Cybersecurity in Digital Twins and AI: An Intelligent Framework for Proactive Threat Detection and Response

Elena Alcaraz *, Oscar Martinez , Patricia Sanchez

Posted Date: 11 September 2025

doi: 10.20944/preprints202509.0994.v1

Keywords: digital twins; cybersecurity; artificial intelligence; threat detection; anomaly detection; cyber-physical systems; federated learning

Article

Cybersecurity in Digital Twins and AI: An Intelligent Framework for Proactive Threat Detection and Response

Elena Alcaraz *, Óscar Martínez and Patricia Sánchez

Department of Computer Science and Artificial Intelligence, Universidad de La Laguna (ULL), Tenerife, Spain
* Correspondence: ealcaraz@ull.es

Abstract

Digital twins (DTs) combined with artificial intelligence (AI) are creating a big change in cybersecurity, by giving new abilities for detecting and reacting to threats in cyber-physical systems in a more proactive way. This paper introduces a detailed framework that uses AI-based digital twins for real-time security observation, anomaly detection, and prediction of possible cyber threats. The method we propose mixes machine learning techniques with digital twin systems to develop autonomous security solutions that are able to learn continuously and adapt to new situations. This integration helps to deal with important cybersecurity problems such as data accuracy, privacy protection, and the ability to grow in Industry 4.0 systems. We have tested our approach with real-world data and the results show good improvements in threat detection accuracy (99.2%) and also in reducing the response time (67%). The framework we suggest can serve as a strong basis for the future of cybersecurity, where systems can predict, model, and reduce cyber threats in real-time.

Keywords: digital twins; cybersecurity; artificial intelligence; threat detection; anomaly detection; cyber-physical systems; federated learning

1. Introduction

The fast development of digital transformation together with Industry 4.0 is changing in a deep way how companies think about cybersecurity. Cyber-physical systems are now more and more connected, which also makes them more open to advanced and smart attacks [1,2]. The old security methods, which mostly react after the attack happens, are not enough anymore to fight against the new kinds of persistent threats. For example, it takes companies about 199 days to find a security problem, and 73 more days to control it [3]. Because of this, combining digital twins with artificial intelligence is becoming a very important and new method. It gives the chance to make security more proactive, predictive, and even automatic. Also, as we move closer to Industry 5.0 and future technologies like 6G, it becomes even more necessary to protect digital twin systems from advanced cyber attacks [7].

The digital twin technology was first made by NASA for space systems, but now it has become a strong framework to make virtual copies of real systems. These digital twins allow for real-time observation, simulation, and also improving the performance of physical systems [4]. When we combine digital twins with AI technologies, it opens many new possibilities for cybersecurity. With features like continuous learning, recognizing patterns, and predicting threats, this combination can improve security in a big way. According to new market research, the digital twin simulation market could grow to $379 billion by the year 2034. This is mostly because of the fast integration of IoT, AI, and cloud technologies [5].

Today's cybersecurity systems are having many problems with the fast-changing cyber threat environment. This includes things like zero-day attacks, strong long-term threats, and even attacks made using AI. One of the main problems is the lack of standard security rules for digital twin systems,

 doi:10.20944/preprints202509.0994.v1

which makes them weak and easy targets for hackers [6]. New opinions from international standard groups are now highlighting the urgent need to build special security frameworks that are made only for digital twin technology [8].

In addition, as cyber-physical systems are becoming more complex, we also need smarter security tools that can understand how different system parts work together and behave. The special double nature of digital twins makes the security even harder, especially in industrial systems. Here, both the physical and digital parts have their own risks, so we must build protection methods that can cover both sides together [9].

The gap between traditional cybersecurity methods and the new types of cyber threats is getting bigger, and this situation needs new solutions that can change, learn, and grow together with the changing attack techniques. In this research, we try to solve important problems in today's security systems by introducing a digital twin architecture improved by artificial intelligence, which is made specially for cybersecurity uses. Our method brings together real-time threat simulation, prediction using data analysis, and automatic response actions to build a complete defense system.

In this paper, we show a new framework that combines AI-based cybersecurity functions with digital twin technology. We demonstrate that this combination can greatly improve how fast we detect threats, how quickly the system can react, and how strong and flexible the system becomes. The organization of this paper is as follows: in Section 2, we talk about previous works and current advanced methods; in Section 3, we explain our proposed method and give the mathematical details; Section 4 shows the experiments and the performance analysis; Section 5 discusses the results and also the limitations; and finally, in Section 6, we give the conclusion and ideas for future research.

2. Related Works

2.1. Digital Twin Architectures for Cybersecurity

Several researchers have studied how digital twin technology can be used in cybersecurity, using different methods for architecture design and practical implementation. Eckhart and Ekelhart [10] introduced the idea of security-aware virtual environments for digital twins, which became an important base for building secure digital twin systems. In their research, they focused on keeping the physical and digital parts of the system in high accuracy, while also making sure that security isolation is applied. Later, Dietz and Pernul [11] suggested a complete digital twin model for enterprise cybersecurity. Their framework showed how digital twins can help improve security actions by using simulation, data analysis, and system copying techniques. As we move forward to Industry 5.0 and also the future of 6G-connected systems, the need to protect digital twin networks from advanced cyber attacks becomes even more serious [7].

More recently, Itäpelto et al. [14] designed a reference architecture for cybersecurity digital twins. Their work gives a standardized structure that can be used in different organizations. This architecture focuses on solving important problems like system compatibility, growing system size, and adding strong security parts. These earlier studies help form the theoretical background for our proposed framework, and they also show some open problems, especially in real-time data handling and using AI more effectively. Also, many researchers agree that situational awareness in cyber-physical systems is very important for managing security and reacting quickly to threats [15].

2.2. AI-Enhanced Threat Detection Systems

The use of artificial intelligence in cybersecurity has changed a lot during the last ten years. Machine learning algorithms have shown better results in finding threats compared to old rule-based systems. This change is because of the fast increase in cyber attacks and the more complex ways attackers use, which old security methods cannot manage well. Sarker et al. [16] gave detailed classifications of explainable AI methods for cybersecurity, showing how important it is to have models that people can understand when making security decisions. Their important work showed

 doi:10.20944/preprints202509.0994.v1

that explainable AI can build more trust and openness in AI security systems, which is one of the main worries about using "black-box" AI models in important security areas.

Using explainable AI in cybersecurity is a big change from secretive decision-making to clear and understandable systems that security experts can check and trust. This is very important in big companies where security choices must be clear and explainable to managers and law regulators. Also, explainable AI helps security teams to learn from the system's decisions, which makes both human skills and overall security better.

Luo et al. [17] did wide surveys about deep learning use in cyber-physical systems, showing big improvements in finding unusual activities and faster reactions in different industries. Their study showed that deep learning can find small signals in network traffic and system behavior that older statistical methods often miss. Their research covered many areas such as smart grids, self-driving cars, and industrial control systems, proving AI security methods work in many places. More recently, Adjei et al. [18] combined K-nearest neighbors algorithms with digital twins to improve detection of network problems, reaching better precision and recall scores for unbalanced network data, which is a common problem in real cybersecurity cases.

The joining of digital twin technology and artificial intelligence is opening new ways for active cybersecurity management. Digital twins make virtual copies of real systems, so security teams can practice attack scenarios and test defenses without risking real equipment. When AI is added, these virtual models become strong tools for predicting threats and organizing automatic responses [1].

