributed Power CPS environments.

In Power CPSs, FL facilitates distributed intelligence by supporting both cross-silo and cross-device collaborations [29,30]. In cross-silo settings, FL enables coordinated model training among regional transmission system operators or different divisions within a utility, while in cross-device scenarios, it allows edge devices—such as smart meters, phasor measurement units (PMUs) [31], and DER controllers—to collaboratively learn without exposing raw data [32,33]. These modes fit the distributed and autonomous nature of power grids. However, canonical FL formulations overlook domain-specific requirements such as safety, reliability, and adversarial threat models. Hence, embedding trustworthiness into FL becomes indispensable [34].

2.3. Dimensions of Trust in Federated Learning

Trust in FL is multi-dimensional and interdependent, particularly in mission-critical CPS environments [35,36]. To establish a foundational understanding of what constitutes trust in Federated Learning for Power CPSs, Table 1 delineates the key dimensions of trustworthiness, highlighting their relevance and critical roles in secure and equitable collaborative intelligence.

Table 1. Core Dimensions of Trustworthiness in Federated Learning for Power CPSs.

Dimension	Description
Privacy	Prevents leakage of sensitive operational data (e.g., load profiles, topology, market signals).
Robustness	Resists poisoning, backdoor insertion, and communication disruption across training cycles.
Fairness	Ensures that all clients—irrespective of size or data heterogeneity—are equitably represented.
Explainability	Provides interpretable model decisions for operators and facilitates compliance auditing.
Accountability	Enables forensic traceability of malicious updates, client behaviors, and federated outcomes.

Each dimension demands domain-aware implementation, often involving new protocols or architectural layers. For instance, privacy may require differential privacy tailored to grid signals, while robustness might require Byzantine-resilient aggregation in environments with variable network latency and node dropout [37,38].

2.4. Architectural Paradigms of TFL in Power Systems

To categorize the deployment strategies of Federated Learning tailored to different operational contexts in power systems, Table 2 presents various FL architectures and their corresponding applications within Power CPSs environments. The increasing complexity of hybrid AC/DC grids and their multi-objective optimization requirements further highlight the need for federated approaches that can accommodate heterogeneous architectures while ensuring privacy and resilience [39].

Table 2. Federated Learning Architectures and Their Applications in Power CPSs.

Architecture	Application in Power CPSs
Horizontal FL	Training across substations with common feature spaces (e.g., voltage, frequency, power flow).
Vertical FL	Used in multi-sector applications (e.g., coupling power data with traffic or building systems).
Hybrid FL	Supports multi-view integration in smart cities with coupled energy infrastructures.
Cross-Silo FL	Utility collaboration (e.g., among TSOs, DSOs, and markets) with moderate-sized, reliable clients.
Cross-Device FL	Edge-level coordination among meters, sensors, and DERs; often constrained by bandwidth and energy.

While Table 2 outlines the taxonomy of federated learning architectures, it remains unclear how these architectures map to specific trust characteristics and application domains in Power CPS. Table 3 bridges this gap by linking representative architectures with their inherent trust dimensions, typical application scenarios, and their strengths and limitations. This mapping highlights not only the diversity of architectural designs but also the practical trade-offs in deploying FL within energy-critical infrastructures.

Table 3. Mapping of Federated Learning Architectures to Trust Characteristics and Application Scenarios in Power CPS.

FL Architecture	Key Trust Characteristics	Typical Application Scenarios in Power CPS	Strengths	Limitations / Gaps
Horizontal FL (Cross-device)	Privacy preservation, scalability	Smart meter data sharing, EV charging coordination	Supports large-scale clients; low raw data leakage	Vulnerable to non-IID data; high communication overhead
Vertical FL (Cross-silo)	Data integration across domains	Utility–bank collaboration (load prediction + credit scoring)	Enables feature complementarity across entities	Requires strict alignment of data; privacy leakage risk in gradients
Cross-silo FL (Consortium)	Robustness, accountability	Regional power utilities collaboration	Stable communication; easier governance	Limited scalability; possible collusion risks
Cross-device FL	Fairness, personalization	Household-level demand response	Adaptation to heterogeneous devices	High dropout rate; low device reliability

FL Architecture	Key Trust Characteristics	Typical Application Scenarios in Power CPS	Strengths	Limitations / Gaps
Zero-Trust FL	Authentication, verifiability	Substation automation, intrusion detection	End-to-end trust; resistant to insider threats	Implementation complexity; high overhead in large systems
Personalized FL (PFL)	Fairness, heterogeneity adaptation	Distributed renewable forecasting (solar/wind farms)	Tackles data heterogeneity; improved accuracy	Trade-off between personalization and global model generalization
Explainable FL (XFL)	Transparency, accountability	Operator decision support, regulatory compliance	Enhances interpretability; human-in-the-loop	Explainability vs. performance trade-off
Digital Twin–Enhanced FL	Validation, resilience	Grid stability monitoring, EV-grid interaction	Enables simulation-based validation	Model drift between real and simulated environments
Human-in-the-Loop FL	Trust calibration, accountability	Operator-assisted intrusion detection, emergency dispatch	Improves human trust in AI; aligns with regulations	Potentially slow response; reliance on expert input

As illustrated in Table 3, no single FL architecture can simultaneously guarantee privacy, resilience, fairness, and explainability under the heterogeneous and real-time conditions of Power CPS. This observation underscores the necessity for hybrid or adaptive frameworks that integrate multiple trust dimensions, potentially combining zero-trust mechanisms, digital twin validation, and personalized modeling.

2.5. Trust-Aware Protocol Stack in Power CPSs

This paper proposes a layered abstraction to embed trust into FL deployments for Power CPSs:

Data Layer: Includes privacy-preserving encodings, differential privacy, and local perturbations to protect raw measurements.

Communication Layer: Supports secure multiparty computation, homomorphic encryption, and secure aggregation to ensure confidentiality and integrity.

Aggregation Layer: Implements robust federated averaging with trust weighting, anomaly scoring, and client reputation systems.

Model Layer: Enhances interpretability through explainable AI (XAI), ensures robustness via adversarial training and validation.

Governance Layer: Auditing, rollback mechanisms, incentive-compatible participation, and regulatory compliance modules.

This modular design allows plug-and-play of trust primitives suited to different operational scenarios, offering a foundation for dynamic, secure, and accountable learning in real-time power system environments [40–42].

2.5. Why Trustworthiness Is Essential in Power CPSs

Unlike commercial or consumer applications, Power CPSs operate under hard constraints on reliability, availability, and safety [43,44]. Model failures can trigger wide-area blackouts or market distortions. Trustworthiness is not optional—it is a functional requirement embedded in:

Grid Protection: Preventing false triggers from poisoned models affecting relay settings or fault classification.

System Restoration: Ensuring reliable decision-making under contingency or disaster recovery conditions.

Operational Markets: Avoiding bias in price forecasts or demand-response coordination.

Human Oversight: Empowering grid operators with confidence in automated recommendations or anomaly alerts.

3. Emerging Threat Landscape in Federated Power CPSs

While FL offers structural privacy benefits, its deployment in power CPSs introduces a new class of system-level threats [45]. These threats exploit both the distributed nature of FL and the criticality of power infrastructure, demanding a systematic rethinking of defense assumptions. As illustrated in Fig. 2, the past decade has witnessed several major cyber incidents targeting power grids worldwide, from the Stuxnet worm in 2010 to recent ransomware attacks on energy companies in 2022 [4,46,47]. These real-world events highlight the increasing sophistication and frequency of cyber threats, reinforcing the urgent need for a trustworthy federated learning paradigm tailored to the unique requirements of power CPSs.

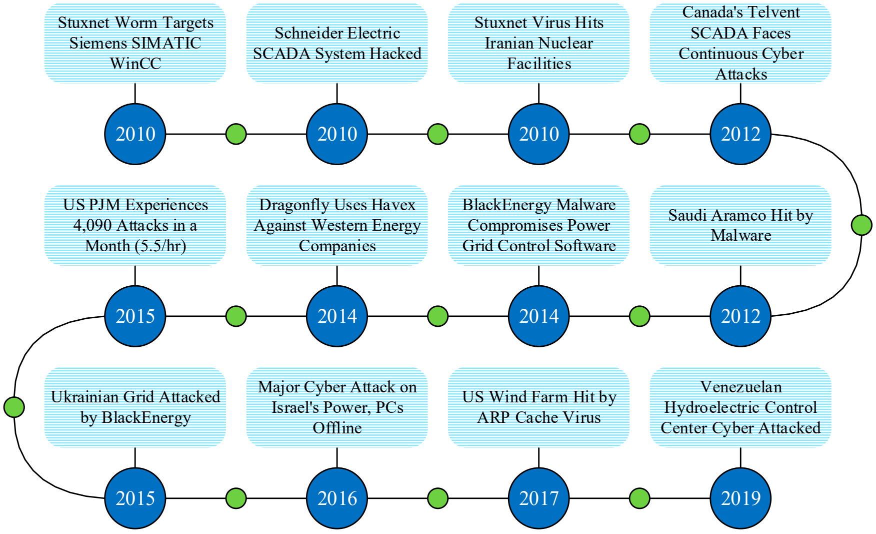
Stuxnet Worm Targets Siemens SIMATIC WinCC
2010
Schneider Electric SCADA System Hacked
2010
Stuxnet Virus Hits Iranian Nuclear Facilities
2010
Canada's Telvent SCADA Faces Continuous Cyber Attacks
2012
Saudi Aramco Hit by Malware
2012
BlackEnergy Malware Compromises Power Grid Control Software
2014
Dragonfly Uses Havex Against Western Energy Companies
2014
US PJM Experiences 4,090 Attacks in a Month (5.5/hr)
2015
Ukrainian Grid Attacked by BlackEnergy
2015
Major Cyber Attack on Israel's Power, PCs Offline
2016
US Wind Farm Hit by ARP Cache Virus
2017
Venezuelan Hydroelectric Control Center Cyber Attacked
2019

Figure 2. Selected Major Cyber Attacks on Power Grids.

3.1. Adversarial Attacks on FL Models

Power CPSs face an expanded adversarial surface under the FL paradigm, primarily due to the potential presence of malicious or compromised clients in collaborative training [48]. One critical threat is model poisoning, where adversaries inject biased gradients to skew the global model—for instance, manipulating forecasting models to underestimate peak loads and thereby jeopardizing operational stability [49–51]. Another concern is backdoor attacks, wherein attackers implant hidden triggers that cause erroneous behavior only under specific conditions, such as phase angle inputs mimicking false fault signatures [52]. Despite FL's decentralization, gradient inversion attacks can still extract sensitive information from shared updates; adversaries may reconstruct load profiles, grid topologies, or event timelines from gradient data. Additionally, free-riding and lazy update behaviors—where certain clients either contribute nothing or provide low-quality updates—can silently degrade model quality and undermine trust in the aggregation process, especially in the absence of effective participation validation mechanisms [53–55].

3.2. Cross-Layer CPS-Specific Threats

Federated Learning in Power CPSs operates beyond the confines of model training—it is deeply intertwined with physical devices, grid operational dynamics, and human-in-the-loop decision processes [56–58]. This tight coupling exposes the system to cross-domain exploits, wherein adversaries may manipulate external data sources such as weather or transportation streams integrated into multimodal FL pipelines, thereby indirectly influencing energy dispatch or DER control actions [59,60]. Moreover, False Data Injection (FDI) via FL presents a growing threat: adversaries may compromise local sensors or edge devices, feeding manipulated inputs into FL-based anomaly detection or state estimation models, which can mislead system responses at scale [61–63]. Equally concerning are control loop exploits, where a compromised FL-enabled control agent—such as one managing voltage/frequency in microgrids—can initiate destabilizing actions that propagate through the system, particularly in scenarios relying on minimal human intervention and real-time closed-loop control [64,65].

3.3. Threats Unique to Federated Architectures

To expose the diverse and evolving threat landscape that undermines the reliability of Federated Learning in power infrastructures, Table 4 summarizes key attack types that target clients, models, and communication in Power CPS.

Table 4. Emerging Threat Types in Federated Learning for Power CPS.

Threat Type	Description
Sybil Attacks	A single adversary controls multiple clients, amplifying malicious gradient contributions.
Model Drift Amplification	Heterogeneous grid dynamics lead to non-IID data, which can exacerbate drift and hide poisoning.
Byzantine Failures	Some clients may behave arbitrarily (crashed, slow, malicious), breaking convergence guarantees.
Communication Interference	Adversaries disrupt or delay model update transmissions, leading to **stale updates** and **aggregation failures**.
Replay Attacks	Attackers reuse previously valid updates to **disrupt convergence** or confuse temporal reasoning.

These threats exploit the lack of centralized validation, temporal inconsistency, and heterogeneity inherent to power grids and federated learning [66,67].

3.4. Case Study: Threat Simulation in FL-Based Load Forecasting

To illustrate the potential vulnerabilities of FL in Power CPS, consider a federated load forecasting framework deployed across a regional grid involving multiple DER operators [68]. In this setting, an adversary controlling a subset of clients—say, three compromised DER nodes—can strategically execute a three-step model poisoning attack. First, the attacker injects manipulated gradients into the local model updates, deliberately skewing the global forecasting model to underestimate demand during peak load periods [69,70]. Second, to avoid detection by conventional anomaly detection or robust aggregation methods, the attacker mimics benign gradient noise patterns derived from historical training data, thereby camouflaging malicious updates [71]. Finally, the compromised forecast leads to insufficient dispatch commands from the central energy management system, resulting in frequency instability and potential cascading failures across the power grid [72–74]. This example underscores the need for trust-aware defenses that go beyond privacy and consider adaptive robustness and adversarial behavior modeling in federated environments. This scenario shows how FL-specific attacks, when contextualized within Power CPS operations, can amplify physical risks far beyond typical AI domains.

