Review

Redefining Adversarial Dynamics: Co-Evolution of Attack and Defense Strategies in AI-Enabled Power Cyber-Physical Systems

Shuo Sun *

Posted Date: 24 June 2025

doi: 10.20944/preprints202506.1885.v1

Keywords: power cyber-physical systems; adversarial machine learning; co-evolutionary defense; reinforcement learning-based attacks; digital twin-based validation; human-in-the-loop security

 doi:10.20944/preprints202506.1885.v1

Review

Redefining Adversarial Dynamics: Co-Evolution of Attack and Defense Strategies in AI-Enabled Power Cyber-Physical Systems

Shuo Sun

College of Electrical Engineering, Guizhou University, Guiyang, China; s.sun95@outlook.com

Abstract

The rise of Artificial Intelligence (AI) in Power Cyber-Physical Systems (Power CPS) has created both transformative capabilities and new security vulnerabilities. While AI enables advanced monitoring, control, and optimization, it also empowers adversaries to craft adaptive, stealthy, and highly effective cyber-physical attacks. This review redefines the cybersecurity paradigm for Power CPS by introducing co-evolutionary defense frameworks that dynamically adapt to adversarial learning and strategy evolution. It systematically analyzes emerging AI-augmented attack techniques, including adversarial machine learning, reinforcement learning-based policy optimization, model poisoning, and federated learning exploitation. The review also critiques current AI-based defense mechanisms, highlighting their limitations in explainability, robustness, and adaptability. To address these challenges, it proposes game-theoretic modeling, digital twin-based validation, and human-in-the-loop adaptive defense as core components of a resilient security architecture. Key research gaps and collaboration needs are identified, offering a roadmap for developing scalable, trustworthy, and mission-centric cyber-physical defense ecosystems. This work aims to advance the field toward dynamic, co-evolutionary security strategies capable of safeguarding next-generation energy systems.

Keywords: power cyber-physical systems; adversarial machine learning; co-evolutionary defense; reinforcement learning-based attacks; digital twin-based validation; human-in-the-loop security

---

1. Introduction

1.1. Power CPS in the AI Era: Opportunities and New Risks

The evolution of power systems into highly interconnected Power Cyber-Physical Systems (Power CPS) represents one of the most transformative advancements in modern energy infrastructure [1-3]. Power CPS tightly integrate physical grid components—such as generators, substations, transmission lines, and distributed energy resources (DERs)—with digital layers of computation, communication, and control [4-5]. This integration enhances the grid's capability to monitor, optimize, and manage energy flows in real time, supporting emerging paradigms such as smart grids [6], microgrids [7], integrated energy systems [8], and transactive energy markets [9].

The deployment of Artificial Intelligence (AI) technologies has further accelerated this transformation. AI enables Power CPS to process vast amounts of real-time data, automate control strategies, and optimize operations across multiple spatial and temporal scales [10-12]. Applications include load forecasting, fault detection, renewable energy integration, and market optimization [13-15]. AI techniques such as machine learning (ML), deep learning, reinforcement learning (RL), and federated learning (FL) are increasingly embedded in energy management systems, supervisory control and data acquisition (SCADA) platforms, and distributed control frameworks [16-18].

 doi:10.20944/preprints202506.1885.v1

However, these same AI capabilities introduce new attack surfaces and adversarial risks. Malicious actors can exploit AI's data dependencies, model vulnerabilities, and control influence to launch AI-augmented cyber-physical attacks [19-20]. Techniques such as adversarial machine learning [21], model poisoning, backdoor insertion, and reinforcement learning-based attack policy optimization have emerged as credible threats. These attacks can mislead detection systems, manipulate grid control actions, and cause widespread disruptions [22-23].

1.2. From Static Defense to Dynamic Adversarial Co-Evolution

Traditional security strategies in power systems have largely focused on static defenses, such as cryptographic protocols, access control mechanisms, and rule-based anomaly detection. While these methods are essential, they often assume a static threat model, where known vulnerabilities are patched, and known attack patterns are detected [24-25].

In contrast, modern adversaries leverage dynamic, adaptive, and AI-driven strategies [26]. They continuously observe system behavior, learn detection patterns, and evolve attack techniques to bypass defenses. This creates a co-evolutionary arms race, where attackers and defenders engage in continuous strategy adaptation and counter-adaptation [27-38].

For example, a defender may deploy a machine learning-based anomaly detector, only to find that an attacker uses generative adversarial networks (GANs) to craft undetectable attack patterns. In response, the defender might improve detection with adversarial training, prompting the attacker to shift to model poisoning or federated learning attacks [39-40]. This adversarial dynamic demands defense strategies that evolve as quickly as the threats they face.

Recognizing this challenge, this review proposes to redefine the cybersecurity paradigm for Power CPS. Rather than viewing defense as a static barrier, we advocate for dynamic, co-evolutionary security frameworks that continuously adapt to adversarial strategies, leveraging AI not only as a target of attacks but as a strategic defense enabler.

1.3. Contributions and Structure of This Review

This review aims to fill critical gaps in the existing literature by:

- Characterizing the emerging class of AI-augmented cyber-physical threats, including adversarial ML, RL-based attacks, and coordinated multi-agent strategies.
- Analyzing the limitations of current AI-based defense mechanisms, particularly their vulnerability to adversarial manipulation and their lack of adaptive resilience.
- Proposing co-evolutionary defense frameworks, integrating adversarial game theory, digital twins, and human-in-the-loop mechanisms to counter evolving threats.
- Identifying research gaps, including theoretical, experimental, and organizational challenges in realizing dynamic, adaptive cyber-physical security for energy systems.

The remainder of this review is organized as follows. Section 2 analyzes the evolution of adversarial threats in AI-enabled Power CPS, moving beyond traditional attack models. Section 3 reviews the current state of AI-based defense capabilities and their limitations. Section 4 shifts focus to the offensive use of AI, exploring how adversaries leverage AI to craft advanced attack strategies. Section 5 introduces co-evolutionary defense frameworks that integrate adaptive learning, game-theoretic reasoning, and human oversight. Section 6 outlines future research directions and collaboration needs. Finally, Section 7 concludes the review with key insights and a call to action for advancing dynamic, co-evolutionary security paradigms in Power CPS.

2. Adversarial Threat Evolution in AI-Enabled Power CPS

 doi:10.20944/preprints202506.1885.v1

The integration of AI into Power CPS has enabled transformative operational improvements, yet it has simultaneously amplified the sophistication, stealth, and impact potential of cyber-physical attacks [41-43]. This section systematically analyzes how adversarial threat models have evolved—from conventional cyber-physical exploits to multi-agent, AI-enhanced, and game-theoretically optimized attack strategies—posing unprecedented challenges to grid security and resilience.

2.1. Traditional Cyber-Physical Attacks Revisited

Early cyber-physical attacks on power systems predominantly targeted three interrelated domains: the cyber layer, the physical infrastructure, and their cyber-physical interactions [44]. At the cyber level, attackers exploited vulnerabilities SCADA systems, energy management systems, and substation automation platforms [45]. Techniques such as network intrusions, denial-of-service (DoS) attacks, and credential theft were commonly employed to gain unauthorized access, disrupt communication flows, or escalate privileges within control networks [46-47].

On the physical side, adversaries focused on direct manipulation of sensors and actuators. Sensor tampering or data injection could mislead state estimation algorithms, while malicious control of physical components—such as remotely opening breakers—could cause service interruptions or equipment damage. Although physical access posed a higher logistical barrier, the consequences of successful intrusions were often severe [48].

Most critically, attackers began exploiting the interplay between cyber and physical domains through cyber-physical interaction exploits [49-51]. One of the most notable techniques was the False Data Injection Attack (FDIA), wherein adversaries manipulated measurement data in a way that remained physically plausible, thereby evading traditional bad data detection mechanisms [52]. Replay attacks also gained prominence, wherein previously recorded valid data was resent to mislead control systems and conceal ongoing intrusions [53-55].

Despite their impact, these early attacks were typically static, manually orchestrated, and reliant on pre-identified system weaknesses. Their success hinged on the attacker's system knowledge and timing precision, which limited their scalability and adaptability. As a result, traditional cyber-physical threats, while disruptive, lacked the dynamic learning and strategic sophistication exhibited by today's AI-enhanced adversaries [56].

2.2. The Emergence of AI-Augmented Attacks

As AI technologies are increasingly integrated into power grid operations and defense, adversaries have begun to exploit these same tools to develop more adaptive, stealthy, and intelligent cyber-physical attacks [57-59]. AI-augmented threats rely on data-driven modeling to understand system behavior, identify detection blind spots, and optimize attack timing and impact under operational constraints. Among the most significant developments in this domain is the rise of adversarial machine learning. By targeting machine learning models used for anomaly detection or forecasting, attackers can craft inputs that evade detection yet induce unsafe system states [60-61]. Poisoning attacks, for instance, involve contaminating training datasets to impair model performance, while backdoor attacks embed hidden triggers that cause misclassifications when activated under specific conditions [62-64].

Reinforcement learning (RL) further amplifies attack sophistication by enabling adversaries to train agents that learn optimal attack policies through interaction with simulated environments [65]. These agents can identify the most disruptive moments to launch false data injection attacks (FDIAs), sequentially manipulate control points to induce cascading failures, or exploit feedback loops by learning how system controls respond to perturbations. In parallel, GANs empower attackers to produce highly realistic yet falsified sensor data [66]. These synthetic signals are carefully crafted to mimic legitimate system behavior, allowing attacks to bypass traditional AI-based detectors through statistical camouflage [67].

 doi:10.20944/preprints202506.1885.v1

The decentralized nature of federated learning (FL)—intended to protect data privacy by distributing model training across edge entities—also introduces new vulnerabilities [68-70]. Adversaries can poison model updates, conduct Sybil attacks by impersonating multiple benign nodes, or even infer sensitive data from model parameters shared during the aggregation process [71-72]. The same mechanisms that FL relies on to enable collaboration can thus be weaponized to undermine global model integrity [73].

2.3. Multi-Agent, Multi-Stage, and Coordinated Threat Dynamics

In addition to individual AI-enhanced attacks, modern adversaries increasingly adopt coordinated strategies that span multiple agents, stages, and system layers [74]. These multi-agent threats involve distributed entities working in concert to simultaneously compromise cyber and physical infrastructure, synchronize attacks across geographically dispersed assets, and exploit operational interdependencies between market dynamics, control hierarchies, and communication protocols. Such coordinated behavior enables system-wide disruption while minimizing the risk of early detection [75-76].

These campaigns often unfold in carefully planned stages. Initially, attackers engage in reconnaissance to gather information about system topology, data flows, and defensive mechanisms. This is followed by privilege escalation through credential theft or exploitation of software vulnerabilities, allowing adversaries to gain access to critical nodes. Once entrenched, they may engage in stealthy manipulation—gradually introducing false data or altering control signals to destabilize the system. At the peak of the campaign, coordinated actions aim to maximize operational impact, such as triggering cascading outages or disrupting electricity market prices. Even after achieving their primary objectives, attackers may persist within the system, erasing traces of compromise and maintaining access for future operations [77-78].

