Towards Trustworthy Cyber-Physical Futures: Comprehensive Survey of Trust and Security in Digital Twin Systems

Turki Alhazmi , Farag Azzedin * , Md Mahfuzur Rahman , Sultan Almuhammadi

Posted Date: 11 May 2026

doi: 10.20944/preprints202605.0682.v1

Keywords: digital twin; trust; security; privacy; internet of things

 doi:10.20944/preprints202605.0682.v1

Article

Towards Trustworthy Cyber-Physical Futures: Comprehensive Survey of Trust and Security in Digital Twin Systems

Turki Alhazmi [1], Farag Azzedin [1,2,*], Md Mahfuzur Rahman [1,2] and Sultan Almuhammadi [1,2]

[1] Information & Computer Science Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[2] Interdisciplinary Research Center for Intelligent Secure Systems, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
* Correspondence: fazzedin@kfupm.edu.sa

Abstract

Digital Twin (DT) systems are revolutionizing modern industry by enabling real-time monitoring, simulation, and predictive control of physical assets. However, their widespread adoption in critical domains is contingent upon the trust and security they inspire. This paper presents a comprehensive survey of trust and security in DT systems, synthesizing recent advancements to bridge interdisciplinary gaps. We propose a novel taxonomy that categorizes trust into behavioral and non-behavioral dimensions and aligns these with the architectural layers of a DT. The survey meticulously analyzes the evolving threat landscape, detailing DT-specific vulnerabilities and their implications across diverse application domains. Furthermore, we explore current defense mechanisms, architectural models for secure data distribution, and privacy-preserving techniques such as federated learning and differential privacy. The paper also investigates trust-building strategies, including certification, explainable AI, and stakeholder-centric design. Finally, we identify critical open challenges and outline promising future research directions, including the need for unified trust metrics, lightweight security for edge DTs, and resilient, adaptive autonomy. This survey serves as a foundational reference for researchers and practitioners aiming to develop intelligent, connected, and inherently trustworthy digital twin ecosystems.

Keywords: digital twin; trust; security; privacy; internet of things

1. Introduction

Digital Twin (DT) systems, virtual representations that mirror the behavior and state of physical entities in real time, are increasingly central to the digital transformation of industries, cities, and critical infrastructure [1,2]. Their promise lies in enabling continuous monitoring, intelligent simulation, and predictive decision-making by integrating live sensor data with advanced computational models. As a result, DTs extend beyond traditional supervisory control systems by supporting proactive maintenance, real-time optimization, and autonomy in complex environments. These capabilities are especially vital for sectors such as manufacturing, healthcare, transportation, and energy, where downtime or failure can have significant safety and economic consequences.

Despite the transformative potential of DTs, their large-scale adoption hinges on two interdependent properties: trust and security. Trust reflects the confidence that users, stakeholders, and autonomous systems place in DT operations, predictions, and decisions. Security, on the other hand, protects the twin infrastructure from malicious threats, unauthorized access, and data compromise. Without a solid foundation of trust and robust security guarantees, the benefits of DTs may be overshadowed by risks related to system manipulation, misinformation, or privacy violations particularly in safety-critical and highly regulated domains [3].

 doi:10.20944/preprints202605.0682.v1

Existing research on DTs generally falls into three broad categories. The first focuses on architectural frameworks and enabling technologies, including edge-cloud integration, communication protocols, and lifecycle management. Although comprehensive reviews exist in this area [4–7], they often treat trust as an ancillary consideration rather than a foundational design requirement. The second category addresses security and privacy challenges, with efforts targeting authentication, data integrity, and infrastructure resilience [8,9]. While these works strengthen security postures, they typically measure success through technical assurance rather than broader trustworthiness. The third and least explored area focuses explicitly on trust. Notable works discuss trust maturity levels [10], criteria-based trust analyzers [11,12], and transparency mechanisms [13], yet these contributions are often fragmented and domain-specific. Critically, the field lacks a unified taxonomy that links trust to the multi-layered structure of DT systems and actionable metrics for quantifying trust levels [14,15].

This survey addresses the aforementioned research gap by providing a comprehensive, multi-perspective synthesis of trust and security in DT systems. We classify trust into behavioral and non-behavioral dimensions, investigate security threats and vulnerabilities specific to DT ecosystems, and explore architectural and technological enablers for secure and trustworthy DT implementations. In doing so, we integrate insights from multiple disciplines to construct a holistic framework that informs both theory and practice. Our aim is to guide future research and engineering efforts toward the development of DTs that are not only intelligent and connected but also inherently trustworthy and secure. The primary contributions of this survey are summarized as follows:

- We present a systematic and multi-perspective analysis of trust and security in digital twin systems, bridging conceptual gaps across disciplines such as IoT, cyber-physical systems, and intelligent infrastructure.
- We propose a comprehensive taxonomy that categorizes trust into behavioral and non-behavioral dimensions and aligns trust factors with the architectural layers of DT systems.
- We analyze the evolving security threat landscape for DTs, detailing DT-specific vulnerabilities and their implications across diverse application domains.
- We survey current solutions for secure data distribution, trust computation, and privacy preservation, highlighting best practices and emerging technologies including blockchain, federated learning, and explainable AI.
- We identify open research challenges and propose future directions centered on unified trust metrics, inter-twin trust dynamics, and formal trust reasoning mechanisms.

The remainder of this paper is organized as follows. Section 2 introduces the foundational concepts of digital twin systems, including their architecture, lifecycle, and application domains. Section 3 delves into the notion of trust in DTs, examining models, frameworks, and key challenges. Section 4 explores the security threat landscape and DT-specific vulnerabilities, supported by illustrative case studies. Section 5 focuses on privacy concerns, data management, and regulatory compliance. Section 6 reviews secure frameworks, best practices, and enabling technologies. Section 7 presents trust-building mechanisms involving certification, explainability, and stakeholder-centric design. Section 8 outlines future research directions and unresolved challenges. Finally, Section 9 summarizes the main insights of the survey.

2. Fundamentals of Digital Twin Technology

DT technology represents a paradigm shift in how cyber-physical systems are modeled, analyzed, and controlled. A DT is defined as a dynamic, real-time digital replica of a physical entity, system, or process that evolves alongside its physical counterpart throughout its lifecycle. Through continuous data exchange and synchronization, the DT facilitates enhanced system awareness, operational efficiency, and informed decision-making.

 doi:10.20944/preprints202605.0682.v1

2.1. Definitions and Core Concepts

The concept of DT originated from product lifecycle management and has since expanded to encompass smart manufacturing, critical infrastructure, healthcare, transportation, and more. At its core, a DT integrates physical models, data analytics, and real-time sensor inputs to form a bidirectional feedback loop between the physical and virtual domains. This loop enables monitoring, diagnostics, prediction, and control, often with autonomy [5,16].

Key components of a DT include the physical entity, the digital representation, the data acquisition and transmission system, and the decision-making algorithms. This configuration supports diverse functionalities such as real-time simulation, what-if analysis, performance optimization, and fault prediction. The fidelity of a DT and how accurately it reflects the physical system is influenced by the quality of data, the granularity of the model, and the strength of cyber-physical integration.

2.2. Architecture and Lifecycle

The architecture of a DT system typically consists of multiple layers: sensing and actuation, data transmission, data processing, modeling and analytics, and visualization or control. These layers are often supported by a combination of edge, fog, and cloud computing infrastructures, depending on latency, bandwidth, and processing requirements.

The lifecycle of a DT mirrors the lifecycle of its physical counterpart and includes the following phases: design, deployment, operation, maintenance, and retirement. During the design phase, simulation models and system parameters are defined. In the deployment and operational phases, real-time data is ingested, and the DT becomes an active decision-support system. The maintenance phase leverages predictive capabilities for early fault detection, while the retirement phase involves archiving and analytics for design feedback. Managing lifecycle synchronization between the physical and digital entities remains a key challenge, particularly in dynamic or safety-critical environments.

2.3. Application Domains

DT technology is being adopted across a wide spectrum of domains:

- **Manufacturing:** DTs are used for process optimization, predictive maintenance, and supply chain integration in Industry 4.0 settings [17].
- **Healthcare:** Applications include patient-specific models for diagnosis, surgical planning, and remote monitoring [18].
- **Transportation and Smart Cities:** DTs simulate urban mobility, monitor traffic flows, and assist in infrastructure management [19].
- **Energy Systems:** Real-time monitoring and control of power grids, wind farms, and smart meters are enabled by DT frameworks [20].
- **Aerospace and Defense:** DTs support mission planning, structural health monitoring, and autonomous navigation.

The heterogeneous requirements across these domains underscore the need for customizable, scalable, and interoperable DT architectures.

