Early version, also known as pre-print

Link to published version (if available):
10.48550/arXiv.2103.11585

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Blockchain-based Digital Twins: Research Trends, Issues, and Future Challenges

SABAH SUHAIL, University of Tartu, Estonia
RASHEED HUSSAIN, Innopolis University, Russia
RAJA JURDAK, Queensland University of Technology, Australia
ALMA ORACEVIC, Innopolis University, Russia
KHALED SALAH, Khalifa University, UAE
RAIMUNDAS MATULEVIČIUS, University of Tartu, Estonia
CHOONG SEON HONG, Kyung Hee University, South Korea

Industrial processes rely on sensory data for decision-making processes, risk assessment, and performance evaluation. Extracting actionable insights from the collected data calls for an infrastructure that can ensure the dissemination of trustworthy data. For the physical data to be trustworthy, it needs to be cross-validated through multiple sensor sources with overlapping fields of view. Cross-validated data can then be stored on the blockchain, to maintain its integrity and trustworthiness. Once trustworthy data is recorded on the blockchain, product lifecycle events can be fed into data-driven systems for process monitoring, diagnostics, and optimized control. In this regard, Digital Twins (DTs) can be leveraged to draw intelligent conclusions from data by identifying the faults and recommending precautionary measures ahead of critical events. Empowering DTs with blockchain in industrial use-cases targets key challenges of disparate data repositories, untrustworthy data dissemination, and the need for predictive maintenance. In this survey, while highlighting the key benefits of using blockchain-based DTs, we present a comprehensive review of the state-of-the-art research results for blockchain-based DTs. Based on the current research trends, we discuss a trustworthy blockchain-based DTs framework. We highlight the role of Artificial Intelligence (AI) in blockchain-based DTs. Furthermore, we discuss current and future research and deployment challenges of blockchain-supported DTs that require further investigation.

Authors' addresses: Sabah Suhail, University of Tartu, Tartu, Estonia, sabah.suhail@ut.ee; Rasheed Hussain, Innopolis University, Innopolis, Russia, r.hussain@innopolis.ru; Raja Jurdak, Queensland University of Technology, Brisbane, Australia, r.jurdak@qut.edu.au; Alma Oracevic, Innopolis University, Innopolis, Russia, a.oracevic@innopolis.ru; Khaled Salah, Khalifa University, Abu Dhabi, UAE, khaled.salah@ku.ac.ae; Raimundas Matulevičius, University of Tartu, Tartu, Estonia, raimundas.matulevicius@ut.ee; Choong Seon Hong, Kyung Hee University, Yongin, South Korea, cshong@khu.ac.kr.

1 INTRODUCTION

Traditional Industrial Control Systems (ICS) include operational technology (OT) such as supervisory control and data acquisition (SCADA) systems, programmable logic controllers (PLCs), and other field devices that are able to command and control physical processes within industrial environments [12]. In the course of Industry 4.0, this industrial infrastructure is increasingly integrated with general-purpose Information Technology (IT) systems by wireless connections with cost-effective miniaturized devices having sensing, computation, and communication capabilities to collect, process, analyze, and interpret data. Whereas integrating IT and OT systems provides promising solutions in a plethora of ubiquitous industrial ecosystems, it introduces novel attack vectors. Loopholes in the system infrastructure lure adversaries to launch different attacks; for instance, denial of participation, false data injection, data forging, and replay attacks, to name a few [66]. Such attacks can have catastrophic consequences for the high-assurance applications that are involved in crucial decision-making processes based on sensor data [51]. Thus, to provide protection against data forgery, potential countermeasures for data security highlight the need for rigorous scrutiny to thwart vulnerabilities, enforce verifiable capabilities, and ensure integrity.

Considering the involvement of multiple entangled participating entities in intricate industrial processes, the overarching question is: *How to ensure the trustworthiness of data collated from disparate data repositories?* In this regard, blockchain technology gained ground to address challenging issues in the industry with regard to product lifecycle data management and data security [64]. Utilizing blockchain allows industries to manage data on a shared distributed ledger to solve important glitches in traceability, and record events in a secure, immutable, and irrevocable way [11]. Relying on the trustworthy data from the blockchain and inputting the retrieved data to a data-driven system to draw intelligent information can certainly open ways for the industries that consider several factors for maximizing their revenues, such as risk prognosis, cost management, and quality assurance [65].

For predictive maintenance, one of the promising solutions is to create a digital fingerprint (or virtual replica) of the underlying product, process, or service. Such a replica can be used to analyze, predict, and optimize all operations before carrying out its real-world implementation. Following a closed-loop, the simulation data are fed back to the physical object to calibrate the operations and to enhance the system performance. Such a bi-directional mapping between *physical space* and *virtual space* is called *Digital Twin* (DT) [68]. As physical assets begin to operate, DTs collect and integrate data from multiple sources, such as sensor data from factory floor, historical production data acquired from product lifecycle data, and domain knowledge to generate comprehensive data in the form of models, simulations, replications, or behavioral analytics. During industrial processes, DTs run synchronously with their physical counterparts where the primary objective is to track and trace *data inconsistencies* between the physical and virtual objects. Data inconsistencies between the physical object and virtual object call for the adoption of better calibration and testing strategies that evolve DT models and physical counterparts to support problem-solving in manufacturing, more accurate estimation, prediction, and optimization of the industrial processes [71].

Combining DTs and blockchain can reshape the industry such that blockchain ensures secure data management and DTs use trustworthy data as input to extract actionable insights for predictive maintenance. For instance, in a manufacturing plant for asset management, on the one hand, upon utilizing a trust architecture (such as [10]) for sensor data cross-validation, trustworthy key threshold data from the sensors and actuators (affixed with equipment) can be stored and retrieved from the blockchain and can be used to detect trends (such as excessive vibration or heat) to avoid failures or operator injury. On the other hand, through DTs, maintenance and operations personnel can acquire more thorough and accurate insight into asset performance based on the analytics and cognitive data generated on the factory floor. Also, other regulators and suppliers of plant

equipment can monitor the data logged at blockchain for preventive maintenance, conduct timely inspections to ensure equipment reliability in addition to record their work or analysis on the blockchain ledger [45].

In this survey, we try to answer the following research questions (RQs.):

RQ1. Do we really need blockchain in DTs?
RQ2. What are the current research trends of DT-, blockchain-, and security-related design solutions?
RQ3. How AI will set a course for blockchain-based DTs?
RQ4. What are the open research and deployment challenges with blockchain-based DTs?

1.1 Existing Works

To the best of our knowledge, no survey has been conducted that thoroughly investigates design and implementation issues in blockchain-based DTs. However, some of the recent work emphasized the integration of blockchains for DTs. Starting with the most relevant paper, in [78], the authors taxonomize DTs while providing insights into DTs levels, design phases, enabling technologies, and their potential in core applications. The authors explore the benefits of using blockchains for DTs and discuss DTs synergies and case studies for the automobile industry including Genie, Volvo, and Volkswagen. This study outline technical and social barriers and identify open issues to the adoption of blockchain in DTs, for instance, scalability, data privacy, limitations due to resource-constrained devices, and interoperability issues due to heterogeneous consensus models, transaction schemes, and smart contracts. Also, it outlines security features such as confidentiality and integrity to protect DTs design and operational information, however does not consider other security-related design details such as access control and quantum immunity. A more recent theoretical work [58] covers applications and benefits of DTs in different industry verticals including manufacturing, automotive, logistics, utilities, aircraft, and healthcare. The focal point of this work is to create blockchain-based DTs to facilitate digital identity and data tracking through blockchain traceability. Besides, it also highlights other features of blockchain including decentralization, safe and secure data transfer, data immutability, authentication of users and data sources in connection with DTs industrial use cases. Despite providing a succinct overview of the blockchain-based DTs, these works do not cover the existing literature on blockchain-based DTs approaches.

In [25], the authors explore the interoperability and integrability of blockchains and DTs for asset life cycle management. They proposed a model of the blockchain-based DT framework based on surveys of existing literature and questionnaire to industry professionals and mapping of technological use cases for DTs, IoT, blockchains, and smart contracts. This work also outlines integration constraints of blockchain-based DTs with possible solutions. For instance, focus on separable blockchains for easy troubleshooting and combating errors arising during scaling networks and collaboration of the different system areas. Nevertheless, the paper does not survey works on blockchain-based DTs in particular. Similarly, a blockchain-oriented work [41] focuses more on the role of blockchain in manufacturing with fewer insights into DTs, whereas a DT-oriented study [60] merely touches blockchain technology as a key to solve trust in information issues.

By narrowing down our survey to blockchain-based DTs featuring industrial use cases and focusing on more high-level issues, we present a holistic approach towards blockchains for DTs while presenting design components, fusion with enabling technologies, benefits, and challenges. To this end, we also cover the existing works based on the proposed design and development strategies through a survey of blockchain-based DTs. We primarily focus on the existing literature that covers the design and implementation aspects of blockchain-based DTs. We choose manifold criteria that cover different design aspects related to DTs, blockchain, and security. In addition to comparison of current solutions, we also emphasize challenges that are yet to be addressed.

1.2 Scope of This Survey

In this survey, we present a thoroughly comprehensive, and systematic review of state-of-the-art results for blockchain-based DTs. We also focus on blockchain-based DTs design and implementation issues in detail and provide possible solutions to address such issues. Additionally, we investigate technical, logistical, and social issues that arise due to the integration of blockchain with DTs and must be addressed by stakeholders. We summarize the main contributions of this paper below:

- We present an in-depth comprehensive and systematic review of the state-of-the-art blockchain-based DTs approaches covering design and implementation issues in terms of blockchain, DTs, and security.
- Based on the limitations of the current works, we propose a framework for trustworthy blockchain-based DTs.
- In addition to ongoing standardization efforts, we provide an in-depth review of various research challenges such as technical, logistical, and social challenges. In contrast to existing surveys, we also map such requirements from both blockchain and DTs perspectives, provide pragmatic approaches for the adoption of blockchain-based DTs, and highlight the open-ended challenges that need to be addressed by the industry and research community.

The rest of the paper is organized as follows: Section 2 covers a quick high-level overview of DTs, blockchain, and the major benefits of using blockchain for DTs. Considering blockchain, DTs, and security as the key factors, we present the design and implementation details of existing research in Section 3. Section 4 discusses the role of enabling technologies in blockchain-based DTs. Section 5 describes the technical, logistical, and social challenges and requirements of blockchain-based DTs. Finally, Section 6 concludes the paper with takeaways from this study.

2 BACKGROUND

This section outlines the necessary background information for this survey. We first provide the main design components and applications of DTs in the industry followed by a discussion on blockchain and its role in ensuring secure data management in the industry.

2.1 Digital Twins

The concept of DT was first introduced by Micheal Grieves in 2003 during his course on Product Lifecycle Management (PLM). In 2012, the National Aeronautics and Space Administration (NASA) formalized the definition of DTs as an integrated multi-physics, multi-scale, probabilistic simulation mirroring the physical system based on models, sensors, history data, etc. [24] in the field of the aerospace industry. In 2014, Grieves' white paper [26] proposed a preliminary form of a three-dimensional (3D) DT model that includes a physical space containing a physical entity, a virtual space containing a virtual entity, and a connection to allow synchronization of data between both spaces (as shown in Fig. 1). In continuation of the role of DTs in the aviation industry, DTs lead to the notion of Airframe Digital Twin (ADT) for predicting the life of aircraft structure and maintenance needs [73]. In parallel, continuous advances in DT enabling technologies, for instance, IoT, big data processing techniques, cloud computing, and AI further motivate the adoption of DTs beyond the aerospace industry and thus are promoted in numerous applications including manufacturing, supply chain, smart cities, industrial IoT, servitization, healthcare, etc. According to Gartner's survey report, 13% of enterprises are already using DTs, whereas 62% are in the process of deploying DTs [23].

Based on the Grieves' 3D DT model, Tao et al. proposed a five-dimensional (5D) model that includes: physical space, virtual space, connection, DT data, and service (as shown in Fig. 2) [72]. The physical part supports the virtual part through assets data while the virtual part controls the physical part through simulation and decision-making. Service supports the management of both spaces during operation, evolution, and optimization.
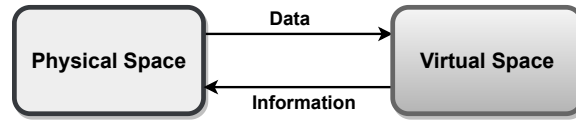
Fig. 1. Adaption of Grieves' 3D DT model [26].