These attack strategies often traverse multiple layers—cyber infrastructure, physical assets, market mechanisms, and human operators—forming cross-layer threats that are exceedingly difficult to detect [79]. Signals indicating an attack may be weak or fragmented across disparate domains, requiring defenders to perform real-time correlation across system components, time scales, and organizational boundaries. The rise of such coordinated, AI-driven, and cross-domain adversaries underscores the inadequacy of static or isolated defense mechanisms and reinforces the need for holistic, adaptive, and intelligence-driven security architectures [80-82].

2.4. Summary of Evolving Adversarial Dynamics

To highlight the escalating sophistication of threats facing Power CPS, Table 1 contrasts traditional cyber-physical attacks with emerging AI-enhanced adversarial strategies across key dimensions.

Table 1. Comparative Evolution of Threats: Traditional vs. AI-Enhanced Attacks.

Threat Evolution Dimension	**Traditional Attacks**	**AI-Enhanced Attacks**
Attack Design	Scripted, rule-based	Data-driven, optimized using AI
Adaptability	Static strategies	Dynamic, adaptive, and learning-based
Coordination	Isolated attacks on single points	Multi-agent, cross-layer, and synchronized campaigns
Attack Objectives	Disrupt single processes or components	Maximize system-wide impact with stealth and persistence
Defense Evasion Techniques	Basic obfuscation or replay attacks	GAN-based evasion, adversarial ML, reinforcement learning

 doi:10.20944/preprints202506.1885.v1

Threat Evolution Dimension	Traditional Attacks	AI-Enhanced Attacks
Exploitation of AI-Defense Limitations	Rare or non-existent	Model poisoning, backdoors, federated learning manipulation

This emerging landscape underscores the urgent need for defense paradigms that move beyond static detection. Defenders must anticipate adversarial learning, coordinated attack policies, and game-theoretic behavior that continuously evolve in response to deployed defenses [83].

3. AI for Defense: Current Capabilities and Limitations

While AI-augmented attacks present unprecedented threats, the same technological advancements can also be harnessed to enhance defense. Power CPS increasingly adopt AI to detect anomalies, orchestrate coordinated responses, and learn optimal defense policies [84-85]. However, these defenses are not without limitations. This section critically reviews the current state of AI-based defense mechanisms, including anomaly detection, RLfor defense, FL for distributed coordination, and the inherent challenges of interpretability, robustness, and adaptability.

3.1. Machine Learning-Based Anomaly Detection

AI-powered anomaly detection has emerged as one of the most widely adopted defense strategies in Power CPS [86-88]. By learning from historical and real-time operational data, ML models are capable of identifying deviations that may signal cyber-physical intrusions. These approaches include supervised learning methods such as support vector machines, random forests, and neural networks, which rely on labeled attack and benign samples to learn decision boundaries [89-90]. Unsupervised methods, on the other hand, such as clustering and density-based techniques like k-means and DBSCAN, attempt to flag outliers in unlabeled data distributions. Semi-supervised models bridge the two by learning from normal operational data and detecting deviations from it. These techniques eliminate the need for extensive rule-based programming and can scale to complex systems, making them particularly effective in detecting known patterns of abnormal behavior [91-93].

However, ML-based anomaly detectors are not without significant limitations. Grid operations are inherently dynamic and diverse, often triggering false positives that can overwhelm system operators and erode trust in automated alerts [94-95]. In addition, attackers can exploit these models by training surrogate systems that reveal detection blind spots, enabling the generation of adversarial inputs that bypass the detectors. The heavy dependence of ML models on the representativeness of training data also limits their generalization to unseen scenarios [96-98]. Most critically, the black-box nature of many ML architectures hinders interpretability, leaving operators without insights into why a particular alert was generated—thus complicating decision-making and response prioritization [99].

3.2. Reinforcement Learning for Autonomous Defense

RLoffers a promising pathway toward adaptive, self-learning defense mechanisms in Power CPS [100-102]. By interacting with a simulated environment, RL agents can learn optimal policies through trial-and-error feedback. This capacity allows them to dynamically tune anomaly detection thresholds based on evolving system conditions, orchestrate automated mitigation strategies such as rerouting power flows or isolating compromised nodes, and balance defense actions with system performance goals to minimize service disruption [103-104]. The ability of RL to adapt over time makes it especially attractive in confronting previously unseen or evolving attack tactics [105].

Nevertheless, RL-based defenses introduce new and serious risks. The effectiveness of learned policies is often limited by the realism of the training environment, and a significant simulation-to-reality gap can arise when deployed in actual grid settings [106]. Moreover, RL systems typically

 doi:10.20944/preprints202506.1885.v1

require a large number of iterations to converge on effective policies, which may not be practical in rapidly evolving threat landscapes [107]. These systems are also vulnerable to adversarial manipulation—attackers can poison the training environment or guide the learning process toward ineffective or even harmful behaviors. Perhaps most concerning is the lack of human oversight; fully autonomous RL agents, if not carefully constrained, may take actions that compromise system safety or contradict operator intentions, especially in high-stakes scenarios [108-110].

3.3. Federated Learning and Collaborative Defense

FL has emerged as an attractive approach for distributed defense coordination, particularly in privacy-sensitive and geographically dispersed grid environments [111-112]. In FL, individual nodes such as substations or microgrids train local models on their own data and share only model updates with a central aggregator. This decentralized framework supports privacy preservation by keeping raw data local, enhances scalability by enabling parallel learning across heterogeneous devices, and improves context-specific detection by tailoring models to localized operational characteristics [113-115].

Yet, despite these advantages, FL introduces several unique vulnerabilities. Malicious participants can submit corrupted model updates, a strategy known as model poisoning, which can degrade the accuracy and robustness of the global model [116]. Additionally, an attacker may launch Sybil attacks by posing as multiple legitimate nodes, thereby dominating the aggregation process and steering the learning trajectory toward compromised outcomes [117-118]. The iterative exchange of model updates also incurs significant communication overhead, particularly in large-scale or latency-sensitive deployments. Compounding these issues is the limited explainability of the resulting aggregated models, which often function as complex ensembles that are difficult for operators to interpret or audit—thus undermining trust and real-time validation [119-121].

3.4. Challenges in Interpretability, Robustness, and Adaptation

Across all AI-based defense strategies, systemic challenges persist in the domains of interpretability, robustness, and adaptability [122-124]. Many machine learning and deep learning models operate as opaque black boxes, offering little transparency into how or why certain predictions are made [125-126]. This lack of interpretability impairs operator confidence and hinders the adoption of human-in-the-loop defense architectures, where timely and informed decision-making is critical [127-129].

Equally problematic is the vulnerability of AI defenses to adversarial manipulation. Sophisticated attackers can craft inputs specifically designed to exploit model blind spots (evasion attacks), corrupt the training process via poisoned data or manipulated updates (model poisoning), or embed hidden triggers that activate malicious behavior under specific conditions (backdoor attacks) [130-131]. These threats erode the reliability of AI defenses and expose systems to subtle, long-term degradation in security performance [132-134].

Furthermore, most existing AI models are static in nature, trained on historical data and incapable of adapting to rapidly changing attack landscapes [135]. Overfitting to past scenarios renders these models ineffective against novel threats. While continual learning and online adaptation frameworks have been proposed, their deployment in operational Power CPS remains limited due to concerns over stability, system compatibility, and the lack of well-established safety guarantees [136-138].

To evaluate the current landscape of intelligent defense approaches in Power CPS, Table 2 summarizes representative AI-driven strategies, along with their strengths and inherent limitations.

 doi:10.20944/preprints202506.1885.v1

Table 2. Evaluation of AI-Enabled Defense Strategies in Power CPS.

Defense Strategy	Strengths	Limitations
ML-Based Anomaly Detection	Effective for known patterns; scalable to large datasets	High false positives; poor explainability; blind spots
Reinforcement Learning for Defense	Learns adaptive policies; minimizes operational impact	Simulation-reality gap; slow convergence; risk of unsafe decisions
Federated Learning for Collaboration	Privacy-preserving; scalable; context-specific learning	Model poisoning; Sybil attacks; communication overhead
General Challenges	Enhances detection and response automation	Vulnerable to adversarial attacks; lacks explainability; poor adaptation

In summary, while AI offers promising capabilities for enhancing Power CPS defense, these mechanisms are not inherently resilient to the very adversarial tactics they aim to counter. Without co-evolutionary adaptation, explainability, and human oversight, AI-based defenses risk being outpaced by dynamic, AI-augmented attackers [139-140].

4. Adversarial AI for Attack Strategy Evolution

As defenders increasingly rely on AI to secure Power CPS, adversaries have also begun leveraging AI to accelerate, optimize, and evolve their offensive strategies [141-143]. This section explores how attackers exploit advanced AI techniques—such as GANs, Reinforcement Learning (RL), model poisoning, and federated learning manipulation—to craft stealthy, adaptive, and highly impactful cyber-physical attacks.

4.1. GANs for Stealthy Attack Crafting

GANs represent a powerful tool for adversaries seeking to craft stealthy and evasive attack strategies [144-145]. A typical GAN architecture involves two neural networks—the generator and the discriminator—engaged in a minimax competition. The generator aims to produce synthetic data indistinguishable from real observations, while the discriminator attempts to distinguish genuine data from fake inputs [146]. Through this adversarial training loop, GANs can learn to generate highly realistic data patterns that mimic legitimate system behavior [147-149].

In the context of Power CPS, attackers can exploit GANs to generate falsified measurement data—such as voltage, current, or frequency readings—that conform to physical laws while steering the system toward unsafe operating states [150]. These data streams may pass unnoticed through conventional validation checks or anomaly detectors. Furthermore, by learning the statistical distribution of normal operations, GANs can craft attack signatures that are camouflaged within routine noise, thereby reducing the likelihood of detection [151-153]. Such techniques are especially effective in evading ML-based defenses, as the generated data targets specific weaknesses in model generalization.

However, the use of GANs in cyber-physical attacks also presents challenges. Training effective models requires access to high-quality system data, which may not always be available to the attacker [154-156]. Overfitting to specific system states can limit a GAN's ability to generalize under dynamic conditions. Moreover, defenders may employ advanced countermeasures—such as multi-source data fusion, temporal consistency checks, or physics-aware anomaly detection—to expose GAN-generated anomalies [157-159]. These open questions highlight the evolving arms race between generative deception and intelligent detection [160].

4.2. Reinforcement Learning-Based Attack Policy Optimization

Adversarial RLenables attackers to develop sophisticated, sequentially optimized attack strategies by interacting with a simulated representation of the target system [161-163]. In this

 doi:10.20944/preprints202506.1885.v1

framework, the attacker is modeled as an intelligent agent that explores the system's response space and exploits learned policies to maximize disruption while minimizing the risk of detection. By defining custom reward functions—such as maximizing power imbalances, inducing cascading failures, or manipulating market dynamics—RL agents can autonomously discover highly effective attack trajectories [164-165].

Such agents can learn to time false data injection attacks to coincide with vulnerable operational windows, such as peak loads or renewable variability [166]. They can also coordinate actions across multiple control points to exploit system interdependencies and trigger cascading failures [167]. Additionally, adversaries may adapt their strategies in response to observed defense behaviors, creating a dynamic co-evolution in which the attacker continuously refines its policy to stay ahead of the defender [168-170].