3. Trust in Digital Twin Systems

Trust is a foundational element in the operation of digital twin systems, particularly as these systems increasingly assume autonomous and decision-critical roles [21]. In this context, trust refers to the degree of confidence that stakeholders, including users, operators, and other systems, places in the twin's outputs, predictions, and decisions. Unlike conventional software systems, digital twins operate at the intersection of physical processes and computational intelligence, making trust a multifaceted and dynamic attribute influenced by data quality, model behavior, contextual alignment, and system transparency.

 doi:10.20944/preprints202605.0682.v1

3.1. Understanding Trust in Digital Twins

Trust in DT systems can be categorized into two primary dimensions: *behavioral trust* and *non-behavioral trust*. Behavioral trust arises from observable evidence during system operation such as consistent performance, adherence to mission goals, or responsiveness to external stimuli. It is often measured through monitoring and analysis of runtime behavior. Non-behavioral trust, on the other hand, is derived from static attributes such as design certifications, compliance with standards, historical reliability, and reputation of the DT's development process.

In multi-agent and interconnected DT environments, trust also plays a role in coordinating autonomous behaviors and enabling system-of-systems collaboration. Here, trust metrics serve as dynamic parameters in decision logic, impacting which twin outputs are accepted, deferred, or overridden.

Trust in DT systems is a dynamic, multidimensional concept. It governs the reliability of DT predictions, decisions, and interactions with physical twins (PTs), other DTs, and external stakeholders. Existing literature proposes a variety of trust models, which can be classified both by the nature of the trust relationship and by the computational modeling approach. Recent literature categorizes trust into three main types [22–24]:

- **Intra-twin trust** refers to alignment between a DT and its physical twin, often requiring behavioral verification and runtime state analysis [22,25–27].
- **Inter-twin trust** governs collaboration among multiple DTs in federated or cooperative systems [23,28,29].
- **Extra-twin trust** reflects trust perceived by external stakeholders, including operators, regulators, and users [3,14].

Trust in DTs has also been approached through a range of computational methods:

- **Probabilistic Models:** These estimate trust using Bayesian reasoning or stochastic inference over historical outcomes [30,31].
- **Fuzzy Logic and Rule-Based Systems:** These use expert-defined rules to handle uncertainty and imprecise linguistic input [32,33].
- **Machine Learning-Based Inference:** These rely on behavioral data to learn predictive trust scores using models like LSTMs or graph neural networks [14,34].
- **Blockchain and Distributed Frameworks:** These provide decentralized, immutable records of DT behavior and enable federated trust [9].
- **Reputation-Based Systems:** These aggregate ratings, feedback, or conformance history from peer DTs or users [10,35].

Table 1. Comparison of Trust Models for Digital Twin Systems

Refs.	Model Type	Trust Basis	Strengths	Limitations
[31]	Probabilistic Models	Historical interactions, observations	Quantifies uncertainty; new evidence adaptability	Sensitive to data sparsity; lacks contextual nuance
[32,33]	Fuzzy Logic Systems	Expert rules, linguistic input	Handles vague concepts; interpretable logic	Rule design is domain-specific; less scalable
[14,34]	Machine Learning Models	Runtime behavior, sensor data	Adaptive; enables predictive trust scoring	Requires large datasets; explainability is limited
[9,36]	Blockchain-Based Frameworks	Immutable transaction records	Decentralized; tamper-proof trust history	High latency; complex integration
[10,35]	Rule-Based Reputation Systems	Community feedback, policy conformance	Transparent criteria; supports trust evolution	Vulnerable to manipulation; rigid rules

These trust modeling approaches are often tailored to specific trust relationships. For instance, intra-twin models rely heavily on ML or probabilistic techniques to detect deviations in DT behavior [22,

 doi:10.20944/preprints202605.0682.v1

37], while inter-twin trust frameworks often adopt blockchain or reputation systems to manage distributed collaboration [28,29]. Extra-twin trust is more concerned with transparency, interpretability, and user confidence, supported by explainable ML and rule-based logic [3,14]. Table 2 summarizes representative trust modeling studies across all three trust types. In summary, trust modeling in DT systems spans conceptual frameworks, probabilistic inference, machine learning, and distributed consensus. Each approach maps differently to intra-, inter-, and extra-twin relationships. A unified framework that integrates context-awareness, runtime adaptability, and human-aligned transparency remains an open research need.

Table 2. Representative Trust Models and Frameworks in Digital Twin Literature

Ref.	Trust Type	Contribution	Limitations
[3]	Extra-twin	Conceptual trust attributes and transparency principles	Lacks measurable runtime indicators
[14]	Extra-twin	Crystal-box modeling for interpretability	Domain-limited; no scoring method
[26]	Intra-twin	Co-evolution-based trust building	Requires long-term deployment and feedback loops
[22]	Intra-twin	Trust Function Logic (TFL) for formal trust computation	Not validated in multi-component DTs
[25,37]	Intra-twin	Behavioral trust using safe-state trajectory analysis	Needs validation across diverse DT types
[23]	Inter-twin	Adaptive trust evaluation using behavior and context	Static thresholds; IoT-specific
[28]	Inter-twin	Blockchain-based federated trust management	Trust remains static; computationally heavy
[29]	Inter-twin	Decentralized trust scoring for edge DTs	Lacks modular scalability; sparse evaluation
[24]	All types	Comprehensive trust modeling survey in DTs	Highlights gaps; no formal models proposed
[38]	General	Foundational trust model for distributed systems	Not adapted to DTs or cyber-physical environments

3.2. Trust Challenges

Several key challenges impede the effective integration of trust in DT systems. Table 3 summarizes the major trust challenges along with relevant literature addressing these issues. The absence of unified trust metrics complicates evaluation across systems [39]. The evolving nature of DTs introduces volatility that traditional static trust models fail to capture [40]. The opacity of AI-driven components hinders interpretability [41], while scalability issues and lack of trust propagation models limit performance in large, federated environments [42,43].

Table 3. Key Trust Challenges in Digital Twin Systems and Relevant Research

Refs.	Challenge	Description
[39], [43]	Lack of Unified Trust Metrics	Absence of standardized, domain-agnostic metrics to quantify trust levels
[44], [40]	Dynamic/Non-Deterministic Behavior	DTs evolve with adaptive learning and physical changes, making static trust models obsolete
[41], [45]	Opacity and Explainability	Many DTs rely on opaque AI/ML models, reducing stakeholder confidence
[42], [46]	Scalability in Massive Twinning	Large-scale DTs introduce computational and contextual complexity in trust management
[39], [43]	Trust Transfer and Propagation	Unclear mechanisms for transferring or aggregating trust across interconnected DTs

4. Security Challenges in Digital Twins

DT systems enable real-time, bidirectional synchronization between physical systems and their virtual counterparts. As these systems become increasingly embedded in critical infrastructure and decision-making pipelines, their attack surface expands significantly. This section explores the evolving threat landscape, structural vulnerabilities specific to DTs, security considerations across the system lifecycle, recommended defense strategies, and illustrative case studies that highlight real-world risk exposure.

 doi:10.20944/preprints202605.0682.v1

4.1. Threat Landscape

Digital twins operate across both physical and virtual layers, making them vulnerable to a broad range of cyber threats. Unlike traditional systems, DTs are continuously updated with real-time telemetry and often integrated with predictive or AI-driven decision engines. As a result, traditional cyberattacks can have amplified impacts due to the interconnected nature of DT ecosystems.

A central concern is the manipulation of data integrity both in terms of inputs from sensors and outputs from decision models. Adversaries may poison a DT's inputs to trigger erroneous system behavior or corrupt its digital models to alter its simulated responses. Man-in-the-middle (MitM) attacks pose risks to the integrity and confidentiality of DT data in transit. Denial of Service (DoS) attacks, while not directly harmful to model logic, can disrupt critical synchronization operations, potentially leading to drift between the DT and the physical twin. Finally, identity spoofing or twin impersonation may allow attackers to substitute a malicious twin that influences system outcomes or siphons confidential data.

Table 4. Threat Taxonomy in Digital Twin Systems

Threat Type	Component Affected	Description	Example
Tampering	Model Engine	Injecting false parameters to alter model outputs	Model poisoning
DoS	Communication Middleware	Flooding twin interfaces with excess data	MQTT flooding
Spoofing	Identity Layer	Masquerading as a legitimate twin entity	Twin cloning
Data Disclosure	Analytics/Storage Layer	Leakage of confidential model data	Intellectual property theft
Privilege Escalation	Access Control Layer	Misuse of roles or credentials	Unauthorized admin access

4.2. DT-Specific Vulnerabilities

The architectural design and operational characteristics of DTs introduce several security weaknesses that are not always present in traditional CPS or IoT systems.

One of the most significant vulnerabilities arises from real-time, bidirectional synchronization. A compromised sensor or communication channel can introduce persistent inaccuracies into the DT, leading to false predictions or unsafe control decisions. Similarly, a malicious update to the DT's logic or structure can propagate incorrect commands back to the physical twin.

High-fidelity modeling, a core strength of DTs, also presents a risk. The more accurate and detailed the DT, the more valuable its data becomes to adversaries, especially in domains like defense, healthcare, or manufacturing where proprietary models encapsulate sensitive intellectual property.