3.5. Summary and Key Insights

Federated Learning in Power CPSs introduces a unique set of vulnerabilities not typically encountered in centralized machine learning frameworks [75–77]. These include challenges in inter-client trustworthiness, communication reliability, and dynamic aggregation integrity, especially under adversarial or resource-constrained conditions [78]. The attack surface in FL deployments is inherently multi-dimensional, spanning cyber (e.g., model poisoning, gradient leakage), physical (e.g., control loop manipulation), and organizational (e.g., insider threats across utilities) layers. These complexities necessitate multi-layered threat models that account for both digital and operational interdependencies [79,80]. Furthermore, the safety-critical nature of power systems, coupled with non-independent and identically distributed (non-IID) data and human-AI decision-making interfaces, amplifies the difficulty of ensuring robust and trustworthy federated learning. Addressing these challenges is pivotal for the reliable integration of FL into real-world Power CPS applications [81–83].

The next section will explore how current defense mechanisms—particularly in federated learning—perform in the face of these emerging threats and what limitations must be addressed. Yet, despite the diversity of proposed countermeasures, most existing studies either isolate cyber or physical aspects, lack scalability under non-IID conditions, or fail to capture human-in-the-loop uncertainties. This fragmentation underscores the pressing need for integrated defense strategies that can simultaneously ensure robustness, explainability, and operational feasibility in Power CPS deployments.

4. Defense Mechanisms: State of the Art and Limitations

In response to the evolving threat landscape outlined in Section 3, a variety of defense mechanisms have been proposed within the FL community. However, their adaptation and effectiveness in Power CPSs remain constrained by operational, architectural, and physical-system-level factors. This section categorizes these defense strategies and critically analyzes their applicability and limitations.

4.1. Robust Aggregation Strategies

To mitigate the impact of model poisoning attacks, robust aggregation mechanisms have been introduced as alternatives to the conventional weighted average used in FedAvg. A widely adopted approach is the geometric median aggregation, formulated as:

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \sum_{i \in \mathcal{S}_t} \| \mathbf{g} - \mathbf{g}_i \|_2 \tag{3}$$

where $\mathbf{g}_i$ denotes the local update from client i, and $\mathcal{S}_t$ is the set of participating clients at round t. Unlike simple averaging, this operator seeks an update vector $\hat{\mathbf{g}}$ that minimizes the total Euclidean distance to all local updates, thereby **reducing the influence of outliers or malicious clients**. Such robust aggregation is particularly relevant in power CPSs, where adversarial manipulation of even a few clients could compromise the safety and reliability of system-wide operations.

Robust aggregation mechanisms are essential in TFL to mitigate the impact of poisoned, anomalous, or strategically manipulated updates during the model aggregation phase [84–86]. Traditional federated averaging becomes unreliable in adversarial settings, prompting the use of robust statistical techniques such as median and trimmed mean, which aim to suppress outliers by considering central tendencies rather than full distributions. Krum and Multi-Krum algorithms improve resilience by selecting client updates closest to the geometric center of the majority cluster, under the assumption that most participants behave honestly. More adaptive approaches introduce trust-aware dynamic weighting, where client contributions are scaled based on their historical reliability, anomaly scores, or similarity to consensus trends. However, applying these strategies to Power CPSs introduces unique challenges [87]. First, the non-IID nature of grid data—influenced by regional demand profiles, DER variability, and market operations—can cause legitimate updates to appear as outliers, thereby reducing accuracy if overly filtered. Second, the assumption of an honest majority may not hold in sparsely connected or cross-organizational settings where client populations are small or loosely monitored. Third, current schemes often lack situational awareness, ignoring contextual factors such as DER criticality or grid emergency states, which are vital for ensuring reliability in safety-critical applications. Hence, robust aggregation in Power CPSs demands more than statistical filtering—it requires integration with domain knowledge and context-sensitive trust calibration [88–90].

4.2. Differential Privacy and Gradient Masking

Differential Privacy (DP) is a cornerstone technique in FL for safeguarding sensitive client data during model updates [91–93]. By injecting mathematically bounded noise into shared gradients or parameters, DP guarantees that an individual client's data contribution remains indistinguishable within a defined privacy budget, typically denoted as (*ε*,).

By clipping each local gradient and injecting calibrated Gaussian noise, DP ensures that the contribution of any individual client remains indistinguishable, even in adversarial settings. This mechanism is formally expressed as:

$$\tilde{\mathbf{g}}_i = clip(\mathbf{g}_i; C) + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I}) \tag{4}$$

where $\mathbf{g}_i$ is the local gradient of client i, C is the clipping threshold, and σ controls the privacy–utility trade-off through noise variance.

In the context of Power CPSs, local DP mechanisms perturb gradients at the client level before transmission, offering stronger privacy at the cost of greater noise [94,95]. Conversely, global DP strategies apply noise after aggregation, reducing the distortion but requiring trust in the aggregator. Gradient masking complements these approaches by obfuscating update directions, further impeding data reconstruction attempts [96]. Despite their theoretical robustness, these methods face substantial limitations in power systems. First, the privacy–utility trade-off is particularly acute in real-time applications—such as short-term forecasting or frequency regulation—where even minor performance degradation may have cascading operational impacts [97]. Second, noise-induced instability can impair convergence and control fidelity, especially in volatile environments with high DER penetration, such as solar PV-dominated microgrids [98]. Third, DP mechanisms treat all features equally, lacking semantic awareness of grid variables—introducing noise to frequency or protection-related signals could disproportionately compromise system reliability [99,100]. As such, deploying DP in Power CPS demands task-specific calibration, utility-aware noise shaping, and integration with domain constraints to ensure both privacy and operational viability.

4.3. Fairness and Heterogeneity in Federated Learning

Beyond privacy and robustness, another critical dimension of trustworthy federated learning in Power CPSs is fairness. In heterogeneous power networks, clients such as regional substations, microgrids, and EV clusters may have vastly different data distributions and computational capabilities. If fairness is not explicitly enforced, global models can become biased toward data-rich or dominant clients, leaving others underrepresented and degrading overall reliability. To address this issue, fairness can be formulated as a variance-regularized optimization problem:

$$\min_{\mathbf{w}} \sum_{i} p_i F_i(\mathbf{w}) + \lambda Var\left(\{F_i(\mathbf{w})\}\right) \tag{5}$$

where the first term ensures standard empirical risk minimization across clients, while the second penalizes disparities in local losses. Here, λ serves as a trade-off coefficient that balances global accuracy with cross-client equity.

While fairness seeks to balance performance across all clients, it is equally important to recognize that heterogeneity in Power CPSs may require personalized solutions. Different regions or assets—such as renewable plants, substations, and EV fleets—often operate under distinct load profiles, environmental conditions, and data distributions. A single global model may thus fail to capture these localized patterns. To accommodate such diversity, personalization is introduced by coupling local objectives with a shared global reference model:

$$\min_{\{\mathbf{w}_i\},\mathbf{w}_g} \sum_{i} p_i \left[F_i(\mathbf{w}_i) + \frac{\lambda}{2} \| \mathbf{w}_i - \mathbf{w}_g \|^2 \right]. \tag{6}$$

Here, $\mathbf{w}_i$ denotes the personalized model parameters for client i, while $\mathbf{w}_g$ represents the shared global model. The regularization term ensures that local models remain anchored to the global knowledge base, while still allowing them to adapt to their specific data distributions.

4.4. Personalization for Heterogeneous Clients

While fairness regularization seeks to balance performance across clients, federated learning in Power CPSs must also account for the fact that different clients may require individualized models due to highly heterogeneous environments. For example, renewable plants operate under different weather conditions, while substations or EV clusters experience distinct load patterns. A single global model may therefore fail to capture these diverse local characteristics. To address this limitation, personalization is introduced by coupling local objectives with a global reference model:

$$\min_{\{\mathbf{w}_i\},\mathbf{w}_g} \sum_{i} p_i \left[F_i(\mathbf{w}_i) + \frac{\lambda}{2} \| \mathbf{w}_i - \mathbf{w}_g \|^2 \right]. \tag{7}$$

In this formulation, $\mathbf{w}_i$ denotes the personalized model parameters of client i, while $\mathbf{w}_g$ represents the shared global model. The term $F_i(\mathbf{w}_i)$ is the local objective function of client i, and the quadratic penalty term ensures that local models remain sufficiently aligned with the global model. The parameter λ controls the trade-off between local adaptation and global consistency, while p_i denotes the normalized weight of each client.

4.5. Byzantine-Resilient Algorithms

Byzantine-resilient FL algorithms are explicitly designed to withstand arbitrary, untrustworthy, or malicious behavior from a fraction of participating clients—referred to as Byzantine faults [101–103]. In the context of Power CPSs, such resilience is vital due to the potential for compromised edge nodes or adversarial insiders. Algorithms like Bulyan employ a two-tier strategy combining robust client selection and trimmed aggregation to ensure that malicious updates are excluded from the final model. Zeno++ takes a validation-based approach, filtering updates based on their contribution to decreasing a surrogate loss function, thereby suppressing harmful gradients. Robust Stochastic Aggregation (RSA) further mitigates poisoning by introducing a regularization term that penalizes large deviations from the global model, promoting coherence across distributed updates [104,105]. Despite their theoretical robustness, these methods face several practical limitations in power system environments. First, their computational and communication overhead can be prohibitive for low-power or bandwidth-constrained edge devices such as smart meters, PMUs, or DER controllers [106–108]. Second, some schemes may over-filter non-IID but legitimate updates, a common occurrence in power systems due to geographic, temporal, or operational diversity, thereby impairing accuracy in forecasting or control [109]. Third, many Byzantine-resilient strategies rely on synchronous update rounds, which conflicts with the inherently asynchronous and event-driven nature of distributed grid assets [110,111]. Addressing these limitations calls for lightweight, adaptive, and grid-aware Byzantine mitigation strategies tailored to the unique constraints and dynamics of Power CPSs.

4.6. Adversarial Training and Certification

Adversarial training is a proactive defense technique that enhances the robustness of federated models by exposing them to deliberately crafted perturbations during the training process [112,113]. In the context of FL for Power CPSs, this method helps prepare models against worst-case perturbations or adversarial manipulations that could arise from compromised clients or noisy sensor environments. Typical implementations include the Fast Gradient Sign Method (FGSM) [114,115] and Projected Gradient Descent (PGD) [116], which are used to generate adversarial examples at the client level, effectively simulating malicious behavior during model updates. More recent developments also explore certified defenses, which aim to offer formal robustness guarantees under norm-bounded perturbations, ensuring that model predictions remain unchanged within a specific threat radius [117].

However, the practical application of adversarial training in Power CPSs faces multiple challenges. First, the lack of interpretability in adversarial perturbations limits operator trust and makes it difficult to integrate such defenses into human-in-the-loop control workflows or regulatory auditing pipelines [118]. Second, most certified defense frameworks are built upon fixed and simplistic threat assumptions, rendering them ineffective against the adaptive, co-evolving attack patterns characteristic of real-world power systems—where attackers may exploit time-varying conditions or cross-domain data fusion (e.g., weather plus grid data) [119]. Third, adversarial training inherently requires extended convergence time and additional computational resources, which contradicts the real-time constraints of critical tasks such as frequency regulation, fault detection, and DER coordination [120]. To make adversarial defenses viable in Power CPSs, future work must prioritize lightweight, explainable, and dynamically adaptable training frameworks that align with both physical grid constraints and operational trust requirements.

4.7. Blockchain and Secure Multiparty Computation

Blockchain and Secure Multiparty Computation (SMPC) have emerged as promising techniques for enhancing trust in FL by providing verifiable, tamper-resistant computation and preserving confidentiality of model updates [121–123]. In the context of Power CPSs, these technologies are often employed to address critical concerns around update integrity, auditability, and secure collaboration among heterogeneous entities. Blockchain frameworks, particularly permissioned blockchains, can be used to log FL-related metadata such as model versions, update timestamps, and authenticated client identities, forming a transparent audit trail that supports accountability and regulatory compliance [124]. Simultaneously, SMPC protocols enable multiple clients to collaboratively train models without ever revealing their raw gradients or local data, using cryptographic primitives such as secret sharing or homomorphic encryption to ensure privacy-preserving aggregation [125,126].

Despite these strengths, integrating blockchain and SMPC into Power CPS FL workflows faces substantial practical barriers. Firstly, latency overheads introduced by blockchain consensus algorithms or SMPC encryption schemes can delay model convergence and compromise real-time control or dispatch decisions, which are often time-critical in power systems. Secondly, high computational and energy costs make these methods ill-suited for deployment on resource-constrained edge devices, such as smart inverters or smart meters, which are prevalent in DER settings [127]. Finally, while permissioned blockchains improve efficiency compared to public chains, they may introduce a single point of failure or trust bottleneck—particularly if the validating authority is compromised or experiences downtime. These limitations suggest that while blockchain and SMPC offer strong theoretical guarantees, their successful application in Power CPS requires lightweight, grid-aware variants and hybrid architectures that can balance trust enforcement with operational responsiveness [128,129].