Yet, despite their power, RL-based attacks are constrained by practical limitations. Training effective attack policies requires access to high-fidelity simulators that accurately capture the cyber-physical dynamics of real-world systems [171]. Even with such simulators, policies developed in virtual environments may not transfer reliably to live systems due to noise, latency, or model inaccuracies [172]. Furthermore, the computational overhead and domain expertise required to train robust RL agents can be substantial. These barriers make RL-based attacks feasible primarily for well-resourced adversaries but also underscore the growing need for adaptive, simulation-informed defense planning [173-174].

4.3. Model Poisoning and Backdoor Attacks on AI Defenses

Model poisoning and backdoor attacks represent insidious threats that compromise the integrity of AI-based defense mechanisms [175]. In model poisoning, attackers subtly inject malicious data into the training process, biasing the learned model to misclassify targeted attack patterns as normal behavior [176-177]. In FL settings, this can be achieved by submitting corrupted model updates during aggregation, degrading the global model's performance across the network [178].

Backdoor attacks take this one step further by embedding hidden triggers into the model itself [179-180]. When these specific inputs are encountered, the model behaves incorrectly—either ignoring malicious activity or raising false alarms for benign behavior [181]. For example, a compromised reinforcement learning controller might operate normally under most conditions but misbehave catastrophically when it receives a particular control signal. Similarly, a poisoned anomaly detector may consistently overlook threats targeting specific substations or assets [182-183].

Detecting and mitigating these threats remains a significant challenge. Backdoors and poisoned models often appear benign under standard evaluation protocols, and their activation patterns may remain dormant until explicitly triggered [184-185]. Although defensive techniques such as robust aggregation, differential privacy, or adversarial training can offer partial protection, they are far from foolproof. The covert nature of these attacks—and their compatibility with collaborative learning paradigms like FL—makes them especially dangerous in distributed Power CPS environments [186].

4.4. Exploiting Federated Learning in Distributed Defense

Federated learning, while beneficial for preserving data privacy and enabling decentralized collaboration, presents a fertile ground for adversarial exploitation [187-188]. Attackers can leverage the very principles that make FL attractive to subvert model integrity and leak sensitive operational insights [189-191]. One prominent strategy is the Sybil attack, in which a single adversary creates multiple identities to dominate the model aggregation process, thereby skewing the global model's behavior in favor of the attacker's goals.

Another subtle method is model drift manipulation, where malicious clients incrementally introduce degradation into the global model over successive training rounds. This gradual erosion of accuracy may go unnoticed until significant damage has already occurred [192-193]. Additionally, attackers may analyze shared model updates to infer private information about other nodes, raising serious concerns about privacy leakage [194].

 doi:10.20944/preprints202506.1885.v1

For instance, a compromised FL aggregator could tamper with aggregation logic to prioritize malicious updates, extract operational trends from model patterns, or inject misleading updates from fake clients to destabilize the overall learning process. These exploits highlight the urgent need for resilient defenses tailored to the distributed nature of FL [195-196].

Mitigation strategies include secure aggregation protocols such as Byzantine-resilient methods, which can tolerate a bounded number of adversarial clients without compromising the model [197]. Enforcing strong client authentication mechanisms helps prevent Sybil-based infiltration, while anomaly detection algorithms applied to model updates can identify suspicious contributions. Nevertheless, implementing these defenses at scale remains an open research challenge, especially when balancing privacy, scalability, and real-time response [198-200].

The emergence of adversarial AI techniques poses significant threats to Power CPS by enhancing the stealth, coordination, and effectiveness of cyber-physical attacks. Table 3 outlines key adversarial AI methods, their offensive capabilities, and corresponding defense implications.

Table 3. Adversarial AI Techniques and Their Implications for Power CPS Defense.

Adversarial AI Technique	Offensive Capabilities	Defense Implications
GAN-Based Attack Crafting	Generates realistic but malicious data to evade detection	Requires temporal and cross-source anomaly validation
RL-Based Attack Optimization	Learns optimal multi-stage attack policies under operational constraints	Demands adaptive, game-theoretic defense strategies
Model Poisoning and Backdoors	Corrupts AI defenses to ignore or misclassify attacks	Necessitates secure training and model validation pipelines
Federated Learning Exploits	Manipulates distributed learning processes to degrade defense performance	Requires Byzantine-resilient aggregation and participant authentication

The growing use of AI by attackers highlights the need for co-evolutionary defense frameworks that anticipate and counter these strategies. The next section will introduce such frameworks, leveraging adversarial game theory, digital twin-based co-simulation, and human-in-the-loop adaptation to build dynamic and resilient security architectures.

5. Co-Evolutionary Defense Frameworks

Given the dynamic and evolving nature of AI-augmented cyber-physical attacks, Power CPS require defense paradigms that continuously adapt to adversarial strategies [201-203]. Static defense models quickly become obsolete as attackers evolve their techniques. In response, this section introduces co-evolutionary defense frameworks, where defenders leverage game-theoretic reasoning, digital twin-based simulation platforms, and human-in-the-loop decision support to engage in ongoing strategy refinement and resilience validation.

5.1. Adversarial Game Theory for Cyber-Physical Defense Design

Game theory offers a rigorous mathematical framework to analyze strategic interactions between adversaries and defenders in Power CPS environments [205-207]. It enables defenders to anticipate how attackers might adapt their strategies in response to deployed defenses and to proactively design countermeasures that are robust even under worst-case assumptions [208]. Importantly, game-theoretic models can help balance defense investments against operational costs and risk exposure, supporting more informed and strategic resource allocation [209-210].

Various forms of adversarial games have been applied to power system security. In static games, attackers and defenders select their strategies simultaneously, offering insights into one-shot

 doi:10.20944/preprints202506.1885.v1

interactions under complete information [211]. Sequential games extend this framework by modeling turn-based interactions, reflecting more realistic attack-defense-response cycles [212-213]. Repeated games capture long-term engagements where both sides adapt their strategies over time. Bayesian games, on the other hand, incorporate uncertainty about the opponent's capabilities, goals, or knowledge, better reflecting real-world asymmetries [214-215].

These models have found concrete application in Power CPS through resource allocation games—optimizing the placement of scarce monitoring or encryption tools—and timing games that identify optimal moments for deploying defenses or triggering attacks. Information disclosure games have also emerged, evaluating the trade-offs between transparency and security risk when revealing system status to external stakeholders [216].

Despite their promise, applying game theory to real-world CPS defense is not without challenges. Modeling high-dimensional, multi-agent systems with realistic assumptions is computationally intensive, and deriving stable equilibria can be intractable [217]. Moreover, the simplifications required to make game models analytically solvable often fail to capture the full behavioral complexity of adaptive, intelligent adversaries [218-220].

5.2. Digital Twins and Co-Simulation Platforms for Dynamic Validation

Digital twins offer a transformative capability for dynamic security validation by creating real-time, high-fidelity digital representations of physical power systems [221-223]. In the context of cyber-physical defense, digital twins can concurrently simulate physical grid dynamics, cyber layer behaviors, and even human operator responses under varying attack scenarios. This comprehensive representation allows for detailed exploration of how AI-augmented attacks propagate and how defense strategies might adapt in response [224].

A key advantage of digital twins lies in their capacity to facilitate co-evolutionary strategy testing in a safe and controlled environment [225]. Unlike traditional static testbeds, digital twins can model dynamic changes over time, allowing defenders to simulate and refine adaptive strategies before deploying them in operational systems. Moreover, these simulations provide valuable insights into the trade-offs between defense effectiveness and system performance, enabling more precise operational impact assessments [226].

To achieve this, co-simulation platforms integrate multiple domain-specific simulation engines, allowing for the synchronized modeling of cyber operations, physical infrastructure, market behaviors, and human decision-making. This multi-layer integration is critical for analyzing cross-domain interactions, training operators in realistic environments, and coordinating defense across diverse subsystems [227-228].

However, large-scale co-simulation comes with significant implementation challenges. Achieving scalability requires high computational capacity, and maintaining model fidelity is essential to avoid overlooking critical dynamics. Synchronizing heterogeneous simulation platforms—each with its own time resolution, data structures, and communication protocols—adds another layer of technical complexity that must be addressed to ensure reliable system behavior emulation [229-230].

5.3. Adaptive Defense Mechanisms with Human-in-the-Loop

While AI has demonstrated great promise in automating defensive tasks, human oversight remains indispensable in Power CPS defense [231-233]. Operators play a critical role in interpreting ambiguous alerts, making context-aware decisions during high-stakes situations, and managing trade-offs between security, reliability, and economic performance. Incorporating human-in-the-loop (HITL) mechanisms into adaptive defense systems ensures that automation remains aligned with operational priorities and ethical standards [234-235].

Key enablers of HITL defenses include explainable AI (XAI), which provides human-understandable justifications for AI-generated alerts or decisions. Such interpretability enhances operator trust and facilitates more effective decision-making [236]. Visualization interfaces, such as

 doi:10.20944/preprints202506.1885.v1

operator-centered dashboards, are also essential in distilling complex, multi-domain information into cognitively digestible formats. Interactive decision support tools further allow human operators to simulate the consequences of various defensive actions before execution, improving situational awareness and response precision [237-238].

Additionally, adaptive defense architectures can incorporate human feedback directly into learning loops. For example, operators can annotate AI decisions as correct or flawed, which in turn refines the underlying models and improves future performance. However, achieving the right balance between automation and oversight remains a nuanced challenge [239]. Fully autonomous systems risk initiating unsafe or suboptimal responses without human validation, while overreliance on human review can lead to slow reaction times in time-critical scenarios. Hybrid frameworks that automate routine responses while escalating complex cases to human operators offer a promising path forward [240].

5.4. Metrics for Evaluating Co-Evolutionary Resilience

Traditional performance metrics—such as accuracy or detection rates—are insufficient to assess the effectiveness of adaptive cyber-physical defense systems [241]. Co-evolutionary resilience requires a broader set of evaluation metrics that capture not only the ability to detect attacks, but also to adapt over time, preserve operational continuity, and maintain user trust [242].

One critical metric is defense adaptability, which reflects how well a system can modify its detection or response strategies in reaction to changing adversarial behavior [243]. Operational continuity, or the system's ability to deliver essential services under attack, is equally important—especially for mission-critical infrastructures. Cost-effectiveness metrics help evaluate whether the benefits of a given defense strategy justify the required investment in terms of resources, manpower, or downtime. Finally, metrics for explainability and trust gauge how transparent and reliable the defense system is from an operator's perspective, which is crucial for long-term human-AI collaboration [244].

Simulation-based benchmarks provide practical means to evaluate these dimensions. Co-evolution curves, for instance, visualize the shifting performance of both attackers and defenders over time, offering insight into their dynamic interplay. Time-to-compromise versus time-to-response metrics quantify the temporal efficiency of threat detection and mitigation. Service degradation profiles, meanwhile, capture the operational impact of sustained attacks under various defense configurations. Together, these metrics offer a holistic view of system resilience and inform the development of next-generation adaptive security architectures [245].

To counter sophisticated adversarial threats in Power CPS, advanced defense components are being integrated across technical and human dimensions. Table 4 summarizes key components, their core capabilities, and the challenges they face in deployment and validation.

Table 4. Defense Components and Their Deployment Challenges in Power CPS.