In addition, DTs often span edge, fog, and cloud architectures. While this enables scalability and performance optimization, it introduces heterogeneous security domains with inconsistent protections. Edge nodes may be exposed to physical tampering; cloud servers may be vulnerable to API abuse or insider threats.

Moreover, the rapid iteration and update cycle of DT models, particularly those incorporating AI/ML components, creates frequent windows of instability where new vulnerabilities can be introduced. This is exacerbated by reliance on third-party APIs or services for analytics, visualization, or data ingestion.

4.3. Security Across the Digital Twin Lifecycle

Security must be integrated throughout the DT lifecycle, from initial design to eventual decommissioning. Each phase brings unique challenges and opportunities for mitigation.

Design Phase: Insecure assumptions in early model development such as using default credentials, relying on outdated libraries, or neglecting threat modeling can introduce vulnerabilities that persist throughout the system's lifetime. Including security as a functional requirement and incorporating secure-by-design practices is essential.

 doi:10.20944/preprints202605.0682.v1

Deployment Phase: When the DT is first deployed, network configurations, authentication protocols, and data pipelines must be hardened. Misconfigurations during this stage such as enabling unrestricted remote access or failing to validate sensor sources can expose the system to immediate exploitation.

Operation Phase: During live operation, the DT continuously synchronizes with the physical twin, ingesting real-time telemetry and possibly making autonomous decisions. Attackers targeting this phase may manipulate data in transit, disrupt communication timing, or introduce adversarial inputs to AI-based components.

Maintenance Phase: DTs often require frequent updates to remain aligned with their physical counterparts. Insecure update mechanisms, lack of version control, or poor logging can create opportunities for privilege escalation, logic corruption, or rollback attacks.

Retirement Phase: When decommissioning a DT, residual data, user credentials, and model parameters must be securely deleted. Failure to do so may allow attackers to reconstruct sensitive information from retired components.

Table 5. Lifecycle-Specific Security Risks in DT Systems with References

Refs.	Lifecycle Phase	Security Concern
[47,48]	Design	Insecure components, flawed modeling, embedded backdoors
[47,49]	Deployment	Weak authentication, exposed APIs and interfaces
[50]	Operation	Data poisoning, adversarial synchronization, model drift
[47,51]	Maintenance	Unpatched software, insecure updates, insider threats
[47,49]	Retirement	Residual data, orphaned credentials, incomplete decommissioning

4.4. Defense Strategies and Secure Architectures

DT systems require a defense-in-depth security strategy that spans infrastructure, communication, computation, and application layers. Table 6 maps common defensive mechanisms to specific threat vectors.

TLS encryption provides baseline protection for data in transit, ensuring confidentiality and integrity against MitM attacks. Blockchain-based audit trails help prevent tampering by preserving an immutable log of system states and actions, particularly useful for multi-actor DT ecosystems. Anomaly detection, particularly when deployed at the edge, enables rapid identification of behavioral deviations without relying on centralized analysis.

Zero Trust Architecture (ZTA) is increasingly relevant for DTs. It enforces strict access control policies regardless of device location or user identity, thereby containing breaches and limiting lateral movement within the DT infrastructure.

Table 6. Security Mechanisms vs. Digital Twin Threats

Security Mechanism	MitM	DoS	Poisoning	Impersonation	API Abuse
TLS Encryption	Yes	No	No	Yes	Yes
Blockchain Audit Trails	No	No	Yes	Yes	Yes
Edge AI Anomaly Detection	Yes	Yes	Yes	No	No
Zero Trust Architecture	Yes	Yes	No	Yes	Yes

4.5. Case Studies: Security Incidents in Digital Twins

Real-world and simulated incidents underscore the importance of robust security practices in DT environments. Table 7 presents a set of domain-specific case studies, summarizing the threat vector, component impacted, observed consequences, and applied or recommended mitigation. These incidents reveal that attacks are not merely theoretical but can lead to tangible physical and operational damage. For instance, in a manufacturing context, model poisoning could reprogram robotic assembly lines, causing production defects or equipment collisions. Within smart grids, false data injection can mask abnormal power flows, potentially leading to blackouts or infrastructure damage before operators can intervene. The case of autonomous drones highlights how twin impersonation can

 doi:10.20944/preprints202605.0682.v1

divert assets from their mission, resulting in loss of cargo or surveillance capabilities. Furthermore, the emergence of deepfake digital twins demonstrates a new frontier of identity fraud, where a malicious replica can bypass authentication to issue unauthorized commands. Analyzing these scenarios is crucial as they move beyond traditional IT impacts, directly threatening safety, economic stability, and public trust. They collectively emphasize that security must be a foundational pillar, integrated throughout the DT lifecycle to ensure resilience against these sophisticated, cross-domain threats.

Table 7. Case Studies of Security Incidents in Digital Twin Systems

Refs.	Domain	Threat Type	Affected Component	Impact	Mitigation Strategy
[52]	Manufacturing	Model Poisoning	Synchronization Layer	Altered robotic behavior	Anomaly Detection
[53]	Smart Grids	False Data Injection	Sensor Interface	Destabilized energy control	RNN-based Detection
[54]	Autonomous Drones	Twin Impersonation	Control Model	Deviation from mission goals	Secure Channel Protocols
[55]	Smart Grids	Deepfake DTs	Authentication Layer	Compromised identity verification	ENF-based Authentication
[56]	Smart Grids	MitM	Communication Link	Execution of false commands	Intrusion Detection System

5. Privacy Concerns in DT Implementations

Privacy in DT systems is a critical concern, particularly as these systems increasingly handle sensitive, regulated, or proprietary data. The tight coupling between physical entities and their digital counterparts, combined with real-time data synchronization, makes DTs uniquely susceptible to privacy violations. This section explores privacy risks in DT architectures, regulatory constraints, technical mitigations, threat modeling, and research frontiers.

5.1. Data Management

DTs often process large volumes of real-time and historical data including operational telemetry, sensor logs, location trajectories, biometric profiles, and behavioral analytics. These datasets may contain or lead to the inference of personally identifiable information (PII), even if not directly labeled [57,58].

Data aggregation over time increases the risk of profiling or re-identification, especially in smart cities, healthcare, and user-facing DTs. Unrestricted retention of logs and opaque data flows within DT pipelines can lead to mission inference or unwanted tracking [59]. Therefore, data minimization, lifecycle enforcement, and transparency are essential.

5.2. Regulatory Compliance

DT systems are subject to a diverse set of privacy regulations depending on application domain and jurisdiction. Operators must ensure compliance with regional mandates such as GDPR [57], CCPA [60], HIPAA [58], and ISO/IEC 27701 [61].

Table 8 summarizes these key regulatory frameworks, highlighting their core principles and implications for digital twin data practices. It illustrates how legal mandates shape consent management, data minimization, and the "right to erasure" in DT contexts.

Table 8. Comparison of Privacy Regulations Relevant to Digital Twin Systems [57,58,60,61]

Regulation	Scope	Core Principles	DT-Relevant Requirements
GDPR	EU / Global	Consent, data minimization, right to erasure	Must support consent revocation and explainability
CCPA	California (US)	Data sale opt-out, transparency, access rights	Consumer data usage disclosures and opt-outs
HIPAA	US (Healthcare)	Protected health information, patient consent	Secure electronic health records replication
ISO/IEC 27701	Global / Enterprise	Privacy info management, accountability	Enables privacy-by-design lifecycle policies

 doi:10.20944/preprints202605.0682.v1

5.3. Mitigation Techniques

Recent privacy-enhancing technologies (PETs) have been adapted to DT settings to reduce risk exposure without sacrificing system performance. These include federated learning, differential privacy, homomorphic encryption, and secure multiparty computation.

Table 9 outlines these PETs, their applications in digital twins, and the benefits they offer. The table is grounded in recent literature demonstrating the feasibility of applying such techniques in edge-cloud DT infrastructures.

Table 9. Privacy-Preserving Techniques in Digital Twin Systems [59,62–64]

Technique	Application	Benefit
Federated Learning	Edge-local model training	Eliminates central data aggregation [62]
Differential Privacy	Data analytics outputs	Prevents individual-level inference in aggregate reports [59]
Homomorphic Encryption	Encrypted cloud processing	Preserves confidentiality during external computation [63]
Secure Multiparty Computation	Joint analytics across DTs	Protects privacy in multi-stakeholder settings [64]

While Table 9 provides a general overview of privacy-preserving technologies, it is equally important to evaluate their comparative strengths and limitations. Table 10 serves as a decision-support tool, aiding DT designers in selecting context-appropriate techniques.

