electronics

MDPI

Review

Advanced Fault Detection and Diagnosis Exploiting Machine Learning and Artificial Intelligence for Engineering Applications

Davide Paolini, Pierpaolo Dini *, Abdussalam Elhanashi and Sergio Saponara

Department of Information Engineering, University of Pisa, Via Girolamo Caruso n.16, 56100 Pisa, Italy; davide.paolini@ing.unipi.it (D.P.); abdussalam.elhanashi@ing.unipi.it (A.E.); sergio.saponara@unipi.it (S.S.)
* Correspondence: pierpaolo.dini@unipi.it

Abstract

Modern engineering systems require reliable and timely Fault Detection and Diagnosis (FDD) to ensure operational safety and resilience. Traditional model-based and rule-based approaches, although interpretable, exhibit limited scalability and adaptability in complex, data-intensive environments. This survey provides a systematic overview of recent studies exploring Machine Learning (ML) and Artificial Intelligence (AI) techniques for FDD across industrial, energy, Cyber-Physical Systems (CPS)/Internet of Things (IoT), and cybersecurity domains. Deep architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Transformers, and Graph Neural Networks (GNNs) are compared with unsupervised, hybrid, and physics-informed frameworks, emphasizing their respective strengths in adaptability, robustness, and interpretability. Quantitative synthesis and radar-based assessments suggest that AI-driven FDD approaches offer increased adaptability, scalability, and early fault detection capabilities compared to classical methods, while also introducing new challenges related to interpretability, robustness, and deployment. Emerging research directions include the development of foundation and multimodal models, federated learning (FL), and privacy-preserving learning, as well as physics-guided trustworthy AI. These trends indicate a paradigm shift toward self-adaptive, interpretable, and collaborative FDD systems capable of sustaining reliability, transparency, and autonomy across critical infrastructures.

Keywords: fault detection and diagnosis; machine learning; deep learning; cyber-physical systems; explainable AI; federated learning; industrial applications

Check for updates

Academic Editors: Panayiotis Bozanis and Leonidas Akritidis

Received: 13 December 2025
Revised: 6 January 2026
Accepted: 9 January 2026
Published: 22 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Modern engineering systems have reached unprecedented levels of complexity and interconnectivity, making timely and reliable fault detection and diagnosis (FDD) critically important. Recognizing its strategic value, industries and research institutions have significantly increased their investments in FDD solutions over recent years as shown in Figure 1. In safety-critical infrastructures such as industrial plants, cyber–physical systems (CPS), smart grids, and cyber-security, undetected faults may lead to cascading failures, severe safety risks, and costly downtime. The growing integration of renewable energy sources and the evolution of smart power grids demand advanced diagnostic techniques to maintain reliability and prevent disruptive outages [1]. Similarly, CPSs in domains ranging from manufacturing to transportation require rapid anomaly detection to ensure operational continuity and safety. Traditional FDD approaches, including model-based observers, rule-based expert systems, and threshold-based techniques, have well-known

limitations in modern large-scale and nonlinear environments. Model-based methods depend on accurate analytical representations and high-fidelity models that are challenging to derive and maintain for complex systems, and often degrade under modeling uncertainties. Rule-based systems, conversely, rely on extensive expert knowledge and predefined failure modes, which limit scalability and adaptability. Consequently, classical methods struggle to identify unforeseen or incipient faults and are prone to the curse of dimensionality when monitoring high-dimensional sensor networks operating under varying conditions. These limitations have motivated a paradigm shift toward data-driven and learning-based FDD. By leveraging large-scale operational data, machine learning (ML) techniques can autonomously learn fault patterns, enabling the detection of both known and previously unseen fault signatures with improved adaptability and robustness. The widespread adoption of the Industrial Internet of Things (IIoT) and high-frequency sensor networks has made vast amounts of heterogeneous data available in real time, facilitating the development of advanced algorithms for large-scale condition monitoring. Among these, deep learning (DL) has demonstrated outstanding performance in FDD tasks. Deep neural networks automatically extract discriminative and hierarchical representations from complex sensor data—such as vibration, acoustic, or electrical signals—achieving higher sensitivity and accuracy than conventional classifiers. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs/LSTMs) are widely employed in predictive maintenance and process monitoring, while attention-based CNNs have achieved high-precision fault detection in rotating machinery. Furthermore, graph-based deep learning, particularly Graph Neural Networks (GNNs), has emerged as a powerful paradigm for modeling spatial and topological dependencies within complex infrastructures such as power grids and industrial systems. Zhang and He [2] proposed a graph-embedded recurrent network for compound-fault diagnosis in integrated energy systems, demonstrating how topological reasoning enhances diagnostic performance. Even in data-scarce contexts, semi-supervised and transfer-learning strategies have achieved accuracy exceeding 99% in detecting mechanical faults with minimal labeled samples, outperforming traditional ML models. Despite these advances, the opacity of purely data-driven models remains a major concern in safety-critical applications. Their black-box nature hinders interpretability, trust, and certification. As a result, Explainable Artificial Intelligence (XAI) has become increasingly relevant for FDD, providing mechanisms to attribute predictions to meaningful input features and fault sources. Recent studies integrate post-hoc interpretability tools and inherently transparent models to enhance the traceability and accountability of AI-based fault diagnostics, thereby improving user confidence and compliance with industrial safety standards [3]. A complementary line of research focuses on hybrid FDD approaches, which integrate analytical redundancy from physical models with the adaptability of ML inference. Rather than replacing physical knowledge, these physics-informed or grey-box frameworks exploit the strengths of both paradigms: the robustness and interpretability of model-based reasoning and the flexibility of data-driven learning. For instance, Carbone et al. [4] proposed a hybrid anomaly detection framework for spacecraft power systems that combines rule-based logic for major fault isolation with ML modules for subtle deviation detection. Similar strategies have improved diagnostic coverage, uncertainty management, and real-time performance in smart grids, industrial HVAC systems, and process control applications [5].

Beyond industrial and energy infrastructures, the aerospace domain represents one of the most mature and safety-critical application fields for Fault Detection and Diagnosis. Spacecraft, launch vehicles, and aviation systems operate under extreme constraints in terms of reliability, redundancy, limited sensing, and absence of human intervention, making autonomous and trustworthy FDD a fundamental requirement. Historically, aerospace systems have pioneered model-based and hybrid diagnostic frameworks, including ana-

lytical redundancy, observer-based fault isolation, and fault-tolerant control architectures. In recent years, these approaches have increasingly integrated machine learning and AI techniques to enhance sensitivity to incipient faults, adaptivity to mission phases, and robustness under uncertainty. AI-driven FDD has been successfully applied to spacecraft power systems, attitude determination and control systems, avionics, and propulsion subsystems, often combining physics-based models with data-driven inference to ensure explainability and certification compliance. Owing to stringent safety and verification requirements, aerospace applications have also been among the first to explore explainable AI, hybrid reasoning, and human-in-the-loop diagnostic strategies, making this domain a key reference for the development of trustworthy AI-based FDD methodologies.

The convergence of physics-based modeling, big data analytics, and explainable AI now defines the frontier of intelligent FDD. In light of these developments, an updated and comprehensive survey is both timely and necessary. This paper systematically reviews recent advances in AI- and ML-based fault detection and diagnosis across multiple domains. It compares modern data-driven approaches with classical techniques, emphasizing their performance, interpretability, and real-time applicability to complex systems. Special focus is devoted to deep learning architectures, explainable AI frameworks, and hybrid data/model-driven strategies. Drawing on representative works in industrial automation, energy systems, and CPS/IoT applications, this survey identifies persistent challenges—including data imbalance, real-time constraints, and trustworthiness—and discusses emerging directions such as federated learning, physics-informed neural networks, and large-model-based diagnostic frameworks that are shaping the next generation of intelligent and resilient fault detection systems.

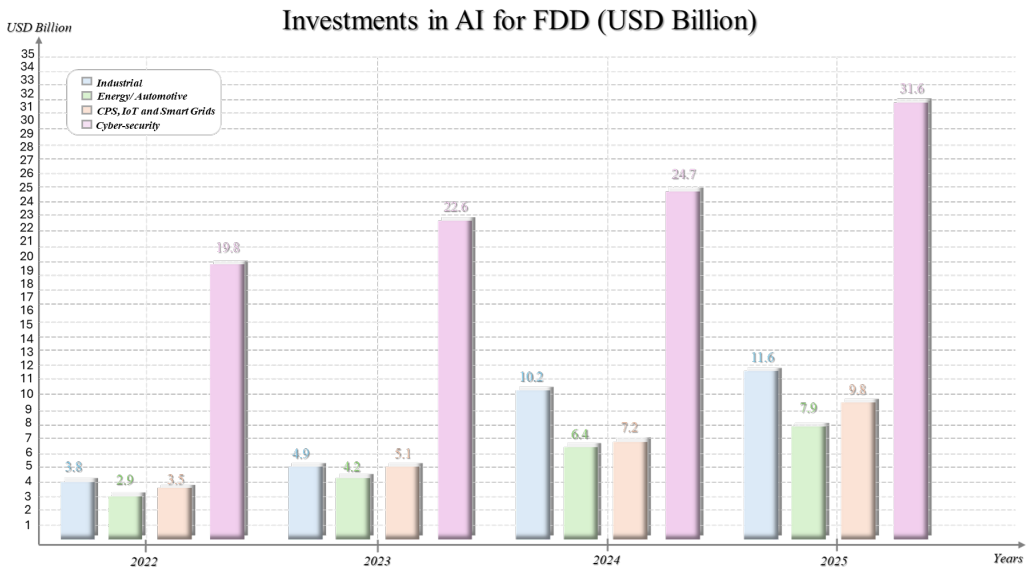


Figure 1. Estimated investments in Artificial Intelligence (AI) for Fault Detection and Diagnosis (FDD) across industrial, energy/automotive, Cyber-Physical Systems (CPS)/IoT and smart grids, and cybersecurity sectors from 2022 to 2025, highlighting the rapid growth of AI-driven solutions, with particularly strong investment trends in cybersecurity-related FDD applications.

2. Overview Methodology

This systematic review analyzes the current state of Machine Learning (ML) and Artificial Intelligence (AI) techniques applied to Fault Detection and Diagnosis (FDD) across the main domains of Industry 4.0, namely, industrial systems, energy systems, and cyber-physical/IoT systems, which have witnessed the most significant technological and research advancements in recent years [6,7], with particular focus on Deep Learning,

Explainable AI (XAI), and hybrid data/model-based strategies during the recent period. The adopted methodology consists of four main stages: definition of research questions, selection of sources, quality assessment, and metadata extraction for comparative analysis.

2.1. Research Questions

The guiding questions define the scope of the review and steer the analysis. In line with the goal of mapping recent advances in AI for FDD, the following research questions were formulated:

- How many and what types of studies (2022–2025) apply ML/AI techniques to FDD across industrial, energy and CPS/IoT domains, and for what purposes (detection, diagnosis, prognosis)?
- Which families of algorithms and architectures are employed (SVM, Random Forest, Autoencoder, CNN, RNN/LSTM/GRU, Transformer, GNN, TCN), and what complementary strategies are adopted (XAI, physics-informed, hybrid approaches, federated learning, TinyML/edge)?
- What is the nature and origin of the data used (real telemetry, SCADA, vibration/electrical sensors, test benches, digital twin simulations, public datasets)? What supervision regimes are applied (supervised, unsupervised, semi-supervised)?
- What results have been achieved in terms of accuracy, false alarm rate, robustness, latency, generalization, and operational reliability?
- What emerging challenges and research directions are identified regarding data imbalance, scarcity of real fault samples, interpretability, safety, and certification in safety-critical environments?

2.2. Selection Criteria

The literature search was conducted across major scientific databases. The time window 2022–2025 was selected, including early-access publications, as it represents the most recent evolution of ML/AI-based FDD methods. Earlier studies were considered only for theoretical context or to reference established benchmarks.

- Search strategy: Boolean combinations and semantic variants were used to ensure comprehensive coverage of the targeted domains, including terms such as: “fault detection” OR “anomaly detection” OR “diagnosis” AND “machine learning” OR “deep learning” OR “graph neural network” OR “transformer” AND “industrial” OR “power grid” OR “renewable” OR “CPS” OR “IoT” AND “XAI” OR “hybrid” OR “physics-informed” OR “federated learning” OR “TinyML”.
- Inclusion criteria:
  - Application-oriented or methodological studies with experimental validation on real data, test benches, or high-fidelity simulations;
  - Relevance to at least one of the four targeted domains;
  - Availability of quantitative results and validation protocol;
  - Explicit contributions on XAI, hybrid modeling, or edge/on-board deployment.
- Exclusion criteria:
  - Purely conceptual or review works without experimental validation;
  - Articles with unverifiable datasets or incomplete methodological details;
  - Studies unrelated to cyber-physical contexts or overly generic treatments.

2.3. Quality Assessment

All selected studies were evaluated based on methodological quality using the following main criteria:

- Clarity of application context and fault definition;
- Detailed description of the ML/AI methodology and training strategies;
- Dataset characteristics (source, size, sampling frequency, train/test split, cross-validation, imbalance handling);
- Performance metrics (precision, recall, F1-score, ROC-AUC, false alarm rate, mean time to detect);
- Consideration of latency, computational constraints, and edge/on-board deployment aspects;
- Inclusion of interpretability (XAI), robustness, and uncertainty analysis;
- Degree of reproducibility and availability of datasets or code.

Studies lacking scientific rigor or sufficient documentation were excluded or only referenced for contextual purposes. The adopted selection and quality assessment methodology ensures a high level of scientific rigor by prioritizing studies with experimental validation, clear methodological descriptions, and relevance to safety-critical engineering applications. This approach enables consistent comparison across heterogeneous domains and reduces the risk of over-representing purely conceptual contributions. However, this methodology also presents limitations. First, the focus on recent (2022–2025) literature may underrepresent earlier foundational works, which are referenced primarily for context. Second, qualitative dimensions such as interpretability or deployment readiness inherently involve expert judgment, which—although mitigated through cross-domain consistency, cannot be entirely eliminated. Finally, performance metrics reported across studies are not always directly comparable due to differences in datasets, evaluation protocols, and operational constraints.

2.4. Data Characterization and Metadata Extraction

For each selected paper, the following metadata were extracted and normalized to enable structured comparison and taxonomy development:

- Authors, year, and publication venue;
- Application domain and target system (industrial, energy, CPS/IoT);
- Fault type and task (detection, diagnosis, prognosis);
- Data source and type (real, simulation, digital twin, public dataset);
- ML algorithm or architecture used (SVM, RF, AE, CNN, RNN/LSTM, Transformer, GNN, TCN, etc.);
- Optimization and generalization strategies (data augmentation, transfer learning, GAN, hybrid modeling);
- Validation protocol and performance metrics;
- Operational constraints (latency, computational resources, edge/on-board implementation);
- Explainability and robustness methods (SHAP, saliency maps, sensitivity analysis);
- Key results, reported limitations, and future directions.

This structure enables a homogeneous comparison of heterogeneous approaches, highlights cross-domain trends such as the effectiveness of hybrid methods and edge inference in real time contexts and identifies current gaps and opportunities for future research and industrial deployment.

Positioning and Novelty of This Survey

Unlike existing surveys that focus on a single application domain (e.g., industrial systems, smart grids, or cybersecurity) or a specific algorithmic family, this work provides a cross-domain, engineering-oriented synthesis of AI- and ML-based FDD techniques. The novelty of this survey lies in three main aspects:

1. A unified comparative framework that evaluates FDD methods across heterogeneous domains using consistent criteria (interpretability, robustness, scalability, real-time feasibility);
2. The explicit integration of deployment constraints, explainability, and human-in-the-loop considerations, which are often marginal in algorithm-centric reviews;
3. A quantitative-inspired qualitative synthesis, through radar-based assessments, designed to highlight structural trade-offs rather than absolute performance rankings.

By bridging industrial, energy, CPS/IoT, cybersecurity, and emerging transportation applications, this survey positions itself as a decision-support reference for system designers and engineers rather than a purely algorithmic catalog.

3. Contribution of This Work

The main contributions of this survey are as follows:

- Comprehensive Overview: A synthesis of advanced fault detection and diagnosis methods using ML/AI across industries, including supervised, unsupervised, and semi-supervised approaches.
- Comparison with Traditional Methods: An analytical evaluation of classical FDD approaches versus modern ML-based techniques, identifying scenarios where hybrid methods offer superior robustness.
- Highlight of Key Advances: A structured taxonomy covering deep learning architectures, explainable AI mechanisms, and hybrid frameworks that define the state-of-the-art in FDD.
- Future Outlook: Identification of open challenges, such as dataset standardization, trustworthiness, and real-time deployment, with a discussion of emerging research directions including physics-informed and federated learning approaches.

The remainder of this paper is organized as follows: Section 4 introduces theoretical foundations and classical fault detection methods; Section 5 presents a taxonomy of ML and AI-based FDD techniques and explores domain-specific implementations across industrial, energy and CPS/IoT systems; Section 6 discusses challenges, gaps, and emerging trends, while Section 7 concludes with key findings and future directions.

4. Background on Classical Fault Detection Methods in Industrial Systems

Fault detection and diagnosis (FDD) in industrial systems aims to identify abnormal behaviors and incipient faults to ensure operational safety, reliability, and maintainability. This section introduces the theoretical foundations and classical model-based approaches to fault detection, highlighting their mathematical formulation and intrinsic limitations that motivate the transition toward data-driven and AI-based solutions.

4.1. Problem Formulation

An industrial dynamic system can be modeled in continuous-time state-space form as:

$$\dot{x}(t) = f(x(t), u(t)) + B_f f(t), \qquad y(t) = h(x(t), u(t)) + D_f f(t), \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}^m$ is the input vector, $y(t) \in \mathbb{R}^p$ is the output vector, and $f(t)$ represents unknown fault signals (e.g., actuator, process, or sensor faults). The matrices B_f and D_f describe the fault distribution within the system dynamics and

outputs, respectively. The primary goal of fault detection is to generate a residual signal $r(t)$ that is sensitive to faults while being robust to disturbances and modeling uncertainties:

$$r(t) = y(t) - \hat{y}(t), \tag{2}$$

where $\hat{y}(t)$ is the estimated output from the nominal model. Under fault-free conditions, the residual $r(t) \approx 0$, whereas a deviation from zero indicates the presence of a fault. The residual evaluation process typically involves statistical or threshold-based decision logic:

$$\text{If } \|r(t)\| > \gamma \Rightarrow \text{Fault detected}, \tag{3}$$

where γ is a detection threshold defined according to the desired false-alarm rate [8].

4.2. Model-Based Fault Detection

Classical fault detection methods rely on analytical redundancy rather than hardware redundancy. A mathematical model of the process is used to reproduce the expected system behavior, and any inconsistency between the measured and estimated outputs is interpreted as a possible fault [9].

4.2.1. Observer-Based Methods

Observer-based methods employ a dynamic estimator—such as a Luenberger observer—to reconstruct the system's internal states and generate residuals. The linear time-invariant form is:

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L\big(y(t) - \hat{y}(t)\big), \qquad \hat{y}(t) = C\hat{x}(t), \tag{4}$$

where $\hat{x}(t)$ and $\hat{y}(t)$ are the estimated state and output, respectively, and L is the observer gain matrix. The residual signal is defined as:

$$r(t) = y(t) - \hat{y}(t) = Ce(t), \tag{5}$$

where $e(t) = x(t) - \hat{x}(t)$ is the estimation error. The observer gain L is typically designed such that $(A - LC)$ is stable and the residual is decoupled from noise and modeling uncertainties as much as possible. When unknown disturbances are present, an Unknown Input Observer (UIO) can be designed to minimize the residual's sensitivity to those disturbances while preserving fault sensitivity. In practice, a bank of observers is often employed, each tuned to a specific fault type, allowing both detection and isolation capabilities [8].

4.2.2. Parity Space Methods

The parity space approach eliminates explicit state estimation by exploiting redundant sensor equations or analytical consistency relations. A residual vector $r_p(t)$ is constructed from parity relations among measured outputs:

$$r_p(t) = Wy(t) + Vu(t), \tag{6}$$

where W and V are designed such that $r_p(t) = 0$ in the nominal case. Any inconsistency due to a fault breaks this parity relation, producing $r_p(t) \neq 0$. The method is algebraically equivalent to certain observer formulations, offering a computationally efficient alternative particularly suited for linear time-invariant systems [9].

4.2.3. Kalman Filter and Stochastic Methods

When the process and measurement noises are stochastic with known covariance, the Kalman filter provides an optimal state estimation framework:

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K\big(y(t) - C\hat{x}(t)\big), \tag{7}$$

$$r_k(t) = y(t) - C\hat{x}(t), \tag{8}$$

where K is the Kalman gain minimizing the estimation error covariance. The residual $r_k(t)$, also called the innovation, follows a zero-mean white-noise distribution under nominal conditions. Faults typically induce a bias or change in the covariance structure of $r_k(t)$. Statistical tests such as χ^2 or Generalized Likelihood Ratio (GLR) tests are often applied to the innovations to detect abnormal deviations [8,10]. Nonlinear variants, such as the Extended and Unscented Kalman Filters (EKF, UKF), extend these principles to nonlinear industrial systems.

4.3. Limitations of Classical Approaches

Despite their solid theoretical foundations, classical FDD techniques face several limitations in modern industrial environments:

- Model Dependency: High-fidelity process models are difficult and costly to derive. Inaccurate modeling or parameter drift can cause residuals to respond to modeling errors rather than real faults [9].
- Sensitivity to Operating Conditions: Classical methods are often designed around a single operating point, making them less robust to varying loads, nonlinearities, and time-varying parameters.
- Noise and Uncertainty: Designing residual generators that are fault-sensitive yet noise-robust is inherently challenging. The trade-off between sensitivity and robustness complicates threshold selection [8].
- Scalability: In large-scale industrial systems, modeling every subsystem and designing observers for all components is impractical. Fault propagation and interaction effects are difficult to capture analytically.

These challenges have motivated the research community to explore data-driven and AI-based fault detection approaches, which learn fault signatures directly from process data. Machine learning and deep learning models can generalize across operating conditions and provide adaptive, scalable solutions to complex fault patterns—addressed in the next section of this paper.

5. Applications of AI-Based Fault Detection Across Industry, Energy, and CPS/IoT

Advanced fault and error detection techniques are being applied across diverse domains to improve system reliability and safety. We highlight four key areas: industrial systems, energy systems, cyber-physical/IoT systems, and cyber-security, where we focus on describing representative applications, ML/AI techniques, recent case studies, and domain-specific challenges.

5.1. Industrial Systems

In manufacturing plants and process industries, AI-driven fault detection has become integral for predictive maintenance and quality control. Modern production lines employ machine learning models to continuously monitor equipment condition via diverse sensor streams (vibration, temperature, pressure, etc.) and detect subtle anomalies that indicate incipient faults. Advanced deep learning methods, such as convolutional neural networks

(CNNs) and recurrent networks, are widely used to capture the complex temporal patterns in time-series sensor data. Recently, Transformer-based neural architectures have also been applied to rotating machinery (motors, bearings, gearboxes) fault diagnosis, achieving high accuracy in detecting early-stage bearing defects under varying loads [11]. Beyond these, graph neural networks (GNNs) have emerged as a novel technique to exploit relational structure in the data for example, by representing sensor signals or machine components as nodes in a graph, GNN models can fuse information from correlated sources. A recent study on bearing fault detection demonstrated that a GNN-based approach improved the area-under-curve by over 6% compared to conventional deep models, effectively detecting faults with minimal signature by leveraging signal relationships [12]. In addition to vibration-based analysis, acoustic monitoring is gaining traction: machine sounds captured via microphones can reveal faults through their acoustic signatures [13]. Unlike physical sensors, audio-based techniques are non-invasive (no hardware modification to the machine) and can be used when other sensors are impractical; modern AI methods apply spectral and time-frequency analysis on sound signals to classify equipment faults in real-time, although robust denoising is required in noisy factory environments to ensure a high signal-to-noise ratio for reliable feature extraction [14]. Computer vision is another mature AI application in manufacturing, used for automated visual inspection of products on assembly lines. Here, CNN-based vision systems can identify surface defects (e.g., scratches, dents) or assembly errors in real-time, often with high precision, enabling 100% quality control of components [15]. Thermal imaging cameras are likewise employed to detect abnormal heat patterns in equipment (e.g., identifying hotspots in motors or electrical panels), which may signal impending failures [16]. In practice, these vision-based solutions help maintain product quality and detect process anomalies that are not evident from sensor readings alone.

More recently, intelligent manufacturing systems have witnessed the emergence of multimodal and foundation-model-based approaches for fault detection and diagnosis, particularly in complex and highly automated environments such as CNC machining. Unlike traditional single-modality monitoring, these methods integrate heterogeneous data sources—including visual inspection, sensor signals, machine logs, and textual process descriptions—within unified learning architectures. A representative example is the CNC-VLM framework, an industrial vision–language model optimized via Reinforcement Learning from Human Feedback (RLHF) for imbalanced CNC fault detection. By jointly processing visual features from machining operations and semantic descriptions of fault patterns, CNC-VLM demonstrates enhanced robustness under severe class imbalance and limited fault samples, outperforming conventional vision-only or signal-based models. This class of multimodal large models highlights a paradigm shift toward intelligent manufacturing systems capable of contextual reasoning, cross-modal knowledge transfer, and adaptive diagnostic interpretation, aligning with the broader trend of foundation and multimodal AI in industrial FDD.