DT data repository acts as a primary source for knowledge creation and connection bridges all parts of the DT model.
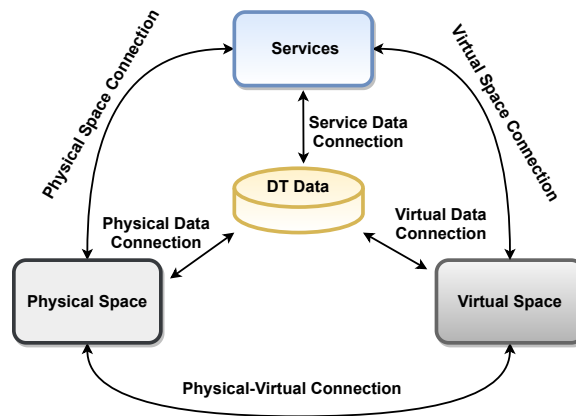

Fig. 2. Adaption of 5D DT model [72].

The data-driven DTs focus on the areas of product design, production, manufacturing, operational control, Prognostic Health Management (PHM), etc. in the industrial domain [70]. For instance, DTs enable more accurate predictions, rational decisions, and informed plans by reinforcing the collaboration between design and manufacturing phases. PHM plays a significant role in the lifecycle monitoring of equipment conditions where a DT of the equipment supported by data is used to depict the erratic behavior of the physical object, perform fault diagnosis, and provide design rules for maintenance [22]. Similarly, dispatching optimization and operational control are among other promising application areas of DTs [70]. Industry leaders including International Business Machines Corporation (IBM), Siemens, General Electric (GE), Microsoft Azure, Airbus, Systems, Applications and Products in Data Processing (SAPSE) have deployed DTs in various industrial application areas.

2.2 Blockchain

The emerging technology of blockchain has attracted immense attention from academics and practitioners in diverse disciplines including finance, law, and computer science due to its promising secure distributed framework that facilitates sharing and synchronization of information across all participating entities [55]. In the blockchain, transactions among nodes serve as the basic communication primitives that are recorded as append-only time-stamped logs in a shared digital ledger. To establish a secure trusted network among untrusted nodes, a consensus mechanism is employed that involves solving a resource-demanding hard-to-solve and easy-to-verify cryptographic puzzle. Following the consensus algorithm, transactions are collated to form a block which is then appended to the ledger. To enforce immutability in blockchain structure, each block contains the hash of the previous block, thereby can readily detect tampering to a block and hence transactions.

The rapid adoption of blockchain has largely been spurred by the success of Bitcoin (the first prototype of cryptocurrency) [50] and successfully replaced economic transaction systems in many organizations. Other alternatives (Ethereum and Ripple) continue to mature, thus proving to be a highly disruptive force in finance and e-commerce sectors [28]. Later on, numerous non-monetary application scenarios are identified as potential candidates of blockchain-based solutions, for instance, intellectual property, proof of location, voting, healthcare, to name a few. Recently, blockchain-based solutions in combination with IoT are gaining ground to revamp heterogeneous business models (such as Business to Business (B2B) and Business to Consumer (B2C)) in a variety of industrial sectors, including supply chain and manufacturing plant asset management [45].
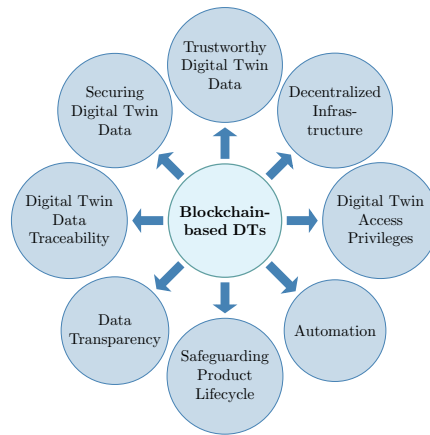

Fig. 3. Benefits of using blockchains for digital twins.

2.3 Major Benefits of Using Blockchain-based Digital Twins in Industry

Although DTs enable the realization of futuristic services and applications of industry in a transparent and efficient way, these benefits of DT are based on an assumption about data trust, integrity, and security. Nevertheless, in real-life scenarios, data breaches could occur due to a number of reasons, both malicious and otherwise. Therefore, to fully harness the features of DT, the data used by the DTs needs to be trustworthy and secure. In this regard, leveraging blockchain technology allows industries to manage data on a distributed ledger while ensuring trusted DTs data coordination. Nevertheless, we need to critically analyze the question, *do we really need blockchain for DTs?* based on the assessment elements including the number of readers/writers of the data, presence of a trusted third party, identification of readers/writers, and the nature of verifiability of the exchanged data [77]. According to the model of Wust et al. [77], a permissioned blockchain can be the right candidate for DT to cope with these issues because of, but not limited to, the cryptographic primitives of blockchain guarantee data immutability and the consensus mechanism of the blockchain (where different stakeholders of the data come to a common consensus while managing the data) provides transparency and accountability. In the following, we discuss the key benefits (summarized in Fig. 3) of using blockchain-based DTs.

2.3.1 Trustworthy Sources. The physical-digital mapping in the industrial processes is achieved through IoT-based technologies that might be vulnerable to cyber threats [67]. Since, the accumulated data accounts for rational decision making, therefore, it is important to manifest the trustworthiness of data-generating sources. Blockchain can play a significant role by collecting data from registered devices to proactively circumvent device

tampering or utilizing a blockchain-based layered trust architecture [10] for cross-validation of sensor data. A more detailed discussion is provided in Subsection 3.2.8.

2.3.2 Distributed Infrastructure. The lifecycle of DT span over different phases where multiple stakeholders (may be geographically dispersed) carry out various operations on it. The centralized data management architectures are not well suited for such environments. Additionally, the disparate data silos act as a barrier in orchestrating the DT lifecycle particularly in digital thread that requires linking data throughout various phases of the lifecycle. To enable the storage of information in a distributed manner with no single point of control can be established through blockchain. Thus, not only the DTs' data but data from other sources (such as data generated within the production facility or during the supply chain trade events), provenance data, etc. flowing across multiple tiers or entities can be stored and accessed through the blockchain ledger [64].

2.3.3 Securing Digital Twins Data. The increased digitization and connectivity opens up new attack vectors that may put industry assets or process at risk whereas DTs are required to be built on trusted data. The question how to provide integrity and trustworthiness of data collected from the physical world that are then shared with the digital world needs to be answered. The tamper-proof and immutable nature of blockchain can be leveraged to procure data across multiple participating entities in a secure manner. For instance, the cryptographic strength blockchain keeps the irrevocable history of DTs creation or access transactions. Furthermore, the use of pseudonyms, such as public keys for representing node identities, for blockchain transactions can support the coexistence of transparency and privacy. Nodes can maintain multiple pseudonyms to remain anonymous, to enable confidential data flows, and to cope with industrial espionage for competitive advantage. Thus, DT data can be shared among multiple untrusted (or partially trusted) lifecycle parties while ensuring confidentiality, integrity, and availability.

2.3.4 Data Traceability in Digital Twins. Additionally, provenance-enabled blockchain-based DTs can reason about the current state of a data object such as *why*, *where*, and *how* throughout the process chain [64]. The complete lineage of DTs ranging from their origin to the present can be accessed to identify the cause of deviations between the physical and digital space in addition to data assessment (such as data debugging, reconciliation, performance tuning, auditing, etc.).

2.3.5 Access Privileges of DTs. In the case of the enterprise industry, both reading and modification to the data are subject to access rights. This argument can be true for DT as well where the data can be viewed by the selected set of entities and the modifications must be done by only authorized entities. Blockchain-based access control manages policies and rules defined for DTs. Moreover, blockchain inherently retains the history of modifications and thus can circumvent the problem of illegal data modification that may lead to other data security-related problems.

2.3.6 Safeguarding Product Lifecycle Data. To facilitate an abstract view of overall phenomena, DTs must support the correlation of different DTs associated with different physical sub-components [67]. For example, Digital Twin Instance (DTI) is a singular entity that mirrors the physical asset and/or activity, whereas Digital Twin Aggregate (DTA) is the combination of DTIs (preferably similar DTIs) to mimic the bigger picture of the physical world [58]. The DTA represents an attractive target for attacks and may lead to severe consequences as the whole product line can be exposed to data breach (for example, manipulated updates to put the DT into a malicious state). Blockchain-empowered DTs provide adequate security measures to protect DTIs within the DTA by allowing only authorized entities to make changes in addition to data integrity and trust as discussed in 2.3.3.

2.3.7 Automation Through Smart Contract. The industrial processes demand automation to trigger timely and seamless actions. For instance, logging data in the course of change in physical state and its counterpart. To

enforce automation, smart contracts from blockchain can be deployed to automate scenarios depending on the underlying application requirements. Subsection 3.2.4 presents more examples on this.

2.3.8 Combating the Problem of Counterfeits. Creating counterfeit products that are near-replicas of the original products not only affect the investment of the industry stakeholders but also badly affects the brand value and reputation of the products. To overcome this challenge, the combination of blockchain and DT can ensure that counterfeit products are identified by providing product authenticity. More precisely, the transparency will come from the blockchain (for instance, creating digital certificates on the blockchain) whereas the digitization of the products and assets will be guaranteed by the DT. Therefore, the integration of DT and blockchain will play a pivotal role in ensuring the Return on Investment (RoI) for the industry by removing the problem of counterfeits. Additionally, blockchain-based solutions for DTs can provide verifiable digital certificates to solve important glitches regarding the legitimacy of DTs by identifying counterfeiters attempting to imitate or distribute fake replicas of DTs [58].

3 BLOCKCHAIN-BASED DIGITAL TWINS: RESEARCH TRENDS

In this section, we review the existing works that focus on the design and implementation aspects of blockchain-based DTs. To discuss the existing works, we choose manifold criteria that cover different design aspects related to DTs, blockchain, and security. In addition to discussion on current solutions, we also emphasize challenges that are yet to be addressed. Furthermore, we provide a detailed discussion on a trustworthy blockchain-based DT that tries to address some of the outstanding challenges in the existing works. Fig. 4 shows the research trends of exploring the integration of blockchain with DTs by the research community. It can be clearly seen from Fig. 4 that there is an increased interest in the integration of blockchain with DT. Fig. 5 depicts the taxonomy which is devised based on the design solutions for blockchain-based DTs in industry. Table 1 comprehensively summarizes relevant recent works on blockchain-based DTs. This table shows three main interesting design considerations for DT, blockchain, and security from the blockchain-based DT literature. Firstly, it covers DT-related design details including DT definition, component composability and integration, lifecycle phases, and implementation aspects. Secondly, it outlines blockchain-related design details including the type of blockchain in the proposed solution, implementation aspects, off-line data storage mechanism, traceability, and trustworthy data sources. Thirdly, it highlights the security-related design details including encryption, access control, and quantum resilience. In the following, we discuss each of them in detail.
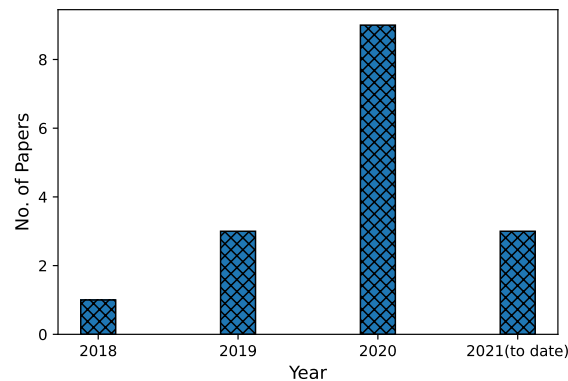

Fig. 4. Blockchain-based DTs in industry: Research trends.

Table 1. Comparison of blockchain-based digital twin approaches.