Important earlier research on AI in digital twin cybersecurity created basic rules for making smart security systems that can detect and respond to threats by themselves [1,19]. These rules include real-time data updating between real and virtual systems, learning methods that get better over time to detect threats, and automatic response actions that stop attacks before they cause damage. These basic frameworks help develop new cybersecurity solutions that use AI's power to predict with digital twins' wide monitoring ability.

2.3. Anomaly Detection in Cyber-Physical Systems

Anomaly detection is a very important part of modern cybersecurity systems, especially in cyber-physical environments where old signature-based methods are not enough to deal with the complex and changing threats today. These signature-based methods show their limits when facing zero-day attacks, advanced long-lasting threats, and smart attack plans that use new techniques to avoid being detected. In cyber-physical systems, where digital and physical parts come together and create new weak points, anomaly detection must find unusual changes not only in network traffic but also in physical processes and sensor data that may show signs of attack.

Detecting anomalies is more difficult because the data comes from many different sources, such as network messages, system logs, sensor values, and control signals. Traditional statistical methods often cannot handle this mixed data well, causing many false alarms and missing small but important anomalies. Xu et al. [21] suggested the new LATTICE framework, which uses curriculum learning ideas to improve anomaly detection with digital twins in cyber-physical systems. Their important work showed better results in finding complex attack behaviors while lowering the false alarm rates that have been a problem in industrial anomaly detection.

The LATTICE framework's use of curriculum learning is a big step forward in how anomaly detection systems are trained. By slowly giving training examples from easy to hard, the system learns better what is normal and becomes stronger at telling harmless differences from real threats. This is very useful in cyber-physical systems, where normal behavior and attack signs can be very close and depend on the situation.

Recent studies showed the great success of digital twin-based anomaly detection for industrial use, showing strong potential for better threat finding in critical infrastructure where security failures can cause very serious damage beyond just data loss [20]. These works explained that digital twins help security systems keep full awareness by watching both the real physical processes and their

 doi:10.20944/preprints202509.0994.v1

digital copies, which helps find differences that might mean cyber attacks on operational technology systems.

Using digital twin-assisted anomaly detection in industry solves important problems that old cybersecurity methods cannot solve well. These problems include understanding normal working patterns of many different industrial processes, finding attacks that only show as small changes in physical measurements, and keeping security monitoring without stopping important industrial work. Digital twins also give a safe place to test and check anomaly detection methods before using them in real systems.

Wang et al. [22] created smart methods to combine space and time features for finding IoT attack behavior, using advanced digital twin technology to better recognize patterns in large sensor networks. Their new method solved big challenges about changing network structures and fast new attack ways in IoT systems. The space-time method understands that attacks in IoT usually happen as planned actions across many devices and time moments, so the detection system needs to connect events in both space and time.

This spatio-temporal feature fusion method is a new way of thinking about IoT security, moving from watching single devices to watching the whole system. This lets the system find smart attack plans that try to take control of many devices one by one or all at once to do bad things. Adding digital twins makes this better by giving a virtual space to model and study how IoT devices work together in real time [22].

These important previous works build a strong foundation for the advanced anomaly detection abilities in our proposed framework. They show how important it is to combine many useful technologies like machine learning, digital twins, and advanced data fusion. The joining of these technologies gives new chances to make next-generation cybersecurity systems that can protect complex cyber-physical systems from smart and changing threats while still keeping industrial systems working well and reliably.

2.4. Industrial Control Systems Security

Industrial control systems (ICS) have special and very hard security problems because they need to work continuously and often use old equipment that was made mainly for function and reliability, not for cybersecurity. These systems, like supervisory control and data acquisition (SCADA), distributed control systems (DCS), and programmable logic controllers (PLCs), are very important parts of critical sectors such as power plants, water treatment, factories, and transportation networks. The difficulty in securing these systems comes from many reasons: they must run all time without stopping, they use decades-old devices that are hard to update or replace, IT and OT networks are now connected together, and if security measures cause problems, it could lead to very serious damage.

Traditional ICS security often depends on air-gapping and separating networks, but with Industry 4.0, remote monitoring, and need for better efficiency, these isolation methods are less useful now. Modern ICS must carefully balance keeping systems running, safety rules, laws, and security protection, making the security environment very complex and needing new ways to detect and respond to threats.

Varghese et al. [23] showed a new method using digital twins for intrusion detection specially made for ICS. Their work explained how virtual copies can greatly improve security monitoring without interrupting important processes that cannot stop. They showed the main need to keep systems working while still having full security coverage, solving one of the hardest problems in ICS security where normal security tools often conflict with operation needs. Digital twins allow security teams to use advanced monitoring and analysis in a virtual system that copies the real system's behavior, so threats can be found in real time without slowing down or disturbing the real industrial systems.

This new method solves many big problems of traditional ICS security. Normal intrusion detection systems have trouble with special communication protocols, time-sensitive tasks, and special hardware in industrial environments. Digital twins fix these problems by making exact virtual models that can be watched and checked by advanced security tools without affecting real system's work or trust.

 doi:10.20944/preprints202509.0994.v1

Using digital twin security monitoring also brings other benefits, like being able to simulate attacks, test security settings, and train workers on how to respond to threats safely without risking real operations. This is very important for critical infrastructure where training and testing must not cause safety or availability problems.

Krishnaveni et al. [24] developed TwinSec-IDS, a better intrusion detection system made for software-defined network (SDN) digital twin-based industrial cyber-physical systems. This work is an important step in combining modern networking technology with cybersecurity solutions. They showed big improvements in how accurately and fast threats are found by smartly mixing SDN and digital twin technologies. This creates a powerful way to use SDN's programmable network with digital twin's prediction ability.

TwinSec-IDS solves special problems from SDN networks that change fast in industrial settings because network setup often changes to meet operation or security needs. Normal intrusion detection often makes mistakes with these changes, either giving false alarms or missing attacks hidden by network changes.

Using SDN and digital twins together in TwinSec-IDS gives new powers that were hard before in industrial cybersecurity. The SDN controller can see all network traffic and quickly change network paths to block threats or send suspicious traffic for checking. The digital twin keeps a correct model of normal network and system behavior, helping find small strange activities that may show smart attack plans.

Also, TwinSec-IDS supports automatic responses that quickly take action to stop attacks without waiting for people. This fast automatic response is very important in industrial places where threats spread quickly and manual response is too slow, which could cause big damage or stop operations. Combining fast threat detection, good behavior modeling, and automatic reaction is a big improvement to protect critical industrial systems from new cyber threats while keeping system working well and reliable.

2.5. Federated Learning and Privacy-Preserving Techniques

The combination of privacy-preserving technologies with digital twins has become very important as many organizations in different sectors try to balance the strong security benefits of digital twin technologies with strict data protection rules such as the General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA), and other industry-specific regulations. This problem is especially serious in industries that manage sensitive personal data, important operational information, or classified materials, where usual centralized data sharing can break privacy laws or create big legal and reputation risks. The need for privacy solutions is even stronger because digital twin systems are growing to include supply chain partners, industry groups, and cross-sector cooperation for security.

A main difficulty between cybersecurity and privacy comes from the fact that good threat detection needs detailed information about systems, user actions, and operations, which often contain sensitive or personal data. Traditional cybersecurity mostly uses centralized data collection and analysis, but this causes risks for privacy breaches and may break rules that limit data movement or storage across borders.

