frontiers | Frontiers in Energy Research

TYPE Review
PUBLISHED 19 March 2025
DOI 10.3389/fenrg.2025.1531655

Check for updates

OPEN ACCESS

EDITED BY
Nicholas Kolokotronis,
University of Peloponnese, Greece

REVIEWED BY
Panagiotis Radoglou-Grammatikis,
K3Y, Bulgaria
Shampa Banik,
Tennessee Technological University,
United States

*CORRESPONDENCE
Rasha Kashef,
rkashef@torontomu.ca

RECEIVED 20 November 2024
ACCEPTED 25 February 2025
PUBLISHED 19 March 2025

CITATION
Ibrahim N and Kashef R (2025) Exploring the emerging role of large language models in smart grid cybersecurity: a survey of attacks, detection mechanisms, and mitigation strategies.
Front. Energy Res. 13:1531655.
doi: 10.3389/fenrg.2025.1531655

COPYRIGHT
© 2025 Ibrahim and Kashef. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Exploring the emerging role of large language models in smart grid cybersecurity: a survey of attacks, detection mechanisms, and mitigation strategies

Nourhan Ibrahim[1,2] and Rasha Kashef[1]*

[1]Electrical, Computer, and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON, Canada, [2]Faculty of Engineering, Alexandria University, Alexandria, Egypt

Smart grids are modernizing the future of providing energy for everyone, allowing us to increase the efficiency of power generation, transmission, or distribution using information and communication technologies. However, the network structure of smart grids makes them vulnerable to varying levels of cyber threats. This paper provides a broad overview of cyber threats against smart grids, considering attack surfaces, communication network layers, and the core security triad of confidentiality, integrity, and availability. This survey also outlines emerging threats and covers current protection, prevention, detection, mitigation, and recovery measures, focusing on emerging technologies such as artificial intelligence and large language models (LLMs) in smart grid security. We analyze and show how previous work has tackled and approached similar themes in this area. Amongst our contributions are categorizing the critical parts of smart grids that are most vulnerable to attack, several threat taxonomies, and a review of the increasing importance of LLMs for enhancing grid security. This evaluation underscores the need for effective and robust security technologies to avoid the compromises that result from more sophisticated cyber attacks.

KEYWORDS

cybersecurity, cyber attacks, intrusion detection, smart grids, large language models (LLMs), deep learning (DL), machine learning (ML), and the internet of things (IoT)

1 Introduction

Smart grids combine information and communication technologies (ICT) to provide an efficient, reliable, and sustainable electric energy service, aiming for greater systemic or multisectoral decarbonization (Ghiasi et al., 2023). This increased interconnection and dependence on digital systems in smart grids expand the attack surface for cyber threats. Smart grid cybersecurity serves as the front-line defense against potential denial-of-service attacks, data integrity breaches, and unauthorized control interventions (El Mrabet et al., 2018).

The integration of artificial intelligence (AI), machine learning (ML), and deep learning (DL) approaches into smart grid cybersecurity has garnered significant attention. AI-powered solutions are easily scalable and can help organizations detect,

analyze, and respond to cyber threats (Gunduz and Das, 2020). ML approaches have been applied to intrusion detection and anomaly detection, enabling systems to recognize patterns that may indicate cyber attacks when anomalies occur (Berghout et al., 2022). DL models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), utilize layer-wise feature extraction from high-dimensional data to achieve advanced detection accuracy in complex, large-scale data landscapes (Ruan et al., 2023).

However, three main limitations highlight the need for more flexible and contextual cyber threat detection solutions. The first is the dependence on extensively labeled datasets for training, which are often sparse in specific disciplines like smart grid cybersecurity due to the absence of long-standing records detailing cyber attacks. The second is the inability of current models to adapt dynamically to the evolving trends of vulnerabilities exploited by threat actors. The third is their "black box" nature and resource-intensive operation, generating results with limited explainability (Divakaran and Peddinti, 2024).

There is a growing need for more adaptive, interpretable, and data-efficient approaches in cybersecurity, which large language models (LLMs) can address. LLMs have potential applications in cybersecurity-related tasks such as anomaly detection, vulnerability assessment, and log analysis (Ferrag et al., 2024), helping to detect complex attack patterns and enable automated decision-making by processing vast and heterogeneous information. However, the applicability of LLMs in smart grid cybersecurity remains largely unexplored, presenting a research gap.

One major challenge posed by LLMs is their propensity for producing "hallucinations." Therefore, it is crucial to address these inaccuracies before LLMs can be reliably deployed in critical infrastructure scenarios such as smart grids (Li et al., 2024b). To mitigate such issues, several methods have been proposed, including fine-tuning LLMs on domain-specific data (Liu et al., 2024), employing reinforcement learning from human feedback (RLHF) (Liu et al., 2024), utilizing retrieval-augmented generation (RAG), leveraging Knowledge Graphs (KGs) (Ibrahim et al., 2024), and incorporating verification layers that cross-check model outputs against trusted data sources. This paper aims to explore existing research on integrating LLMs with smart grid cybersecurity and identify gaps for further investigation.

1.1 Related surveys

Several surveys have examined the cybersecurity landscape of smart grids in recent years, each contributing unique perspectives. One notable review paper (Gunduz and Das, 2020) focuses on the cybersecurity of Internet of Things (IoT)-enabled smart grids, categorizing cyber attacks by their impact on confidentiality, integrity, and availability. It identifies IoT-based communication systems as both beneficial and vulnerable, particularly concerning critical infrastructure threats. The survey categorizes attack types, assesses network vulnerabilities, and evaluates defense strategies. It presents IoT-driven cybersecurity solutions and outlines future research directions, emphasizing the importance of protecting smart grids as crucial infrastructure for national security.

The review in Ding et al. (2022) addresses cyber threats in smart grids by analyzing hardware, software, and data communication vulnerabilities. It categorizes attacks and highlights potential solutions, particularly blockchain and AI techniques. The paper examines historical cyber attacks, such as ransomware and SCADA (Supervisory Control and Data Acquisition) system breaches, and advocates for advanced detection and response measures. Future directions focus on addressing protection gaps for evolving grid complexities and integrating distributed energy sources.

The survey in Tala et al. (2022) investigates cyber attacks on smart grids, focusing on open system interconnection (OSI) model layers and categorizing attacks by their impact on network security. It proposes a classification for detection and countermeasures, addressing attacks across different layers of the communication model. The study emphasizes confidentiality, integrity, availability, and accountability. It reviews techniques such as ML and cryptographic methods for mitigating cyber threats and discusses open challenges, including detection and defense strategies tailored to smart grids.

The work in Tatipatri and Arun (2024) explores cyber attacks on power systems, focusing on impact, detection, and mitigation methods. It examines IoT and machine-to-machine communications within smart grids, emphasizing security vulnerabilities in data transmission and IoT components. The review covers cryptographic solutions, blockchain, and artificial intelligence for securing communication channels. Key contributions include insights into real-world cyber incidents and the economic impacts of attacks on deregulated energy markets, as well as specific recommendations for improving grid resilience. This paper builds on these findings by exploring the integration of LLMs and advanced ML techniques, which offer promising capabilities for adaptive cybersecurity in smart grids. Table 1 summarizes the related surveys in the literature.

1.2 Research contributions

In this paper, the following contributions to the study of cybersecurity in smart grids, with a particular emphasis on the potential and challenges of LLMs, are provided:

- This paper provides an overview of smart grid architectures and their unique cybersecurity needs, focusing on their layered structures and specific security requirements for maintaining grid resilience against cyber threats.
- It categorizes and analyzes various cyber threats relevant to smart grids using multiple taxonomies, focusing on vectors, target layers, and CIA principles.
- The paper also investigates current cybersecurity techniques for smart grids, focusing on protection, prevention, detection, mitigation, and recovery mechanisms, with a particular emphasis on machine learning-based models.
- The paper examines and analyzes the role of LLMs in cybersecurity, surveying current literature on their general applications and potential benefits for smart grid security.
- The paper identifies challenges in applying LLMs to smart grid security, including issues related to model reliability, data

TABLE 1 Comparison of cybersecurity surveys for smart grids.

Criteria	Gunduz and Das (2020)	Ding et al. (2022)	Tala et al. (2022)	Tatipatri and Arun (2024)
Focus	IoT-enabled smart grid vulnerabilities, attack types, and defenses	Threat taxonomy, blockchain, and AI-based defenses	Cyber-attack classification on OSI layers and detection techniques	Cyber-attack impact, detection, and mitigation in power systems
Attack Taxonomy	CIA triad categorization (Confidentiality, Integrity, Availability)	Taxonomy by system vulnerabilities and threat types	Categorized by OSI model layers	Focus on distribution-level attacks, emphasizing real-world cases
Emerging Techniques	IoT-based security solutions, cryptography	Blockchain, AI, ML	ML, cryptography, accountability focus	Blockchain, cryptography, IoT-based secure communication
Challenges Identified	IoT vulnerabilities in public communication networks	Distributed energy resource vulnerabilities, SCADA system risks	OSI-layer accountability, advanced threat detection	IoT dependency risks, secure transmission and data integrity
Future Directions	Enhanced IoT security measures, multi-layer defenses	Blockchain integration, adaptive AI strategies	Detailed OSI layer defenses, accountability in smart grids	Real-world case applications, cryptographic enhancements
Unique Contribution	IoT vulnerabilities and solutions, CIA-based taxonomy	Blockchain and AI solutions with a focus on SCADA	OSI model-driven attack taxonomy and accountability	Emphasis on economic impacts, cryptographic methods

integrity, interpretability, and the risk of adversarial attacks like data poisoning and "hallucination."
- Based on our analysis, we suggest future directions for research and development in smart grid cybersecurity.

1.3 Organization

The rest of the paper is organized as follows: Section 2 provides a detailed overview of the smart grid architecture, including key components and cyber-physical interfaces. Section 3 categorizes cybersecurity threats. Section 4 reviews current cybersecurity techniques and discusses their limitations. Section 5 focuses on the potential of LLMs in grid security, pointing out their advantages and limitations. Section 6 provides summaries of future research directions. Finally, Section 8 concludes our findings.

2 Smart grid overview

In this section, we delve into the architecture and the fundamental components of the smart grid, providing an understanding of its cybersecurity requirements. This includes examining the main hardware systems that support key grid operations and outlining the architecture model proposed by the National Institute of Standards and Technology (NIST). Furthermore, the communication networks, protocols, and technologies integral to smart grid operations will be discussed in terms of their vulnerabilities and the specific cybersecurity challenges posed by the grid's multiple networking and distribution mechanisms.

2.1 Smart grid architectures

The advent of smart grids has enabled instantaneous communication between different entities in the grid, allowing demand management, local generation, and monitoring, among other functionalities. This transformation has become crucial to meeting the growing electricity demand, integrating clean energy sources, and addressing challenges in the current energy market (Yadav et al., 2016). To ensure effective and efficient smart grid operations, there needs to be a well-defined architecture focusing on standards, data exchange, and security aspects. The smart grid architecture consists of three interconnected layers: the physical layer, the cyber-physical layer, and the cyber layer, as illustrated in Figure 1.

The physical layer includes core infrastructure for energy generation, transmission, and distribution, such as power plants, substations, and distribution networks. Although essential for electricity delivery, this infrastructure relies heavily on control systems, making it vulnerable to cyberattacks that could disrupt power flows or destabilize operations, particularly through distributed energy resources (DERs).

The cyber-physical layer bridges the physical infrastructure and digital control systems, incorporating sensors, actuators, phasor measurement units (PMUs), and smart meters to provide real-time data for automated decision-making. However, these interconnected devices increase the vulnerability to data falsification and injection attacks, which can mislead grid operators and disrupt stability. The extensive deployment of IoT in this layer broadens the attack surface, as unsecured devices can serve as entry points for attackers.

The cyber layer serves as the digital backbone, consisting of SCADA, energy management systems (EMS), distribution management systems (DMS), and communication protocols such as IEC 61850 and DNP3 (Distributed Network Protocol 3) that

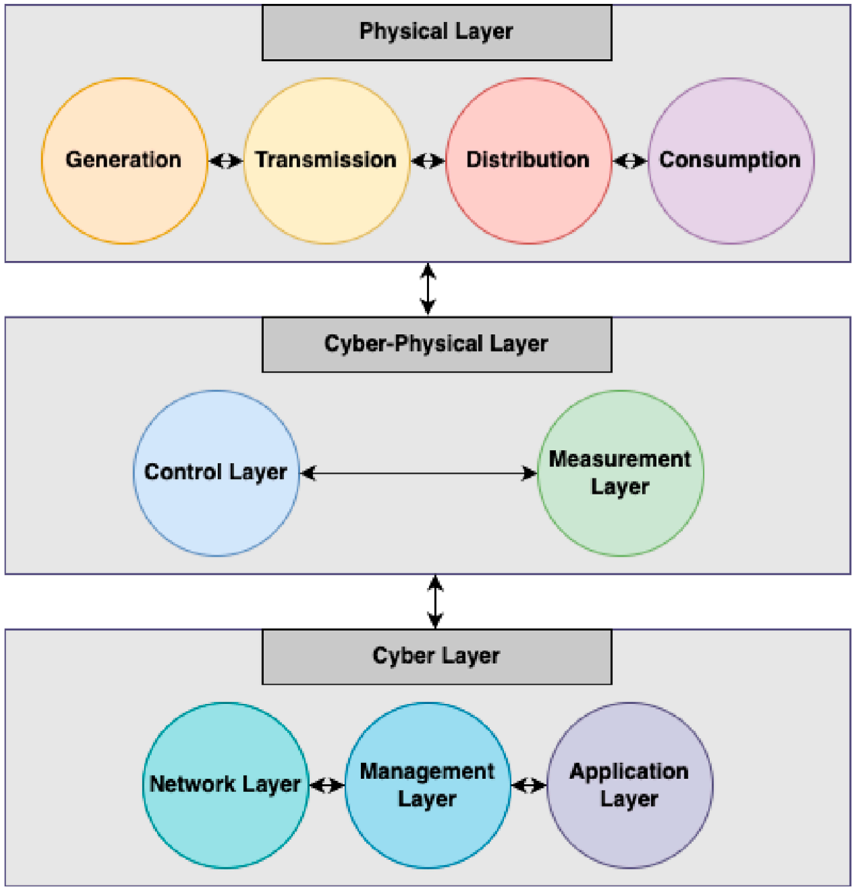

FIGURE 1
Smart grid layered architecture.

facilitate data exchange between field devices and control centers. The reliance of this layer on network communication makes it susceptible to man-in-the-middle attacks, data breaches, and ransomware, allowing attackers to intercept, alter, or inject malicious commands into critical systems. Together, these layers enable smart grid functionality but necessitate robust, multi-layered defenses to mitigate sophisticated cyber threats.