Ref.	Digital Twin-related Design Details									Blockchain-related Design Details								Security-related Design Details		
	Definition	Component	Lifecycle Phases	Implementation						Implementation								Encryption	Access Control	Quantum Resilient
				Data Fusion	Inter-operable	S&S Rules	Data Sync	Granular Data	Interface	Type	Scalability	Smart Contract	Consensus Protocol	Prototype	Data Storage	Traceability	Trustworthy Sources			
[1]	machine events	✗	M	✗	✗	✗	✗	✗	✗	ETH	off-chain	✓	PoW PoA	◑	◑	◑	✗	✗	◑	✗
[13]	any asset	✗	E&M	◑	◑	◑	◑	✗	✓	?	off-chain	✓	✗	✗	◑	◑	✗	◑	◑	✗
[44]	product	✗	E&M	✗	✗	✗	✗	✗	✗	?	✗	✗	✗	◑	✗	✗	✗	✗	✗	✗
[81]	manufacturing events	✗	E&M	◑	◑	✗	✗	✗	✗	HLF	off-chain	✓	XFT	◑	◑	◑	✗	✗	◑	✗
[30]	product	✗	E&M	✗	✗	✗	◑	✗	✗	?	✗	✓	PoW	◑	✗	✗	✗	◑	✗	✗
[27]	any asset	✗	E&M	✗	✗	✗	✗	✗	✗	ETH	off-chain	✓	✗	◑	◑	◑	✗	✗	✗	✗
[42]	manufacturing events	✗	E	✗	✗	✗	◑	✗	✗	HLF	off-chain	✓	XFT	◑	◑	✗	✗	✗	✗	✗
[37]	product	✗	E	✗	✗	✗	✗	✗	✗	variant	✗	✗	PoW	✗	✗	✗	✗	✗	✗	✓
[56]	any asset	✗	L	✗	✗	✗	◑	✗	✓	ETH	off-chain	✓	✗	✓	✓	✗	✗	✓	✓	✗

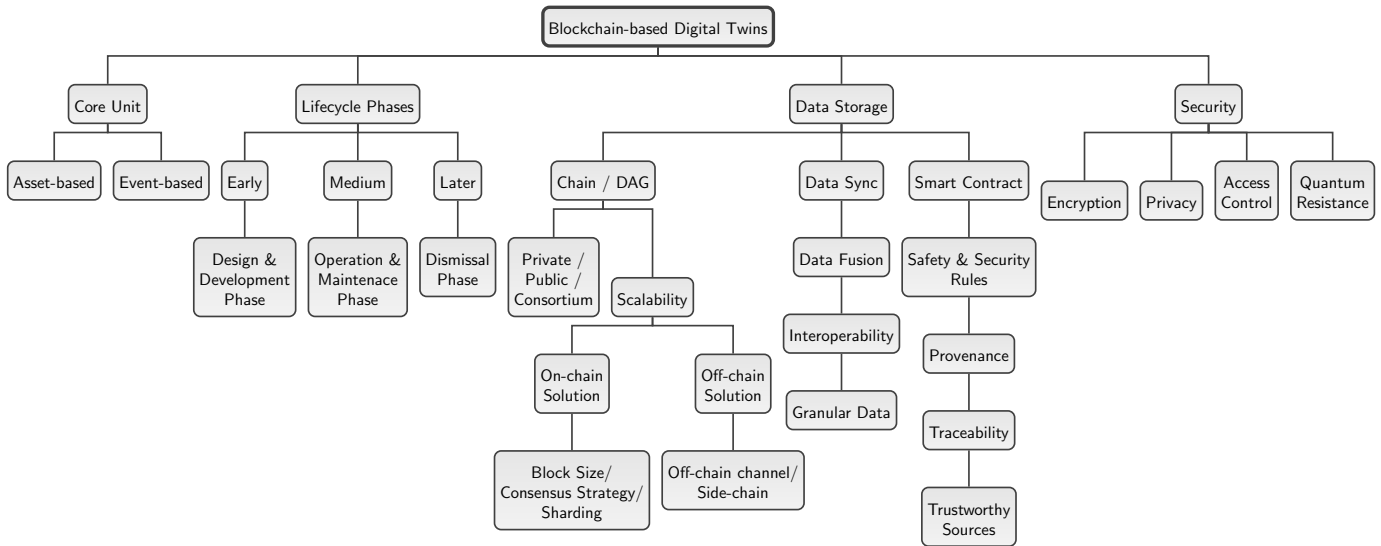Legend: ? Unknown; ✗ Not covered; ◑ Partially covered; ✓ Covered; **E** Early; **M** Medium, **L** Later.

Fig. 5. Taxonomy: Blockchain-based digital twins

3.1 Design Solutions for Digital Twins

The majority of existing works consider DTs as a representation of any asset, product, machine, or manufacturing events. However, the existing works do not cover the integration and interoperability of multiple components or sub-components to provide an aggregated view of the complex physical system. For instance, combining DTIs of the physical asset and/or activity to form DTA to generate the bigger picture of the physical world. As a result, the lack of correlation among different elementary DTs leads to the failure of abstract views of larger phenomena. In the following, we discuss the design solutions for DTs from different perspectives such as design lifecycle, fusion and interoperability, security and safety, and synchronization and consistency. These are the cornerstones of DT design and we discuss both the existing works in these directions and the outstanding challenges.
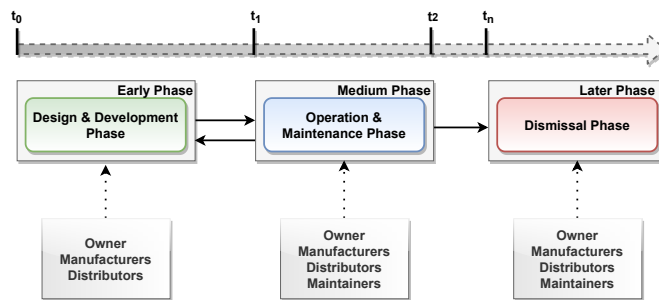

Fig. 6. Lifecycle of digital twins: From design to dismissal.

3.1.1 Digital Twin Lifecycle. Fig. 6 represent a DTI where the data linkage throughout various phases of the lifecycle can be performed through digital thread. The lifecycle phases of a DTI start from design to dismissal, in

addition to the potentially involved lifecycle parities (i.e., owner, manufacturers, distributors, and maintainers) [4]. Firstly, the asset prototype is designed and is evolved based on historical data, static data, and simulation data during the *design phase*. Then the physical asset is tailored based on the asset prototype and other technical requirements during the *development phase*. Secondly, following a synergic and continuous interaction between the prototype and the asset, the asset further evolves and optimizes its functionalities during *operational and maintenance phase* while the prototype grows and self-adapts to the asset during its existence. Finally, due to obsolescence or other reasons, the asset along with its DTs is demolished during the *dismissal phase*. The existing works have usually focused on tackling early ([37, 42]), medium ([1]), or both of these phases ([13, 27, 30, 44, 81]) of the DTs lifecycle. While developing a prototypical blockchain-based DT solution called *EtherTwin*, Putz et al. [56] mentioned the dismissal phase during the DT lifecycle; however, they do not discuss the details such as how the object is disassembled and the stored historical data of the DT product are backed-up and made available to other objects or domain experts to optimize the production of future devices.

3.1.2 Data Fusion and Interoperability. Regarding the implementation details of DTs, fusion and interoperability of data from multimodal and heterogeneous sensors is one of the significant prerequisites. This involves cleaning invalid, duplicate, or missing data, converting heterogeneous data formats into a unified data structure, and then fusion of such data to generate a consistent interpretation of a certain object. As depicted from Table 1, none of the existing works consider the process of data fusion from technical perspective. Similarly, interoperability is partially addressed by [13, 81]. For example, a hardware infrastructure proposed by Zhang et al. [81], introduced adopters for parsing the data collected from systems with different communication protocols. Additionally, it includes edge gateway nodes to pre-process the perceived data through filtration and aggregation. Similarly, Dietz et al. [13] discussed the integration of data in multiple formats to support data variety. They categorized the data as semi-structured data to model physical asset (static properties of the device) by using AutomationML (AML) format, structured data (sensor data), and unstructured data (human-readable documents). However, the above-mentioned theoretical solutions lack implementation details to demonstrate their feasibility in practice. Interoperability and data fusion provides a more comprehensive, consistent and accurate representation compared with the single perspective of data. However, the existing works still lack due attention to the relation of data fusion with the DT functionality.

3.1.3 Safety and Security (S&S). Some other implementation aspects including Safety and Security (S&S) threats, malicious or accidental, must be integrated to detect misconfigurations and to mitigate the outcomes of malfunctioning components and malicious activities in a production system. For example, safety rules may define upper- and lower-speed thresholds on a motor [20]. Inability to impose such rules leads to devastating results on the system in the form of cyberattacks or Advanced Persistent Threats (APT), for instance, Stuxnet worm [39] and Triton [46]. Despite being a critical construction parameter of "security by design", S&S rules are overlooked by existing works. To the best of our knowledge, only one of the existing works by Dietz et al. [13], suggested adopting S&S works by referring to interesting work by Eckhart and Ekelhart [20]. In this work, the authors consider monitoring of S&S rules and demonstrated the detection of the notorious Man-in-The-Middle attack (MiTM) attack by targeting the manipulation of a motor's speed. Moreover, to further explore this parameter, future directions such as modeling language for S&S rules, investigating device templates with security rules, behavior learning, and analysis in conjunction with anomaly detection are also highlighted. Introducing S&S mechanisms at the DT design phase can help to lower security and incident-response costs thereby making later lifecycle phases less prone to errors and incidents [12]. Whereas, S&S rules must be stored and retrieved from the ledger to ensure their reliability. This further strengthens the rationale for integrating blockchain with DTs.

3.1.4 Synchronization and Consistency. Another key factor of DT-related design is the synchronization and consistency of data between the physical and digital spaces. Both of these factors are highly associated with the

volume and velocity of data generation and data storage. DT data can be categorized as descriptive (or static) data and behavioral (or dynamic) data. The latter changes infrequently with time along the lifecycle of the real-world counterpart whereas the former evolves frequently with each lifecycle phase resulting in different versions and a dynamic information structure [61]. In the existing works, the issue of supporting data volume and velocity is partially considered. For example, Huang et al. [30] suggested storing sensor data based on preset time intervals. However, they did not consider the fundamental requirement of real-time data synchronization. On the other hand, Dietz et al. [13] discussed a more feasible solution to store the hash reference to the data stream on the ledger while storing data off-chain. Although this solution addresses the data storage issues, the details subjected to performance metrics are still missing. Similarly, Putz et al. [56] suggested the off-chain data storage pattern to meet the data volume and latency requirements and suggested sharing only non-nominal or aggregated sensor data to reduce the data volume.

Leng et al. [42] suggested a multi-view synchronization-based DT model that uses blockchain as an indexing server to record the transactions that are critical for the life-cycle quality control and product anti-counterfeiting. They also use traditional databases for the fast retrieval of history transactions. Considering the existing solutions, the crux of the problem is that the dynamic data produced by the deployed sensors and actuators providing real-time updates about the asset's actual state usually amass in intervals ranging from minutes to milliseconds. However, sensor data streams' velocity and volume exceed the performance characteristics of mainstream blockchain frameworks. Furthermore, deciding which data to store on-chain and which to off-chain, also needs due attention.

To address data synchronization challenges, the blockchain-based DTs solutions must exhibit high throughput and low sharing latency to efficiently sync data between physical and virtual space in addition to improving data collection and using it intelligently to manage big data challenges.

3.1.5 Data Granularity. Data is considered as the fuel for ensuring precise decision-making while at the same time quality of the data matters at par with quantity for decision-making. Therefore, the amount of collected data holds as an important design parameter. The granularity level (fine- and coarse-grained) of collected data which in turn also depends on the optimal number and placing of sensors [69], is still missing in the existing works. Further gaps that are required to be addressed include what types of data to sample, which sensors to use and where to install them [38].

3.2 Design Solutions for Blockchain

In the following, we discuss blockchain-related design solutions (for instance, selection of blockchain, implementation aspects, on-off chain storage, traceability, and trustworthy data sources) proposed by the existing works and associated challenges.

3.2.1 Implementation. While most of the existing schemes proposed only a theoretical blockchain-based framework ([13, 37, 81]) neglecting the prototypical implementation. However, several works such as [27, 44] only partially accomplished prototype implementation. To date, only Putz et al. [56] carried out a thorough implementation. EtherTwin [56], an open-source prototype, is based on blockchain design patterns and DApp technologies with performance and cost measurements. Some other works such as [30, 44] do exist in the literature but do not provide any technical details about the blockchain type and other evaluation metrics. Similarly, Khan et al. [37] proposed a quantum-immune blockchain variant; however, it lacks technical and other supporting implementation details. On the other hand, in addition to other evaluation metrics, providing a user interface is another important aspect of implementation that is suggested by [13], however, in [56] they design and additionally implement it.

Overall, the existing works are not extensively evaluated to gain additional practical insights in the industry, and thus require focusing more on the prototype to identify potential adoption barriers and to provide avenues

for future research. Moreover, validation of the prototype's practical suitability by the industrial experts can further bridge the gap between industry and academia.

3.2.2 Blockchain Type. The blockchain type can be categorized with respect to structure (chain and Directed Acyclic Graph (DAG)) and with respect of access mechanism (public, private, consortium). The blockchain type must be selected depending on the underlying application requirements. However, the current DT-based blockchain solutions mostly adopt mainstream sequential chain-structured public or private blockchains, such as Ethereum (ETH) or Hyperledger (HLF). For instance, [1, 27, 56] use ETH while [42, 81] use HLF as the underlying framework for DTs. The main reason for such trend is due to the increasing number of implementations targeting different application domains while other Distributed Ledger Technologies (DLT) may not have reached the required level of maturity in terms of implementation and may not be mature enough for mass adoption yet.

3.2.3 Scalability. The selection of apropos blockchain is of significant importance to avoid scalability snag that may arise due to the futuristic surge based on the growing number of actors and activities in the Cyber-Physical Systems (CPS). The existing solutions of solving the scalability of blockchains include on-chain solutions (increasing block size, improved consensus strategies, sharding, and blockchain data structure) and off-chain solutions (off-chain channel and side-chain) [83]. Among the current blockchain-based DTs solutions, scalability issues are addressed by off-chain solutions are (as discussed in Subsection 3.2.6). Considering the data structure of blockchain, the DAG-structured blockchains (IOTA [53], NANO [40], Byteball [9]), lightweight chain-structured blockchains (LSB [19], PoET [32]), or alternative structures (Tree-Chain [16]), must be considered.