Recent progress in federated learning for digital twin cybersecurity shows very good results in protecting privacy while allowing shared threat intelligence between different organizations [26]. Federated learning changes the old way of central machine learning by letting many participants train shared models together without sharing raw data. For digital twin cybersecurity, this lets organizations get benefits from common threat information while keeping their sensitive data safe inside their own security borders.

Federated learning solves many important problems in shared cybersecurity. First, it helps organizations with less security knowledge or resources to use the combined knowledge and detection skills of bigger, more expert groups. Second, it creates better and more flexible threat detection models by using different data and attack examples from many places without risking data privacy. Third, it helps follow data protection laws by making sure sensitive data does not leave the organization.

 doi:10.20944/preprints202509.0994.v1

Using federated learning in digital twin systems also gives new chances to improve cybersecurity with distributed intelligence. Digital twins can act as local training areas where federated learning algorithms test and improve themselves with synthetic data that looks like real system behavior but does not expose real operational data. This lets organizations join shared threat intelligence projects while fully controlling their private data.

He et al. [27] suggested new federated meta-learning methods for intrusion detection in consumer electronics inside metaverse environments. They showed how distributed learning can greatly improve security while keeping sensitive data private in complex virtual systems. Their work solves special problems in metaverse security, where old methods cannot handle the large scale, variety, and fast changes of virtual interactions with millions of users and many connected devices.

Federated meta-learning is better than normal federated learning because it uses meta-learning ideas to quickly adapt to new threats and changing conditions. This is very useful in metaverse systems where threats change fast due to new technologies, user habits, and interaction types. Meta-learning helps the system update its detection skills when it sees new kinds of attacks or works in new virtual places with different rules.

This research about consumer electronics solves important problems in Internet of Things (IoT) security where devices have limited power and computing but still must join security collaborations without harming user privacy or device performance. The federated method lets these devices share and gain threat intelligence while working inside their limited resources.

These new methods give important ideas for making full privacy-aware security frameworks that can work well in many technical systems and meet very strict privacy rules. Combining federated learning with meta-learning opens new ways for shared cybersecurity that were not possible before because of privacy and data sharing limits.

Recent progress in digital twin community frameworks has clearly solved problems of secure data sharing, giving a strong base for shared threat intelligence work while keeping strong data integrity and privacy across organizations [25]. These community-based approaches understand that many modern threats attack several organizations at once or spread through linked systems, so coordinated defense that goes beyond single organizations is needed.

The idea of digital twin community frameworks is an important step forward in shared cybersecurity. It goes beyond simple information sharing to create joint virtual environments where many organizations can work together on threat detection, analysis, and response while keeping full control over their own sensitive data. These frameworks use advanced cryptography, secure multi-party computation, and blockchain trust methods to make sure that collaboration does not reduce any participant's security or privacy.

Secure data sharing in digital twin communities solves many important problems for cybersecurity cooperation between organizations. These include building trust between participants, keeping data safe during joint analysis, stopping unauthorized access, and making sure activities are recorded for rules and responsibility. These frameworks also have flexible trust systems that change with new threats and participant behavior, making sure cooperation stays safe and useful over time.

2.6. Blockchain Integration and Trust Management

The integration of blockchain technology into digital twin architectures has become a promising method for improving security, trust, and data integrity in distributed cyber-physical systems. Blockchain's immutable and decentralized features help solve key weaknesses in traditional digital twin designs, especially in environments involving multiple stakeholders where building trust is difficult. The cryptographic protections of blockchain safeguard against data tampering and unauthorized changes, which is crucial because digital twin accuracy directly affects the safety and reliability of physical systems.

Liu et al. [28] developed blockchain-based secure communication protocols for digital twins in intelligent transportation systems. Their work showed how distributed ledger technology can strengthen security in complex cyber-physical environments. In transportation, digital twins must

 doi:10.20944/preprints202509.0994.v1

accurately reflect fast-changing conditions while ensuring communication between components is tamper-proof. Since transportation systems are safety-critical, security breaches could cause severe accidents or service failures. Blockchain protocols provide cryptographic verification of messages, distributed data validation, and immutable logging of system activities.

Lv et al. [29] recently proposed blockchain-based centralized learning methods to secure digital twin networks, tackling the challenges of data integrity and trust management in distributed environments. Their approach combines the decentralized security of blockchain with the efficiency of centralized machine learning, achieving higher threat detection accuracy and faster responses than traditional distributed methods.

Blockchain's trust management features are especially useful in digital twin networks where participants have different security policies and business goals. The framework can dynamically assess participant reliability and adjust collaboration based on past behavior and current security status. These advances emphasize blockchain's potential in our proposed security framework.

Recent developments in blockchain-enhanced digital twin security have also been applied to smart water infrastructure. Here, combining AI with distributed ledger technology improves security coordination and trust management [30,31]. Protecting smart water systems is particularly challenging because breaches can have catastrophic effects, and these systems require real-time monitoring over large areas. AI provides advanced threat detection, while blockchain guarantees data integrity and authenticity across all system operations, forming a strong security architecture for critical infrastructure protection [32,33].

3. Methodology

3.1. AI-Enhanced Digital Twin Security Architecture

Our proposed methodology creates a full framework to combine artificial intelligence (AI) with digital twin technology for making autonomous cybersecurity systems. These systems can change and adapt to new threats while protecting important infrastructure all the time. This new method solves problems of old cybersecurity systems that use fixed rules and need people to control them by making self-learning and self-adapting security methods. The framework uses the predicting power of digital twins with the pattern recognition and decision-making skills of AI to find threats early and make automatic responses in complex cyber-physical systems.

The design has four main layers working together to give full security coverage: Physical Asset Layer (PAL), Digital Twin Core (DTC), AI Processing Engine (APE), and Security Orchestration Layer (SOL). Each layer has its own jobs but works closely with others to get best performance and security results.

The Physical Asset Layer (PAL) is the base of our framework. It includes all physical parts like sensors, actuators, control systems, and network devices that need protection. This layer collects real-time data and applies security controls at the physical level. The Digital Twin Core (DTC) makes and keeps accurate virtual models of all physical parts. It allows continuous monitoring and simulations without stopping real operations. The AI Processing Engine (APE) uses advanced machine learning to study data patterns, find anomalies, and predict possible security problems before they affect the physical system. Last, the Security Orchestration Layer (SOL) manages response actions in all layers, making sure security steps are done quickly and well while keeping system stable and working smoothly.

The connection between these layers gives some advanced features not possible in old security methods. The digital twin part lets security systems try response plans in virtual spaces before doing them in real systems, lowering risk of unwanted problems. The AI engine learns from new threats and system behavior all the time, making detection better and reducing false alarms. The orchestration layer makes sure security responses work well between many systems and teams, avoiding conflicts between different security actions and improving total protection.

 doi:10.20944/preprints202509.0994.v1

The math base of our framework is about state synchronization between real and virtual parts. This makes sure digital twins keep correct copies of real systems in real time. This synchronization is very important for predicting security problems and making sure virtual tests match real system conditions. The state synchronization must consider many things like sensor accuracy, communication delays, and system behavior that can change the quality of digital twin models. The math model also uses uncertainty measurement methods to check how reliable digital twin predictions and security advice are, shown as:

$$S_{DT}(t) = f(S_{PA}(t), H(t - \Delta t), M_{AI}(t)) \tag{1}$$

where $S_{DT}(t)$ represents the digital twin state at time t, $S_{PA}(t)$ denotes the physical asset state, $H(t - \Delta t)$ captures historical data patterns, and $M_{AI}(t)$ represents the AI model predictions. The synchronization function f ensures real-time fidelity between physical and virtual components while incorporating predictive intelligence.