Additionally, a widely recognized model by NIST partitions the smart grid into seven broad domains, which define its core functionalities: bulk generation, transmission, distribution, customers, markets, service providers, operations, and distribution control (Gopstein et al., 2021). This model helps classify the grid's vast interconnectedness while emphasizing that communication between devices and across domains must be secure and seamless. Figure 2 depicts the seven domains of the smart grid as proposed by NIST.

The cybersecurity needs of the smart grid are complex and multifaceted. Demand response (DR) is a key technology that helps manage electricity demand by requesting consumers to reduce consumption during peak periods. However, it also introduces risks, as attackers can inject false signals into the DR system, potentially destabilizing the electricity grid (Gunduz and Das, 2020; Ding et al., 2022). DERs, such as solar and wind-based energy sources, diversify non-centralized energy generation and strengthen the grid. However, cyberattacks targeting DERs can harm grid integrity, induce instability, and disrupt power distribution (Liu et al., 2023).

Smart grids have transformed power utilities by improving disaster recovery, enabling cost-effective energy consumption, and reducing power outages. Smart meters and advanced metering infrastructure (AMI) facilitate detailed data exchange between suppliers and consumers, but incorrect information or unauthorized access can pose threats to grid integrity and consumer data privacy (Achaal et al., 2024). SCADA-based grid framework systems, which integrate remote terminal units (RTUs) and PMUs, face significant threats from cyberattacks, as they play a crucial role in maintaining grid stability (Tala et al., 2022).

EMS and DMS are essential for controlling transmission and distribution networks, managing energy flow, identifying faults, and minimizing energy wastage. Attacks on these systems could lead to resource mismanagement, power outages, and system destabilization (Ding et al., 2022; Tatipatri and Arun, 2024).

Despite the increasing demand for IoT-supported functionalities in smart homes and grids, low security awareness presents a significant challenge. The growing number of integrated devices provides attackers with more opportunities to exploit control systems, making cyber warfare an emerging threat for smart energy

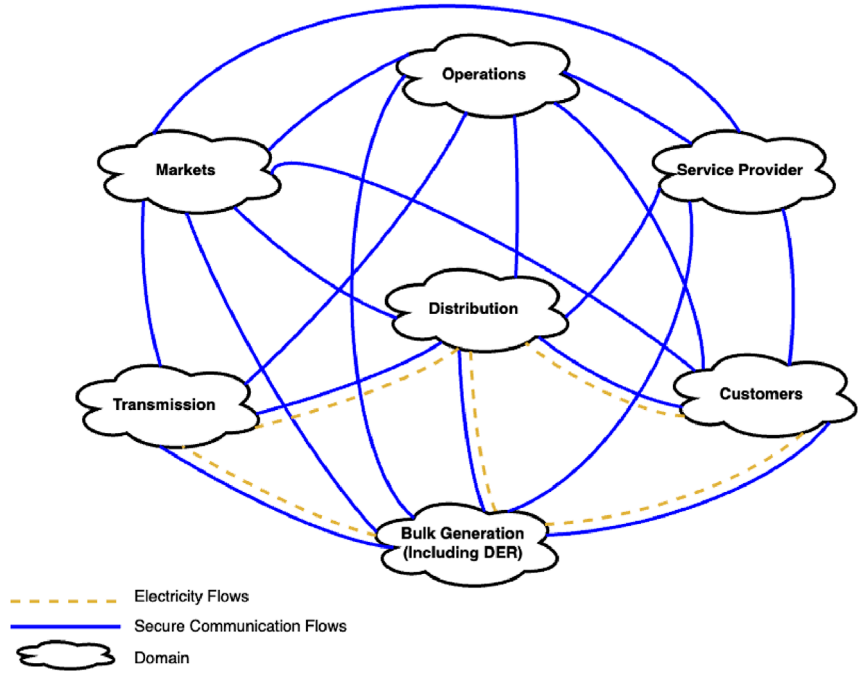

FIGURE 2
NIST smart grid model (Gopstein et al., 2021).

grids. Implementing robust control systems and security measures will enhance the ability to remotely monitor and manage critical energy sector infrastructures (Singhal et al., 2021).

2.2 Communication networks and protocols in smart grids

Communication networks are vital for integrating smart grid components, enabling data transmission, remote control, and real-time monitoring. However, each network type presents unique security challenges that must be addressed.

Wide Area Networks (WANs) connect central management systems with DERs and substations, ensuring rapid data transfer but posing risks of interception (Elkhorchani et al., 2013). Field Area Networks (FANs) link field devices like sensors and meters for local monitoring but are vulnerable to interference and unauthorized access (Budka et al., 2014).

Local Area Networks (LANs) in substations connect SCADA system RTUs and other control devices but face risks of cyber and physical attacks, potentially impacting grid stability (Tala et al., 2022). Home Area Networks (HANs) connect consumer devices for DR and energy management but are susceptible to eavesdropping and jamming due to weak encryption (Xiaocheng et al., 2023). Neighborhood Area Networks (NANs) aggregate multiple HANs for improved system visibility but require strong security measures (Noorwali et al., 2016).

Smart grid communication relies on various protocols with distinct security challenges. DNP3 lacks encryption, making it vulnerable to interception (Padilla et al., 2014). IEC 61850 ensures secure substation communication but has implementation inconsistencies (Kush et al., 2010). Modbus lacks built-in security, making it susceptible to data injection and man-in-the-middle attacks (Kush et al., 2010). Zigbee, used for low-power wireless IoT devices, is prone to signal spoofing and jamming (Elkhorchani et al., 2013). MQTT, widely used in IoT applications, requires additional encryption and access control (Salvadori et al., 2013).

Both wired and wireless technologies are used for data transmission. Power line communication (PLC) and fiber optics offer secure, high-speed transmission but can be compromised (Salvadori et al., 2013). Wireless technologies like LTE/3G, WiMAX, and WiFi facilitate communication but are vulnerable to interception and interference (Elkhorchani et al., 2013).

To protect the smart grid from cyber threats, robust encryption, authentication, and continuous monitoring are essential, ensuring secure communication and resilience against attacks.

2.3 Cybersecurity requirements and challenges for smart grids

Smart grids are data-centric systems that require sophisticated cybersecurity strategies to protect against cyber warfare. These strategies include the triangle of CIA, accountability, safety, and resilience (Arpilleda, 2023). Confidentiality is crucial to prevent unauthorized access to sensitive data, while integrity ensures data accuracy and unmodifiedness, preventing false data injection attacks and disrupting grid operations. Redundancy planning, load balancing, and DoS detection are widely applied to ensure availability (Gunduz and Das, 2020). Authentication and authorization processes are essential for user identity and access control, with multi-factor authentication and role-based

access control being helpful security improvements (Gunduz and Das, 2020).

Accountability is necessary to track activities within the smart grid, allowing for detailed investigations following security events. Safety and criticality aim to prevent physical damage or widespread disruption, as the grid infrastructure supports public welfare. Systems must be designed to withstand catastrophic events from external forces such as cyber and natural attacks (Gunduz and Das, 2020).

Smart grids must also be resilient and survivable, enabling them to resist and recover from cyber attacks. Resilience involves developing systems that can quickly adapt to and reduce interruptions, while survivability ensures the system continues functioning even in damaged conditions. Identification of affected zones, redundancy, and other multilayered defense methods increase the grid's resilience and survivability (Tala et al., 2022).

The distributed and interconnected characteristics of smart grids present new cybersecurity threats, with interoperability between devices and legacy systems and infrastructure being significant challenges (Arpilleda, 2023). To safeguard smart grids, a holistic strategy beyond traditional security procedures is needed, including complex and multi-layered defense systems, continuous surveillance, and interaction between relevant parties (Ghiasi et al., 2023).

3 Cyber threats in smart grid

The variety of interrelated systems that make up smart grids subject them to a wide spectrum of cyber threats targeting multiple elements and data flows. Cyber threats in smart grids have been categorized based on several criteria, including security objectives that the threat impacts, layers that the threat targets, domains affected by the threat, attack vectors, and the level of expertise required to deploy an attack. The following subsections will also cover these classifications, their principal attacks, and the problems encountered in their detection and elimination.

3.1 Classification of cyber threats in smart grids

This subsection investigates different classifications of cyber threats in smart grids, as shown in Figure 3.

- Classification Based on the CIA Triad: Cyber threats to smart grids are typically categorized according to their impact on the CIA Triad (Confidentiality, Integrity, and Availability) (Gajanan and Kirar, 2022). This aspect of security encompasses various dimensions, each categorized under the subsequent areas:
- Confidentiality: A major objective of attackers is to obtain sensitive data, such as consumer usage statistics and configurations of the grid. Data breaches and eavesdropping are among the most common threats against confidentiality, which can lead to privacy violation or disclosure of operational data (Gajanan and Kirar, 2022; Qureshi et al., 2023).
- Integrity: Integrity threats can modify data or disrupt the accurate flow of data within the grid, leading to erroneous operational decisions. For example, false data injection (FDI) attacks can mislead SCADA systems or demand forecasting models by injecting false data to destabilize the operations of the grid (Gajanan and Kirar, 2022; Qureshi et al., 2023).
- Availability: Availability-related threats are designed to disrupt the use of grid services. Common examples include denial of service (DoS) and distributed denial of service (DDoS) attacks, which flood systems with traffic and can result in outages (Gajanan and Kirar, 2022; Qureshi et al., 2023).

Table 2 summarizes the CIA triad attacks.

- Classification Based on Layers of Attack: Smart grid threats can also be categorized by their target layers: cyber layer or cyber-physical layer, as summarized in Table 3.
- Cyber Layer Attacks: These attacks target IT infrastructure, such as communication networks, data systems, and software applications. Phishing, malware, and SQL injection attacks compromise the digital components of the grid, potentially giving attackers access to control systems or sensitive information (Simonthomas et al., 2024).
- Cyber-Physical Layer Attacks: These attacks focus on the elements that connect the digital and physical realms, such as SCADA systems and distributed energy resources. For instance, if power generation becomes unstable because of the SCADA system or any control device, there is a significant likelihood of other devices malfunctioning or equipment being disrupted (Simonthomas et al., 2024).
- Classification Based on NIST Model Domains: According to the NIST model, the smart grid comprises seven domains: generation, transmission, distribution, customers, markets, service providers, and operations. Each domain has specific functions and is vulnerable to various cyber threats (Ding et al., 2022). Table 4 summarizes these attacks.
- Generation: Cyber attacks on power generation, especially renewables, disrupt energy output by tampering with control signals, affecting supply-demand balance and damaging assets (Ding et al., 2022; Tala et al., 2022).
- Transmission: High-voltage networks, including substations and PMUs, face *MITM* and *DoS* attacks, leading to grid instability and potential blackouts (Ding et al., 2022; Tala et al., 2022).
- Distribution: Attacks on smart transformers and SCADA systems disrupt energy delivery, increasing outage risks. Unauthorized access to control devices can trigger malfunctions across connected equipment (Ding et al., 2022; Tala et al., 2022).
- Customer: Smart meter tampering and intrusions into Home Area Networks (HANs) compromise billing accuracy and user privacy, potentially exposing in-home devices to attackers (Ding et al., 2022; Tala et al., 2022).

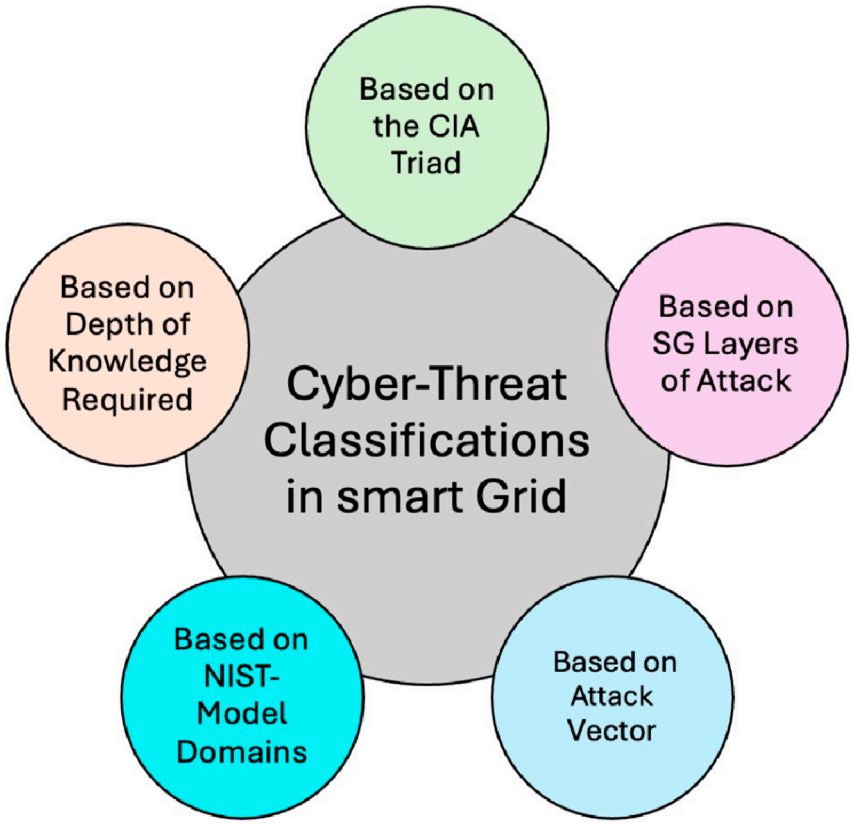

FIGURE 3
Classification of cyber threats in smart grids.