4.8. Summary Table of Defense Strategies

To address the multifaceted security and reliability risks in federated deployments, Table 5 presents key defense strategies and critically evaluates their strengths and specific limitations within the operational constraints of Power CPSs.

Table 5. Comparative Analysis of Defense Strategies for FL in Power CPSs.

Defense Strategy	Strengths	Limitations in Power CPSs
Robust Aggregation	Mitigates outlier updates	Poor handling of non-IID grid conditions
Differential Privacy	Prevents data leakage	Accuracy–privacy trade-off; operational instability
Byzantine-Resilient FL	Tolerates arbitrary client behavior	High cost; convergence degradation in asynchronous settings
Adversarial Training	Prepares models for worst-case perturbations	Training time increase; interpretability challenges
Blockchain / SMPC	Tamper-proof, privacy-preserving updates	Latency, scalability, and energy inefficiency

While Tables 4 and 5 have separately outlined major threats and corresponding defense strategies in federated learning, a direct mapping is necessary to reveal the extent to which current solutions address the unique requirements of Power CPS. Table 6 provides an integrated perspective by linking representative threats with their defense mechanisms, evaluating their effectiveness, and highlighting remaining gaps. This comparative view not only clarifies the coverage of existing approaches but also exposes unresolved challenges that should drive future research efforts.

Table 6. Mapping of Threats, Defense, and Remaining Gaps in FL for Power CPS.

Threat Type	Typical Manifestation in Power CPS	Defense Mechanisms	Effectiveness	Remaining Gaps / Limitations
Data Poisoning	Manipulated PMU/AMI data	Robust aggregation (Krum, Trimmed	Mitigates simple outliers	High-dimensional attacks remain

Threat Type	Typical Manifestation in Power CPS	Defense Mechanisms	Effectiveness	Remaining Gaps / Limitations
	injection to bias model updates	Mean, Median), Outlier detection		stealthy; real-time detection cost is high
Backdoor Attacks	Trigger-based malicious model behavior in load forecasting or intrusion detection	Differential privacy, Anomaly-based gradient filtering, Blockchain audit	Blocks simple triggers	Adaptive backdoors bypass filters; limited explainability of detection
Sybil/Byzantine Attacks	Malicious clients flooding FL server with corrupted gradients	Byzantine-resilient aggregation (Bulyan, Multi-Krum), Reputation mechanisms	Tolerates up to 30–40% adversaries	Degrades with non-IID data; communication overhead remains unsolved
False Data Injection (FDI)	Compromised measurements in state estimation or frequency control	Cross-validation with Digital Twins, Secure multiparty computation	Detects abnormal patterns	Hard to scale for large CPS; lacks theoretical robustness guarantees
Model Inference/Privacy Leakage	Membership inference, gradient inversion against AMI datasets	Differential Privacy (DP), Homomorphic Encryption, Secure Enclaves	Provides provable privacy	Accuracy loss (DP noise), high computation cost (HE), deployment challenges

As shown in Table 6, although several mechanisms—such as robust aggregation, differential privacy, and Byzantine-resilient learning—mitigate certain attack vectors, critical gaps remain in scalability, explainability, and adaptability to non-IID and real-time conditions of Power CPS. These limitations underscore the necessity for hybrid frameworks that combine algorithmic robustness with system-level resilience and policy-driven governance.

4.9. Key Takeaways

The defense strategies surveyed in this section reveal a fundamental mismatch between general-purpose FL robustness methods and the unique demands of Power CPSs. Most existing techniques—ranging from robust aggregation to adversarial training and secure computation—were primarily designed for generic, cloud-based FL environments [130]. As such, they often neglect the real-time operational constraints, system heterogeneity, and safety-critical nature inherent to power systems [131]. For instance, defenses that assume synchronized client participation or tolerate slow convergence may be unacceptable in grid control applications, where even minor delays can destabilize voltage, frequency, or load balance.

To close this gap, future research must move toward domain-adaptive defense designs that are explicitly tailored to grid contexts. These designs should account for heterogeneous DER behaviors, temporal volatility (e.g., solar ramp rates), and communication limitations, while also supporting human interpretability and intervention [132,133]. Importantly, the effectiveness of defenses in Power CPS should be measured not solely by improvements in model accuracy or convergence, but by their impact on physical system safety and resilience—such as the ability to maintain frequency stability, avoid blackouts, or prevent cascading failures under attack. In this light, FL defense becomes not just a machine learning challenge, but a cross-disciplinary effort involving control theory, grid operations, and adversarial modeling [134–136].

In the next section, we turn to emerging architectures and paradigms that aim to embed trustworthiness directly into FL frameworks for Power CPSs. Yet, in the absence of unified benchmarks, standardized evaluation metrics, and realistic testbeds spanning both cyber and physical layers, research progress risks remaining fragmented and hard to translate into practical deployment. Addressing this gap requires community-driven efforts that integrate algorithmic advances with system-level validation under real-world grid conditions.

5. Architectures and Design Paradigms for Trustworthy FL in Power CPSs

As discussed in Section 4, the limitations of conventional FL defense mechanisms underscore the necessity for architectural innovations tailored to the unique constraints of Power CPSs. This section explores emerging architectural designs and paradigms that embed trustworthiness, scalability, and resilience into FL workflows—from foundational trust mechanisms to human-in-the-loop and simulation-integrated systems.

5.1. Zero-Trust Federated Learning Frameworks

The Zero-Trust Federated Learning (ZTFL) paradigm introduces a transformative shift in how participants interact within FL ecosystems, especially in safety-critical environments like Power CPSs [137,138]. Departing from traditional FL assumptions—where clients and aggregators are implicitly trusted—ZTFL adopts the "never trust, always verify" doctrine, assuming that any participant can be compromised and must be continuously validated [139].

Key architectural features include:

Identity Verification: Before any training session, clients must present verifiable cryptographic credentials or participate in decentralized identity frameworks (e.g., blockchain-based identity proofs). This ensures that only **authorized grid entities**, such as substations or DER controllers, participate in model updates.

Dynamic Trust Scoring: Trust is no longer binary or static. ZTFL dynamically adjusts trust levels based on behavioral analytics—such as update consistency, anomaly detection, and participation history—often integrated with **reputation systems or anomaly scoring algorithms**. These scores directly influence model aggregation weights or access privileges [140].

Policy Enforcement Points (PEPs): Instead of relying on historical reputation or organizational affiliation, PEPs act as **real-time gatekeepers**, enforcing fine-grained rules on data usage, update frequency, and contribution thresholds for each training round. This ensures **per-session accountability**, preventing misuse even from previously trusted entities [141,142].

Relevance to Power CPSs:

In the power grid domain, ZTFL holds particular significance due to the **increasingly decentralized and multi-stakeholder nature** of modern energy systems [143–145]. For instance:

It **prevents insider threats**, where compromised DER vendors, third-party service providers, or even internal operator terminals may inject malicious updates under assumed trust.

It enables **fine-grained access control**, crucial when coordinating learning across mixed criticality assets (e.g., between high-voltage substations and low-impact residential microgrids).

Through cryptographic audit trails and trust enforcement, ZTFL also supports **regulatory compliance** and **post-incident forensics**, aligning with critical infrastructure governance policies such as NERC CIP or IEC 62351.

Overall, Zero-Trust FL introduces a **proactive trust infrastructure** that aligns with both the **cybersecurity requirements** and **operational demands** of modern power systems, making it a foundational pillar in the evolution of trustworthy AI-enabled grid intelligence.

5.2. Personalized Federated Learning (PFL)

PFL challenges the foundational assumption of conventional FL—that a single global model is optimal for all participants. This assumption often breaks down in Power CPSs, where client nodes such as substations, microgrids, EV charging stations, or DER controllers operate under vastly different physical conditions, grid topologies, and control objectives [146–148]. PFL seeks to balance shared learning with local specialization, enabling each client to benefit from global knowledge while maintaining model fidelity to its own operational context [149].

Key methodological approaches in PFL include meta-learning, clustering, and multi-task optimization. Meta-learning methods such as FedMeta or Reptile aim to learn a global initialization that can be quickly adapted to each client's local data through fine-tuning, which is especially valuable in power systems requiring rapid retraining under dynamic conditions like weather changes or shifting load profiles [150,151]. Clustered FL organizes clients into groups based on data similarity, operational roles, or geographic proximity—such as microgrids in the same climate zone or feeders with comparable load characteristics—enabling localized yet collaborative learning [152]. Meanwhile, multi-task FL acknowledges that clients may face different but related objectives, for instance, one optimizing renewable generation forecasting while another ensures voltage control; such models share generalizable features while tailoring solutions to task-specific goals [154,155].

5.3. Explainable Federated Learning

XFL aims to embed transparency into the federated learning process, ensuring that both training dynamics and model outputs are understandable to human operators, auditors, and regulators [156]. Key techniques include model attribution methods such as SHAP, LIME, or gradient-based explanations, which help quantify the influence of specific features on predictions, thereby demystifying black-box behaviors. Audit trails and decision logs capture the sequence of model updates and the rationale behind aggregation outcomes, supporting traceability and accountability. Additionally, operator-facing dashboards deliver intuitive visualizations of model behavior, anomalies, and confidence levels, empowering grid operators to make informed and trustworthy decisions in real time [157,158].

In the context of Power CPS, explainable federated learning enhances human-AI collaboration by enabling grid operators to investigate, interpret, or override suspicious outcomes, especially during critical operations. It supports regulatory compliance with emerging AI governance frameworks such as the EU AI Act, where transparency and accountability are mandated. Furthermore, explainability plays a vital role in incident forensics, allowing post-event analysis to trace adversarial manipulations or system faults back to specific model decisions or client contributions [159].

5.4. Digital Twin-Augmented FL Architectures

Digital twin–augmented federated learning integrates high-fidelity simulations of power system components into the FL training and validation pipeline, enabling risk-free experimentation and enhanced model generalization [160–162]. Through sim-to-real transfer, digital twins generate synthetic data under rare or hazardous conditions—such as line faults or cyber intrusions—to strengthen FL robustness. Real-time synchronization ensures that these twin models dynamically update based on live telemetry, supporting continuous model refinement. Moreover, cross-scale co-simulation bridges SCADA-level grid behavior with edge-level FL agents, ensuring consistency between system-wide operations and local control intelligence in Power CPS environments [163].

In the context of Power CPS, digital twin–augmented federated learning offers several key advantages: it enables risk-free pre-training of FL models prior to deployment, ensuring safety in critical grid applications; it supports resilience testing by simulating synthetic attacks or control anomalies, helping identify vulnerabilities before real-world impact; and it facilitates coordinated defense simulation, allowing joint training of distributed agents such as EV charging nodes, WAMS units, and substation controllers within a unified and realistic virtual environment [164–166].

5.5. Human-in-the-Loop Federated Defense

Human-in-the-loop FL integrates expert oversight into automated decision loops, ensuring safety and adaptability without compromising scalability. This approach enables active querying, where FL models identify uncertain or high-impact predictions and defer to human judgment; supports collaborative labeling, leveraging operator insights to refine datasets and improve model quality over time; and allows dynamic role adjustment, modulating the degree of human involvement based on grid stress levels or detected cyber threats—thus promoting resilient and accountable AI-driven operations in Power CPS [167,168].

In Power CPSs, human-in-the-loop federated learning plays a critical role in preventing automation bias in high-stakes control scenarios by enabling operators to scrutinize and override model decisions when necessary. It also reduces false positives in intrusion detection by blending statistical outputs with human intuition and domain expertise. Over time, this collaborative interaction supports trust calibration, allowing operators to gradually build confidence in AI-driven processes, thereby fostering safer, more transparent, and more resilient grid operations [169].

5.6. Comparative Table: Architectural Paradigms

To strengthen trust across the lifecycle of federated learning in critical infrastructure, Table 7 outlines emerging architectural paradigms that enhance explainability, adaptability, and operational safety in Power CPS applications.

Table 7. Emerging Architectures Enhancing Trust in Federated Learning for Power CPS.

Architecture	Trust Contributions	Power CPS Benefits
Zero-Trust FL	Mitigates insider threats; enforces strict access controls	Secure multi-organization collaboration
Personalized FL	Aligns with local heterogeneity; resists model drift	Adaptive control and forecasting across distributed assets
Explainable FL	Builds operator trust; supports audits	Human-centric operations and incident response
Digital Twin-Augmented FL	Risk-free training and validation	Safe testing under rare/extreme scenarios
Human-in-the-Loop FL	Merges AI automation with expert oversight	Reliable decision-making under uncertainty

5.7. Summary and Design Principles

Across the surveyed emerging architectures, several unifying principles crystallize for the effective design of trustworthy federated learning systems in Power CPSs. First, modular trust anchors are essential—verifications of identity, behavior, and outcomes should be independently enforced across system layers to avoid single points of failure. Second, hybrid autonomy is key, requiring a careful and dynamic balance between automated FL processes and human operator oversight, particularly in safety-critical grid applications. Third, simulation-real coupling must be embedded into continuous FL workflows, allowing digital twins to simulate edge cases and validate models prior to deployment. Lastly, interpretability-by-design mandates that explainability mechanisms be integrated into the architecture from the outset, rather than appended as afterthoughts, to support human-AI collaboration and regulatory accountability [170–172]. In the next section, we move from architectural innovations to practical **use cases**, detailing how these paradigms are applied in real-world power systems and what lessons have emerged.