3.2. Intelligent Threat Detection Mechanism

The threat detection system uses a complex multi-layered method that carefully combines both supervised and unsupervised learning algorithms. This helps to find abnormal behaviors and possible security threats across many types of attacks and system situations. This mixed method solves the problems of using only one algorithm by using the strong points of different machine learning methods. It gives better detection results and lowers false alarms that can bother security teams with too many unnecessary alerts. The multi-layered design allows the system to find both known attacks using supervised learning and new or unknown threats using unsupervised anomaly detection.

Using many detection algorithms together has important advantages over old single-method ways. Supervised learning parts are good at finding threats that look like known attack patterns, giving high accuracy for known threats but may miss new attack types. Unsupervised learning methods can find strange system behaviors that might show new or changed attacks, but they may have more false alarms because they are sensitive to normal changes in system operation. By mixing these methods, our system gets the good parts of both and reduces their weaknesses.

The detection algorithm combines these different methods using a weighted ensemble approach that changes the importance of each part based on current system status and threat types. This adaptive weighting makes sure the system works well in different situations and with changing attack patterns. The math form of this combined method is written as:

$$T_{score}(x) = \alpha \cdot \text{SVM}(x) + \beta \cdot \text{LSTM}(x) + \gamma \cdot \text{KNN}(x) \tag{2}$$

where $T_{score}(x)$ represents the comprehensive threat confidence score for input vector x, providing a unified metric that quantifies the likelihood of a security threat based on the combined assessment of all detection components. The weighting parameters α, β, and γ are dynamically optimized through reinforcement learning techniques that continuously adapt to changing threat patterns and system characteristics, ensuring that the detection algorithm maintains optimal performance as both legitimate system behavior and attack strategies evolve over time.

The Support Vector Machine (SVM) part is the base for handling known threat patterns by making decision boundaries that separate harmful and safe activities using past training data. SVM works well in high-dimensional feature spaces, so it is good for processing the complex and multi-dimensional data produced by modern cyber-physical systems. The Long Short-Term Memory (LSTM) network handles sequences over time to find attack patterns that happen over longer periods, like advanced persistent threats or multi-stage attacks that other algorithms might miss because they look only at single events or short time frames.

The LSTM's memory allows the system to keep context about earlier system states and actions, so it can detect smart attacks that spread activities over time to avoid being found by normal security systems. The K-Nearest Neighbors (KNN) algorithm finds threats by comparing current system

 doi:10.20944/preprints202509.0994.v1

behavior to known examples of both normal and harmful activities stored in the system's knowledge base.

KNN is very useful for finding attack types that are similar to known threats but do not exactly match old patterns. This lets the system detect changed or evolved attacks that might bypass signature-based detection. Combining these three methods creates a strong threat detection system that can adapt to many attack types while keeping high accuracy and low false alarms in complicated real-world environments.

The anomaly threshold is dynamically adjusted using:

$$\theta(t) = \mu(T_{score}) + k \cdot \sigma(T_{score}) \cdot e^{-\lambda \cdot \Delta t} \tag{3}$$

where μ and σ represent mean and standard deviation of historical threat scores, k is the sensitivity parameter, and λ controls the temporal decay factor.

3.3. Real-Time Security Orchestration

The security orchestration part works as the main coordination center that smartly manages automated response actions based on full evaluation of threat seriousness and system importance across the whole cyber-physical infrastructure. This advanced orchestration solves one of the hardest problems in autonomous cybersecurity systems: making sure security responses match the threats without causing too much disruption to important operations. The orchestration system must balance many goals like reducing threats well, keeping operations running, using resources wisely, and ensuring long-term system strength to make the best decisions in complex and urgent situations.

The orchestration part uses a multi-criteria decision-making system that checks possible response actions on many factors to find the best security steps for each threat case. This method understands that different threats may need very different responses, and the same threat might need different actions depending on the current work situation, system status, and business needs. The system includes real-time checks of system importance, current operations, resource limits, and possible chain reactions caused by different responses.

Because the orchestration system changes with time, it can update response plans based on changing threat levels and operational needs. When threats grow or system conditions change, the orchestration part keeps reviewing all response options and changes its decision rules to keep the best protection while lowering operational problems. This flexible ability is very important in complex cyber-physical systems where fixed rules cannot cover all possible threats and situations.

The response choice method uses a smart optimization process that carefully looks at all possible response actions to find the one that best balances security strength with operational needs. This optimization is written as a multi-objective optimization problem that looks at many goals at the same time, like cost to implement, response speed, and effect on operations. The math form of this optimization is written as:

$$R_{optimal} = \arg\min_{r \in R} [w_1 \cdot C(r) + w_2 \cdot T(r) + w_3 \cdot I(r)] \tag{4}$$

where R represents the comprehensive set of available response actions that the system can implement, ranging from simple alerting mechanisms to complex network reconfiguration and system isolation procedures. The cost function $C(r)$ includes several important factors such as the computational resources needed to carry out the response, the possible financial costs related to the response action, and the opportunity costs from not being able to perform other actions at the same time.

The response time component $T(r)$ measures how long each response action takes, including both the time to implement it and how long its effects last on the system's operation. This timing aspect is especially important in cyber-physical systems where delays or timing issues can affect the success of security measures and the smooth running of critical processes.

 doi:10.20944/preprints202509.0994.v1

The operational impact function $I(r)$ estimates the possible effects of each response on normal system operations, such as changes in system performance, availability, and functionality.

The weighting parameters w_1, w_2, and w_3 allow flexible adjustment of the importance given to each optimization criterion, depending on the current threat environment and changing system priorities. These weights are updated continuously using machine learning techniques that analyze past response success, current threat data, and operational feedback. This adaptive weighting helps the system make the best decisions as threats and operational needs change over time, supporting reliable and autonomous cybersecurity management in complex cyber-physical environments.

3.4. Federated Learning for Collaborative Security

Our framework uses advanced federated learning methods to allow sharing of threat intelligence together while keeping strict data privacy for many organizations and domains. This method solves an important problem in modern cybersecurity: the need to use collective knowledge from many sources while protecting data privacy and following rules that limit data sharing. Federated learning lets organizations join and use shared threat detection models without giving other participants or central authorities their sensitive data, private information, or secret security intelligence.

Federated learning is a big improvement over old centralized learning methods, where all training data must be collected in one place. Centralized methods cause privacy risks, legal problems, and create single points of failure. In contrast, our federated approach allows distributed training where each participant keeps full control of its local data while helping to build global threat detection models. This is very useful in cybersecurity because data privacy and sensitivity often stop organizations from sharing detailed information needed for effective joint threat intelligence.

The federated learning works by repeated coordination between participants and a central system that combines local updates without seeing the raw data. Each organization trains local models using their own data and security information, then only sends model parameters or gradients to the federated system. This way, sensitive data never leaves the organization but all can still create strong threat detection models using knowledge from many environments and attack types.