Defense Component	Capabilities	Challenges
Adversarial Game Theory	Models strategic interactions; optimizes defense resource allocation	Model realism and computational complexity
Digital Twins and Co-Simulation	Realistic multi-layer validation; operator training platforms	Scalability and integration complexity
Human-in-the-Loop Adaptive Defense	Combines AI automation with human judgment and oversight	Balancing speed and reliability in decision-making

 doi:10.20944/preprints202506.1885.v1

Defense Component	Capabilities	Challenges
Resilience Evaluation Metrics	Quantifies dynamic defense effectiveness and operational impacts	Defining standardized, actionable, and validated resilience metrics

In summary, co-evolutionary defense frameworks provide a holistic, dynamic, and adaptive approach to securing Power CPS. By anticipating adversarial learning, validating strategies in realistic simulations, and integrating human expertise, these frameworks offer a promising path toward long-term resilience against AI-augmented cyber-physical threats.

6. Research Gaps and Future Directions

While the concept of co-evolutionary cyber-physical defense provides a promising framework for safeguarding Power CPS, its realization remains largely theoretical. Bridging the gap between research and real-world deployment requires addressing several unresolved theoretical, experimental, organizational, and policy challenges [246]. This section identifies key research gaps and proposes future directions to advance practical, scalable, and trustworthy co-evolutionary defense for AI-enabled Power CPS.

6.1. Theoretical Gaps in Adversarial Dynamics Modeling

Despite growing interest in co-evolutionary cyber defense, the theoretical foundations for modeling attacker-defender dynamics in Power CPS remain underdeveloped [247]. A major limitation in existing literature is the prevalent assumption of static or one-sided adaptation, which overlooks the bidirectional learning process that characterizes modern adversarial engagements. In reality, attackers and defenders iteratively refine their strategies in response to each other's actions, necessitating the formalization of game-theoretic models that explicitly capture long-term strategy evolution. Such models must go beyond traditional equilibria and incorporate continuous adaptation, strategic uncertainty, and incomplete information [248-249].

Moreover, current models often treat attackers and defenders as isolated agents, failing to account for coordination among multiple adversarial or defensive actors [250]. In practice, both sides may involve distributed agents operating across cyber, physical, and market layers. Extending existing frameworks to multi-agent settings is essential for capturing the true complexity of cyber-physical interactions. Additionally, theoretical models rarely account for human behavior, which plays a central role in both defense operations and attacker deception. Incorporating cognitive factors such as bias, fatigue, and varying expertise into strategic modeling can enhance realism and improve the applicability of theoretical insights.

Another key gap lies in the use of overly simplistic reward and cost functions. Many models are optimized for narrow objectives such as maximizing detection accuracy or minimizing attack success probability, without considering operational trade-offs or stakeholder-specific priorities [251]. Realistic frameworks must incorporate multi-objective optimization that balances system resilience, service continuity, security investment, and adversary incentives. Accounting for both economic motivations and resource constraints on both sides will be critical for building models that are not only theoretically rigorous but also operationally relevant [252].

6.2. Experimental Gaps in Large-Scale, Realistic Testing

A major barrier to the practical advancement of co-evolutionary defense strategies is the lack of high-fidelity experimental validation [253]. Most proposed techniques are evaluated within small-scale testbeds or idealized simulations that fail to capture the operational complexity of real-world Power CPS environments. In particular, these simplified settings often ignore the interdependencies between cyber, physical, and market layers, neglect the role of human operator behavior, and exclude cross-sectoral influences from related infrastructures such as telecommunications or logistics.

 doi:10.20944/preprints202506.1885.v1

To bridge this gap, there is an urgent need to develop modular, scalable, and open-access cyber-physical testbeds that enable realistic evaluation of attack-defense interactions [254]. These platforms should integrate digital twins, adversarial agents, and dynamic control systems to faithfully replicate the conditions under which co-evolutionary strategies must operate. Only by stress-testing defense mechanisms in such environments can researchers identify failure modes, validate adaptability, and assess resilience under operational constraints [255].

In addition to testbed limitations, progress is also hindered by a lack of standardized datasets and benchmarking protocols. Few publicly available datasets capture AI-driven attack scenarios or the evolution of coordinated cyber-physical campaigns. Without access to shared benchmarks, it becomes difficult to compare different strategies or reproduce key findings. Community-driven initiatives to curate, anonymize, and share realistic data—alongside agreed-upon evaluation frameworks—are essential to advance the field with scientific rigor and transparency.

6.3. Organizational and Policy Challenges in Adaptive Defense Adoption

Even as co-evolutionary defense frameworks mature in theory and experimentation, their practical adoption faces considerable organizational resistance. Many utility operators and regulatory bodies remain hesitant to deploy dynamic or AI-driven defenses, instead favoring static, proven technologies that align with existing compliance requirements. This conservatism is driven by risk aversion in mission-critical environments, where the cost of unanticipated behavior is high, and by regulatory regimes that emphasize checklist-style verification over adaptive performance [256].

Overcoming this institutional inertia requires both technical demonstration and policy innovation. Pilot deployments and field trials must be conducted to showcase the operational advantages of adaptive security mechanisms—particularly in detecting advanced persistent threats or mitigating multi-stage campaigns. In parallel, regulatory frameworks need to evolve to recognize and certify AI-based security solutions, potentially through sandbox environments or tiered certification models that incentivize gradual adoption.

Another persistent obstacle is the fragmentation of defense responsibilities across sectors, regions, and stakeholders [257]. AI-augmented attacks increasingly target interconnected infrastructures, yet current defense efforts remain siloed between electric utilities, telecom providers, transportation agencies, and other actors. Likewise, cross-border coordination among national grid operators and international regulators remains limited. Future research must address governance mechanisms that facilitate shared threat intelligence, coordinated response exercises, and joint development of secure architectures across both sectoral and geographic boundaries.

6.4. Toward Fully Autonomous but Human-Supervised Defense Ecosystems

As the complexity and velocity of cyber-physical threats continue to increase, fully autonomous defense agents are being explored as a means of enabling rapid, scalable, and real-time responses. These agents could continuously monitor system states, execute mitigation strategies, and learn from unfolding attack patterns with minimal human intervention [258]. However, while promising in theory, fully autonomous systems introduce their own risks—including unintended behavior due to misinterpreted inputs, opaque decision-making processes, and a lack of accountability in critical scenarios.

To mitigate these risks, future architectures must embrace hybrid defense ecosystems that blend automation with human oversight. In such frameworks, AI agents can manage routine or time-sensitive tasks, such as isolating compromised nodes or rerouting communication paths, while reserving high-stakes or ambiguous decisions for human operators [259]. This division of labor ensures both responsiveness and accountability, particularly when system stability or public safety is at risk. Feedback loops that allow human input to inform AI retraining—and vice versa—will be essential to maintaining alignment between human intent and machine behavior over time [260].

 doi:10.20944/preprints202506.1885.v1

Developing such systems also requires new research into human-AI teaming. Key questions include how to optimize task allocation between humans and machines, how to design interfaces that provide intuitive and actionable explanations of AI reasoning, and how to train operators to effectively interact with adaptive, learning-enabled systems. Without thoughtful integration of human cognition, ethical reasoning, and transparent design, autonomous security systems may remain untrusted, underutilized, or even dangerous when deployed at scale [261].

Despite advancements in federated and AI-driven defense strategies, several critical gaps remain across theoretical, experimental, organizational, and human-integration dimensions. Table 5 outlines these gaps and highlights promising directions for future research and implementation.

Table 5. Key Gaps and Future Directions for Federated Defense in Power CPS.

Category	Key Gaps	Future Directions
Theoretical Modeling	Limited formalization of co-evolutionary dynamics and multi-agent adversarial games	Develop mathematically grounded, multi-objective, and human-aware models
Experimental Validation	Lack of high-fidelity testbeds, data availability, and standardized benchmarks	Build open, scalable testbeds and community-driven evaluation frameworks
Organizational and Policy Adoption	Resistance to dynamic defense; fragmented cross-sector coordination	Develop adaptive security policies and cross-sector governance structures
Human-AI Integration	Risks of full autonomy; lack of human-AI teaming frameworks	Design hybrid human-AI defense ecosystems with explainable interfaces

In summary, realizing co-evolutionary cyber-physical defense requires multi-disciplinary research, operational validation, and cross-sector collaboration. The next and final section will conclude this review with a synthesis of key insights and a strategic call to action.

7. Conclusion

The integration of AI into Power CPS has fundamentally reshaped both the capabilities and the vulnerabilities of modern energy infrastructure. This review has systematically examined how adversaries exploit AI techniques—such as Generative Adversarial Networks, Reinforcement Learning, and federated learning manipulation—to craft adaptive and stealthy attacks. In response, it has evaluated the limitations of current AI-based defenses and proposed a co-evolutionary defense paradigm that incorporates game-theoretic reasoning, digital twin-based validation, and human-in-the-loop adaptability. Key research challenges were identified across theoretical modeling, experimental testing, organizational readiness, and human-AI collaboration, offering a comprehensive roadmap for advancing resilient Power CPS security.

Looking ahead, building secure and trustworthy Power CPS requires coordinated action across disciplines, sectors, and policy domains. Future efforts should focus on developing scalable testbeds, fostering international research collaboration, and embedding explainable AI into operational workflows. By embracing dynamic, co-evolutionary defense architectures, the energy sector can move beyond static protection models and proactively safeguard critical infrastructures against evolving cyber-physical threats.

 doi:10.20944/preprints202506.1885.v1

References

1. Hassan Y G, Collins A, Babatunde G O, et al. AI-powered cyber-physical security framework for critical industrial IoT systems[J]. Machine learning, 2023: 27: 1158-1164.
2. Whig P, Aggarwal A, Ganeshan V, et al. AI for Secure and Resilient Cyber-Physical Systems[M]//Artificial Intelligence Solutions for Cyber-Physical Systems. Auerbach Publications, 40-63.
3. Qu Z, Zhao T, Zhang Y, et al. Determination Method of Network Risk Propagation Threshold in Power CPS Based on Percolation Theory[J]. Automation of Electric Power Systems, 2020, 44(4): 16-23.
4. Li Y, Yang Z, et al. Optimal scheduling of an isolated microgrid with battery storage considering load and renewable generation uncertainties[J]. IEEE Transactions on Industrial Electronics, 2018, 66(2): 1565-1575.
5. Chang J, Li Z, Kaveh M, et al. A Survey on AI-Enabled Attacks and AI-Empowered Countermeasures in Physical Layer[C]//2023 IEEE 9th World Forum on Internet of Things (WF-IoT). IEEE, 2023: 1-7.
6. Zhang F, Huang Z, Kou L, et al. Data Encryption Based on a 9D Complex Chaotic System with Quaternion for Smart Grid[J]. Chinese Physics B, 2023, 32(1): 010502.
7. 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.
8. Li Y, Wang C, Li G, et al. Improving operational flexibility of integrated energy system with uncertain renewable generations considering thermal inertia of buildings[J]. Energy Conversion and Management, 2020, 207: 112526..
9. Zeng Y, Dong P, Shi Y, et al. Analyzing the co-evolution of green technology diffusion and consumers' pro-environmental attitudes: An agent-based model[J]. Journal of Cleaner Production, 2020, 256: 120384.
10. Lalouani W. AI Cyber Threat in Cyber Physical Systems[D]. University of Maryland, Baltimore County, 2022.
11. Cao J, Wang Q, Qu Z, et al. Method for identifying false data injection attacks in power grid based on improved CNN-LSTM[J]. Electrical Engineering, 2025: 1-26.
12. Kavitha D, Thejas S. Ai enabled threat detection: Leveraging artificial intelligence for advanced security and cyber threat mitigation[J]. IEEE Access, 2024, 12: 173127-173136.
13. Tooki O O, Popoola O M. A critical review on intelligent-based techniques for detection and mitigation of cyberthreats and cascaded failures in cyber-physical power systems[J]. Renewable Energy Focus, 2024: 100628.
14. Jiang Y, Wu S, Ma R, et al. Monitoring and defense of industrial cyber-physical systems under typical attacks: From a systems and control perspective[J]. IEEE Transactions on Industrial Cyber-Physical Systems, 2023, 1: 192-207.
15. 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, 2019, 34: 3651-3660.
16. 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.
17. 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.
18. 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.
19. 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.
20. Asevameh I O, Dopamu O M, Adesiyan J S. Enhancing resilience and security in the US power grid against cyber-physical attacks[J]. World Journal of Advanced Research and Reviews, 2024, 22(2): 1043-1052.
21. 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.
22. Lee S, Chae J, Jeon H, et al. Cyber-Physical AI: Systematic Research Domain for Integrating AI and Cyber-Physical Systems[J]. ACM Transactions on Cyber-Physical Systems, 2025, 9(2): 1-33.