6. Real-World Applications and Lessons Learned

While the prior sections focused on conceptual foundations and architectural innovations, this section grounds the discussion in real-world deployments and operational insights. By reviewing representative applications of trustworthy FL in CPSs, we extract key lessons on feasibility, effectiveness, and open bottlenecks.

6.1. Privacy-Preserving Load and Renewable Forecasting

Accurate forecasting of load demand and renewable generation, particularly wind and photovoltaic power, is fundamental for reliable scheduling and secure operation of modern power grids [173,174]. Grid operators and DER owners collaboratively train time-series forecasting models—such as for wind power generation or residential load patterns—without exchanging raw consumption or generation data. This is typically achieved using federated learning frameworks based on LSTM or Transformer architectures, enhanced with differential privacy and PFL techniques. A representative deployment comes from a multi-utility pilot within the EU Horizon 2020 edge computing project, where participants maintained approximately 95% forecasting accuracy while preserving data locality. This not only ensured operational effectiveness but also alleviated compliance challenges with GDPR and other privacy regulations [175,176].

Personalization strategies that account for regional heterogeneity—such as varying climatic conditions or load profiles [177,178]—substantially enhance model utility and local adaptability. Additionally, implementing communication sparsity techniques, such as event-triggered updates or periodic aggregation, effectively reduces communication overhead while preserving forecasting accuracy, making FL more feasible for real-world Power CPS deployments.

6.2. Collaborative Intrusion Detection for Power CPSs

SCADA systems, substations, and microgrids collaboratively detect cyber-attacks—such as FDIAs and Distributed Denial-of-Service (DDoS)—by training anomaly detection models in a federated manner. This approach leverages federated autoencoders and graph neural networks (GNNs) enhanced with adversarial training defenses to counter model poisoning [179,180]. A simulated cross-regional testbed using WADI and ICS-AD datasets, involving multiple data owners, demonstrated a 23% improvement in attack detection rates compared to isolated models and showed strong resilience against poisoning attacks.

Trust-aware aggregation methods such as Krum and Median are crucial for mitigating the impact of malicious client contributions in federated training. Additionally, incorporating model explainability techniques—such as SHAP values or saliency maps—not only enhances operator confidence but also accelerates incident response by making anomalies and model decisions more transparent [181,182].

6.3. FL for EV Charging and V2G Coordination

EV aggregators coordinate charging profiles and provide grid support services—such as frequency regulation—through federated learning without disclosing user-specific behaviors. The approach leverages reinforcement learning–enhanced FL (FedRL) with personalized policies tailored to each fleet. In a regional deployment, EV clusters participated in demand response using a distributed learning controller [183–185]. This setup led to an 18% improvement in peak shaving and a 12% reduction in operational costs, all while preserving user privacy.

Key lessons from this use case highlight that trust calibration—through transparent operations and well-designed participation incentives—is crucial for sustaining aggregator engagement, while personalized federated learning policies are necessary to prevent negative transfer across heterogeneous EV clusters, such as those in urban versus rural settings [186,187].

6.4. Federated Voltage and Frequency Control in Microgrids

Ensuring stable voltage and frequency in microgrids has emerged as a focal research area, driven by increasing renewable integration and decentralized operation [188–190]. Microgrid agents (e.g., inverters, storage systems) collaboratively learn distributed control policies using actor-critic federated learning with simulation pre-training and asynchronous real-time updates, achieving a significant reduction in costs and CO2 emissions [191,192].

Bridging the simulation-to-reality gap demands integration of digital twins and robust pretraining, while incorporating local grid codes into FL loss functions is essential to ensure regulatory compliance and stable control behavior in real-world deployments.

6.5. Federated State Estimation in Power Grids

A representative application of trustworthy federated learning in Power CPSs is state estimation (SE), which plays a central role in monitoring and control. In practice, measurement data from SCADA systems or PMUs are often geographically distributed and sensitive, making them well-suited for federated settings. By enabling collaborative training across substations or regions without exposing raw data, TFL can provide accurate and privacy-preserving estimates of system states. The state estimation problem can be formulated as:

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \epsilon, \quad \min_{\mathbf{w}} \sum p_i \mathbb{E}\left[\| \hat{\mathbf{x}}_{\mathbf{w}}(\mathbf{z}) - \mathbf{x} \|_2^2 \right]. \tag{8}$$

Here, $\mathbf{z}$ denotes the measurement vector, $\mathbf{H}$ is the measurement matrix, and $\mathbf{x}$ represents the true system state to be estimated. The term ϵ accounts for measurement noise, while $\hat{\mathbf{x}}_{\mathbf{w}}(\mathbf{z})$ is the state estimator parameterized by the federated model $\mathbf{w}$. The client weights p_i ensure that contributions from different regions or assets are properly balanced. This formulation highlights how TFL can be concretely embedded into CPS operations, bridging abstract trust dimensions with real-world power system applications.

State estimation underpins monitoring and control in power grids but is highly vulnerable to False Data Injection (FDI) attacks [193–195]. Traditional centralized SE aggregates raw SCADA and PMU data at control centers, raising privacy and scalability concerns [196–199]. FL enables collaborative SE across substations or regional operators without sharing raw data, while TFL strengthens resilience against compromised nodes through Byzantine-resilient aggregation, explainability, and digital twin–based validation.

Recent studies show that federated SE can achieve accuracy comparable to centralized methods while preserving data locality and reducing communication overhead [200,201]. Key challenges remain in handling non-IID measurement distributions, synchronization under delays, and embedding observability constraints into FL objectives, but TFL-based SE offers a path toward secure and privacy-preserving grid monitoring [202,203].

6.6. Substation Automation and Federated Lifelong Learning

Substation intelligent electronic devices adaptively enhance fault classification and control actions over time using federated lifelong learning (FedLL), which incorporates elastic weight consolidation and replay buffers. A field-tested deployment in industrial substations with legacy devices achieved sustained fault detection accuracy exceeding 90% over 12 months, while effectively avoiding catastrophic forgetting [204–206].

Substation intelligent electronic devices adaptively enhance fault classification and control actions over time using FedLL, which incorporates elastic weight consolidation and replay buffers. A field-tested deployment in industrial substations with legacy devices achieved sustained fault detection accuracy exceeding 90% over 12 months, while effectively avoiding catastrophic forgetting [207–210]. These results demonstrate that lifelong learning enables models to continuously adapt to aging equipment and evolving fault signatures, while also highlighting the necessity of secure update channels and robust device authentication mechanisms in low-trust substation environments to ensure the integrity and reliability of deployed models [211–213].

6.7. Summary Table: Use Case Comparison

To illustrate how TFL can be applied across critical power system tasks, Table 8 presents representative use cases, highlighting the specific FL techniques, trust-enhancing features, and their operational impacts in Power CPSs. These cases cover a broad spectrum ranging from forecasting and intrusion detection to real-time control and fault management, thereby reflecting the versatility of TFL in addressing both data-driven prediction and safety-critical operation. By juxtaposing different application domains, the comparison emphasizes how trust mechanisms such as privacy preservation, robustness, explainability, and continual learning can be selectively integrated to meet domain-specific requirements. Moreover, the table underscores that while individual tasks benefit from targeted TFL adaptations, the overarching trend points toward building a cohesive, trustworthy learning ecosystem that balances accuracy, resilience, and compliance in modern power grids.

Table 8. Representative Applications of TFL in Power CPS.

Application	FL Technique	Trust Feature	Key Impact
Load/Renewable Forecasting	PFL + DP	Privacy-preserving local models	High accuracy + GDPR compliance
CPS Intrusion Detection	Robust FL + Explainability	Model integrity + interpretability	Higher attack detection + operator trust
EV Charging and V2G	FedRL + personalization	Human-in-the-loop scheduling	Demand response with privacy guarantees
Voltage/Frequency Control	Actor-Critic FL	Digital twin–based validation	Stability with renewable uncertainty
Substation Fault Management	Lifelong FL	Continual adaptation + device authenticate	Resilience under long-term equipment drift

While Table 8 provides an overview of representative application cases, it does not explicitly connect performance outcomes with trust-enhancing features. Table 9 addresses this gap by linking key application domains with specific FL techniques, performance metrics, and the trust dimensions they strengthen. This mapping highlights both the potential benefits and the practical limitations of applying federated learning in diverse Power CPS tasks.

Table 9. Mapping of Application Domains, Performance Metrics, and Trust-Enhancing Features of Federated Learning in Power CPS.

Application Domain	FL Technique / Variant	Key Performance Metrics	Trust-Enhancing Features	Limitations / Open Issues
Load Forecasting	Horizontal FL, Personalized FL	MAE, RMSE, MAPE	Privacy preservation, Non-IID adaptation	Limited explainability; vulnerable to poisoning under heterogeneous data
Intrusion Detection	Robust FL, Byzantine-resilient FL	Accuracy, F1-score, Detection latency	Robustness, Resilience	High false positive rate under adaptive attacks; heavy communication overhead

Application Domain	FL Technique / Variant	Key Performance Metrics	Trust-Enhancing Features	Limitations / Open Issues
EV Coordination / V2G	Vertical FL, Cross-silo FL	Charging efficiency, Grid stability index	Fairness, Accountability	Data alignment challenges; privacy leakage risk during coordination
Microgrid Energy Management	Hybrid FL with Digital Twin	Energy cost reduction, Frequency deviation	Validation, Resilience, Human-in-the-loop	Digital twin drift; scalability to multi-microgrids remains limited
State Estimation	Zero-Trust FL, Secure Multi-Party Computation	Estimation accuracy, Latency, Robustness index	Authentication, Privacy protection	High computational cost; integration with legacy SCADA not seamless
Substation Automation	Explainable FL, Edge-enabled FL	Reliability index, Fault detection rate	Transparency, Accountability	Explainability–accuracy trade-off; deployment constraints in resource-limited devices

As seen in Table 9, different application domains emphasize distinct trust features—privacy in forecasting, robustness in intrusion detection, explainability in substation automation—yet none achieve a comprehensive trust guarantee. This fragmentation reveals the need for unified frameworks that balance accuracy, efficiency, and multi-dimensional trust in real-world power systems.

6.7. Insights and Cross-Cutting Observations

Insights from both real-world and simulated deployments reveal several recurring themes that underpin the successful application of Trustworthy Federated Learning in Power CPSs [214–216]. First, context-aware trust mechanisms—such as personalized federated learning and explainability—are essential not only for achieving high technical performance but also for gaining user acceptance and operational trust. Second, resilience testing through digital twins and hybrid architectures enables risk-free validation under rare or extreme conditions, thus enhancing the robustness of deployments. Third, providing trust incentives—including transparency, privacy guarantees, and local control over models—plays a pivotal role in sustaining long-term stakeholder engagement, especially in collaborative ecosystems [217]. Lastly, in multi-party settings involving diverse actors such as regional utilities and third-party vendors, adopting a zero-trust assumption is imperative to mitigate insider threats and enforce rigorous verification throughout the learning lifecycle [218–220].

7. Research Challenges and Future Directions

To operationalize trustworthy FL in CPSs, researchers and practitioners must address a range of open challenges spanning theory, system design, human factors, and policy. This section categorizes these challenges and proposes future research directions across four key dimensions: technical robustness, human-centered trust, system integration, and governance frameworks.

7.1. Technical Robustness and Adversarial Resilience

A core challenge in deploying federated learning within Power CPS lies in ensuring robustness against a wide spectrum of adversarial threats [221,222]. Byzantine vulnerabilities continue to pose significant risks in multi-party federated settings, where malicious participants may inject poisoned updates or manipulate aggregation processes to compromise global model integrity. Despite the use of differential privacy, gradient leakage attacks remain a concern—particularly in time-series applications—where adversaries can partially reconstruct sensitive operational data from shared updates. Moreover, adaptive adversaries can dynamically evolve their attack strategies over time, exploiting phenomena such as model drift and overfitting to bypass static defenses, further undermining the reliability and security of FL-based systems in critical power infrastructures.

To overcome these threats, future efforts should focus on designing provably robust aggregation mechanisms that can maintain model integrity even when over 30% of participating clients behave maliciously—a critical threshold for resilience in untrusted environments [223]. Furthermore, adopting adversarial co-training paradigms, where defensive strategies evolve in tandem with adaptive attack behaviors within a game-theoretic framework, offers a promising pathway to preemptively mitigate sophisticated threat patterns. Equally important is the advancement of differential privacy techniques that are not only mathematically rigorous but also context-aware—specifically tailored to the spatiotemporal dependencies and graph-structured data inherent in energy systems—to ensure both privacy protection and task relevance in federated Power CPS applications.

7.2. Human-Centered Trust and Explainability

Building trust in federated learning systems for Power CPSs extends beyond technical robustness—it critically depends on human interpretability and operational alignment. One major challenge lies in the inherent opacity of black-box FL models, particularly those trained across decentralized and non-transparent participants, which diminishes operator confidence and limits practical deployment in safety-critical environments. Moreover, discrepancies between model-generated recommendations and human expert judgments can lead to decision conflicts, especially in protection systems or emergency control scenarios where accountability and real-time responsiveness are essential. Compounding these issues is the absence of standardized trust metrics tailored to the unique demands of power grid operations, making it difficult to quantify or benchmark the trustworthiness of FL deployments across heterogeneous stakeholders [224].