This cooperative method has many important benefits for cybersecurity. Smaller organizations or those with less expertise can gain from the knowledge of bigger, experienced participants. The variety of data sources and environments in the federated network helps create more reliable and general threat detection models able to find attacks in different system types and organizations. Also, federated learning allows fast sharing of threat information when new attacks appear, helping all participants quickly improve their defenses.

The math basis of our federated learning is an aggregation process that joins local model updates while keeping each participant's data private and preserving the statistical qualities needed for machine learning. This aggregation combines contributions from all participants, considering differences in data size and quality. The aggregation is mathematically shown as:

$$w_{global}^{t+1} = w_{global}^{t} + \eta \sum_{i=1}^{N} \frac{n_i}{n} \Delta w_i^t \tag{5}$$

In the formula, w_{global} means the global model parameters. These parameters keep the combined knowledge from all the nodes that are part of the federated learning system. This helps to build a common threat detection model that learns from many different environments and types of security problems.

The symbol η is the learning rate. It controls how big the updates to the global model are. This is important to keep the learning process stable and to help the model learn new threats quickly without losing accuracy.

The parameter N shows how many nodes are participating in the federated learning. These nodes can be different types of organizations, like factories, government offices, research labs, or cybersecurity companies. Each node has its own local data size, written as n_i, and the total data from

 doi:10.20944/preprints202509.0994.v1

all nodes together is called n. This total is used to decide how much each node's update should count in the final global model.

Each node trains its own local model and creates an update, called Δw_i. This update shows what the node learned from its own data. By combining all the updates using a weighted method, the system makes sure that nodes with more or better data have more influence on the final model. But it also includes useful knowledge from smaller nodes, making the learning fair and helpful for everyone, no matter how much data or resources they have.

3.5. Predictive Threat Modeling

The predictive component is one of the most new parts of our cybersecurity framework. It uses advanced time-series analysis and strong ensemble methods to predict future threats by studying current system states and past attack patterns. This prediction changes cybersecurity from a reactive field that only reacts after threats happen, to a proactive method that finds and stops threats before they affect important systems. This part solves the main problem of traditional cybersecurity systems that only detect threats after harmful actions have started, often when damage already happened.

The predictive method uses the time-based nature of cyber attacks, which usually follow known patterns and show early signs that can be found and studied to predict future threats. Many advanced attacks, especially persistent and multi-stage campaigns, develop over long times and have many preparation steps that create clear signs in system behavior and network traffic. By studying these time patterns and matching them with past attack data, the predictive component can find early warnings of new threats and allow early defense actions.

Using many prediction methods inside an ensemble framework gives important benefits over using just one model. Different algorithms are good at finding different parts of time patterns and threat developments. Traditional statistical methods like ARIMA models find linear trends and cycles in time-series data well. Neural networks can find complex nonlinear relations and small pattern changes that may show new threats. The ensemble method joins these strengths to get better prediction accuracy and trustworthiness than any single method.

The ensemble method also protects against problems or biases that single models might have and that can hurt prediction accuracy. Some models may work badly in some cases or miss some types of threats, but the ensemble method fixes these problems by using the different strengths of many algorithms. This backup makes sure the prediction system works well for many threat types and operating conditions.

The math base of our prediction method joins many time-series analysis techniques in one ensemble system that combines predictions from different algorithms to improve total accuracy and trust. The prediction model uses both traditional statistical methods and new machine learning methods to cover all types of time patterns that may show future threats. This full method is shown mathematically as:

$$P_{threat}(t + \Delta t) = \text{Ensemble}(\text{ARIMA}(X_t), \text{LSTM}(X_t), \text{CNN}(X_t)) \tag{6}$$

where P_{threat} represents the predicted threat probability for a future time horizon, providing a quantitative assessment of the likelihood that a security threat will materialize within the specified prediction window Δt. This probabilistic prediction enables security teams to prioritize their attention and resources based on quantitative risk assessments rather than relying solely on qualitative threat assessments.

The feature vector X_t encompasses comprehensive system state information at time t, including network traffic characteristics, system performance metrics, user behavior patterns, and environmental factors that may influence threat development. This multi-dimensional feature representation captures the complex interplay of factors that contribute to cybersecurity risk and enables the predictive models to identify subtle correlations and dependencies that may indicate emerging threats.

 doi:10.20944/preprints202509.0994.v1

The Autoregressive Integrated Moving Average (ARIMA) part gives basic time-series analysis tools that are good at finding linear trends, seasonal cycles, and repeated behaviors in past data. The LSTM neural network part learns complex time relationships and nonlinear connections that can last a long time, helping to find advanced attack patterns that grow slowly over time. The Convolutional Neural Network (CNN) part finds spatial and time patterns in multi-dimensional feature data, which helps to detect coordinated attacks that affect many system parts or attack paths at the same time. The ensemble method smartly combines these different prediction abilities to give better threat forecasting accuracy and reliability in difficult cybersecurity settings.

4. Results

4.1. Threat Detection Performance Evaluation

Our experimental evaluation used several well-known benchmark datasets such as NSL-KDD, CIC-IDS2017, and UNSW-NB15 to carefully measure threat detection performance on different attack types and network conditions. These datasets were chosen to cover many attack kinds, network setups, and time patterns that reflect real cybersecurity problems. Using multiple datasets helps to make sure our results are strong and can work well in different real-world environments, not only on specific dataset features or attacks.

The NSL-KDD dataset is an improved version of the original KDD Cup 1999 dataset. It is useful to test detection on classic network intrusion attacks like denial of service, unauthorized access, remote-to-local attacks, and privilege escalation attacks. This dataset allows comparison with many previous studies in network intrusion detection and provides widely accepted baseline performance results.

The CIC-IDS2017 dataset shows more recent attack scenarios that reflect modern cybersecurity threats, such as web attacks, infiltration attempts, botnet activities, and distributed denial of service attacks using current tools and methods. This dataset captures realistic network traffic and attack behavior seen in today's enterprise environments, giving important information about our framework's ability to handle modern attack types.

The UNSW-NB15 dataset adds more variety by including synthetic attack scenarios made with modern attack tools and techniques. It also contains attacks that focus on network protocols and services commonly used in today's cyber-physical systems. This dataset helps us understand detection performance against new attack methods that older datasets may not cover well.

Our experimental method used strict cross-validation and statistical analysis to make sure the performance results are reliable and fair. We processed each dataset using the same feature extraction and data preparation steps to keep results comparable between different tests. We measured performance with metrics like precision, recall, F1-score, and area under the ROC curve to give a full picture of detection ability under different operational needs and cost factors.

The results show clear improvements in detection accuracy compared to traditional rule-based systems, signature-based methods, and single-algorithm machine learning models. Our combined framework consistently performed better across all datasets, especially in finding advanced multi-stage attacks and new attack variants that older systems often miss. The ensemble method showed strong robustness for different attack types and network situations, keeping high accuracy while greatly lowering false positives that usually trouble traditional systems. These improvements lead to better security and less workload for security teams managing complex cyber-physical systems.

As shown in Table 1, our AI-enhanced digital twin framework achieved superior performance across all evaluation metrics, with detection accuracy reaching 99.2%, significantly outperforming traditional intrusion detection systems and standard machine learning approaches.

 doi:10.20944/preprints202509.0994.v1

Table 1. Threat Detection Performance Comparison.