3.2.4 Smart Contracts. A smart contract is an automatically executable code running on the blockchain to control digital assets by enforcing an agreement between untrustworthy parties [76]. Smart contracts can be deployed to automate scenarios depending on the underlying application requirements. When deployed into a DT environment, the smart contracts can be used to track data sharing mechanism [30], to represent twin-creation transactions [27], to store authorization information for all involved parties [13, 56], or to automate event-based interaction among machines [1]. Other examples that may involve smart contracts utilization in CPSs are triggering S&S rules (as discussed in Subsection 3.1.3), invoking PLC functions due to changing states of physical processes, modifying simulation setup parameters, or auditing DTs by actively or retroactively monitoring smart contract transactions.

3.2.5 Consensus Mechanism. To keep the ledger state synchronized across the network while establishing a secure trusted network among untrusted nodes, consensus mechanism is performed by the participating nodes. Consensus algorithm involve solving a resource-intensive hard-to-solve and easy to verify puzzle. We do not describe the working of consensus algorithms here, as there are other works describing different consensus strategies of blockchain, for example [5, 83]. Among the existing works on blockchain-based DTs, Proof of Work (PoW) is used by [30, 37] without stating any convincing reason for this adoption. [1] tested both PoW and Proof-of-Authority (PoA) to support the argument that the consensus mechanism must be chosen based on the underlying application. For example, PoW exhibits high computing power in comparison to PoA whereas PoW is more suitable for partially trusted environment and PoA is well-suited for permissioned environment. However, [42, 81] use Cross Fault Tolerance (XFT) to achieve higher throughput and lower latency. Similarly, [56] suggested using resource-efficient byzantine-fault tolerant consensus algorithm such as Istanbul Byzantine Fault Tolerant (IBFT) as a part of future improvements to their prototype.

3.2.6 Data Storage. In terms of data storage, most of the existing works suggest using off-chain storage [1, 13, 27], however, do not properly implement it. For instance, Hasan et al. [27] suggested using InterPlanetary File System (IPFS) storage but the details on how IPFS is used, are still missing. In comparison to other works, Putz et al. [56] implemented off-chain storage based on swarm Distributed Hash Table (DHT). Following this approach, the

off-chain data storage can meet the data volume and latency requirements where sensor feed is stored off-chain after registering sensor on-chain.

The idea of off-chain storage can help to solve the scalability issues in addition to reducing on-chain storage costs in the traditional blockchain. However, integrating off-chain storage solutions with blockchain-based DTs have several key challenges, e.g., there is no guarantee on the validity of the off-chain transactions, off-chain storage requires centralization, state consistency with on-chain storage in a real-time manner, and security issues [75].

3.2.7 Traceability. Though blockchain ensures traceability through its append-only nature, it requires the selection of key parameters to track and trace the entire chain throughout the process. Some of the existing works ([13, 27, 81]) discuss traceability but do not illustrate its working mechanism or actual implementation. To enforce traceability in the industrial system, provenance data based on ⟨*source, destination*⟩ pair and other key identifiers can be constructed (as suggested in [65]) while traversing through the process and can be stored on the blockchain to reconstruct the process chain on demand.

3.2.8 Trustworthy Sources. While stringent security solutions are inherited from the blockchain, ensuring the *trustworthiness* of data-generating sources and data in transit is, by far, not considered by most existing works. Maliciously or mistakenly, data adulteration may occur due to various reasons such as device tampering, data forging, or maliciously participating entity thereby causing the Garbage In Garbage Out (GIGO) problem [65]. Given that blockchain mechanisms do not guarantee the trustworthiness of data at the origin, [10] proposed a blockchain-based layered trust architecture for IoT for improving end-to-end trust. In this proposed framework, at the data layer, the data trust module quantifies the confidence in observational data based on (i) evidence from neighboring data sources, (ii) reputation of data sources depending on their long-term behavior, and (iii) confidence level of the observation reported by data sources. To address such issues in blockchain-based DTs, Suhail et al. [64] suggested the following measures:

- To circumvent device tampering, the sensor nodes are required to register as authorized devices through their public keys and digital certificates in the blockchain. Such measures enable the gateways to collect data from authorized devices only by first confirming their authenticity from the blockchain. Furthermore, firmware updates are also carried out through blockchain to avoid malicious code injection or malware.
- To diagnose data forgery during data transition, potential solutions involve, i) reliance on provenance data to identify the faulty entity in the underlying chained processes and ii) leveraging twinned data to identify anomalies in the system that can be detected based on the digital-physical mapping.
- To identify a rogue participating entity through, i) issuing or revoking participation certificates by conducting a periodic on-site inspection of the units to ascertain that industrial processes are abiding by environment-friendly operations and ethical practices, and ii) deploying a monetary punishment mechanism [59] to reward honest entities while penalizing or revoking trader's participation.

3.3 Design Solutions for Security

In this subsection, we discuss the existing works pertaining to security-related design issues including encryption, access control, and quantum resistance.

3.3.1 Encryption. Ensuring confidentiality through encryption is proposed by some of the works but is not described in detail or properly implemented. For example, Huang et al. [30] suggested encrypting the product lifecycle data; however, the implementation details of such encryption mechanisms are not provided which makes it hard to analyze. Similarly, Putz et al. [56] encrypt data before uploading it to the DHT using AES-256 whereas permissions are enforced by encrypting a symmetric key with the corresponding entity's public key. It is also an established fact that the security of de-facto cryptographic primitives is at risk of being broken by

quantum computers [8]. Therefore, to address this critical issue, one set of solutions is to use quantum-resistant public-key cryptography (also called Post-Quantum Cryptography (PQC)) as drop-in-replacements [7]. Based on mathematical problems other than factoring and discrete logarithms (used by the traditional cryptographic algorithms), such solutions are yet believed to be secure against quantum attacks. Another set of solutions is quantum cryptography, most notably, Quantum-Key Distribution (QKD), which is a method for establishing symmetric keys [48]. However, due to certain limitations (i.e., presence of quantum channel for sending bits), it is difficult to use such a solution as a plug-and-play replacement for the current security methods [49].

3.3.2 Access Control. The current literature on blockchain-based DTs has a few authentication and authorization mechanisms as the concomitant of access control. However, we could not find their implementation details. For instance, Dietz et al. [13] proposed a stored role-permission mapping to grant or deny access based on Role-Based Access Control (RBAC). Similarly, Putz et al. [56] implemented a hybrid access control model by combining RBAC and Attribute-Based Access Control (ABAC). The access control capability provides authentication and authorization to enable data sharing among the involved parties and therefore must be considered carefully.

3.3.3 Quantum Resilience. With the arrival of quantum computers, wrongdoers are able to forge digital signatures and sabotage transactions thereby putting blockchain security at risk [21]. To prepare for the quantum apocalypse, blockchain-enabled schemes that already support post-quantum techniques such as IOTA [53] (using WOTS), Quantum Resistant Ledger (QRL) [57] (using XMSS), and Corda (using BPQS: a single-chain variant of XMSS) should be considered for long term security [66]. Moreover, to further enhance security and enable blockchain to become faster and more efficient, requires quantum-linked features including quantum Internet, quantum-safe communication, and QKD for both communication and computational processing of blockchain data [21]. Among the existing works, Khan et al. [37] proposed a variant of blockchain called Twinchain which replaces Elliptic Curve Digital Signature (ECDSA) with quantum-safe Hash-Based Signature (HBS). However, the proposed quantum-resilient blockchain does not provide sufficient details about the chosen HBS. To this end, there are still outstanding challenges and research avenues that could be investigated by the research community.
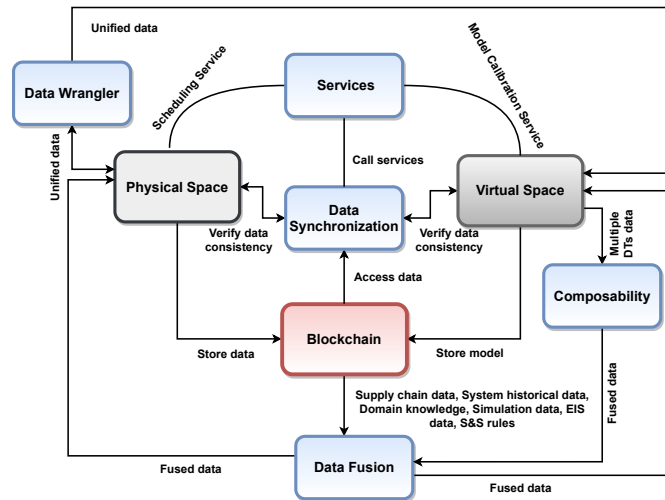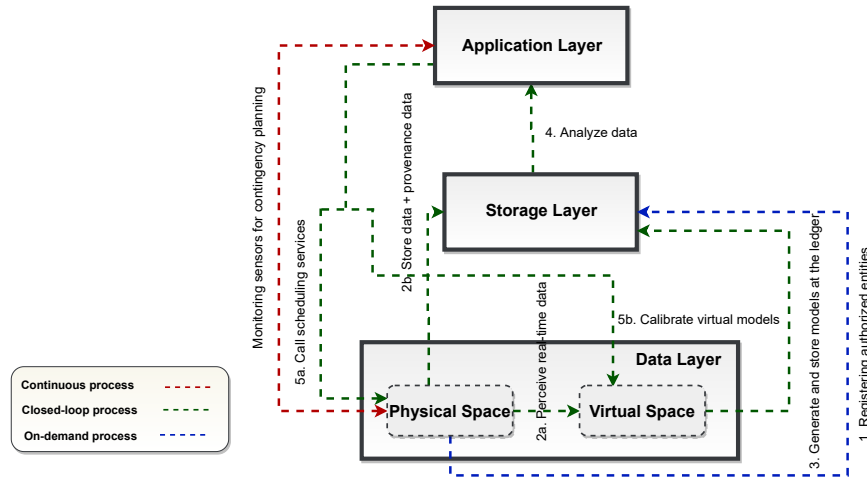

Fig. 7. A trustworthy 7D DT model.

Fig. 8. Blockchain-based digital twin: A framework. (Adapted from [64])

In the following, we discuss a seven-dimension (7D) DT framework (Fig. 7) based on the limitations of the existing works. We first provide a high-level layered structure of blockchain-based DTs (Fig. 8). Then we further discuss the main and supporting components and their functionalities in connection with DTs-, blockchain-, and security-related design issues outlined in Table 1.

3.4 Trustworthy Blockchain-based Digital Twins: A Unified Framework

Curating and mining information from the collected data calls for an infrastructure that can manifest the trustworthiness of data-generating sources, data in transit, and data at rest (storage) [64]. In this regard, Suhail et al. [64] recently addressed the requirements of trustworthy data management, data security, and predictive maintenance through a provenance-enabled blockchain-based DT framework. The main components of the proposed framework are presented in Fig. 8. At first, the participating entities such as sensors, machines, and humans register as authorized entities at the blockchain (step 1). Next, at the *data layer*, cyber-physical mapping is performed by monitoring, collecting, and processing designated parameters at the physical space that are mirrored at the virtual space (step 2a). The collected data along with provenance (i.e., a complete lineage of data along with a set of actions [80]) is sent to the *storage layer* (step 2b). At the twinned system, models are generated based on collected data, domain knowledge, and system historical data and are stored at the storage layer (step 3). The *application layer* continuously analyzes data to detect manufacturing disturbances (step 4) and calls for the scheduling services (step 5a) or model calibration services (step 5b), thereby completing the agile feedback loop. Additionally, physical security countermeasures can be exercised through monitoring devices such as surveillance cameras, power monitoring, fire alarms, etc. Thus, blockchain ensures secure data management and DTs use provenance-enabled trustworthy data as input to predict events and their associated causes by means of data analysis [64].

In the following, we discuss the 7D blockchain-based DT model which enhances the functionalities of the 5D DT model (shown in Fig. 2) by introducing data integration and interoperability components (such as data fusion, data wrangler, and composability) and data synchronization. Also, it introduces blockchain (as an alternative to DT Data) to facilitate secure data management to all components. Table 2 presents a summary of the blockchain-based DT model by mapping them on the desirable properties of DTs. Furthermore, we map the high-level

Table 2. Mapping DT properties with the proposed blockchain-based DTs model

Properties	Blockchain-based DT model
Clone counterpart	Virtual Space
Entanglement	Data Synchronization
Contextualization	Composability
Memorization	Blockchain
Availability	Blockchain
Data Integration and Interoperability	Data Fusion, Data Wrangler
Aggregation	Composability
Secure Data Management and Accountability	Blockchain, Services
Input (Historic) Knowledge Base	Blockchain

functionalities of the components in 7D blockchain-based DT model to the desirable characteristics outlined in Table 1. Finally, to illustrate the working of the proposed framework, we mapped the components of 7D DT model (Fig. 7) to the production process of autonomous vehicles as a CPS use case. The proposed framework can be significantly beneficial for the safe (or fail-safe) operation of autonomous vehicles where safety and reliability are highly deemed.