To address these limitations, future research should focus on designing human-in-the-loop federated learning pipelines that integrate visual analytics, interactive model exploration, and operator override mechanisms, thereby embedding human judgment into the FL decision-making loop. Establishing formalized trust scoring systems—grounded in explainability, calibration accuracy, update integrity, and the degree of user control—will also be essential for quantifying and communicating the trustworthiness of FL models in operational contexts. In parallel, advancing sociotechnical interface research is needed to bridge the gap between human cognitive models and AI-driven reasoning, ensuring that FL systems not only perform accurately but also align with the expectations, workflows, and mental models of power system operators [225].

7.3. Scalable System Integration and Digital Twin Coupling

Despite its promise, the large-scale deployment of federated learning in Power CPS is constrained by several integration challenges. First, embedding FL capabilities into edge devices—such as remote sensors, substations, and DER controllers—remains difficult due to limited computational resources, energy constraints, and intermittent connectivity, all of which undermine training stability and model synchronization. Second, the absence of standardized middleware for orchestrating FL workflows across heterogeneous infrastructures—including SCADA systems, Advanced Metering Infrastructure (AMI), and Distributed Energy Resource Management Systems (DERMS)—creates interoperability bottlenecks and limits scalability. Finally, validating FL models under realistic operating conditions is hampered by the persistent sim-to-real gap and the lack of co-simulation environments that can replicate the complex dynamics of physical power systems in conjunction with federated AI behavior [226].

To overcome these barriers, future research should prioritize the development of energy-efficient federated learning protocols tailored to edge environments—leveraging techniques such as quantized model updates, asynchronous training, and sparse representations to reduce computational and communication overhead. In parallel, the creation of digital twin–augmented FL platforms will be vital for enabling closed-loop validation, “what-if” scenario analysis, and dynamic co-simulation across both cyber and physical layers of power systems. Furthermore, advancing FL-as-a-service infrastructures that support cross-vendor compatibility and modular integration with existing utility platforms will accelerate scalable adoption and facilitate seamless orchestration of federated intelligence across diverse energy assets and stakeholders [227].

7.4. Policy, Regulation, and Multi-Stakeholder Governance

The deployment of federated learning in Power CPSs also confronts significant governance and regulatory challenges that go beyond technical implementation. Current cybersecurity and data protection frameworks—such as NERC CIP in North America and GDPR in Europe—provide limited guidance on the roles, responsibilities, and liabilities associated with decentralized AI systems, leaving critical gaps in compliance and risk attribution. Moreover, trust asymmetry among stakeholders—particularly between private technology vendors, public utilities, and regulatory bodies—undermines collaborative data federation efforts, often stalling progress due to concerns over data misuse, competitive exposure, or unclear governance authority. Compounding these issues is the lack of well-defined auditing and accountability mechanisms for FL-driven decisions, especially in high-stakes scenarios such as system blackouts or operational misjudgments, where tracing causality across distributed and opaque model updates remains a major hurdle [228].

Addressing these governance challenges requires the establishment of auditable federated learning compliance frameworks that align not only with existing cybersecurity and data protection regulations but also with emerging energy justice principles, ensuring equitable access, transparency, and accountability. Promoting multi-tenant governance architectures—with clearly defined participation rules, traceable data provenance, and predefined fallback mechanisms—will be essential to enable trustworthy collaboration across utilities, vendors, regulators, and consumers. Additionally, fostering international cooperation through the development of cross-border federated testbeds and standardized trust certification schemes can help harmonize regulatory approaches, support interoperability, and accelerate the global adoption of trustworthy FL in critical energy infrastructures [229,230].

7.5. Summary Table: Key Gaps and Proposed Directions

To outline the current limitations and research frontiers in deploying TFL in real-world power systems, Table 7 summarizes key challenge categories, existing gaps, and proposed research directions to advance the field.

Table 7. Open Challenges and Future Directions for TFL in Power CPSs.

Category	Key Gaps	Proposed Directions
Adversarial Robustness	Lack of provable defense against poisoning and inference attacks	Game-theoretic defenses, resilient aggregation, and contextual DP
Human-Centric Trust	Poor explainability and operator alignment	Visualized FL interfaces, trust metrics, and override mechanisms
System Integration	Sim-to-real gap; constrained devices; lack of orchestration tools	Digital twin–based validation, FL-as-a-service platforms, efficient edge protocols
Governance and Policy	Ambiguous compliance rules; stakeholder mistrust; no audit trail	Trust governance frameworks, multi-tenant traceability, cross-border standardization

8. Conclusion

This review provides a holistic framework for federated learning (FL) in power cyber-physical systems, emphasizing trust dimensions of security, resilience, and explainability. Unlike prior surveys focused mainly on privacy or generic FL, it systematically maps threats to defenses, links architectures to trust features, and contextualizes them in real-world power system applications.

For researchers, the review highlights the need for hybrid approaches that combine robust aggregation, zero-trust validation, digital twin verification, and human-in-the-loop calibration. For practitioners, it underscores the trade-offs among accuracy, latency, and robustness in applications such as intrusion detection, EV coordination, and microgrid optimization. For policymakers, it calls for governance frameworks and auditing standards to guide responsible adoption. Trustworthy FL must evolve into adaptive frameworks that withstand adversarial dynamics, integrate heterogeneous resources, and align with regulatory and ethical principles. Positioned this way, FL becomes not just an algorithmic tool but a cornerstone for secure, resilient, and explainable intelligence in next-generation power systems. These contributions—multi-dimensional trust framework, architecture–trust–application mapping, and threat–defense–gap synthesis—jointly provide a roadmap for advancing trustworthy FL in Power CPS.

Reference

1. Tariq A, Serhani M A, Sallabi F, et al. Trustworthy federated learning: A survey[J]. arXiv preprint arXiv:2305.11537, 2023.
2. Zhang Y, Zeng D, Luo J, et al. A survey of trustworthy federated learning: Issues, solutions, and challenges[J]. ACM Transactions on Intelligent Systems and Technology, 2024, 15(6): 1-47.
3. Sánchez P M S, Celdrán A H, Xie N, et al. Federatedtrust: A solution for trustworthy federated learning[J]. Future Generation Computer Systems, 2024, 152: 83-98.
4. Goebel R, Yu H, Faltings B, et al. Trustworthy Federated Learning[J]. Cham, Switzerland: Springer, 2023.
5. Chen C, Liu J, Tan H, et al. Trustworthy federated learning: privacy, security, and beyond[J]. Knowledge and Information Systems, 2025, 67(3): 2321-2356.
6. Li Y, Gao J, et al. Physical informed-inspired deep reinforcement learning based bi-level programming for microgrid scheduling[J]. IEEE Transactions on Industry Applications, 2025, 61(1): 1488-1500.
7. Wang Y, Cui Y, et al. Collaborative optimization of multi-microgrids system with shared energy storage based on multi-agent stochastic game and reinforcement learning[J]. Energy, 2023, 280: 128182.
8. Li Y, Han M, Yang Z, et al. Coordinating flexible demand response and renewable uncertainties for scheduling of community integrated energy systems with an electric vehicle charging station: A bi-level approach[J]. IEEE Transactions on Sustainable Energy, 2021, 12(4): 2321-2331.
9. Li Y, He S, Li Y, et al. Probabilistic charging power forecast of EVCS: Reinforcement learning assisted deep learning approach[J]. IEEE Transactions on Intelligent Vehicles, 2022, 8(1): 344-357.
10. Khan H, Consul P, Jabbari A, et al. Asynchronous Federated Learning Based Energy Scheduling for Microgrid-Enabled MEC Network[J]. IEEE Transactions on Consumer Electronics, 2025.
11. Veerasamy V, Sampath L P M I, Singh S, et al. Blockchain-based decentralized frequency control of microgrids using federated learning fractional-order recurrent neural network[J]. IEEE transactions on smart grid, 2023, 15(1): 1089-1102.
12. Li Y, He S, et al. Federated multiagent deep reinforcement learning approach via physics-informed reward for multimicrogrid energy management[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 35(5): 5902-5914.
13. Li G, Wu Y, Yesha Y. Decentralized condition monitoring for distributed wind systems: A federated learning-based approach to enhance SCADA data privacy[C]//Energy Sustainability. American Society of Mechanical Engineers, 2024, 87899: V001T01A010.
14. Shang Y, Li D, Li Y, et al. Explainable spatiotemporal multi-task learning for electric vehicle charging demand prediction[J]. Applied Energy, 2025, 384: 125460.

15. Li Y, Zhao B, Li Y, et al. Safe-AutoSAC: AutoML-enhanced safe deep reinforcement learning for integrated energy system scheduling with multi-channel informer forecasting and electric vehicle demand response[J]. Applied Energy, 2025, 399: 126468.
16. Yuan Z, Tian Y, Zhou Z, et al. Trustworthy federated learning against malicious attacks in web 3.0[J]. IEEE Transactions on Network Science and Engineering, 2024, 11(5): 3969-3982.
17. Wehbi O, Arisdakessian S, Guizani M, et al. Enhancing mutual trustworthiness in federated learning for data-rich smart cities[J]. IEEE Internet of Things Journal, 2024.
18. Rodríguez-Barroso N, García-Márquez M, Luzón M, et al. Challenges of Trustworthy Federated Learning: What's Done, Current Trends and Remaining Work[J]. arXiv preprint arXiv:2507.15796, 2025.
19. Cai C, Fang Y, Liu W, et al. FedCov: enhanced trustworthy federated learning for machine RUL prediction with continuous-to-discrete conversion[J]. IEEE Transactions on Industrial Informatics, 2024.
20. Gao X, Yang X, Yu H, et al. Fedprok: Trustworthy federated class-incremental learning via prototypical feature knowledge transfer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 4205-4214.
21. Rashid M M, Xiang Y, Uddin M P, et al. Trustworthy and fair federated learning via reputation-based consensus and adaptive incentives[J]. IEEE Transactions on Information Forensics and Security, 2025.
22. Zhang Z, Wu L, Jin J, et al. Secure federated learning for cloud-fog automation: Vulnerabilities, challenges, solutions, and future directions[J]. IEEE Transactions on Industrial Informatics, 2025, 21(5): 3528-3540.
23. Wu L, Jin Y, Yan Y, et al. FL-OTCSEnc: Towards secure federated learning with deep compressed sensing[J]. Knowledge-Based Systems, 2024, 291: 111534.
24. Kim D, Oh K, Lee Y, et al. Overview of fair federated learning for fairness and privacy preservation[J]. Expert Systems with Applications, 2025: 128568.
25. Han Z, Wang W, Huang J, et al. Distributed adaptive formation tracking control of mobile robots with event-triggered communication and denial-of-service attacks[J]. IEEE Transactions on Industrial Electronics, 2022, 70(4): 4077-4087.
26. Wang L, Qu Z, et al. Method for Extracting Patterns of Coordinated Network Attacks on Electric Power CPS Based on Temporal–Topological Correlation[J]. IEEE Access, 2020, 8: 57260-57272.
27. Chen L, Zhang W, Dong C, et al. Feddrl: Trustworthy federated learning model fusion method based on staged reinforcement learning[J]. Computing and Informatics, 2024.
28. Wang Y, Lu J, Liang J. Security control of multiagent systems under denial-of-service attacks[J]. IEEE Transactions on Cybernetics, 2020, 52(6): 4323-4333.
29. Bo X, Qu Z, Liu Y, et al. Review of active defense methods against power cps false data injection attacks from the multiple spatiotemporal perspective[J]. Energy Reports, 2022, 8: 11235-11248.
30. Ghosal M, Rao V. Fusion of multirate measurements for nonlinear dynamic state estimation of the power systems[J]. IEEE Transactions on Smart Grid, 2017, 10(1): 216-226.
31. Li Y, Zhang S, Li Y, et al. PMU measurements-based short-term voltage stability assessment of power systems via deep transfer learning[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-11.
32. Wu T, Chung C Y, Kamwa I. A fast state estimator for systems including limited number of PMUs[J]. IEEE Transactions on Power Systems, 2017, 32(6): 4329-4339.
33. Putra M A P, Karna N B A, Alief R N, et al. PureFed: An Efficient Collaborative and Trustworthy Federated Learning Framework Based on Blockchain Network[J]. IEEE Access, 2024, 12: 82413-82426.
34. Qu Z, Dong Y, Qu N, et al. Survivability Evaluation Method for Cascading Failure of Electric Cyber Physical System Considering Load Optimal Allocation[J]. Mathematical Problems in Engineering, 2019, 2019: 2817586.
35. Daole M, Ducange P, Marcelloni F, et al. Trustworthy AI in heterogeneous settings: federated learning of explainable classifiers[C]//2024 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE, 2024: 1-9.
36. Celdran A H, Feng C, Sanchez P M S, et al. Assessing the sustainability and trustworthiness of federated learning models[J]. arXiv preprint arXiv:2310.20435, 2023.