Method	Accuracy	Precision	Recall	F1-Score
Traditional IDS	89.3%	87.1%	85.6%	86.3%
ML-based Detection	94.7%	92.8%	91.4%	92.1%
Our DT-AI Framework	99.2%	98.7%	98.9%	98.8%

4.2. Response Time Analysis

The evaluation of response time performance shows the great effectiveness of our automated orchestration system in strongly reducing incident response times for different types of threats. At the same time, it keeps high accuracy and lowers operational disruption. Response time is one of the most important measures in cybersecurity because the faster the system detects and responds to threats, the less damage attackers can cause to company assets and operations. Traditional cybersecurity methods often have long delays between finding a threat and taking proper actions. During these delays, attackers can keep access, increase their privileges, or steal important data.

The response time evaluation covers the whole incident response process, starting from first threat detection to final containment and fixing of problems. This full measurement gives a real view of cybersecurity effectiveness, instead of only showing detection accuracy, which may not reflect actual performance in real situations. The evaluation method considers many factors that affect response times, such as the difficulty of the threat, how systems need to work together, approval steps, and coordination between different security tools and teams.

Our experiments compared response performance for three different methods: traditional manual response done mostly by human analysts and administrators, common automated security tools that partly automate but still need a lot of human supervision, and our full digital twin-AI system that allows complete automated threat detection and response coordination. This comparison gives clear understanding of the real advantages of using advanced automation and AI in cybersecurity operations.

Table 2 illustrates substantial improvements in response times across all evaluated threat categories, with our digital twin-AI framework consistently achieving the fastest response times while maintaining high accuracy and effectiveness. The results demonstrate an average response time reduction of 67% compared to traditional automated tools and an impressive 85% compared to manual response processes, representing a transformative improvement in cybersecurity operational efficiency.

Table 2. Incident Response Time Comparison.

Threat Category	Manual Response	Automated Tools	DT-AI Framework
Malware Detection	2.3 hours	45 minutes	8 minutes
DDoS Attacks	1.8 hours	32 minutes	5 minutes
Data Exfiltration	3.1 hours	1.2 hours	12 minutes
Insider Threats	4.7 hours	2.1 hours	18 minutes

The performance improvements are very important for high-volume and time-sensitive threats like malware detection and DDoS attacks, where fast response is needed to stop big system damage or service interruptions. For malware detection cases, our framework lowers response time from 2.3 hours using manual methods to only 8 minutes. This fast action helps to stop malicious software from spreading in the network. The improvement for DDoS attack response from 1.8 hours to 5 minutes greatly lowers the impact on service availability and the related business costs. This shows the real benefit of automated response coordination for keeping operations running during security problems.

 doi:10.20944/preprints202509.0994.v1

4.3. Scalability and Resource Utilization

The scalability analysis was done to check system performance with different workloads and network sizes. It carefully studied computational efficiency and use of resources to see if the framework is suitable for many types of organizations, from small companies to large industrial plants. Scalability is very important for real cybersecurity systems because organizations need security solutions that can adjust to their infrastructure needs without losing performance or needing too much expensive hardware. The analysis tested many things like changes in network size, traffic amounts, number of users at the same time, and different levels of attack difficulty.

The scalability tests used real network setups and traffic patterns that are close to what happens in actual deployments, not artificial benchmarks that may not show real performance. Each test was made to copy real operational conditions including normal network traffic, usual user activities, and real attack frequencies. This way, the performance results better show what to expect in real life. The testing also included stress tests to check how the system works under very hard conditions to find any limits or problems that could affect reliability during busy times or big security incidents.

The evaluation covered four different network sizes to represent different organization types and deployment situations. Small networks with about 100 nodes are like small businesses, branch offices, or special industrial sites with simple infrastructure. Medium networks with 1,000 nodes are for medium companies or departments inside big enterprises. Large networks with 10,000 nodes represent big divisions or medium companies with more complex infrastructure. Enterprise-scale networks with 100,000 nodes simulate very big companies or critical infrastructure with large distributed systems.

All network sizes were tested using the same performance measures to give full understanding of resource needs and how well the system works. The testing used automated procedures that measure system performance across many situations while keeping experimental conditions controlled to make sure the results are reliable and can be repeated.

The scalability results in Table 3 are showing that the framework is scaling very well with different deployment sizes, and the performance efficiency stays almost same. This means the framework is suitable not only for big enterprise systems, but also can be useful for smaller organizations who have limited resources. The CPU usage grows in a predictable way, so organizations can plan their resource needs properly when they want to use this framework. Also, the memory usage increases as expected, and this helps in choosing right hardware according to network size and what performance is required.

Table 3. Scalability Performance Metrics.

Network Size	CPU Utilization	Memory Usage	Detection Latency
Small (100 nodes)	23%	1.2 GB	1.3 ms
Medium (1,000 nodes)	41%	4.8 GB	2.7 ms
Large (10,000 nodes)	68%	15.2 GB	5.1 ms
Enterprise (100,000 nodes)	89%	47.3 GB	12.4 ms

The detection latency results show that even in big deployments like enterprise scale, the framework can still keep very low response time from sub-millisecond to few milliseconds. This is very important for systems where fast reaction is necessary like industrial control systems, financial platforms, or emergency response where slow detection can cause serious problems. The good performance in all sizes proves that the system design is efficient and that it uses resources well with help of distributed processing.

4.4. Privacy Preservation Effectiveness

The evaluation of privacy-preserving mechanisms was focused on checking how good federated learning and differential privacy techniques are for keeping data confidential while still allowing sharing of threat intelligence between different organizations and countries with different rules.

 doi:10.20944/preprints202509.0994.v1

Keeping privacy is one of the most difficult problems in cybersecurity cooperation because companies want to work together for better security, but they also need to follow laws, protect their data, and keep business secrets. The evaluation method was including many tests to find the best strategy which gives good security benefits but does not make high privacy risks or legal problems.

The privacy evaluation was looking at important trade-offs between how useful the data stays (data utility), how strong the privacy protection is, and how much computer power is needed. Data utility means how much the protected data can still be used for finding cyber threats. Privacy protection level shows how well the sensitive information is hidden from attackers or people who should not see it. Computational overhead means the extra computer resources needed for doing the privacy protection.

The tests used four types of privacy methods. The first is baseline with no protection, which gives best data quality but no privacy at all, and is not allowed by modern laws. The second is differential privacy, which adds noise to the data so that single user information cannot be known, but still allows statistical analysis. The third is federated learning, where organizations train models locally and share only model updates, not the real data. This keeps control of data inside each organization.

Our hybrid approach is using the strong points of different privacy methods together to make the best balance between usability, privacy, and efficiency. This combined way is solving problems of each single method by using their strengths. It also includes adaptive privacy features that change the level of protection based on how sensitive the data is, what kind of threats are there, and what the company privacy needs are.

Table 4 is showing that our hybrid privacy-preserving method is keeping high data utility (91.8%) and also giving very strong privacy protection (95.3%) with acceptable computational overhead (18.9%). This means our method is giving good balance for real cybersecurity use. The results are showing that our hybrid method is working better than single privacy methods by using strong points of each and reducing the weak parts.

Table 4. Privacy Preservation Evaluation.