3.4.1 Design of Digital Twin. In addition to the existing components of the 5D DT model (Fig. 2), Suhail et al.'s framework [64] introduces the following three components: (i) data wrangler, (ii) data synchronization, and (iii) data fusion. *Data wrangler* is responsible for cleaning erroneous (invalid or missing) data and converting different data formats into a unified format before inputting data to the physical and virtual spaces as mentioned in subsection 4.1.2. For instance, during the vehicle production process, data from multiple physical sources such as sensors and actuators (attached to the robotic arms or vehicles structures) is gathered and is inputted to both the physical and virtual spaces. *Data fusion* is used to combine data from multiple sources to generate consistent interpretation for a certain object. For instance, before initiating the production process of vehicles, the necessary details such as order information (e.g., material stock, production quantity, estimated cost), supply chain data (e.g., consignment information), simulation data (equipment historical data, prediction of equipment fault), thresholds and consistency checks (S&S rules), etc. are already available in the storage system for usage. Furthermore, data wrangler and data fusion play a significant role in collecting and aggregating data after the vehicle production process. For example, to provide rich data about the car environment, the sensors attached to the connected vehicle (such as camera, GPS, infrared detectors, etc.) accumulate data during or after drive and then fed to the DTs for optimizing the operational control [67]. Such data can be used by threat intelligence to study the adversarial space and to update S&S rules for stealthy attacks. To facilitate the creation of an abstract view of overall phenomena, DTs must support the correlation of different DTs associated with different physical sub-components. For instance, different physical process during the vehicle production process along with their

twins must be combined. To do so, the framework uses *composability* that aggregates multiple DTs (i.e., DTA as discussed in Subsection 2.3.6) to integrate and interoperate multiple DTs or replicas.

To register any instantaneous change (based on requirements of application) in the system status, DTs and their physical counterparts must be "constantly connected". To do so, *data synchronization* enforces self-adaptation by performing a continuous digital-physical mapping between the pre-defined system performance parameters retrieved from the storage, and real-time sensor data obtained from the manufacturing unit, to verify data consistency. For instance, during manufacturing of vehicles, to minimize the makespan and production cost, DTs keep monitoring and analyzing the physical machine's current status based on constraints such as tool wear data, machine performance in the ongoing process, etc. To make DT reflecting the behavior and the status of the physical object, an effective and reliable communication suitable to the rate of changes are among other desirable properties to support entanglement between the physical space and the DT [47]. Depending on the data anomalies (defined in S&S rules) or uncertain data inconsistencies, the service component calls the corresponding scheduling services (in the physical space) or model calibration service (in the virtual space) to carry out the necessary measures. Based on the above-mentioned phases, the lifecycle of DT evolves from the early phase (design and development) to the medium phase (operation and maintenance). Whereas the existence of DT may occur before the creation of the product and even after the end of product life. The existence of DT may occur before the creation of the product and it may continue to exist even after the end of product life. Usually, the dismissal of the object is followed by the dismantling of the DT.

3.4.2 Design of Blockchain. The storage layer provides secure distributed data storage through a lightweight, scalable, and quantum-immune blockchain. Data can be securely accessed to and from blockchain, for example, the application layer uses it for analysis and decision-making while DT uses it for generating and uploading updated virtual models. Also, the data synchronization can access the updated data directly from the ledger, to eliminate the qualms of untrustworthy data. Empowering DTs with blockchain solves the key issues, i.e., disparate data repositories, and untrustworthy data dissemination; however, to reason about the current *source* and *state* of a data object entails a complete lineage of processes chain. Therefore, to diagnose data forgery during data acquisition and transition and to support "track and trace solutions", Suhail et al.'s framework relies on provenance data and twinned data to identify the cause of deviations between the physical-digital space in addition to data assessment.

Considering big data storage issues, we need to carefully consider what type of data needs to be stored on the blockchain. For instance, storing sensor data streams may not be an ideal situation due to data velocity and volume. The proposed framework relies on a data synchronization process that limits the frequent time-consuming access to the storage system, where the data flow of real-time sensor data and the less dynamic production and provenance data are explicitly separated.

3.4.3 Design of Security Features. Currently, most of the blockchain-based solutions heavily rely on de-facto cryptographic primitives. However, blockchains that are not quantum-resistant could pose long-tail data risk by potentially enabling attackers with high-powered quantum computers to monopolize the network by forging digital signatures and sabotage transactions [21]. Therefore, to prepare for the quantum world, blockchain schemes (e.g., [53]) that support post-quantum cryptography must be adopted [66]. Also, to allow multi-party use of DTs data, distributed data accessibility and auditability based on ownership, roles, and access levels must be enforced.

4 INTEGRATING BLOCKCHAIN-BASED DIGITAL TWINS WITH AI: WAY FORWARD

Although DTs offer a variety of advantages, they may pose limitations. For instance, *how to analyze the huge volume of data for creating actionable insights in real-time [67]?* AI can address these limitations through the

predictive capability of Machine Learning (ML) algorithms. For instance, blockchain and AI can provide intelligent and trusted DTs as shown in Fig. 9. However, integrating AI for blockchain-based DT requires consideration of several aspects. In this section, we provide a way forward for AI-driven blockchain-based DTs. We start with the rationale for AI in DTs and blockchain. Then we look into the role of AI in blockchain-based DTs in a holistic way and the outstanding challenges that need to be addressed.
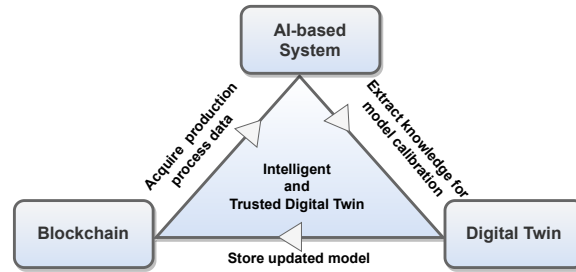

Fig. 9. Integrating AI and blockchain for intelligent and trusted DTs.

4.1 How AI Will Set a Course for Blockchain-based Digital Twins

In the following, we discuss some of the benefits brought by AI in blockchain-based DTs.

4.1.1 AI-based anomaly detection solutions for DTs. The complex, heterogeneous, and interconnected nature of smart CPS is prone to errors and attacks due to the integration of multiple un-trusted but interdependent participating entities and disjoint data from a number of sources. For example, sequential interdependence of processes (such as assembly line) or reciprocal interdependence (such as vehicles passing back and forth between multiple units during manufacturing). Due to the increasing interdependence of the underlying environment, predicting the effects of apparently innocent actions or events could become more challenging. Whereas ignoring such small variations in the system behavior may collapse the whole system or may incur drastic effect on the long-term behavior of the system. One such example within the ICS is Stuxnet where, without being detected, malicious code is able to intercept and modify the data sent to and from the PLC.

In this regard, implementing intelligence-driven solutions (such as data analytics and threat intelligence) provide autonomic security solutions to detect zero-day threats and can help to identify vulnerabilities, faulty models, threat actors, existing and potential attack vectors, thereby minimize the threat landscape [34]. Additionally, threat intelligence covers the broader perspective of *'who'*, *'why'*, *'when'*, *'how'*, and the *'magnitude of damage'* about the data breaches and attacks. In literature, ML has proven to be useful in detecting security threats by analyzing security and log data. For instance, Castellani et al. [6] used weak supervised learning in DTs for anomaly detection in the industry settings. The training data is generated using DTs whereas labeled anomalous data is generated from the physical components in the industry. More precisely, the authors used Siamese AutoEncoders, especially tailored for weak supervised learning with synthetic data to detect anomalies in the real data.

With reference to the proposed model represented in the Fig. 7, it is prudent to employ ML models in DTs data synchronization and model calibration services to ensure that stand-alone components have the ability to realize when it is being fed incorrect or malicious external data, detect the occurrence of an attack, and carry out a safe mission in the face of the attack [67]. Furthermore, the S&S rules must be stored and retrieved from the ledger to ensure their reliability. For instance, a provenance-aware blockchain-based system can track and trace the accountable entity for adding or updating the S&S rules [67]. Thus, based on the "learn and prevent" mechanism,

a blockchain-based DT governed by an intelligent AI-based system can detect the presence of malicious or mistaken disruptions in the system and can invoke the appropriate defence mechanisms automatically. Even if the damage is unavoidable, the AI-based solution, at the very least, isolates the attacked physical (sub) component, thereby keeping the rest of the infrastructure safe from further damage attack.

4.1.2 Improving Data Quality. In addition to predictions, AI helps in improving the data quality by implementing intelligent automated methods to fill in the missing data and to clean the data (such as data denoising, data de-duplication). For instance, AI-enabled data curation at the data wrangler method can help in achieving data completeness and data accuracy which ultimately improves the data analytics and optimization through the twinned system. To this end, Generative Adversarial Networks (GANs) can be used to improve the data quality by imputing missing data [79]. Similarly, data augmentation can be exploited to expand the amount of relevant data by adding minor alterations to the existing data. Through algorithms such as neural network, Kalman filter, and bayesian inference, data collected from multiple heterogeneous sources can be fused to generate consistent and comprehensive interpretations of the object [71]. While data quality can be improved through AI, data traceability can be enforced through blockchain-based solutions as discussed in Subsection 4.2.

4.1.3 AI-aided Predictive Maintenance in DTs. Creating future scenarios in the absence of direct observation data is very challenging for DTs. In this regard, to predict anomaly detection or to perform risk assessment, the predictive capability of ML algorithms can play an important role. For instance, supervised learning algorithms (such as LSTM), unsupervised learning algorithms (such as autoencoder), and deep learning (such as reinforcement learning) can be used for projections, predictions, or conducting sensitive analysis in a system through DTs. The appropriate algorithm can be chosen depending on the availability of labelled and unlabelled data in DTs. However, in such data-deprived situations, deep learning can be used. The blockchain can be used to record parameters and hyperparameters of ML and to log outcome in order to pinpoint trends in correct predictions.

4.2 Challenges for AI in Blockchain-based Digital Twins

Despite the afore-mentioned benefits of AI in blockchain-based DT, there are still challenges that need to be addressed before the realization of AI-driven blockchain-based DTs. In essence, we focus mainly on the data-related challenges faced by AI in blockchain-based DT. Identification of these challenges is important to avoid pitfalls of AI in such environments.

4.2.1 Need for Explainable AI. Despite the wide success of AI-based systems that are capable of perceiving, learning, predicting, and acting autonomously, there is a reluctance to adopt these systems in practice due to the black-box nature of most of the AI model. Decisions made by the resulting systems suffer from a loss of transparency and comprehensibility, particularly for the users who are not ML experts and hence can lead to a negative impact on the trustworthiness of the system [67]. Hesitation is even higher in safety-critical systems where explainability becomes crucial as slight dysfunction of an AI algorithm in DTs may lead to wrong decisions in the physical counterpart that could imply loss of life or economic disaster. Therefore, it is essential to remove the opaqueness of black-box-models by explaining their inner working. In this regard, explainable AI provides a set of techniques and methods to convert the black-box models to white-box approaches such that the overall process adopted by the algorithms and models to reach a decision can be explained and interpreted [31]. However, democratizing the data require decentralized AI while assuring traceability and solidity of data and algorithms. In this regard, consolidation of AI and blockchain supports transparency for AI operations by tracking data flow and complex behaviours of AI-based systems to establish more understanding and confidence in the decisions made by the underlying systems [15]. The application of AI with federated learning algorithms is one such example that requires data traceability favored by blockchain to lurk privacy issues.

4.2.2 Data Scarcity. Data serve as the lifeblood of AI models. To be able to learn the behavior and make accurate predictions, a sufficient (large) amount of data is necessary and oftentimes becomes the bottleneck for AI. Abstractly, there are two ways to generate data: (i) acquire data from the real environmental components (for instance sensors and actuators), and (ii) generate data synthetically (for instance, using GANs). Another reason for data scarcity might be siloed data by tech-giants or reluctance to share data due to competition. Since DTs consumes real-time and historical data to simulate or replicate the real-world environment, the use of AI in DT should take data scarcity into account. Uniform, consistent, healthy, and on-time data sources must be present to be fed to the AI models. One possible solution to solve data scarcity problem is to incentivize users for crowdsourcing and crowdselling data, for instance, through blockchain-based data marketplace, streamr [1]. The industry stakeholders may utilize the crowdsourced data during early phase (design and development) of DTs lifecycle. Similarly, the crowdsourced data may provide hypothetical conditions for simulation mode particularly when the current physical state of the system is not known. Federated learning can be used to solve the data scarcity problems by means of training local models at the user's devices with the user's private data.