- Markets: Cyber threats in energy trading manipulate real-time prices and disrupt demand-response programs, leading to financial losses and market instability (Ding et al., 2022; Tala et al., 2022).
- Service Providers: Supply chain attacks exploit third-party relationships to inject malware or breach defenses, making this domain highly vulnerable to system-wide compromises (Ding et al., 2022; Tala et al., 2022).
- Operations: Attacks on real-time management systems like SCADA and EMS (e.g., ransomware) can shut down grid operations, causing load imbalances and widespread blackouts (Ding et al., 2022; Tala et al., 2022).
- Classification Based on Attack Vectors: Cyber threats in smart grids can also be categorized by their attack vectors, each exploiting unique vulnerabilities (Uma and Padmavathi, 2013). Table 5 summarizes these vectors.
- Network-Based Attacks: Exploit weaknesses in communication protocols, overwhelming channels with DoS/DDoS or intercepting messages via MITM attacks. These threats disrupt real-time grid communication, compromising stability (Uma and Padmavathi, 2013; Singh et al., 2016).
- Data-Based Attacks: Manipulate grid information to distort decision-making. Examples include False Data Injection (FDIAs), AI data poisoning, and protocol corruption, making detection difficult due to subtle alterations (Uma and Padmavathi, 2013; Singh et al., 2016).
- User-Based Attacks: Leverage social engineering, phishing, and credential theft to gain unauthorized access. Attackers use compromised credentials for long-term infiltration (e.g., APT campaigns), enabling further network or data breaches (Uma and Padmavathi, 2013; Singh et al., 2016).
- Classification Based on Depth of Knowledge Required: Smart grid threats can also be classified by the expertise needed to execute them, as summarized in Table 6. This classification highlights the accessibility of different attack types to various threat actors (Qureshi et al., 2023).
- Low Knowledge (Opportunistic Attacks): Require minimal expertise, often leveraging free tools for phishing or malware injection without smart grid-specific knowledge (Ding et al., 2022; Qureshi et al., 2023).
- Moderate Knowledge (Intermediate Attacks): Demand familiarity with smart grid systems, control protocols, and software vulnerabilities. Attacks on smart meters and HANs involve data manipulation and basic programming skills (Ding et al., 2022; Qureshi et al., 2023).
- High Knowledge (Advanced Attacks): Require deep understanding of control protocols, network configurations, and grid operations. APTs and coordinated attacks, often state-sponsored, involve prolonged infiltration and complex evasion tactics (Ding et al., 2022; Qureshi et al., 2023).

TABLE 2 Classification of cyber threats based on the CIA triad.

Classification	Description	Example attacks
Confidentiality	Protects access to sensitive data, such as user data and grid configuration	Data breaches, eavesdropping
Integrity	Ensures data accuracy and protects against unauthorized alterations	FDI, data tampering
Availability	Ensures system and data availability to authorized users	DoS, DDoS, service disruption

TABLE 3 Classification of cyber threats based on layers of attack.

Layer targeted	Description	Example attacks
Cyber Layer	Focuses on IT infrastructure, networks, and software systems	Phishing, malware, SQL injection
Cyber-Physical Layer	Targets the integration between digital systems and physical control components	SCADA tampering, remote device manipulation

TABLE 4 Cyber threats across the NIST smart grid domains.

Domain	Description	Example attacks
Generation	Targets generation facilities, including renewable sources, to disrupt energy output or stability	Manipulation of generation settings, command injection
Transmission	Involves high-voltage networks like substations, where attacks can disrupt large-scale distribution	man-in-the-middle (MITM) attacks, denial of service (DoS)
Distribution	Focuses on delivering electricity to consumers; attacks here cause localized disruptions	tampering with smart transformers, load misconfigurations
Customer	Targets consumer devices such as smart meters and HANs, risking privacy and billing accuracy	Data tampering in smart meters, unauthorized HAN access
Markets	Manipulates energy pricing and trading mechanisms, impacting financial stability and supply-demand balance	Price manipulation, unauthorized access to market data
Service Providers	Exploits third-party services to introduce malware or breach data, posing supply chain risks	Supply chain attacks, third-party data breaches
Operations	Attacks on real-time management systems like SCADA and EMS can disrupt grid functionality	Ransomware, SCADA tampering, and unauthorized control

3.2 Cyber threats in smart grids

Smart grids face numerous cyber threats across three key layers: physical, cyber-physical, and cyber. These attacks disrupt operations, steal sensitive data, or manipulate grid information, jeopardizing integrity and security. Table 7 summarizes key attacks and their properties.

- False Data Injection (FDI) Attacks: Attackers manipulate control system data, leading to incorrect decisions and grid instability, targeting the cyber layer (Khare et al., 2023; Jin, 2024).
- Denial of Service (DoS/DDoS) Attacks: Overwhelm control systems with traffic, rendering services unavailable, affecting the cyber layer (Naqvi et al., 2024).
- Man-in-the-Middle (MITM) Attacks: Intercept and modify communication, compromising data integrity and confidentiality in the cyber layer (Tala et al., 2022; Gao et al., 2023).
- Smart Meter Tampering: Unauthorized modification of meters to alter consumption data, breaching integrity in the physical layer (Tala et al., 2022;Ilokanuno, 2024).
- Puppet Attacks: Use infected devices as proxies to perform hidden actions, targeting the cyber-physical layer (Tala et al., 2022; Yi et al., 2014).
- Message Replay Attacks: Replay valid messages to mislead systems, threatening data integrity in the cyber layer (Sriranjani et al., 2023).
- Masquerade Attacks: Attackers impersonate legitimate devices, compromising credibility in the cyber layer (Tala et al., 2022; Tatipatri and Arun, 2024).

TABLE 5 Comparison of attack vectors in smart grids.

Attack vector	Description	Common examples
Network-Based	Exploits vulnerabilities in communication protocols and network connections, disrupting real-time communication across grid systems	DoS/DDoS attacks, MITM attacks
Data-Based	Focuses on compromising data integrity, affecting grid decision-making processes by altering critical information	FDIAs, data poisoning for AI models, manipulation of control signals
User-Based	Involves social engineering and credential exploitation to gain unauthorized access to control systems, often as the initial step in broader attack campaigns	Phishing, social engineering, exploitation of user credentials

TABLE 6 Classification of cyber threats based on depth of knowledge required.

Depth of knowledge	Description	Example attacks
Low Knowledge	Opportunistic attacks using basic tools and minimal understanding of grid operations	Phishing, basic malware
Moderate Knowledge	Intermediate attacks requiring some knowledge of smart grid operations	FDI, data tampering
High Knowledge	Advanced attacks requiring in-depth expertise, often from skilled or organized adversaries	APTs, coordinated DoS with FDI

- Eavesdropping & Traffic Analysis: Monitor network traffic to collect sensitive data, breaching confidentiality in the cyber layer (Pandey and Kalra, 2022).
- Advanced Persistent Threats (APTs): Long-term infiltration for observation and control, compromising the cyber-physical layer (e.g., Stuxnet, Industroyer) (Chen et al., 2014).
- Adversarial Machine Learning Attacks: Manipulate ML models used in anomaly detection, affecting the cyber layer's integrity (Zhang et al., 2024).
- Coordinated Attacks: Simultaneous or sequential multi-layer attacks (e.g., combining DoS with FDI) amplify impact across all tiers (Ding et al., 2022).

These evolving cyber threats emphasize the need for robust, multi-layered security strategies to protect smart grids from sophisticated adversarial techniques.

3.3 Complexity of attack detection and mitigation

Several critical cyber incidents have exposed vulnerabilities in smart grid infrastructure. The 2010 Stuxnet malware (Baezner and Robin, 2017) demonstrated the potential for malware to disrupt critical systems. The 2015 Ukraine power grid attack, using BlackEnergy malware, leveraged spear phishing to infiltrate SCADA systems, causing a blackout for 230,000 people (Alert, 2016). In 2016, Industroyer malware exploited industrial protocols to shut down substations (Geiger et al., 2020). The 2017 Triton/Trisis malware (Giles, 2019) targeted safety systems, while a 2019 DoS attack exposed renewable energy vulnerabilities (Walton, 2019). The 2021 Colonial Pipeline ransomware attack (Easterly and Fanning, 2023) further highlighted threats to critical infrastructure. These incidents underscore the evolving cyber threats to smart grids and the need for proactive, multi-layered defenses.

4 Cybersecurity techniques in smart grid

Protection of smart grids from cyber threats and attacks requires a multi-stage strategy. Protecting the assets against vulnerabilities, preventing attacks that exploit vulnerabilities, detecting intrusion in real-time as well as reducing or recovering the impact of an attack are some ways this could be achieved.

4.1 Protection and prevention techniques

Smart grid cybersecurity consists of protection and prevention techniques that aim to harden the infrastructure to possible threats before they occur. This subsection will examine different protection and prevention techniques.

The authors in Vidya et al. (2023) presented a smart grid analyzer that simulates and analyzes the effects of cyber attacks on components like PMUs and IP camera sensors, identifying vulnerabilities in the smart grid. It used various security measures such as attack trees, attack graph generation, attack success probability, attack cost, and attack impact to assess the vulnerabilities and risks of cyber attacks. It also employed a common vulnerability score system (CVSS) to rank vulnerability severity with PMUs as the most sensitive components of power grid stability. The study combined the graphical security model with different metrics to survey the attacks in detail, thereby increasing knowledge about possible attack channels and informing IT specialists' preparation

TABLE 7 Summary of cyber attacks on smart grids.

Attack name	Objective/Purpose	Layer	NIST model domain	Security requirement affected	Attack vector	Level of knowledge
False Data Injection (FDI)	Corrupt data for grid miscalculation	Cyber	Operations	Integrity	Data-based	Moderate
Denial of service (DoS)	Prevent access to grid resources	Cyber	Operations	Availability	Network-based	Moderate
Man-in-the-Middle (MITM)	Intercept and modify communications	Cyber	Operations	Confidentiality, Integrity	Network-based	High
Smart Meter Tampering	Manipulate meter readings	Physical	Customer, Distribution	Integrity	Data-based	Low
Puppet Attack	Remote control of compromised device	Cyber-Physical	Operations	Integrity, Confidentiality	User-based	High
Message Replay	Repeat valid messages to mislead system	Cyber	Operations, Distribution	Integrity, Availability	Data-based	Moderate
Adversarial ML Attack	Manipulate ML models for incorrect predictions	Cyber	Operations	Integrity, Availability	Data-based	High
Coordinated Attack	Amplify impact with multiple simultaneous attacks	Physical, Cyber, Cyber-Physical	Multiple domains	Multiple (Conf., Int., Avail.)	Multiple	High
Masquerade Attack	Impersonate devices to mislead system	Cyber	Customer, Distribution	Integrity, Confidentiality	User-based	Moderate
Advanced Persistent Threats	Long-term monitoring and manipulation	Cyber-Physical	Operations	Confidentiality, Integrity	Network-based	High
Eavesdropping	Monitor traffic to steal data	Cyber	Customer, Operations	Confidentiality	Network-based	Low
Spoofing	Impersonate devices to mislead system	Cyber	Customer, Distribution	Integrity, Availability	User-based	Moderate
Buffer Overflow	Overload buffer to disrupt systems	Cyber	Generation, Operations	Availability	Data-based	Low
Time Synchronization	Desynchronize time-sensitive devices (PMUs)	Physical	Transmission	Integrity	Network-based	High
Malware/Ransomware	Disrupt or lock systems, data theft	Cyber	Operations	Integrity, Availability	Data-based	Moderate
Command Injection	Manipulate control systems	Cyber	Operations	Integrity	User-based	High
SQL Injection	Unauthorized database access	Cyber	Customer, Operations	Confidentiality, Integrity	Data-based	Moderate
Phishing & Social Engineering	Unauthorized access via deception	Cyber	Operations	Confidentiality, Integrity	User-based	Low

10.3389/fenrg.2025.1531655

for cybersecurity. The study uses a database that assists in inputting reconfiguration, allowing smart grid topology analysis under different attack strategies. It also stressed the need for cybersecurity actions such as having network intrusion prevention systems (NIPS) and firewalls against potential threats.

PMUs is used (Pourahmad and Hooshmand, 2023) to improve the security of power networks against cyber attacks. The method deployed optimal PMU placements to minimize the risk of attack and not compromise network observability using the Tabu Search algorithm (TS). The minimum number of PMUs necessary for full network observability and their optimal locations for enhanced cybersecurity were determined. The methodology consisted of an attack optimization problem that assesses potential disruptions originating from cyber attacks using the attack criterion based on the infinite norm of the attack vector. Simulations were conducted on IEEE bus networks, namely, IEEE 14-bus, IEEE 30-bus, and IEEE 57-bus networks within this study. The proposed method's effectiveness in shielding against cyber attacks is measured by two evaluation metrics: the number and location of PMU deployments in the network. H, denoted as the attack criterion, is another important metric that helps determine how far-reaching a possible attack can be within a given network. Observability conditions are assessed to ensure the network remains detectable and manageable even with probable attacks. Compared with previous works, simulation results verified that this technique outperforms them, proving its worthiness in improving network security.

A comprehensive and ongoing risk assessment methodology for smart grid systems is presented (Sharma et al., 2023). It uses attack-defense trees (ADTs) to show the complex interconnections between threats and responses across distributed components. The research stressed the importance of continuous risk evaluation and adaptation as a response to evolving threats and system states, which will aid in making security protection decisions by analyzing the sensitivity of system risk to various attack and defense scenarios to optimize security measures. Furthermore, this method encouraged standardized security and privacy protection taxonomies to facilitate security certifications. Evaluation metrics within the framework of the proposed approach focused on quantitative assessment of cyber risks in smart grid systems. Thus, it was validated using a real-life case study scenario, proving its usefulness for initial risk calculations and continuing assessments.

The research in Kumar et al. (2024) introduced a novel approach to secure smart grid systems by integrating cryptographic methods and AI techniques. They proposed an intelligent IDS that employed a combination of ML algorithms, including Convolutional Neural Networks (CNNs) and XGBoost Classifiers, to enhance anomaly detection and cyber attack identification. The Cryptographic algorithms, including Asymmetric algorithms like Rivest Shamir and Adleman (RSA) and Secure Hash Algorithm (SHA-512), were utilized to secure IoT devices. They utilized a dataset with 128 features and 29 types of estimations from PMUs to analyze smart grid data. The evaluation metrics used in the study included accuracy, F1 score, recall, and precision, which were critical for assessing the performance of ML algorithms in detecting cyber attacks on smart grids. The Random Forest Classifier achieved a higher performance compared to the XGB Classifier.

Table 8 summarizes the prevention and protection techniques discussed in this section.

4.2 Detection techniques

Detection techniques are critical in identifying cyber threats within smart grid systems as early as possible, enabling timely responses to minimize potential damage. In this subsection, we will investigate various detection techniques, examining their methodologies, effectiveness, and potential for application in smart grid cybersecurity to provide a comprehensive understanding of their role in enhancing grid resilience. Table 9 provides a summary of the detection methods presented in this section.