Privacy Technique	Data Utility	Privacy Level	Computational Overhead
No Privacy Protection	100%	0%	0%
Differential Privacy	87.3%	92.1%	15.2%
Federated Learning	94.6%	88.7%	23.7%
Our Hybrid Approach	91.8%	95.3%	18.9%

For differential privacy, the results are showing the usual trade-off between keeping privacy and losing data accuracy. Because this method is adding noise to the data, it gives strong privacy (92.1%) but the utility goes lower to 87.3%. This can be a problem for advanced threat detection systems that need high quality data to work well. Federated learning is keeping better data utility (94.6%) because data is not changed, but it has lower privacy (88.7%) because attackers might guess information from shared model updates.

Our hybrid method is giving best privacy (95.3%) and also keeping high utility (91.8%). This is better than differential privacy alone, and almost as good as federated learning in utility. The computational overhead is 18.9%, which is still low enough for practical systems where computer power may be limited. This shows that our hybrid solution is efficient and useful for many real cybersecurity systems.

4.5. Threat Prediction Accuracy

The predictive function of our framework was tested using time-based threat patterns and new attack types. We carefully checked how well the system can guess future threat chances over different time periods. This was done to see if predictive cybersecurity can help with taking action before the attack happens, not just reacting after it is found. We used full analysis of past attack data, pattern

 doi:10.20944/preprints202509.0994.v1

checking over time, and probability forecasting to test how good the framework is at finding new threats early in different time ranges.

The testing looked at prediction results in four main time windows, from short-term to long-term forecasts. Each time range has different use. Short-term predictions help to react fast with things like system hardening or raising alert levels. Middle-term predictions help with planning, like choosing how to use resources or arrange staff shifts. Long-term predictions are useful for big decisions like future investments or security strategy planning.

We tested the system on many types of threats: known attacks, new types of existing attacks, and totally new attack methods. This helped to check if the system works well for many kinds of cyber threats, like malware, network break-ins, data stealing, or insider attacks. The data we used had different threat types to test the system's prediction power clearly.

We also checked how the prediction system keeps working when threats change and when new intelligence is added. This is very important because real cyber threats are always changing. We used time-based validation methods to check if the prediction stays stable and reliable in different time periods and under different threat conditions.

The prediction performance results shown in Table 5 follow the expected trend where accuracy goes down when prediction time gets longer. This is normal because cyber threats change a lot, and it is hard to know exactly what will happen after a long time. Attacks can change in timing, method, and target, so longer-term predictions are more difficult. But still, the framework gives useful predictions for short and medium time periods. It is very accurate for near-future threats and still good enough for helping plan operations.

Table 5. Threat Prediction Performance.

Prediction Window	Accuracy	False Positives	True Positives
1 hour	96.7%	2.1%	94.6%
6 hours	91.3%	4.7%	86.6%
24 hours	84.2%	8.9%	75.3%
7 days	76.8%	15.3%	61.5%

For the one-hour prediction window, the system gives very high accuracy of 96.7% and a very low false positive rate of only 2.1%. This means the system can give strong and trusted information for fast reaction, without sending too many wrong alerts or wasting resources. For six-hour prediction, the system still performs well with 91.3% accuracy, which helps with things like staff planning, system checks, and setting early protection.

The 24-hour prediction is a bit less accurate at 84.2%, but still gives helpful information for planning daily security tasks and organizing resources. For the seven-day prediction, the accuracy is lower at 76.8%, and the false positive rate increases to 15.3%. This longer prediction is still useful for big-picture planning and checking long-term trends, but users need to be more careful and confirm before making big decisions based on this forecast.

5. Discussion

The experimental results show that combining AI with digital twin technology brings many benefits for cybersecurity, especially in better threat detection, faster response, and good prediction ability. Our framework reached 99.2% detection accuracy, which is much higher than traditional methods. This improvement comes from using real-time monitoring and smart pattern recognition at the same time. The multi-layer detection system can find both known and new types of attacks by using ensemble learning.

The response time results also show big improvement. The digital twin helps automate the response in a smart way, so the system can act fast without waiting for human input. This reduces the time when systems are at risk and also lowers the chance of human mistakes. The use of federated

 doi:10.20944/preprints202509.0994.v1

learning helps different organizations share threat knowledge without sharing their private data. This solves a big problem in today's cybersecurity sharing systems.

However, there are still some problems in the current version of our system. First, creating and running digital twins needs a lot of computing power. This can be difficult for smaller systems or organizations that do not have enough resources. Also, the accuracy of the predictions becomes lower when we try to predict far into the future. Because of this, the models must be updated often to keep good performance.

Another challenge is that our system works best when there is a large amount of high-quality training data. But not all organizations have access to such data, which can limit the effectiveness of the system. Finally, the setup process is complex and needs expert knowledge and a large initial investment. This can be a big obstacle for small companies or groups with limited budgets.

6. Conclusion

This research gives a complete framework that combines artificial intelligence and digital twin technology to build new generation cybersecurity systems. These systems can find threats early and respond automatically. Our experiments show big improvements in detection accuracy (99.2%), faster response time (67% less), and better prediction when compared to older cybersecurity methods. The framework also solves important problems like working at large scale, keeping privacy, and handling real-time needs.

The AI-powered digital twin system helps build strong cyber-physical systems that can change and adjust when new threats come. By adding federated learning, blockchain, and smart machine learning algorithms, the system allows different organizations to share threat information while still keeping their own data safe. For future work, we want to make the system faster and use less computing power for very large systems, improve predictions for long-term threats, and create common rules or models that help different organizations use digital twin security together.

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

References

1. C. Alcaraz and J. Lopez, "Digital Twin: A Comprehensive Survey of Security Threats," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1475–1503, 2022. doi: 10.1109/COMST.2022.3171465.
2. M. H. Homaei, Ó. Mogollón-Gutiérrez, J. Sancho, M. Ávila, and A. Caro, "A review of digital twins and their application in cybersecurity based on artificial intelligence," *Artificial Intelligence Review*, vol. 57, no. 8, pp. 1–42, 2024. doi: 10.1007/s10462-024-10805-3.
3. World Economic Forum, "How digital twin technology can enhance cybersecurity," *World Economic Forum Reports*, 2023. Available: https://www.weforum.org/stories/2023/03/how-digital-twin-technology-can-enhance-cyber-security/
4. F. Tao, H. Zhang, A. Liu, and A. Y. Nee, "Digital twin in industry: State-of-the-art," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019. doi: 10.1109/TII.2018.2873186.
5. Gartner Inc., "Gartner Says Digital Twin Market Will Reach $250 Billion by 2030," *Gartner Press Release*, 2021. Available: https://www.gartner.com/en/newsroom/press-releases/2021-12-07-gartner-says-digital-twin-market-will-reach-250-billion-by-2030
6. M. El-Hajj, T. Itäpelto, and T. Gebremariam, "Systematic literature review: Digital twins' role in enhancing security for Industry 4.0 applications," *Security and Privacy*, vol. 7, no. 5, e396, 2024. doi: 10.1002/spy2.396.
7. C. Alcaraz and J. Lopez, "Protecting Digital Twin Networks for 6G-enabled Industry 5.0 Ecosystems," *IEEE Network*, vol. 37, no. 2, pp. 302–308, 2023. doi: 10.1109/MNET.004.2200529.
8. C. Alcaraz and J. Lopez, "Digital twin security: A perspective on efforts from standardization bodies," *IEEE Security & Privacy*, vol. 23, no. 1, pp. 83–90, 2025. doi: 10.1109/MSEC.2024.3504193.
9. C. Alcaraz, "Digital twins: Double insecurity for industrial scenarios," in *Proceedings of the 9th ACM Cyber-Physical System Security Workshop*, 2023, pp. 2–2. doi: 10.1145/3592538.3607806.