4.2.3 Model Errors and Bias Problem. Model errors are closely coupled with the bias problem and are inherently related to the data inputted to AI. If the AI model is trained on the imbalanced or poor quality data, then the outcome of the learning can be biased thereby causing prediction errors and compromising the efficacy of AI. Since the data from both the physical and its twinned environment is used for learning, it is important to consider the source of data, the environment in which it is collected, the context of data collection, and the way the data was collected to avoid biasness in data and eventually model errors. To enforce data quality, employing blockchain can solve the bias in data, and model error problems in two ways: (i) incentivizing the participants to contribute good data [36], and (ii) gathering data from multiple siloed sources through blockchain-based federated learning [74]. It is important to note that incentivisation of contributing data will depend on the specific blockchain type used and the built in incentives to contribute data. For data bias through sharing data from multiple sources, this in itself does not solve the problem. Data fusion methods that incorporate trust are needed on top of blockchain to overcome the data bias problem. Additionally, collective decision-making processes can be carried out through decentralized infrastructures and consensus protocols. Thus, human users can re-calibrate their trust in the system by tracing back the decision process and making justifications that are essential to identify which entity (human or machine) is faulty.

5 RESEARCH AND DEPLOYMENT CHALLENGES

In the quest to enforce secure data management in the Industry 4.0 era, the following technical, logistical, and social challenges of blockchain-based DTs schemes require further investigation. In this section, we highlight the open-ended challenges that hinder the successful implementation of blockchain-supported DTs in the industry. These challenges open new avenues for future research in these directions. Fig. 10 presents the detailed taxonomy of the current and future research and deployment challenges for blockchain-driven DTs. We summarize the challenges along with causes and possible solutions in Table 3.

5.1 Technical Issues

Here we discuss technical challenges with reference to blockchains, DTs, and their integration.

5.1.1 Representation of Digital Twins. "High-fidelity" model is one of the critical challenges for a virtual model that represents the real-world situation. For instance, in a production unit, the quintessential DT for a robotic arm should mirror an exact image of all the behavior and functionalities of the physical component, synchronized in real-time throughout its lifecycle [64]. However, in industry, uncertain scenarios transpire due to the dynamism

---

[1]https://streamr.network/

Challenges for Blockchain-based Digital Twins

Technical Challenges
- Representing High-fidelity Models
- Data-related Issues
- Model Challenges
- Uncertain Events
- Interoperability

Logistical Challenges
- Standardization Barriers
- Scattered Ownership of Data
- Infrastructure Cost
- Skepticism in Blockchain

Social Challenges
- Ethical and Moral Issues
- Environmental Issues
- Lack of Collaboration
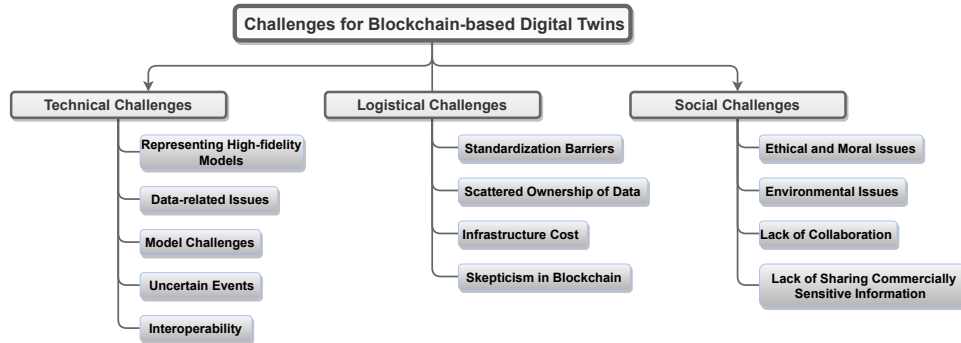- Lack of Sharing Commercially Sensitive Information

Fig. 10. Current and future challenges for blockchain-based digital twins in the industrial domain.

and complexity of underlying (sub) systems, machinery malfunctioning, human errors, hidden vulnerabilities, erroneous data, knowledge gaps, and variability, fuzziness, and uncertainty of physical asset. Moreover, while maintaining backward compatibility, a corresponding evolution of the models is required as the physical object evolves over time [60]. Therefore, to enable DTs to unearth the unknown unknowns, a pragmatic approach must be adopted by integrating cognitive DT solutions [35].

Considering representativeness and contextualization in DTs, feature simplifications or generalizations are possible, provided that they do not affect the behavior in the specific context under a specific environment [47]. In this regard, replication modes are used for tracing physical objects and reproducing activities by emulating their behavior [12]. Initiating replicas in the framework or in the application space poses issues. For example, depending on the time period defined by the application, all replicas of a physical object should have the same status as that of the physical object. Otherwise, it may lead to the problem of synchronization and consistency of information between the physical object and its different replicas. To solve such issues, one possible solution is to integrate transaction management functions through blockchain. Other issues that may arise with replication are temporal dependencies and the need for input knowledge (such as events and activities) in advance to produce the same stimuli. Also, it requires experts to solve complex and intertwined problems.

To emphasize socio-technical design implications, Barricelli et al. [4] focused on Human Work Interaction Design (HWID) and End-User Development (EUD) to bridge the communication gaps among domain experts, designers, and stakeholders. In this regard, DTs must be equipped with well-designed and accessible interfaces such that it appears indistinguishable from the physical asset and allows end-user (domain experts or non-informatic experts) to intuitively create, modify, test, and operate digital artifacts.

5.1.2 Data-related Challenges. Data challenges mainly emerge due to the following four factors: i) types of data to collect, ii) rate of data collection, iii) data storage, and iv) merging of disparate data types. The first issue requires a trade-off between excessive data volumes and limited data volumes, i.e., too few data may result in inaccurate predictions, and too many data may result in mired down in details or may pose decision paralysis due to information overload [64]. The second issue involves the frequency at which data samples are collected. For instance, recording vibrations from a turbine gearbox every minute would miss shorter glitches but sampling every second could result in too much data thereby causing transmission bottlenecks [69]. Furthermore, the duration for which data is stored needs to be determined as high volumes of rapid measurements lead to high storage cost but at the same time, long-term data are required for modeling [38]. Therefore, depending on the system requirements, it is important to define the spatio-temporal granularity of the data. The third issue corresponds to storing *big data* on the blockchain that requires careful consideration of factors such as high cost

associated with high volumes of rapid measurements, block size, latency, etc. Alternative storage mechanisms and solutions include off-chain storage, storing only data hash to reduce storage costs, or removable blockchain designs [18]. Despite these solutions, the big data storage problem considering the volume and velocity of data are still open issues. The fourth issue involves disparate data types that are hard to merge, for instance, vibrations, temperatures, videos, or images all have different data types, formats, structure, and scale. Under such scenarios, missing or erroneous data can distort results and obscure faults, timings can get out of step during data sampling at different rates, and averaging fine data may lead to loss of details [69]. In a nutshell, to embed viable and size-efficient data into the ecosystem, the data-related challenges including volume, variety, veracity, generation rate, spatio-temporal resolution of sensor data, and efficient processing, archival, and retrieval require further investigation.

5.2 Logistical Issues

In the following, we outline some logistical challenges faced by the adoption of blockchain-based DTs in the industrial domain.

5.2.1 Expenditures on Infrastructure. The adoption of emerging technological trends, such as IoT, High-performance computing (HPC), blockchain, DTs, AI, etc. aids in reducing operational costs, reinforce productivity, provide predictive and preventive maintenance of industrial machines, and attain higher operational efficiency. However, owing to such innovations, companies have to spend a considerable amount of capital on the day-to-day management of networked devices (maybe geographically dispersed). Such practices require the balancing of the cost-benefit analysis. Otherwise, it could drain operational resources for managing industry assets and business processes. Despite the significance of DTs in reducing risks, increasing safety, and improving efficiency, many enterprises realize that the connectivity, computing, bandwidth, and storage required to process massive volumes of data involved in creating DTs are cost-prohibitive [64]. Likewise, for blockchain technology, perceived risks associated with the immature technology, high costs of initial implementation, and the likelihood of disrupting existing practices may pose significant challenges to businesses. Another addition to the Capital EXpenditures (CapEXs) and Operational EXpenditures (OpEXs) involves the issue to train security specialists and institutions with the state-of-the-art applied tools and knowledge of the data used to establish the model. Otherwise, improper or missing influencing factors can lead to fallacious results on which the critical security decisions and actions are based [12]. In such situations, shifting to the new paradigm is more beneficial for medium-to-large scale industries in contrast to small-scale industries. However, sooner or later, small-scale industries have to adopt the current technologies as they may scale in the future or ultimately due to technological obsolescence.

5.2.2 Standardization Barriers and Efforts. Due to the lack of applicable laws, the absence of unifying data and model standards, and integration problems into business processes, an internationally operational framework and infrastructure for DTs is still a long way away. Typically incompatibilities exist throughout the entire development process and life cycle. For instance, scattered ownership of data involves different data interfaces, proprietary file formats, object definitions, encoding, and sharing of knowledge strewn throughout the product lifecycle. Moreover, it includes a multitude of vendors, contractors, and customers of all sizes and specializations that do not necessarily use the same software/tools, etc. Therefore, the impetus to create DT standard requires improved interoperability where each step of the workflow ingests artifacts from the previous stage and output artifacts that can be consumed by the following phase [52]. With firm guidelines for cross-disciplinary standardization, a universal non-proprietary design and development platform for DTs should be developed with user-friendly and API-based interfaces on which all models can run without data monetization or fixed vendor partnerships.

The common denominator among blockchain platforms is their reliance on a distributed and decentralized peer-to-peer network. They nevertheless vary substantially in terms of the underlying data structure, scalability solutions, fault tolerance, degree of decentralization, and consensus approaches. The disadvantages due to non-interoperable blockchain implementations are manifold, for example, information restrictions, highly fragmented industrial ecosystems, stand-alone and disconnected networks, to name a few. In essence, allowing seamless integration of data by enforcing open data structure and processing harmonization through common linked data language and multiple agile, scalable and mergeable blockchain ledgers for logging different data types can minimize obstacles in the widespread adoption of blockchain. The Enterprise Ethereum Alliance (EEA), the International Organization for Standardization (ISO), the standardization sector of the International Telecommunications Union (ITU-T), the World Wide Web Consortium (W3C), and the Society for Worldwide Interbank Financial Telecommunication (SWIFT) carried out notable efforts in this regard.

5.2.3 Skepticism in Blockchain Technology. Many current research works are raising concerns about the practical adoption of blockchain technology in the industry. Some of their concerns include how to ensure the reliability of data from the participating entities and/or sensors [77] considering the connection between the physical and digital world, latency issues with the increasing number of nodes in the network [29, 54], lack of information leading to the existence of gray markets [3], security concerns due to quantum computing [21], lack of privacy and GIGO problem [2], decision paralysis due to information overload, high energy consumption [82], throughput and latency issues [43], lack of standardization and shifting to new infrastructure from legacy systems [62], etc. [65]. Paradoxically, blockchain technology supports the full gamut of industrial applications while struggling to address such challenges.

Additionally, problems stemming from the blockchain "black-box" nature force consumers to trust the integrity and fairness of the processes without understanding the technical underpinnings. Due to this reason, companies are hesitant to make use blockchain technology in DTs. Innovation hubs and maker spaces organize webinars, workshops, tutorials, and certifications that can enable the teams to learn blockchain technology and explore its use-cases in their respective areas, for example, learning platform provided by IBM [33].

5.3 Social Issues

There are several risks involved with the proliferation of the blockchain and DTs on an individual level which eventually summed up as issues in their integration. In the following, we discuss the social challenges faced by the integration of blockchain-based DTs.

5.3.1 Ethical and Legal Issues. Due to industry rivalry, companies often do not want to share commercially sensitive information with their competitors. Moreover, commercial pressures dissuade companies from sharing models, smaller firms lose out, resulting in monopolies or oligopolies. For instance, giant companies such as Siemens, GE, IBM, etc. own most of the DTs. Additionally, on one hand, due to a lack of common space between academia and industry, data scientists, cybersecurity experts, engineering, and business strategists cannot communicate and share knowledge and software. On the other hand, researchers and practitioners patent or even hoard knowledge, ensuing limited access to the data. Such problems call for a platform managed by government funding agencies, consortia of private corporations, or by a coalition of universities to reconcile issues about data governance, data ownership, and openness [69]. In this regard, a close-knit team of specialists spanning disciplines should collaborate, build precise DTs, and share data and models on a public database for widespread usage. Currently, such examples include openVertebrate (share data and models of vertebrate anatomies), Siemens NX software (package design, simulation, and manufacturing tools), LlamaZOO (develop interactive 3D visualization, digital twinning, and training solutions for Industry 4.0 through VR/AR), and IBM Model-Based Systems Engineering (streamline product design and development). Some of the open challenges

such as how to resolve the ownership rights in scenarios related to DTs of retired or former human workers, or regulations for using healthcare data with consent and anonymously, still need to be addressed. Also, these challenges motivate the design of blockchains for DTs that protect sensitive information while providing sufficient transparency.