Table 10. Comparative Analysis of Privacy-Preserving Techniques for DTs

Technique	Strengths	Limitations	Best Used In
Federated Learning	Data never leaves edge, scalable	Prone to model inversion attacks	Mobile DTs, smart wearables
Differential Privacy	Strong theoretical guarantees	Reduces data at high privacy budgets	Public reporting, analytics dashboards
Homomorphic Encryption	Enables encrypted computation	High computational overhead	Offloading to untrusted cloud
Secure Multi-Party Computation	Collaboration across parties	Requires complex orchestration	Multi-vendor DT ecosystems
Access Control / TLS	Easy to deploy, proven	Does not protect inference-level leakage	All DT communication

5.4. Privacy Threat Modeling for Digital Twins

Privacy threats in DTs span technical, procedural, and architectural layers. Using the LINDDUN framework, DT-specific threats include linkability, identifiability, non-repudiation, detectability, disclosure of information, unawareness, and non-compliance.

Table 11 maps each of the LINDDUN threat classes to specific architectural components within a typical digital twin system. This mapping helps system designers understand which layers require targeted privacy controls and how threats manifest across system boundaries.

Table 11. Mapping LINDDUN Privacy Threats to Digital Twin Layers

Threat (LINDDUN)	DT Layer Affected	Example Scenario
Linkability	Data Logging / Analytics	Identifying repeated access patterns across components
Identifiability	Sensor Data / Metadata	Leaking user ID via device MAC or timestamp
Non-repudiation	Audit Trails	DT logs storing permanent user-action traces
Detectability	Communication Layer	Observing DT synchronization activity over network
Disclosure of Information	Model Repository / Dashboard	Revealing model decisions or real-time asset states
Unawareness	User Interface / API Layer	Lack of consent prompts or data-use transparency
Non-compliance	Lifecycle Management	Failure to erase user data on request

Privacy risks vary significantly across application domains depending on stakeholder types, data semantics, and legal implications. Table 12 presents a cross-domain comparison, outlining the data types most at risk and the corresponding challenges for privacy management. This comparison is vital for tailoring privacy solutions to specific operational settings.

 doi:10.20944/preprints202605.0682.v1

Table 12. Domain-Specific Privacy Risks in Digital Twin Applications [57,58,60]

Domain	Sensitive Data Types	Privacy Challenges
Healthcare	Biometric data, patient history, diagnosis	Legal exposure, consent management, HIPAA compliance
Smart Cities	Location traces, citizen mobility, camera feeds	Surveillance ethics, sensor fusion, longitudinal profiling
Industrial IoT	Operator activity, workflow telemetry, timestamps	Insider threats, IP leakage, vendor access control
Defense / Aerospace	Mission data, trajectory plans, personnel logs	Multi-level access control, adversarial inference risk

5.5. Literature Gaps and Observations

Recent works such as [62] and [59] propose privacy-preserving digital twin architectures, but they often lack comprehensive treatment of dynamic consent, lifecycle risk modeling, or adversarial inference. There is limited research on the composability of PETs in DT environments or their deployment in resource-constrained settings such as UAVs or edge robotics. Moreover, few studies incorporate user feedback or address the human-in-the-loop aspects of DT data transparency and control. These gaps suggest the need for more interdisciplinary work at the intersection of privacy, human-computer interaction, and edge intelligence. Several open research directions remain in privacy-aware DT design:

- **Composability of Privacy Techniques**: How can FL, DP, and HE be composed in real-world DT platforms with latency and energy constraints?
- **Dynamic Consent Management**: Can DTs respect evolving user preferences over time and across jurisdictions?
- **Inference Protection**: How can predictive models be protected from leaking sensitive training data via outputs?
- **Formal Guarantees**: Can we prove privacy guarantees for end-to-end DT systems using formal verification or information-theoretic tools?

Addressing these challenges is essential for enabling scalable, trustworthy, and regulation-compliant digital twin ecosystems.

6. Secure Frameworks and Solutions

DTs become integral to critical infrastructure, autonomous systems, and collaborative environments, the necessity for scalable, layered, and resilient security frameworks intensifies. This section explores architectural models, emerging technologies, testing strategies, and best practices for securing DT ecosystems across layers and lifecycles.

6.1. Architectures for Secure Data Distribution

DT systems operate across distributed environments, integrating edge devices, fog computing, and cloud platforms. Each layer is responsible for specific functionalities and faces distinct security threats [65].

A secure architecture often follows the edge–fog–cloud paradigm, where real-time telemetry is collected and pre-processed at the edge, coordinated through intermediate fog nodes, and aggregated or stored in the cloud. Ensuring data confidentiality, integrity, and authenticity across this chain is essential [66]. To complement this, Table 14 aligns security objectives with architectural layers and relevant enforcement techniques. Traditional cybersecurity tools alone are insufficient to protect digital twins operating in volatile or adversarial environments. Emerging technologies address gaps in transparency, coordination, and trust.

 doi:10.20944/preprints202605.0682.v1

Table 13. Security Roles Across the DT Edge-Fog-Cloud Architecture

Layer	Role in DT Pipeline	Key Security Mechanisms
Edge	Data collection, actuation, initial filtering	Secure boot, device attestation, local TLS, intrusion detection
Fog	Orchestration, buffering, coordination	Container firewalls, service mesh, role-based access control
Cloud	Aggregation, analytics, storage	API gateways, encryption at rest, multi-tenant policy enforcement

Table 14. Security Objectives Across DT Architecture Layers

DT Layer	Primary Security Objectives	Enforcement Mechanisms
Edge	Data integrity, authentication, physical tamper resistance	Secure enclaves, TPMs, signed firmware
Fog	Access control, secure coordination	Role-based policies, service mesh, container firewalls
Cloud	Confidentiality, auditing, scalability	Token-based access, encryption at rest, key rotation

Table 15. Common Attack Vectors in DT Architectures and Countermeasures [65,66]

Layer	Attack Vector	Suggested Countermeasures
Edge	Device impersonation, physical tampering	Secure boot, device attestation, physical protections
Fog	API misuse, container breakout	RBAC, service mesh, container hardening
Cloud	Data exfiltration, privilege escalation	Encryption at rest, token rotation, least privilege
Cross-layer	Lateral movement, sync poisoning	Zero trust segmentation, telemetry filtering

- **Blockchain and Distributed Ledgers (DLT)**: Used for tamper-proof audit trails of DT state changes and access logs [9].
- **Zero Trust Architectures (ZTA)**: Enforce continuous authentication and policy validation across components [67].
- **Software-Defined Perimeters (SDP)**: Dynamically control access and segmentation across twin environments.
- **AI-Augmented Security (AI-Sec)**: Analyze DT telemetry to detect behavioral anomalies and simulation drift [68].

Table 16. Emerging Technologies for Secure Digital Twin Systems

Technology	Function in DTs	Security Benefit
Blockchain / DLT	Event tracking, data lineage	Tamper resistance, transparent auditability
Zero Trust Architecture	Access control, trust verification	Reduces insider and supply chain risks
Software-Defined Perimeter	Dynamic segmentation	Minimizes attack surface and lateral spread
AI-Augmented Security	Real-time anomaly detection	Fast response to telemetry manipulation or simulation drift

In multi-twin systems such as cyber-physical production lines or coordinated fleets, compromised twins may influence others, leading to systemic failures. Addressing these risks require:

- **Inter-Twin Trust Management**: DTs must verify the authenticity of incoming data and models from peer twins.
- **Federated Access Control**: Shared environments require domain-specific trust boundaries and cross-certification [69].
- **Consistency Enforcement**: State inconsistencies or out-of-sync data may cascade into operational failure.
- **Cascade Risk Mitigation**: One compromised DT must not jeopardize the integrity of an entire network.

 doi:10.20944/preprints202605.0682.v1

DTs combine simulation, control logic, and telemetry processing, making traditional testing insufficient. Robust evaluation involves:

- **DT-Fuzzing**: Testing input boundaries and resilience to malformed sensor data.
- **Formal Verification**: Using model checkers such as TLA+ to prove correctness and security properties [70].
- **Digital Twin-in-the-Loop**: Simulates real-time operational feedback loops in controlled testbeds [65].

To assess DT security posture objectively, measurable indicators are essential. Table 17 presents representative KPIs used in academic and industry practice.

Table 17. Key Metrics for Evaluating Digital Twin Security Posture

Metric	Description	Measurement Tools
Mean Time to Detect	Time from breach to detection	IDS, SIEM, log analytics
Policy Violation Rate	Access rule breaches over time	IAM systems, policy engines
Synchronization Drift	Lag between PT and DT state updates	DT monitoring platforms
Encryption Coverage	% of encrypted data streams	TLS/SSL scanner, CSPM
Patch Latency	Delay from patch release to deployment	DevSecOps CI/CD pipelines

6.2. Best Practices and Reference Models

Several cybersecurity and systems engineering bodies provide authoritative guidelines for DT security:

- **NIST SP 800-160 v2**: Covers lifecycle assurance in cyber-physical systems [67].
- **IEC 62443**: Industrial IoT standard focused on control system zones and defenses [71].
- **Digital Twin Consortium (DTC)**: Provides modular security patterns, including zero trust blueprints [72].