The research (Rahul et al., 2024) introduced a novel hybrid DL model, CMGFD-DL, for intrusion detection in smart grids, achieving a test accuracy of 98.20%. This model integrated convolutional and recurrent layers to enhance detection capabilities against cyber threats. The study utilized the Edge-IIoT cybersecurity Dataset, encompassing diverse IoT devices and attack categories, providing a rich foundation for evaluating intrusion detection mechanisms. The evaluation of the proposed intrusion detection model employs several key performance metrics, including accuracy, precision, and recall, which are essential for assessing its effectiveness in identifying cybersecurity threats. The proposed model demonstrates robustness and reliability, contributing significantly to the field of smart grid security by addressing the unique challenges posed by emerging technologies.

The work in Mobini et al. (2024) investigated a cyber-attack detection technique in smart grids using the windowed online dynamic mode decomposition (WODMD). It focused on three criteria for detection: the absolute Frobenius norm difference of system matrices, the absolute norm difference of the eigenvalue vector, and the one-step-ahead prediction error. The WODMD method is purely data-driven, operates online without prior learning of attack scenarios, and demonstrates superiority over three commonly used model-free methods in simulation studies. The study evaluated the proposed WODMD-based cyber-attack detection method using an IEEE 14-bus power system as the dataset. The dataset comprises voltage phase angles considered as states, with one bus designated as the reference. The evaluation of the cyber-attack detection problem is framed as a binary classification issue, assessed through various metrics. The evaluation of the cyber-attack detection problem is framed as a binary classification issue, assessed through various metrics.

Authors in Naqvi et al. (2024) proposed a novel reconstructive DL technique for detecting DDoS attacks in smart grid networks, which minimizes disruptions during introducing new attack classes. It explored three types of autoencoders for DDoS attack detection: deep autoencoder, extreme learning machine autoencoder, and marginalized denoising autoencoder. It evaluated the proposed method using standardized benchmark datasets, namely, UNB ISCX Intrusion Detection Evaluation 2012 dataset, and UNSW-NB15 dataset. The evaluation metrics employed for assessing the proposed method include false positive (FP), true negative (TN), false negative (FN), and true positive (TP) rates. These metrics are crucial for calculating overall accuracy. The model demonstrated its effectiveness in achieving high accuracy without requiring full model retraining.

The authors in Sweeten et al. (2023) highlighted the critical need for advanced intrusion detection systems (IDS) in smart power grids due to increasing cyber attacks targeting this infrastructure.

TABLE 8 Comparative analysis of protection and prevention techniques in smart grid Cybersecurity.

Ref	Focus of technique	Dataset used	Key metrics	Notable features
Vidya et al. (2023)	Vulnerability analysis using Attack Trees and Graphs for PMUs and IP sensors	Internal database for smart grid simulations	Attack Success Probability, Attack Cost, Attack Impact, CVSS	Uses graphical security models to identify vulnerabilities, emphasizes NIPS and firewalls for protection
Pourahmad and Hooshmand (2023)	Optimal PMU placement to minimize attack risk	IEEE 14-bus, IEEE 30-bus, IEEE 57-bus networks	Number and location of PMUs, Attack Criterion (H)	Uses Tabu Search Algorithm for optimal PMU placement, ensures network observability under cyber threats
Sharma et al. (2023)	Risk assessment using Attack-Defense Trees (ADTs)	Real-world smart grid scenario	Quantitative cyber risk assessment	Emphasizes continuous risk evaluation, promotes standardized protection taxonomies for certification
Kumar et al. (2024)	Integration of cryptographic methods and AI in IDS for smart grids	Dataset with 128 features, 29 types of PMU estimations	Accuracy, F1 Score, Recall, Precision	Combines CNNs and XGBoost with cryptographic methods like RSA and SHA-512 for enhanced security

They proposed solution involved a multi-modal IDS that fuses cyber and physical data, leveraging graph neural networks (GNNs) to enhance detection performance by exploiting spatial and temporal correlations. They developed a cyber-physical power system testbed that emulates the power grid's physical and cyber layers using OPAL-RT and RT-Lab, based on the ModbusTCP protocol. A comprehensive multi-modal dataset was created, covering normal operations and operations under cyber attacks, including false data injection (FDI) and ransomware attacks. Experimental results demonstrated that the GNN-based IDS significantly outperformed benchmark models, achieving a 5%–13%

The work in Shahid et al. (2023) proposed a novel detection technique called the nonlinear function-based variable dummy value model (NF-VDVM) to address the limitations of the existing variable dummy value model (VDVM) for detecting false data injection (FDI) attacks. NF-VDVM is designed to handle the vulnerabilities of the VDVM technique, particularly against attackers who can use multivariate linear regression to predict dummy values. The model was evaluated using the IEEE 14-bus test system, demonstrating its capability to enhance the security of the smart grid's measurement infrastructure. Data generation for this system is based on standard realistic load curves for four different seasons: winter, summer, spring, and fall. The evaluation metrics included the accuracy of attack detection and the system's resilience against stealth FDI attacks, ensuring the security of the smart grid's measurement. The findings indicated that the proposed method improves detection accuracy against stealth FDI attacks.

The authors in Habib et al. (2023) employed a hybrid ML approach for detecting DDoS attacks in smart grid systems, explicitly targeting PMU data within the wide-area measurement system (WAMS). Various ML algorithms are utilized, including Support Vector Machine (SVM), Artificial Neural Network (ANN), Logistic Regression (LR), Naive Bayes (NB), and Random Forest (RF). The study utilized a dataset from the Kaggle data center, designed explicitly for DDoS attack detection in WAMS. This dataset included both "normal" and "malicious" PMU data, essential for training various ML algorithms. The performance of these algorithms is evaluated based on accuracy, precision, recall, and F1-score, with the hybrid model achieving the highest accuracy.

4.3 Mitigation and recovery techniques

Mitigation and recovery techniques focus on minimizing the impact of detected cyber threats and swiftly restoring smart grid operations to normal. In this subsection, we will investigate various mitigation and recovery techniques.

The research in Zhu et al. (2022) presented a generalized data recovery model to address the challenges posed by false data injection attacks (FDIA) in smart grids. This model could be activated immediately upon detecting an attack, eliminating the need for impractical assumptions typically required by previous methods. It utilized the Measurement Data Inertia (MDI) concept to infer preliminary measurement values post-attack, enhancing recovery accuracy. An optimization model is introduced to refine the recovery process, ensuring the recovered data is closely aligned with real values. The recovery efficiency was evaluated using an indicator system that incorporates various error criteria, including Mean Absolute Error (MAE) and Mean Square Deviation (MSE). Extensive simulations performed on the IEEE 30-bus benchmark demonstrated that the proposed model can recover data closely to real values, enhancing the cybersecurity of smart grids.

The work in Cao et al. (2022) focused on a distributed resilient mitigation strategy for false data injection attacks (FDIA) in cyber-physical microgrids, addressing the vulnerabilities posed by cyber threats to physical system operations. It introduced a synchronous mitigation framework that utilized local detection to ensure data reliability and maintain control over reactive power in microgrids. The study categorized FDIAs into deception and disruption attacks, analyzing their impacts on voltage and reactive power stability. Simulations were conducted

TABLE 9 Comparative analysis of detection techniques in smart grid cybersecurity.

Ref	Attack type	Dataset used	Metrics used	Models used
Rahul et al. (2024)	Intrusion detection in smart grids	Edge-IIoT cybersecurity Dataset	Accuracy, Precision, Recall	CMGFD-DL (hybrid DL model with convolutional and recurrent layers)
Mobini et al. (2024)	Cyber attack detection in smart grids using WODMD	IEEE 14-bus power system (voltage phase angles)	TPR, FPR, binary classification metrics	Windowed Online Dynamic Mode Decomposition (WODMD)
Naqvi et al. (2024)	DDoS attack detection	UNB ISCX Intrusion Detection Evaluation 2012, UNSW-NB15	False Positive (FP), True Negative (TN), False Negative (FN), True Positive (TP), Accuracy	Deep autoencoder, Extreme Learning Machine autoencoder, Marginalized Denoising autoencoder
Sweeten et al. (2023)	Intrusion detection in smart power grids (cyber-physical IDS)	Multi-modal dataset (normal and attack scenarios including FDI, ransomware)	DR, False Alarm (FA) rate	Multi-modal IDS with Graph Neural Networks (GNNs)
Shahid et al. (2023)	FDI attack detection	IEEE 14-bus test system (standard realistic load curves)	Accuracy, system resilience, stealth attack detection	Nonlinear Function-based Variable Dummy Value Model (NF-VDVM)
Habib et al. (2023)	DDoS attack detection in PMU data within WAMS	Kaggle dataset for DDoS in WAMS (normal and malicious PMU data)	Accuracy, Precision, Recall, F1-score	Hybrid ML (SVM, ANN, LR, NB, RF)

using MATLABSimulink to validate the effectiveness of the proposed strategies. The proposed methodology aimed to restore system performance while minimizing communication overhead, enhancing the resilience of microgrid operations. Simulations validated the distributed resilient consensus cooperative control method's effectiveness under deception and disruption attacks.

Authors in Rahiminejad et al. (2023) proposed a cyber attack recovery scheme that focuses on enhancing the resilience of smart grids following cyber attacks on substations. The framework aimed to obtain the optimal recovery path based on the attacker's capability, including multi-stage attacks. The research also incorporated practical power system limitations and characteristics, such as generator Automatic Generation Control (AGC) capability and transient stability, enhancing overall system resilience. The proposed evaluation metrics encompassed five distinct metrics, four related to Power-side Resilience (PsR) and one to Cyber-side Resilience (CsR). PsR metrics include load restoration, reserve recovery, line capacity recovery, and reliability. The final cyber-physical resilience metric is termed CPARM, which integrates these metrics to assess the overall resilience of the smart grid. The evaluation also considered the maximum potential damage from cyber attacks, ensuring a comprehensive physical and cyber resilience assessment. The proposed approach was tested on the 39-bus New England test system and compared with existing methods to demonstrate its effectiveness in improving system resilience by 22%.

A novel multi-stage cyber intelligence technique (MSCIT) is designed in Muneeswari et al. (2024) to enhance security against multi-stage cyber attacks targeting the smart grid. It proposed a system comprising a pre-processing stage where a Chebyshev filter was used to filter the signals coming through the noise from an intrusion detection system (IDS) sensor. A set of components for risk identification, estimation, and evaluation were included in the MSCIT system as well, which in unison amplifies the capability of flagging possible threats. For comprehensive treatment, the security identification block utilized a cyber database to assess the risks. If a risk is not recorded, the data is forwarded to a Bi-LSTM network for such an attack to be investigated more closely. The methodology is software based on a MATLAB simulator, and evaluation metrics are precision, F1-score, specificity, accuracy, and detection rate, which are essential measures. The research involved the DARPA 2000 dataset, well-known to cybersecurity researchers. The obtained experimental results proved that the efficiency of the MSCI method was as high as 99%, which is a considerable improvement over CDS, AD-IOT, SVM, and other existing methods.

The study (Wang et al., 2020) proposed the reconstruction of the operating states to improve the resilience of cyber-physical smart grids (CPSG) against any prospective cyber attacks. The scheme incorporated an attack separation mechanism, a state forecasting algorithm, and a state recovery approach to filter out the prospective threat. Worst-case analysis and P-Q decomposition were employed in the attack separation method to determine and mark grid states that were out of the normal. They used the state forecasting algorithm, a particle filtering technique to estimate the actual operating levels of the marked states, thus eliminating the adverse effects that cybernetic attacks would have caused. The state recovery mechanism's approach concerned

TABLE 10 Comparative analysis of mitigation and recovery techniques in smart grid cybersecurity.

Ref	Focus of technique	Dataset used	Key metrics	Notable features
Zhu et al. (2022)	Data recovery model for False Data Injection Attacks (FDIA)	IEEE 30-bus benchmark	Mean Absolute Error (MAE), Mean Square Deviation (MSE)	Uses Measurement Data Inertia (MDI) to infer preliminary values and optimization for accurate data recovery
Cao et al. (2022)	Distributed resilient mitigation for FDIA in microgrids	MATLABSimulink for cyber-physical microgrids	System reliability, voltage stability, reactive power control	Introduces a synchronous mitigation framework with local detection and categorization of deception and disruption attacks
Rahiminejad et al. (2023)	Cyber attack recovery scheme to improve resilience post-attack	39-bus New England test system	Power-side Resilience (PsR), Cyber-side Resilience (CsR), CPARM metric	Incorporates multi-stage attack recovery, considering generator AGC and transient stability
Muneeswari et al. (2024)	Multi-Stage Cyber Intelligence Technique (MSCIT) for multi-stage attacks	DARPA 2000 dataset	Precision, F1 Score, Specificity, Accuracy, Detection Rate	Chebyshev filter for pre-processing IDS data, Bi-LSTM Network for attack verification
Wang et al. (2020)	State reconstruction for Cyber-Physical Smart Grids (CPSG) against cyber attacks	IEEE 9-, 14-, 30-, 57-, and 118-bus power systems, Dongguan dispatch center load data	Operating state accuracy	Combines attack separation, state forecasting (particle filtering), and state recovery for enhanced security

getting the forecasted states' corrected states by filtering out the smearing effects of uncontaminated states. The works employed various datasets, including the IEEE standard 9-, 14-, 30-, 57-, and 118-bus power systems, to test the proposed scheme for state reconstruction. System load curves were collected from the Dongguan dispatch center in China, representing actual operating conditions in the power industry by constructing the typical daily load curve.

The list of mitigation and recovery techniques discussed in this section is summarized in Table 10.

4.4 Challenges and limitations

As smart grids grow more sophisticated and interdependent, securing them becomes increasingly difficult for existing cybersecurity systems. One such issue includes the never-ending changes in the methods used by attackers, which advance quicker than older mechanisms such as intrusion detection systems based on recognizing specific signatures and incapable of zero-day attacks or APTs (Achaal et al., 2024). In addition, ML and DL models that are often deployed for the detection of anomalies are model-driven and require large sets of features. Nonetheless, both limited past cyber attack events in the area of smart grid and the overwhelming presence of normal operations biased those datasets, thereby hurting the models, performance (Divakaran and Peddinti, 2024). Another critical limitation is the lack of interpretability in ML and DL systems, which frequently function as "black boxes," making their decisions challenging to understand and trust—a major concern in critical infrastructures such as smart grids (Naiho et al., 2024).