5.3.2 Environmental Impact. Inefficient by design, blockchain consensus protocols are considered the Achilles' heel in terms of energy constraints. The high cost incurred on the consensus phase is due to massive electricity consumption in performing compute-intensive mining phase (running cryptographic algorithms); draws a worrying figure for energy footprints. For example, with an estimated energy expenditure of almost 10 TWh, Ethereum matches the total energy consumption of some nations [14]. To curtail such an impact on the environment, possible solutions involve switching from energy-hungry to less power-consuming consensus mechanisms, such as TreeChain [17] or deploying blockchains featuring miner-free solutions, such as IOTA [53, 63]. Moreover, the AI-supported energy planning of blockchain-based systems could be optimized in how to allocate and maximize resources.

Table 3. Current and future research and deployment challenges in blockchain-based DTs.

Class	Key challenges	Possible solutions
T1: Representing DTs	• Accurate representation of physical object • Contextualization in DTs through replications • Design implications	• Integrate cognitive DT solutions • Integrate transaction management functions • Focus on HWID and EUD
T2: Data-related issues	• Types of data to collect • Rate of data collection • Data storage • Merging disparate data types	• Consider data granularity • Off-chain storage, removable blockchain designs
L1: Expenditure on infrastructure	• High cost on managing infrastructures • Training personnel cost	• Plan expenditure on software and hardware costs
L2: Standardization barriers	• Absence of unifying data and models • Scattered ownership of data • Non-interoperable blockchains	• Developing universal platform for DTs • Integration of data among different blockchains
L3: Skepticism in blockchain technology	• Black-box	• Supporting learning platforms

S1: Ethical and moral issues	• Lack of commercially sensitive information • Companies owning DTs may monopolize global market • Limited access to data and models	• Collaborating and encouraging widespread knowledge of data and models
S2: Environmental issues	• Energy consumption due to compute-intensive mining phase	• Use of efficient consensus mechanism
T: *Technical challenges*	**L**: *Logistical challenges*	**S**: *Social challenges*

6 DISCUSSION AND CONCLUSION

Due to the involvement of multiple entangled entities and ongoing intricate industrial processes, industries have to consider several challenging factors for maximizing their revenues, such as cost management, risk prognosis, and improving efficiency. Therefore, to maximize industrial efficiency, one of the promising solutions is to create a digital twin that mirrors every facet of the underlying system or process in order to analyze, predict, and optimize all operations before its real-world implementation. Driven by data collected and accessed from heterogeneous data sources and data silos owned by multiple participating entities calls for a secure distributed infrastructure. In this regard, leveraging blockchain allows industries to manage data on a distributed ledger that securely records product lifecycle data management events. Thus, DTs improve system maintenance by focusing more on preventing problems rather than on solving problems while blockchain provides the creation of secure DT frameworks.

In this survey, we have conducted a detailed and comprehensive review of the design and implementation issues of the current state-of-the-art blockchain-based DTs solutions in the industry. The key takeaways from this survey are as follows:

- The integrating blockchain in DTs allows product lifecycle stakeholders to manage data on a shared distributed ledger to solve data trust, integrity, and security challenges.
- The comparative analysis of existing blockchain-based DTs identified the research gaps in the design solutions related to DTs, blockchain, and security. Based on the comparison shortcomings in the existing literature, a trustworthy 7D blockchain-based DT framework (Fig. 7) has been proposed.
- The integration of AI in blockchain-based DTs provide intelligent twins to support predictive maintenance and threat intelligence.
- Finally, the research community needs to investigate the research and deployment challenges mentioned in Table 3 for the realization of the successful implementation of blockchain-based DTs.

We conjecture that this survey provides close insights to researchers to overcome the challenges and pave the way for the standardization of blockchain-based DTs in industrial applications.

ACKNOWLEDGMENTS

REFERENCES

[1] Atin Angrish, Benjamin Craver, Mahmud Hasan, and Binil Starly. 2018. A case study for Blockchain in manufacturing:"FabRec": A prototype for peer-to-peer network of manufacturing nodes. *Procedia Manufacturing* 26 (2018), 1180–1192. https://doi.org/10.1016/j.promfg.2018.07.154

[2] Volodymyr Babich and Gilles Hilary. 2020. OM Forum - Distributed Ledgers and Operations: What Operations Management Researchers Should Know About Blockchain Technology. *Manufacturing & Service Operations Management* 22, 2 (2020), 223–240. https://doi.org/10.1287/msom.2018.0752

[3] Volodymyr Babich, Gilles Hilary, et al. 2019. Blockchain and other distributed ledger technologies in operations. *Foundations and Trends® in Technology, Information and Operations Management* 12, 2-3 (2019), 152–172. https://doi.org/10.2139/ssrn.3232977

[4] Barbara Rita Barricelli, Elena Casiraghi, and Daniela Fogli. 2019. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access* 7 (2019), 167653–167671. https://doi.org/10.1109/ACCESS.2019.2953499

[5] Umesh Bodkhe, Dhyey Mehta, Sudeep Tanwar, Pronaya Bhattacharya, Pradeep Kumar Singh, and Wei-Chiang Hong. 2020. A Survey on Decentralized Consensus Mechanisms for Cyber Physical Systems. *IEEE Access* 8 (2020), 54371–54401. https://doi.org/10.1109/ACCESS.2020.2981415

[6] Andrea Castellani, Sebastian Schmitt, and Stefano Squartini. 2020. Real-World Anomaly Detection by using Digital Twin Systems and Weakly-Supervised Learning. *IEEE Transactions on Industrial Informatics* (2020), 1–1. https://doi.org/10.1109/tii.2020.3019788

[7] L. Chen. 2017. Cryptography Standards in Quantum Time: New Wine in an Old Wineskin? *IEEE Security & Privacy* 15, 04 (jul 2017), 51–57. https://doi.org/10.1109/MSP.2017.3151339

[8] Chi Cheng, Rongxing Lu, Albrecht Petzoldt, and Tsuyoshi Takagi. 2017. Securing the Internet of Things in a quantum world. *IEEE Communications Magazine* 55, 2 (2017), 116–120. https://doi.org/10.1109/MCOM.2017.1600522CM

[9] Anton Churyumov. 2016. Byteball: A decentralized system for storage and transfer of value. Retrieved December 20, 2020 from https://obyte.org/Byteball.pdf Byteball.org, Moscow, Russia.

[10] Volkan Dedeoglu, Raja Jurdak, Guntur D. Putra, Ali Dorri, and Salil S. Kanhere. 2019. A Trust Architecture for Blockchain in IoT. In *Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (Houston, Texas, USA) *(MobiQuitous '19)*. Association for Computing Machinery, New York, NY, USA, 190–199. https://doi.org/10.1145/3360774.3360822

[11] Deloitte. 2019. Continuous interconnected supply chain. https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/technology/lublockchain-internet-things-supply-chain-traceability.pdf

[12] Marietheres Dietz and Günther Pernul. 2020. Unleashing the Digital Twin's Potential for ICS Security. *IEEE Security & Privacy* 18, 4 (2020), 20–27. https://doi.org/10.1109/MSEC.2019.2961650

[13] Marietheres Dietz, Benedikt Putz, and Günther Pernul. 2019. A Distributed Ledger approach to Digital Twin secure data sharing. In *IFIP Annual Conference on Data and Applications Security and Privacy*, Vol. 11559. Springer, Cham, 281–300. https://doi.org/10.1007/978-3-030-22479-0_15

[14] Digiconomist. 2020. Ethereum Energy Consumption Index (beta). Retrieved November 10, 2020 from https://digiconomist.net/ethereum-energy-consumption

[15] T. N. Dinh and M. T. Thai. 2018. AI and Blockchain: A Disruptive Integration. *Computer* 51, 9 (2018), 48–53. https://doi.org/10.1109/MC.2018.3620971

[16] A. Dorri and R. Jurdak. 2020. Tree-Chain: A Fast Lightweight Consensus Algorithm for IoT Applications. In *2020 IEEE 45th Conference on Local Computer Networks (LCN)*. IEEE Computer Society, Los Alamitos, CA, USA, 369–372. https://doi.org/10.1109/LCN48667.2020.9314831

[17] A. Dorri and R. Jurdak. 2020. Tree-Chain: A Fast Lightweight Consensus Algorithm for IoT Applications. In *2020 IEEE 45th Conference on Local Computer Networks (LCN)*. IEEE, Sydney, NSW, Australia, 369–372. https://doi.org/10.1109/LCN48667.2020.9314831

[18] Ali Dorri, Salil S Kanhere, and Raja Jurdak. 2019. MOF-BC: A memory optimized and flexible blockchain for large scale networks. *Future Generation Computer Systems* 92 (2019), 357–373. https://doi.org/10.1016/j.future.2018.10.002

[19] Ali Dorri, Salil S Kanhere, Raja Jurdak, and Praveen Gauravaram. 2019. LSB: A Lightweight Scalable Blockchain for IoT security and anonymity. *J. Parallel and Distrib. Comput.* 134 (2019), 180–197. https://doi.org/10.1016/j.jpdc.2019.08.005

[20] Matthias Eckhart and Andreas Ekelhart. 2018. Towards Security-Aware Virtual Environments for Digital Twins. In *Proceedings of the 4th ACM Workshop on Cyber-Physical System Security* (Incheon, Republic of Korea) *(CPSS '18)*. Association for Computing Machinery, New York, NY, USA, 61–72. https://doi.org/10.1145/3198458.3198464

[21] Aleksey K Fedorov, Evgeniy O Kiktenko, and Alexander I Lvovsky. 2018. Quantum computers put blockchain security at risk. *Nature* 563 (2018), 465–467. https://doi.org/10.1038/d41586-018-07449-z

[22] R. Gao, L. Wang, R. Teti, D. Dornfeld, S. Kumara, M. Mori, and M. Helu. 2015. Cloud-enabled prognosis for manufacturing. *CIRP Annals* 64, 2 (2015), 749–772. https://doi.org/10.1016/j.cirp.2015.05.011

[23] Gartner. 2019. Gartner Survey Reveals Digital Twins Are Entering Mainstream Use. Retrieved November 10, 2020 from https://www.gartner.com/en/newsroom/press-releases/2019-02-20-gartner-survey-reveals-digital-twins-are-entering-mai

[24] Edward Glaessgen and David Stargel. 2012. The digital twin paradigm for future NASA and US Air Force vehicles. In *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA.* Honolulu, Hawaii, 1818. https://doi.org/10.2514/6.2012-1818

[25] Christopher Santi Götz, Patrik Karlsson, and Ibrahim Yitmen. 2020. Exploring applicability, interoperability and integrability of Blockchain-based digital twins for asset life cycle management. *Smart and Sustainable Built Environment* (2020). https://doi.org/10.1108/SASBE-08-2020-0115

[26] Michael Grieves. 2014. Digital twin: manufacturing excellence through virtual factory replication. *White paper* 1 (2014), 1–7. https://doi.org/10.5281/zenodo.1493930

[27] Haya R Hasan, Khaled Salah, Raja Jayaraman, Mohammed Omar, Ibrar Yaqoob, Saša Pesic, Todd Taylor, and Dragan Boscovic. 2020. A Blockchain-Based Approach for the Creation of Digital Twins. *IEEE Access* 8 (2020), 34113–34126. https://doi.org/10.1109/ACCESS.2020.2974810

[28] R. Henry, A. Herzberg, and A. Kate. 2018. Blockchain Access Privacy: Challenges and Directions. *IEEE Security & Privacy* 16, 4 (2018), 38–45. https://doi.org/10.1109/MSP.2018.3111245

[29] Matt Higginson, Marie-Claude Nadeau, and Kausik Rajgopal. 2019. Blockchain's Occam problem. Retrieved December 29, 2019 from https://www.mckinsey.com/industries/financial-services/our-insights/blockchains-occam-problem#

[30] Sihan Huang, Guoxin Wang, Yan Yan, and Xiongbing Fang. 2020. Blockchain-based data management for digital twin of product. *Journal of Manufacturing Systems* 54 (2020), 361–371.

[31] F Hussain, R Hussain, and E Hossain. 2021. Explainable Artificial Intelligence (XAI): An Engineering Perspective. arXiv:2101.03613 [cs.CV]

[32] HyperLedger Sawtooth 2020. HyperLedger Sawtooth. https://www.hyperledger.org/use/sawtooth

[33] IBM. 2018. IBM Blockchain Training. https://www.ibm.com/training/blockchain

[34] Amani Ibrahim, Dhananjay Thiruvady, Jean-Guy Schneider, and Mohamed Abdelrazek. 2020. The Challenges of Leveraging Threat Intelligence to Stop Data Breaches. *Frontiers in Computer Science* 2 (2020), 36. https://doi.org/10.3389/fcomp.2020.00036

[35] Muhammad Intizar Ali, Pankesh Patel, John G. Breslin, Ramy Harik, and Amit Sheth. 2021. Cognitive Digital Twins for Smart Manufacturing. *IEEE Intelligent Systems* 36, 2 (2021), 96–100. https://doi.org/10.1109/MIS.2021.3062437

[36] D Justin and BW Harris. 2019. Decentralized & Collaborative AI on Blockchain. In *Proceedings of the 2019 IEEE International Conference on Blockchain (Blockchain), Atlanta, GA, USA.* 14–17.