As AI models proliferate on the factory floor, explainable AI (XAI) techniques are increasingly explored to ensure the models' predictions are interpretable to human engineers [17]. This is crucial in industrial settings for gaining operators' trust in AI-driven diagnostics and for compliance with safety standards. Data-driven fault detection in industry must also overcome several key challenges, and new emerging techniques have been developed to address them [18,19]. One major hurdle is the scarcity of labeled failure data catastrophic machinery breakdowns are rare, making it hard to train conventional supervised models on every fault mode [20]. To mitigate this, researchers are combining traditional statistical process control with modern machine learning and developing hybrid and semi-supervised solutions [21]. A representative case study is the use of deep neural

networks for early fault detection in industrial pump systems, which demonstrated that data-driven models can predict pump failures several hours in advance, thereby reducing unplanned downtime by enabling timely maintenance [22]. Similarly, in chemical process industries, advanced signal decomposition techniques have been paired with ML: for instance, a dynamic Independent Component Analysis (ICA) combined with a machine learning approach was able to detect anomalies in a chemical plant with near-perfect detection rates, outperforming traditional fixed-threshold methods in catching subtle process deviations [23]. Beyond supervised learning on labeled faults, anomaly detection via unsupervised and self-supervised methods is common when labeled fault examples are scarce [24]. Techniques such as autoencoder networks, one-class SVMs, and clustering-based outlier detection are used to learn a model of "normal" operation and flag deviations without requiring prior fault examples. Indeed, a variety of deep autoencoder architectures have been employed to reconstruct normal sensor patterns and identify when the reconstruction error spikes, indicating an anomaly [25,26]. To further address data scarcity, transfer learning and data augmentation strategies are proving effective. In transfer learning, models pre-trained on one machine or operating condition can be adapted to another machine with minimal new data, leveraging commonalities in their behavior. This prevents having to train from scratch for each piece of equipment. Data augmentation using generative models has also shown promise: generative adversarial networks (GANs) can synthesize realistic fault data for training [27,28], alleviating the class imbalance between normal and fault examples. These approaches improve model robustness across different operating conditions (e.g., a model trained on one motor type can be quickly fine-tuned to a new motor with only a few real fault samples). In addition, few-shot learning and meta-learning techniques have gained traction to tackle the paucity of failure examples [29,30]. Instead of requiring large datasets, few-shot approaches train models to rapidly learn new fault classes from only a handful of examples by leveraging knowledge learned from related diagnostic tasks [31]. Such methods (e.g., model-agnostic meta-learning and metric learning frameworks) enable an ML system to generalize to unseen fault modes with minimal data, which is highly valuable in industrial domains where we often encounter novel failure patterns [32]. Another emerging trend is federated learning for distributed fault diagnosis [33]. In many industrial scenarios, data reside on different machines or plants and cannot be easily centralized due to privacy, security, or bandwidth constraints [34]. Federated learning allows multiple sites to collaboratively train a global anomaly detection model without sharing raw data—only model updates are exchanged—thereby protecting proprietary information [35]. Recent works suggest that federated training can also help capture a wider diversity of operating conditions from different factories, improving the generalization of fault detectors across sites. For example, a cross-factory study on paper-making machines combined transfer learning with federated learning to diagnose paper breakage faults under varying process conditions: the approach used parameter-sharing and feature transfer to adapt a model across different paper production lines, and applied federated averaging to integrate learning from each line. This federated fault diagnosis system achieved over 94% fault classification accuracy across different operating conditions, significantly better than single-line models, and it maintained data privacy by using model compression techniques during aggregation [36]. Overall, these innovations in leveraging data from multiple sources (while respecting confidentiality) and synthesizing additional training examples have substantially improved the reliability of AI models for fault detection in situations with limited fault data [37]. Industrial environments also impose strict performance and safety requirements on fault detection systems, spurring further technical advances in the field [38]. False alarms must be minimized to avoid unnecessary production halts, yet missed detections can lead to costly machine damage

or safety incidents. Achieving a low false-positive rate at the same time as a high true fault detection rate is difficult, especially under highly imbalanced data [39]. To strike this balance, researchers have explored cost-sensitive learning (assigning higher penalty to misclassifying a fault vs. a false alarm) and ensemble methods that combine multiple models for cross-verification of an anomaly before triggering an alert [40]. By fusing outputs from diverse detectors—for example, an ensemble might require concurrence between a vibration-based model and an acoustic-based model—the system can reduce spurious alarms [41]. Another critical concern is real-time responsiveness: many industrial control systems (PLC/SCADA) require fault detection and diagnosis within milliseconds to reliably trigger protective actions [42,43]. Therefore, AI models for FDD must be optimized for fast inference and often deployed on edge computing hardware located near the machines [44]. Techniques such as model quantization, knowledge distillation, and efficient neural architecture design (e.g., using smaller CNN kernels or shallow networks) are applied to achieve the low latency and small footprint necessary for on-device deployment [45–47]. In some cases, dedicated edge AI accelerators or microcontroller-based ML (TinyML) are used to run inference directly on sensors or controllers, eliminating network delays [48]. Moreover, any AI-driven diagnostic tool needs to be integrated with existing industrial automation workflows and comply with industry standards [49] (for instance, interfacing with protocols like OPC-UA or MQTT for IIoT data streams). This integration ensures that anomaly alerts from an ML model can seamlessly initiate appropriate responses such as equipment shutdowns, maintenance work orders, or human operator notifications through the plant's HMIs [50] (Human-Machine Interfaces). Finally, as noted above, interpretability remains important—plant operators and maintenance staff need actionable explanations rather than opaque "black-box" alerts [51]. In practice, this means the AI system should ideally indicate *which component is likely failing and why*. Recent research efforts in XAI for industrial fault diagnosis [52,53] are addressing this by extracting human-understandable features from complex models. For example, an explainable model might highlight that a specific vibration frequency band or waveform pattern is the reason for a bearing failure prediction, linking the ML output to known failure signatures that engineers recognize. Providing such reasoning (e.g., "high energy at 5 kHz suggests an inner-race bearing defect") helps experts validate and trust the model's outputs [54]. Despite these challenges and requirements, numerous industrial trials and deployments have shown that AI-driven fault detection can significantly reduce maintenance costs and unplanned downtime, while improving safety. These successes validate the promise of machine learning in enabling intelligent manufacturing and ushering in the era of Industry 4.0.

Table 1 summarizes representative AI-based FDD techniques across industrial domains, highlighting their strengths, limitations, and key references to support comparison in terms of performance, robustness, data requirements, and deployment complexity.

Table 1. Summary of AI-based Fault Detection Techniques in Industrial Applications.

Method	Main Application Domain	Key Strengths	Limitations	References
CNN/RNN/Transformer	Vibration-based condition monitoring	Captures temporal and spatial patterns; high accuracy	Requires large labeled datasets	[11,25]
Graph Neural Networks (GNN)	Multi-sensor and relational data fusion	Exploits sensor correlations; robust to missing data	High computational cost	[12]

Table 1. *Cont.*

Method	Main Application Domain	Key Strengths	Limitations	References
Acoustic Monitoring (AI-based)	Non-invasive sound-based diagnosis	Simple deployment; real-time detection	Sensitive to noise, requires denoising	[14]
Computer Vision (Thermal/Visual)	Surface and thermal defect inspection	High precision; visual interpretability	Limited to visible defects	[15,16]
Unsupervised/Self-supervised Learning	General anomaly detection	Works with unlabeled data; adaptable	May produce false alarms	[24,26]
Transfer/Federated Learning	Cross-factory and distributed systems	Enhances generalization; preserves privacy	Communication and synchronization overhead	[35,36]
Explainable AI (XAI)	Model transparency and trust	Improves interpretability; compliance with standards	Trade-off with accuracy	[17,52]
Edge/TinyML	Real-time on-device deployment	Low latency; energy efficient	Limited model complexity	[46,48]
Vision–Language/ Multimodal Foundation Models	Intelligent manufacturing (CNC machining)	Multimodal reasoning; robustness to class imbalance; contextual fault interpretation	High computational cost; training complexity	[55]

Figure 2 outlines the overall framework, from multi-modal data sources to learning paradigms and deployment constraints, highlighting key aspects such as real-time operation, privacy, and explainability.

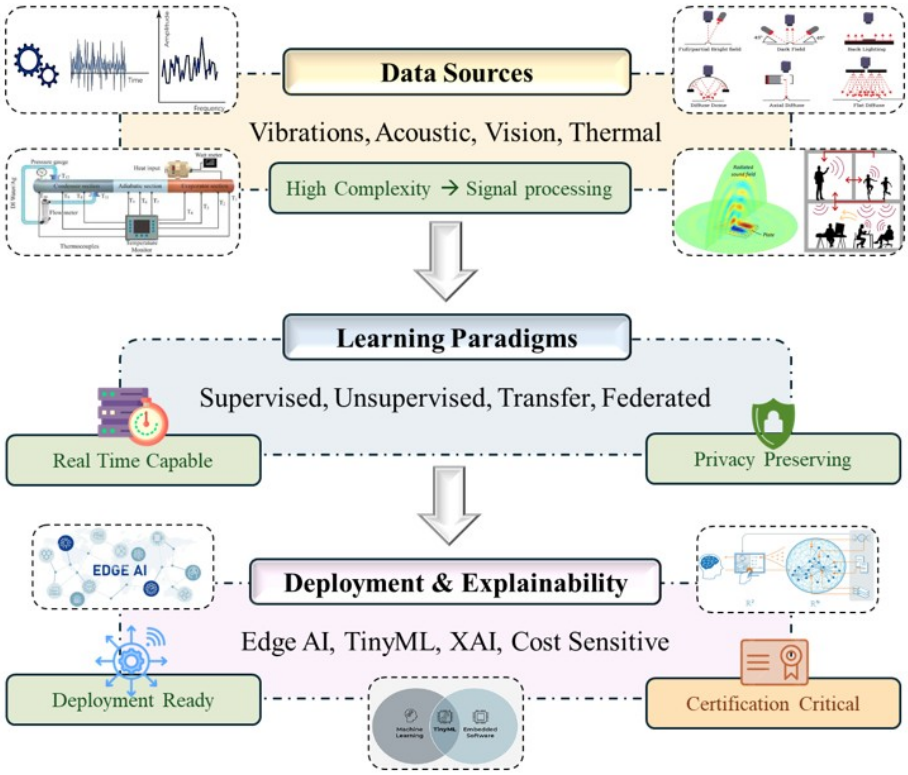


Figure 2. Taxonomy of AI-driven FDD techniques, highlighting not only model families but also their typical data sources, deployment constraints, and explainability mechanisms across Industrial Systems.

5.2. Energy Systems

Energy infrastructures, including electric power grids, renewable energy plants, and energy storage systems, are increasingly adopting artificial intelligence (AI) to enhance fault detection, improve reliability, and ensure operational continuity [56]. Traditional protection mechanisms such as distance and differential relays, while effective, are constrained by their predefined thresholds and limited adaptability to evolving grid dynamics [57]. In contrast, AI-based fault detection enables adaptive, data-driven decision-making capable of capturing complex nonlinear patterns in sensor data streams [58]. High-voltage transmission networks and distribution systems are now instrumented with phasor measurement units (PMUs) and IoT-based monitoring devices that continuously acquire high-frequency data on voltages, currents, and equipment states [59]. Machine learning (ML) models, both supervised classifiers and unsupervised anomaly detectors, can process these multivariate data streams to identify incipient events such as line faults, transformer failures, or voltage instabilities with significantly reduced response times compared to static rule-based methods [60]. Recent advances have focused on hybrid deep learning architectures that capture spatiotemporal dependencies in grid signals. An interesting study, demonstrated the use of CNN–RNN hybrid networks, comparing CNN-LSTM, CNN-RNN, and CNN-GRU configurations for real-time analysis of electrical measurements, showing that the CNN-GRU variant achieved the highest classification accuracy with minimal prediction loss [61]. These models effectively learn correlations between voltage and current oscillations preceding faults, enabling operators to isolate a malfunctioning line or substation before a cascading blackout occurs. Reinforcement learning (RL) techniques have also emerged as promising tools for autonomous fault management [62]. For example, a graph-based deep RL controller incorporating graph attention networks and a Soft Actor–Critic agent was shown to achieve millisecond-level restoration in power distribution networks [63]. Such agents can learn optimal switching and load-shedding strategies; however, ensuring safety, stability, and compliance with operational constraints remains essential, and AI systems are therefore typically employed under human or traditional relay supervision [64,65]. The same paradigm shift is occurring across renewable energy systems, where AI-driven fault detection contributes to the reliability and performance optimization of assets such as wind turbines and solar photovoltaic (PV) farms [66–68]. Wind turbine farms produce vast supervisory control and data acquisition (SCADA) logs and high-frequency vibration data, which provide valuable information for predictive maintenance [69]. Modern turbines, equipped with dozens of sensors measuring temperature, pressure, mechanical strain, and power output, allow ML algorithms to detect subtle signatures of component degradation [70]. A recent and interesting example is the Transformer-based HARO model (Huber–Adam Regression Optimizer), which combines a Transformer neural network with Lasso regression and the Adam optimizer to detect early-stage turbine faults [71]. This method demonstrated the ability to forecast gearbox or bearing failures several days in advance, facilitating proactive maintenance interventions and minimizing unplanned downtime. Moreover, combining physics-based models with ML further strengthens the interpretability and reliability of turbine health assessments [72]. Similarly, solar PV farms benefit from AI-based monitoring techniques that leverage both visual and electrical data. Computer vision algorithms applied to drone or satellite imagery can identify defects such as cracks, soiling, or electrical anomalies in panels [73–75]. Regarding this, a recent work proposed an explainable CNN-based classifier capable of distinguishing between physical and electrical faults in PV modules, achieving over 91% accuracy [76]. The integration of thermal-imaging drones allows for rapid, autonomous inspection of large-scale solar farms, locating hotspots and defective panels in minutes [77]. Beyond image analysis, time-series ML models are increasingly utilized to monitor inverter outputs, string currents,

and voltage levels, detecting patterns associated with shading, wiring faults, or inverter degradation [78,79]. The fusion of visual and electrical analytics thus provides comprehensive fault coverage, enabling real-time diagnostics and maximizing energy yield [80]. AI-based fault detection has also proven instrumental in battery energy storage systems, from grid-scale lithium-ion banks to electric vehicle (EV) battery packs [81]. These systems face risks associated with overheating, internal short circuits, or cell degradation, potentially leading to thermal runaway [82]. Modern battery management systems (BMS) incorporate ML algorithms trained on cell voltage, temperature, impedance, and pressure data to detect early deviations from normal operational profiles [83]. Recently, researchers focus the attention on study and developed a wireless sensing approach, where AI algorithms interpret Wi-Fi and millimeter-wave signal reflections from battery modules to infer internal temperature buildup non-intrusively [84]. This methods achieved an high accuracy in detecting thermal anomalies that precede runaway events, offering a safe, sensorless means of continuous monitoring. Integrating such predictive intelligence into BMS frameworks enables early interventions - such as cell isolation or active cooling to improve both operational safety and system longevity. Despite their advantages, AI-driven fault detection systems in energy domains face several technical and operational challenges [85]. First, the stringent latency requirements of electrical protection demand models capable of executing decisions within a few power cycles (20–40 milliseconds) [86]. To meet these constraints, edge computing architectures are increasingly employed, deploying lightweight deep learning models directly within substations or controllers to enable local inference without cloud dependencies [87]. Second, fault events in critical infrastructure are rare, resulting in class imbalance that complicates model training and validation [88]. Synthetic data generation, physics-informed simulations, and domain randomization are therefore used to enrich datasets and improve generalization [89]. Moreover, the consequences of misclassification are severe: false positives can cause unnecessary equipment shutdowns and service interruptions, while false negatives may lead to physical damage or widespread outages [90]. Consequently, AI-based detectors are often integrated in supervisory configurations, complementing but not replacing conventional protection logic until their reliability is thoroughly validated. Another major consideration is interpretability [91]. Operators require transparent and explainable AI models to justify critical decisions. In this context, a recent study, proposed an interpretable adaptive fault detection framework for smart grids based on belief rule bases, which achieves high detection accuracy while providing human-understandable reasoning behind each decision [92]. Finally, given the cyber-physical nature of energy systems, robustness against malicious interference is paramount. Recent work in cybersecurity-aware fault detection focuses on identifying false data injection attacks designed to mimic legitimate fault signatures [93], ensuring that AI models can discriminate between true equipment failures and adversarial manipulations [94]. The integration of AI into energy system fault detection enables more predictive, adaptive, and efficient monitoring of critical infrastructure [95]. When combined with edge computing, explainable models, and cyber-resilient architectures, these intelligent detection systems are poised to become foundational components of future smart energy grids and renewable energy networks.

Table 2 summarizes representative AI-based FDD approaches for energy systems, organized by application domain and highlighting strengths, limitations, and key references, with emphasis on trade-offs among accuracy, interpretability, computational complexity, and real-time deployment.

Figure 3 presents a high-level taxonomy of AI-driven fault detection and diagnosis approaches, linking data sources, learning paradigms, and deployment constraints, with emphasis on real-time operation, physics-informed modeling, and certification requirements.

Table 2. Summary of AI-based Fault Detection Techniques in Energy Systems.

Method	Main Application Domain	Key Strengths	Limitations	References
CNN/RNN/Transformer	Grid and renewable sensor data analysis (e.g., voltage, current, SCADA logs)	Captures spatiotemporal patterns; early fault prediction	Requires large labeled datasets; limited interpretability	[61,71]
Graph Neural Networks (GNN)	Topology-aware power grid fault detection	Incorporates relational structure of grid; scalable to complex networks	High computational complexity	[63]
Reinforcement Learning	Grid reconfiguration and autonomous fault management	Learns optimal control actions; fast restoration policies	Hard to enforce safety constraints; data-hungry	[62,63]
Computer Vision (Thermal/Visual)	PV module inspection (cracks, soiling, hotspots)	High fault localization accuracy; drone-compatible; interpretable outputs	Dependent on weather/image quality	[76,77]
Time-Series Anomaly Detection	Solar inverters, battery diagnostics, SCADA streams	Detects subtle electrical anomalies; applicable across modalities	Needs tuning; possible false alarms	[78,83]
Wireless Sensing + AI	Non-intrusive battery pack thermal monitoring	Sensorless monitoring; early thermal fault detection	Novel technique; limited field validation	[84]
Explainable AI (XAI)	Grid and PV diagnostics (interpretable decisions)	Improves operator trust; rule-based justification	May reduce model complexity/accuracy	[92]
Edge/Embedded AI	Real-time on-device fault detection (grids, solar, BMS)	Fast inference (<40 ms); removes cloud dependency	Constrained resources; requires model compression	[86,87]
Cybersecurity-aware Detection	Smart grid anomaly discrimination (FDIA vs. real fault)	Defends against spoofing/data injection; improves trust	Needs secure integration; adversarial testing	[93,94]

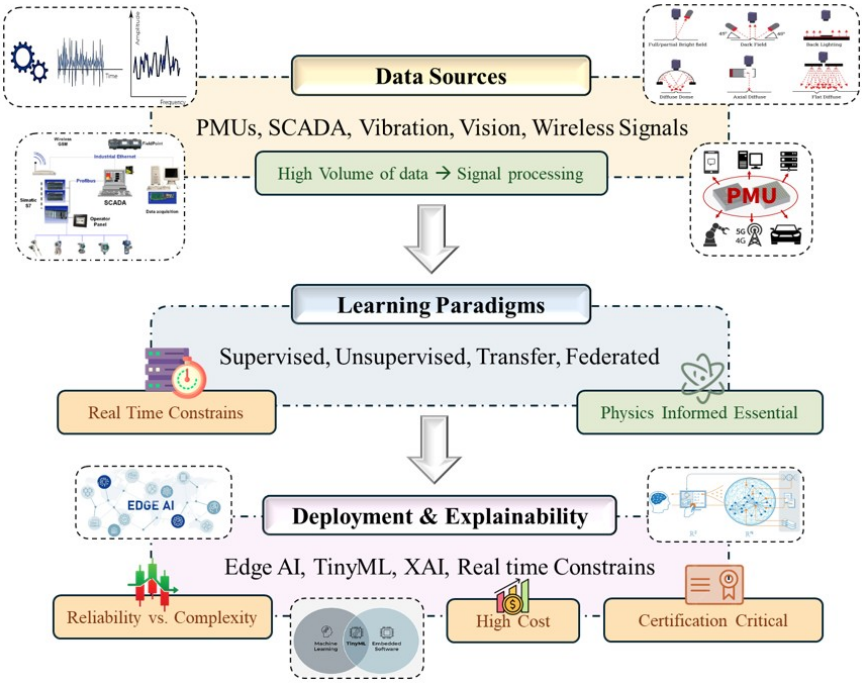


Figure 3. Taxonomy of AI-driven FDD techniques, highlighting not only model families but also their typical data sources, deployment constraints, and explainability mechanisms across Energy Systems.

5.3. Cyber-Physical and IoT Systems

Cyber-physical systems (CPS) and the Internet of Things (IoT) constitute a broad class of distributed, networked environments encompassing smart buildings, autonomous vehicles, industrial automation, and city-scale sensor networks [96]. Within such ecosystems, fault and anomaly detection must often occur in a decentralized manner at the network edge, as real-time responses are essential to maintain safety and operational continuity [97]. The intrinsic scale and heterogeneity of IoT devices, often numbering in the thousands and spanning various sensing modalities such as temperature, vibration, images, or flow, pose significant challenges to centralized fault detection as shown in previous industrial system section. Traditional cloud-based monitoring, which requires continuous transmission of raw data to a central server, is typically infeasible due to bandwidth limitations, latency constraints, and privacy concerns [98]. Consequently, recent advancements have emphasized edge intelligence, federated learning, and resource-efficient models to enable distributed and privacy-preserving fault detection across heterogeneous CPS and IoT infrastructures [99]. A key technological development in this domain is the adoption of Federated Learning (FL) for anomaly detection [100]. Federated approaches allow distributed devices or gateways to collaboratively train models without sharing raw data, thus preserving privacy and reducing network overhead [101]. Each node contributes model updates that are aggregated into a global model capable of detecting faults locally while benefiting from collective learning. Studies have shown that federated anomaly detectors can achieve accuracy comparable to centralized systems [102]. For instance, the FedGroup framework demonstrated that group-based federated learning in smart-home IoT environments could match or even surpass centralized baselines in detection accuracy while maintaining privacy guarantees [103]. Similarly, FL-based methods have been successfully applied in smart grid and sensor network contexts, where the ability to retain data locally avoids massive upstream traffic [104]. Ongoing research further explores FL robustness against non-identically distributed data, device heterogeneity, and limited label availability, aiming to ensure reliable model convergence across distributed IoT nodes [105–107]. Complementary to FL, the TinyML paradigm has emerged to address on-device analytics in highly resource-constrained environments [108]. By deploying lightweight machine learning models directly on microcontrollers, TinyML enables real-time fault detection even in the absence of network connectivity [109]. An interesting work, demonstrated a one-dimensional convolutional neural network (1D-CNN) for vibration-based machine fault diagnosis implemented on low-cost embedded hardware such as Raspberry Pi and ESP32 boards [110]. The proposed model, enhanced with transfer learning, achieved effective cross-domain generalization across equipment types. These results highlight the feasibility of executing efficient deep learning models on microcontroller-class devices for continuous anomaly monitoring [111]. Similar studies have employed compact autoencoders and quantized CNNs on Cortex-M architectures, illustrating that even limited hardware can sustain adaptive diagnostic capabilities [112–114]. TinyML thus mitigates latency and reliability concerns by enabling autonomous local detection while minimizing dependency on remote computation [115]. The increasing interconnectivity of CPS and IoT systems has also prompted the adoption of graph-based and multi-modal learning for fault detection [116]. Graph neural networks (GNNs) can model the system's topology, treating devices as nodes and their interactions as edges, thereby capturing spatial and relational dependencies between sensors [117]. The IoT-GRAF framework exemplifies this approach by integrating heterogeneous graph representations that combine physical sensor data with network traffic information to detect both operational faults and cyber anomalies [118]. This fusion of cyber and physical data domains allows the detection system to differentiate between natural component malfunctions and malicious network activities [119]. In an