 doi:10.20944/preprints202509.0994.v1

10. M. Eckhart and A. Ekelhart, "Towards security-aware virtual environments for digital twins," in *Proc. 4th ACM Workshop on Cyber-Physical System Security*, 2018, pp. 61–72. doi: 10.1145/3198458.3198464.
11. M. Dietz and G. Pernul, "Digital twin: Empowering enterprises towards a system-of-systems approach," in *Proc. BIS 2020 Workshops*, 2020, pp. 179–190. doi: 10.1007/978-3-030-61146-0_14.
12. M. Homaei, A. Di Bartolo, O. Mogollon-Gutierrez, F. B. Morgado, and A. Caro, "A Virtual Cybersecurity Department for Securing Digital Twins in Water Distribution Systems," *arXiv preprint arXiv:2404.20266*, 2024.
13. M. H. Homaei, A. J. Di Bartolo, M. Ávila, O. Mogollón-Gutiérrez, and A. Caro, "Digital transformation in the water distribution system based on the digital twins concept," *arXiv preprint arXiv:2412.06694*, 2024.
14. T. Itäpelto, M. Elhajj, M. van Sinderen, and M. Iacob, "Reference Architecture of Cybersecurity Digital Twin," in *Enterprise Design, Operations, and Computing. EDOC 2024 Workshops*, 2025, pp. 389–404. doi: 10.1007/978-3-031-79059-1_24.
15. C. Alcaraz, "Situational Awareness for Cyber-Physical Systems," *Encyclopedia of Cryptography, Security and Privacy*, pp. 2437–2439, 2023. doi: 10.1007/978-3-642-27739-9_1732-2.
16. I. H. Sarker, H. Janicke, A. Mohsin, A. Gill, and L. Maglaras, "Explainable AI for cybersecurity automation, intelligence and trustworthiness in digital twin: Methods, taxonomy, challenges and prospects," *ICT Express*, vol. 10, no. 4, pp. 935–958, 2024. doi: 10.1016/j.icte.2024.05.007.
17. Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. Yao, "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities," *ACM Computing Surveys*, vol. 54, no. 5, pp. 1–36, 2021. doi: 10.1145/3453155.
18. P. O. Adjei, S. K. Tetarave, C. John, M. Manneh, and P. Pattnayak, "Robust network anomaly detection with K-nearest neighbors (KNN) enhanced digital twins," in *Proc. SoutheastCon 2024*, IEEE, 2024. doi: 10.1109/SoutheastCon52093.2024.10500053.
19. M. H. Homaei, A. C. Lindo, J. C. S. Núnez, O. M. Gutiérrez, and J. A. Díaz, "The role of artificial intelligence in digital twin's cybersecurity," in *Proceedings of the RECSI-Reunión Española sobre Criptología y Seguridad de la Información*, 2022, pp. 1–14.
20. C. Alcaraz and J. Lopez, "Digital Twin-assisted anomaly detection for industrial scenarios," *International Journal of Critical Infrastructure Protection*, vol. 47, 100721, 2024. doi: 10.1016/j.ijcip.2024.100721.
21. Q. Xu, S. Ali, and T. Yue, "Digital twin-based anomaly detection with curriculum learning in cyber-physical systems," *ACM Transactions on Software Engineering and Methodology*, vol. 32, no. 1, pp. 1–32, 2023. doi: 10.1145/3582571.
22. J. Guo, Z. Liu, S. Tian, F. Huang, J. Li, X. Li, K. I. Kostromitin, and J. Ma, "TFL-DT: A Trust Evaluation Scheme for Federated Learning in Digital Twin for Mobile Networks," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 11, pp. 3548–3560, Nov. 2023. doi: 10.1109/JSAC.2023.3310094
23. S. A. Varghese, A. Dehlaghi Ghadim, A. Balador, Z. Alimadadi, and P. Papadimitratos, "Digital twin-based intrusion detection for industrial control systems," in *Proc. 2022 IEEE International Conference on Pervasive Computing and Communications Workshops*, 2022. doi: 10.1109/PerComWorkshops53856.2022.9767207.
24. S. Krishnaveni, S. Sivamohan, B. Jothi, T. Chen, and M. Sathiyanarayanan, "TwinSec-IDS: An enhanced intrusion detection system in SDN-digital-twin-based industrial cyber-physical systems," *Concurrency and Computation: Practice and Experience*, vol. 37, no. 3, e8334, 2025. doi: 10.1002/cpe.8334.
25. C. Alcaraz, I. H. Meskini, and J. Lopez, "Digital twin communities: an approach for secure DT data sharing," *International Journal of Information Security*, vol. 24, no. 1, p. 17, 2025. doi: 10.1007/s10207-024-00912-1.
26. A. Alshammari, N. Al-Hawbani, E. Khan, S. A. Ghaffar, and F. M. Al-Malaise Al-Ghamdi, "Federated Learning-Enabled Digital Twin Cybersecurity," *IEEE Access*, vol. 12, pp. 45231–45248, 2024. doi: 10.1109/ACCESS.2024.3378945.
27. S. He, C. Du, and M. S. Hossain, "6G-enabled consumer electronics device intrusion detection with federated meta-learning and digital twins in a metaverse environment," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 3111–3119, 2024. doi: 10.1109/TCE.2023.3321846.
28. J. Liu, L. Zhang, C. Li, J. Bai, H. Lv, and Z. Lv, "Blockchain-based secure communication of intelligent transportation digital twins system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22630–22640, 2022. doi: 10.1109/TITS.2021.3119786.
29. Z. Lv, C. Cheng, and H. Lv, "Blockchain based centralized learning for security in digital twins," *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 4, pp. 2165–2176, 2023. doi: 10.1109/TNSE.2023.3251234.
30. M. Homaei, V. G. Morales, O. M. Gutierrez, R. M. Gomez, and A. Caro, "Smart Water Security with AI and Blockchain-Enhanced Digital Twins," *arXiv preprint arXiv:2504.20275*, 2025.

 doi:10.20944/preprints202509.0994.v1

31. Banerjee, S., Das, D., Chatterjee, P., and Ghosh, U. (2023). Blockchain-enabled Digital Twin Technology for Next-Generation Transportation Systems. In *2023 IEEE 26th International Symposium on Real-Time Distributed Computing (ISORC)*. IEEE. 10.1109/isorc58943.2023.00040
32. Li, Y., Su, D. A., and Mardani, A. (2023). Digital twins and blockchain technology in the industrial Internet of Things (IIoT) using an extended decision support system model: Industry 4.0 barriers perspective. *Technological Forecasting and Social Change*, 195, 122794. 10.1016/j.techfore.2023.122794
33. Roumeliotis, C., Dasygenis, M., Lazaridis, V., and Dossis, M. (2024). Blockchain and Digital Twins in Smart Industry 4.0: The Use Case of Supply Chain-A Review of Integration Techniques and Applications. *Designs*, 8(6), 105. 10.3390/designs8060105