37. Quan M K, Pathirana P N, Wijayasundara M, et al. Federated learning for cyber physical systems: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2025.
38. War M R, Singh Y, Sheikh Z A, et al. Review on the Use of Federated Learning Models for the Security of Cyber-Physical Systems[J]. Scalable Computing: Practice and Experience, 2025, 26(1): 16-33.
39. Li Y, Li Y, Li G, et al. A multi-objective optimal power flow approach considering economy and environmental factors for hybrid AC/DC grids incorporating VSC-HVDC[J]. Power System Technology, 2016, 40(9): 2661-2667.
40. Han S, Ding H, Zhao S, et al. Practical and robust federated learning with highly scalable regression training[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 35(10): 13801-13815.
41. Consul P, Budhiraja I, Chaudhary R, et al. FLBCPS: Federated learning based secured computation offloading in blockchain-assisted cyber-physical systems[C]//2022 IEEE/ACM 15th International Conference on Utility and Cloud Computing (UCC). IEEE, 2022: 412-417.
42. Pene P, Musa A A, Musa U, et al. Edge intelligence in smart energy CPS[M]//Edge Intelligence in Cyber-Physical Systems. Academic Press, 2025: 169-192.
43. Difrina S, Ramkumar M P, Selvan G S R E. Trust Based Federated Learning for Privacy-preserving in Connected Smart Communities[C]//2025 Second International Conference on Cognitive Robotics and Intelligent Systems (ICC-ROBINS). IEEE, 2025: 409-418.
44. Qu Z, Qu N, Zhou Y, et al. Extraction of Typical Operating Scenarios of New Power System Based on Deep Time Series Aggregation[J]. CAAI Transactions on Intelligence Technology, 2024, 1-17. DOI: 10.1049/cit2.12369.
45. Chen D, Jiang X, Zhong H, et al. Building trusted federated learning: Key technologies and challenges[J]. Journal of Sensor and Actuator Networks, 2023, 12(1): 13.
46. Shafi S, Tariq N, Ashraf M, et al. Cyber Defense in Energy: Federated Learning Solution for Malicious Intrusions in Smart Grids[C]//International Conference on Data Analytics & Management. Singapore: Springer Nature Singapore, 2024: 177-191.
47. Majidi S H, Asharioun H. Privacy preserving federated learning solution for security of industrial cyber physical systems[M]//AI-Enabled Threat Detection and Security Analysis for Industrial IoT. Cham: Springer International Publishing, 2021: 195-211.
48. Čaušević S, Snijders R, Pingen G, et al. Flexibility prediction in smart grids: Making a case for federated learning[C]//IET Conference Proceedings CP785. Stevenage, UK: The Institution of Engineering and Technology, 2021, 2021(6): 1983-1987.
49. Li Y, Gu X. Feature selection for transient stability assessment based on improved maximal relevance and minimal redundancy criterion[J]. Proceedings of the CSEE, 2013, 33: 179-186.
50. Uddin M P, Xiang Y, Hasan M, et al. A Systematic Literature Review of Robust Federated Learning: Issues, Solutions, and Future Research Directions[J]. ACM Computing Surveys, 2025, 57(10): 1-62.
51. Li Y, Yang Z. Application of EOS-ELM with binary Jaya-based feature selection to real-time transient stability assessment using PMU data[J]. IEEE Access, 2017, 5: 23092-23101.
52. Tavallaie O, Thilakarathna K, Seneviratne S, et al. SHFL: Secure Hierarchical Federated Learning Framework for Edge Networks[J]. arXiv preprint arXiv:2409.15067, 2024.
53. Amin M, El-Sousy F F M, Aziz G A A, et al. CPS attacks mitigation approaches on power electronic systems with security challenges for smart grid applications: A review[J]. IEEE Access, 2021, 9: 38571-38601.
54. Shalabi E, Khedr W, Rushdy E, et al. A comparative study of privacy-preserving techniques in federated learning: A performance and security analysis[J]. Information, 2025, 16(3): 244.
55. Parizad A, Baghaee H R, Rahman S. Overview of Smart Cyber-Physical Power Systems: Fundamentals, Challenges, and Solutions[J]. Smart Cyber-Physical Power Systems: Fundamental Concepts, Challenges, and Solutions, 2025, 1: 1-69.
56. Chen L, Gu S, Wang Y, et al. Stacked Autoencoder Framework of False Data Injection Attack Detection in Smart Grid[J]. Mathematical Problems in Engineering, 2021, 2021(1): 2014345.
57. Lin G, Rehtanz C, Wang S, et al. Review on the key technologies of power grid cyber-physical systems simulation[J]. IET Cyber-Physical Systems: Theory & Applications, 2024, 9(1): 1-16.

58. Lin S, Qiu Y, Hu C, et al. A Scheme for Improving the Observability of Power Grid by Weakening the Coupling of CPS[C]//2024 IEEE 7th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, 2024, 7: 489-493.
59. Zhang X, Yan K, Guo Q, et al. A CPS Cyber Security Situation Assessment Model Against Cyberattacks[C]//2025 IEEE International Conference on Power Systems and Smart Grid Technologies (PSSGT). IEEE, 2025: 242-247.
60. Salehpour A, Al-Anbagi I. RTAP: A Real-time Model for Attack Detection and Prediction in Smart Grid Systems[J]. IEEE Access, 2024,12: 130425-130443.
61. Li Y, Wei X, et al. Detection of false data injection attacks in smart grid: A secure federated deep learning approach[J]. IEEE Transactions on Smart Grid, 2022, 13(6): 4862–4872.
62. Li Y, Li Y. Semi-supervised federated learning for collaborative security threat detection in control system for distributed power generation[J]. Engineering Applications of Artificial Intelligence, 2025, 148: 110374.
63. Dehbozorgi M R, Rastegar M. A deep learning deviation-based scheme to defend against false data injection attacks in power distribution systems[J]. Electric Power Systems Research, 2025, 238: 111076.
64. Shabbir A, Manzoor H U, Zoha A, et al. Smart grid security through fusion-enhanced federated learning against adversarial attacks[J]. Engineering Applications of Artificial Intelligence, 2025, 157: 111169.
65. Zhang Z, Rath S, Xu J, et al. Federated learning for smart grid: A survey on applications and potential vulnerabilities[J]. ACM Transactions on Cyber-Physical Systems, 2024.
66. Kumar G H, Reddy S, Saxena S, et al. FL-DPCSA: Federated learning with differential privacy for cache side-channel attack detection in edge-based smart grids[J]. e-Prime-Advances in Electrical Engineering, Electronics and Energy, 2025: 101057.
67. Sonani R, Govindarajan V, Verma P. Federated Learning-Driven Privacy-Preserving Framework for Decentralized Data Analysis and Anomaly Detection in Contract Review[J]. International Journal of Advanced Computer Science & Applications, 2025, 16(3).
68. Popli M S, Singh R P, Popli N K, et al. A federated learning framework for enhanced data security and cyber intrusion detection in distributed network of underwater drones[J]. IEEE Access, 2025, 13: 12634-12646.
69. Abdullah M, Mengash H A, Maray M, et al. Federated learning with Blockchain on Denial-of-Service attacks detection and classification of edge IIoT networks using Deep Transfer Learning model[J]. Computers and Electrical Engineering, 2025, 124: 110319.
70. Salsabil S, Kabir M S, Mitra R. Secure and Decentralized Homomorphic Federated Learning for Smart Microgrid Stability[C]//2024 27th International Conference on Computer and Information Technology (ICCIT). IEEE, 2024: 281-286.
71. Manojkumar R, Vaidya S P, Jena P K, et al. Machine Learning and Federated Learning in Industrial Cybersecurity[M]//AI-Enhanced Cybersecurity for Industrial Automation. IGI Global Scientific Publishing, 2025: 407-438.
72. Li Y, Cao J, Xu Y, et al. Deep learning based on Transformer architecture for power system short-term voltage stability assessment with class imbalance[J]. Renewable and Sustainable Energy Reviews, 2024, 189: 113913.
73. Alohali M A, Dafaalla H, Baihan M, et al. Leveraging self attention driven gated recurrent unit with crocodile optimization algorithm for cyberattack detection using federated learning framework[J]. Scientific Reports, 2025, 15(1): 23805.
74. Li Y, Zhang S, Li Y. AI-enhanced resilience in power systems: Adversarial deep learning for robust short-term voltage stability assessment under cyber-attacks[J]. Chaos, Solitons & Fractals, 2025, 196: 116406.
75. Du R, Li X, He D, et al. Toward secure and verifiable hybrid federated learning[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 2935-2950.
76. Namakshenas D, Yazdinejad A, Dehghantanha A, et al. IP2FL: Interpretation-based privacy-preserving federated learning for industrial cyber-physical systems[J]. IEEE Transactions on Industrial Cyber-Physical Systems, 2024,2: 321-330.

77. Shafi S, Tariq N, Khan F A, et al. Federated learning for enhanced malware threat detection to secure smart power grids[C]//International Conference on Ubiquitous Computing and Ambient Intelligence. Cham: Springer Nature Switzerland, 2024: 692-703.
78. Marfo W, Tosh D K, Moore S V. Federated learning for efficient condition monitoring and anomaly detection in industrial cyber-physical systems[C]//2025 International Conference on Computing, Networking and Communications (ICNC). IEEE, 2025: 740-746.
79. Kausar F, Deo S, Hussain S, et al. Federated Deep Learning Model for False Data Injection Attack Detection in Cyber Physical Power Systems[J]. Energies, 2024, 17(21): 5337.
80. Busari W A, Bello A A. Security, Trust, and Privacy in Cyber-physical Systems (CPS)[C]//2024 2nd International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV). IEEE, 2024: 1-6.
81. Pereira L, Nair V, Dias B, et al. Federated learning forecasting for strengthening grid reliability and enabling markets for resilience[C]//IET Conference Proceedings CP882. Stevenage, UK: The Institution of Engineering and Technology, 2024, 2024(27): 246-250.
82. Yang N, Wang S, Chen M, et al. A Privacy Preserving and Byzantine Robust Collaborative Federated Learning Method Design[C]//ICC 2024-IEEE International Conference on Communications. IEEE, 2024: 3598-3603.
83. Li Y, Guo Z, Yang N, et al. Threats and defenses in the federated learning life cycle: a comprehensive survey and challenges[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025.
84. Mohanty R, Dash R K, Tripathy P K. FedAgri: A Federated Learning Framework for Sustainable Agriculture in Cyber-Physical Systems[C]//2025 International Conference on Intelligent and Cloud Computing (ICoICC). IEEE, 2025: 1-6.
85. Adil M, Farouk A, Abulkasim H, et al. NG-ICPS: Next generation Industrial-CPS, Security Threats in the era of Artificial Intelligence, Open challenges with future research directions[J]. IEEE Internet of Things Journal, 2025, 12(2): 1343-1367.
86. Yu B, Zhao J, Zhang K, et al. Lightweight and dynamic privacy-preserving federated learning via functional encryption[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 2496-2508.
87. Adnan M, Syed M H, Anjum A, et al. A framework for privacy-preserving in iov using federated learning with differential privacy[J]. IEEE Access, 2025, 13: 13507-13521.
88. Khraisat A, Alazab A, Singh S, et al. Survey on federated learning for intrusion detection system: Concept, architectures, aggregation strategies, challenges, and future directions[J]. ACM Computing Surveys, 2024, 57(1): 1-38.
89. GK S K, Muniyal B, Rajarajan M. Explainable Federated Framework for Enhanced Security and Privacy in Connected Vehicles Against Advanced Persistent Threats[J]. IEEE Open Journal of Vehicular Technology, 2025, 6: 1438-1463.
90. Lei S, Xia X, Sha J. Day-ahead Demand Response Potential Forecasting Method for Data Centers Based on Federated Learning[C]//2024 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia). IEEE, 2024: 490-495.
91. Fatema K, Dey S K, Anannya M, et al. Federated XAI IDS: An explainable and safeguarding privacy approach to detect intrusion combining federated learning and SHAP[J]. Future Internet, 2025, 17(6): 234.
92. Ding Z, Wang W, Li X, et al. Identifying alternately poisoning attacks in federated learning online using trajectory anomaly detection method[J]. Scientific Reports, 2024, 14(1): 20269.
93. Alcaraz C, Martinelli F. Trusted Platform and Privacy Management in Cyber Physical Systems: The DUCA Framework[J]. IFIP Annual Conference on Data and Applications Security and Privacy, vol. 15722, 2025.
94. Mazid A, Kirmani S, Manaullah, et al. FL-IDPP: A Federated Learning Based Intrusion Detection Approach With Privacy Preservation[J]. Transactions on Emerging Telecommunications Technologies, 2025, 36(1): e70039.
95. Torre D, Chennamaneni A, Jo J, et al. Toward enhancing privacy preservation of a federated learning CNN intrusion detection system in IoT: method and empirical study[J]. ACM Transactions on Software Engineering and Methodology, 2025, 34(2): 1-48.