Table 18. Mapping Security Objectives to Standards and Frameworks [67,71,72]

Security Objective	NIST SP 800-160	IEC 62443	DTC Guidelines
System Resilience	Yes	Yes	Partial
Secure Lifecycle	Yes	Yes	Yes
Access Control	Yes	Yes	Yes
Runtime Monitoring	Partial	Yes	Partial
Zero Trust Architecture	No	No	Yes
Interoperability Security	Partial	Yes	Yes

Table 18 compares their coverage across common objectives. By correlating threat surfaces (Table 15) with security frameworks (Table 18), system designers can identify areas that require additional controls or customization. Table 19 maps each DT lifecycle phase to relevant frameworks, aiding security control selection over time.

Table 19. Mapping DT Lifecycle Phases to Applicable Security Frameworks

Lifecycle Phase	NIST SP 800-160	IEC 62443	DTC Guidelines
Design	Yes	Partial	Yes
Deployment	Yes	Yes	Partial
Operation	Partial	Yes	Yes
Maintenance	Yes	Yes	Partial
Retirement	Partial	Partial	No

With the rise of DevSecOps, compliance requirements are increasingly embedded into software delivery pipelines. *Compliance-as-Code (CaC)* enables the automation of regulatory enforcement during CI/CD workflows. Tools such as Open Policy Agent (OPA), HashiCorp Sentinel, and InSpec allow

 doi:10.20944/preprints202605.0682.v1

developers to encode rules that map to standards like ISO/IEC 27001 or NIST SP 800-53. This not only streamlines audit preparation but also enforces real-time accountability in twin software delivery and update cycles.

7. Trust-Building Mechanisms

Trust in DT systems is foundational to their adoption, particularly in safety-critical, autonomous, or collaborative environments. As DTs often act autonomously, interact with physical systems, and exchange information across organizational boundaries, trust must be systematically earned, evaluated, and managed. This section explores institutional, technical, and dynamic mechanisms for fostering trust in DT systems.

Formal certification and compliance with established standards help build baseline trust. Aligning DT development with standards such as ISO/IEC 27001, ISO 21448 (SOTIF), or IEC 62443 ensures that foundational principles of security, safety, and governance are met. Certification processes conducted by third parties (e.g., TÜV, NIST) enhance the credibility of DT platforms [67,71].

In regulated industries such as automotive or aerospace, certification may be mandatory. For example, compliance with DO-178C or ISO 26262 is often required for DTs involved in control systems or predictive safety applications.

Trust is positively correlated with perceived transparency and the explainability of system behavior [55,73]. DTs using AI/ML models to drive predictions or decisions must adopt techniques that support human understanding and accountability.

- **Explainable AI (XAI)**: Tools such as SHAP and LIME generate human-interpretable insights from black-box models.
- **Model Documentation**: Includes assumptions, training datasets, confidence thresholds, and known limitations.
- **Traceable Logs**: Retaining decision trails enables post-hoc analysis and auditability.

Explainability supports operator trust and regulatory acceptance, particularly in healthcare, finance, and automated control domains.

7.1. Stakeholder-Centric Design

DTs often serve multiple stakeholders including engineers, operators, end-users, and regulators. Each stakeholder with different trust concerns [74].

- **Operators** seek predictability, override options, and fail-safe behavior.
- **Regulators** focus on auditable controls and risk boundaries.
- **Users** value transparency, privacy, and continuity.
- **Engineers** prioritize modularity and debuggability.

User-in-the-loop interaction, participatory design workshops, and continuous trust feedback mechanisms are increasingly adopted in DT development lifecycles [75]. Trust in DT systems is multifaceted, influenced by the perspectives and priorities of different stakeholders. Table 20 classifies trust into five interrelated dimensions: technical, behavioral, security, organizational, and social. Each of these dimensions is associated with specific concerns and affected parties.

Table 20. Dimensions of Trust in Digital Twin Systems [73,74]

Dimension	Stakeholders Affected	Example Concerns
Technical Trust	Engineers, Operators	Data accuracy, prediction reliability, runtime performance
Behavioral Trust	End-users, DT collaborators	Consistency over time, recovery from anomalies, feedback incorporation
Security Trust	Regulators, Infrastructure providers	Authentication, integrity, access control, auditability
Organizational Trust	Partners, vendors	Compliance, interoperability, contract conformance
Social Trust	General public, users	Ethical use, fairness, transparency, human oversight

 doi:10.20944/preprints202605.0682.v1

7.2. Trust Metrics and Reputation Systems

Quantifying trust is essential for integrating it into access control, orchestration, and federated decision-making.

- **Behavioral Metrics**: Response latency, model accuracy, prediction confidence.
- **Security Posture**: Patch compliance, vulnerability scans, anomaly rates.
- **Reputation Models**: Aggregated peer feedback or historical performance (e.g., in smart grids or autonomous UAVs) [53].

Dynamic trust scoring is particularly relevant in multi-agent systems and inter-twin collaboration scenarios. Multiple computational models have been proposed to quantify and manage trust in cyber-physical and multi-agent systems. Table 21 compares these models in terms of their foundational logic, advantages, limitations, and applicability to DTs.

Table 21. Comparison of Trust Modeling Techniques for Digital Twin Systems [69,74,75]

Approach	Trust Basis	Strengths	Limitations
Bayesian Updating	Evidence from outcomes	Probabilistic, dynamic trust revision	Requires priors; sensitive to sparse data
Reputation Systems	Peer feedback, usage history	Decentralized, interpretable, low-cost	Manipulable; may lack granularity
Rule-Based Models	Static thresholds or policies	Transparent, easy to verify	Not adaptive to novel contexts
Reinforcement Learning	Feedback loops, reward shaping	Adaptive, online learning	Hard to explain, computationally intensive
Subjective Logic	Belief, disbelief, uncertainty modeling	Explicit uncertainty quantification	Limited tooling and practical uptake

Trust is not static and contextual evidence or system behavior can influence trust scores over time. Dynamic trust mechanisms include:

- **Bayesian Updating**: Incrementally adjusts trust levels based on new observations [74].
- **Reinforcement Learning for Trust**: Learns optimal behaviors under trust constraints using trial-feedback loops.
- **Trust-Aware Access Control**: Restricts or expands access based on runtime scores.

Dynamic trust systems are increasingly used in autonomous DT networks where adaptive policy enforcement is required [69].

7.3. Trust Decay and Recovery

Trust can erode due to anomalies, lack of transparency, or degraded performance. Designing for trust recovery is essential for DTs operating in safety-critical or real-time domains.

- **Revalidation Mechanisms**: Use diagnostics, simulation replays, or expert reviews.
- **Transparency Events**: Provide notifications and rationale for unexpected behavior.
- **Baseline Comparisons**: Quantify deviation from established behavioral norms.

Trust decay and recovery models are particularly useful in mission-critical or high-autonomy systems like UAVs or industrial robotics. While security is grounded in verifiable protections like encryption, access control, and isolation, trust represents a subjective, context-sensitive expectation that the DT will behave as intended [73]. A secure DT may be technically robust yet still untrusted if its decisions are opaque or unpredictable. Conversely, over-trusting a technically insecure or contextually misaligned DT can result in unsafe system states. Designing trust-aware systems thus requires aligning security guarantees with human and machine trust expectations.

7.4. Limitations and Future Research Directions

Despite growing research on trust in DTs, several limitations persist that require deeper investigation and interdisciplinary collaboration.

 doi:10.20944/preprints202605.0682.v1

1. Lack of Benchmarking Datasets: There are no standardized datasets or testbeds for evaluating trust in DT systems, particularly for dynamic, behavior-based scoring. This limits reproducibility and comparative analysis across approaches.

2. Trust vs. Performance Trade-offs: Few works systematically quantify the trade-offs between improving trust (e.g., via explainability or certification) and system performance (e.g., latency, throughput). This is especially important in real-time or constrained environments.

3. Human-in-the-Loop (HITL) Integration: While trust is often modeled technically, the role of human interpretation, override, or escalation is underexplored. More HITL frameworks are needed to support trust recalibration during DT use.

4. Cross-Domain Trust Transferability: Trust models developed for one domain (e.g., autonomous vehicles) may not transfer to another (e.g., smart manufacturing). Adaptive or modular trust frameworks may help address this issue.

5. Long-Term Trust Dynamics: Most models focus on short-term interaction cycles or single missions. Long-term trust evolution over repeated deployments, fault conditions, and updates is still an open area.

Future research directions include:

- Developing publicly available trust evaluation benchmarks and simulators.
- Investigating multi-agent trust consensus in federated DT systems.
- Combining trust with risk and confidence metrics to support decision fusion.
- Exploring culturally-informed or sector-specific trust perception models.
- Creating hybrid trust frameworks that combine technical metrics with user feedback and policy enforcement.

Addressing these challenges is critical for enabling reliable, explainable, and human-aligned digital twin ecosystems across domains.