IoT greenhouse case study, multi-modal graph-based learning improved fault detection performance by over 20% compared to unimodal baselines [120]. Such approaches demonstrate the power of contextual reasoning across diverse data modalities and inter-device dependencies, paving the way for more holistic fault detection architectures. In parallel, there has been a growing emphasis on privacy-preserving and unsupervised anomaly detection methods [121]. Because labeled anomalies are rare and manual annotation is infeasible at scale, autoencoders, one-class SVMs, and generative models are employed to learn "normal" operational patterns from unlabeled data streams [122]. Federated autoencoder frameworks, for instance, have been applied to distributed power systems, enabling local model training without centralizing sensitive measurements [123]. Other emerging strategies employ hyperdimensional computing and sketch-based techniques to efficiently detect outliers in streaming data with minimal computation [124]. These methods address both the scarcity of labels and the privacy restrictions inherent to large-scale IoT networks. The applicability of these AI-driven methods spans a variety of real-world domains. In smart buildings, machine learning–based Fault Detection and Diagnostics (FDD) frameworks are integrated into building management systems to detect issues such as sensor drift, valve malfunctions, and suboptimal control behavior [125]. Decision-tree and rule-based analytics on time-series measurements—such as temperature, flow rate, and valve position—enable early identification of anomalies, preventing energy losses and enhancing occupant comfort [126]. Similarly, IoT-based monitoring in municipal infrastructure has revolutionized fault detection in water networks [127]. A 2024 study employing vibration sensors and ensemble models (e.g., XGBoost) achieved an high accuracy in identifying pipeline leaks [128], demonstrating the potential for automated, high-fidelity leak diagnostics compared to traditional manual inspection. In transportation and autonomous systems, on-board anomaly detection models continuously monitor sensor health and actuator performance [129]. These diagnostic networks are capable of identifying early-stage failures—such as actuator degradation or sensor bias—allowing proactive fault mitigation [130]. Edge-embedded inference and fail-operational architectures ensure uninterrupted system safety, distinguishing genuine faults from environmental outliers (e.g., temporary LiDAR occlusion or sensor glare) [131]. Despite these advancements, CPS and IoT environments present distinctive challenges for reliable and scalable fault detection. The first concerns resource constraints: many IoT nodes operate on limited computational power and battery capacity, necessitating extreme model optimization through quantization, pruning, and approximate inference [132]. There exists a critical trade-off between detection accuracy and energy efficiency, as overcomplex models can deplete device resources [133]. Secondly, the decentralized nature of IoT introduces difficulties in global situational awareness, since individual nodes have only partial observations [134]. Coordinated decision-making thus relies on multi-tier architectures, where local nodes detect anomalies independently, fog or edge gateways aggregate alerts, and cloud layers provide global analysis [135]. This hierarchical structure balances latency and scalability but introduces synchronization and model consistency challenges. Moreover, communication efficiency is vital: transmitting all sensor data would overload networks, so edge models are designed to transmit only aggregated statistics or alerts when anomaly thresholds are exceeded, prioritizing critical information and reducing congestion [136]. Data heterogeneity and quality further complicate anomaly detection. IoT sensors often suffer from drift, noise, or degradation, while diverse device types yield inconsistent data formats and sampling rates [137]. Robust fault detection requires models capable of domain adaptation and online learning to remain effective despite sensor recalibration or replacement [138]. The boundary between fault detection and cybersecurity also blurs in CPS: adversarial attacks such as false data injection can mimic legitimate sensor faults. Discriminating

between natural failures and malicious tampering thus necessitates the integration of secure, trust-aware detection frameworks [139]. Finally, scalability remains a persistent issue [140]. As IoT deployments expand to thousands or millions of nodes, centralized data aggregation becomes untenable. Federated, hierarchical, and swarm intelligence–inspired paradigms offer scalable solutions, though they introduce new concerns around model drift, synchronization, and versioning across distributed learners [141]. Despite these challenges, AI-driven fault detection in CPS and IoT ecosystems is rapidly advancing toward autonomy, resilience, and scalability. Recent research has also extended AI-based FDD to intelligent transportation systems, particularly railway and autonomous vehicle platforms. For instance, adaptive fault diagnosis frameworks based on Large Language Models (LLMs) have been proposed for railway vehicle on-board controllers, enabling contextual reasoning over multivariate sensor data, logs, and operational constraints. These approaches demonstrate the potential of LLMs to support fault diagnosis and decision-making in complex transportation CPS, further broadening the applicability of AI-driven FDD beyond traditional industrial and energy domains [142]. The confluence of federated learning, TinyML, graph-based reasoning, and privacy-aware unsupervised analytics is reshaping how distributed systems self-monitor and adapt. Real-world deployments in smart buildings, urban infrastructures, and autonomous vehicles demonstrate tangible benefits, including reduced downtime, enhanced energy efficiency, and improved safety [143]. Future CPS and IoT systems will likely evolve into self-healing networks where edge intelligence, explainable models, and secure coordination mechanisms jointly ensure that cyber-physical infrastructures remain dependable under both natural faults and adversarial threats.

Table 3 summarizes representative AI-based FDD techniques for CPS and IoT, categorized by deployment and architectural paradigms and highlighting trade-offs among privacy, scalability, computational constraints, and detection reliability.

Table 3. Summary of AI-based Fault Detection Techniques in Cyber-Physical and IoT Systems.

Method	Main Application Domain	Key Strengths	Limitations	References
Federated Learning (FL)	Smart homes, sensor networks, smart grids	Privacy-preserving; decentralized training; scalable	Model drift; requires communication synchronization	[100,103,104]
TinyML	Embedded sensors, edge devices, autonomous systems	On-device inference; ultra-low power; low latency	Limited model capacity and memory	[110,111,115]
Graph Neural Networks (GNN)	IoT topologies, cyber-physical fusion	Captures inter-device relationships; multi-modal analysis	High complexity; requires system graph modeling	[116,118]
Unsupervised/Self-supervised Models	Smart infrastructure, streaming analytics	No need for labeled data; adaptive to evolving patterns	Prone to false alarms; difficult to interpret	[121,123]
Hyperdimensional/ Sketch-based	Resource-constrained streaming nodes	Efficient memory footprint; fast anomaly computation	Lower accuracy; requires calibration	[124]
Edge/Fog Hierarchical Architectures	Smart cities, industrial IoT, transportation CPS	Scalable; balances latency and bandwidth; modular	Model consistency across layers; fog infrastructure needed	[135,141]

Table 3. *Cont.*

Method	Main Application Domain	Key Strengths	Limitations	References
Secure and Trust-aware Detection	Cyber-physical intrusion detection	Distinguishes faults from attacks; trust modeling	Complexity in distinguishing ambiguous scenarios	[139]
Domain Adaptation/ Online Learning	Multi-device heterogeneous environments	Adapts to sensor drift and new devices	Requires continuous training and monitoring	[138]

Figure 4 summarizes a taxonomy of AI-driven FDD approaches for cyber-physical and IoT systems, relating data sources, learning paradigms, and deployment aspects under real-time, low-power, and security constraints.

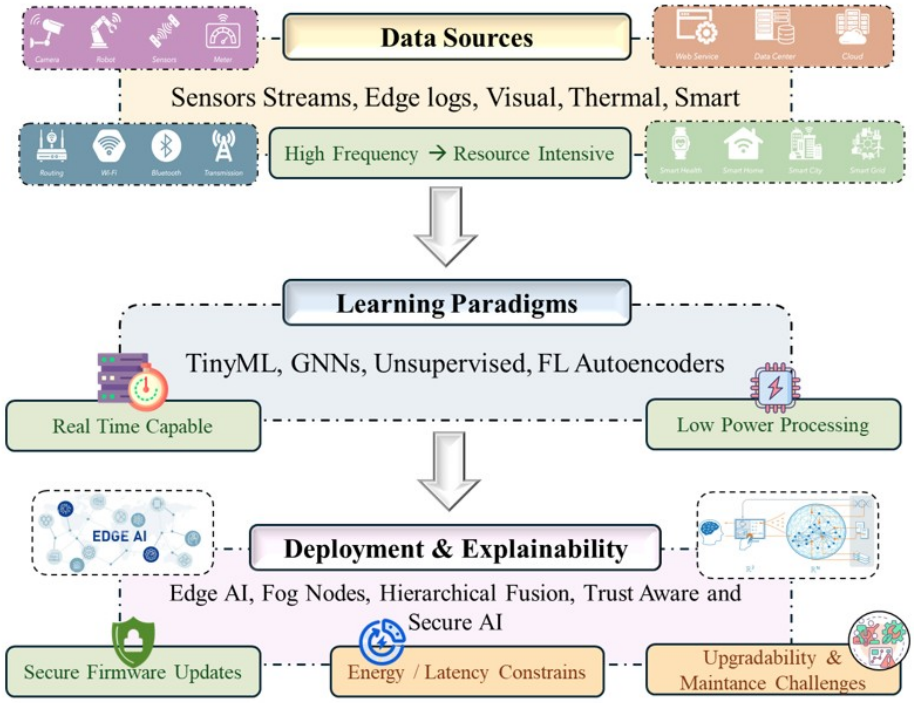


Figure 4. Taxonomy of AI-driven FDD techniques, highlighting not only model families but also their typical data sources, deployment constraints, and explainability mechanisms across Cyber-Physical and IoT Systems.

5.4. Cybersecurity

In modern enterprise IT networks and critical digital infrastructures, timely detection of malicious intrusions and anomalies is vital to ensure data integrity, service continuity, and operational resilience. Classical cybersecurity monitoring has traditionally relied on signature-based intrusion detection systems (IDS) and rule-based firewalls, which match known attack patterns and static thresholds to identify malicious behavior. While such methods provide clear and interpretable logic for analysts, they are inherently limited to previously observed threats and fail to generalize to evolving, zero-day, or polymorphic attacks [144]. Consequently, sophisticated adversaries can easily evade detection by modifying attack payloads or communication patterns. The integration of AI-based Fault Detection and Diagnosis (FDD) into cybersecurity represents a paradigm shift toward adaptive, predictive, and autonomous defense mechanisms. Machine learning and deep learning models trained on large-scale datasets of network traffic, system logs, or user

activities can identify subtle deviations from normal behavior, revealing early signs of intrusion or compromise even without predefined signatures [145]. In this sense, cybersecurity anomaly detection mirrors the concept of FDD in physical systems—detecting deviations from normal operating states that indicate latent faults or failures. Recent years have witnessed a surge in the application of advanced deep learning architectures for intrusion and anomaly detection. Convolutional and recurrent neural networks (CNNs, LSTMs, GRUs) remain fundamental for time-series packet analysis, while emerging Transformer-based models (e.g., BERT4IDS, CyberViT) leverage attention mechanisms to capture long-range dependencies in network flows, yielding superior detection of stealthy attacks such as advanced persistent threats (APTs) and data exfiltration campaigns [146]. Moreover, Graph Neural Networks (GNNs) have gained prominence for modeling communication topologies—representing hosts, sensors, or applications as interconnected nodes—and have demonstrated effectiveness in identifying coordinated, multi-hop attacks or botnet propagation in IoT and industrial networks [147,148]. Parallel advances in Generative AI, such as the use of Generative Adversarial Networks (GANs) and diffusion models, are also being explored to generate synthetic intrusion samples, augment imbalanced datasets, and model realistic attack behaviors [149,150]. In addition, recent studies have shown the emergence of Large Language Models (LLMs)—including domain-specialized variants like SecGPT and CyberBERT—for log analysis, vulnerability summarization, and cross-modal reasoning in cybersecurity fault detection. These models can process unstructured text (e.g., system logs, incident reports) to perform semantic anomaly detection and facilitate context-aware threat intelligence [151,152]. Unsupervised and semi-supervised approaches, such as autoencoders, contrastive learning, and self-supervised Transformers, further enable detection of rare or previously unseen attack types with minimal labeled data [153,154]. Despite these advancements, several challenges persist. The high class imbalance of security data (where normal traffic vastly outnumbers attacks) often leads to biased models with high false-positive rates. Furthermore, cybersecurity operates in an adversarial environment: attackers can deliberately manipulate inputs or poison datasets to evade AI-based detection. To counter this, research has focused on adversarially robust learning, ensemble models, and uncertainty-aware detection to maintain reliability under attack [155]. The real time constraint remains another major limitation—transformer and graph-based architectures, though powerful, require significant computational resources for inference in high-throughput networks. Therefore, lightweight and edge-deployable models (TinyML, quantized CNNs) are increasingly being investigated for on-device intrusion monitoring in IoT and embedded systems [156]. Explainable AI (XAI) has become a cornerstone of trustworthy AI-based cybersecurity. Analysts require interpretable outputs to validate and act upon model predictions. Recent explainability frameworks integrate methods such as SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-agnostic Explanations), and attention-based visualization to attribute model decisions to relevant network features or log entries [157]. Hybrid explainable architectures that combine rule-based reasoning with deep learning (e.g., knowledge-graph augmented XAI, rule-extraction from transformers) are also emerging, enhancing both interpretability and diagnostic traceability [158]. These systems allow analysts to identify not only that an intrusion occurred but also which features—such as protocol fields, packet timing, or host relationships—triggered the alert. Finally, integrating AI-driven FDD with traditional cybersecurity workflows requires a layered and hybrid approach. In practice, AI detectors are deployed in parallel with classical systems, where rule-based IDS handle known threats, and AI components expand coverage to unknown or anomalous behaviors. This combination provides defense-in-depth, improving both detection breadth and system robustness [159]. In summary, AI-based FDD in cybersecurity enables predictive and au-

tonomous threat detection, but widespread adoption hinges on achieving a balance among accuracy, interpretability, robustness, and computational efficiency [160]. The convergence of deep learning, GNNs, transformers, and explainable AI defines the frontier of intelligent, resilient, and trustworthy cyber fault detection.

5.5. Representative Real-World Engineering Case Studies

While this survey does not introduce new experimental results, several representative studies demonstrate the real-world applicability and validation of AI/ML-based FDD in engineering systems. In industrial environments, CNN- and autoencoder-based diagnostic frameworks have been deployed on centrifugal pumps, rotating machinery, and HVAC systems, achieving early fault detection several hours before threshold-based alarms, thus enabling predictive maintenance and reducing unplanned downtime [22,25]. In energy systems, hybrid physics–ML models and transformer-based architectures have been validated on real SCADA and PMU datasets from power grids and wind farms, demonstrating reliable fault localization and early-stage degradation detection under variable operating conditions [61,71]. These studies confirm the feasibility of AI-based diagnostics under real-time and safety-critical constraints. In CPS/IoT contexts, federated and TinyML-based solutions have been experimentally validated on embedded platforms (e.g., ESP32, Raspberry Pi), showing that lightweight models can perform on-device anomaly detection with millisecond-level latency and limited energy consumption [110,111]. Finally, in cybersecurity, transformer and GNN-based intrusion detection systems have been evaluated on large-scale real traffic datasets and industrial control system (ICS) testbeds, achieving robust detection of zero-day and coordinated attacks while maintaining interpretability through XAI techniques [146–148].

Several of these methodologies originate from or are validated in aerospace contexts, where autonomous FDD has been extensively studied for spacecraft health management, fault-tolerant avionics, and on-board power systems.

Table 4 summarizes representative AI-based FDD paradigms for cybersecurity, organized by analytical scope and highlighting trade-offs among detection accuracy, robustness, interpretability, and computational requirements.

Table 4. Summary of AI-based Fault Detection and Diagnosis Techniques in Cybersecurity.

Method/Paradigm	Main Application Domain	Key Strengths	Limitations	References
Deep Neural Networks (CNN, LSTM, GRU)	Network traffic and packet sequence analysis	Captures temporal and spatial patterns; effective for flow-based intrusion detection	High computational cost; limited explainability	[145]
Transformer-based Models (BERT4IDS, CyberViT)	Advanced Persistent Threat (APT) and exfiltration detection	Long-range dependency modeling; high detection accuracy	Expensive inference; requires large datasets	[146]
Graph Neural Networks (GNN)	Topology-aware attack detection, IoT/ICS networks	Models inter-host relationships; detects coordinated multi-hop attacks	Complex graph construction; scalability issues	[147,148]
Generative Models (GANs, Diffusion)	Synthetic data generation, data augmentation	Addresses class imbalance; simulates realistic attacks	Risk of generating unrealistic samples	[149,150]

Table 4. *Cont.*

Method/Paradigm	Main Application Domain	Key Strengths	Limitations	References
Large Language Models (LLMs, SecGPT, CyberBERT)	Log analysis, vulnerability reasoning, text-based anomaly detection	Handles unstructured data; contextual reasoning and summarization	High resource demand; potential data leakage	[151,152]
Unsupervised/Self-supervised Learning (Autoencoders, Contrastive, Transformers)	Detection of unknown and zero-day attacks	Works with minimal labels; adapts to new threats	Sensitive to noise; false positives possible	[153,154]
Adversarially Robust and Ensemble Models	Resilient intrusion detection under adversarial manipulation	Robustness to evasion and poisoning; uncertainty estimation	Training complexity; reduced efficiency	[155]
Explainable AI (XAI) and Hybrid Reasoning	Trustworthy intrusion detection and analyst support	Interpretable decisions; feature attribution and rule extraction	Performance trade-offs; partial automation	[157,158]
Lightweight Edge Models (TinyML, Quantized CNNs)	On-device intrusion monitoring for IoT/embedded systems	Low latency; deployable on constrained devices	Limited model capacity; reduced accuracy	[156]

Figure 5 depicts a taxonomy of AI-driven FDD techniques for cybersecurity systems, connecting data sources, learning models, and deployment aspects under cost, latency, and trustworthiness constraints.

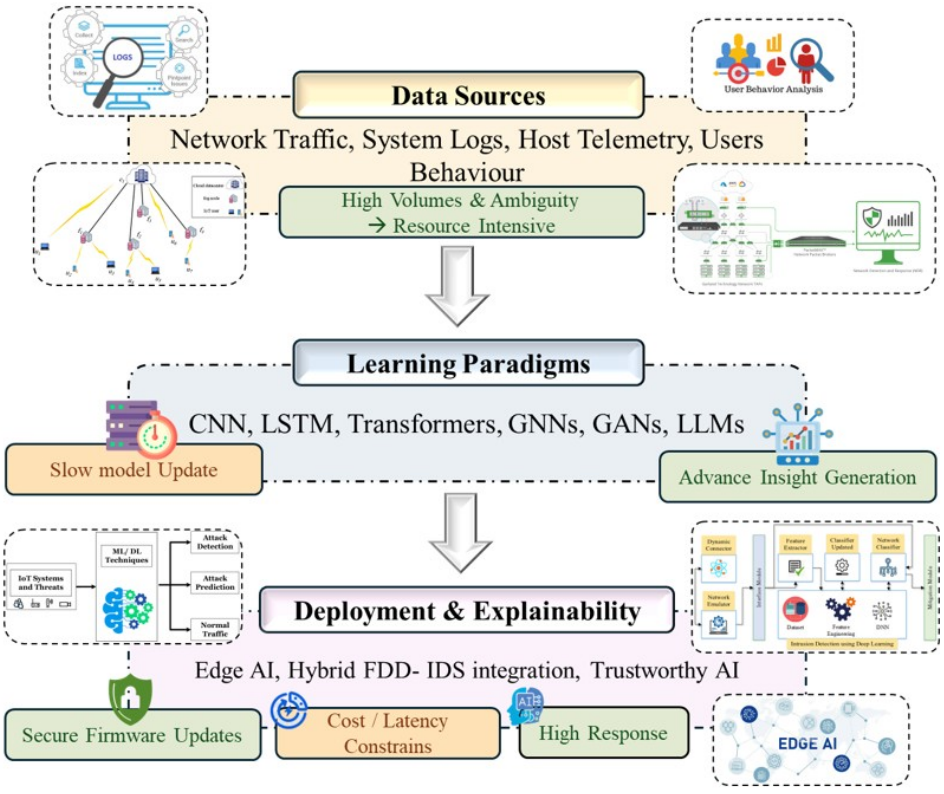


Figure 5. Taxonomy of AI-driven FDD techniques, highlighting not only model families but also their typical data sources, deployment constraints, and explainability mechanisms across Cybersecurity Systems.

6. Challenges, Gaps, and Emerging Trends in Fault Detection Using AI and Classical Techniques

Building on the previous sections, we now provide a comparative analysis of classical (model-based or heuristic) techniques versus AI/ML-based approaches across different domains, followed by a discussion of cross-cutting challenges and gaps (e.g., interpretability, data scarcity, robustness). We then highlight emerging trends that aim to bridge these gaps, such as foundation models, physics-informed learning, and federated learning. This discussion is tightly connected to the techniques and case studies presented earlier, ensuring a cohesive narrative.

6.1. Domain-Specific Comparisons: Classical vs. AI Approaches

Industrial Systems

In industrial automation and manufacturing, classical methods have long relied on mechanistic models and rule-based decision logic (e.g., PID controllers for process control, statistical quality control charts, and scheduled maintenance based on expert knowledge). These approaches are typically grounded in physics of machinery or empirical rules, offering transparency but limited adaptability [161]. AI/ML techniques are transforming this sector by enabling data-driven optimization and predictive analytics at scale. For instance, in predictive maintenance, machine learning models (like anomaly detectors or prognostic models) can learn failure patterns from sensor data to predict equipment faults earlier than threshold-based alarms. AI-driven process control can dynamically optimize operations using reinforcement learning or neural network surrogates, surpassing static setpoints. The benefits include improved decision-making and efficiency, as evidenced by Industry 4.0 implementations. However, challenges remain: many factories have legacy equipment and fragmented data infrastructures, making integration of AI non-trivial. The need to retrofit or interface new ML systems with legacy industrial control systems can introduce technical complexities and compatibility issues [162]. Moreover, safety and reliability requirements demand extensive testing before deploying learning-based controllers on the factory floor. In summary, while AI techniques promise significant gains in productivity and flexibility for industrial systems, ensuring smooth integration with existing processes and maintaining trustworthiness (through explainability and fail-safes) are key gaps to address.

Energy Systems

Traditional approaches in the energy domain (power generation, smart grids, etc.) have centered on well-understood physical models and deterministic optimization [163]. Examples include load flow calculations [164], auto-regressive (ARIMA) models for load forecasting [165], and rule-based grid operation policies designed by human experts for worst-case scenarios [166]. These classical methods are often transparent and supported by decades of operational experience, but they struggle to cope with the increasing complexity introduced by renewable energy sources and real-time management of distributed resources. AI and machine learning are increasingly adopted to enhance forecasting (e.g., using deep neural networks for solar/wind power prediction) and to optimize grid control in smart grids [167]. ML models can analyze the massive data from smart meters, sensors, and weather feeds to improve demand response and energy storage management. For example, deep learning has achieved lower error rates in short-term load and renewable generation forecasting than classical time-series models, helping grid operators handle volatility [168]. Data-driven optimization (such as reinforcement learning for energy distribution) can adapt to complex, dynamic conditions that were not envisioned by static rules [169]. The gains in efficiency and reliability (e.g., reducing energy waste by better matching supply and demand) align with sustainable energy goals. Nonetheless, several

challenges impede widespread ML deployment in energy systems [170]. A primary concern is *model validity and safety*: grid operators require guarantees that an AI agent will not violate physical constraints or jeopardize stability. Pure black-box models lacking physical insight might propose control actions that, while optimal in training data, could be unsafe in reality. This has led to interest in *physics-informed ML* and hybrid approaches in this domain (discussed later) to ensure consistency with power system laws [171]. Another challenge is data quality and availability: while the energy sector generates vast data, labeled data for abnormal events (e.g., blackouts, rare grid contingencies) are scarce, making it hard for supervised models to learn rare-event behavior. Moreover, interoperability with legacy grid infrastructure is an issue, similar to industrial settings [172]. Finally, regulators and stakeholders in energy demand interpretable solutions for critical decisions, so the opaqueness of complex ML models remains a barrier. In summary, AI/ML holds great promise for modernizing energy systems (improving forecasting, efficiency, and automation), but must overcome trust, data, and integration hurdles for mission-critical adoption.

Cyber-Physical Systems and IoT

In the broad domain of CPS and IoT (spanning smart cities, autonomous vehicles, distributed sensor networks, etc.), classical solutions have typically involved simplified models or preset logic running on embedded devices [173]. For example, many IoT sensor networks use fixed threshold triggers or basic signal processing for anomaly detection, and control systems in CPS (like automotive or robotic systems) use model-based controllers or observers (e.g., Kalman filters) tuned to known dynamics [174]. These approaches benefit from low computational cost and predictable behavior but often lack adaptability—they cannot easily handle novel conditions or complex sensor fusion beyond their design assumptions [175]. The advent of data-driven ML has rapidly spread in CPS/IoT applications for enhanced perception and decision-making. Machine learning models (from shallow classifiers to deep neural networks) are now used for tasks like human activity recognition from wearable sensors, predictive maintenance in IoT-enabled infrastructure, and anomaly detection in networked control systems. These models can automatically learn patterns from high-dimensional sensor data, replacing or augmenting manually engineered features. For instance, in smart buildings or factories, ML-based analytics can detect subtle anomalies or optimize energy usage more effectively than any static rule [176]. Similarly, autonomous vehicles rely on AI for vision, sensor fusion, and planning, far exceeding the capabilities of classical kinematic models alone [177]. The flexibility of ML comes with significant challenges in CPS/IoT contexts, however. A first issue is the diversity of operating conditions and sensor modalities in real-world CPS: initial generation ML models that work well in the lab often face *scalability and generalization limits* when deployed across many diverse IoT devices and environments [178]. Each new task or deployment may require re-training and careful tuning (a "task-specific" model), which does not scale well when there are thousands of different IoT tasks and settings. The scarcity of labeled data for every possible context is acute; many CPS sensors produce data that are unlabeled or not easily interpretable by humans (e.g., raw vibration signals), making supervised learning difficult. Additionally, data may be sensitive (e.g., personal data from wearables or cameras), raising privacy concerns if aggregated centrally. Real-time constraints are also critical in CPS: models must often run on edge devices with limited computation and respond within stringent time bounds for safety [179]. Classical algorithms were usually designed with these real-time and compute limits in mind, whereas large neural networks can be resource-hungry. Device heterogeneity and network unreliability further complicate centralized AI solutions [180]. In summary, while AI techniques can greatly enhance CPS/IoT systems by leveraging rich sensor data (enabling, for example, more autonomous

and adaptive operation), ensuring reliability, timeliness, and privacy in distributed deployments presents a challenge. Techniques like model compression, edge computing, and federated learning (discussed later) are emerging to tackle these issues. Moreover, the need for continuous learning and adaptation in CPS has spurred research into online learning and domain adaptation so that models remain robust as environments evolve. To summarize the domain-specific analysis, Table 5 provides a high-level comparison of traditional versus AI-based techniques in each domain, highlighting their typical use-cases and limitations.