Resource constraints exacerbate these challenges, as many cybersecurity solutions require significant computational power, which is incompatible with the limited processing capabilities of edge devices commonly deployed in smart grids. One of the issues is caused by the limited scaling up of traditional structures that integrate protection against the entire set of the rapidly growing thousands of grid elements, such as DERs, IoT, and other devices (Zhang et al., 2020). In addition, ML and DL models are susceptible to so-called adversarial attacks, in which slight perturbations of the input data can evade the detection systems, thereby threatening the overall security of the grid (Lakhani and Rohit, 2024). Also, these methods are context-blind, which means they cannot identify advanced multi-stage attacks exploiting specific features of the operations of smart grid systems.

Furthermore, many existing solutions have high false positive rates, which overwhelm operators with frequent alerts and cause desensitization or alert fatigue, increasing the likelihood that critical incidents are missed (Zhang et al., 2020). Another weakness is in real-time capabilities, where some of the techniques have high latencies and computational requirements, which have a great impact on timely detection and response that are critical in averting avalanche collapses in smart grids (Gunduz and Das, 2020). These constraints together emphasize the necessity for developing more dynamic, contextually relevant, and cost-effective approaches to ensure cybersecurity, thus making it possible to use LLMs (Ferrag et al., 2024). The problem of the current smart grid cybersecurity paradigm can be addressed due to the possibilities afforded by LLMs due to their ability and capacity to deeply understand context, synthesize data, and adapt to changing conditions.

5 The role of large language models in cybersecurity

The emergence of AI and NLP has shown that LLMs could be used to better the efforts of cybersecurity practice. The section will provide an overview of the basic principles of LLM, their typical use case scenarios within the cybersecurity domain, and evaluate their promise, but also the present shortcoming in particular the area of smart grid systems security.

5.1 Background on large language models

LLMs have fundamentally altered the realm of NLP by, without a doubt, demonstrating the ability to understand, generate, and evaluate intricate human language. By employing the transformer architecture (Vaswani et al., 2017), LLMs like GPT (Zaboli et al., 2024a), BERT (Kenton and Toutanova, 2019), or T5 (Raffel et al., 2020) can concentrate on whole sequences of text through self-attention mechanisms. Through this self-attention mechanism, it is evident that LLMs can understand many contexts within long texts, hence producing more coherent and relevant responses. Unlike previous models like Recurrent Neural Networks (RNNs), transformers can handle and operate on a text sequence more efficiently and simultaneously scale, increasing the potential of LLMs (Vaswani et al., 2017).

The scale is an essential feature of LLMs: The models in question consist of billions of parameters, or learned weights, which refine understanding and generating. The convexity of GPT-3, which has 175 billion parameters, brings unique abilities such as few-shot or zero-shot learning, making it possible to address entirely new tasks with minimal to zero data for the target task. Also, the increase in the volume of LLMs is associated with the improvement of reasoning processes in these models, directing and being able to be instructed to 'think' in a chain of reasoning, which improves their precision in performing complex tasks. The onset of these capabilities makes LLMs ideal in a wide range of tasks, as they are able to perform text generation and even more complex tasks, finding their way out of even complex problems (Brown et al., 2020). LLMs are generally created in two phases: initial training and subsequent refinement. During the pre-training stage, LLMs learn language structures and concepts by sifting through a substantial corpus, enabling them to articulate fluent language (Liu et al., 2024). Fine-tuning improves the task completion of the model, for example, by answering questions or summarizing, so that it is trained to available labelled data related to fulfilling such tasks. This two-step mechanism allows LLMs to be utilized as universal and task-oriented instruments; hence, their usage flexibility across various tasks increases (Liu et al., 2024). However, the use of LLMs also raises some relevant ethical problems. Since LLMs are trained using large volumes of data from different sources, they may reproduce the biases embedded in that data, resulting in biased and possibly harmful results (Liu et al., 2025). Additionally, considerable environmental and economic costs are associated with the large energy required to train these models. Individuals in the profession work actively to solve these problems, for example, by employing bias reduction or energy-efficient training processes (Asesh and Dugar, 2023). LLMs represent a more mature stage in the development of AI-implemented language understanding systems, as they have demonstrated a notable performance in language-focused tasks that require the incorporation of many contexts. Their potential impact in various areas, especially cyberspace protection, enables them to be key technologies for any language-oriented big data in analysis, generation, and comprehension in innovative and complex ways.

5.2 Applications of LLMs in cybersecurity

The cybersecurity community progressively acknowledges LLMs' capabilities, leading to their adoption for analyzing extensive datasets, identifying emerging threats, and improving automated response functionalities. The study by Ferrag et al. (2024) goes into detail on the use cases and the reasons for the adoption of LLMs, considering that such tools have the capability of reducing work overload for cybersecurity personnel through automating activities such as vulnerability scanning, network mapping, and the exploitation of known vulnerabilities. This research includes the performance evaluation of 42 LLMs on some cybersecurity datasets to identify the strongholds and the deficiencies of these models and define the scope of future research. This paper examined prompt injection and data poisoning issues that emerged through LLM use and researched approaches intended to secure such models. Besides identifying bottlenecks, the general discussion favoured the crucial point of devising safe and efficient models with the implementation of LLMs. The latter concerns advanced technologies such as half-quadratic quantization and reinforcement learning with human feedback that could effectively enhance cybersecurity measures in real-time against new risks.

A novel technique in Mudassar Yamin et al. (2024) was introduced that utilized LLMs to create rolling and complex scenarios of cybersecurity exercises that improved the training and the awareness by mimicking various cyberspace threats, both existent and new. This approach also stems from Turing's work on machine intelligence, and it attempts to apply some form of machine intelligence simulation to human intelligence. The study highlighted the capability of LLMs to create complex scenarios that question conventional cybersecurity training approaches, turning the intrinsic "hallucination" of LLMs into a beneficial aspect. The produced scenarios were subjected to a thorough evaluation, encompassing assessments with GPT models and expert analysis to guarantee their authenticity and pertinence, utilizing a RAG methodology to enhance the intricacy of the tasks.

Authors in Bhatt et al. (2024) proposed CYBERSECEVAL2, which is a benchmark suite developed to evaluate cybersecurity vulnerabilities of LLMs considering prompt injection and interpreter abuse attack. Adversarial methods, including gradient and heuristic optimization techniques, were used to induce attack behavior within the LLMs. All evaluated LLMs exhibited weaknesses to prompt injections, with success rates between 26% and 41%, highlighting a considerable difficulty in maintaining compliance with system prompts. The investigation underscored the significance of quantifying the false refusal rate (FRR) to comprehend the safety-utility balance in LLM responses to cybersecurity-related tasks. The findings showed that while LLMs are responsive to benign requests, they do, however, have significant weaknesses when inflating injections are used. Hence, further developments in enhancing security are more necessary during their use.

The authors in Tseng et al. (2024) described an automated AI agent for analysts to work with cyber threat intelligence (CTI) reports in security operations centers (SOCs), which today is rather time-consuming and requires effort. This agent used the capabilities of GPT-4 and other LLMs to extract information and create regular expressions (RegEx) for building the necessary SIEM rules by itself. The procedure involved a purification phase intended to improve the accuracy of the recognized identified indicators of compromise (IOCs) and build relationship graphs to depict interrelations among the different IOCs. The investigation underscored the constraints of current ML methods in adapting to the progression of attack techniques, stressing the necessity for more sophisticated solutions.

A framework named SEVENLLM (Ji et al., 2024) aimed at benchmarking, eliciting, and improving the capabilities of LLMs in analyzing and responding to cybersecurity incidents. Developing a bilingual educational corpus, SEVENLLM-Instruct, successfully addressed the problem of the scarcity of quality, cohesive datasets through a cybersecurity text-based approach. With the original texts as raw material, supervised corpora were created to train different foundational LLMs with a multi-task learning objective using automatic retrieval of tasks from a task pool. A new evaluation benchmark, SEVENLLM-Bench, has been developed to measure the performance of LLMs in CTI, addressing the existing gaps between traditional domains and cybersecurity. Thorough investigations on the specialized benchmark, SEVENLLM-Bench, validate SEVENLLM's effectiveness in enhancing analytical capabilities and delivering strong responses to emerging cyber threats.

The usage of LLMs within cybersecurity was brought forth in Divakaran and Peddinti (2024), where the authors suggest that LLMs could be used to improve or even automate some security classifiers. The aim was to use LLMs to augment the data so that many training samples could be created without large data collection effort. The study investigates the applications of LLMs in phishing detection, highlighting systems such as D-Fence and ChatSpamDetector that leverage LLMs for efficient email classification. They also discussed the continuous endeavours to reduce risks linked to LLMs via frameworks and cooperative initiatives among corporations and governmental bodies.

The authors in Li et al. (2024b) drew attention to the primary weaknesses of LLMs in the context of smart grid applications, with a particular focus on bad data injection and knowledge extraction. It notes that even LLMs can be exploited by adversaries to insert textual information of fabric nature or to withdraw domain confidential knowledge, thus creating a risk to data confidentiality. The analysis argues that there is a need to evaluate these risks before deploying LLMs in critical infrastructure deployments, as new attack vectors may emerge with the accelerated development of LLM technologies. Future studies should focus on tracking such emerging threats.

The work in Huang and Zhu (2023) presents PenHeal, a two-stage framework utilizing LLM technology that independently detects and addresses security vulnerabilities via its integrated Pentest Module and Remediation Module, thereby improving the automation of penetration testing and vulnerability remediation processes. The combination of the two modules is enhanced by methods like Counterfactual Prompting and an Instructor module, which directs the LLMs by leveraging external knowledge. This enables the framework to investigate various possible attack routes thoroughly, thus enhancing the overall efficiency of identifying and addressing vulnerabilities. Experimental results demonstrated that PenHeal enhanced vulnerability coverage by 31%, boosted the effectiveness of remediation strategies by 32%, and decreased associated costs by 4% in comparison to baseline models, highlighting the framework's considerable influence on cybersecurity practices.

ShieldGPT (Wang et al., 2024b) is a framework designed to help overcome DDoS attacks using LLMs to facilitate detection and answering mechanisms. ShieldGPT consisted of four main components: attack detection, traffic representation, domain-knowledge injection and role representation. It also provided a representation scheme that could capture both global and local traffic features effectively and prompt the generation of easy-to-understand and specific explanations and mitigation measures. Preliminary results showed that ShieldGPT effectively provided helpful information and strategies for dealing with DDoS attacks.

The study (Guastalla et al., 2023) explored the efficacy of LLMs, including OpenAI's ChatGPT variants (GPT-3.5, GPT-4, and Ada), in improving the detection capabilities of DDOS attacks, showcasing their promise as a solution for network security challenges. The findings indicated that LLMs, incredibly when fine-tuned, attained impressive accuracy rates of around 95% on the CICIDS 2017 dataset and nearly 96% on the Urban IoT Dataset for aggressive DDoS attacks, surpassing conventional neural networks such as multi-layer perceptrons (MLP) trained with comparable data.

Net-GPT (Piggott et al., 2023), an LLM-driven offensive chatbot crafted to comprehend network protocols and carry out MITM attacks on communications between unmanned aerial vehicles (UAV) and ground control stations (GCS), highlighting the capabilities of LLMs in the realm of cybersecurity threats. The results demonstrated the generative performance of Net-GPT which was on average 95.3% with Llama-2-13B and 94.1% with Llama-2-7B. Furthermore, it emphasized the efficacy of low-end models like Distil-GPT-2 that, with a speed improvement of 47 folds, can achieve 77.9% predictive ability of Llama-2-7B model. This illustrates the trade-offs between model size, speed, and accuracy in edge-computing environments.

AURORA (Wang et al., 2024a), an automatic end-to-end framework for constructing and emulating cyber attacks. This system autonomously develops multi-stage cyber attack plans derived from CTI reports, establishes the required emulation infrastructures, and carries out the attack procedures independently, thereby greatly minimizing the time needed for attack simulation. By integrating 40% more attack techniques than earlier solutions, AURORA enhanced the diversity and quality of constructed attacks. This has been evaluated by designing and deploying more than 20 cyber attacks encompassing the whole cycle, thus establishing the efficiency and effectiveness of advanced cyber attacks simulation.

Table 11 summarizes the efforts in utilizing LLMs in cybersecurity.

5.3 Exploring LLMs for smart grid cybersecurity

The increasing application of LLMs and natural language processing (NLP) in smart grid cybersecurity is welcomed as it enhances the security and protection of crucial resources. While

TABLE 11 Summary of surveyed literature on LLM applications in cybersecurity.

Ref	Focus area	Key contribution	Use case/Application
Ferrag et al. (2024)	Performance Evaluation of LLMs	Evaluated 42 LLMs on cybersecurity datasets; introduced techniques like RLHF	Automating vulnerability scanning and network mapping
Mudassar Yamin et al. (2024)	Cybersecurity Training	Leveraged LLMs to create complex cyber scenarios for training purposes	Simulation of advanced cyberspace threats for training
Bhatt et al. (2024)	Vulnerability Benchmarking	Developed CYBERSECEVAL2 benchmark suite for LLM vulnerabilities	Assessment of LLM weaknesses against prompt injections
Tseng et al. (2024)	CTI	Automated AI agent for processing CTI reports in SOCs	Extracting IOCs and automating SIEM rule creation
Ji et al. (2024)	Bilingual Cybersecurity Framework	Created SEVENLLM framework and evaluation benchmark for LLMs in cybersecurity	Improved analysis and response for CTI
Divakaran and Peddinti (2024)	Phishing Detection	Explored LLM applications in augmenting phishing classifiers like D-Fence	Automating email classification for phishing detection
Li et al. (2024b)	Smart Grid Security	Highlighted risks like bad data injection and knowledge extraction in LLMs	Critical infrastructure protection using LLMs
Huang and Zhu (2023)	Penetration Testing and Remediation	Introduced PenHeal framework for automating penetration testing and remediation	Enhanced automation of penetration testing
Wang et al. (2024b)	DDoS Mitigation	Developed ShieldGPT framework to detect and mitigate DDoS attacks	Addressing DDoS attacks with LLM-driven detection
Guastalla et al. (2023)	DDoS Detection	Showcased LLMs' performance in detecting DDoS attacks on datasets like CICIDS	Improved accuracy in network security challenges
Piggott et al. (2023)	Offensive Cybersecurity	Introduced Net-GPT chatbot for simulating offensive tasks like MITM attacks	Simulated MITM attacks on UAV communications
Wang et al. (2024a)	Cyber attack Simulation	Developed AURORA framework for end-to-end emulation of cyber attacks	Automated multi-stage cyber attack construction

LLMs are extensively used in cybersecurity concerning threat recognition, logging, and reporting irregularities, their use in smart grid cybersecurity seems still under-researched. Employing LLMs for information assurance has emerged as a hot area of investigation. Still, there are relatively few studies that seek to respond to the information assurance challenges in smart grids using LLMs. This section reviews the literature on the subject to assess the status of LLM deployment within the context of smart grid information security, along with the issue's boundaries that require more investigation.