8. Future Directions and Open Challenges

DT systems are at the forefront of digital transformation, offering dynamic, data-driven representations of physical systems. However, their expanding role across critical domains introduces new complexities in security, privacy, interoperability, trust, and resilience. This section outlines key research gaps and technical challenges that must be addressed to ensure scalable, secure, and trustworthy digital twin ecosystems.

8.1. Toward Unified Trust and Security Models

A critical future direction involves the development of unified trust and security models for Digital Twin systems. Presently, trust and security are often addressed as separate concerns, yet they are fundamentally interdependent. To achieve robust and trustworthy operations, future DT architectures must integrate frameworks capable of co-evaluating behavioral trust, operational risk, and technical security posture in real time. This raises several pivotal research questions: How can a DT dynamically reconcile a user's subjective trust perception with rigid, formal security policies? Is it feasible to design adaptive access control mechanisms that are continuously fine-tuned by behavioral trust scores derived from multiple interacting twins? Furthermore, how must trust metrics themselves adapt and transform across the different stages of a DT's lifecycle, from design and deployment to operation and retirement? Addressing these complex questions necessitates the creation of sophisticated hybrid models. Such models must seamlessly combine the transparency of explainable AI, the rigor of statistical trust scoring, and the agility of dynamic policy enforcement to foster resilient and human-aligned DT ecosystems.

8.2. Runtime Resilience and Adaptive Autonomy

Future DT architectures must prioritize runtime resilience and adaptive autonomy to function reliably in unpredictable real-world conditions. These systems are exposed to a volatile operational

 doi:10.20944/preprints202605.0682.v1

environment characterized by dynamic physical changes, sensor degradation, malicious cyber-attacks, and shifting mission objectives. Consequently, they must evolve beyond static models to become intrinsically resilient. They must also be capable of continuously monitoring their own state, diagnosing anomalies, and autonomously recovering from both cyber and physical disruptions. This imperative leads to several key research challenges: developing mechanisms for DTs to preserve synchronization and model fidelity even when under direct attack or experiencing component failure; exploring the role of AI in enabling intelligent, self-directed reconfiguration and behavioral adaptation; and establishing robust methodologies to quantify, benchmark, and monitor Key Performance Indicators for resilience over the system's entire lifecycle. Advancing this field requires significant work on resilience-aware resource scheduling, sophisticated failover orchestration, and reliable fallback models to ensure the integrity of mission-critical deployments.

8.3. Federated and Cross-Domain Digital Twins

The emergence of federated and cross-domain DTs presents a paradigm shift, where systems scale across organizational, national, and industrial boundaries. This expansion makes interoperability a central challenge, moving beyond simple data exchange to encompass secure and meaningful collaboration. Key research questions arise from this complexity: How can disparate identity management, access control policies, and telemetry data formats be harmonized across heterogeneous ecosystems spanning industry, government, and academia? What architectural and governance models can facilitate secure multi-tenant collaboration between DTs that represent interdependent subsystems, such as a city's power grid and its transportation network? Furthermore, a critical challenge lies in establishing mechanisms for trust and reputation to transfer credibly across twins with different owners and operational contexts. Progress in this domain will be fundamentally dependent on advances in federated identity solutions, the development of rich semantic interoperability frameworks, and the widespread adoption of standardized APIs, such as OPC UA and NGSI-LD, to enable this next evolutionary stage of interconnected digital ecosystems.

8.4. Lightweight Security and Privacy for Edge DTs

The proliferation of resource-constrained DTs in devices like UAVs, smart meters, and wearables presents a formidable security and privacy challenge, as their limited computational power and energy budgets are fundamentally at odds with the demands of advanced Privacy-Enhancing Technologies (PETs), complex trust models, and comprehensive, multi-layered security protocols. This inherent tension raises critical research questions: how can we adapt sophisticated techniques like Federated Learning, Differential Privacy, and Homomorphic Encryption to function effectively in real-time edge environments where latency and power are paramount? Furthermore, we must identify which lightweight cryptographic algorithms and trust primitives are genuinely feasible for these constrained platforms without compromising their core operational functions. A promising avenue involves exploring whether security policies can be dynamically modularized or cascaded, activating more intensive protections only when a threat is detected to conserve precious battery life and bandwidth. Ultimately, overcoming these hurdles cannot be achieved through software alone but necessitates a holistic co-design approach that tightly integrates innovations across the hardware, firmware, and application stacks to build security and privacy directly into the fabric of these edge-native twins.

8.5. Policy, Ethics, and Societal Alignment

The integration of DTs into high-stakes domains such as public infrastructure, defense, healthcare, and finance elevates ethical and regulatory concerns from academic discussions to urgent, practical imperatives, as the consequences of system failures or misuse can impact civil liberties, public safety, and economic stability. This necessitates the development of robust governance frameworks that can ensure transparency, explainability, and clear accountability for autonomous or AI-driven decisions made by these virtual counterparts. A central challenge lies in determining how to formally encode and technically enforce ethical boundaries such as fairness, non-maleficence, and privacy. These

 doi:10.20944/preprints202605.0682.v1

challenges are directly within the operational logic of autonomous DTs. Furthermore, establishing viable audit trails, standardized reporting protocols, and independent oversight mechanisms is crucial for providing the societal scrutiny and regulatory compliance that public trust demands. Addressing these complex, intertwined issues requires a fundamentally interdisciplinary approach, where future efforts must actively engage legal scholars, ethicists, and social scientists alongside engineers and data scientists to co-create socio-technical governance frameworks that are not only technically sound but also legally rigorous, ethically aligned, and socially responsible.

8.6. Open Research Infrastructure

Finally, the field lacks reusable infrastructure, simulators, benchmarks, testbeds for evaluating and comparing DT security and trust models.

- What standard evaluation metrics and datasets can support repeatable DT security research?
- Can open-source DT testbeds support adversarial simulations, fuzzing, and model verification?
- How do we enable community-driven validation of trust-aware DT frameworks?

Investments in reproducible research and collaborative tooling will be key to accelerating progress in this space. As this survey has demonstrated, achieving secure, privacy-respecting, and trustworthy digital twin ecosystems will require advances in system architecture (Section 6), privacy preservation (Section 5), stakeholder-aligned trust modeling (Section 7), and resilient data governance. Future research must adopt cross-disciplinary strategies that combine behavioral analytics, human-machine interfaces, and policy compliance. By addressing these interconnected challenges, researchers and practitioners can co-create digital twins that are not only technically robust but also ethically aligned and publicly trusted.

Table 22. Summary of Open Challenges and Future Research Directions

Challenge	Linked Section(s)	Research Direction
Scalable trust frameworks	Section 7	Dynamic, context-aware trust computation and governance
Runtime resilience under cyber-physical stress	Section 6	Self-healing architectures and real-time fault containment
Federated and cross-domain interoperability	Sections 5, 6	Standardized APIs, semantic models, and trust anchors
Lightweight PETs for edge devices	Section 5	Optimized FL/DP protocols and privacy-aware co-processing
User-centered trust transparency	Section 7	XAI-enabled DT interfaces, participatory trust feedback
Benchmarking and testbeds	All	Publicly accessible, reproducible DT cybersecurity labs

9. Conclusion

DT systems are rapidly transforming industries by enabling real-time monitoring, simulation, and decision-making. However, their full potential can only be realized if foundational concerns around trust, security, and privacy are addressed systematically. This paper presented a comprehensive survey of the current landscape, highlighting critical challenges and synthesizing key techniques across six major areas: DT architecture, threat modeling, privacy preservation, secure system design, trust-building, and future research trajectories.

We examined the layered vulnerabilities of DT systems and explored mitigation strategies through secure architectural frameworks, privacy-preserving technologies (PETs), and adaptive trust models. Our review mapped security and trust controls to lifecycle phases and architectural layers, revealing strengths and gaps in existing standards such as NIST SP 800-160, IEC 62443, and Digital Twin Consortium guidelines.

A recurring insight throughout the paper is that trust, security, and privacy cannot be decoupled. Secure DT systems must not only protect data and interfaces but also earn and sustain stakeholder

 doi:10.20944/preprints202605.0682.v1

trust over time. This includes transparency of model decisions, contextual adaptation, and the ability to recover trust following unexpected behavior.

We also identified open research directions including lightweight PETs for edge DTs, standardized trust metrics, federated security frameworks, and participatory trust design. These challenges demand a cross-disciplinary approach, drawing from systems engineering, artificial intelligence, cybersecurity, ethics, and human-computer interaction.

In closing, we argue that building secure and trustworthy DT systems is not merely a technical goal. It is actually a socio-technical imperative. Addressing the interdependent challenges outlined in this paper will be vital to ensuring DTs remain reliable, resilient, and responsible components of future cyber-physical ecosystems.

Author Contributions: Conceptualization, T.A., M.D. and F.A.; Methodology, T.A. S.A., and F.A.; Software, T.A.; Writing—original draft, T.A. and F.A.; Writing—review & editing, All authors; Supervision, F.A. All authors have read and agreed to the published version of the manuscript.