96. Xiong H, Zhao Y, Xia Y, et al. DA-FL: Blockchain Empowered Secure and Private Federated Learning With Anonymous Authentication[J]. IEEE Transactions on Reliability, 2025.
97. Raza M, Saeed M J, Riaz M B, et al. Federated learning for privacy-preserving intrusion detection in software-defined networks[J]. IEEE Access, 2024, 12: 69551-69567.
98. Moussaoui J E, Kmiti M, El Gholami K, et al. A Systematic Review on Hybrid AI Models Integrating Machine Learning and Federated Learning[J]. Journal of Cybersecurity and Privacy, 2025, 5(3): 41.
99. Jerkovic F, Sarkar N I, Ali J. Smart Grid IoT Framework for Predicting Energy Consumption Using Federated Learning Homomorphic Encryption[J]. Sensors, 2025, 25(12): 3700.
100. Shah K, Kumari A, Solanki B, et al. ZTSec-FedSDN: A Privacy-Preserving Federated Framework for SDN Attack Detection Using Zero-Trust Blockchain and 6G Terahertz Networks[C]//2025 International Conference on Computer, Information and Telecommunication Systems (CITS). IEEE, 2025: 1-6.
101. Gupta P, Sengupta B, Nandi S. Federated Learning-Driven Intrusion Detection for Cybersecurity in Smart Distribution system[C]//2024 IEEE Globecom Workshops (GC Wkshps). IEEE, 2024: 1-6.
102. Han J, Wang L, Liu Z, et al. PPFL: Privacy-Preserving Federated Learning Based on Differential Privacy and Personalized Data Transformation[J]. IEEE Internet of Things Journal, 2025.
103. An Z, Johnson T T, Ma M. Formal logic enabled personalized federated learning through property inference[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(10): 10882-10890.
104. Huang S, Li Y, Yan X, et al. Scope: On Detecting Constrained Backdoor Attacks in Federated Learning[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 3302-3315.
105. Ari I, Balkan K, Pirbhulal S, et al. Ensuring Security Continuum from Edge to Cloud: Adaptive Security for IoT-based Critical Infrastructures using FL at the Edge[C]//2024 IEEE International Conference on Big Data (BigData). IEEE, 2024: 4921-4929.
106. Gokcen A, Boyaci A. Robust Federated Learning with Confidence-Weighted Filtering and GAN-Based Completion under Noisy and Incomplete Data[J]. arXiv preprint arXiv:2505.09733, 2025.
107. Li Z, Yao W, Luo J, et al. Flow-Based IoT Intrusion Detection via Improved Generative Federated Distillation Learning[J]. IEEE Internet of Things Journal, 2025, 12(10): 14797-14811.
108. Mejdi H, Elmadssia S, Koubaa M, et al. A comprehensive survey on game theory applications in cyber-physical system security: attack models, security analyses, and machine learning classifications[J]. IEEE Access, 2024, 12: 163638-163653.
109. Yu X, Xue Y. Smart grids: A cyber–physical systems perspective[J]. Proceedings of the IEEE, 2016, 104(5): 1058-1070.
110. Han S, Ding H, Zhao S, et al. Fed-GAN: Federated Generative Adversarial Network with Privacy-Preserving for Cross-Device Scenarios[J]. IEEE Transactions on Dependable and Secure Computing, 2025.
111. Swearingen M, Brunasso S, Weiss J, et al. What you need to know (and don't) about the Aurora vulnerability[J]. Power, 2013, 157(9): 52-52.
112. Li Y, Ma W, Li Y, et al. Enhancing cyber-resilience in integrated energy system scheduling with demand response using deep reinforcement learning[J]. Applied Energy, 2025, 379: 124831.
113. Liu T, Wu H, Sun X, et al. FL-APB: Balancing Privacy Protection and Performance Optimization for Adversarial Training in Federated Learning[J]. Electronics, 2024, 13(21): 4187.
114. Musleh A S, Chen G, Dong Z Y. A survey on the detection algorithms for false data injection attacks in smart grids[J]. IEEE Transactions on Smart Grid, 2019, 11(3): 2218-2234.
115. Lee S. Adaptive selection of loss function for federated learning clients under adversarial attacks[J]. IEEE Access, 2024, 12: 96051-96062.
116. Konstantinou C, Maniatakos M. A case study on implementing false data injection attacks against nonlinear state estimation[C]//Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy. 2016: 81-92.
117. Bondok A H, Mahmoud M, Badr M M, et al. A Distillation-Based Attack Against Adversarial Training Defense for Smart Grid Federated Learning[C]//2024 IEEE 21st Consumer Communications & Networking Conference (CCNC). IEEE, 2024: 963-968.

118. Macwan R, Drew C, Panumpabi P, et al. Collaborative defense against data injection attack in IEC61850 based smart substations[C]//2016 IEEE Power and Energy Society General Meeting (PESGM). IEEE, 2016: 1-5.
119. Pan S, Morris T, Adhikari U. Developing a hybrid intrusion detection system using data mining for power systems[J]. IEEE Transactions on Smart Grid, 2015, 6(6): 3104-3113.
120. Liu X, Chang P, Sun Q. Detection of False Data Injection Attacks in Power Grids Based on XGBoost and Unscented Kalman Filter Adaptive Hybrid Prediction[J]. Proceedings of the CSEE, 2021, 41(16): 5462-5476.
121. Kubo R. Detection and mitigation of false data injection attacks for secure interactive networked control systems[C]//2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR). IEEE, 2018: 7-12.
122. Peng L, Cao X, Shi H, et al. Optimal jamming attack schedule for remote state estimation with two sensors[J]. Journal of the Franklin Institute, 2018, 355(14): 6859-6876.
123. Liang G, Zhao J, Luo F, et al. A review of false data injection attacks against modern power systems[J]. IEEE Transactions on Smart Grid, 2016, 8(4): 1630-1638.
124. Wehbi O, Arisdakessian S, Guizani M, et al. Enhancing mutual trustworthiness in federated learning for data-rich smart cities[J]. IEEE Internet of Things Journal, 2024.
125. Hu Z, Wang Y, Tian X, et al. False data injection attacks identification for smart grids[C]//2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE). IEEE, 2015: 139-143.
126. Yang F, Wang J, Pan Q, et al. Resilient Event-Triggered Control for Cyber-Physical Integrated Power Systems Under Network Attacks[J]. Acta Automatica Sinica, 2019, 45(1): 110-119.
127. Rodríguez-Barroso N, García-Márquez M, Luzón M, et al. Challenges of Trustworthy Federated Learning: What's Done, Current Trends and Remaining Work[J]. arXiv preprint arXiv:2507.15796, 2025.
128. Jin M, Lavaei J, Johansson K H. Power grid AC-based state estimation: Vulnerability analysis against cyber attacks[J]. IEEE Transactions on Automatic Control, 2018, 64(5): 1784-1799.
129. Liu X, Li Z. False data attacks against AC state estimation with incomplete network information[J]. IEEE Transactions on Smart Grid, 2016, 8(5): 2239-2248.
130. Zhao J, Mili L, Wang M. A generalized false data injection attacks against power system nonlinear state estimator and countermeasures[J]. IEEE Transactions on Power Systems, 2018, 33(5): 4868-4877.
131. Jorjani M, Seifi H, Varjani A Y. A graph theory-based approach to detect false data injection attacks in power system AC state estimation[J]. IEEE Transactions on Industrial Informatics, 2020, 17(4): 2465-2475.
132. James J Q, Hou Y, Li V O K. Online false data injection attack detection with wavelet transform and deep neural networks[J]. IEEE Transactions on Industrial Informatics, 2018, 14(7): 3271-3280.
133. Zhang Y, Wang J, Chen B. Detecting false data injection attacks in smart grids: A semi-supervised deep learning approach[J]. IEEE Transactions on Smart Grid, 2020, 12(1): 623-634.
134. Yin X, Zhu Y, Hu J. A subgrid-oriented privacy-preserving microservice framework based on deep neural network for false data injection attack detection in smart grids[J]. IEEE Transactions on Industrial Informatics, 2021, 18(3): 1957-1967.
135. Li S, Yılmaz Y, Wang X. Quickest detection of false data injection attack in wide-area smart grids[J]. IEEE Transactions on Smart Grid, 2014, 6(6): 2725-2735.
136. Liu Y, Garg S, Nie J, et al. Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach[J]. IEEE Internet of Things Journal, 2020, 8(8): 6348-6358.
137. Li B, Wu Y, Song J, et al. DeepFed: Federated deep learning for intrusion detection in industrial cyber–physical systems[J]. IEEE Transactions on Industrial Informatics, 2020, 17(8): 5615-5624.
138. Li Y, Li J, Wang Y. Privacy-preserving spatiotemporal scenario generation of renewable energies: A federated deep generative learning approach[J]. IEEE Transactions on Industrial Informatics, 2021, 18(4): 2310-2320.
139. Liu X, Li Z, Liu X, et al. Masking transmission line outages via false data injection attacks[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(7): 1592-1602.
140. Jia L, Kim J, Thomas R J, et al. Impact of data quality on real-time locational marginal price[J]. IEEE Transactions on Power Systems, 2013, 29(2): 627-636.

141. Song M, Zhou J, Gao C, et al. Coordinated operation of urban buildings and distribution networks from a CPSS perspective: A review and outlook[J]. Automation of Electric Power Systems, 2023, 47(23): 105–121.
142. Qu Z, Bo X, Yu T, et al. Active and Passive Hybrid Detection Method for Power CPS False Data Injection Attacks with Improved AKF and GRU-CNN[J]. IET Renewable Power Generation, 2022, 16: 1490-1508.
143. Liang G, Weller S R, Luo F, et al. Generalized FDIA-based cyber topology attack with application to the Australian electricity market trading mechanism[J]. IEEE Transactions on Smart Grid, 2017, 9(4): 3820-3829.
144. Jokar P, Arianpoo N, Leung V C M. Electricity theft detection in AMI using customers' consumption patterns[J]. IEEE Transactions on Smart Grid, 2015, 7(1): 216-226.
145. Qu Z, Xie Q, Liu Y, et al. Power Cyber-Physical System Risk Area Prediction Using Dependent Markov Chain and Improved Grey Wolf Optimization[J]. IEEE Access, 2020, 8: 82844-82854.
146. Chang Z, Wu J, Liang H, et al. A review of Power System False data attack Detection Technology based on Big data[J]. Information, 2024, 15(8): 439.
147. Xie L, Mo Y, Sinopoli B. False data injection attacks in electricity markets[C]//2010 First IEEE International Conference on Smart Grid Communications. IEEE, 2010: 226-231.
148. Tajer A. False data injection attacks in electricity markets by limited adversaries: Stochastic robustness[J]. IEEE Transactions on Smart Grid, 2017, 10(1): 128-138.
149. He S, Li Y, et al. Boosting communication efficiency in federated learning for multiagent-based multimicrogrid energy management[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(5): 8592–8605.
150. Qu Z, Shi H, Wang Y, et al. Active and Passive Defense Strategies of Cyber-Physical Power System against Cyber Attacks Considering Node Vulnerability[J]. Processes, 2022, 10(7): 1351.
151. Javed F, Mangues-Bafalluy J, Zeydan E, et al. Blockchain and trustworthy reputation for federated learning: Opportunities and challenges[C]//2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom). IEEE, 2024: 578-584.
152. Luo X, Li Y, Wang X, et al. Interval observer-based detection and localization against false data injection attack in smart grids[J]. IEEE Internet of Things Journal, 2020, 8(2): 657-671.
153. Zhang L, Xu Y, Wu X, et al. Distributed resilient control of AC microgrids against false data injection attacks [J/OL]. Automation of Electric Power Systems, 2023, 47(8):44-52.
154. Dong Y, Wang Q, Cao J, et al. Identification of false data injection attacks in power grids based on oversampling and cascaded machine learning [J/OL]. Automation of Electric Power Systems, 2023, 47(8): 179-188.
155. Zheng Y, Zhang J, Yao W, et al. Detection of false data injection attacks in power grids based on spatial features of synchronized measurements [J/OL]. Automation of Electric Power Systems,2023,47. (10):128-13.
156. Konstantinou C, Sazos M, Musleh A S, et al. GPS spoofing effect on phase angle monitoring and control in a real-time digital simulator-based hardware-in-the-loop environment[J]. IET Cyber-Physical Systems: Theory & Applications, 2017, 2(4): 180-187.
157. Li A, Chang X, Ma J, et al. Vtfl: A blockchain based vehicular trustworthy federated learning framework[C]//2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2023, 6: 1002-1006.
158. Liu J, Ma B, Yu Q, et al. Efficient Federated Learning with Heterogeneous Data and Adaptive Dropout[J]. ACM Transactions on Knowledge Discovery from Data, 2025.
159. Tan R, Nguyen H H, Foo E Y S, et al. Optimal false data injection attack against automatic generation control in power grids[C]//2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS). IEEE, 2016: 1-10.
160. Musleh A S, Khalid H M, Muyeen S M, et al. A prediction algorithm to enhance grid resilience toward cyber attacks in WAMCS applications[J]. IEEE Systems Journal, 2017, 13(1): 710-719.
161. Deng R, Xiao G, Lu R, et al. False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey[J]. IEEE Transactions on Industrial Informatics, 2016, 13(2): 411-423.
162. Tao M, Liao L, Zhang Y, et al. EDT-SaFL: Semi-Asynchronous Federated Learning for Edge Digital Twin in Industrial Internet-of-Things[J]. IEEE Transactions on Mobile Computing, 2025.