The authors in Zaboli et al. (2024a) advocated for using LLMs such as ChatGPT, to enhance the augmentation of cybersecurity within the IEC 61850-based communication in digital substations. The paper also provides evidence of a comparison between different LLMs based on performance evaluation metrics including the true positive rate (TPR), the false positive rate (FPR), the false negative rate (FNR), and the precision and F1 score. From the analysis results, it was found out that LLM ChatGPT 4.0 was more integrated with the detection of anomalies in IEC 61850 communications than other LLMs, including Anthropic's Claude 2 and Google BardPaLM 2, attaining TPRs of 98.18% for GOOSE (Generic Object Oriented Substation Event) and 96.67% for SV (Sampled Value) messages at the maximum trained levels. The paper developed a hardware-in-the-loop (HIL) testbed to fabricate and retrieve datasets for GOOSE and SV communication, which enabled the realistic case studies.

The authors in Zaboli et al. (2024a) extended their work in Zaboli et al. (2024b). They introduced an extended task-oriented dialogue (ToD) system named CyberGridToD, which utilized LLMs for anomaly detection (AD) in multicast messages within digital substations. It automates decision-making processes by simulating human choice patterns, which may reduce error rates over time as it learns from new data. The study used two main datasets: GOOSE and SV packets, which are essential for communication

TABLE 12 Comparative analysis of LLM-based cybersecurity applications in smart grids.

Ref	Attack	Dataset used	Metrics used	Models used
Zaboli et al. (2024a)	Anomaly detection in IEC 61850-based communications	GOOSE and SV datasets generated using hardware-in-the-loop (HIL) testbed	TPR, FPR, FNR, Precision, F1-Score	ChatGPT 4.0, Claude 2, Google Bard PaLM 2
Zaboli et al. (2024b)	Anomaly detection in multicast messages within digital substations	GOOSE and SV packets for communication in digital substations	TPR, FPR, FNR, Precision, accuracy, F1-Score, PercMarkedness, Informedness, MCC	Anthropic Claude Pro model, Microsoft Copilot AI

in digital substations. The evaluation metrics used in the proposed LLM-based ToD framework include true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs), true positive rate (TPR), and false positive rate (FPR). Advanced metrics such as markedness, informedness, and Matthews correlation coefficient (MCC) were employed to evaluate the model's reliability and decision-making quality (Zaboli et al., 2024b). Compared to traditional human-in-the-loop (HITL) processes, the model outperformed them regarding scalability, adaptability, and error rate. It set a new standard for evaluating intrusion detection systems (IDSs) with less effort and greater adaptability than previous methods (Zaboli et al., 2024b).

Table 12 summarizes the LLM-based models based on the attached type, datasets, metrics, and models.

5.4 LLMs as an enabler of cyber attacks

While LLMs play a pivotal role in enhancing cybersecurity defenses, they also introduce new attack vectors that adversaries can exploit. Cybercriminals leverage adversarial prompt engineering, fine-tuning, and automation capabilities of LLMs to scale cyberattacks, bypass security mechanisms, and enhance malware generation. These capabilities make LLMs an enabler of cyber threats such as phishing, malware obfuscation, adversarial AI attacks, and automated reconnaissance.

One of the most significant threats posed by LLMs is their ability to generate highly sophisticated phishing emails and conduct automated social engineering attacks. By leveraging advanced prompt engineering techniques, adversaries can craft persuasive and highly targeted phishing emails that evade traditional security filters by mimicking legitimate communication patterns (Brown et al., 2020; Fakhouri et al., 2024). Fine-tuning LLMs on leaked corporate email datasets further enhances the personalization of phishing attacks, increasing the likelihood of credential theft and unauthorized access to SCADA (Supervisory Control and Data Acquisition) systems.

Beyond phishing, LLMs can optimize malware development by automating code generation, obfuscation, and polymorphism. Fine-tuned models trained on cybersecurity datasets, such as VirusTotal and CISA Malware Archives, can assist attackers in creating malware that evades signature-based detection mechanisms (Chen et al., 2021; Pearce et al., 2025). By leveraging transformer-based architectures such as GPT-4, Codex, and LLaMA, threat actors can generate or modify malware variants in real-time, making detection and mitigation significantly more challenging. Additionally, LLMs enable reverse engineering by analyzing firmware, software binaries, and network logs, aiding adversaries in identifying zero-day vulnerabilities within smart grid control systems. The ability to generate dynamic exploit code using AI significantly lowers the technical expertise required to launch cyberattacks.

A growing concern is adversarial attacks on AI-driven cybersecurity mechanisms, such as IDS and anomaly detection models. Attackers can craft adversarial examples—subtly modified inputs designed to deceive AI-based security defenses (Biggio and Roli, 2018; Zhang and Li, 2019). By fine-tuning LLMs on cybersecurity logs and attack patterns, adversaries can manipulate IDS models into misclassifying malicious traffic as benign, allowing stealthy and persistent access to critical infrastructure. Furthermore, reinforcement learning-based fine-tuning enables attackers to optimize AI-generated attack methods based on system responses, making traditional defense mechanisms less effective.

LLMs are also used to automate reconnaissance and vulnerability exploitation. AI-driven cyber agents can process cybersecurity knowledge bases, network topologies, and open-source intelligence (OSINT) datasets to generate detailed attack plans (Browne et al., 2024; Mishra et al., 2022). This includes extracting organizational structures from employee profiles and leaked databases, automating penetration testing by identifying misconfigurations and security gaps, and generating step-by-step attack execution scripts, reducing the need for manual attack planning. Moreover, LLMs can automate vulnerability exploitation by generating tailored payloads for system-specific weaknesses, significantly lowering the barrier to entry for cybercriminals.

Despite their increasing use in cybercrime, LLMs also have inherent limitations that impact their effectiveness in adversarial applications. First, data constraints hinder their efficiency, as attackers require high-quality datasets, such as real-world exploit samples and leaked credentials, to fine-tune LLMs effectively (Ferrag et al., 2024; Ji et al., 2024). However, access to proprietary threat intelligence is often limited. Second, hallucination issues present a significant challenge, as LLMs frequently generate incorrect or misleading attack strategies, reducing their reliability in real-world hacking scenarios (Li et al., 2024b). Third, as cybersecurity tools integrate adversarial training, AI-generated

attacks are becoming more detectable, requiring attackers to constantly refine their methods.

In summary, LLMs present a dual-use challenge in cybersecurity—while they enhance defensive strategies, they also empower cybercriminals by automating phishing, malware development, adversarial attacks, and reconnaissance. Understanding these adversarial applications is critical for developing proactive AI-driven countermeasures and securing smart grid infrastructures against AI-enabled cyber threats.

5.5 Critical analysis of LLMs in cybersecurity

Despite their capabilities, LLMs face several limitations that hinder their widespread adoption in cybersecurity defense. These challenges stem from data availability constraints, security vulnerabilities, model limitations, and integration difficulties, all of which must be addressed to ensure LLMs contribute effectively to smart grid security.

One of the primary obstacles to using LLMs in cybersecurity is the limited availability of high-quality training data. Cybersecurity threats are dynamic and complex, requiring extensive and up-to-date datasets for effective training. However, organizations are often reluctant to share cybersecurity-related data, leading to data scarcity that hampers the development of robust cybersecurity LLMs (Ferrag et al., 2024; Ji et al., 2024). The non-sharing of threat intelligence and network security data slows the generation of relevant datasets, making it difficult to train effective LLM-based threat detection systems.

Additionally, LLMs introduce new cybersecurity vulnerabilities. One major risk is adversarial attacks, in which maliciously crafted inputs manipulate LLMs to generate incorrect or misleading outputs. These attacks pose a significant threat to AI-driven cybersecurity systems, as threat actors can exploit them to bypass automated detection mechanisms or manipulate security recommendations (Ferrag et al., 2024; Ji et al., 2024). Prompt injection attacks represent another major concern, as attackers can exploit vulnerabilities in LLM-generated responses to extract confidential information or manipulate system behavior, leading to data leaks, unauthorized access, or misinformation (Ferrag et al., 2024; Ji et al., 2024).

A critical limitation of LLMs in cybersecurity is the phenomenon of hallucination, where models generate coherent but factually incorrect or misleading information. In cybersecurity, even minor inaccuracies in threat intelligence reports or vulnerability assessments can lead to misguided decisions and system compromise (Ferrag et al., 2024; Li et al., 2024b). Hallucination mitigation strategies, such as fact verification mechanisms, retrieval-augmented generation (RAG), and reinforcement learning from human feedback (RLHF), are essential for improving LLM reliability (Li et al., 2024b).

6 Future research directions

The integration of LLMs in the cybersecurity of smart grids presents several promising research directions to enhance resilience, adaptability, and security. To fully develop the potential of LLMs in smart grid security, future research must focus on their robustness, explainability, integration with emerging technologies, and protection against evolving cyber threats. Key areas that require further exploration include:

- Development of Robust and Explainable LLM Models: Future research should aim at improving the robustness of LLMs against adversarial attacks, ensuring interpretability and trustworthiness in cybersecurity applications. This includes designing explainable AI (XAI) frameworks that provide insights into LLM decision-making, enabling operators to validate responses effectively (Jha, 2023).
- Advanced Anomaly Detection Mechanisms: LLMs can be further optimized for anomaly detection in smart grids by leveraging self-learning mechanisms that identify novel cyber threats in real-time. This entails enhancing adaptive security systems capable of detecting and mitigating evolving cyber threats without frequent retraining.
- Integration with Emerging Technologies: Combining LLMs with blockchain, edge computing, and federated learning can enhance security frameworks by improving data integrity, decentralization, and efficient threat detection. Blockchain can ensure tamper-proof logs, while edge computing can enable real-time processing of security data closer to the source (Jha, 2023).
- Standardized Frameworks and Best Practices: Establishing standardized guidelines and best practices for the deployment of LLMs in smart grid cybersecurity is crucial. Future research should develop regulatory-compliant frameworks to ensure interoperability and reliability while addressing ethical considerations in AI-driven security applications.
- Privacy-Preserving Techniques: Ensuring data privacy while using LLMs in cybersecurity is critical. Techniques such as differential privacy, secure multiparty computation, and homomorphic encryption should be explored to enhance data confidentiality without compromising model performance (Li et al., 2024a).
- Defensive Strategies Against LLM Exploitation: Research should focus on developing advanced threat models to protect LLMs against data poisoning, prompt injection, and adversarial perturbations. Implementing validation mechanisms and real-time anomaly detection can prevent LLMs from being manipulated by attackers (Nakhleh et al., 2024).
- Addressing Hallucinations in LLMs: To mitigate the risks posed by LLM hallucinations in smart grid applications, retrieval-augmented generation (RAG) and knowledge graphs (KGs) should be incorporated to ground LLM outputs in reliable datasets, ensuring factual accuracy (Ibrahim et al., 2024; Perković et al., 2024).
- Advancements in Energy Sector Cybersecurity: Future research should consider the application of LLMs in protecting critical energy infrastructure such as industrial control systems (ICS), SCADA systems, microgrids, and islanding operations. Additionally, the potential of energy honeypots for deception-based cyber defense mechanisms should be explored.

By focusing on these research directions, LLM-based security solutions can be significantly enhanced, making smart grids more resilient against cyber threats while maintaining high levels of efficiency and reliability.

7 Future research directions

The integration of LLMs in the cybersecurity of smart grids presents several promising research directions to enhance resilience, adaptability, and security. To fully develop the potential of LLMs in smart grid security, future research must focus on their robustness, explainability, integration with emerging technologies, and protection against evolving cyber threats. Additionally, it is critical to explore the role of LLMs in securing energy infrastructure, including industrial control systems (ICS), SCADA environments, microgrids, islanding mechanisms, and energy honeypots. Key areas that require further exploration include:

- Development of Robust and Explainable LLM Models: Future research should aim at improving the robustness of LLMs against adversarial attacks, ensuring interpretability and trustworthiness in cybersecurity applications. This includes designing explainable AI (XAI) frameworks that provide insights into LLM decision-making, enabling operators to validate responses effectively (Jha, 2023; Arrieta et al., 2020).
- Advanced Anomaly Detection Mechanisms for Smart Grids: LLMs can be further optimized for anomaly detection in smart grids by leveraging self-learning mechanisms that identify novel cyber threats in real time. Reinforcement learning (RL) techniques can be employed to continuously train LLM-based anomaly detection models on real-world cyber threats (Mukherjee et al., 2023). Future research should explore how RL-enhanced LLMs can adapt to evolving attack patterns in ICS/SCADA environments and microgrid operations (Yadav and Paul, 2021).
- LLMs for ICS and SCADA Security: Industrial control systems (ICS) and SCADA environments are critical components of smart grids that require specialized cybersecurity measures. Future research should focus on how LLMs can be fine-tuned on ICS-specific datasets to detect anomalous commands, unauthorized access, and cyber-physical threats (Yadav and Paul, 2021). Additionally, AI-driven threat modeling can enhance SCADA resilience by predicting attack vectors before exploitation occurs (Bhamare et al., 2020).
- LLMs for Microgrid and Islanding Cybersecurity: With the increasing adoption of decentralized energy networks, microgrid security has become a key concern. Future studies should investigate how LLMs can assist in microgrid anomaly detection, network segmentation security, and protection against false data injection attacks (FDIAs) in islanding scenarios (Zhang et al., 2021). Additionally, AI-based decision support systems for energy transition planning during islanding events should be explored (Dutta et al., 2021).
- Energy Honeypots for Deceptive Cyber Defense: Deploying LLM-assisted honeypots in smart grid infrastructures can serve as a deception-based security strategy. Future research should examine how generative AI can create realistic honeypot environments to attract and analyze adversaries, thereby improving threat intelligence (Mashima and An, 2019). Additionally, integrating LLM-driven deception tactics with ICS/SCADA security can enhance early threat detection (Mashima and An, 2017).
- Integration with Emerging Technologies: Combining LLMs with blockchain, edge computing, and federated learning can enhance security frameworks by improving data integrity, decentralization, and efficient threat detection. Blockchain can ensure tamper-proof logs, while edge computing can enable real-time processing of security data closer to the source (Zhuang et al., 2021; Alazab et al., 2022).
- Standardized Frameworks and Best Practices: Establishing standardized guidelines and best practices for the deployment of LLMs in smart grid cybersecurity is crucial. Future research should develop regulatory-compliant frameworks to ensure interoperability and reliability while addressing ethical considerations in AI-driven security applications (Jawhar et al., 2024).
- Privacy-Preserving Techniques: Ensuring data privacy while using LLMs in cybersecurity is critical. Techniques such as differential privacy, secure multiparty computation, and homomorphic encryption should be explored to enhance data confidentiality without compromising model performance (Li et al., 2024a; Gentry, 2009).
- Defensive Strategies Against LLM Exploitation: Research should focus on developing advanced threat models to protect LLMs against data poisoning, prompt injection, and adversarial perturbations. Implementing validation mechanisms and real-time anomaly detection can prevent LLMs from being manipulated by attackers (Nakhleh et al., 2024; Ghimire and Thapaliya, 2024).
- Addressing Hallucinations in LLMs: To mitigate the risks posed by LLM hallucinations in smart grid applications, retrieval-augmented generation (RAG) and knowledge graphs (KGs) should be incorporated to ground LLM outputs in reliable datasets, ensuring factual accuracy (Ibrahim et al., 2024; Perković et al., 2024).