Table 5. Comparison of classical and AI/ML fault detection across domains.

Domain	Classical Approaches	AI/ML Approaches	Key Benefits/Challenges
Industrial	Physics-based modeling and rule-based control (PID, SPC); scheduled maintenance; deterministic thresholds.	Predictive maintenance via ML and digital twins; reinforcement learning for adaptive control; computer vision for inspection.	Benefits: Early fault prediction, reduced downtime. Challenges: Model interpretability and data quality in heterogeneous environments.
Energy	Deterministic simulation, ARIMA forecasting, static safety margins.	Deep learning for load and renewable forecasting; reinforcement learning for grid optimization; hybrid physics–ML energy models.	Benefits: Improved forecasting accuracy, dynamic control. Challenges: Data imbalance, explainability for safety-critical systems.
CPS/IoT	Threshold-based alarms; simple empirical or rule-based diagnostics on embedded devices.	Edge ML for distributed anomaly detection; CNN/RNN for sensor fusion; federated and TinyML approaches for low-power analytics.	Benefits: Real-time local fault detection, scalability, privacy preservation. Challenges: Limited compute resources, model synchronization, data heterogeneity.
Cybersecurity	Signature- and rule-based Intrusion Detection Systems (IDS); expert-crafted firewall policies; static heuristics.	Deep learning for anomaly detection (CNN/LSTM); Transformer and GNN architectures for topology-aware intrusion detection; GANs for synthetic data; LLMs for log analysis; XAI for interpretable decision support.	Benefits: Detection of zero-day and polymorphic attacks; semantic log reasoning; automation of threat analysis. Challenges: High false positives, adversarial evasion, computational overhead, explainability gaps.

Cyber-Security

In cybersecurity and network defense, classical approaches have centered on static signatures and predefined rules to identify malicious activities—examples include rule-based intrusion detection systems that match known attack patterns or firewall policies crafted by experts [181]. These methods are transparent and reliable for familiar attack types, but they are fundamentally reactive and struggle against novel or evolving threats: a clever attacker can slightly modify an exploit to evade a signature, rendering traditional IDS blind to the new variant [182]. AI-driven techniques are reshaping this landscape by introducing adaptability and data-driven insight into threat detection. Machine learning models can generalize from vast historical data to detect anomalies or attack indicators that were not explicitly programmed into the system. For instance, an ML-based anomaly detector might learn typical network traffic profiles and trigger an alert when it encounters subtle deviations indicative of a potential breach, catching incidents that static rules would miss. The benefit is a dramatically improved coverage of unknown or polymorphic

attacks and the ability to analyze high-volume data streams in real time, far beyond human capacity [183]. Building upon these foundations, recent advances have diversified the AI toolbox for cybersecurity. Deep neural networks such as CNNs, LSTMs, and GRUs enable high-fidelity analysis of temporal network flows, while Transformer-based architectures (e.g., BERT4IDS, CyberViT) leverage attention mechanisms to capture long-range dependencies and subtle contextual cues in packet sequences [184]. Graph Neural Networks (GNNs) extend this capability by modeling the relational structure of hosts, services, or IoT devices, making them particularly effective for detecting coordinated or multi-hop attacks [185]. Generative AI models, including GANs and diffusion architectures, augment training datasets by synthesizing realistic intrusion samples, mitigating class imbalance and improving robustness against rare threats [186]. Meanwhile, Large Language Models (LLMs) such as SecGPT and CyberBERT introduce semantic reasoning into security analytics, parsing unstructured logs or incident reports to extract high-level attack narratives and contextual anomalies [187]. Lightweight TinyML and quantized CNN implementations further enable deployment of such models directly on edge devices and embedded systems, providing near-real-time intrusion monitoring in resource-constrained environments [188]. However, these advantages come with trade-offs. Learning-based systems can generate increased false positives if not carefully tuned, as normal behavior in networks can be highly variable; each spurious alert incurs a cost by diverting analyst attention [189]. Moreover, AI models often operate as "black boxes," making it difficult for security teams to interpret why a given alert was raised—unlike a signature (which directly points to a known threat), an anomaly alert might require additional analysis to verify. To address this limitation, Explainable AI (XAI) methods such as SHAP, LIME, and attention visualization are increasingly integrated to provide interpretability and feature attribution, helping analysts understand model reasoning and improving trust in automated alerts [190,191]. There is also the issue of adversarial resistance: attackers may attempt to probe and manipulate ML models (for example, by injecting crafted traffic) to avoid detection, a challenge less applicable to deterministic, rule-based systems. Consequently, in practice, organizations are cautious about fully replacing classical cybersecurity tools with AI. Instead, AI-based detectors are usually deployed alongside traditional methods, serving as an early-warning layer for suspicious behaviors while established signature-based systems handle the well-characterized threats [192]. In hybrid architectures, AI-driven Fault Detection and Diagnosis (FDD) frameworks act as adaptive, predictive layers complementing deterministic rule engines, forming multi-tier defense systems that balance interpretability, robustness, and computational efficiency [193]. The cybersecurity domain exemplifies a balance—AI approaches significantly extend detection capabilities beyond the static scope of classical methods, but ensuring their reliability, explainability, and adversarial robustness is essential before they can attain the same level of trust and operational maturity as conventional solutions.

Figure 6 outlines predominant FDD techniques and motivating factors across industrial, energy/automotive, CPS/smart grid, and cybersecurity domains, highlighting the influence of application-specific requirements and constraints.

Figure 7 provides a high-level view of FDD application domains, illustrating typical fault types and objectives and the transition from model- and rule-based approaches to data-driven and AI-based methods across industrial, energy/automotive, CPS/IoT, and cybersecurity contexts.

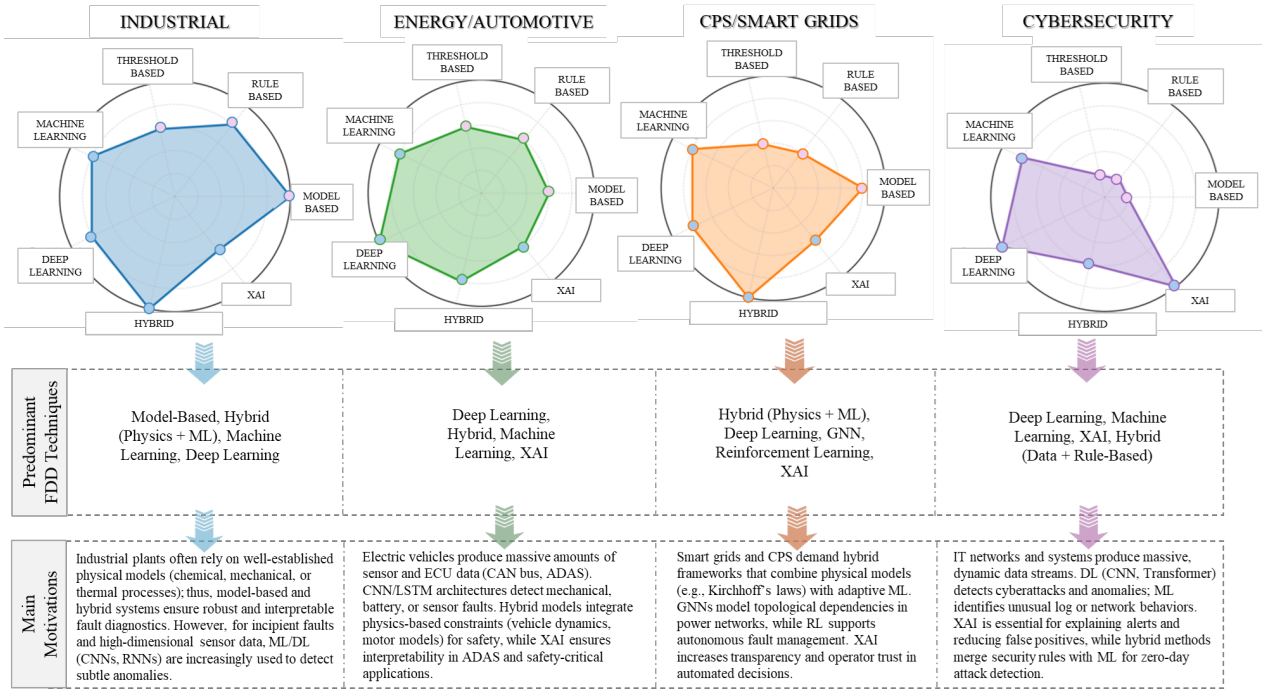


Figure 6. Overview of predominant FDD techniques and main motivations across industrial, automotive, CPS/smart grid, and cybersecurity domains. Radar charts illustrate the relative use of different approaches, while the table summarizes key drivers in each sector.

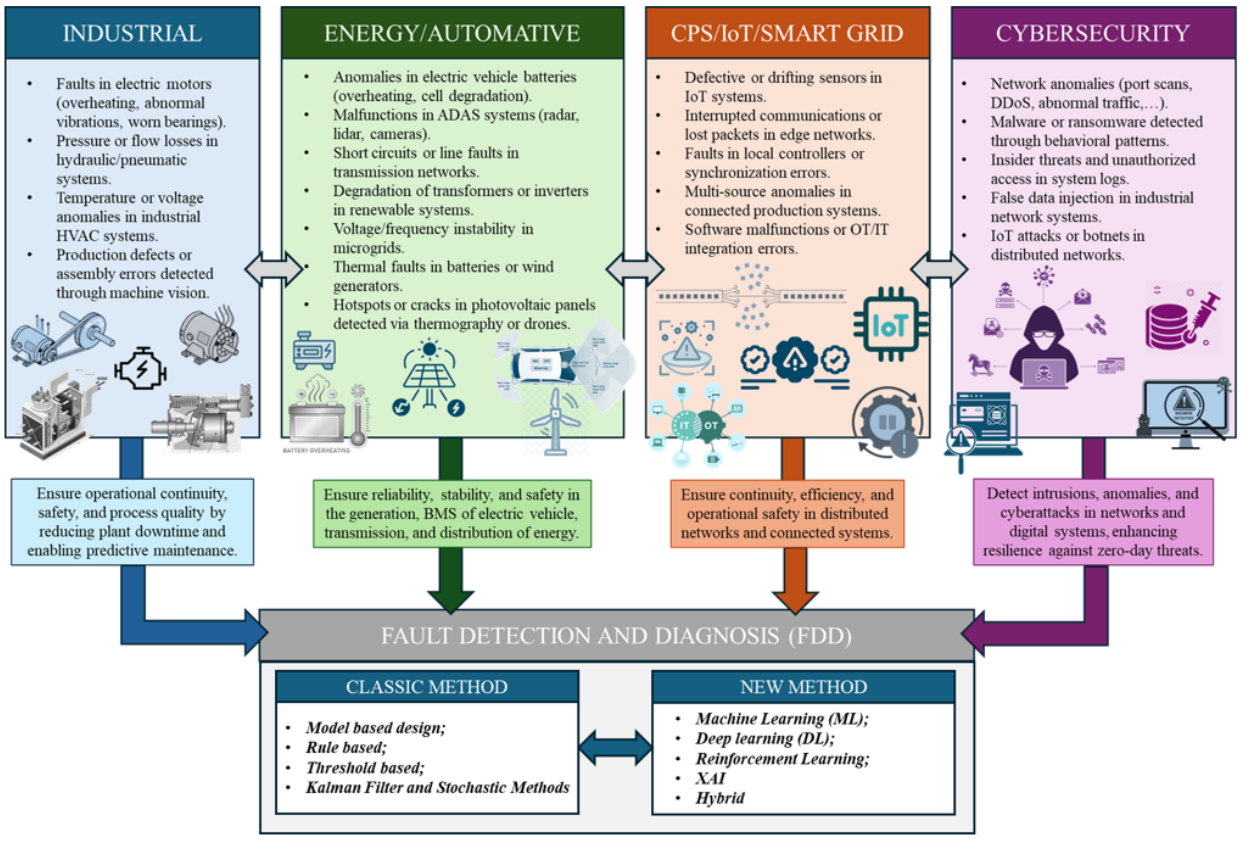


Figure 7. Overview of main application domains of Fault Detection and Diagnosis (FDD) techniques, highlighting typical faults, objectives, and the transition from classic model- and rule-based methods to modern data-driven and AI-based approaches.

6.2. Thematic Challenges and Gaps

Beyond the specificities of each domain, several *cross-cutting* challenges emerge when comparing classical techniques to AI/ML approaches. We discuss these thematically below, as they represent common gaps or requirements that need to be addressed for AI-based methods to fully realize their potential in engineering systems.

Interpretability and Transparency

One of the most cited challenges of data-driven AI methods is their often opaque decision-making process. Classical methods, rooted in physical laws or simple logic, are usually transparent by construction—an engineer can examine a formula or threshold and understand its effect. This interpretability is crucial in domains where safety and accountability matter: for instance, a control law derived from first principles can be vetted and understood failure modes, whereas a complex neural network controlling the same process may defy intuitive explanation [194]. The "black-box" nature of many ML models erodes trust and poses hurdles for certification in regulated industries [195]. Lack of interpretability also makes debugging and refining models difficult. Therefore, the gap in explainability is a major barrier to replacing classical approaches with AI in practice. Efforts to mitigate this include developing *explainable AI (XAI)* techniques and inherently interpretable models [196]. For instance, research has demonstrated the use of SHAP values or LIME in explaining feature importance for ML models in structural health monitoring, and "white-box" model designs (like decision trees or rule-based learners) are sometimes favored over slightly more accurate but inscrutable deep networks for critical tasks. Nonetheless, achieving a level of clarity comparable to physics-based models remains an open challenge. In many cases, a hybrid strategy is suggested, where the core of the decision logic remains interpretable, and ML is used in a limited, observable way (or with monitors) to ensure transparency [197]. Table 6 (first row) summarizes the stark contrast in interpretability: classical systems are typically explainable (and thus easier to certify and trust), whereas many AI systems require additional work to make their reasoning understandable.

Table 6. Comparison of classical vs. AI/ML approaches by key aspects across domains.

Aspect	Classical Techniques	AI/ML Techniques
Interpretability	High; physics- or rule-based, deterministic and explainable to domain experts.	Often low; deep and ensemble models behave as black boxes, requiring Explainable AI (XAI) tools such as SHAP, LIME, or attention visualization to ensure analyst trust.
Adaptability and Generalization	Limited to predefined conditions or known faults; requires manual re-tuning for new scenarios.	Learns from data and adapts to evolving patterns; capable of detecting zero-day faults or attacks through unsupervised, self-supervised, or transfer learning.
Data Requirements	Minimal; relies on expert knowledge, system models, and calibration data.	High; requires large-scale and diverse datasets—possibly augmented via synthetic data generation (GANs, diffusion models) or federated data sharing for privacy.

Table 6. *Cont.*

Aspect	Classical Techniques	AI/ML Techniques
Robustness and Security	Predictable within design envelope; resistant to random noise and interpretable under formal verification.	Vulnerable to data drift, poisoning, and adversarial inputs; mitigated via adversarial training, ensemble models, and uncertainty-aware learning.
Computational Efficiency	Lightweight; suitable for embedded or real-time systems with constrained resources.	Can be resource-intensive; modern approaches (TinyML, quantization, pruning) enable efficient deployment on edge and IoT devices.
Scalability and Deployment	Centralized processing or rule-based systems; limited scalability in distributed contexts.	Highly scalable through edge/fog architectures, federated learning, and cloud orchestration; enables distributed and collaborative detection.
Explainability and Trustworthiness	Transparent reasoning directly linked to known physical laws or rules.	Requires hybrid reasoning (symbolic + neural) and explainability layers to justify alerts, especially in cybersecurity and safety-critical applications.
Human-in-the-Loop Integration	Strong dependence on expert judgment and manual rule updates.	Supports analyst augmentation via AI recommendations, semantic reasoning (LLMs), and automated incident prioritization.

Data Availability and Quality

Modern ML is hungry for data: large quantities of labeled examples are often needed to train high-performing models. Classical engineering approaches, by contrast, do not rely on data in the same way; they leverage knowledge of the system (e.g., equations, design tolerances) and may only need a small amount of calibration data. In many industrial and CPS settings, obtaining massive labeled datasets is impractical, especially for failure or anomaly cases which are rare by nature. This creates a gap: AI methods promise finer insight but are constrained by *data scarcity* [198]. For example, developing a predictive maintenance model for turbine failures requires many examples of turbines approaching failure, which could take years of run-to-failure data that companies simply do not have for new equipment [199]. Similarly, an ML-based grid fault detector would ideally train on many examples of grid faults under varied conditions, but real grid blackouts are (hopefully) too infrequent to supply this [200]. Data quality is another issue: sensor data can be noisy, incomplete, or subject to biases (e.g., different operating regimes underrepresented in the training set) [201]. Classical methods, by encoding physical truths, can often work with minimal data and still be reliable within design regimes. When data is limited, AI models also risk overfitting and poor generalization. Recent approaches to mitigate data issues include transfer learning (pretraining models on related tasks and fine-tuning), semi-supervised and unsupervised learning (to exploit plentiful unlabeled data), and synthetic data generation (e.g., using simulations or generative models to create additional training examples) [202]. In fact, the abundance of unlabeled operational data in domains like IoT has driven interest in self-supervised learning and foundation models (discussed later) to reduce dependence on annotated data [203]. Another strategy is to incorporate domain knowledge to reduce data needs, which leads to physics-informed methods [204]. Despite these mitigations, ensuring that AI models have sufficient high-quality data covering all

scenarios remains a significant challenge, especially in safety-critical applications where one cannot simply "fail more often" to gather data. Table 6 illustrates this difference: classical methods require relatively little data (but much expert input), whereas ML methods thrive on big data and struggle when data is small or skewed.

Robustness and Generalization

Engineering systems must operate reliably under a wide range of conditions, including those not seen during design or testing. Classical control and diagnostic techniques are usually designed with worst-case assumptions and safety margins, and their behavior outside nominal conditions can often be analyzed (e.g., via robust control theory) [205]. In contrast, learned models excel in the conditions similar to their training data but can behave unpredictably when faced with novel inputs or disturbances (the *distribution shift* problem) [206]. For example, an ML-based fault detector might perform impressively on known fault types but fail to recognize a new failure mode it was never trained on, whereas a physics-based limit check might still catch an anomaly if it violates a fundamental threshold (like temperature beyond a limit) [207]. Ensuring robustness of AI models is a critical gap: they can be sensitive to noise, adversarial inputs (especially in cyber domains), or changing environmental factors [208]. In IoT and CPS, even a change in sensor calibration or network latency could degrade an ML model's performance if not accounted for [209]. Moreover, many AI models do not inherently know how to say "I'm not sure" when encountering an out-of-distribution input; they might extrapolate in a dangerous way [210]. Techniques to improve robustness include adversarial training (to handle malicious perturbations), uncertainty estimation and anomaly detection (so the model can flag when it is unsure or when data looks unlike anything seen before), and continual learning to update models as new data comes in [211]. Combining models with physics constraints can also improve reliability by preventing physically impossible outputs. Another aspect of robustness is *fault tolerance*: classical systems often have built-in redundancy (e.g., triple-modular redundancy in avionics) to handle component failures [212]. ML models deployed in such architectures need similar redundancy or fail-safe modes (for instance, defaulting to a safe baseline policy if the ML outputs seem aberrant). Achieving the high confidence levels required in domains like cybersecurity or power grid operation is an ongoing challenge for AI. Researchers are developing verification methods for neural networks and other ML models to provide guarantees on their behavior, but this is still an evolving field [213]. In summary, whereas classical methods offer predictable (if sometimes conservative) performance guaranteed by design, ML methods must overcome issues of generalization and brittleness. Robustness must be built and proven, not assumed. Table 6 compares these aspects, noting that classical techniques tend to be robust within their design envelope and easier to certify, whereas ML techniques offer adaptability but can fail in unforeseen ways if not carefully engineered.

Integration and Lifecycle Considerations

A practical challenge when introducing AI/ML into established engineering domains is how to integrate these techniques into existing workflows, infrastructure, and human processes [214]. Classical solutions are often deeply ingrained in industry standards and operator training. Replacing or augmenting them with AI means addressing compatibility with legacy systems, as well as the need for new tools and expertise (e.g., data engineering pipelines, model maintenance over time) [215]. Organizations may face a skills gap, where engineers need training in ML to validate and interpret model outputs. Additionally, AI models may need frequent updates as data distributions change (for instance, a predictive model may drift as machinery ages or as usage patterns shift), whereas classical methods

might only need recalibration occasionally [216]. This introduces a lifecycle maintenance issue: models must be monitored, revalidated, and possibly retrained, which is a new kind of ongoing effort in engineering management. There are also *ethical and compliance* considerations as identified in some domains: AI systems can inadvertently introduce bias or violate privacy if not carefully controlled (for example, a smart grid ML algorithm might inadvertently favor certain users or draw on personal data) [217]. Ensuring that AI deployments meet regulatory requirements (data protection laws, nondiscrimination, etc.) is part of the integration challenge [218]. While these considerations go somewhat beyond purely technical comparisons, they are important gaps to acknowledge: the successful application of AI requires more than just performance metrics in isolation; it needs a holistic approach to integrate into the socio-technical system [219]. This often means hybrid strategies in early adoption: using ML as a decision support or advisory tool rather than fully autonomous controller, until confidence and experience grow [220]. Over time, as standards and best practices develop (for example, standards for ML model validation in automotive or aviation are now emerging), these integration hurdles will diminish, but currently they represent a non-trivial gap between the promise of AI techniques and real-world deployment. To recap these themes, Table 6 summarizes how classical and AI-based approaches compare on key factors of interpretability, data reliance, and robustness.

6.3. Emerging Trends and Future Directions

Given the above challenges and gaps, current research is actively exploring several emerging directions to combine the strengths of classical and AI approaches and to address the limitations identified. We highlight a few prominent trends: foundation models, physics-informed learning, and federated learning (as well as related concepts), which are expected to play a significant role in the next generation of intelligent engineering systems. These methods and concepts are summarized in the Figure 8.

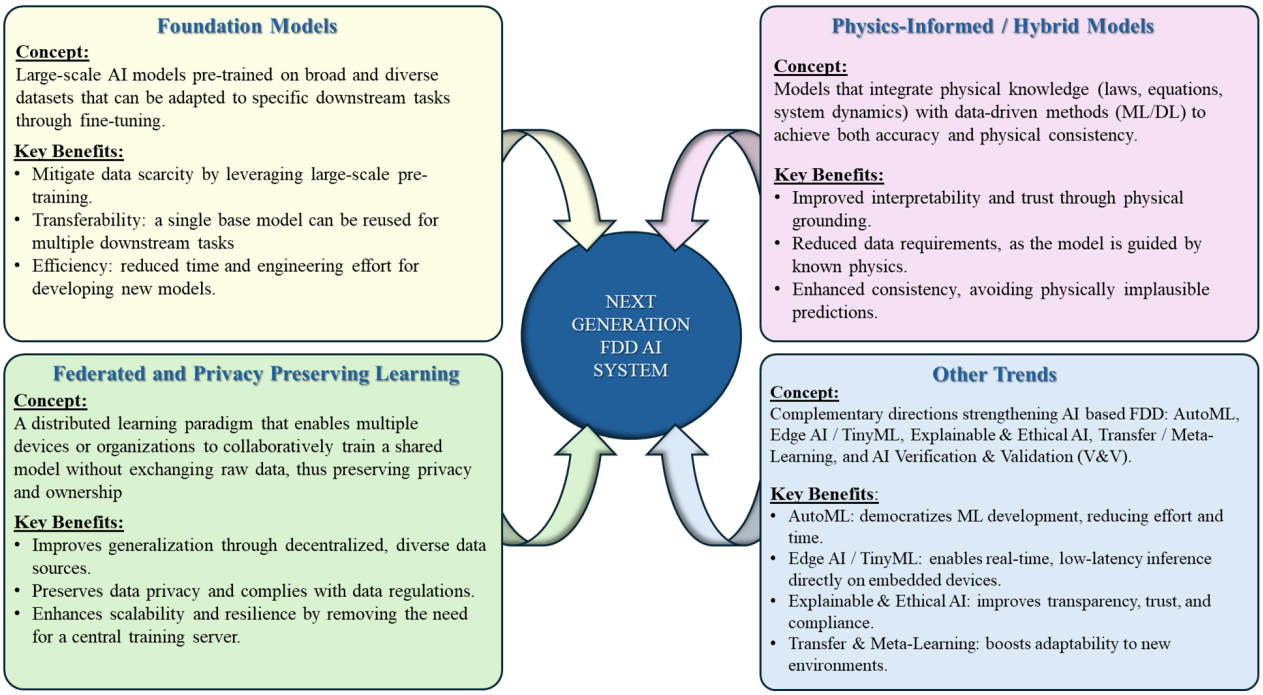


Figure 8. Conceptual overview of emerging research directions for next-generation Fault Detection and Diagnosis (FDD). Foundation models address data scarcity and transferability, physics-informed and hybrid models enhance interpretability and reliability, and federated learning enables collaborative and privacy-preserving AI. Other complementary trends support automation, efficiency, and trust in intelligent engineering systems.