163. Liu Y, Ning P, Reiter M K. False data injection attacks against state estimation in electric power grids[J]. ACM Transactions on Information and System Security (TISSEC), 2011, 14(1): 1-33.
164. Manandhar K, Cao X, Hu F, et al. Detection of faults and attacks including false data injection attack in smart grid using Kalman filter[J]. IEEE Transactions on Control of Network Systems, 2014, 1(4): 370-379.
165. Le J, Lang H, Tan T, et al. A Review of Information Security Issues in Distributed Economic Dispatch of New Distribution Systems[J]. Automation of Electric Power Systems, 2024, 48(12): 177-191.
166. Telçeken M, Bozkaya-Aras E. A Trustworthy Federated Learning Model: Client Selection for IoT Edge Networks[C]//2024 Innovations in Intelligent Systems and Applications Conference (ASYU). IEEE, 2024: 1-6.
167. Bo X, Chen X, Li H, et al. Modeling Method for the Coupling Relations of Microgrid Cyber-Physical Systems Driven by Hybrid Spatiotemporal Events[J]. IEEE Access, 2021, 9: 19619-19631.
168. Zuo X, Wang M, Zhu T, et al. Federated learning with blockchain-enhanced machine unlearning: A trustworthy approach[J]. IEEE Transactions on Services Computing, 2025.
169. Wang L, Xu P, Qu Z, et al. Coordinated Cyber-Attack Detection Model of Cyber-Physical Power System Based on the Operating State Data Link[J]. Frontiers in Energy Research, 2021, 9: 666130.
170. Qu Z, Zhang Y, Qu N, et al. Method for Quantitative Estimation of the Risk Propagation Threshold in Electric Power CPS Based on Seepage Probability[J]. IEEE Access, 2018, 6: 68813-68823.
171. Alomari M A, Al-Andoli M N, Ghaleb M, et al. Security of Smart Grid: Cybersecurity Issues, Potential Cyberattacks, Major Incidents, and Future Directions[J]. Energies, 2025, 18(1): 141.
172. Guan Y, Ge X. Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks[J]. IEEE Transactions on Signal and Information Processing over Networks, 2017, 4(1): 48-59.
173. Kou L, Wu J, Zhang F, et al. Image encryption for Offshore wind power based on 2D-LCLM and Zhou Yi Eight Trigrams[J]. International Journal of Bio-Inspired Computation, 2023, 22(1): 53-64.
174. Tang Z, Meng Q, Cao S, et al. Wind power ramp prediction algorithm based on wavelet deep belief network[J]. Acta Energiae Solaris Sinica, 40(11): 3213-3220.
175. Li Y, Wang R, Li Y, et al. Wind power forecasting considering data privacy protection: A federated deep reinforcement learning approach[J]. Applied Energy, 2023, 329: 120291.
176. Long X, Ding Y, et al. Privacy-Preserving Graph Inference Network for Multi-Entity Wind Power Forecast: A Federated Learning Approach[J]. IEEE Transactions on Network Science and Engineering, 2025, 12(4): 2428-2444.
177. Zhou X, Feng J, et al. Non-intrusive load decomposition based on CNN–LSTM hybrid deep learning model[J]. Energy Reports, 2021, 7: 5762-5771.
178. Qu Z, Dong Y, Mugemanyi S, et al. Dynamic Exploitation Gaussian Bare-Bones Bat Algorithm for Optimal Reactive Power Dispatch to Improve the Safety and Stability of Power System[J]. IET Renewable Power Generation, 2022, 16: 1401-1424.
179. Le Vinh T, Tran H T, Phan H T, et al. Federated Learning-Based Trust Evaluation with Fuzzy Logic for Privacy and Robustness in Fog Computing[J]. IEEE Access, 2025, 13: 137952-137972.
180. Deng R, Liang H. False data injection attacks with limited susceptance information and new countermeasures in smart grid[J]. IEEE Transactions on Industrial Informatics, 2018, 15(3): 1619-1628.
181. Qu Z, Dong Y, Qu N, et al. Quantitative Assessment of Survivability of Power CPS Considering Load Optimization and Reconfiguration[J]. Automation of Electric Power Systems, 2019, 43(6): 15-24.
182. Zhang Z, Deng R, Yau D K Y, et al. Analysis of moving target defense against false data injection attacks on power grid[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 2320-2335.
183. He Z, Gao S, Wei X, et al. Research on Attack-Defense Game Model of False Topology Attacks with Branch and Protection Coordination[J]. Power System Technology, 2022, 46(11): 4346-4355.
184. Li Y, Long X, Li Y, et al. A demand–supply cooperative responding strategy in power system with high renewable energy penetration[J]. IEEE transactions on control systems technology, 2023, 32(3): 874-890.
185. Tahir B, Jolfaei A, Tariq M. Experience-driven attack design and federated-learning-based intrusion detection in industry 4.0[J]. IEEE Transactions on Industrial Informatics, 2021, 18(9): 6398-6405.

186. Li Y, Quevedo D E, Dey S, et al. SINR-based DoS attack on remote state estimation: A game-theoretic approach[J]. IEEE Transactions on Control of Network Systems, 2016, 4(3): 632-642.
187. Zhang H, Qi Y, Wu J, et al. DoS attack energy management against remote state estimation[J]. IEEE Transactions on Control of Network Systems,2018,5(1):383-394.
188. Yuan Z, Tian Y, Zhou Z, et al. Trustworthy federated learning against malicious attacks in web 3.0[J]. IEEE Transactions on Network Science and Engineering, 2024, 11(5): 3969-3982.
189. Jin P, Li Y, Li G, et al. Optimized hierarchical power oscillations control for distributed generation under unbalanced conditions[J]. Applied energy, 2017, 194: 343-352.
190. Wu Z, Cheng H. Voltage and frequency control of microgrids considering state constraints and attacks[J]. IEEE Transactions on Network Science and Engineering, 2024, 11(5): 4979-4989.
191. Zhong C, Li H, Zhou Y, et al. Virtual synchronous generator of PV generation without energy storage for frequency support in autonomous microgrid[J]. International Journal of Electrical Power & Energy Systems, 2022, 134: 107343.
192. Rezazadeh F, Bartzoudis N. A federated DRL approach for smart micro-grid energy control with distributed energy resources[C]//2022 IEEE 27th international workshop on computer aided modeling and design of communication links and networks (CAMAD). IEEE, 2022: 108-114.
193. Lin S W, Chu C C, Tung C F. Distributed actor-critic neural networks-based structure for frequency synchronization in isolated AC microgrids[J]. IEEE Transactions on Industry Applications, 2024, 60(3): 4433-4445.
194. Li Y, Li J, Chen L. Dynamic state estimation of synchronous machines based on robust cubature Kalman filter under complex measurement noise conditions[J]. Transactions of china electrotechnical society, 2019, 34(17): 3651-3660.
195. Li X., Li W., Du D., et al. Dynamic state estimation of smart grid under denial-of-service attacks based on UKF [J]. Acta Automatica Sinica, 2019, 45(1): 120–131.
196. Chen L, Jin P, Yang J, et al. Robust Kalman Filter-Based Dynamic State Estimation of Natural Gas Pipeline Networks[J]. Mathematical Problems in Engineering, 2021, 2021(1): 5590572.
197. Li Y, Li Z, Chen L, et al. A false data injection attack method for generator dynamic state estimation[J]. Transactions of China Electrotechnical Society, 2020, 35(7): 1476-1488.
198. Chen L, et al. Robust Dynamic State Estimator of Integrated Energy Systems Based on Natural Gas Partial Differential Equations[J]. IEEE Transactions on Industry Applications, 2022, 58(3): 3303-3312.
199. Chen L, Hui X, et al. Dynamic state estimation for integrated natural gas and electric power systems[C]//2021 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia). IEEE, 2021: 397-402.
200. Li Y, Li Z, Chen L. Dynamic State Estimation of Generators Under Cyber Attacks[J]. IEEE Access, 2019, 7: 125252-125267.
201. Wu H, Xu Z, Ruan J, et al. FedDSSE: A personalized federated learning approach for distribution system state estimation[J]. CSEE Journal of Power and Energy Systems, 2024, 10(5): 226 -2270.
202. Qu Z, Dong Y, Li Y, et al. Localization of Dummy Data Injection Attacks in Power Systems Considering Incomplete Topological Information: A Spatio-Temporal Graph Wavelet Convolutional Neural Network Approach[J]. Applied Energy, 2024, 360: 122736.
203. Piperigkos N, Gkillas A, Anagnostopoulos C, et al. Federated Data-Driven Kalman Filtering for State Estimation[C]//2024 IEEE 26th International Workshop on Multimedia Signal Processing (MMSP). IEEE, 2024: 1-6.
204. Li Y, Li J, Qi J, et al. Robust cubature Kalman filter for dynamic state estimation of synchronous machines under unknown measurement noise statistics[J]. IEEE Access, 2019, 7: 29139-29148.
205. McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. PMLR, 2017: 1273-1282.
206. Zhu L, Liu Z, Han S. Deep leakage from gradients[J]. Advances in Neural Information Processing Systems, 2019, 32.
207. Zumtaugwald L. Designing and Implementing an Advanced Algorithm to Measure the Trustworthiness Level of Federated Learning Models[D]. University of Zurich, 2023.

208. Paillier P. Public-key cryptosystems Based on composite degree residuosity classes[C]//Advances in Cryptology—EUROCRYPT'99: International Conference on the Theory and Application of Cryptographic Techniques Prague, Czech Republic, 1999 Proceedings 18. Springer Berlin Heidelberg, 1999: 223-238.
209. Luo B, Beuran R, Tan Y. Smart grid security: attack modeling from a CPS perspective[C]//2020 IEEE Computing, Communications and IoT Applications (ComComAp). IEEE, 2020: 1-6.
210. Sarkar A, Mathur B, Kushwah V S, et al. Smart grid and energy management in smart cyber-physical systems (SCPS)[M]//Smart Cyber-Physical Systems. CRC Press, 277-298.
211. Meera V M, Arjun K P. Challenges in Ensuring Security for Smart Energy Management Systems Based on CPS[J]. Cyber Physical Energy Systems, 2024: 217-254.
212. Pene P, Musa A A, Musa U, et al. Secured edge intelligence in smart energy CPS[M]//Edge Intelligence in Cyber-Physical Systems. Academic Press, 2025: 325-351.
213. Zhang G, Gao W, Li Y, et al. A two-stage recovery strategy against false data injection attacks in smart grids[J]. Electric Power Systems Research, 2025, 245: 111632.
214. Chandu G, Karthik T, Parag B. Federated Learning for Distributed IoT Security: A Privacy-Preserving Approach to Intrusion Detection[J]. IEEE Access, 2025, 13: 135863-135875.
215. Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. ArXiv Preprint ArXiv:1412.6980, 2014.
216. Chen Y, Yang Y, Liang Y, et al. Federated learning with privacy preservation in large-scale distributed systems using differential privacy and homomorphic encryption[J]. Informatica, 2025, 49(13).
217. Zimmerman R D, Murillo-Sánchez C E, Thomas R J. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education[J]. IEEE Transactions on Power Systems, 2010, 26(1): 12-19.
218. Jiang J, Tian B, Yuan X, et al. Optimal DoS Attack Energy Allocation in Cyber-Physical Systems Based on Deep Reinforcement Learning[J]. IEEE Internet of Things Journal, 2025.
219. Custodio P M, Putra M A P, Lee J M, et al. TLFed: Federated Learning-based 1D-CNN-LSTM Transmission Line Fault Location and Classification in Smart Grids[C]//2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE, 2024: 026-031.
220. Chen Z, Tian P, Liao W, et al. Resource-aware knowledge distillation for federated learning[J]. IEEE Transactions on Emerging Topics in Computing, 2023, 11(3): 706-719.
221. Bhatia K, Bhattacharya S, Sharma I. Privacy-preserving detection of DDoS attacks in IoT using federated learning techniques[C]//2024 IEEE International Conference on Big Data & Machine Learning (ICBDML). IEEE, 2024: 234-239.
222. Maya P, Thomas M, Salam P A. P2P Energy Sharing with Federated Learning and Blockchain[C]//2024 International Conference on Sustainable Energy: Energy Transition and Net-Zero Climate Future (ICUE). IEEE, 2024: 1-4.
223. Lu Z, Wang L, Zhang Z, et al. TMT-FL: Enabling trustworthy model training of federated learning with malicious participants[J]. IEEE Transactions on Dependable and Secure Computing, 2024.
224. He Y, Mendis G J, Wei J. Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism[J]. IEEE Transactions on Smart Grid, 2017, 8(5): 2505-2516.
225. Gholami A, Torkzaban N, Baras J S. Trusted decentralized federated learning[C]//2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC). IEEE, 2022: 1-6.
226. Nariman G S, Hamarashid H K. Communication overhead reduction in federated learning: a review[J]. International Journal of Data Science and Analytics, 2025, 19(2): 185-216.
227. Wang S, Bi S, Zhang Y J A. Locational detection of the false data injection attack in a smart grid: A multilabel classification approach[J]. IEEE Internet of Things Journal, 2020, 7(9): 8218-8227.
228. Yogi M K, Chakravarthy A S N. Privacy-Preserving Deep Reinforcement Learning for Secure Resource Orchestration in Cyber-Physical Systems[J]. International Journal of Scientific Research in Network Security and Communication, 2025, 13(2): 12-21.
229. Androutsopoulou M, Carayannis E G, Askounis D, et al. Towards AI-Enabled Cyber-Physical Infrastructures—Challenges, Opportunities, and Implications for a Data-Driven eGovernment Theory, Policy, and Practice[J]. Journal of the Knowledge Economy, 2025: 1-38.

230. Nweke L O, Yayilgan S Y. Opportunities and challenges of using artificial intelligence in securing cyber-physical systems[J]. Artificial Intelligence for Security: Enhancing Protection in a Changing World, 2024: 131-164.