By focusing on these research directions, LLM-based security solutions can be significantly enhanced, making smart grids more resilient against cyber threats while maintaining high levels of efficiency and reliability.

8 Conclusion

The electric grid's incorporation of smart grid technology has improved performance and increased sustainability. However, the addition of ICT poses complex challenges in terms of cybersecurity that must be dealt with to protect critical infrastructure. This article takes a comprehensive analysis of cyber warfare with a focus on smart grids, detects and analyses countermeasures and investigates the potential value of LLMs for enhancing the grid's security with LLMs. This review emphasizes the range and variety of cyber attacks, including data integrity attacks of the FDI type and

complex multi-layered advanced persistent threats (APTs). Existing countermeasures for mitigating, for example, those based on machine-learning algorithms for intrusion detection and employing cryptographic means, are promising but have limitations, such as high demand for computation power and complexity of sharing understandable results. As new technologies advance, for example, in LLMs, it can change the dynamics of anomaly identification, pattern recognition and response mechanisms in real time. However, several questions, such as the model's reliability, the hallucinations of the models and the context in which the models will be deployed, should be answered before more extensive application in critical infrastructure. Looking ahead, robust strategies for smart grids could be implemented by applying adaptive ML models. These strategies could be revolutionary in conjunction with advanced domain-specific fine-tuning and verification processes for LLMs. Joint efforts among the experts and decision-makers in the relevant industries will have to be made against the weaknesses in detecting, preventing, and recovering threats. Multi-facetted cybersecurity structures coupled with smart grids can change the face of energy distribution while ensuring operational safety and reliability.

Author contributions

NI: Conceptualization, Formal Analysis, Investigation, Methodology, Writing–original draft, Writing–review and editing. RK: Conceptualization, Project administration, Supervision, Validation, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research is funded by Toronto Metropolitan University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. We used it for English Improvements.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Achaal, B., Adda, M., Berger, M., Ibrahim, H., and Awde, A. (2024). Study of smart grid cyber-security, examining architectures, communication networks, cyber-attacks, countermeasure techniques, and challenges. *Cybersecurity* 7, 10. doi:10.1186/s42400-023-00200-w

Alazab, M., Swarna Priya, R. M., Parimala, M., Maddikunta, P. K. R., Gadekallu, T. R., and Pham, Q.-V. (2022). Federated learning for cybersecurity: concepts, challenges, and future directions. *IEEE Trans. Industrial Inf.* 18, 3501–3509. doi:10.1109/TII.2021.3119038

Alert, I.-C. (2016). "Cyber-attack against Ukrainian critical infrastructure," in *Cybersecurity infrastruct. Secur. Agency* (Washington, DC, USA). Tech. Rep. ICS Alert (IR-ALERT-H-16-056-01).

Arpilleda, J. Y. (2023). Cybersecurity in the smart grid: vulnerabilities, threats, and countermeasures. *Int. J. Adv. Res. Sci. Commun. Technol.* 3, 743–750. doi:10.48175/ijarsct-12364

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable artificial intelligence (xai): concepts, taxonomies, opportunities and challenges toward responsible ai. *Inf. Fusion* 58, 82–115. doi:10.1016/j.inffus.2019.12.012

Asesh, A., and Dugar, M. (2023). "Computational optimizations in llms," in *2023 IEEE international conference on machine learning and applied network technologies (ICMLANT)* (IEEE), 1–5.

Baezner, M., and Robin, P. (2017). *Stuxnet*. Zurich, Switzerland: Center for Security Studies (CSS), ETH Zürich. doi:10.3929/ethz-b-000200661

Berghout, T., Benbouzid, M., and Muyeen, S. (2022). Machine learning for cybersecurity in smart grids: a comprehensive review-based study on methods, solutions, and prospects. *Int. J. Crit. Infrastructure Prot.* 38, 100547. doi:10.1016/j.ijcip.2022.100547

Bhamare, D., Zolanvari, M., Erbad, A., Jain, R., Khan, K., and Meskin, N. (2020). Cybersecurity for industrial control systems: a survey. *Comput. and Secur.* 89, 101677. doi:10.1016/j.cose.2019.101677

Bhatt, M., Chennabasappa, S., Li, Y., Nikolaidis, C., Song, D., Wan, S., et al. (2024). *Cyberseceval 2: a wide-ranging cybersecurity evaluation suite for large language models*. arXiv preprint arXiv:2404.13161.

Biggio, B., and Roli, F. (2018). Wild patterns: ten years after the rise of adversarial machine learning. *Pattern Recognit.* 84, 317–331. doi:10.1016/j.patcog.2018.07.023

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Adv. neural Inf. Process. Syst.* 33, 1877–1901. doi:10.48550/arXiv.2005.14165

Browne, T. O., Abedin, M., and Chowdhury, M. J. M. (2024). A systematic review on research utilising artificial intelligence for open source intelligence (opsin) applications. *Int. J. Inf. Secur.* 23, 2911–2938. doi:10.1007/s10207-024-00868-2

Budka, K. C., Deshpande, J. G., Thottan, M., Budka, K. C., Deshpande, J. G., and Thottan, M. (2014). A communication network architecture for the smart grid. *Commun. Netw. Smart Grids Mak. Smart Grid Real*, 149–167. doi:10.1007/978-1-4471-6302-2_6

Cao, G., Jia, R., and Dang, J. (2022). Distributed resilient mitigation strategy for false data injection attack in cyber-physical microgrids. *Front. Energy Res.* 10, 845341. doi:10.3389/fenrg.2022.845341

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H.P.D.O., Kaplan, J., et al. (2021). *Evaluating large language models trained on code*. arXiv preprint arXiv:2107.03374.

Chen, P., Desmet, L., and Huygens, C. (2014). "A study on advanced persistent threats," in *Communications and multimedia security: 15th IFIP TC 6/TC 11 international conference, CMS 2014, aveiro, Portugal, september 25-26, 2014. Proceedings 15* (Springer), 63–72.

Ding, J., Qammar, A., Zhang, Z., Karim, A., and Ning, H. (2022). Cyber threats to smart grids: review, taxonomy, potential solutions, and future directions. *Energies* 15, 6799. doi:10.3390/en15186799

Divakaran, D. M., and Peddinti, S. T. (2024). *Large language models for cybersecurity: new opportunities*. IEEE Security and Privacy, 2–9. doi:10.1109/MSEC.2024.3504512

Dutta, S., Sadhu, P. K., Cherikuri, M., and Mohanta, D. K. (2021). Application of artificial intelligence and machine learning techniques in island detection in a smart grid. *Intell. Renew. Energy Syst.*, 79–109. doi:10.1002/9781119786306.ch3

Easterly, J., and Fanning, T. (2023). The attack on colonial pipeline: what we've learned and what we've done over the past two years. *CISA*. Available online at:

https://www.cisa.gov/news-events/news/attack-colonial-pipeline-what-weve-learned-what-weve-done-over-pasttwo-years (Accessed May 7, 2023).

Elkhorchani, H., Idoudi, M., and Grayaa, K. (2013). "Development of communication architecture for intelligent energy networks," in *2013 international conference on electrical engineering and software applications* (IEEE), 1–6.

El Mrabet, Z., Kaabouch, N., El Ghazi, H., and El Ghazi, H. (2018). Cyber-security in smart grid: survey and challenges. *Comput. and Electr. Eng.* 67, 469–482. doi:10.1016/j.compeleceng.2018.01.015

Fakhouri, H. N., Alhadidi, B., Omar, K., Makhadmeh, S. N., Hamad, F., and Halalsheh, N. Z. (2024). "Ai-driven solutions for social engineering attacks: detection, prevention, and response," in *2024 2nd international conference on cyber resilience (ICCR)* (IEEE), 1–8.

Ferrag, M. A., Alwahedi, F., Battah, A., Cherif, B., Mechri, A., and Tihanyi, N. (2024). *Generative ai and large language models for cyber security: all insights you need*. arXiv preprint arXiv:2405.12750.

Gajanan, L. S., and Kirar, M. (2022). "Cyber-attacks on smart grid system: a review," in *2022 IEEE 10th power India international conference (PIICON)* (IEEE), 1–6.

Gao, Z., Illindala, M., and Wang, J. (2023). "Intrusion detection system with updated structure and feature selection algorithm for man-in-the-middle attacks in the smart grid," in *2023 IEEE industry applications society annual meeting (IAS)* (IEEE), 1–6.

Geiger, M., Bauer, J., Masuch, M., and Franke, J. (2020). "An analysis of black energy 3, crashoverride, and trisis, three malware approaches targeting operational technology systems," in *2020 25th IEEE international conference on emerging technologies and factory automation (ETFA)* (IEEE), 1, 1537–1543. doi:10.1109/etfa46521.2020.9212128

Gentry, C. (2009). "Fully homomorphic encryption using ideal lattices," in *Proceedings of the 41st annual ACM symposium on theory of computing*.

Ghiasi, M., Niknam, T., Wang, Z., Mehrandezh, M., Dehghani, M., and Ghadimi, N. (2023). A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: past, present and future. *Electr. Power Syst. Res.* 215, 108975. doi:10.1016/j.epsr.2022.108975

Ghimire, S., and Thapaliya, S. (2024). Al-driven cybersecurity: mitigating prompt injection attacks through adversarial machine learning. *NPRC J. Multidiscip. Res.* 1, 63–69. doi:10.3126/nprcjmr.v1i8.73029

Giles, M. (2019). *Triton is the world's most murderous malware, and it's spreading*. MIT Technology Review.

Gopstein, A., Nguyen, C., O'Fallon, C., Hastings, N., Wollman, D., et al. (2021). *NIST framework and roadmap for smart grid interoperability standards, release 4.0*. Gaithersburg, Department of Commerce. National Institute of Standards and Technology, 10.

Guastalla, M., Li, Y., Hekmati, A., and Krishnamachari, B. (2023). "Application of large language models to ddos attack detection," in *International conference on security and privacy in cyber-physical systems and smart vehicles* (Springer), 83–99.

Gunduz, M. Z., and Das, R. (2020). Cyber-security on smart grid: threats and potential solutions. *Comput. Netw.* 169, 107094. doi:10.1016/j.comnet.2019.107094

Habib, A. A., Hasan, M. K., Hassan, R., Islam, S., Thakkar, R., and Vo, N. (2023). Distributed denial-of-service attack detection for smart grid wide area measurement system: a hybrid machine learning technique. *Energy Rep.* 9, 638–646. doi:10.1016/j.egyr.2023.05.087

Huang, J., and Zhu, Q. (2023). "Penheal: a two-stage llm framework for automated pentesting and optimal remediation," in *Proceedings of the workshop on autonomous cybersecurity*, 11–22.

Ibrahim, N., Aboulela, S., Ibrahim, A., and Kashef, R. (2024). A survey on augmenting knowledge graphs (kgs) with large language models (llms): models, evaluation metrics, benchmarks, and challenges. *Discov. Artif. Intell.* 4, 76. doi:10.1007/s44163-024-00175-8

Ilokanuno, O. A. (2024). Smart meter tampering detection using IoT based unsupervised machine learning. *Int. J. Res. Appl. Sci. Eng. Technol.* 12, 5434–5445. doi:10.22214/ijraset.2024.61153

Jawhar, S., Miller, J., and Bitar, Z. (2024). "Ai-based cybersecurity policies and procedures," in *Proceedings of the 2024 IEEE 3rd international conference on AI in cybersecurity (ICAIC)* (IEEE), 1–5. doi:10.1109/ICAIC60265.2024.10433845

Jha, R. K. (2023). Strengthening smart grid cybersecurity: an in-depth investigation into the fusion of machine learning and natural language processing. *J. Trends Comput. Sci. Smart Technol.* 5, 284–301. doi:10.36548/jtcsst.2023.3.005

Ji, H., Yang, J., Chai, L., Wei, C., Yang, L., Duan, Y., et al. (2024). *Sevenllm: benchmarking, eliciting, and enhancing abilities of large language models in cyber threat intelligence. arXiv preprint arXiv:2405.03446*.

Jin, S. (2024). False data injection attack against smart power grid based on incomplete network information. *Electr. Power Syst. Res.* 230, 110294. doi:10.1016/j.epsr.2024.110294

Kenton, J. D. M.-W. C., and Toutanova, L. K. (2019). "Bert: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of naacL-HLT (minneapolis, Minnesota)*, 1.

Khare, U., Malviya, A., Gawre, S. K., and Arya, A. (2023). "Cyber physical security of a smart grid: a review," in *2023 IEEE international students' conference on electrical, electronics and computer science (SCEECS)* (IEEE), 1–6.

Kumar, D. K., Reddy, K. K., and Kathrine, G. J. W. (2024). "Smart grid protection with ai and cryptographic security," in *2024 3rd international conference on applied artificial intelligence and computing (ICAAIC)* (IEEE), 246–251.