Foundation Models and Large Pre-Trained AI

One trend in the broader AI field is the rise of *foundation models*—very large models (often neural networks) trained on extremely broad data at scale, which can be adapted to a wide range of downstream tasks [221]. Notable examples include large language models (like GPT series) and large vision models. The concept is that a single model trained with self-supervised learning on diverse data can serve as a general-purpose base which, with minimal fine-tuning, can achieve high performance on specific tasks [222]. This approach contrasts with training a separate model from scratch for each task (the "first generation" supervised learning paradigm). Foundation models have had huge success in NLP and computer vision, and researchers are now looking at applying them to CPS/IoT and industrial domains [223]. For example, in the CPS context, a foundation model might learn a general representation of time-series sensor data from a fleet of machines or a city of IoT sensors, and then be specialized for anomaly detection, forecasting, or control tasks as needed [224]. The appeal is that such models could mitigate data scarcity issues by leveraging unlabeled data (through self-supervised pre-training) and provide more universal feature representations that work across different operating conditions. Early efforts include large sensor data models and multimodal systems that integrate vision, audio, and other signals relevant to CPS. There is enthusiasm that foundation models could reduce the need for extensive task-specific data and engineering, essentially providing an AI-native way to incorporate a form of "general knowledge." [225] However, significant challenges remain in bringing this to fruition: foundation models are resource-intensive and their behaviors can be hard to predict or control, especially in safety-critical applications [226]. There is a gap between what these models can do in general domains and the strict requirements of engineering systems. Ongoing research is identifying desiderata for foundation models in these domains (e.g., reliability, physical consistency, safety constraints) [227]. We anticipate that in the near future, specialized foundation models for domains like manufacturing or smart grids will emerge, possibly offered as baseline tools that industry practitioners can leverage (much like pre-trained image or language models today) [228]. If successful, this trend would help address data scarcity and cold-start problems (since the heavy lifting of learning representations is done on broad data once, rather than from scratch for each small project) and could accelerate the deployment of AI solutions in domains that currently lack large labeled datasets [229]. Moreover, foundation models, with their ability to unify tasks, might ease integration: a single model could handle multiple related tasks (detection, diagnosis, prediction) in a system, simplifying the toolchain.

Recent industrial foundation models, such as multimodal vision–language architectures for intelligent manufacturing, demonstrate how large-scale pre-trained models can be adapted to fault detection under data scarcity and class imbalance, bridging industrial and aerospace-grade diagnostic requirements.

Physics-Informed and Hybrid Models

Another vital emerging trend is the blending of data-driven methods with physical models and domain knowledge, often referred to as *physics-informed machine learning* or hybrid modeling. The core idea is to inject physical laws, constraints, or model-based components into the learning process of AI models, aiming to get the best of both worlds: the accuracy and flexibility of ML together with the consistency and interpretability of physics-based methods. One popular approach is the use of physics-informed neural networks (PINNs) [230], where the loss function of a neural network is augmented with terms that enforce differential equation residuals, so that the network learns to satisfy known physics (e.g., conservation laws) while fitting the data [231]. This can dramatically reduce the amount of data needed and ensures the model's outputs obey critical properties

(like energy conservation or boundary conditions). In control and system identification, hybrid models might use a partial physics model (e.g., a few dominant equations) and let an ML model learn the unmodeled dynamics. In fault diagnosis, one might use a bank of observers (classical method) to generate residuals and then use an ML classifier on those residuals to isolate faults: combining analytical redundancy with data-driven pattern recognition [232]. We have already seen examples in the literature of hybrid FEM (finite element method) plus ML approaches for structural monitoring: a finite-element model provides a baseline simulation of structural behavior, and a learning algorithm adjusts or interprets results to match sensor data, yielding an approach that can work with fewer sensors and less data while remaining interpretable in terms of physical mechanics [233]. A recent literature study [234] highlights how a hybrid FEM–ML system achieved comparable damage localization accuracy to pure deep learning but with far less experimental data and with outputs that correlate with physical structural parameters, thus easing interpretation. This demonstrates a general trend: by leveraging physical insight, we can often reduce the data burden and increase trust [235]. Digital Twin technology is a related concept gaining traction across industries: a digital twin is essentially a live digital simulation of a physical asset, which can be continuously updated with data. AI can be integrated into digital twins to adjust the simulation based on data (calibrating the model in real-time) and to predict future states [236]. The twin provides a physically grounded context for the AI, making results more interpretable to engineers [237]. Physics-informed ML also naturally addresses some safety concerns, as known unsafe regions or invariants can be hard-coded into models. Overall, the community expects that purely black-box models will be less common in critical domains; instead, we'll see "grey-box" models with physics-based scaffolding and ML filling in the gaps [238]. This approach directly tackles the interpretability and robustness challenges: by anchoring ML with equations or known constraints, the model is less free to wander into nonsensical or unsafe predictions, and stakeholders can understand and trust the model more easily when they recognize physical laws at work.

Federated Learning and Privacy-Preserving AI

Data privacy and distributed data ownership are significant issues, especially in domains like IoT, smart healthcare, or any multi-stakeholder industrial consortium. Often the most valuable data is siloed (e.g., each factory, each device, or each organization holds its own data and may be unwilling or unable to share it due to privacy, competition, or regulatory reasons). This limits the data available for training centralized ML models. *Federated learning (FL)* has emerged as a promising solution: it allows training machine learning models collaboratively across multiple devices or organizations without exchanging the raw data [239]. In the federated paradigm, each device (or site) computes model updates on its local data and only those updates (e.g., gradient information) are sent to a central server, where they are aggregated to update a global model. The raw data never leaves its origin [240]. This approach has clear advantages for privacy and also can reduce communication costs in IoT settings (since only model parameters are transmitted, not all the raw data) [241]. In the context of CPS/IoT, FL is particularly attractive because sensor data is naturally decentralized and often sensitive (consider personal wearable sensors or security cameras) [242]. By applying FL, a collective model (say, for anomaly detection or predictive maintenance) can be trained using data from many sources (machines, vehicles, etc.) to improve its generalization, while each data silo keeps ownership of its information [243]. This can effectively enlarge the training dataset without centralizing data, addressing the data scarcity problem to some extent. There are challenges, of course: federated learning must deal with heterogeneous data (each client's data distribution might be different), unreliable or low-bandwidth connections, and potential privacy leaks

even from transmitted gradients (research in federated learning introduces techniques like secure aggregation and differential privacy to counter that) [244]. Despite these, we are already seeing federated learning pilots in domains like smart grids (utilities collaboratively training demand forecasting models), healthcare (hospitals training diagnostic models without sharing patient records), and industrial IoT (multiple factories jointly training quality control models). In CPS security, FL can enable cooperative anomaly detection between organizations to detect cyber-attacks on critical infrastructure without divulging sensitive logs [245]. The trend aligns well with the need for collaboration and data sharing in AI: it provides a mechanism to achieve it in a privacy-preserving manner. For future systems, we envision an ecosystem where edge devices are not just data sources but active participants in learning, continuously improving models in the field via federated or online learning [246]. This could dramatically speed up deployment and adaptation of AI models, as new data from each device makes the overall model smarter, benefiting all, without a central authority ever seeing raw data. Federated learning also encourages a modular view of AI deployment: models can be updated in a decentralized fashion, improving scalability and robustness (the system has no single point of failure for learning) [247]. Ultimately, combining federated learning with the other above trends, e.g., federated training of a foundation model across many organizations, could overcome many current limits [248]. It would allow leveraging vast amounts of distributed data to train powerful models while respecting privacy and diversity.

Other Noteworthy Trends

In addition to the above, there are several other emerging directions that are worth mentioning briefly. *AutoML and Automated Deployment* is one: to reduce the skill barrier and integration effort, automated machine learning pipelines (including hyperparameter tuning, neural architecture search, and continuous monitoring) are being developed so that AI models can be more easily produced and maintained in engineering contexts [249]. *Edge AI and TinyML* are important for IoT, focusing on making models lightweight and energy-efficient so that they can run on embedded devices in real time, which addresses the latency and resource constraints of deploying AI at the edge rather than in the cloud [250]. *Explainable and Ethical AI* considerations lead to the development of standards and tools to ensure AI decisions can be audited and are fair [251], for example, methods to detect and mitigate bias in models that control resource allocation in energy or recommend maintenance in industry [252]. *Digital Twins and Simulation-Augmented AI*, as mentioned, provide fertile ground for integrating ML with virtual testing, allowing countless what-if scenarios to be played out safely to improve the models [253]. We also see growing interest in *transfer learning and domain adaptation* techniques, acknowledging that an AI model will often be deployed in an environment different from where it was trained, so it must adapt with minimal additional data [254]. Research in meta-learning (models that learn how to learn) and few-shot learning is relevant here. Finally, the community is actively developing *verification and validation techniques for ML*: Taking advantage of formal methods and software testing, these aim to systematically prove or test that an AI component meets certain safety criteria, an essential step for assurance domains [255]. In conclusion, the landscape of AI/ML in engineering systems is evolving rapidly. Classical techniques and AI approaches each have distinct advantages, and a clear message from recent research is that the future lies in combining these strengths: using AI where it provides value (handling complexity, adapting to data) while using physical insight, human knowledge, and stringent engineering practices to guide and constrain the AI [256]. The emerging trends discussed (from large pre-trained models to physics-informed hybrids and collaborative learning) are all mechanisms to achieve this synergy. They address the core gaps of interpretability,

data, and robustness in different ways: foundation models tackle data scarcity by broad pre-training, physics-informed methods inject interpretability and reliability, and federated learning enables robust models built from wider data without breaching privacy. As these trends mature, we expect to see AI become an integral, yet safe and transparent, part of industrial, energy, CPS/IoT, and cybersecurity. This will lead to smarter and more resilient systems that can meet the growing demands of efficiency, sustainability, and autonomy in the coming decades.

6.4. Summary of Research Gaps and Open Challenges

Despite the significant progress of AI- and ML-based Fault Detection and Diagnosis (FDD) across industrial, energy, CPS/IoT, and cybersecurity domains, the comparative analysis presented in this survey highlights several persistent research gaps and open challenges that currently limit large-scale and safety-critical deployment.

- Data scarcity, imbalance, and representativeness
  A fundamental limitation remains the scarcity of labeled fault data, particularly for rare, incipient, or safety-critical failure modes. In real-world engineering systems, faults occur infrequently by design, resulting in highly imbalanced datasets dominated by normal operating conditions. This severely constrains supervised learning approaches and limits the generalization of trained models. Moreover, available datasets often fail to capture the full variability of operational regimes, aging effects, and environmental disturbances, leading to biased or overfitted models. Although data augmentation, synthetic data generation, and digital twin simulations partially mitigate this issue, ensuring that generated data faithfully represent real failure dynamics remains an open challenge.
- Generalization, robustness, and distribution shift
  Many AI-based FDD models exhibit strong performance under controlled experimental conditions but degrade when exposed to distribution shifts caused by sensor drift, changing operating conditions, component aging, or system reconfiguration. This challenge is particularly critical in CPS, energy systems, and cybersecurity, where operational contexts evolve continuously. Robustness against noise, missing data, and adversarial manipulation, such as false data injection attacks, remains insufficiently addressed in many studies. While ensemble learning, uncertainty estimation, and adversarial training have shown promise, systematic robustness guarantees comparable to those of classical model-based methods are still lacking.
- Explainability, trust, and certification barriers
  Deep learning architectures often outperform classical FDD methods in detection accuracy and adaptability; however, their black-box nature poses a major barrier to trust, certification, and regulatory acceptance in safety-critical engineering applications. Industrial operators, grid managers, and cybersecurity analysts require transparent diagnostic reasoning to validate automated decisions and to support corrective actions. Despite growing interest in Explainable AI (XAI), many current approaches provide post-hoc explanations that may not fully align with domain knowledge or physical causality. Bridging the gap between high-performance learning models and interpretable, auditable diagnostics remains a central research challenge.
- Real-time constraints and edge deployment limitations
  Numerous engineering applications demand fault detection and diagnosis within strict latency bounds, often on resource-constrained embedded or edge devices. However, state-of-the-art models—such as transformers, GNNs, and large ensembles, are computationally intensive and difficult to deploy under real-time and energy constraints. Although TinyML, model compression, quantization, and edge AI accel-

erators enable partial mitigation, there is still a trade-off between model complexity, detection accuracy, and computational efficiency. Designing architectures that natively balance performance with real-time operability remains an open problem.

- Lifecycle management, maintenance, and long-term reliability
  Unlike classical FDD systems, AI-based models require continuous monitoring, validation, and retraining to remain effective over time. Concept drift, evolving fault patterns, and system upgrades can rapidly invalidate previously trained models. However, systematic lifecycle management strategies—covering model versioning, online adaptation, validation after retraining, and safe rollback mechanisms—are rarely addressed in the literature. This gap hinders long-term deployment and raises concerns about reliability, accountability, and maintenance costs in industrial environments.
- Integration with legacy systems and human-in-the-loop operation
  The integration of AI-based FDD within existing industrial control systems, SCADA architectures, and cybersecurity workflows remains challenging. Many deployments still rely on legacy hardware and deterministic logic, requiring hybrid architectures where AI complements rather than replaces classical methods. Ensuring effective human-in-the-loop interaction, where AI systems provide actionable insights rather than opaque alerts, is essential but insufficiently standardized. Developing frameworks that seamlessly integrate AI diagnostics with human expertise and established operational procedures is an ongoing challenge.

These gaps motivate the emerging research directions discussed in the following section, where foundation models, physics-informed learning, and federated intelligence are explored as promising pathways toward scalable, trustworthy, and deployable FDD systems.

7. Results

This section consolidates the quantitative and qualitative findings of this survey, encompassing more than 200 peer-reviewed studies published between 2022 and 2025 on AI and ML based Fault Detection and Diagnosis (FDD). The analysis integrates comparative evidence across industrial, energy, cyber–physical/IoT, and cybersecurity domains. Figures 9–11 and the supplementary radar charts summarize the relative maturity of major AI paradigms (namely CNN/LSTM, Transformers, GNNs, GANs, LLMs, Autoencoders, Federated Learning, TinyML, XAI, and Hybrid models) across eight evaluation dimensions: interpretability, adaptability and generalization, data requirements, robustness and security, computational efficiency, scalability and deployment, explainability and trustworthiness, and human-in-the-loop integration.

Figure 12 presents a radar-based comparison of representative AI/ML approaches for cybersecurity FDD, highlighting trade-offs across interpretability, adaptability, data requirements, robustness, scalability, and computational efficiency.

7.1. Methodology for Radar-Chart Scoring and Quantitative Synthesis

To support cross-domain comparison while avoiding purely qualitative narrative synthesis, this survey adopts radar-chart visualizations (Figures 9–12) as a structured comparative tool rather than as an objective performance benchmark. Each radar chart summarizes the relative maturity and suitability of major AI/ML paradigms for Fault Detection and Diagnosis (FDD) across eight evaluation dimensions previously indicated. Each axis is discretized on a five-level ordinal scale (1–5), where higher values indicate stronger alignment with the corresponding criterion. Importantly, these scores do not represent absolute performance metrics (e.g., accuracy, latency, or F1-score), nor are they intended as direct rankings between algorithms. Instead, they reflect a normalized qualitative

synthesis derived from the literature reviewed in Sections 4–6. For each AI/ML paradigm and application domain, scores were assigned according to the following principles:

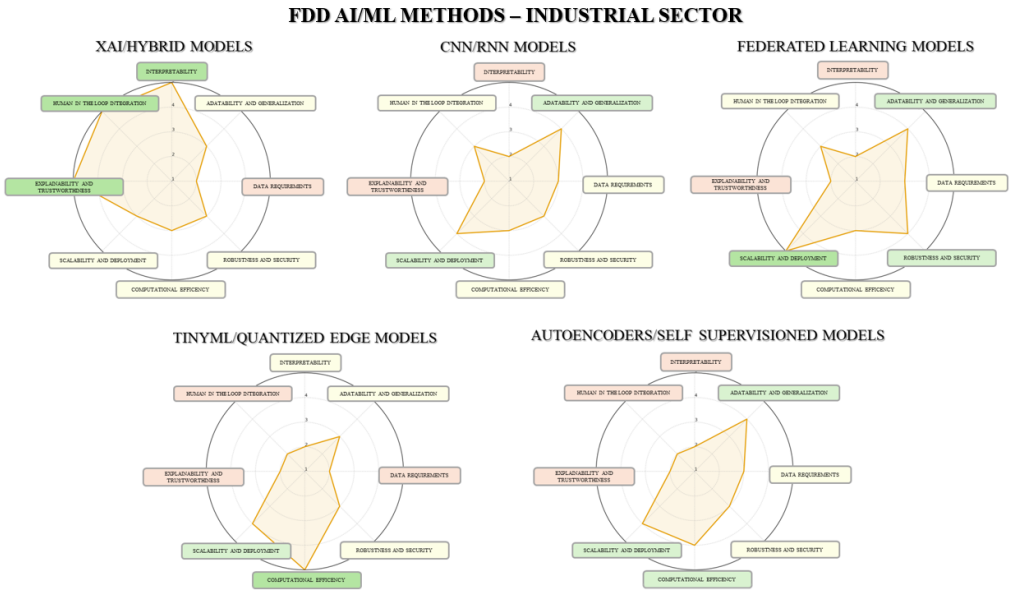


Figure 9. Radar chart comparison of representative AI/ML techniques for Fault Detection and Diagnosis (FDD) in the industrial domain, including XAI/hybrid approaches, CNN/RNN models, federated learning, TinyML/quantized edge models, and autoencoder/self-supervised methods. The charts illustrate trade-offs across multiple criteria such as interpretability, data efficiency, scalability, robustness, and computational performance, emphasizing domain-specific advantages and limitations. Radar charts represent a qualitative, literature-driven synthesis of relative strengths and trade-offs, not an objective performance ranking.

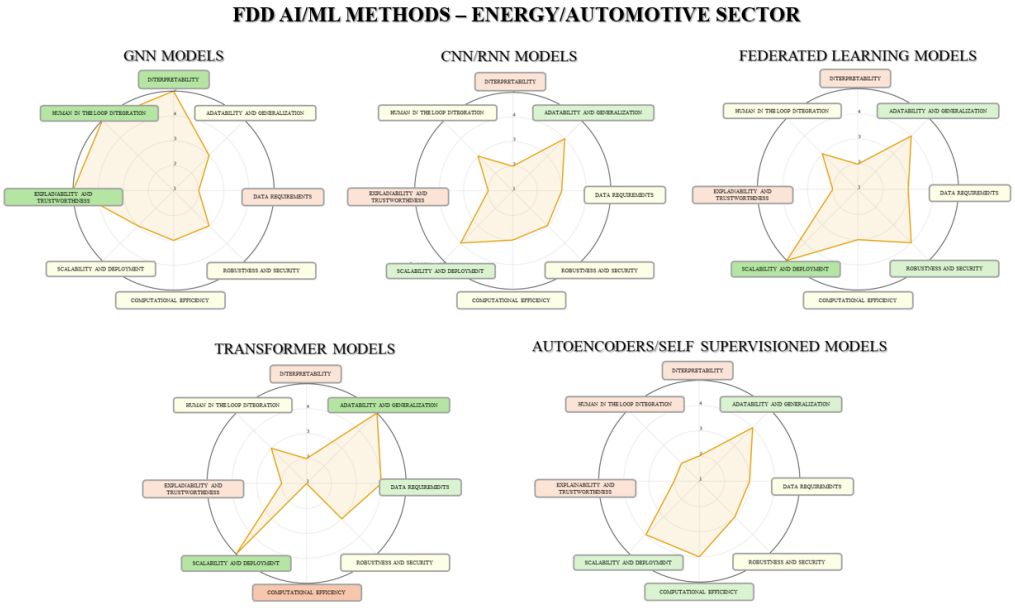


Figure 10. Radar chart comparison of representative AI/ML techniques for Fault Detection and Diagnosis (FDD) in the energy and automotive domains. The figure compares GNNs, CNN/RNNs, federated learning, transformers, and autoencoder/self-supervised models in terms of interpretability, adaptability, data requirements, robustness, scalability, and computational efficiency, highlighting the trade-offs and strengths of each approach. Radar charts represent a qualitative, literature-driven synthesis of relative strengths and trade-offs, not an objective performance ranking.

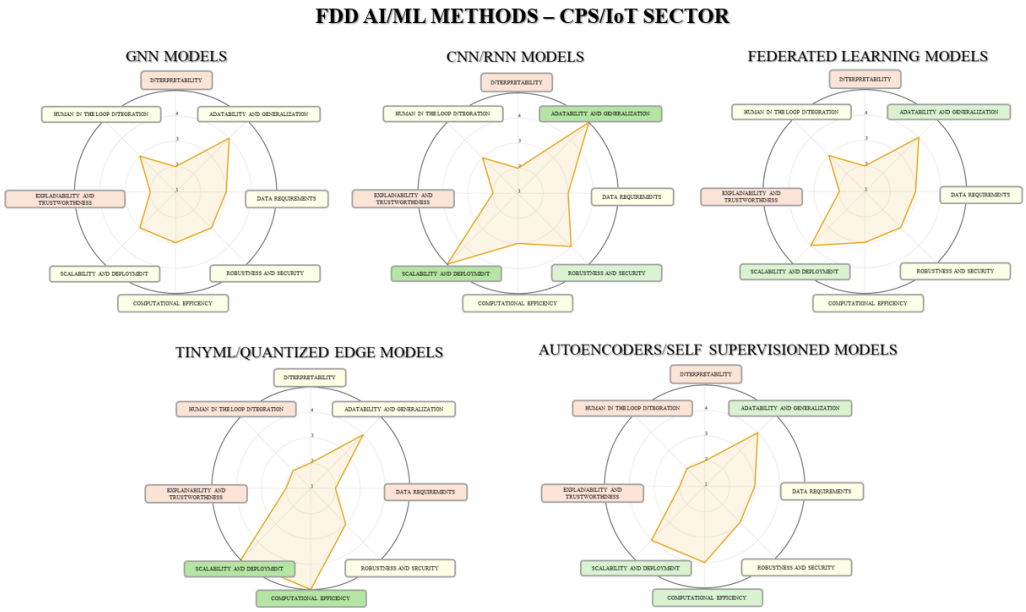


Figure 11. Radar chart comparison of representative FDD AI/ML methods for the CPS/IoT sector, including GNNs, CNN/RNN models, federated learning, TinyML/quantized edge models, and autoencoders/self-supervised approaches. Each chart highlights the trade-offs among key criteria: interpretability, adaptability, data requirements, robustness, scalability, and computational efficiency, providing a visual synthesis of their strengths and limitations. Radar charts represent a qualitative, literature-driven synthesis of relative strengths and trade-offs, not an objective performance ranking.

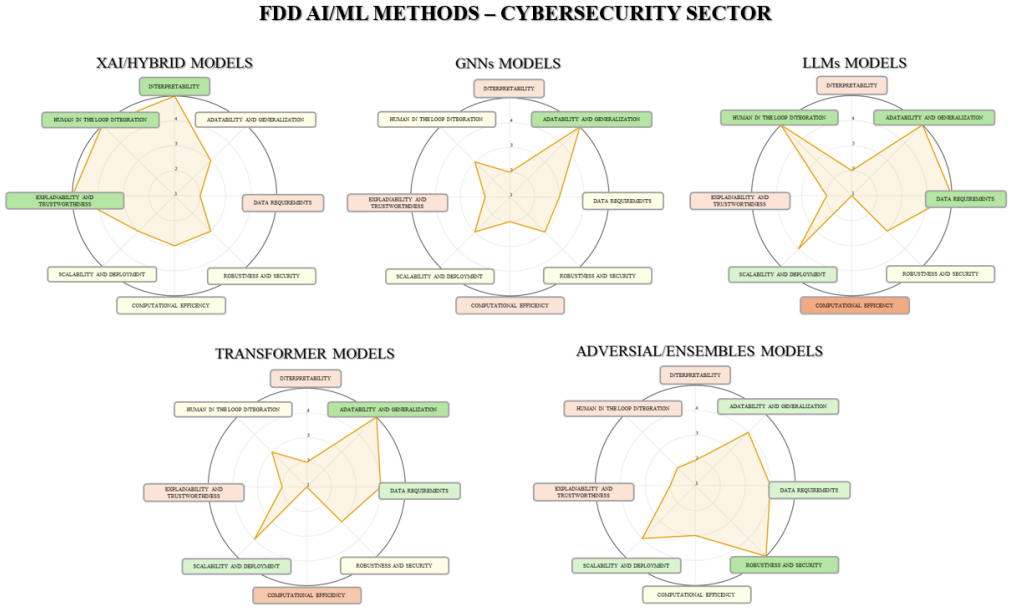


Figure 12. Radar chart comparison of representative AI/ML approaches for Fault Detection and Diagnosis (FDD) in the cybersecurity domain, including XAI/hybrid models, GNNs, large language models (LLMs), transformers, and adversarial/ensemble methods. The comparison highlights trade-offs across multiple dimensions such as interpretability, adaptability, data requirements, scalability, robustness, and computational efficiency, emphasizing their potential and challenges in cyber-physical defense applications.