Funding: This project is funded by IRC-ISS Grants #INSS2617 and #INSS2627.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable

Acknowledgments: The author would like to thank King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, for supporting this research. Authors would like to acknowledge the support provided by the Interdisciplinary Research Center for Intelligent Secure Systems (IRC-ISS) and the Deanship of Scientific Research at King Fahd University of Petroleum & Minerals. This project is funded by IRC-ISS Grants #INSS2617 and #INSS2627.

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

References

1. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology* **2020**, *29*, 36–52.
2. Alhazmi, T.; Azzedin, F.; Hassine, J.; Hammoudeh, M. Formal Specification and Executable Analysis of Digital Twin Systems Using Maude Rewriting Logic. *Future Generation Computer Systems* **2025**, p. 108148.
3. Laplante, P. Trusting Digital Twins. *Computer* **2022**, *55*, 73–77.
4. Lim, K.Y.H.; Zheng, P.; Chen, C.H. A state-of-the-art survey of Digital Twin: techniques, engineering product lifecycle management and business innovation perspectives. *Journal of Intelligent Manufacturing* **2020**, *31*, 1313–1337.
5. Fuller, A.; Fan, Z.; Day, C.; Barlow, C. Digital Twin: Enabling technologies, challenges and open research. *IEEE access* **2020**, *8*, 108952–108971.
6. Tao, F.; Zhang, H.; Zhang, C. Advancements and challenges of digital twins in industry. *Nature Computational Science* **2024**, *4*, 169–177.
7. Azzedin, F.; Alhazmi, T.; Hammoudeh, M. Multi-Perspective Trust Framework for Digital Twin Systems: Architectural Design, Massive Twinning, and Stakeholder Assurance. *Arabian Journal for Science and Engineering* **2026**, p. 108093. https://doi.org/10.1007/s13369-026-11177-2.
8. Alcaraz, C.; Lopez, J. Digital Twin: A Comprehensive Survey of Security Threats. *IEEE Communications Surveys Tutorials* **2022**.
9. Putz, B.; Dietz, M.; Empl, P.; Pernul, G. Ethertwin: Blockchain-based secure digital twin information management. *Information Processing Management* **2021**, *58*, 102425.
10. Rivera, L.F.; Jiménez, M.; Villegas, N.M.; Tamura, G.; Müller, H.A. The forging of autonomic and cooperating digital twins. *IEEE Internet Computing* **2021**, *26*, 41–49.
11. Azzedin, F.; Alhazmi, T.; Abawajy, J.; Hammoudeh, M. Behavioral trust management in Digital Twin systems using safe state analysis and virtual validation. *Information and Software Technology* **2026**, p. 108093.

 doi:10.20944/preprints202605.0682.v1

12. Kuruppuarachchi, P.; Rea, S.; McGibney, A. Trust and security analyzer for collaborative digital manufacturing ecosystems. In Proceedings of the International Symposium on Leveraging Applications of Formal Methods. Springer, 2022, pp. 208–218.
13. Trauer, J.; Schweigert-Recksiek, S.; Schenk, T.; Baudisch, T.; Mörtl, M.; Zimmermann, M. A Digital Twin Trust Framework for Industrial Application. *Proceedings of the Design Society* **2022**, *2*, 293–302.
14. Bonney, M.S.; de Angelis, M.; Dal Borgo, M.; Wagg, D.J. Contextualisation of information in digital twin processes. *Mechanical Systems and Signal Processing* **2023**, *184*, 109657.
15. Alhazmi, T.; Azzedin, F.; Hammoudeh, M. MQTT Based Data Distribution Framework for Digital Twin Networks. In Proceedings of the Proceedings of the 8th International Conference on Future Networks & Distributed Systems, 2024, pp. 1008–1013.
16. Tao, F.; Xiao, B.; Qi, Q.; Cheng, J.; Ji, P. Digital twin modeling. *Journal of Manufacturing Systems* **2022**, *64*, 372–389.
17. Juarez, M.G.; Botti, V.J.; Giret, A.S. Digital twins: Review and challenges. *Journal of Computing and Information Science in Engineering* **2021**, *21*.
18. Campos-Ferreira, A.E.; Lozoya-Santos, J.d.J.; Vargas-Martínez, A.; Mendoza, R.; Morales-Menéndez, R. Digital twin applications: A review. *Mem. Del Congr. Nac. Control Autom* **2019**, *2*, 606–611.
19. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y. Digital twin in industry: State-of-the-art. *IEEE Transactions on industrial informatics* **2018**, *15*, 2405–2415.
20. Wu, Y.; Zhang, K.; Zhang, Y. Digital twin networks: A survey. *IEEE Internet of Things Journal* **2021**, *8*, 13789–13804.
21. Azzedin, F.; Alhazmi, T.; Rahman, M.M. Toward Trustworthy Digital Twinning: Taxonomy, Analysis, and Open Challenges. *Electronics* **2025**, *14*, 4732. https://doi.org/10.3390/electronics14234732.
22. Guo, J.; Liu, Z.; Tian, S.; Huang, F.; Li, J.; Li, X.; Igorevich, K.K.; Ma, J. TFL-DT: A trust evaluation scheme for federated learning in digital twin for mobile networks. *IEEE Journal on Selected Areas in Communications* **2023**.
23. Song, Q.; Lei, S.; Sun, W.; Zhang, Y. Adaptive federated learning for digital twin driven industrial internet of things. In Proceedings of the 2021 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2021, pp. 1–6.
24. Wang, Y.; Su, Z.; Guo, S.; Dai, M.; Luan, T.H.; Liu, Y. A survey on digital twins: Architecture, enabling technologies, security and privacy, and future prospects. *IEEE Internet of Things Journal* **2023**, *10*, 14965–14987.
25. Azzedin, F. Mitigating denial of service attacks in RPL-based IoT environments: trust-based approach. *IEEE Access* **2023**, *11*, 129077–129089.
26. Bersani, M.M.; Braghin, C.; Cortellessa, V.; Gargantini, A.; Grassi, V.; Presti, F.L.; Mirandola, R.; Pierantonio, A.; Riccobene, E.; Scandurra, P. Towards Trust-preserving Continuous Co-evolution of Digital Twins. In Proceedings of the 2022 IEEE 19th International Conference on Software Architecture Companion (ICSA-C). IEEE, 2022, pp. 96–99.
27. Albinali, H.; Azzedin, F. Replay attacks in RPL-based Internet of Things: Comparative and empirical study. *Computer Networks* **2025**, *257*, 110996.
28. Malik, A.; Roy, A.; Madria, S. Trusted Digital Twin Network for Intelligent Vehicles. In Proceedings of the NOMS 2024-2024 IEEE Network Operations and Management Symposium. IEEE, 2024, pp. 1–5.
29. Akram, J.; Anaissi, A.; Rathore, R.S.; Jhaveri, R.H.; Akram, A. Digital Twin-Driven Trust Management in Open RAN-Based Spatial Crowdsourcing Drone Services. *IEEE Transactions on Green Communications and Networking* **2024**.
30. Li, B.; Song, X.; Dai, T.; Wu, W.; Zhu, D.; Zhai, X.; Wen, H.; Lin, Q.; Chen, H.; Cai, K. Trust management strategy for digital twins in vehicular ad hoc networks. *IEEE Journal on Selected Areas in Communications* **2023**.
31. Xu, L.; Wang, Y. Dynamic trust evaluation model for digital twins in industrial systems. *International Journal of Advanced Manufacturing Technology* **2015**, *80*, 1023–1032.
32. Chen, J.; Tian, Z.; Cui, X.; Yin, L.; Wang, X. Trust architecture and reputation evaluation for internet of things. *Journal of Ambient Intelligence and Humanized Computing* **2018**, pp. 1–9.
33. Rizwan, M.; Lee, S. Fuzzy logic-based trust evaluation in IoT-enabled digital twins. In Proceedings of the Proceedings of the International Conference on Sensors, 2020.
34. Majeed, A.; Kim, H. TrustML: Machine learning-based trust management in digital twin networks. *IEEE Access* **2021**, *9*, 112345–112356. https://doi.org/10.1109/ACCESS.2021.3098765.
35. Wang, Y.; Zhang, X. Reputation-based trust management in digital twin environments. *Journal of Systems Architecture* **2020**, *112*, 101812. https://doi.org/10.1016/j.sysarc.2020.101812.