Kush, N., Clark, A. J., and Foo, E. (2010). Smart grid test bed design and implementation

Lakhani, A., and Rohit, N. (2024). "Securing machine learning: understanding adversarial attacks and bias mitigation," in *International journal of innovative science and research technology (IJISRT)*, 2316–2342. doi:10.38124/ijisrt/IJISRT24JUN1671

Li, H., Chi, H., Liu, M., and Yang, W. (2024a). *Look within, why llms hallucinate: a causal perspective*. arXiv preprint arXiv:2407.10153.

Li, J., Yang, Y., and Sun, J. (2024b). *Risks of practicing large language models in smart grid: threat modeling and validation*. arXiv preprint arXiv:2405.06237.

Liu, F., Jiang, J., Lu, Y., Huang, Z., and Jiang, J. (2025). The ethical security of large language models: a systematic review. *Front. Eng. Manag.*, 1–13. doi:10.1007/s42524-025-4082-6

Liu, X.-Y., Zhang, J., Wang, G., Tong, W., and Walid, A. (2024). "Efficient pretraining and finetuning of quantized llms with low-rank structure," in *2024 IEEE 44th international conference on distributed computing systems (ICDCS)* (IEEE), 300–311.

Liu, Y., Fu, Y., Wang, T., Cao, Y., Yan, J., and Snoussi, H. (2023). "A survey on cyber-security in smart grid," in *2023 China automation congress (CAC)*, 8462–8468. doi:10.1109/CAC59555.2023.10451787

Mashima, D., and An, B. (2017). "Honeypot-enabled optimal defense strategy selection for smart grids," in *Proceedings of the 2017 IEEE conference on communications and network security (CNS)*, 1–9doi. doi:10.1109/CNS.2017.8228660

Mashima, D., and An, B. (2019). "Who's scanning our smart grid? empirical study on honeypot data," in *Proceedings of the 2019 IEEE conference on communications and network security (CNS)*, 1–9doi. doi:10.1109/CNS.2019.8802710

Mishra, S., Albarakati, A., and Sharma, S. K. (2022). Cyber threat intelligence for iot using machine learning. *Processes* 10, 2673. doi:10.3390/pr10122673

Mobini, E., Abolmasoumi, A. H., and Daeichian, A. (2024). Online model-free cyber attack detection in smart grid using dynamic mode decomposition. *IEEE Trans. Netw. Sci. Eng.* 11, 4305–4314. doi:10.1109/tnse.2024.3416964

Mudassar Yamin, M., Hashmi, E., Ullah, M., and Katt, B. (2024). Applications of llms for generating cyber security exercise scenarios. *IEEE Access* 12, 143806–143822. doi:10.1109/ACCESS.2024.3468914

Mukherjee, S., Hossain, R. R., Liu, Y., Du, W., Adetola, V., Mohiuddin, S. M., et al. (2023). "Enhancing cyber resilience of networked microgrids using vertical federated reinforcement learning," in *Proceedings of the 2023 IEEE power and energy society general meeting (PESGM) (IEEE)*, 1–5. doi:10.1109/PESGM52003.2023.10252480

Muneeswari, G., Rose, R. M., Balaganesh, S., Prasath, G. J., and Chellam, S. (2024). Mitigation of attack detection via multi-stage cyber intelligence technique in smart grid. *Meas. Sensors* 33, 101077. doi:10.1016/j.measen.2024.101077

Naiho, H. N. N., Layode, O., Adeleke, G. S., Udeh, E. O., and Labake, T. T. (2024). Addressing cybersecurity challenges in smart grid technologies: implications for sustainable energy infrastructure. *Eng. Sci. and Technol. J.* 5, 1995–2015. doi:10.51594/estj.v5i6.1218

Nakhleh, S., Qasaimeh, M., and Qasaimeh, A. (2024). "Character-level adversarial attacks evaluation for arabert's," in *2024 15th international conference on information and communication systems (ICICS)* (IEEE), 1–6.

Naqvi, S. S. A., Li, Y., and Uzair, M. (2024). Ddos attack detection in smart grid network using reconstructive machine learning models. *PeerJ Comput. Sci.* 10, e1784. doi:10.7717/peerj-cs.1784

Noorwali, A., Hamed, A., Rao, R., and Shami, A. (2016). "Modeling and delay analysis of wireless hans in smart grids over fading channels subjected to multiple access schemes and interference," in *2016 IEEE Canadian conference on electrical and computer engineering (CCECE)* (IEEE), 1–6.

Padilla, E., Agbossou, K., and Cardenas, A. (2014). Towards smart integration of distributed energy resources using distributed network protocol over ethernet. *IEEE Trans. Smart Grid* 5, 1686–1695. doi:10.1109/tsg.2014.2303857

Pandey, J. C., and Kalra, M. (2022). A review of security concerns in smart grid. *Innovative Data Commun. Technol. Appl. Proc. ICIDCA* 2021, 125–140. doi:10.1007/978-981-16-7167-8_10

Pearce, H., Ahmad, B., Tan, B., Dolan-Gavitt, B., and Karri, R. (2025). Asleep at the keyboard? assessing the security of github copilot's code contributions. *Communications of the ACM*, 68(2), 96–105. doi:10.1145/3610721

Perković, G., Drobnjak, A., and Botički, I. (2024). "Hallucinations in llms: understanding and addressing challenges," in *2024 47th MIPRO ICT and electronics convention (MIPRO)* (IEEE), 2084–2088.

Piggott, B., Patil, S., Feng, G., Odat, I., Mukherjee, R., Dharmalingam, B., et al. (2023). "Net-gpt: a ll-empowered man-in-the-middle chatbot for unmanned aerial vehicle," in *2023 IEEE/ACM symposium on edge computing (SEC)* (IEEE), 287–293.

Pourahmad, Z., and Hooshmand, R.-A. (2023). "Smart grid protection against cyber-attacks using pus and dc system model," in *2023 13th smart grid conference (SGC)* (IEEE), 1–8.

Qureshi, M. S., Khan, I. U., and Kim, K. (2023). "Securing the smart grid: a comprehensive analysis of recent cyber attacks," in *2023 5th international conference on electrical, control and instrumentation engineering (ICECIE)* (IEEE), 1–6.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 1–67. doi:10.48550/arXiv.1910.10683

Rahiminejad, A., Plotnek, J., Atallah, R., Dubois, M.-A., Malatrait, D., Ghafouri, M., et al. (2023). A resilience-based recovery scheme for smart grid restoration following cyberattacks to substations. *Int. J. Electr. Power and Energy Syst.* 145, 108610. doi:10.1016/j.ijepes.2022.108610

Rahul, R., Sindhu, P., Sundar, G. N., and Venkatesan, R. (2024). Fusing Deep Learning Techniques For Intrusion Detection in Smart Grids. *Fusion: Practice and Applications*, 16(1)). doi:10.54216/FPA.160105

Ruan, J., Liang, G., Zhao, J., Zhao, H., Qiu, J., Wen, F., et al. (2023). Deep learning for cybersecurity in smart grids: review and perspectives. *Energy Convers. Econ.* 4, 233–251. doi:10.1049/enc2.12091

Salvadori, F., Gehrke, C. S., de Oliveira, A. C., de Campos, M., and Sausen, P. S. (2013). Smart grid infrastructure using a hybrid network architecture. *IEEE Trans. Smart Grid* 4, 1630–1639. doi:10.1109/tsg.2013.2265264

Shahid, M. A., Ahmad, F., Nawaz, R., Khan, S. U., Wadood, A., and Albalawi, H. (2023). A novel false measurement data detection mechanism for smart grids. *Energies* 16, 6614. doi:10.3390/en16186614

Sharma, V., Batra, R., and Prabhu, A. (2023). "Accurate attack preventing system implementation in grids," in *2023 international conference on power energy, environment and intelligent control (PEEIC)* (IEEE), 1400–1403.

Simonthomas, S., Subramanian, R., and Mathiew, S. A. (2024). "A survey of enhancing cyber physical system security in smart grid," in *2024 international conference on communication, computing and Internet of Things (IC3IoT)* (IEEE), 1–6.

Singh, J., Kaur, S., Kaur, G., and Kaur, G. (2016). A detailed survey and classification of commonly recurring cyber attacks. *Int. J. Comput. Appl.* 141, 15–19. doi:10.5120/ijca2016909811

Singhal, D., Som, S., and Ahuja, L. (2021). "A review: iot based smart grid," in *2021 9th international conference on reliability, infocom technologies and optimization (trends and future directions) (ICRITO)*, 1–3. doi:10.1109/ICRITO51393.2021.9596157

Sriranjani, R., Saleem, M.D., Hemavathi, N., and Parvathy, A. (2023), "Machine learning based intrusion detection scheme to detect replay attacks in smart grid." in *2023 IEEE international students' Conference on electrical, Electronics and computer science (SCEECS)* (IEEE), 1–5.

Sweeten, J., Takiddin, A., Ismail, M., Refaat, S. S., and Atat, R. (2023). "Cyber-physical gnn-based intrusion detection in smart power grids," in *2023 IEEE international conference on communications, control, and computing technologies for smart grids (SmartGridComm)* (IEEE), 1–6.

Tala, T. K., Hadjar, O. S., and Naima, K. (2022). Cyber-security of smart grids: attacks, detection, countermeasure techniques, and future directions. *Commun. Netw.* 14, 119–170. doi:10.4236/cn.2022.144009

Tatipatri, N., and Arun, S. (2024). A comprehensive review on cyber-attacks in power systems: impact analysis, detection and cyber security. *IEEE Access* 12, 18147–18167. doi:10.1109/access.2024.3361039

Tseng, P., Yeh, Z., Dai, X., and Liu, P. (2024). *Using llms to automate threat intelligence analysis workflows in security operation centers*. arXiv preprint arXiv:2407.13093.

Uma, M., and Padmavathi, G. (2013). A survey on various cyber attacks and their classification. *Int. J. Netw. Secur.* 15, 390–396.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. neural Inf. Process. Syst.* 30. doi:10.48550/arXiv.1706.03762

Vidya, A. M., Sai, D. D., Sarveshwaran, G., Mukesh, S., and Chandrakala, K. V. (2023). "Identification of false data injection and man in the middle cyber-attacks impact on smart grid," in *2023 third international conference on secure cyber computing and communication (ICSCCC)* (IEEE), 678–683.

Walton, R. (2019). First cyberattack on solar, wind assets revealed widespread grid weaknesses, analysts say. *Utility. Dive*. Available at: https://www.utilitydive.com/news/first-cyber-attack-on-solar-wind-assets-revealed-widespread-grid-weaknesse/566505. 4.

Wang, H., Wen, X., Xu, Y., Zhou, B., Peng, J., and Liu, W. (2020). Operating state reconstruction in cyber physical smart grid for automatic attack filtering. *IEEE Trans. Industrial Inf.* 18, 2909–2922. doi:10.1109/tii.2020.3000172

Wang, L., Wang, J., Jung, K., Thiagarajan, K., Wei, E., Shen, X., et al. (2024a). *From sands to mansions: enabling automatic full-life-cycle cyberattack construction with llm*. arXiv preprint arXiv:2407.16928.

Wang, T., Xie, X., Zhang, L., Wang, C., Zhang, L., and Cui, Y. (2024b). "Shieldgpt: an llm-based framework for ddos mitigation," in *Proceedings of the 8th asia-pacific workshop on networking*, 108–114.

Xiaocheng, W., Qiaoni, H., Yuan, Y., and Ma, H. (2023). Energy-efficient data transmission with proportional rate fairness for nans of smart grid communication network. *EURASIP J. Adv. Signal Process.* 2023, 43. doi:10.1186/s13634-023-01007-0

Yadav, G., and Paul, K. (2021). Architecture and security of scada systems: a review. *Int. J. Crit. Infrastructure Prot.* 34, 100433. doi:10.1016/j.ijcip.2021.100433

Yadav, S. A., Kumar, S. R., Sharma, S., and Singh, A. (2016). "A review of possibilities and solutions of cyber attacks in smart grids," in *2016 international conference on innovation and challenges in cyber security (ICICCS-INBUSH)* (IEEE), 60–63.

Yi, P., Zhu, T., Zhang, Q., Wu, Y., and Li, J. (2014). "A denial of service attack in advanced metering infrastructure network," in *2014 IEEE international conference on communications (ICC)* (IEEE), 1029–1034.

Zaboli, A., Choi, S. L., Song, T.-J., and Hong, J. (2024a). "Chatgpt and other large language models for cybersecurity of smart grid applications," in *2024 IEEE power and energy society general meeting (PESGM)* (IEEE), 1–5. doi:10.1109/PESGM51994.2024.10688863

Zaboli, A., Choi, S. L., Song, T.-J., and Hong, J. (2024b). A novel generative ai-based framework for anomaly detection in multicast messages in smart grid communications. *arXiv Prepr. arXiv:2406.05472*. doi:10.48550/arXiv.2406.05472

Zhang, J., and Li, C. (2019). Adversarial examples: opportunities and challenges. *IEEE Trans. neural Netw. Learn. Syst.* 31, 2578–2593. doi:10.1109/TNNLS.2019.2933524

Zhang, Y., Wang, J., and Chen, B. (2020). Detecting false data injection attacks in smart grids: a semi-supervised deep learning approach. *IEEE Trans. Smart Grid* 12, 623–634. doi:10.1109/tsg.2020.3010510

Zhang, Y., Wang, J., and Chen, B. (2021). Detecting false data injection attacks in smart grids: a semi-supervised deep learning approach. *IEEE Trans. Smart Grid* 12, 623–634. doi:10.1109/TSG.2020.3010510

Zhang, Z., Liu, M., Sun, M., Deng, R., Cheng, P., Niyato, D., et al. (2024). Vulnerability of machine learning approaches applied in iot-based smart grid: a review. *IEEE Internet Things J.* 11, 18951–18975. doi:10.1109/jiot.2024.3349381

Zhu, Y., Ruan, J., Fan, G., Wang, S., Liang, G., and Zhao, J. (2022). "A generalized data recovery model against false data injection attack in smart grid," in *2022 IEEE 6th conference on energy Internet and energy system integration (EI2)* (IEEE), 1477–1482.

Zhuang, P., Zamir, T., and Liang, H. (2021). Blockchain for cybersecurity in smart grid: a comprehensive survey. *IEEE Trans. Industrial Inf.* 17, 3–19. doi:10.1109/tii.2020.2998479