- Evidence-driven assessment: Values are grounded in recurring experimental evidence, architectural properties, and deployment characteristics reported across multiple peer-reviewed studies (2022–2025), rather than isolated results.

- Operational definition of axes:
  - Interpretability and Explainability & Trustworthiness capture the intrinsic transparency of the model and the availability of XAI mechanisms (e.g., rule-based reasoning, SHAP, attention visualization);
  - Adaptability and Generalization reflects robustness across operating conditions, transferability, and performance under distribution shift;
  - Data Requirements indicates the typical dependency on labeled data volume and diversity;
  - Robustness and Security encompasses resilience to noise, uncertainty, adversarial manipulation, and fault/attack ambiguity;
  - Computational Efficiency reflects inference cost, suitability for real-time and edge deployment, and model compactness;
  - Scalability and Deployment evaluates feasibility in large-scale, distributed, or federated infrastructures;
  - Human-in-the-loop Integration reflects compatibility with supervisory control, operator validation, and decision support workflows.
- Domain-aware normalization: Scores are contextualized within each domain (industrial, energy/automotive, CPS/IoT, cybersecurity) to avoid misleading cross-domain absolute comparisons. For example, computational efficiency in industrial edge systems is evaluated differently than in cybersecurity analytics pipelines.
- No explicit weighting: All axes are intentionally treated with equal importance to avoid bias toward a specific application objective. The resulting radar shapes should therefore be interpreted as qualitative fingerprints of method families rather than weighted performance indices.

As a consequence, the radar charts are intended to support comparative reasoning and design-space exploration, highlighting structural trade-offs (e.g., interpretability vs. adaptability, efficiency vs. robustness), rather than to provide definitive rankings of AI/ML techniques. This approach aligns with the survey's objective of guiding method selection and system design in heterogeneous FDD contexts, rather than prescribing universally optimal solutions.

7.2. Cross-Domain Performance Overview

The cross-domain analysis reveals distinct specialization patterns among the examined techniques as we can see in Figures 6 and 9–11. Deep neural architectures such as CNNs, RNNs, and LSTMs remain the dominant choice for industrial and energy applications, where temporal pattern extraction and sensor fusion are paramount. Conversely, Transformer-based models and Graph Neural Networks (GNNs) are increasingly prevalent in CPS and cybersecurity settings, owing to their ability to capture long-range dependencies, relational dynamics, and multi-hop interactions within complex networks. Edge-oriented approaches, including Federated Learning (FL) and TinyML, have demonstrated strong potential for distributed fault detection in IoT infrastructures, enabling low-latency and privacy-preserving analytics close to the data source. At the same time, Large Language Models (LLMs) and Explainable AI (XAI) frameworks are revolutionizing cybersecurity fault diagnosis, introducing semantic reasoning, context-aware log analysis, and human-traceable decision explanations. A synthesis of performance scores derived from the radar analysis highlights shown:

- Industrial systems achieve the highest average robustness and computational efficiency, benefiting from mature edge deployment pipelines and stable data acquisition environments.

- Energy systems show strong scalability and generalization through hybrid physics–ML frameworks and reinforcement learning strategies, supporting adaptive grid control and predictive maintenance.
- CPS/IoT infrastructures excel in adaptability and real-time responsiveness via federated and TinyML frameworks, though they remain limited in interpretability due to model opacity at the edge.
- Cybersecurity leads in adaptability and growing explainability thanks to Transformer- and GNN-based intrusion detection; however, real-time deployment is still constrained by computational overhead and false-positive mitigation.

The results confirm that model families emphasize different trade-offs: deep architectures (CNN/LSTM) and Autoencoders are efficient and accurate but require substantial training data, and Transformers and GNNs maximize adaptability and relational modeling, while XAI and hybrid paradigms achieve the highest interpretability and human trust, albeit with moderate scalability. These results collectively indicate that FDD performance is inherently context-dependent rather than paradigm-dependent.

7.3. Integrated Discussion and Observed Trends

Beyond algorithmic advances, the practical deployment of AI/ML-based FDD systems faces several non-trivial challenges. These include data integration with legacy infrastructures, model validation and certification in safety-critical environments, lifecycle management under concept drift, and the need for human-in-the-loop supervision. Addressing these issues is essential for transitioning AI-based diagnostics from laboratory prototypes to reliable industrial solutions. The cross-domain synthesis shows that FDD research is converging toward three complementary directions: (i) deeper temporal–spatial modeling for complex sensor networks, (ii) hybrid interpretability frameworks combining data-driven and physical reasoning, and (iii) distributed, privacy-preserving intelligence for scalable operation.

Specifically:

- CNN/RNN remain optimal for high-frequency, sensor-level fault recognition in industrial and energy applications, offering reliable time-series diagnostics.
- Transformers and GNNs dominate in large-scale, interconnected infrastructures (e.g., smart grids, IoT, cyber defense) due to their superior representation of global dependencies.
- GANs and Diffusion models serve as auxiliary tools, augmenting scarce fault datasets and improving class balance.
- LLMs extend FDD to unstructured textual data (e.g., logs, maintenance reports) through semantic interpretation and reasoning.
- Federated and TinyML paradigms enable decentralized learning and edge autonomy, crucial for real-time diagnostics under limited bandwidth and privacy constraints.
- XAI and hybrid models deliver explainable and auditable decisions, supporting certification and human validation in safety-critical contexts.

Overall, AI-driven FDD frameworks are transitioning from monolithic, centralized architectures to adaptive, distributed ecosystems capable of self-learning and self-explanation. This shift underpins the emergence of scalable, interpretable, and energy efficient diagnostic intelligence that can be securely embedded within industrial, energy, CPS/IoT, and cybersecurity infrastructures.

7.4. Future Directions

Recent studies indicate that FDD research is evolving toward the convergence of large-scale, explainable, and physically consistent AI. Next-generation diagnostic systems will

integrate analytical redundancy from classical models with the adaptability of advanced learning architectures, forming hybrid, distributed, and self-evolving frameworks. Table 7 summarizes the most relevant research frontiers and their expected technological impact.

Table 7. Emerging research trends shaping next-generation FDD systems.

Trend/Paradigm	Core Concept	Expected Advantages and Research Focus	Main Domains
Foundation and Multimodal Models	Large-scale pre-trained models (sensor LLMs, graph foundation models).	Unified multimodal representations and reduced need for labeled data through self-supervised learning; expected to dominate post-2026.	Industrial, CPS/IoT, Cybersecurity.
Physics-Informed and Grey-Box AI	Embedding physical laws, constraints, and digital twins into neural architectures.	Enhanced interpretability and safety with reduced data dependency; PINNs and differentiable simulators enabling certifiable AI.	Energy, Industrial.
Federated and Privacy-Preserving Learning	Collaborative model training across distributed nodes without data centralization.	Cross-organization learning with privacy guarantees; integration with TinyML for decentralized analytics.	CPS/IoT, Smart Grids, Cybersecurity.
Causal and Explainable Learning	Causal inference and feature attribution using graph explainers and SHAP-based frameworks.	Improved fault root-cause analysis and diagnostic traceability.	Cybersecurity, Industrial, Energy.
AutoML and Continual Learning Pipelines	Automated model optimization and adaptive retraining during operation.	Reduced human effort and sustained model relevance over time; key for long-term autonomous FDD.	Industrial, CPS, Energy.
Edge and Tiny Foundation AI	Quantized and distilled foundation models on microcontrollers.	Ultra-low-power and real-time inference; distributed intelligence at device level.	CPS/IoT, Smart Sensors.
Quantum and Neuromorphic FDD	Quantum kernels and spiking neural networks for ultra-fast fault classification.	Potential exponential gains in speed and energy efficiency; promising for mission-critical diagnostics.	Energy, Cyber-security.

From the synthesis of recent contributions, four converging trajectories emerge:

1. Integration of Foundation and Edge AI: Large multimodal models are progressively distilled into compact edge versions, fusing global knowledge with local adaptability to enable self-healing cyber–physical systems.

2. Physics-Guided Trustworthy AI: Hybrid grey-box architectures (combining PINNs, analytical redundancy, and uncertainty quantification) are becoming central to certification, ready diagnostics.
3. Collaborative and Federated Intelligence: Decentralized learning ecosystems will drive cross-factory and cross-network collaboration while preserving data sovereignty.
4. Causal and Agentic Diagnostics: Future systems will transition from reactive anomaly detection to proactive, reasoning-based agents capable of explaining, predicting, and autonomously mitigating faults under human supervision.

In conclusion, FDD research is advancing toward autonomous, physics-informed, and federated AI ecosystems capable of online adaptation, explainability, and resilient operation across industrial, energy, CPS/IoT, and cybersecurity domains. These developments define the emerging paradigm of self-adaptive, trustworthy, and energy-efficient diagnostic intelligence.

Author Contributions: Conceptualization, D.P. and P.D.; methodology, D.P. and P.D.; software, D.P.; validation, D.P. and P.D.; formal analysis, D.P. and P.D.; investigation, D.P. and P.D.; resources, A.E. and S.S.; data curation, D.P.; writing—original draft preparation, D.P. and P.D.; writing—review and editing, A.E. and S.S.; visualization, D.P.; supervision, A.E. and S.S.; project administration, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: The work has been partially supported by National Centre for HPC, Big Data and Quantum Computing, Spoke 6, CUP B83C22002940006, Multiscale modelling & Engineering applications; and by MIUR FoReLab Project, Dipartimenti di Eccellenza.

Data Availability Statement: No new data were created or analyzed in this study. The data supporting the findings of this review are derived from publicly available sources cited throughout the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shakiba, F.M.; Azizi, S.M.; Zhou, M.; Abusorrah, A. Application of machine learning methods in fault detection and classification of power transmission lines: A survey. *Artif. Intell. Rev.* **2023**, *56*, 5799–5836. [CrossRef]
2. Zhang, J.; He, X. Compound-fault diagnosis of integrated energy systems based on graph embedded recurrent neural networks. *IEEE Trans. Ind. Inform.* **2023**, *20*, 3478–3486. [CrossRef]
3. Alsaif, K.M.; Albeshri, A.A.; Khemakhem, M.A.; Eassa, F.E. Multimodal Large Language Model-Based Fault Detection and Diagnosis in Context of Industry 4.0. *Electronics* **2024**, *13*, 4912. [CrossRef]
4. Carbone, M.A.; Muscatello, M.J.; McCormick, J.A.; Csank, J.T.; Teubert, C.A.; Frank, J.D. A Hybrid Model and Data-Driven Approach for Anomaly Detection in Space Power Systems. In Proceedings of the AIAA ASCEND 2025, Las Vegas, NV, USA, 21–25 July 2025; pp. 1–14.
5. Crotti, E.; Colagrossi, A. Machine Learning Approaches for Data-Driven Self-Diagnosis and Fault Detection in Spacecraft Systems. *Appl. Sci.* **2025**, *15*, 7761. [CrossRef]
6. Rana, B.; Rathore, S.S. Industry 4.0–Applications, challenges and opportunities in industries and academia: A review. *Mater. Today Proc.* **2023**, *79*, 389–394. [CrossRef]
7. Raja Santhi, A.; Muthuswamy, P. Industry 5.0 or industry 4.0 S? Introduction to industry 4.0 and a peek into the prospective industry 5.0 technologies. *Int. J. Interact. Des. Manuf. (IJIDeM)* **2023**, *17*, 947–979. [CrossRef]
8. Gao, Z.; Cecati, C.; Ding, S.X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis with Model-Based and Signal-Based Approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767. [CrossRef]
9. Blanke, M.; Kinnaert, M.; Lunze, J.; Staroswiecki, M. *Diagnosis and Fault-Tolerant Control*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2016. [CrossRef]
10. Lee, L.; Yang, Z.; Liu, Y.; Yu, C. Power Switch Fault Diagnosis Technique for Three-Phase Voltage Source Inverters. *IEICE Electron. Express* **2025**, *22*, 20250281. [CrossRef]
11. Su, H.; Xiang, L.; Hu, A. Application of deep learning to fault diagnosis of rotating machineries. *Meas. Sci. Technol.* **2024**, *35*, 042003. [CrossRef]
12. Xiao, L.; Yang, X.; Yang, X. A graph neural network-based bearing fault detection method. *Sci. Rep.* **2023**, *13*, 5286. [CrossRef]

13. Jombo, G.; Zhang, Y. Acoustic-based machine condition monitoring—Methods and challenges. *Eng* **2023**, *4*, 47–79. [CrossRef]
14. Bai, Y.; Cheng, W.; Wen, W.; Liu, Y. Application of time-frequency analysis in rotating machinery fault diagnosis. *Shock Vib.* **2023**, *2023*, 9878228. [CrossRef]
15. Khanam, R.; Hussain, M.; Hill, R.; Allen, P. A comprehensive review of convolutional neural networks for defect detection in industrial applications. *IEEE Access* **2024**, *12*, 94250–94295. [CrossRef]
16. Farooq, M.A.; Shariff, W.; O'callaghan, D.; Merla, A.; Corcoran, P. On the role of thermal imaging in automotive applications: A critical review. *IEEE Access* **2023**, *11*, 25152–25173. [CrossRef]
17. Love, P.E.; Fang, W.; Matthews, J.; Porter, S.; Luo, H.; Ding, L. Explainable artificial intelligence (XAI): Precepts, models, and opportunities for research in construction. *Adv. Eng. Inform.* **2023**, *57*, 102024. [CrossRef]
18. Calabrese, F.; Regattieri, A.; Bortolini, M.; Galizia, F.G. Data-driven fault detection and diagnosis: Challenges and opportunities in real-world scenarios. *Appl. Sci.* **2022**, *12*, 9212. [CrossRef]
19. Patil, A.; Soni, G.; Prakash, A. Data-driven approaches for impending fault detection of industrial systems: A review. *Int. J. Syst. Assur. Eng. Manag.* **2024**, *15*, 1326–1344. [CrossRef]
20. Martínez-Heredia, A.M.; Ventura, S. Weak Supervision: A Survey on Predictive Maintenance. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2025**, *15*, e70022. [CrossRef]
21. Yang, X.; Song, Z.; King, I.; Xu, Z. A survey on deep semi-supervised learning. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 8934–8954. [CrossRef]
22. e Souza, A.C.O.; de Souza, M.B., Jr.; da Silva, F.V. Development of a CNN-based fault detection system for a real water injection centrifugal pump. *Expert Syst. Appl.* **2024**, *244*, 122947. [CrossRef]
23. Ali, H.; Zhang, Z.; Safdar, R.; Rasool, M.H.; Yao, Y.; Yao, L.; Gao, F. Fault detection using machine learning based dynamic ICA-distributed CCA: Application to industrial chemical process. *Digit. Chem. Eng.* **2024**, *11*, 100156. [CrossRef]
24. Wu, Z.; Yang, X.; Wei, X.; Yuan, P.; Zhang, Y.; Bai, J. A self-supervised anomaly detection algorithm with interpretability. *Expert Syst. Appl.* **2024**, *237*, 121539. [CrossRef]
25. Esmaeili, F.; Cassie, E.; Nguyen, H.P.T.; Plank, N.O.; Unsworth, C.P.; Wang, A. Anomaly detection for sensor signals utilizing deep learning autoencoder-based neural networks. *Bioengineering* **2023**, *10*, 405. [CrossRef]
26. Givnan, S.; Chalmers, C.; Fergus, P.; Ortega-Martorell, S.; Whalley, T. Anomaly detection using autoencoder reconstruction upon industrial motors. *Sensors* **2022**, *22*, 3166. [CrossRef] [PubMed]
27. Shao, S.; Wang, P.; Yan, R. Generative adversarial networks for data augmentation in machine fault diagnosis. *Comput. Ind.* **2019**, *106*, 85–93. [CrossRef]
28. Ding, Y.; Ma, L.; Ma, J.; Wang, C.; Lu, C. A generative adversarial network-based intelligent fault diagnosis method for rotating machinery under small sample size conditions. *IEEE Access* **2019**, *7*, 149736–149749. [CrossRef]
29. Gharoun, H.; Momenifar, F.; Chen, F.; Gandomi, A.H. Meta-learning approaches for few-shot learning: A survey of recent advances. *ACM Comput. Surv.* **2024**, *56*, 294. [CrossRef]
30. Vettoruzzo, A.; Bouguelia, M.R.; Vanschoren, J.; Rögnvaldsson, T.; Santosh, K. Advances and challenges in meta-learning: A technical review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 4763–4779. [CrossRef] [PubMed]
31. Lai, P.; Zhang, F.; Li, T.; Guo, J.; Teng, F. Unlocking the power of knowledge for few-shot fault diagnosis: A review from a knowledge perspective. *Inf. Sci.* **2025**, *706*, 121996. [CrossRef]
32. Gao, W.; Xu, Z.; Akoudad, Y. Metric-based meta-learning relation network for cross-domain few-shot bearing fault diagnosis. *IEEE Sens. J.* **2025**, *25*, 13632–13647. [CrossRef]
33. Berghout, T.; Benbouzid, M.; Bentrcia, T.; Lim, W.H.; Amirat, Y. Federated learning for condition monitoring of industrial processes: A review on fault diagnosis methods, challenges, and prospects. *Electronics* **2022**, *12*, 158. [CrossRef]
34. Demertzi, V.; Demertzis, S.; Demertzis, K. An overview of privacy dimensions on the industrial internet of things (iiot). *Algorithms* **2023**, *16*, 378. [CrossRef]
35. Cavallin, F.; Mayer, R. Anomaly detection from distributed data sources via federated learning. In *Proceedings of the International Conference on Advanced Information Networking and Applications*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 317–328.
36. Yu, X.; Chen, G.; Zeng, X.; He, Z. Federated Transfer Learning-Based Paper Breakage Fault Diagnosis. *Adv. Mater. Sustain. Manuf.* **2024**, *1*, 10009. [CrossRef]
37. Leite, D.; Andrade, E.; Rativa, D.; Maciel, A.M. Fault detection and diagnosis in industry 4.0: A review on challenges and opportunities. *Sensors* **2024**, *25*, 60. [CrossRef] [PubMed]
38. Cheng, Y.; Cao, Y.; Yao, H.; Luo, W.; Jiang, C.; Zhang, H.; Shen, W. A comprehensive survey for real-world industrial defect detection: Challenges, approaches, and prospects. *arXiv* **2025**, arXiv:2507.13378. [CrossRef]
39. Khurram, M.; Zhang, C.; Muhammad, S.; Kishnani, H.; An, K.; Abeywardena, K.; Chadha, U.; Behdinan, K. Artificial Intelligence in Manufacturing Industry Worker Safety: A New Paradigm for Hazard Prevention and Mitigation. *Processes* **2025**, *13*, 1312. [CrossRef]

40. Johnson, J.M.; Khoshgoftaar, T.M. Cost-sensitive ensemble learning for highly imbalanced classification. In *Proceedings of the 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*; IEEE: Piscataway, NJ, USA, 2022; pp. 1427–1434.
41. Alharbi, F.; Luo, S.; Zhang, H.; Shaukat, K.; Yang, G.; Wheeler, C.A.; Chen, Z. A brief review of acoustic and vibration signal-based fault detection for belt conveyor idlers using machine learning models. *Sensors* **2023**, *23*, 1902. [CrossRef] [PubMed]
42. Chukwunweike, J.; Abubakar, I.; Anang, A.N. Enhancing Industrial Efficiency through Automation: Leveraging PLC, Scada, HMI, Batch, and DCS Systems for Downtime Mitigation in Industrial Control Projects. *Int. J. Res. Publ. Rev.* **2024**, *5*, 1466–1478. [CrossRef]
43. OV, G.S.; Karthikeyan, A.; Karthikeyan, K.; Sanjeevikumar, P.; Thomas, S.K.; Babu, A. Critical review of SCADA and PLC in smart buildings and energy sector. *Energy Rep.* **2024**, *12*, 1518–1530. [CrossRef]
44. Wang, T.; Guo, J.; Zhang, B.; Yang, G.; Li, D. Deploying AI on Edge: Advancement and Challenges in Edge Intelligence. *Mathematics* **2025**, *13*, 1878. [CrossRef]
45. Alabbasy, F.M.; Abohamama, A.S.; Alrahmawy, M.F. Compressing medical deep neural network models for edge devices using knowledge distillation. *J. King Saud Univ.-Comput. Inf. Sci.* **2023**, *35*, 101616.
46. Zeeshan, M. Efficient Deep Learning Models for Edge IOT Devices—A Review. *Authorea* **2024**. [CrossRef]
47. Chen, Y.; Zheng, B.; Zhang, Z.; Wang, Q.; Shen, C.; Zhang, Q. Deep learning on mobile and embedded devices: State-of-the-art, challenges, and future directions. *ACM Comput. Surv. (CSUR)* **2020**, *53*, 1–37. [CrossRef]
48. Heydari, S.; Mahmoud, Q.H. Tiny machine learning and on-device inference: A survey of applications, challenges, and future directions. *Sensors* **2025**, *25*, 3191. [CrossRef]
49. Qiu, F.; Kumar, A.; Hu, J.; Sharma, P.; Tang, Y.B.; Xu Xiang, Y.; Hong, J. A Review on Integrating IoT, IIoT, and Industry 4.0: A Pathway to Smart Manufacturing and Digital Transformation. *IET Inf. Secur.* **2025**, *2025*, 9275962.
50. Chowdhury, A.; Nuruzzaman, M. Design, Testing, and Troubleshooting of Industrial Equipment: A Systematic Review of Integration Techniques for Us Manufacturing Plants. *Rev. Appl. Sci. Technol.* **2023**, *2*, 53–84.
51. Cao, S.; Sun, X.; Widyasari, R.; Lo, D.; Wu, X.; Bo, L.; Zhang, J.; Li, B.; Liu, W.; Wu, D.; et al. A systematic literature review on explainability for machine/deep learning-based software engineering research. *arXiv* **2024**, arXiv:2401.14617.
52. Chen, H.Y.; Lee, C.H. Vibration signals analysis by explainable artificial intelligence (XAI) approach: Application on bearing faults diagnosis. *IEEE Access* **2020**, *8*, 134246–134256. [CrossRef]
53. Brusa, E.; Cibrario, L.; Delprete, C.; Di Maggio, L.G. Explainable AI for machine fault diagnosis: Understanding features' contribution in machine learning models for industrial condition monitoring. *Appl. Sci.* **2023**, *13*, 2038. [CrossRef]
54. Sanakkayala, D.C.; Varadarajan, V.; Kumar, N.; Karan; Soni, G.; Kamat, P.; Kumar, S.; Patil, S.; Kotecha, K. Explainable AI for bearing fault prognosis using deep learning techniques. *Micromachines* **2022**, *13*, 1471. [CrossRef]
55. Wang, Z.; Chen, J.; Wang, C.; Peng, C.; Xuan, J.; Shi, T.; Zuo, M. CNC-VLM: An RLHF-optimized industrial large vision-language model with multimodal learning for imbalanced CNC fault detection. *Mech. Syst. Signal Process.* **2026**, *245*, 113838. [CrossRef]
56. Iyaniwura, A.A.; Mayaki, C.S. Artificial Intelligence-enabled smart grid systems for real-time load forecasting, fault detection, renewable energy integration and optimization. *Glob. J. Eng. Technol. Adv.* **2025**, *24*, 191–208. [CrossRef]
57. Pujari, R.; Alam, M.N. Review on distance relaying for the protection of modern power system networks. *IEEE Access* **2025**, *13*, 28861–28893. [CrossRef]
58. Sufi, F. Beyond the Sensor: A Systematic Review of AI's Role in Next-Generation Machine Health Monitoring. *Appl. Sci.* **2025**, *15*, 10494. [CrossRef]
59. Pazderin, A.; Zicmane, I.; Senyuk, M.; Gubin, P.; Polyakov, I.; Mukhlynin, N.; Safaraliev, M.; Kamalov, F. Directions of application of phasor measurement units for control and monitoring of modern power systems: A state-of-the-art review. *Energies* **2023**, *16*, 6203. [CrossRef]
60. Ibrahim, A.H.M.; Sadanandan, S.K.; Ghaoud, T.; Rajkumar, V.S.; Sharma, M. Incipient fault detection in power distribution networks: Review, analysis, challenges and future directions. *IEEE Access* **2024**, *12*, 112822–112838. [CrossRef]
61. Almasoudi, F.M. Enhancing power grid resilience through real-time fault detection and remediation using advanced hybrid machine learning models. *Sustainability* **2023**, *15*, 8348. [CrossRef]
62. Quamar, M.M.; Nasir, A. Review on fault diagnosis and fault-tolerant control scheme for robotic manipulators: Recent advances in AI, machine learning, and digital twin. *arXiv* **2024**, arXiv:2402.02980. [CrossRef]
63. Dan, Y.; Zhong, H.; Wang, C.; Wang, J.; Fei, Y.; Yu, L. A graph deep reinforcement learning-based fault restoration method for active distribution networks. *Energies* **2025**, *18*, 4420. [CrossRef]
64. Alhamrouni, I.; Abdul Kahar, N.H.; Salem, M.; Swadi, M.; Zahroui, Y.; Kadhim, D.J.; Mohamed, F.A.; Alhuyi Nazari, M. A comprehensive review on the role of artificial intelligence in power system stability, control, and protection: Insights and future directions. *Appl. Sci.* **2024**, *14*, 6214. [CrossRef]
65. Fathollahi, A. Machine Learning and Artificial Intelligence Techniques in Smart Grids Stability Analysis: A Review. *Energies* **2025**, *18*, 3431. [CrossRef]