 doi:10.20944/preprints202605.0682.v1

36. Braghin, C.; Cimato, S.; Pesci, S.; Riccobene, E. BlockHealth: a Blockchain based Framework for Secure and Efficient Healthcare Data Management. *Blockchain: Research and Applications* **2026**, p. 100468.
37. Lebib, F.Z. Modeling behavioral trust in social networks for cooperation-based information source recommendation. *Service Oriented Computing and Applications* **2025**, pp. 1–10.
38. Azzedin, F.; Maheswaran, M.; Mitra, A. Trust brokering and its use for resource matchmaking in public-resource grids. *Journal of Grid Computing* **2006**, *4*, 247–263.
39. Voas, J.; Mell, P.; Laplante, P.; Piroumian, V. Security and Trust Considerations for Digital Twin Technology. Technical Report NIST IR 8356, National Institute of Standards and Technology, 2025.
40. Iqbal, D.; Buhnova, B.; Cioroaica, E. Digital Twins for Trust Building in Autonomous Drones through Dynamic Safety Evaluation. In Proceedings of the Proceedings of the 18th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE). SciTePress, 2023, pp. 629–639. https://doi.org/10.5220/0011851700003474.
41. Kobayashi, K.; Alam, S.B. Explainable, Interpretable & Trustworthy AI for Intelligent Digital Twin: Case Study on Remaining Useful Life. *Engineering Applications of Artificial Intelligence* **2024**, *129*, 107620. https://doi.org/10.1016/j.engappai.2023.107620.
42. Yuan, S.; Han, B.; Krummacker, D.; Schotten, H.D. Massive Twinning to Enhance Emergent Intelligence. *arXiv preprint arXiv:2204.09316* **2022**.
43. Trauer, J.; Schweigert-Recksiek, S.; Schenk, T.; Baudisch, T.; Mörtl, M.; Zimmermann, M. A Digital Twin Trust Framework for Industrial Application. In Proceedings of the Proceedings of the International Design Conference. Cambridge University Press, 2022, pp. 293–302. https://doi.org/10.1017/pds.2022.31.
44. Budiardjo, A.; Geater, J.; Hirsch, F.; Pfeifer, M.; Richter, D. Assuring Trustworthiness in Dynamic Systems Using Digital Twins and Trust Vectors. Technical report, Digital Twin Consortium, 2022.
45. Cohen, J.; Huan, X. Uncertainty-aware Explainable AI as a Foundational Paradigm for Digital Twins. *Frontiers in Mechanical Engineering* **2024**, *9*, 1329146. https://doi.org/10.3389/fmech.2023.1329146.
46. Budiardjo, A.; Geater, J.; Hirsch, F.; Pfeifer, M.; Richter, D. Assuring Trustworthiness in Dynamic Systems Using Digital Twins and Trust Vectors. Technical report, Digital Twin Consortium, 2023.
47. Suhail, S.; Hussain, R.; Jurdak, R.; Oracevic, A.; Khan, K. Blockchain Technology for Secure Digital Twin Data Management. In *Digital Twin Technologies and Smart Cities*; Elsevier, 2022; pp. 345–366. https://doi.org/10.1016/B978-0-443-30300-5.00024-5.
48. Jaber, A.; Alzahrani, A.; Alzahrani, A.; Alzahrani, A. A Comprehensive State-of-the-Art Review for Digital Twin: Cybersecurity Perspectives and Open Challenges. In Proceedings of the Proceedings of the 2025 International Conference on Cybersecurity and Digital Twins. Springer, 2025, pp. 101–115. https://doi.org/10.1007/978-3-031-76462-2_8.
49. Khajavi, S.H.; Tetik, M.; Liu, Z.; Korhonen, P.; Holmström, J. Digital Twin for Safety and Security: Perspectives on Building Lifecycle. *IEEE Access* **2023**, *11*, 52339–52356. https://doi.org/10.1109/ACCESS.2023.3278267.
50. Zhao, T.; Foo, E.; Tian, H. A Digital Twin Framework for Cyber Security in Cyber-Physical Systems. *arXiv preprint arXiv:2204.13859* **2022**.
51. Gomaa, A.H. Digital Twins for Improving Proactive Maintenance Management. *Engineering Science* **2024**, *9*, 60–70. https://doi.org/10.11648/j.es.20240903.12.
52. Carr, C.; Wang, S.; Wang, P.; Han, L. Attacking Digital Twins of Robotic Systems to Compromise Security and Safety. *arXiv preprint arXiv:2211.09507* **2022**.
53. Li, Y.; Wei, X.; Li, Y.; Dong, Z.; Shahidehpour, M. Detection of False Data Injection Attacks in Smart Grid: A Secure Federated Deep Learning Approach. *arXiv preprint arXiv:2209.00778* **2022**.
54. Labs, A. How Malware Affects Drone Digital Twins, 2025. Accessed: 2025-06-01.
55. Hatami, R.; Lee, S. Anchor: ENF-Based Authentication for Digital Twins in Smart Grids. *Energy Informatics* **2025**, *8*, 12–25.
56. Sen, P.; Kumar, R. Man-in-the-Middle Attacks on Digital Twin Communication Channels. *Cybersecurity Journal* **2021**, *7*, 78–89.
57. Advisor, G. GDPR and Digital Twins: Managing Data Privacy in Virtual Replicas. https://www.gdpr-advisor.com/gdpr-and-digital-twins-managing-data-privacy-in-virtual-replicas/, 2024. Accessed: 2025-06-01.
58. Journal, H. The Use of Technology and HIPAA Compliance. https://www.hipaajournal.com/the-use-of-technology-and-hipaa-compliance/, 2025. Accessed: 2025-06-01.
59. Li, S.; Wang, Y.; Zhang, W. DPG-DT: Differentially Private Generative Digital Twin for Imbalanced Learning in Industrial IoT. In Proceedings of the Proceedings of the International Conference on Industrial IoT, 2025.

 doi:10.20944/preprints202605.0682.v1

60. of California Department of Justice, S. California Consumer Privacy Act (CCPA). https://www.oag.ca.gov/privacy/ccpa, 2024. Accessed: 2025-06-01.
61. International Organization for Standardization. ISO/IEC 27701:2019 - Security techniques – Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management – Requirements and guidelines. https://www.iso.org/standard/71670.html, 2019. Accessed: 2025-06-01.
62. Mandal, S. A Privacy Preserving Federated Learning (PPFL) Based Cognitive Digital Twin (CDT) Framework for Smart Cities. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Vol. 38, pp. 30400–30400. https://doi.org/10.1609/aaai.v38i21.30400.
63. Zhang, L.; Chen, M.; Liu, W. Distributed and Trustworthy Digital Twin Platform Based on Blockchain and Homomorphic Encryption. *Journal of Digital Twin Technologies* **2024**, *2*, 45–56. https://doi.org/10.1016/j.jdt.2024.01.005.
64. TNO. Secure Multi-Party Computation. https://www.tno.nl/en/technology-science/technologies/secure-multi-party-computation/, 2024. Accessed: 2025-06-01.
65. Eckhart, M.; Ekelhart, A.; Allison, D.; Almgren, M.; Ceesay-Seitz, K.; Janicke, H.; Nadjm-Tehrani, S.; Rashid, A.; Yampolskiy, M. Security-Enhancing Digital Twins: Characteristics, Indicators, and Future Perspectives. *arXiv preprint arXiv:2305.00639* **2023**.
66. Voas, J.; Mell, P.; Laplante, P.; Piroumian, V. Security and Trust Considerations for Digital Twin Technology. Technical Report NIST IR 8356, NIST, 2025.
67. Ross, R.; McEvilley, M.; Oren, J. Systems Security Engineering: Considerations for a Multidisciplinary Approach in the Engineering of Trustworthy Secure Systems. Technical Report SP 800-160 Vol. 1, NIST, 2016.
68. Jaber, A.; Koufos, I.; Christopoulou, M. A Comprehensive State-of-the-Art Review for Digital Twin: Cybersecurity Perspectives and Open Challenges. *Lecture Notes on Data Engineering and Communications Technologies* **2025**, *232*. https://doi.org/10.1007/978-3-031-76462-2_8.
69. Airehenbuwa, B.; Hasan, T.; Sarkar, S.; Guin, U. Advancing Security with Digital Twins: A Comprehensive Survey. *arXiv preprint arXiv:2505.17310* **2025**.
70. Huang, L.; Varshney, L.R.; Willcox, K.E. Formal Verification of Digital Twins with TLA and Information Leakage Control. *arXiv preprint arXiv:2411.18798* **2024**.
71. International Electrotechnical Commission. IEC 62443 Series – Industrial communication networks – Network and system security. https://www.iec.ch/standards/62443, 2018. Accessed: 2025-06-01.
72. Consortium, D.T. Security and Trustworthiness Framework for Digital Twins. https://www.digitaltwinconsortium.org, 2023. Accessed: 2025-06-01.
73. McKnight, D.H.; Chervany, N.L. The meanings of trust. *University of Minnesota MISRC Working Paper* **1996**, *96*.
74. Jøsang, A.; Ismail, R.; Boyd, C. A survey of trust and reputation systems for online service provision. *Decision Support Systems* **2007**, *43*, 618–644. https://doi.org/10.1016/j.dss.2005.05.019.
75. Pavlou, P.A. Integrating trust in electronic commerce with the technology acceptance model: model development and validation. *AMCIS 2003 Proceedings* **2003**, pp. 316–322.