 doi:10.20944/preprints202506.1885.v1

23. Mala D J, Dhanapal A C T A, Sthapit S, et al. Integrating AI Techniques into the Design and Development of Smart Cyber-Physical Systems: Defense, Biomedical, Infrastructure, and Transportation[M]. CRC Press, 2025.
24. Yao P, Yan B, Yang Q. Game Theoretical Decision-Making of Dynamic Defense in Cyber-Physical Power Systems under Cyber-Attacks[J]. ACM Transactions on Cyber-Physical Systems, 2025, 9(2): 1-21.
25. Yu Z. Cyber-Physical Security Through the Lens of AI-Enabled Systems[D]. Washington University in St. Louis, 2025.
26. Abdalla* A S, Tang B, Marojevic V. AI at the Physical Layer for Wireless Network Security and Privacy[J]. Artificial Intelligence for Future Networks, 2025: 341-380.
27. 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.
28. Akpolat A N, Kalay M S. Defense Mechanism of PV-Powered Energy Islands Against Cyber-Attacks Utilizing Supervised Machine Learning[J]. Applied Sciences, 2025, 15(9): 5021.
29. Barboni A, Rezaee H, Boem F. Detection of Covert Cyber-Attacks in Interconnected Systems: A Distributed Model-Based Approach[J]. IEEE Transactions on Automatic Control, 2020, 65(9): 3728-3741.
30. Sharma G, Malley D J, Parle D, et al. Analysis and Study of Intelligent Testbed for Safeguarding Nuclear and Defense Industry from AI-Enabled Cyberattacks[C]//2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI). IEEE, 2024, 2: 1-6.
31. 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.
32. Oun A, Wince K, Cheng X. The Role of Artificial Intelligence in Boosting Cybersecurity and Trusted Embedded Systems Performance: A Systematic Review on Current and Future Trends[J]. IEEE Access, 2025.
33. 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.
34. Lian Z, Shi P, Chen M. A Survey on Cyber-Attacks for Cyber-Physical Systems: Modeling, Defense and Design[J]. IEEE Internet of Things Journal, 2024.
35. Rani S, Kataria A, Kumar S, et al. A New Generation Cyber-Physical System: A Comprehensive Review from Security Perspective[J]. Computers & Security, 2024: 104095.
36. Kumari M, Gaikwad M, Chavan S A. A secure IoT-edge architecture with data-driven AI techniques for early detection of cyber threats in healthcare[J]. Discover Internet of Things, 2025, 5(1): 1-16.
37. 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.
38. Degtiareva O, Shyriaieva N, Kuklinova T. Artificial Intelligence Solutions for Cybersecurity in Energy Systems[C]//2024 IEEE International Workshop on Technologies for Defense and Security (TechDefense). IEEE, 2024: 177-182.
39. Wang 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.
40. Wang L, Qu Z, Li Y, 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.
41. 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.
42. 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.
43. Okegbile S D, Gambo I P. Artificial intelligence-driven security framework for internet of things-enhanced digital twin networks[J]. Internet of Things, 2025, 31: 101564.
44. 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.

 doi:10.20944/preprints202506.1885.v1

45. 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.
46. Ijaz A, Raza W, Farooq H, et al. An ai-enabled framework to defend ingenious mdt-based attacks on the emerging zero touch cellular networks[J]. IEEE Network, 2023, 38(1): 228-237.
47. Wang R, Zhou Z, Song J, et al. MORTAR: A Model-based Runtime Action Repair Framework for AI-enabled Cyber-Physical Systems[J]. arXiv preprint arXiv:2408.03892, 2024.
48. 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.
49. Lyu D, Song J, Zhang Z, et al. Autorepair: Automated repair for ai-enabled cyber-physical systems under safety-critical conditions[J]. arXiv preprint arXiv:2304.05617, 2023.
50. 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): 3651-60.
51. Gaba S, Budhiraja I, Kumar V, et al. An innovative multi-agent approach for robust cyber–physical systems using vertical federated learning[J]. Ad Hoc Networks, 2024, 163: 103578.
52. El-Hajj M, Itäpelto T, Gebremariam T. Systematic literature review: Digital twins' role in enhancing security for Industry 4.0 applications[J]. Security and Privacy, 2024, 7(5): e396.
53. 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.
54. Guinea-Cabrera M A, Holgado-Terriza J A. Digital twins in software engineering—a systematic literature review and vision[J]. Applied Sciences, 2024, 14(3): 977.
55. Aghazadeh Ardebili A, Zappatore M, Ramadan A I H A, et al. Digital Twins of smart energy systems: a systematic literature review on enablers, design, management and computational challenges[J]. Energy Informatics, 2024, 7(1): 94.
56. Opranescu V, Ionita A D. Review of Cyber-Physical Systems Modeling with UML, SysML and MARTE[J]. IEEE Access, 2025.
57. Pirbhulal S, Chockalingam S, Abie H, et al. Cognitive digital twins for improving security in IT-OT enabled healthcare applications[C]//International Conference on Human-Computer Interaction. Cham: Springer Nature Switzerland, 2024: 153-163.
58. González J, Barzellay Ferreira da Costa B, Tam V W Y, et al. Integration of 4.0 technologies towards the creation of digital twins[J]. International Journal of Construction Management, 2024: 1-18.
59. Peruzzini M, Bilancia P, Majić T, et al. Human-centric digital twin: a transdisciplinary view[M]//Leveraging Transdisciplinary Engineering in a Changing and Connected World. IOS Press, 2023: 923-932.
60. Li Y, Wei X, Li Y, 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.
61. Ngadi H, Bounceur A, Bezoui M, et al. Toward a digital twin IoT for the validation of AI algorithms in smart-city applications[C]//International Conference on Machine Learning for Networking. Cham: Springer Nature Switzerland, 2023: 108-117.
62. Szymoniak S, Piątkowski J, Kurkowski M. Defense and Security Mechanisms in the Internet of Things: A Review[J]. Applied Sciences (2076-3417), 2025, 15(2).
63. Pourmadadkar M, Lezzi M, Corallo A. Cyber Security for Cyber-Physical Systems in Critical Infrastructures: Bibliometrics Analysis and Future Directions[J]. IEEE Transactions on Engineering Management, 2024.
64. Metibemu O C, Adesokan-Imran T O, Ajayi A J, et al. Developing Proactive Threat Mitigation Strategies for Cloud Misconfiguration Risks in Financial SaaS Applications[J]. Journal of Engineering Research and Reports, 2025, 27(3): 393-413.
65. Li Y, Gao J, Li Y, 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.

 doi:10.20944/preprints202506.1885.v1

66. Zhao Z, Li Z, Xie X, et al. Verify All Traffic: Towards Zero-Trust In-Network Intrusion Detection against Multipath Routing[J]. IEEE Journal on Selected Areas in Communications, 2025.
67. Majeed A, Patni S, Hwang S O. A Comprehensive Analysis of Privacy-Preserving Solutions Developed for IoT-Based Systems and Applications[J]. Electronics, 2025, 14(11): 2106.
68. 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.
69. Routray K, Bera P. ZTAAC: Zero Trust Adaptive Authorization with CP-ABE for Context-Aware Data Protection[C]//2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS). IEEE, 2025: 814-816.
70. He S, 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.
71. Yu J, Shvetsov A V, Alsamhi S H. Leveraging machine learning for cybersecurity resilience in industry 4.0: Challenges and future directions[J]. IEEE Access, 2024.
72. Ullah H, Uzair M, Jan Z, et al. Integrating industry 4.0 technologies in defense manufacturing: Challenges, solutions, and potential opportunities[J]. Array, 2024: 100358.
73. Li Y, Li J, Wang Y, et al. Privacy-preserving spatiotemporal scenario generation of renewable energies: A federated deep generative learning approach[J]. IEEE Transactions on Industrial Informatics, 2022, 18(4): 2310–2320.
74. Rizqi Z U, Chou S Y, Cahyo W N. A simulation-based Digital Twin for smart warehouse: Towards standardization[J]. Decision Analytics Journal, 2024, 12: 100509.
75. Qudus L. Resilient systems: building secure cyber-physical infrastructure for critical industries against emerging threats[J]. Int J Res Publ Rev, 2025, 6(1): 3330-46.
76. Shakeri Z, Benfriha K, Varmazyar M, et al. Production scheduling with multi-robot task allocation in a real industry 4.0 setting[J]. Scientific Reports, 2025, 15(1): 1795.
77. Huang H, Wlazlo P, Mao Z, et al. Cyberattack defense with cyber-physical alert and control logic in industrial controllers[J]. IEEE Transactions on Industry Applications, 2022, 58(5): 5921-5934.
78. 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.
79. Aleisa M A. Enhancing Security in CPS Industry 5.0 using Lightweight MobileNetV3 with Adaptive Optimization Technique[J]. Scientific Reports, 2025, 15(1): 1-21.
80. Acquaah Y T, Kaushik R. Normal-only Anomaly detection in environmental sensors in CPS: A comprehensive review[J]. IEEE Access, 2024.
81. Laxmi Lydia E, Ramesh S N, Denisovich V, et al. African buffalo optimization with deep learning-based intrusion detection in cyber-physical systems[J]. Scientific Reports, 2025, 15(1): 10219.
82. 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: 2526111.
83. Sun J, et al. Indicator & crowding distance-based evolutionary algorithm for combined heat and power economic emission dispatch[J]. Applied Soft Computing, 2020, 90: 106158.
84. Hallaji E, Razavi-Far R, Saif M. Robust Federated Learning for Mitigating Advanced Persistent Threats in Cyber-Physical Systems[J]. Applied Sciences, 2024, 14(19): 8840.
85. Chen L, Li Y, Huang M, 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.
86. Fard N E, Selmic R R, Khorasani K. A review of techniques and policies on cybersecurity using artificial intelligence and reinforcement learning algorithms[J]. IEEE Technology and Society Magazine, 2023, 42(3): 57-68.