66. Ukoba, K.; Olatunji, K.O.; Adeoye, E.; Jen, T.C.; Madyira, D.M. Optimizing renewable energy systems through artificial intelligence: Review and future prospects. *Energy Environ.* **2024**, *35*, 3833–3879. [CrossRef]
67. Hassan, A.; Mohani, S.S.U.H.; Essani, I.Y.; Taj, S.A.; Aslam, A.; Abbas, Y. AI-Enabled Energy Management for Large-Scale Solar Farms: Optimizing Power Distribution, Grid Stability, and Real-Time Performance Monitoring. *PROGRESS J. Multidiscip. Stud.* **2025**, *6*, 66–84. [CrossRef]
68. Hamdan, A.; Ibekwe, K.I.; Ilojianya, V.I.; Sonko, S.; Etukudoh, E.A. AI in renewable energy: A review of predictive maintenance and energy optimization. *Int. J. Sci. Res. Arch.* **2024**, *11*, 718–729. [CrossRef]
69. Soe, H.M.; Htet, A. A Comprehensive Review of SCADA-Based Wind Turbine Performance and Reliability Modeling with Machine Learning Approaches. *J. Technol. Innov. Energy* **2024**, *3*, 68–92. [CrossRef]
70. Bunyan, S.T.; Khan, Z.H.; Al-Haddad, L.A.; Dhahad, H.A.; Al-Karkhi, M.I.; Ogaili, A.A.F.; Al-Sharify, Z.T. Intelligent Thermal Condition Monitoring for Predictive Maintenance of Gas Turbines Using Machine Learning. *Machines* **2025**, *13*, 401. [CrossRef]
71. Raju, S.K.; Periyasamy, M.; Alhussan, A.A.; Kannan, S.; Raghavendran, S.; El-Kenawy, E.S.M. Machine learning boosts wind turbine efficiency with smart failure detection and strategic placement. *Sci. Rep.* **2025**, *15*, 1485. [CrossRef]
72. Farhat, H.; Altarawneh, A. Physics-Informed Machine Learning for Intelligent Gas Turbine Digital Twins: A Review. *Energies* **2025**, *18*, 5523. [CrossRef]
73. Shafik, W. An Overview of Artificial Intelligence Solutions for the Maintenance and Evaluation of Photovoltaic Systems. In *Energy Conversion Systems-Based Artificial Intelligence: Applications and Tools*; Springer: Singapore, 2025; pp. 23–53.
74. Thakfan, A.; Bin Salamah, Y. Artificial-intelligence-based detection of defects and faults in photovoltaic systems: A survey. *Energies* **2024**, *17*, 4807. [CrossRef]
75. Abro, G.E.M.; Ali, A.; Memon, S.A.; Memon, T.D.; Khan, F. Strategies and Challenges for Unmanned Aerial Vehicle-Based Continuous Inspection and Predictive Maintenance of Solar Modules. *IEEE Access* **2024**, *12*, 176615–176629. [CrossRef]
76. Ledmaoui, Y.; El Maghraoui, A.; El Aroussi, M.; Saadane, R. Enhanced fault detection in photovoltaic panels using cnn-based classification with pyqt5 implementation. *Sensors* **2024**, *24*, 7407. [CrossRef]
77. Naser, R.A.; Altahir, A.A.R.; Ahmed, A.A.; Alanbari, M.H. Enhancing Solar Panel Efficiency in Harsh Climates: An UAV-Integrated WSN Cleaning Approach. In *Proceedings of the International Conference on Data Analytics & Management*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 64–80.
78. García-Pérez, D.; Saeed, M.; Díaz, I.; Enguita, J.M.; Guerrero, J.M.; Briz, F. Machine learning for Inverter-Fed motors monitoring and fault detection: An overview. *IEEE Access* **2024**, *12*, 27167–27179. [CrossRef]
79. Roy, S.; Stevenson, A.; Tufail, S.; Riggs, H.; Tariq, M.; Sarwat, A. Machine Learning-Driven Reliability Estimation of PV Inverters Considering Alert-Ambient Variability. *IEEE Trans. Ind. Appl.* **2024**, *61*, 3538–3552. [CrossRef]
80. Said, N.; Mansouri, M.; Al Hmouz, R.; Khedher, A. Deep Learning Techniques for Fault Diagnosis in Interconnected Systems: A Comprehensive Review and Future Directions. *Appl. Sci.* **2025**, *15*, 6263. [CrossRef]
81. Omojola, A.F.; Ilabija, C.O.; Onyeka, C.; Ishiwu, J.; Olaleye, T.G.; Ozoemena, I.J.; Nzereogu, P.U. Artificial Intelligence-Driven Strategies for Advancing Lithium-Ion Battery Performance and Safety. *Int. J. Adv. Eng. Manag.* **2024**, *6*, 452–484.
82. Yang, M.; Rong, M.; Ye, Y.; Zhang, Y.; Yang, A.; Chu, J.; Yuan, H.; Wang, X. A comprehensive study of thermal runaway behavior and early warning subjected to internal short-circuit. *J. Power Sources* **2024**, *620*, 235213. [CrossRef]
83. Ghazali, A.K.; Aziz, N.A.A.; Hassan, M.K. Advanced algorithms in battery management systems for electric vehicles: A comprehensive review. *Symmetry* **2025**, *17*, 321. [CrossRef]
84. Muneer, B.; Palazzi, V.; Alimenti, F.; Mezzanotte, P.; Roselli, L. A Way Towards Energy Autonomous Wireless Sensing for EV Battery Management System. *IEEE J. Microw.* **2025**, *5*, 555–571. [CrossRef]
85. Rana, S. AI-driven fault detection and predictive maintenance in electrical power systems: A systematic review of data-driven approaches, digital twins, and self-healing grids. *Am. J. Adv. Technol. Eng. Solut.* **2025**, *1*, 258–289.
86. Dong, Z.; Tao, Y.; Lai, S.; Wang, T.; Zhang, Z. Powering Future Advancements and Applications of Battery Energy Storage Systems Across Different Scales. *Energy Storage Appl.* **2025**, *2*, 1. [CrossRef]
87. Jouini, O.; Sethom, K.; Namoun, A.; Aljohani, N.; Alanazi, M.H.; Alanazi, M.N. A survey of machine learning in edge computing: Techniques, frameworks, applications, issues, and research directions. *Technologies* **2024**, *12*, 81. [CrossRef]
88. Kumar, A.; Gutierrez, J.A. Impact of Machine Learning on Intrusion Detection Systems for the Protection of Critical Infrastructure. *Information* **2025**, *16*, 515. [CrossRef]
89. Khan, M.T.; Waheed, A. Foundation Model Driven Robotics: A Comprehensive Review. *arXiv* **2025**, arXiv:2507.10087. [CrossRef]
90. Belchandan, R.K.; Cecchi, V. Classification of Distribution System Modeling Inaccuracies and Their Effects on Planning Studies. In *Proceedings of the SoutheastCon 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 747–752.
91. Muneer, S.; Farooq, U.; Athar, A.; Ahsan Raza, M.; Ghazal, T.M.; Sakib, S. A critical review of artificial intelligence based approaches in intrusion detection: A comprehensive analysis. *J. Eng.* **2024**, *2024*, 3909173. [CrossRef]

92. Mohite, R.; Ouarbya, L. Interpretable anomaly detection: A hybrid approach using rule-based and machine learning techniques. In *Proceedings of the 2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–10.
93. Kasri, W.; Himeur, Y.; Alkhazaleh, H.A.; Tarapiah, S.; Atalla, S.; Mansoor, W.; Al-Ahmad, H. From vulnerability to defense: The role of large language models in enhancing cybersecurity. *Computation* **2025**, *13*, 30. [CrossRef]
94. Guo, Y.; Cheng, Z.; Zhang, J.; Sun, B.; Wang, Y. A review on adversarial–based deep transfer learning mechanical fault diagnosis. *J. Big Data* **2024**, *11*, 151. [CrossRef]
95. Halgamuge, M.N. Leveraging deep learning to strengthen the cyber-resilience of renewable energy supply chains: A survey. *IEEE Commun. Surv. Tutor.* **2024**, *26*, 2146–2175.
96. Zhang, T.; Xue, C.; Wang, J.; Yun, Z.; Lin, N.; Han, S. A survey on industrial Internet of Things (IIoT) Testbeds for connectivity research. *arXiv* **2024**, arXiv:2404.17485. [CrossRef]
97. Aghazadeh Ardebili, A.; Hasidi, O.; Bendaouia, A.; Khalil, A.; Khalil, S.; Luceri, D.; Longo, A.; Abdelwahed, E.H.; Qassimi, S.; Ficarella, A. Enhancing resilience in complex energy systems through real-time anomaly detection: A systematic literature review. *Energy Inform.* **2024**, *7*, 96.
98. Singh, N.; Buyya, R.; Kim, H. Securing cloud-based internet of things: Challenges and mitigations. *Sensors* **2024**, *25*, 79. [CrossRef]
99. Zhan, S.; Huang, L.; Luo, G.; Zheng, S.; Gao, Z.; Chao, H.C. A Review on Federated Learning Architectures for Privacy-Preserving AI: Lightweight and Secure Cloud–Edge–End Collaboration. *Electronics* **2025**, *14*, 2512.
100. Quan, M.K.; Pathirana, P.N.; Wijayasundara, M.; Setunge, S.; Nguyen, D.C.; Brinton, C.G.; Love, D.J.; Poor, H.V. Federated learning for cyber physical systems: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2025**, *28*, 3751–3790.
101. Aggarwal, M.; Khullar, V.; Goyal, N. A comprehensive review of federated learning: Methods, applications, and challenges in privacy-preserving collaborative model training. In *Applied Data Science and Smart Systems*; CRC Press: Boca Raton, FL, USA, 2024; pp. 570–575.
102. Zhang, C.; Yang, S.; Mao, L.; Ning, H. Anomaly detection and defense techniques in federated learning: A comprehensive review. *Artif. Intell. Rev.* **2024**, *57*, 150. [CrossRef]
103. Berkani, M.R.A.; Chouchane, A.; Himeur, Y.; Ouamane, A.; Miniaoui, S.; Atalla, S.; Mansoor, W.; Al-Ahmad, H. Advances in federated learning: Applications and challenges in smart building environments and beyond. *Computers* **2025**, *14*, 124. [CrossRef]
104. Ali, A.; Jianjun, H.; Jabbar, A. Recent Advances in Federated Learning for Connected Autonomous Vehicles: Addressing Privacy, Performance, and Scalability Challenges. *IEEE Access* **2025**, *13*, 80637–80665. [CrossRef]
105. Adam, M.; Baroud, U. Federated learning for IoT: Applications, trends, taxonomy, challenges, current solutions, and future directions. *IEEE Open J. Commun. Soc.* **2024**, *5*, 7842–7877. [CrossRef]
106. Mengistu, T.M.; Kim, T.; Lin, J.W. A survey on heterogeneity taxonomy, security and privacy preservation in the integration of IoT, wireless sensor networks and federated learning. *Sensors* **2024**, *24*, 968. [CrossRef] [PubMed]
107. Khan, M.A.; Rais, R.N.B.; Khalid, O.; Deriche, M. Comparative Analysis of Centralized and Federated Intrusion Detection in IoT-Enabled Cyber-Physical Systems Under Data and Label-Skew. *IEEE Access* **2025**, *13*, 160767–160785.
108. da Silva, C.N.; Prazeres, C.V. Tiny Federated Learning for Constrained Sensors: A Systematic Literature Review. *IEEE Sens. Rev.* **2025**, *2*, 17–31. [CrossRef]
109. Wainbuch, R.; Samuel, A.J. TinyML: Deploying Machine Learning on Microcontrollers for IoT Applications. *J. Sci. Technol. Eng. Res.* **2024**, *2*, 44–57.
110. Asutkar, S.; Chalke, C.; Shivgan, K.; Tallur, S. TinyML-enabled edge implementation of transfer learning framework for domain generalization in machine fault diagnosis. *Expert Syst. Appl.* **2023**, *213*, 119016. [CrossRef]
111. Capogrosso, L.; Cunico, F.; Cheng, D.S.; Fummi, F.; Cristani, M. A machine learning-oriented survey on tiny machine learning. *IEEE Access* **2024**, *12*, 23406–23426. [CrossRef]
112. Deutel, M.; Woller, P.; Mutschler, C.; Teich, J. Energy-efficient deployment of deep learning applications on Cortex-M based microcontrollers using deep compression. *arXiv* **2022**, arXiv:2205.10369.
113. Deutel, M.; Hannig, F.; Mutschler, C.; Teich, J. On-Device Training of Fully Quantized Deep Neural Networks on Cortex-M Microcontrollers. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **2024**, *44*, 1250–1261.
114. Lucan Orășan, I.; Seiculescu, C.; Căleanu, C.D. A brief review of deep neural network implementations for ARM cortex-M processor. *Electronics* **2022**, *11*, 2545. [CrossRef]
115. Elhanashi, A.; Dini, P.; Saponara, S.; Zheng, Q. Advancements in TinyML: Applications, limitations, and impact on IoT devices. *Electronics* **2024**, *13*, 3562. [CrossRef]
116. Yasaei, R.; Moghaddas, Y.; Al Faruque, M.A. IoT-GRAF: IoT Graph Learning-Based Anomaly and Intrusion Detection Through Multi-Modal Data Fusion. In *Proceedings of the 2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
117. Dong, G.; Tang, M.; Wang, Z.; Gao, J.; Guo, S.; Cai, L.; Gutierrez, R.; Campbel, B.; Barnes, L.E.; Boukhechba, M. Graph neural networks in IoT: A survey. *ACM Trans. Sens. Netw.* **2023**, *19*, 47. [CrossRef]

118. Villegas-Ch, W.; Govea, J.; Navarro, A.M.; Játiva, P.P. Intrusion Detection in IoT Networks Using Dynamic Graph Modeling and Graph-Based Neural Networks. *IEEE Access* **2025**, *13*, 65356–65375.
119. Tung, N.X.; Son, B.D.; Geun-Jeong, S.; Van Chien, T.; Hanzo, L.; Hwang, W.J. Graph neural networks for next-generation-iot: Recent advances and open challenges. *IEEE Commun. Surv. Tutor.* **2025**, *28*, 2226–2262. [CrossRef]
120. Javadi, S.; Riboni, D.; Borzì, L.; Zolfaghari, S. Graph-Based Methods for Multimodal Indoor Activity Recognition: A Comprehensive Survey. *IEEE Trans. Comput. Soc. Syst.* **2025**, *12*, 3728–3746. [CrossRef]
121. Chen, J.; He, J.; Chen, F.; Lv, Z.; Tang, J.; Li, W.; Liu, Z.; Yang, H.H.; Han, G. Towards General Industrial Intelligence: A Survey of Continual Large Models in Industrial IoT. *arXiv* **2024**, arXiv:2409.01207.
122. Usmani, U.A.; Aziz, I.A.; Jaafar, J.; Watada, J. Deep Learning for Anomaly Detection in Time-Series Data: An Analysis of Techniques, Review of Applications, and Guidelines for Future Research. *IEEE Access* **2024**, *12*, 174564–174590. [CrossRef]
123. Shrestha, R.; Mohammadi, M.; Sinaei, S.; Salcines, A.; Pampliega, D.; Clemente, R.; Sanz, A.L.; Nowroozi, E.; Lindgren, A. Anomaly detection based on lstm and autoencoders using federated learning in smart electric grid. *J. Parallel Distrib. Comput.* **2024**, *193*, 104951. [CrossRef]
124. Guato Burgos, M.F.; Morato, J.; Vizcaino Imacaña, F.P. A review of smart grid anomaly detection approaches pertaining to artificial intelligence. *Appl. Sci.* **2024**, *14*, 1194. [CrossRef]
125. Mercorelli, P. Recent advances in intelligent algorithms for fault detection and diagnosis. *Sensors* **2024**, *24*, 2656. [CrossRef] [PubMed]
126. Baset, A.; Jradi, M. Data-driven decision support for smart and efficient building energy retrofits: A review. *Appl. Syst. Innov.* **2024**, *8*, 5. [CrossRef]
127. Bandara, R.M.P.N.S.; Jayasignhe, A.B.; Retscher, G. The Integration of IoT (Internet of Things) Sensors and Location-Based Services for Water Quality Monitoring: A Systematic Literature Review. *Sensors* **2025**, *25*, 1918.
128. Jagadeesh, V.; Sivakumar, P. Enhanced Pipeline Safety: Cloud-Based Leak Prediction Using XGBoost. In *Proceedings of the 2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1087–1091.
129. Raouf, I.; Khan, A.; Khalid, S.; Sohail, M.; Azad, M.M.; Kim, H.S. Sensor-based prognostic health management of advanced driver assistance system for autonomous vehicles: A recent survey. *Mathematics* **2022**, *10*, 3233. [CrossRef]
130. Hossain, M.; Rahman, M.; Ramasamy, D. Artificial intelligence-driven vehicle fault diagnosis to revolutionize automotive maintenance: A review. *Comput. Model. Eng. Sci.* **2024**, *141*, 951. [CrossRef]
131. Vermesan, O.; Pétrot, F.; Coppola, M.; Schneider, M.; Höß, A. Industrial AI Technologies for Next-Generation Autonomous Operations with Sustainable Performance. In *Intelligent Edge-Embedded Technologies for Digitising Industry*; River Publishers: Gistrup, Denmark, 2022; pp. 1–71.
132. Hudda, S.; Haribabu, K. A review on WSN based resource constrained smart IoT systems. *Discov. Internet Things* **2025**, *5*, 56. [CrossRef]
133. Tekin, N.; Acar, A.; Aris, A.; Uluagac, A.S.; Gungor, V.C. Energy consumption of on-device machine learning models for IoT intrusion detection. *Internet Things* **2023**, *21*, 100670. [CrossRef]
134. Aguzzi, G.; Casadei, R.; Pianini, D.; Viroli, M. Dynamic decentralization domains for the internet of things. *IEEE Internet Comput.* **2022**, *26*, 16–23. [CrossRef]
135. Trigka, M.; Dritsas, E. Edge and cloud computing in smart cities. *Future Internet* **2025**, *17*, 118. [CrossRef]
136. El-Hajj, M. Enhancing communication networks in the new era with artificial intelligence: Techniques, applications, and future directions. *Network* **2025**, *5*, 1. [CrossRef]
137. Morshedi, R.; Matinkhah, S.M. A Comprehensive Review of Deep Learning Techniques for Anomaly Detection in IoT Networks: Methods, Challenges, and Datasets. *Eng. Rep.* **2025**, *7*, e70415. [CrossRef]
138. Wang, Z.; Ragab, M.; Yang, W.; Wu, M.; Pan, S.J.; Zhang, J.; Chen, Z. Overcoming Negative Transfer by Online Selection: Distant Domain Adaptation for Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3538009. [CrossRef]
139. Saber, A.M.; Maheshwari, A.; Youssef, A.; Kundur, D. Adversarial Attacks on Deep Learning-Based False Data Injection Detection in Differential Relays. *IEEE Trans. Smart Grid* **2025**, *16*, 4155–4166. [CrossRef]
140. Shivashankar, K.; Al Hajj, G.; Martini, A. Maintainability and Scalability in Machine Learning: Challenges and Solutions. *ACM Comput. Surv.* **2025**, *57*, 318. [CrossRef]
141. Xu, H.; Seng, K.P.; Ang, L.M.; Smith, J. Decentralized and distributed learning for AIoT: A comprehensive review, emerging challenges, and opportunities. *IEEE Access* **2024**, *12*, 101016–101052. [CrossRef]
142. Peng, C.; Peng, J.; Wang, Z.; Wang, Z.; Chen, J.; Xuan, J.; Shi, T. Adaptive fault diagnosis of railway vehicle on-board controller with large language models. *Appl. Soft Comput.* **2025**, *185*, 113919. [CrossRef]
143. Zakaria, A.A.; Amr, T.; Ragheb, A.A. IoT in Smart Urban Planning: A Comprehensive Review of Applications, Developments and Engineering Perspectives. *IEEE Access* **2025**, *13*, 135316–135335. [CrossRef]

144. Hernández-Rivas, A.; Morales-Rocha, V.; Sánchez-Solís, J.P. Towards autonomous cybersecurity: A comparative analysis of agnostic and hybrid AI approaches for advanced persistent threat detection. In *Innovative Applications of Artificial Neural Networks to Data Analytics and Signal Processing*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 181–219.
145. Ali, M.L.; Thakur, K.; Schmeelk, S.; Debello, J.; Dragos, D. Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study. *Appl. Sci.* **2025**, *15*, 1903. [CrossRef]
146. Latibari, B.S.; Nazari, N.; Chowdhury, M.A.; Gubbi, K.I.; Fang, C.; Ghimire, S.; Hosseini, E.; Sayadi, H.; Homayoun, H.; Salehi, S.; et al. Transformers: A security perspective. *IEEE Access* **2024**, *12*, 181071–181105. [CrossRef]
147. Agrawal, G.; Kaur, A.; Myneni, S. A review of generative models in generating synthetic attack data for cybersecurity. *Electronics* **2024**, *13*, 322. [CrossRef]
148. Ali, M.; Udoidiok, I.; Li, F.; Zhang, J. A review on generative intelligence in deep learning based network intrusion detection. In *Proceedings of the 2024 Cyber Awareness and Research Symposium (CARS)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–9.
149. Dunmore, A.; Jang-Jaccard, J.; Sabrina, F.; Kwak, J. A comprehensive survey of generative adversarial networks (GANs) in cybersecurity intrusion detection. *IEEE Access* **2023**, *11*, 76071–76094. [CrossRef]
150. Kumar, V.; Sinha, D. Synthetic attack data generation model applying generative adversarial network for intrusion detection. *Comput. Secur.* **2023**, *125*, 103054.
151. Yu, Z.; Zeng, J.; Chen, S.; Xu, W.; Xu, D.; Liu, X.; Ying, Z.; Wang, N.; Zhang, Y.; Yang, M. CS-Eval: A Comprehensive Large Language Model Benchmark for CyberSecurity. *arXiv* **2024**, arXiv:2411.16239. [CrossRef]
152. Liu, Z.; Zheng, S.; Zhang, F.; Wang, Z.; Wu, W.; Su, W. Cyber Security Entity Recognition Model Based on Cross-Attention Feature Enhancement. In *Proceedings of the International Conference on Intelligent Computing*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 211–222.
153. Bahlali, A.R.; Bachir, A.; Labed, A. Self-Supervised Learning Meets Custom Autoencoder Classifier: A Semi-Supervised Approach for Encrypted Traffic Anomaly Detection. *IEEE Access* **2025**, *13*, 139141–139154.
154. Koukoulis, I.; Syrigos, I.; Korakis, T. Self-Supervised Transformer-based Contrastive Learning for Intrusion Detection Systems. *arXiv* **2025**, arXiv:2505.08816.
155. Dong, J.; Qu, X.; Wang, Z.J.; Ong, Y.S. Enhancing Adversarial Robustness via Uncertainty-Aware Distributional Adversarial Training. *arXiv* **2024**, arXiv:2411.02871. [CrossRef]
156. Sucipto, W.; Zhou, J.; Kwon, R.S.M.; Chen, F. A Survey of TinyML Applications in Beekeeping for Hive Monitoring and Management. *arXiv* **2025**, arXiv:2509.08822. [CrossRef]
157. Gaspar, D.; Silva, P.; Silva, C. Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron. *IEEE Access* **2024**, *12*, 30164–30175. [CrossRef]
158. Salloum, S.; Norozpour, S. XAI-IDS: A Transparent and Interpretable Framework for Robust Cybersecurity Using Explainable Artificial Intelligence. *SHIFRA* **2025**, *2025*, 69–80. [CrossRef]
159. Khan, N.; Ahmad, K.; Tamimi, A.A.; Alani, M.M.; Bermak, A.; Khalil, I. Explainable AI-based Intrusion Detection System for Industry 5.0: An Overview of the Literature, associated Challenges, the existing Solutions, and Potential Research Directions. *arXiv* **2024**, arXiv:2408.03335.
160. 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. *IEEE Access* **2025**, *13*, 55258–55276. [CrossRef]
161. Tursunalieva, A.; Alexander, D.L.; Dunne, R.; Li, J.; Riera, L.; Zhao, Y. Making sense of machine learning: A review of interpretation techniques and their applications. *Appl. Sci.* **2024**, *14*, 496. [CrossRef]
162. Al-Shetwi, A.Q.; Atawi, I.E.; El-Hameed, M.A.; Abuelrub, A. Digital Twin Technology for Renewable Energy, Smart Grids, Energy Storage and Vehicle-to-Grid Integration: Advancements, Applications, Key Players, Challenges and Future Perspectives in Modernising Sustainable Grids. *IET Smart Grid* **2025**, *8*, e70026. [CrossRef]
163. Ginzburg-Ganz, E.; Horodi, E.D.; Shadafny, O.; Savir, U.; Machlev, R.; Levron, Y. Statistical Foundations of Generative AI for Optimal Control Problems in Power Systems: Comprehensive Review and Future Directions. *Energies* **2025**, *18*, 2461. [CrossRef]
164. El-Fergany, A.A. Reviews on load flow methods in electric distribution networks. *Arch. Comput. Methods Eng.* **2025**, *32*, 1619–1633.
165. Zaboli, A.; Kasimalla, S.R.; Park, K.; Hong, Y.; Hong, J. A Comprehensive Review of Behind-the-Meter Distributed Energy Resources Load Forecasting: Models, Challenges, and Emerging Technologies. *Energies* **2024**, *17*, 2534. [CrossRef]
166. Ginzburg-Ganz, E.; Segev, I.; Balabanov, A.; Segev, E.; Kaully Naveh, S.; Machlev, R.; Belikov, J.; Katzir, L.; Keren, S.; Levron, Y. Reinforcement learning model-based and model-free paradigms for optimal control problems in power systems: Comprehensive review and future directions. *Energies* **2024**, *17*, 5307. [CrossRef]
167. Razak, T.R.; Ismail, M.H.; Darus, M.Y.; Jarimi, H.; Su, Y. Artificial Intelligence in Renewable Energy: A Systematic Review of Trends in Solar, Wind, and Smart Grid Applications. *Res. Rev. Sustain.* **2025**, *1*, 1–22. [CrossRef]
168. Yu, J.; Li, X.; Yang, L.; Li, L.; Huang, Z.; Shen, K.; Yang, X.; Yang, X.; Xu, Z.; Zhang, D.; et al. Deep learning models for PV power forecasting. *Energies* **2024**, *17*, 3973. [CrossRef]