[37] A. Khan, F. Shahid, C. Maple, A. Ahmad, and G. Jeon. 2020. Towards Smart Manufacturing Using Spiral Digital Twin Framework and Twinchain. *IEEE Transactions on Industrial Informatics* (2020), 1–1. https://doi.org/10.1109/TII.2020.3047840

[38] Andrew Kusiak. 2017. Smart manufacturing must embrace big data. *Nature* 544, 7648 (2017), 23–25. https://doi.org/10.1038/544023a

[39] Ralph Langner. 2013. To kill a centrifuge: A technical analysis of what stuxnet's creators tried to achieve. Retrieved January 20, 2020 from https://www.langner.com/wp-content/uploads/2017/03/to-kill-a-centrifuge.pdf

[40] Colin LeMahieu. 2018. Nano: A feeless distributed cryptocurrency network. https://nano.org/en/whitepaper

[41] Jiewu Leng, Guolei Ruan, Pingyu Jiang, Kailin Xu, Qiang Liu, Xueliang Zhou, and Chao Liu. 2020. Blockchain-empowered sustainable manufacturing and product lifecycle management in industry 4.0: A survey. *Renewable and Sustainable Energy Reviews* 132 (2020), 110112. https://doi.org/10.1016/j.rser.2020.110112

[42] Jiewu Leng, Douxi Yan, Qiang Liu, Kailin Xu, J Leon Zhao, Rui Shi, Lijun Wei, Ding Zhang, and Xin Chen. 2019. ManuChain: Combining permissioned blockchain with a holistic optimization model as bi-level intelligence for smart manufacturing. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 50, 1 (2019), 182–192.

[43] Mario Lezoche, Jorge E Hernandez, Maria del Mar Eva Alemany Díaz, Hervé Panetto, and Janusz Kacprzyk. 2020. Agri-food 4.0: a survey of the supply chains and technologies for the future agriculture. *Computers in Industry* 117 (2020), 103187. https://doi.org/10.1016/j.compind.2020.103187

[44] Claudio Mandolla, Antonio Messeni Petruzzelli, Gianluca Percoco, and Andrea Urbinati. 2019. Building a digital twin for additive manufacturing through the exploitation of blockchain: A case analysis of the aircraft industry. *Computers in industry* 109 (2019), 134–152. https://doi.org/10.1016/j.compind.2019.04.011

[45] D. Miller. 2018. Blockchain and the Internet of Things in the Industrial Sector. *IT Professional* 20, 3 (2018), 15–18. https://doi.org/10.1109/MITP.2018.032501742

[46] Steve Miller, Nathan Brubaker, D Kappelmann Zafra, and Dan Caban. 2019. Triton actor TTP profile, custom attack tools, detections, and ATT&CK mapping. Retrieved January 20, 2020 from https://www.fireeye.com/blog/threat-research/2019/04/triton-actor-ttp-profile-custom-attack-tools-detections.html

[47] R. Minerva and N. Crespi. 2021. Digital Twins: Properties, Software Frameworks, and Application Scenarios. *IT Professional* 23, 1 (2021), 51–55. https://doi.org/10.1109/MITP.2020.2982896

[48] M. Mosca. 2018. Cybersecurity in an Era with Quantum Computers: Will We Be Ready? *IEEE Security & Privacy* 16, 5 (2018), 38–41. https://doi.org/10.1109/MSP.2018.3761723

[49] J. Mulholland, M. Mosca, and J. Braun. 2017. The Day the Cryptography Dies. *IEEE Security & Privacy* 15, 4 (2017), 14–21. https://doi.org/10.1109/MSP.2017.3151325

[50] Satoshi Nakamoto. 2019. *Bitcoin: A peer-to-peer electronic cash system*. Technical Report. Manubot.

[51] N. Neshenko, E. Bou-Harb, J. Crichigno, G. Kaddoum, and N. Ghani. 2019. Demystifying IoT Security: An Exhaustive Survey on IoT Vulnerabilities and a First Empirical Look on Internet-Scale IoT Exploitations. *IEEE Communications Surveys Tutorials* 21, 3 (2019), 2702–2733. https://doi.org/10.1109/COMST.2019.2910750

[52] V. Piroumian. 2021. Digital Twins: Universal Interoperability for the Digital Age. *Computer* 54, 1 (2021), 61–69. https://doi.org/10.1109/MC.2020.3032148

[53] S Popov. 2017. The tangle. White paper. https://iota.org/IOTA_Whitepaper.pdf

[54] Kyleen W. Prewett, Gregory L. Prescott, and Kirk Phillips. 2020. Blockchain adoption is inevitable—Barriers and risks remain. *Journal of Corporate Accounting & Finance* 31, 2 (apr 2020), 21–28. https://doi.org/10.1002/jcaf.22415

[55] Deepak Puthal, Nisha Malik, Saraju P Mohanty, Elias Kougianos, and Chi Yang. 2018. The blockchain as a decentralized security framework [future directions]. *IEEE Consumer Electronics Magazine* 7, 2 (2018), 18–21. https://doi.org/10.1109/MCE.2017.2776459

[56] Benedikt Putz, Marietheres Dietz, Philip Empl, and Günther Pernul. 2021. EtherTwin: Blockchain-based Secure Digital Twin Information Management. *Information Processing & Management* 58, 1 (2021), 102425. https://doi.org/10.1016/j.ipm.2020.102425

[57] QRL 2019. The Quantum Resistant Ledger. https://github.com/theQRL/Whitepaper/blob/master/QRL_whitepaper.pdf

[58] Pethuru Raj. 2021. Chapter Thirteen - Empowering digital twins with blockchain. In *The Blockchain Technology for Secure and Smart Applications across Industry Verticals*. Advances in Computers, Vol. 121. Elsevier, 267 – 283. https://doi.org/10.1016/bs.adcom.2020.08.013

[59] Aravind Ramachandran, Dr Kantarcioglu, et al. 2017. Using blockchain and smart contracts for secure data provenance management. arXiv:1709.10000

[60] Adil Rasheed, Omer San, and Trond Kvamsdal. 2020. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access* 8 (2020), 21980–22012. https://doi.org/10.1109/ACCESS.2020.2970143

[61] José Ríos, Juan Carlos Hernandez, Manuel Oliva, and Fernando Mas. 2015. Product Avatar as Digital Counterpart of a Physical Individual Product: Literature Review and Implications in an Aircraft. In *ISPE CE*, Vol. 2. IOS Press, 657–666. https://doi.org/10.3233/978-1-61499-544-9-657

[62] Henrik S. Sternberg, Erik Hofmann, and Dominik Roeck. 2020. The Struggle is Real: Insights from a Supply Chain Blockchain Case. *Journal of Business Logistics* (mar 2020). https://doi.org/10.1111/jbl.12240

[63] D. Strugar, R. Hussain, M. Mazzara, V. Rivera, J. Young Lee, and R. Mustafin. 2018. On M2M Micropayments: A Case Study of Electric Autonomous Vehicles. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, Halifax, NS, Canada, 1697–1700. https://doi.org/10.1109/Cybermatics_2018.2018.00283

[64] Sabah Suhail, Rasheed Hussain, Raja Jurdak, and Choong Seon Hong. 2021. Trustworthy Digital Twins in the Industrial Internet of Things with Blockchain. *IEEE Internet Computing* (2021), 1–10. https://doi.org/10.1109/mic.2021.3059320

[65] Sabah Suhail, Rasheed Hussain, Abid Khan, and Choong Seon Hong. 2020. Orchestrating Product Provenance Story: When IOTA ECOSYSTEM Meets The Electronics Supply Chain Space. *Computers in Industry* 123 (2020), 103334. https://doi.org/10.1016/j.compind.2020.103334

[66] S. Suhail, R. Hussain, A. Khan, and C. S. Hong. 2021. On the Role of Hash-Based Signatures in Quantum-Safe Internet of Things: Current Solutions and Future Directions. *IEEE Internet of Things Journal* 8, 1 (2021), 1–17. https://doi.org/10.1109/JIOT.2020.3013019

[67] Sabah Suhail, Raja Jurdak, Raimundas Matulevicius, and Choong Seon Hong. 2021. Securing Cyber-Physical Systems Through Blockchain-Based Digital Twins and Threat Intelligence. *CoRR* abs/2105.08886 (2021). arXiv:2105.08886 https://arxiv.org/abs/2105.08886

[68] Fei Tao, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, and Fangyuan Sui. 2018. Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology* 94, 9-12 (2018), 3563–3576. https://doi.org/10.1007/s00170-017-0233-1

[69] Fei Tao and Qinglin Qi. 2019. Make more digital twins. *Nature* 573 (2019), 490–491. https://doi.org/10.1038/d41586-019-02849-1

[70] Fei Tao, He Zhang, Ang Liu, and Andrew YC Nee. 2018. Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics* 15, 4 (2018), 2405–2415. https://doi.org/10.1109/TII.2018.2873186

[71] Fei Tao and Meng Zhang. 2017. Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. *IEEE Access* 5 (2017), 20418–20427. https://doi.org/10.1109/ACCESS.2017.2756069

[72] Fei Tao, Meng Zhang, Yushan Liu, and AYC Nee. 2018. Digital twin driven prognostics and health management for complex equipment. *Cirp Annals* 67, 1 (2018), 169–172. https://doi.org/10.1016/j.cirp.2018.04.055

[73] Eric J Tuegel, Anthony R Ingraffea, Thomas G Eason, and S Michael Spottswood. 2011. Reengineering aircraft structural life prediction using a digital twin. *International Journal of Aerospace Engineering* 2011 (2011), 1–14. https://doi.org/10.1155/2011/154798

[74] Muhammad Habib ur Rehman, Khaled Salah, Ernesto Damiani, and Davor Svetinovic. 2020. Towards Blockchain-Based Reputation-Aware Federated Learning. In *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. 183–188. https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9163027

[75] Gang Wang, Zhijie Jerry Shi, Mark Nixon, and Song Han. 2019. SoK: Sharding on Blockchain. In *Proceedings of the 1st ACM Conference on Advances in Financial Technologies* (Zurich, Switzerland) *(AFT '19)*. Association for Computing Machinery, New York, NY, USA, 41–61.

https://doi.org/10.1145/3318041.3355457

[76] Shuai Wang, Liwei Ouyang, Yong Yuan, Xiaochun Ni, Xuan Han, and Fei-Yue Wang. 2019. Blockchain-Enabled Smart Contracts: Architecture, Applications, and Future Trends. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 49, 11 (2019), 2266–2277. https://doi.org/10.1109/TSMC.2019.2895123

[77] Karl Wüst and Arthur Gervais. 2018. Do you Need a Blockchain?. In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*. IEEE, Zug, Switzerland, 45–54. https://doi.org/10.1109/CVCBT.2018.00011

[78] I. Yaqoob, K. Salah, M. Uddin, R. Jayaraman, M. Omar, and M. Imran. 2020. Blockchain for Digital Twins: Recent Advances and Future Research Challenges. *IEEE Network* 34, 5 (2020), 290–298. https://doi.org/10.1109/MNET.001.1900661

[79] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. 2018. GAIN: Missing Data Imputation using Generative Adversarial Nets. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, Stockholmsmässan, Stockholm Sweden, 5689–5698. http://proceedings.mlr.press/v80/yoon18a.html

[80] Faheem Zafar, Abid Khan, Sabah Suhail, Idrees Ahmed, Khizar Hameed, Hayat Mohammad Khan, Farhana Jabeen, and Adeel Anjum. 2017. Trustworthy data: A survey, taxonomy and future trends of secure provenance schemes. *Journal of Network and Computer Applications* 94 (2017), 50–68. https://doi.org/10.1016/j.jnca.2017.06.003

[81] Chao Zhang, Guanghui Zhou, Han Li, and Yan Cao. 2020. Manufacturing Blockchain of Things for the Configuration of a Data-and Knowledge-Driven Digital Twin Manufacturing Cell. *IEEE Internet of Things Journal* 7, 12 (2020), 11884–11894. https://doi.org/10.1109/JIOT.2020.3005729

[82] Guoqing Zhao, Shaofeng Liu, Carmen Lopez, Haiyan Lu, Sebastian Elgueta, Huilan Chen, and Biljana Mileva Boshkoska. 2019. Blockchain technology in agri-food value chain management: A synthesis of applications, challenges and future research directions. *Computers in Industry* 109 (2019), 83–99. https://doi.org/10.1016/j.compind.2019.04.002

[83] Qiheng Zhou, Huawei Huang, Zibin Zheng, and Jing Bian. 2020. Solutions to Scalability of Blockchain: A Survey. *IEEE Access* 8 (2020), 16440–16455. https://doi.org/10.1109/ACCESS.2020.2967218