 doi:10.20944/preprints202506.1885.v1

87. Pimple J F, Sharma A, Mishra J K. Elevating Security Measures in Cyber-Physical Systems: Deep Neural Network-Based Anomaly Detection with Ethereum Blockchain for Enhanced Data Integrity[J]. Journal of Electrical Systems, 2023, 19(2).
88. Al-Jaburi L R, Fadare O A, Al-Turjman F. Integrated AI Architecture and Fog Computing for Cyber-Physical Systems Autonomous Vehicles[M]//Smart Infrastructures in the IoT Era. Cham: Springer Nature Switzerland, 2025: 969-983.
89. Wei L, Zhang Q. Detection of False Data Attacks in Smart Grids Based on Improved UKF[J]. Journal of System Simulation, 2023, 35(7): 1508.
90. Pavão J, Bastardo R, Rocha N P. Cyber Resilience of Cyber-Physical Systems and Machine Learning, a Scoping Review[C]//International Conference on Information Technology and Applications. Singapore: Springer Nature Singapore, 2022: 501-512.
91. Chunduri V, Alsaadi M, Gupta S, et al. Blockchain-Based Secure Trust Management Scheme for Internet of Vehicles Over Cyber-Physical System[J]. IEEE Transactions on Intelligent Transportation Systems, 2024.
92. 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.
93. Byeon H, Jena S R, Bhargavi K N, et al. An intelligent cyber-physical system based-consensus algorithm for sustainable edge service provisioning in 6G-based IIoT applications[J]. Cyber-Physical Systems, 2025: 1-28.
94. Karimipour H, Derakhshan F. Artificial Intelligence for Threat Detection and Analysis in Industrial IoT: Applications and Challenges[J]. AI-Enabled Threat Detection and Security Analysis for Industrial IoT, 2021: 1-6.
95. Chehida S, Rutten E, Giraud G, et al. A model-based approach for self-adaptive security in CPS: Application to smart grids[J]. Journal of Systems Architecture, 2024, 150: 103118.
96. 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.
97. Chen L, Wang B. Robustness assessment of weakly coupled cyber-physical power systems under multi-stage attacks[J]. Electric Power Systems Research, 2024, 231: 110325.
98. Luo X, He J, Wang X, et al. Topology Optimization for Resilient Defense Strategies Against False Data Injection Attacks in Smart Grids[J]. Acta Automatica Sinica, 2023, 49(6): 1326-1338.
99. Bahaa A, Sayed A, Elfangary L, et al. A novel hybrid optimization enabled robust CNN algorithm for an IoT network intrusion detection approach[J]. Plos one, 2022, 17(12): e0278493.
100. Ahsan M S, Islam S, Shatabda S. A systematic review of metaheuristics-based and machine learning-driven intrusion detection systems in IoT[J]. Swarm and Evolutionary Computation, 2025, 96: 101984.
101. Liu F, et al. Bitcoin transaction strategy construction based on deep reinforcement learning[J]. Applied Soft Computing, 2021, 113: 107952.
102. Seema P N, Nair M G. The key modules involved in the evolution of an effective instrumentation and communication network in smart grids: a review[J]. Smart Science, 2023, 11(3): 519-537.
103. Buksh Z, Sharma N A, Chand R, et al. Cybersecurity Challenges in Smart Grid IoT[J]. IoT for Smart Grid: Revolutionizing Electrical Engineering, 2025: 175-206.
104. Mortlock T, Al Faruque M A. Adaptive Data Fusion for State Estimation and Control of Power Grids Under Attack[J]. IEEE Transactions on Industrial Informatics, 2024.
105. Wang Q, Tai W, Tang Y, et al. A review of false data injection attacks on cyber-physical power systems[J]. Acta Automatica Sinica, 2019, 45(1): 72-83.
106. 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.
107. Yang T, Liu Y C, Liu Y Z, et al. A review of the current status and trends in cyber-physical systems technology[J]. Journal of Electronics and Information, 2021, 43(12): 3393-3406.
108. Zhu B Q, Guo Y H, Guo C X, et al. A review of security assessment and defense strategies for power cyber-physical systems under information failure threats[J]. Power System Protection and Control, 2021, 49(1): 178-187.

 doi:10.20944/preprints202506.1885.v1

109. Zhang B, Liu X, Yu Z C, et al. A review of artificial intelligence-based network attack detection in power systems[J]. High Voltage Engineering, 2022, 48(11): 4413-4426.
110. Wang X J, Dou J M, Liu Z H, et al. A review and outlook on the application of explainable artificial intelligence in power systems[J]. Automation of Electric Power Systems, 2024, 48(4): 169-191.
111. Chen Y F, Zhao Q, He Y J, et al. A review of artificial intelligence applications in power systems[J]. Distributed Energy, 2023, 8(6): 49-57.
112. Yang B, Chen Y J, Yao W, et al. A review of power system stability assessment and decision-making based on next-generation artificial intelligence technologies[J]. Automation of Electric Power Systems, 2022, 46(22): 200-223.
113. Patel C D, Aggarwal M, Chaubey N K. Enhancing Cyber-Physical Systems Security Through Advanced Defense Mechanisms[M]//Advancing Cyber Security Through Quantum Cryptography. IGI Global, 2025: 307-342.
114. Li Y, He S, Li Y, 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, 2024, 35(5): 5902 - 5914.
115. Li B, Gao Z Y. Analysis and outlook on the application of artificial intelligence technology in smart grids[J]. China Electric Power, 2017, 50(12): 136.
116. 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.
117. Ding X, Wang H, Zhang X, et al. Dual nature of cyber–physical power systems and the mitigation strategies[J]. Reliability Engineering & System Safety, 2024, 244: 109958.
118. Zhou Z, Zhang J, Zhang X. A review on defense mechanism against the denial of service and false data injection in cyber-physical power systems[C]//2023 IEEE 6th International Electrical and Energy Conference (CIEEC). IEEE, 2023: 4539-4545.
119. Bitirgen K, Filik Ü B. Markov game based on reinforcement learning solution against cyber–physical attacks in smart grid[J]. Expert Systems with Applications, 2024, 255: 124607.
120. Li J, Li H, Su Q. Dynamic load altering attack detection for cyber physical power systems via sliding mode observer[J]. International Journal of Electrical Power & Energy Systems, 2023, 153: 109320.
121. Dong L, Xu H, Park J H, et al. Intermediate-Variable-Based Robust State Estimation for Cyber–Physical Systems Against FDI Attacks[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2024, 71(5): 2719-2723.
122. Bitirgen K, Filik Ü B. Markov game based on reinforcement learning solution against cyber–physical attacks in smart grid[J]. Expert Systems with Applications, 2024, 255: 124607.
123. Huang J, Yang J, Zheng Z, et al. A novel probabilistic modeling for multilateral random attacks in cyber-physical system reliability analysis[J]. Quality and Reliability Engineering International, 2024, 40(5): 2620-2637.
124. Li J, Li Y, Su Q. Sequential Recovery of Cyber-Physical Power Systems Considering Cyber-Attacks[J]. Information Sciences, 2025: 122310.
125. Li Y, Li Z, Chen L. Dynamic State Estimation of Generators Under Cyber Attacks[J]. IEEE Access, 2019, 7: 125252-125267.
126. Liu F, Wu L, Liu Q, et al. Dynamic-Memory Event-Triggered Secure Control for Cyber-Physical Power Systems Under Hybrid Attacks[J]. IEEE Transactions on Network Science and Engineering, 2025.
127. Sun C, Su Q, Li J. Secure Tracking Control and Attack Detection for Power Cyber-Physical Systems based on Integrated Control Decision[J]. IEEE Transactions on Information Forensics and Security, 2024.
128. Sun Z, Chen G, Ding Y, et al. Joint safety and security risk analysis in industrial cyber-physical systems: A survey[J]. IET Cyber-Physical Systems: Theory & Applications, 2024, 9(4): 334-349.
129. 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.
130. 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.

 doi:10.20944/preprints202506.1885.v1

131. Li Y, Zhang M, Chen C. A deep-learning intelligent system incorporating data augmentation for short-term voltage stability assessment of power systems[J]. Applied Energy, 2022, 308: 118347.
132. Zhao A P, Gu C, Bao Z, et al. Optimizing Cyber Insurance and Defense for Multi-Energy Systems Under False Data Injections[J]. IET Renewable Power Generation, 2025, 19(1): e70011.
133. Yang B, He G, Xin L, et al. Multi-sensor CPS security state estimation based on distributed detection and alarm switching[J]. Control Theory & Applications, 2024, 41(11):11-21.
134. Bahrami J, Ebrahimabadi M, Younis M, et al. Digital Twin Based Topology Fingerprinting for Detecting False Data Injection Attacks in Cyber-Physical Systems[C]//ICC 2024-IEEE International Conference on Communications. IEEE, 2024: 2053-2058.
135. Krishnaveni S, Chen T M, Sathiyanarayanan M, et al. CyberDefender: an integrated intelligent defense framework for digital-twin-based industrial cyber-physical systems[J]. Cluster Computing, 2024, 27(6): 7273-7306.
136. Zhang Y, Du C, Chen Z, et al. Digital Twin-Based Resilient Model Predictive Control for Industrial Cyber-Physical Systems[J]. IEEE Journal of Emerging and Selected Topics in Industrial Electronics, 2025.
137. Ma J, Guo Y, Fang C, et al. Digital-twin-based CPS anomaly diagnosis and security defense countermeasure recommendation[J]. IEEE Internet of Things Journal, 2024.
138. Li Q, Zhao F, Zhuang L, et al. Steady-state risk prediction analysis of power system based on power digital twinning[J]. Sustainability, 2023, 15(3): 2555.
139. Aziz A, Schelén O, Bodin U. Digital twin as a proxy for industrial cyber-physical systems[C]//Proceedings of the 2023 10th International Conference on Wireless Communication and Sensor Networks. 2023: 85-92.
140. Adeyemo H B, Bahsoon R, Tiňo P. Surrogate-based digital twin for predictive fault modelling and testing of cyber physical systems[C]//2022 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT). IEEE, 2022: 166-169.
141. Kabir M R, Halder D, Ray S. Digital Twins for IoT-Driven Energy Systems: A Survey[J]. IEEE Access, 2024.
142. Erceylan G, Akbarzadeh A, Gkioulos V. Leveraging digital twins for advanced threat modeling in cyber-physical systems cybersecurity[J]. International Journal of Information Security, 2025, 24(3): 1-18.
143. 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. DOI: 10.1109/TNSE.2025.3547227.
144. Lucchese M, Salerno G, Pugliese A. A Digital Twin-Based Approach for Detecting Cyber–Physical Attacks in ICS Using Knowledge Discovery[J]. Applied Sciences, 2024, 14(19): 8665.
145. 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.
146. Safavat S, Rawat D B. Digital Twin Based Asynchronous Federated Learning Enabled IDS for False Data Injection Attacks in Vehicular CPS[C]//2024 8th International Conference on Computer, Software and Modeling (ICCSM). IEEE, 2024: 19-23.
147. Danilczyk W C. Digital Twin for Intelligent Cyber-Physical Systems[M]. University of Rhode Island, 2023.
148. Krishnaveni S, Sivamohan S, Jothi B, et al. TwinSec-IDS: An Enhanced Intrusion Detection System in SDN-Digital-Twin-Based Industrial Cyber-Physical Systems[J]. Concurrency and Computation: Practice and Experience, 2025, 37(3): e8334.
149. Ahmed C M, Umer M A, Liyakkathali B S S B, et al. Machine Learning for CPS Security: Applications, Challenges[J]. Machine Intelligence and Big Data Analytics for Cybersecurity Applications, 2020, 919: 397.
150. Patel H, Akbarfam A J, Maleki H. A Survey on Digital Twin: From Industrial Applications to Cybersecurity[C]//2024 IEEE Smart World Congress (SWC). IEEE, 2024: 2111-2118.
151. Choi W, Hudachek K, Koskey S, et al. Digital twin in the power generation industry[J]. JMST Advances, 2024, 6(1): 103-119.
152. Jayasinghe H, Gunawardane K, Nicholson R. Applications of Electrical Load Modelling in Digital Twins of Power Systems[J]. Energies (19961073), 2025, 18(4).
153. Zhou H, Li M, Sun Y, et al. Digital Twin-Based Cyber Range for Industrial Internet of Things[J]. IEEE Consumer Electronics Magazine, 2022, 12(6): 66-77.