169. Rahmani, S.; Aghalar, H.; Jebreili, S.; Goli, A. Optimization and computing using intelligent data-driven approaches for decision-making. In *Optimization and Computing Using Intelligent Data-Driven Approaches for Decision-Making*; CRC Press: Boca Raton, FL, USA, 2024; pp. 90–176.
170. El Rhatrif, A.; Bouihi, B.; Mestari, M. AI-based solutions for grid stability and efficiency: Challenges, limitations, and opportunities. *Int. J. Internet Things Web Serv.* **2024**, *9*, 16–28.
171. Meng, C.; Griesemer, S.; Cao, D.; Seo, S.; Liu, Y. When physics meets machine learning: A survey of physics-informed machine learning. *Mach. Learn. Comput. Sci. Eng.* **2025**, *1*, 20. [CrossRef]
172. Strasser, T.I.; Widl, E.; Kuchenbuch, R.A.; Lázaro-Elorriaga, L.; Laraudogoitia, B.T.; Ginocchi, M.; Penthong, T.; Ponci, F.; Gyrard, A.; Kung, A.; et al. Towards interoperability testing of smart energy systems–an overview and discussion of possibilities. In *Proceedings of the IET Conference Proceedings CP904*; IET: London, UK, 2024; Volume 2024, pp. 263–268.
173. Mishra, D.; Mishra, R.K.; Agarwal, R. Cyber-Physical Systems and Internet of Things (IoT): Convergence, Architectures, and Engineering Applications. In *Advances in Engineering Science and Applications*; Bhumi Publishing: Kolhapur, India, 2025; pp. 17–46.
174. Abshari, D.; Sridhar, M. A survey of anomaly detection in cyber-physical systems. *arXiv* **2025**, arXiv:2502.13256. [CrossRef]
175. Segovia-Ferreira, M.; Rubio-Hernan, J.; Cavalli, A.; Garcia-Alfaro, J. A survey on cyber-resilience approaches for cyber-physical systems. *ACM Comput. Surv.* **2024**, *56*, 1–37. [CrossRef]
176. Amangeldy, B.; Imankulov, T.; Tasmurzayev, N.; Dikhanbayeva, G.; Nurakhov, Y. A Review of Artificial Intelligence and Deep Learning Approaches for Resource Management in Smart Buildings. *Buildings* **2025**, *15*, 2631. [CrossRef]
177. Yeong, D.J.; Panduru, K.; Walsh, J. Exploring the unseen: A survey of multi-sensor fusion and the role of explainable ai (xai) in autonomous vehicles. *Sensors* **2025**, *25*, 856. [CrossRef]
178. Shivashankar, K.; Hajj, G.S.A.; Martini, A. Scalability and Maintainability Challenges and Solutions in Machine Learning: Systematic Literature Review. *arXiv* **2025**, arXiv:2504.11079. [CrossRef]
179. Pene, P.; Musa, A.A.; Musa, U.; Liao, W.; Yu, W. Edge intelligence in smart energy CPS. In *Edge Intelligence in Cyber-Physical Systems*; Elsevier: Amsterdam, The Netherlands, 2025; pp. 169–192.
180. Walia, G.K.; Kumar, M.; Gill, S.S. AI-empowered fog/edge resource management for IoT applications: A comprehensive review, research challenges, and future perspectives. *IEEE Commun. Surv. Tutor.* **2023**, *26*, 619–669. [CrossRef]
181. Okoli, U.I.; Obi, O.C.; Adewusi, A.O.; Abrahams, T.O. Machine learning in cybersecurity: A review of threat detection and defense mechanisms. *World J. Adv. Res. Rev.* **2024**, *21*, 2286–2295. [CrossRef]
182. Ceviz, O.; Sen, S.; Sadioglu, P. A survey of security in uavs and fanets: Issues, threats, analysis of attacks, and solutions. *IEEE Commun. Surv. Tutor.* **2024**, *27*, 3227–3265.
183. Mallick, M.A.I.; Nath, R. Navigating the cyber security landscape: A comprehensive review of cyber-attacks, emerging trends, and recent developments. *World Sci. News* **2024**, *190*, 1–69.
184. Wu, Y.; Zou, B.; Cao, Y. Current status and challenges and future trends of deep learning-based intrusion detection models. *J. Imaging* **2024**, *10*, 254. [CrossRef]
185. Moorthy, S.K.; Jagannath, J. Survey of graph neural network for internet of things and nextg networks. *IEEE Commun. Surv. Tutor.* **2025**, *28*, 1803–1844.
186. Al-Ajlan, M.; Ykhlef, M. A Review of Generative Adversarial Networks for Intrusion Detection Systems: Advances, Challenges, and Future Directions. *Comput. Mater. Contin.* **2024**, *81*, 2053–2076. [CrossRef]
187. Karras, A.; Theodorakopoulos, L.; Karras, C.; Theodoropoulou, A.; Kalliampakou, I.; Kalogeratos, G. LLMs for Cybersecurity in the Big Data Era: A Comprehensive Review of Applications, Challenges, and Future Directions. *Information* **2025**, *16*, 957. [CrossRef]
188. Abbas, N.A.B.; Ahmad, M.R.B. A Comprehensive Review of Tiny Machine Learning: Enabling AI on Resource-Constrained Devices. *ELEKTRIKA-J. Electr. Eng.* **2025**, *24*, 81–97.
189. Alzaabi, F.R.; Mehmood, A. A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. *IEEE Access* **2024**, *12*, 30907–30927.
190. Verma, S.; Prabakeran, S. A Hybrid Deep Learning Approach to Network Traffic Anomaly Detection Enhanced by SHAP and LIME Interpretability. In *Proceedings of the 2025 8th International Conference on Trends in Electronics and Informatics (ICOEI)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1254–1261.
191. Salih, A.M.; Raisi-Estabragh, Z.; Galazzo, I.B.; Radeva, P.; Petersen, S.E.; Lekadir, K.; Menegaz, G. A perspective on explainable artificial intelligence methods: SHAP and LIME. *Adv. Intell. Syst.* **2025**, *7*, 2400304.
192. Mallidi, S.K.R.; Ramisetty, R.R. Advancements in training and deployment strategies for AI-based intrusion detection systems in iot: A systematic literature review. *Discov. Internet Things* **2025**, *5*, 8.
193. Maican, C.A.; Pană, C.F.; Pătrașcu-Pană, D.M.; Rădulescu, V.M. Review of Fault Detection and Diagnosis Methods in Power Plants: Algorithms, Architectures, and Trends. *Appl. Sci.* **2025**, *15*, 6334. [CrossRef]
194. Trillo, J.R.; González-López, F.; Morente-Molinera, J.A.; Magán-Carrión, R.; García-Sánchez, P. Evaluation of Explainable, Interpretable and Non-Interpretable Algorithms for Cyber Threat Detection. *Electronics* **2025**, *14*, 3073. [CrossRef]

195. Matus, K.J.; Veale, M. Certification systems for machine learning: Lessons from sustainability. *Regul. Gov.* **2022**, *16*, 177–196.
196. Kalasampath, K.; Spoorthi, K.; Sajeev, S.; Kuppa, S.S.; Ajay, K.; Angulakshmi, M. A Literature review on applications of explainable artificial intelligence (XAI). *IEEE Access* **2025**, *13*, 41111–41140. [CrossRef]
197. Pelosi, D.; Cacciagrano, D.; Piangerelli, M. Explainability and Interpretability in Concept and Data Drift: A Systematic Literature Review. *Algorithms* **2025**, *18*, 443. [CrossRef]
198. Abdalla, H.B.; Kumar, Y.; Marchena, J.; Guzman, S.; Awlla, A.; Gheisari, M.; Cheraghy, M. The Future of Artificial Intelligence in the Face of Data Scarcity. *Comput. Mater. Contin.* **2025**, *84*, 1073–1099. [CrossRef]
199. Jankauskas, M.; Serackis, A.; Šapurov, M.; Pomarnacki, R.; Baskys, A.; Hyunh, V.K.; Vaimann, T.; Zakis, J. Exploring the limits of early predictive maintenance in wind turbines applying an anomaly detection technique. *Sensors* **2023**, *23*, 5695. [CrossRef]
200. Zaben, M.M.; Worku, M.Y.; Hassan, M.A.; Abido, M.A. Machine learning methods for fault diagnosis in ac microgrids: A systematic review. *IEEE Access* **2024**, *12*, 20260–20298. [CrossRef]
201. Whang, S.E.; Roh, Y.; Song, H.; Lee, J.G. Data collection and quality challenges in deep learning: A data-centric ai perspective. *VLDB J.* **2023**, *32*, 791–813. [CrossRef]
202. Fuhg, J.N.; Padmanabha, G.A.; Bouklas, N.; Bahmani, B.; Sun, W.; Vlassis, N.N.; Flaschel, M.; Carrara, P.; De Lorenzis, L. A review on data-driven constitutive laws for solids. *arXiv* **2024**, arXiv:2405.03658. [CrossRef]
203. Gui, J.; Chen, T.; Zhang, J.; Cao, Q.; Sun, Z.; Luo, H.; Tao, D. A survey on self-supervised learning: Algorithms, applications, and future trends. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 9052–9071. [CrossRef]
204. Farea, A.; Yli-Harja, O.; Emmert-Streib, F. Understanding physics-informed neural networks: Techniques, applications, trends, and challenges. *AI* **2024**, *5*, 1534–1557. [CrossRef]
205. Dini, P.; Saponara, S.; Chakraborty, S.; Hegazy, O. Modeling, Control and Monitoring of Automotive Electric Drives. *Electronics* **2025**, *14*, 3950. [CrossRef]
206. Javed, H.; El-Sappagh, S.; Abuhmed, T. Robustness in deep learning models for medical diagnostics: Security and adversarial challenges towards robust AI applications. *Artif. Intell. Rev.* **2024**, *58*, 12. [CrossRef]
207. Bittla, S.R. Predicting Failures with AI/ML Analytics. In *AI-Driven Software Testing: Transforming Software Testing with Artificial Intelligence and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 435–454.
208. Roy, P. Enhancing Real-World Robustness in AI: Challenges and Solutions. *J. Recent Trends Comput. Sci. Eng.* **2024**, *12*, 34–49. [CrossRef]
209. Anitha, C.; Tellur, A.; Rao, K.; Kumbhar, V.; Gopi, T.; Jadhav, S.; Vidhya, R. Enhancing Cyber-Physical Systems Dependability through Integrated CPS-IoT Monitoring. *Int. Res. J. Multidiscip. Scope* **2024**, *5*, 706–713. [CrossRef]
210. Manchingal, S.K. Epistemic Deep Learning: Enabling Machine Learning Models to Know When They Do Not Know. *arXiv* **2025**, arXiv:2510.22261. [CrossRef]
211. Ennaji, S.; De Gaspari, F.; Hitaj, D.; Kbidi, A.; Mancini, L.V. Adversarial challenges in network intrusion detection systems: Research insights and future prospects. *IEEE Access* **2025**, *13*, 148613–148645. [CrossRef]
212. Mohapatra, H. A Comprehensive Review on Urban Resilience via Fault-Tolerant IoT and Sensor Networks. *Comput. Mater. Contin.* **2025**, *85*, 221–247. [CrossRef]
213. Liu, R.; Huang, J.; Lu, B.; Ding, W. Certified Neural Network Control Architectures: Methodological Advances in Stability, Robustness, and Cross-Domain Applications. *Mathematics* **2025**, *13*, 1677. [CrossRef]
214. Adeyeye, O.J.; Akanbi, I. Artificial intelligence for systems engineering complexity: A review on the use of AI and machine learning algorithms. *Comput. Sci. IT Res. J.* **2024**, *5*, 787–808. [CrossRef]
215. Ayyat, M.; Osman, M.; Nadeem, T. Opportunities and Challenges of Foundation Models in Industrial Manufacturing. *IEEE Access* **2025**, *13*, 138745–138775. [CrossRef]
216. Li, J.L.; Hsu, C.F.; Chang, M.C.; Chen, W.C. A Comprehensive Review of Machine Learning Advances on Data Change: A Cross-Field Perspective. *arXiv* **2024**, arXiv:2402.12627. [CrossRef]
217. Miah, M.A.; Faruque, M.; Akter, S.; Jahan, I. Ethical Considerations in AI and Information Technology Privacy and Bias. *Management* **2024**, *1*, 33–37.
218. Orwat, C.; Bareis, J.; Folberth, A.; Jahnel, J.; Wadephul, C. Normative challenges of risk regulation of artificial intelligence. *NanoEthics* **2024**, *18*, 11. [CrossRef]
219. Prather, J.; Leinonen, J.; Kiesler, N.; Gorson Benario, J.; Lau, S.; MacNeil, S.; Norouzi, N.; Opel, S.; Pettit, V.; Porter, L.; et al. Beyond the hype: A comprehensive review of current trends in generative AI research, teaching practices, and tools. In *2024 Working Group Reports on Innovation and Technology in Computer Science Education*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 300–338.
220. Narne, S.; Adedoja, T.; Mohan, M.; Ayyalasomayajula, T. AI-driven decision support systems in management: Enhancing strategic planning and execution. *Int. J. Recent Innov. Trends Comput. Commun.* **2024**, *12*, 268–276.
221. Yang, T.; Chang, L.; Yan, J.; Li, J.; Wang, Z.; Zhang, K. A survey on foundation-model-based industrial defect detection. *arXiv* **2025**, arXiv:2502.19106.

222. Awais, M.; Naseer, M.; Khan, S.; Anwer, R.M.; Cholakkal, H.; Shah, M.; Yang, M.H.; Khan, F.S. Foundation models defining a new era in vision: A survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 2245–2264. [CrossRef]
223. Tang, J.; Chen, J.; He, J.; Chen, F.; Lv, Z.; Han, G.; Liu, Z.; Yang, H.H.; Li, W. Towards General Industrial Intelligence: A Survey of Large Models as a Service in Industrial IoT. *IEEE Commun. Surv. Tutor.* **2025**, *28*, 2054–2086.
224. Baris, O.; Chen, Y.; Dong, G.; Han, L.; Kimura, T.; Quan, P.; Wang, R.; Wang, T.; Abdelzaher, T.; Bergés, M.; et al. Foundation models for cps-iot: Opportunities and challenges. *arXiv* **2025**, arXiv:2501.16368.
225. Xu, Q.; Stok, L.; Drechsler, R.; Wang, X.; Zhang, G.L.; Markov, I.L. Revolution or Hype? Seeking the Limits of Large Models in Hardware Design. *arXiv* **2025**, arXiv:2509.04905. [CrossRef]
226. Awaisi, K.S.; Ye, Q.; Sampalli, S. A survey of industrial AIoT: Opportunities, challenges, and directions. *IEEE Access* **2024**, *12*, 96946–96996. [CrossRef]
227. Schneider, J.; Meske, C.; Kuss, P. Foundation models: A new paradigm for artificial intelligence. *Bus. Inf. Syst. Eng.* **2024**, *66*, 221–231.
228. Chen, H.; Chen, H.; Zhao, Z.; Han, K.; Zhu, G.; Zhao, Y.; Du, Y.; Xu, W.; Shi, Q. An overview of domain-specific foundation model: Key technologies, applications and challenges. *Sci. China Inf. Sci.* **2026**, *69*, 111301.
229. Jangid, M.; Kumar, R. Deep learning approaches to address cold start and long tail challenges in recommendation systems: A systematic review. *Multimed. Tools Appl.* **2025**, *84*, 2293–2325.
230. Khalid, S.; Yazdani, M.H.; Azad, M.M.; Elahi, M.U.; Raouf, I.; Kim, H.S. Advancements in physics-informed neural networks for laminated composites: A comprehensive review. *Mathematics* **2024**, *13*, 17. [CrossRef]
231. Dashtbayaz, N.H.; Farhani, G.; Wang, B.; Ling, C.X. Physics-informed neural networks: Minimizing residual loss with wide networks and effective activations. *arXiv* **2024**, arXiv:2405.01680. [CrossRef]
232. Kasilingam, S.; Yang, R.; Singh, S.K.; Farahani, M.A.; Rai, R.; Wuest, T. Physics-based and data-driven hybrid modeling in manufacturing: A review. *Prod. Manuf. Res.* **2024**, *12*, 2305358.
233. Zhang, S.; Lomazzi, L.; Ma, D.; Manes, A. A data-driven hybrid method combining experiments, finite element modeling and machine learning for impact response prediction of TPU composites. *Int. J. Struct. Integr.* **2025**, 1–21. [CrossRef]
234. Khatir, A.; Capozucca, R.; Khatir, S.; Magagnini, E.; Le Thanh, C.; Riahi, M.K. Advancements and emerging trends in integrating machine learning and deep learning for SHM in mechanical and civil engineering: A comprehensive review. *J. Braz. Soc. Mech. Sci. Eng.* **2025**, *47*, 419. [CrossRef]
235. Khoei, T.T.; Singh, A. Data reduction in big data: A survey of methods, challenges and future directions. *Int. J. Data Sci. Anal.* **2025**, *20*, 1643–1682.
236. Sajadieh, S.M.M.; Noh, S.D. From Simulation to Autonomy: Reviews of the Integration of Artificial Intelligence and Digital Twins. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2025**, *12*, 1597–1628.
237. Kobayashi, K.; Alam, S.B. Explainable, interpretable, and trustworthy AI for an intelligent digital twin: A case study on remaining useful life. *Eng. Appl. Artif. Intell.* **2024**, *129*, 107620. [CrossRef]
238. Quarteroni, A.; Gervasio, P.; Regazzoni, F. Combining physics-based and data-driven models: Advancing the frontiers of research with scientific machine learning. *arXiv* **2025**, arXiv:2501.18708. [CrossRef]
239. Lazaros, K.; Koumadorakis, D.E.; Vrahatis, A.G.; Kotsiantis, S. Federated learning: Navigating the landscape of collaborative intelligence. *Electronics* **2024**, *13*, 4744. [CrossRef]
240. Alharbey, R.A.; Jamil, F. Federated learning framework for real-time activity and context monitoring using edge devices. *Sensors* **2025**, *25*, 1266. [CrossRef]
241. Vyas, A.; Lin, P.C.; Hwang, R.H.; Tripathi, M. Privacy-preserving federated learning for intrusion detection in IoT environments: A survey. *IEEE Access* **2024**, *12*, 127018–127050. [CrossRef]
242. Ahmad, I.; Rodriguez, F.; Kumar, T.; Suomalainen, J.; Jagatheesaperumal, S.K.; Walter, S.; Asghar, M.Z.; Li, G.; Papakonstantinou, N.; Ylianttila, M.; et al. Communications security in Industry X: A survey. *IEEE Open J. Commun. Soc.* **2024**, *5*, 982–1025. [CrossRef]
243. Shubyn, B.; Maksymyuk, T.; Gazda, J.; Rusyn, B.; Mrozek, D. Federated Learning: A Solution for Improving Anomaly Detection Accuracy of Autonomous Guided Vehicles in Smart Manufacturing. In *Proceedings of the IEEE International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 746–761.
244. Barona López, L.I.; Borja Saltos, T. Heterogeneity challenges of federated learning for future wireless communication networks. *J. Sens. Actuator Netw.* **2025**, *14*, 37. [CrossRef]
245. War, M.R.; Singh, Y.; Sheikh, Z.A.; Singh, P.K. Review on the Use of Federated Learning Models for the Security of Cyber-Physical Systems. *Scalable Comput. Pract. Exp.* **2025**, *26*, 16–33.
246. Cajas Ordóñez, S.A.; Samanta, J.; Suárez-Cetrulo, A.L.; Carbajo, R.S. Intelligent Edge Computing and Machine Learning: A Survey of Optimization and Applications. *Future Internet* **2025**, *17*, 417. [CrossRef]

247. Tariq, A.; Serhani, M.A.; Sallabi, F.M.; Barka, E.S.; Qayyum, T.; Khater, H.M.; Shuaib, K.A. Trustworthy federated learning: A comprehensive review, architecture, key challenges, and future research prospects. *IEEE Open J. Commun. Soc.* **2024**, *5*, 4920–4998. [CrossRef]
248. Ji, S.; Tan, Y.; Saravirta, T.; Yang, Z.; Liu, Y.; Vasankari, L.; Pan, S.; Long, G.; Walid, A. Emerging trends in federated learning: From model fusion to federated x learning. *Int. J. Mach. Learn. Cybern.* **2024**, *15*, 3769–3790. [CrossRef]
249. Pandhare, H.V. Future of Software Test Automation Using AI/ML. *Int. J. Eng. Comput. Sci.* **2025**, *13*, 27159–27182. [CrossRef]
250. Trigkas, A.; Piromalis, D.; Papageorgas, P. Edge Intelligence in Urban Landscapes: Reviewing TinyML Applications for Connected and Sustainable Smart Cities. *Electronics* **2025**, *14*, 2890. [CrossRef]
251. Bahangulu, J.K.; Owusu-Berko, L. Algorithmic bias, data ethics, and governance: Ensuring fairness, transparency and compliance in AI-powered business analytics applications. *World J. Adv. Res. Rev.* **2025**, *25*, 1746–1763. [CrossRef]
252. Alshkeili, H.M.H.A.; Almheiri, S.J.; Khan, M.A. Privacy-Preserving Interpretability: An Explainable Federated Learning Model for Predictive Maintenance in Sustainable Manufacturing and Industry 4.0. *AI* **2025**, *6*, 117. [CrossRef]
253. Cirelli, M.; Canonico, P.; Cellupica, A.; Valentini, P.P. Reduced-order models and augmented reality for real-time interactive structural digital twin exploration and interrogation. *Int. J. Interact. Des. Manuf. (IJIDeM)* **2025**, *19*, 7263–7281.
254. Sajjadi, P.; Dinmohammadi, F.; Shafiee, M. Fault Detection of Cyber-Physical Systems Using a Transfer Learning Method Based on Pre-Trained Transformers. *Sensors* **2025**, *25*, 4164. [CrossRef] [PubMed]
255. Baqar, M.; Khanda, R. The Future of Software Testing: AI–Powered Test Case Generation and Validation. In *Proceedings of the Intelligent Computing-Proceedings of the Computing Conference*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 276–300.
256. Zong, Z.; Guan, Y. AI-driven intelligent data analytics and predictive analysis in Industry 4.0: Transforming knowledge, innovation, and efficiency. *J. Knowl. Econ.* **2025**, *16*, 864–903. [CrossRef]

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