 doi:10.20944/preprints202506.1885.v1

154. Ribas Monteiro L F, Rodrigues Y R, Zambroni de Souza A C. Cybersecurity in cyber–physical power systems[J]. Energies, 2023, 16(12): 4556.
155. 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. DOI: 10.1049/rpg2.12432.
156. Heininger R, Jost T E, Stary C. Developing and Operating Digital Process Twins While Preserving Autonomy in CPS[C]//2024 26th International Conference on Business Informatics (CBI). IEEE, 2024: 50-59.
157. Atalay M. A Service-Oriented Digital Twins Framework for Smart Grid Management[C]//2022 International Workshop on Secure and Reliable Microservices and Containers (SRMC). IEEE, 2022: 9-17.
158. Van Dinter R, Tekinerdogan B, Catal C. Reference architecture for digital twin-based predictive maintenance systems[J]. Computers & Industrial Engineering, 2023, 177: 109099.
159. Pires F, Ahmad B, Moreira A P, et al. Digital twin based what-if simulation for energy management[C]//2021 4th IEEE international conference on industrial cyber-physical systems (ICPS). IEEE, 2021: 309-314.
160. Jeong D Y, Baek M S, Lim T B, et al. Digital twin: Technology evolution stages and implementation layers with technology elements[J]. IEEE Access, 2022, 10: 52609-52620.
161. Negri E, Cattaneo L, Pandhare V, et al. Integrating PHM into production scheduling through a Digital Twin-based framework[J]. IFAC-PapersOnLine, 2022, 55(19): 31-36.
162. Wang J, Jin C, Lee Y, et al. Energy Use Prediction Model in Digital Twin[C]//International conference on construction engineering and project management. Korea Institute of Construction Engineering and Management, 2022: 1256-1263.
163. Jovanovic V, Kuzlu M, Cali U, et al. Digital twin in industry 4.0 and beyond applications[M]//Digital Twin Driven Intelligent Systems and Emerging Metaverse. Singapore: Springer Nature Singapore, 2023: 155-174.
164. Wang Y, Wang S, Yang W, et al. A digital-twin-based adaptive multi-objective Harris Hawks Optimizer for dynamic hybrid flow green scheduling problem with dynamic events[J]. Applied Soft Computing, 2023, 143: 110274.
165. Li J, Wang J. Digital twin-driven management strategies for logistics transportation systems[J]. Scientific Reports, 2025, 15(1): 12186.
166. Li Y, Han M, Shahidehpour M, et al. Data-driven distributionally robust scheduling of community integrated energy systems with uncertain renewable generations considering integrated demand response[J]. Applied Energy, 2023, 335: 120749.
167. Sheikh Z A, Singh Y, Singh P K, et al. Defending the defender: Adversarial learning based defending strategy for learning based security methods in cyber-physical systems (cps)[J]. Sensors, 2023, 23(12): 5459.
168. Bai M, Liu P, Lv F, et al. Adversarial Attack against Intrusion Detectors in Cyber-Physical Systems With Minimal Perturbations[C]//2024 IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA). IEEE, 2024: 816-825.
169. Yang X, et al. Gaussian Mixture Model Uncertainty Modeling for Power Systems Considering Mutual Assistance of Latent Variables[J]. IEEE Transactions on Sustainable Energy, 2024, 1-4. DOI: 10.1109/TSTE.2024.3356259.
170. Feng X, Liu Y, Hu S. Machine learning for cyber-physical power system security[M]//Machine Learning for Embedded System Security. Cham: Springer International Publishing, 2022: 105-124.
171. Sharma A, Kejriwal D, Pakina A K. Adversarial AI and cyber–physical system resilience: Protecting critical[J]. International Journal of Artificial Intelligence and Data Research, 2023, 14(2).
172. Tian P, Liao W, Qian C, et al. Foundation of secured edge intelligence in CPS[M]//Edge Intelligence in Cyber-Physical Systems. Academic Press, 2025: 297-323.
173. Bhattacharjee A, Bai G, Tushar W, et al. Deebbaa: A benchmark deep black box adversarial attack against cyber-physical power systems[J]. IEEE Internet of Things Journal, 2024.
174. Alqaralleh B A Y, Aldhaban F, AlQarallehs E A, et al. Optimal machine learning enabled intrusion detection in cyber-physical system environment[J]. Computers, Materials & Continua, 2022, 72(3): 4691-4707.
175. Alqaralleh B A Y, Aldhaban F, AlQarallehs E A, et al. Optimal machine learning enabled intrusion detection in cyber-physical system environment[J]. Computers, Materials & Continua, 2022, 72(3): 4691-4707.

 doi:10.20944/preprints202506.1885.v1

176. Figueroa H, Wang Y, Giakos G C. Adversarial attacks in industrial control cyber physical systems[C]//2022 IEEE international conference on imaging systems and techniques (IST). IEEE, 2022: 1-6.
177. 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.
178. Vähäkainu J P, Lehto M J, Kariluoto A J E. Adversarial attack's impact on machine learning model in cyber-physical systems[J]. Journal of Information Warfare, 2020, 19(4): 57-69.
179. 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.
180. 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.
181. Gipiškis R, Chiaro D, Preziosi M, et al. The impact of adversarial attacks on interpretable semantic segmentation in cyber–physical systems[J]. IEEE Systems Journal, 2023, 17(4): 5327-5334.
182. Chan M. Trustworthy machine learning for securing cyber-physical systems[D]. Rutgers University-School of Graduate Studies, 2024.
183. Wu Y, Wei-Kocsis J. A practical and stealthy adversarial attack for cyber-physical applications[C]//The AAAI-22 Workshop on Adversarial Machine Learning and Beyond. 2022.
184. Barve Y, Karve P, Gokhale A, et al. Research challenges in the design and composition of surrogate models for robust CPS: position paper[C]//Proceedings of the Workshop on Design Automation for CPS and IoT. 2021: 26-29.
185. Madabhushi S, Dewri R. Mitigating Over-Generalization in Anomalous Power Consumption Detection using Adversarial Training[J]. ACM Transactions on Cyber-Physical Systems, 2025.
186. Lakshmiranganatha S, Manjunatha K A. Machine Learning and Deep Learning Approaches in Cyber-Physical Systems[M]//Secure and Smart Cyber-Physical Systems. CRC Press, 2024: 1-29.
187. Zhang S, Xu Y, Xie X. Universal Adversarial Perturbations Against Machine Learning-Based Intrusion Detection Systems in Industrial Internet of Things[J]. IEEE Internet of Things Journal, 2024.
188. 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.
189. Sayin B, Zoppi T, Marchini N, et al. Bringing Machine Learning Classifiers Into Critical Cyber-Physical Systems: a Matter of Design[J]. IEEE Access, 2025.
190. Haque N I, Shahriar M H, Dastgir M G, et al. A survey of machine learning-based cyber-physical attack generation, detection, and mitigation in smart-grid[C]//2020 52nd North American Power Symposium (NAPS). IEEE, 2021: 1-6.
191. Olowononi F O, Rawat D B, Garuba M, et al. Security engineering with machine learning for adversarial resiliency in cyber physical systems[C]//Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications. SPIE, 2019, 11006: 605-611.
192. Li F, Shen W, Bi Z, et al. Sparse Adversarial Learning for FDIA Attack Sample Generation in Distributed Smart Grids[J]. Computer Modeling in Engineering & Sciences (CMES), 2024, 139(2).
193. Neema H, Volgyesi P, Koutsoukos X, et al. Online testbed for evaluating vulnerability of deep learning based power grid load forecasters[C]//2020 8th Workshop on Modeling and Simulation of Cyber-Physical Energy Systems. IEEE, 2020: 1-6.
194. Olowononi F O, Rawat D B, Liu C. Trust-based adversarial resiliency in vehicular cyber physical systems using reinforcement learning[C]//International Symposium on Security in Computing and Communication. Singapore: Springer Singapore, 2020: 139-151.
195. Tahsini A, Dunstatter N, Guirguis M, et al. Deepbloc: a framework for securing cps through deep reinforcement learning on stochastic games[C]//2020 IEEE Conference on Communications and Network Security (CNS). IEEE, 2020: 1-9.
196. Riya F F, Hoque S, Sun J S, et al. Mitigating Adversarial Effects of False Data Injection Attacks in Power Grid[J]. arXiv preprint arXiv:2301.12487, 2023.
197. Broda-Milian K. The ART of CityLearn: Observation Perturbation Attacks and Mitigation for Reinforcement Learning Agents in a Cyber Physical Power System[J]. 2024.

 doi:10.20944/preprints202506.1885.v1

198. Song Q, Tan R, Ren C, et al. On credibility of adversarial examples against learning-based grid voltage stability assessment[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 21(2): 585-599.
199. Kou L, Li Y, Zhang F, et al. Review on monitoring, operation and maintenance of smart offshore wind farms[J]. Sensors, 2022, 22(8): 2822.
200. Verma N, Kumar N, Sheikh Z A, et al. Machine Learning for the Cybersecurity of Robotic Cyber-Physical Systems: A Review[J]. Procedia Computer Science, 2025, 259: 1817-1826.
201. Edara P, Banerjee S, Joardar B K. Localization of Data Compromised by Hardware Attacks in Machine Learning enabled Cyber-Physical Edge Devices[J]. ACM Transactions on Cyber-Physical Systems, 2024.
202. Sarkar S. CAREER: Robustifying Machine Learning for Cyber-Physical Systems[J]. NSF Award Number 1845969. Directorate for Computer and Information Science and Engineering, 2019, 18(1845969): 45969.
203. Cao Y, Yan J, Feng H, et al. SCAAE: Using Self-Supervised Contrastive Learning in Adversarial AutoEncoder for Anomaly Detection of Multivariate Time Series in Cyber Physics Systems[C]//2023 China Automation Congress (CAC). IEEE, 2023: 8102-8107.
204. Ramachandran A, Gayathri K, Alkhayyat A, et al. Aquila Optimization with Machine Learning-Based Anomaly Detection Technique in Cyber-Physical Systems[J]. Computer Systems Science & Engineering, 2023, 46(2).
205. Jaisingh W, Nanjundan P, George J P. Machine learning in cyber threats intelligent system[M]//Artificial Intelligence for Cyber Defense and Smart Policing. Chapman and Hall/CRC, 2024: 1-20.
206. Zhong K, Yang Z, Yu S, et al. Deep Reinforcement Learning-based Multi-Layer Cascaded Resilient Recovery for Cyber-Physical Systems[J]. IEEE Transactions on Services Computing, 2024.
207. Meyer T, Kaloudi N, Li J. A systematic literature review on malicious use of reinforcement learning[C]//2021 IEEE/ACM 2nd International Workshop on Engineering and Cybersecurity of Critical Systems (EnCyCriS). IEEE, 2021: 21-28.
208. Selvi K, Dilip G. Enhancing Cyber-Physical Systems Security: A Review of Deep Learning and Blockchain Integration[C]//2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN). IEEE, 2024: 725-734.
209. Li Y, Wang B, Yang Z, et al. Hierarchical stochastic scheduling of multi-community integrated energy systems in uncertain environments via Stackelberg game[J]. Applied Energy, 2022, 308: 118392.
210. Hudani D, Haseeb M, Taufiq M, et al. A data-centric approach to generate invariants for a smart grid using machine learning[C]//Proceedings of the 2022 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems. 2022: 31-36.
211. Li J, Jia C, Zhu X, et al. Dual-layer model for capacity optimization of hybrid energy storage system to reduce thermal power frequency modulation loss[J]. High Voltage Engineering, 2023, 49: 3965-3976.
212. [212 Li K, Li F, Wang B, et al. False data injection attack sample generation using an adversarial attention-diffusion model in smart grids[J]. AIMS Energy, 2024, 12(6).
213. Babadi N, Karimipour H, Islam A. An Ensemble Learning to Detect Decision-Based Adversarial Attacks in Industrial Control Systems[C]//2023 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2023: 879-884.
214. Elsisi M, Su C L, Lin C H, et al. Enhancing resilient operation of distributed energy resources using reliable machine learning-based iot connectivity[C]//2024 IEEE/IAS 60th Industrial and Commercial Power Systems Technical Conference (I&CPS). IEEE, 2024: 1-6.
215. Fang Z, Zhao D, Chen C, et al. Nonintrusive Appliance Identification with Appliance-Specific Networks[J]. IEEE Transactions on Industry Applications, 2020, 56(4): 3443-3452.
216. Fernandes J M, Rivadeneira J E, Rodrigues A, et al. People 4.0—A model for Human-in-the-Loop CPS-based systems[J]. Computer Standards & Interfaces, 2025, 91: 103895.
217. 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.
218. Maity A, Banerjee A, Gupta S K S. Detection of Unknown-Unknowns in Human-in-Loop Human-in-Plant Safety Critical Systems[J]. IEEE Transactions on Artificial Intelligence, 2025.
219. Cambeiro J M R A P. The Human in the loop in Cyber-Physical Systems: the case of Building Automation[D]. Universidade NOVA de Lisboa (Portugal), 2019.

 doi:10.20944/preprints202506.1885.v1

220. Banerjee A, Lamrani I, Gupta S K S. Synthesizing Operationally Safe Controllers for Human-in-the-Loop Human-in-the-Plant Hybrid Close Loop Systems[C]//International Conference on Pattern Recognition. Cham: Springer Nature Switzerland, 2024: 17-35.
221. Nota G, Petraglia G. The Design of Human-in-the-Loop Cyber-Physical Systems for Monitoring the Ecosystem of Historic Villages[J]. Smart Cities, 2024, 7(5): 2966-2994.
222. Makri C, Gürdür Broo D, Neely A. Human-in-Loop Decision-Making and Autonomy: Lessons Learnt from the Aviation Industry Transferred to Cyber-Physical Systems[J]. Technologies, 2022, 10(6): 120.
223. Manolova A, Poulkov V, Tonchev K, et al. Challenges in the design of smart vehicular cyber physical systems with human in the loop[M]//Breakthroughs in smart city implementation. River Publishers, 2017: 165-186.
224. Qin B, Liu D. Research Progress and Prospects on Analysis and Control of Power Grid Cyber-Physical Systems[J]. Proceedings of the CSEE, 2020, 40(18): 5816-5826.
225. Gil M, Albert M, Fons J, et al. Engineering human-in-the-loop interactions in cyber-physical systems[J]. Information and software technology, 2020, 126: 106349.
226. 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.
227. Rivadeneira J E, Borges G A, Rodrigues A, et al. A unified privacy preserving model with AI at the edge for Human-in-the-Loop Cyber-Physical Systems[J]. Internet of Things, 2024, 25: 101034.
228. Eskandar S, Wang J, Razavi S. Human-in-the-loop cyber-physical systems for construction safety[J]. Cyber-physical systems in the built environment, 2020: 161-173.
229. Zhong C, Ding Y, Wang H, et al. Model predictive control strategy in waked wind farms for optimal fatigue loads[J]. Electric Power Systems Research, 2023, 224: 109793.
230. Jena S, Sundarrajan S, Meena A, et al. Human-in-the-loop control and security for intelligent cyber-physical systems (CPSs) and IoT[C]//XVIII International Conference on Data Science and Intelligent Analysis of Information. Cham: Springer International Publishing, 2022: 393-403.
231. Cao J, Jin J, Ming Y, et al. Human-cyber-physical systems for energy internet—A review[J]. Energies, 2023, 16(15): 5624.
232. Karunamurthy A, Kiruthivasan R, Gauthamkrishna S. Human-in-the-Loop Intelligence: Advancing AI-Centric Cybersecurity for the Future[J]. Quing: International Journal of Multidisciplinary Scientific Research and Development, 2023, 2(3): 20-43.
233. Yang L, Sun Q, Zhang N, et al. Optimal energy operation strategy for we-energy of energy internet based on hybrid reinforcement learning with human-in-the-loop[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 52(1): 32-42.
234. Zeng Q, Nait-Abdesselam F. Enhancing UAV Network Security: A Human-in-the-Loop and GAN-Based Approach to Intrusion Detection[J]. IEEE Internet of Things Journal, 2025.
235. Han Q L. Secure and Safe MetaControl for Cyber Physical Systems[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(12): 2177-2178.
236. Sakhnini J, Karimipour H. AI and security of cyber physical systems: Opportunities and challenges[J]. Security of Cyber-Physical Systems: Vulnerability and Impact, 2020: 1-4.
237. Guidotti D, Pandolfo L, Pulina L. A Systematic Literature Review of Supervised Machine Learning Techniques for Predictive Maintenance in Industry 4.0[J]. IEEE Access, 2025.
238. Parizad A, Baghaee H R, Alizadeh V, et al. Emerging Technologies and Future Trends in Cyber-Physical Power Systems: Toward a New Era of Innovations[J]. Smart Cyber-Physical Power Systems: Solutions from Emerging Technologies, 2025, 2: 525-565.
239. Trivedi S, Aggarwal V, Rastogi R. Enhancing the Power of Cyber-Physical Systems Enabled with AI: An Introduction—Facts and Myths along with Modular Approach[J]. Artificial Intelligence Solutions for Cyber-Physical Systems, 1-39.
240. Menon U V, Kumaravelu V B, Kumar C V, et al. AI-Powered IoT: A Survey on Integrating Artificial Intelligence with IoT for Enhanced Security, Efficiency, and Smart Applications[J]. IEEE Access, 2025.

 doi:10.20944/preprints202506.1885.v1

241. Cui J, Jin Y, Yu R, et al. A robust approach for the decomposition of high-energy-consuming industrial loads with deep learning[J]. Journal of Cleaner Production, 2022, 349: 131208.
242. Mohamed N. Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms[J]. Knowledge and Information Systems, 2025: 1-87.
243. Akinsola J E T, Abimbola R O, Adeagbo M A, et al. Application of Artificial Intelligence for DDoS Attack Detection and Prevention on Cyber Physical Systems Using Deep Learning[M]//Internet of Things and Cyber Physical Systems. CRC Press, 2022: 83-126.
244. Ding D, Han Q L, Ge X, et al. Privacy-preserving filtering, control and optimization for industrial cyber-physical systems[J]. Science China Information Sciences, 2025, 68(4): 1-17.
245. Abshari D, Sridhar M. A survey of anomaly detection in cyber-physical systems[J]. arXiv preprint arXiv:2502.13256, 2025.
246. Tooki O O, Popoola O M. A critical review on intelligent-based techniques for detection and mitigation of cyberthreats and cascaded failures in cyber-physical power systems[J]. Renewable Energy Focus, 2024: 100628.
247. 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.
248. 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.
249. Zhong K, Yang Z, Yu S, et al. Deep Reinforcement Learning-based Multi-Layer Cascaded Resilient Recovery for Cyber-Physical Systems[J]. IEEE Transactions on Services Computing, 2024.
250. Ghorbani M, Ghassemi A, Alikhani M, et al. Using Kolmogorov–Arnold network for cyber-physical system security: A fast and efficient approach[J]. International Journal of Critical Infrastructure Protection, 2025: 100768.
251. Liu F, Lao K W, Gao L, et al. Predicting Power Outage at Low-Lying Area Substations During Storm Surge Disasters Using Multi-Grained Cascaded Forest[J]. IEEE Transactions on Industry Applications, 2025, 61(2): 2803-2812.
252. Ye X, Esnaola I, Perlaza S M, et al. Decentralized Stealth Attacks on Cyber-Physical Systems[J]. arXiv preprint arXiv:2505.07029, 2025.
253. Mäkelburg J, Perez-Palacin D, Mirandola R, et al. Surveying Uncertainty Representation: A Unified Model for Cyber-Physical Systems[J]. arXiv preprint arXiv:2503.23892, 2025.
254. Reiter H, Rathje P, Landsiedel O, et al. DPM-Bench: Benchmark for Distributed Process Mining Algorithms on Cyber-Physical Systems[J]. arXiv preprint arXiv:2502.09975, 2025.
255. Bashir A, Shamszaman Z U, Song Z, et al. Co-evolutionary Dynamics of Attack and Defence in Cybersecurity[J]. arXiv preprint arXiv:2505.19338, 2025.
256. Yamany W. Optimised Learning Models for Defending Against Cyber Attacks in Cyber-Physical Systems[D]. University of New South Wales (Australia), 2023.
257. Sun Q, Yang G H. Anti-Disturbance Secure State Estimation for Continuous-Time Cyber-Physical Systems Under Sparse Sensor Attacks[J]. IEEE Internet of Things Journal, 2025.
258. Kaloudi N, Li J. The ML-based sensor data deception targeting cyber–physical systems: A review[J]. Computer Science Review, 2025, 57: 100753.
259. Niu H, Jagannathan S. Optimal defense and control of dynamic systems modeled as cyber-physical systems[J]. The Journal of Defense Modeling and Simulation, 2015, 12(4): 423-438.
260. Casola V, De Benedictis A, Mazzocca C, et al. Designing secure and resilient cyber-physical systems: A model-based moving target defense approach[J]. IEEE Transactions on Emerging Topics in Computing, 2022.

 doi:10.20944/preprints202506.1885.v1

261. 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.