Research Article
Towards Public Opinion Digital Twin: A Conceptual Prototype

Jing He [ID],[1,2] Yuanbo Qi,[3] Jie Feng,[4] and Anling Xiang[4]

[1]*Institute for Advanced Studies in Humanities and Social Science, Beihang University, Beijing 100083, China*
[2]*Beijing Key Laboratory of Urban Spatial Information Engineering, Beijing 100000, China*
[3]*Humanities Institute, Donghua University, Shanghai 201620, China*
[4]*School of Journalism and Communication, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Jing He; 287773664@qq.com

This paper proposes a novel modeling concept, the "public opinion digital twin," for public opinion analysis. The public opinion digital twin can be regarded as an experimental sandbox for social science. By digitalizing public data acquired from cyberspace into digital models, the modeling enables practical simulation, data analytics, scenario reflection, and decision support in a digital space with fine controllability, so that all possible evolutions of the research target can be analyzed. By simply inputting or filtering variables, any number of future scenarios are simulated, the effect models of each strategy for coping with public opinion are presented, and the optimized solution can be derived from continuous deep learning. If a robust digital twin is established and the required digital replicas are constantly updated, the system can perform risk assessments and trend predictions for social events. In this case, public opinion information can provide intelligent decision support for governments or enterprises and significantly facilitate social loss aversion, which will greatly advance the revolution in production, dissemination, and guidance.

1. Introduction

An emergent internet public opinion event may not only cause direct losses but also cause profound "secondary disasters." The public situation changes rapidly, and large-scale public opinion data has multidimensional attributes. Traditional mathematical modeling, social network, and other information mining methods seem to be slightly insufficient to support such complex data analysis and plan selection, which require more interaction and intuitive solutions to help users understand the impact of multiple scenarios. The combination of big data technology and digital twin technology is one of the available solutions. Therefore, we propose a new conceptual model, the "Digital Twin of Public Opinion," as a test sandbox for social development. It conducts linkage analysis of public opinion prediction, alarming, and planning and performs three-dimensional visualization modeling to realize the early alarming of negative public opinion event risk, the prediction of tendency, and the plan for response. The contributions to this article are as follows:

(1) We deconstructed the data situation of public opinion digital twins, including twin design data, twin manufacturing data, twin management data, and twin recovery data.

(2) We constructed three important spaces: real space, virtual space, and virtual subspace. Among them, the real space is the information mapping space about the offline physical objects of the public opinion event; the virtual space is based on the real space, which is the user behaviors integrated space formed by the online public opinion event; and the virtual subspace is based on the virtual space and is oriented to public opinion case library, which is a digital space for simulating the development scenarios of public opinion events and obtaining solutions for public opinion events.

For the specific content, we will describe each component of the framework in detail in Section 4. To illustrate the usefulness of the proposed solution, we apply it to the case in

Section 5. This article will end in Section 6, providing conclusions and ideas for future work.

2. Related Works

The application and popularization of the Internet have continuously deepened the impact of social information transmission, promoted the continuous expansion of the research scope of online public opinion, formed three research results in the fields of journalism, communication, and information science.

The first is the study of the law of public opinion spreading on the Internet. Mathematical models are usually constructed based on data sources generated from hot topics, and content analysis methods, numerical simulations, random forest algorithms, and statistical analysis methods are used to explore the law of the spread and evolution of online public opinion. For example: Daley and Kendall [1] initially applied the infectious disease dynamics model to the spread of rumors, divided individuals into three categories: susceptible, communicator, and immune, and then constructed a classic D-K model. In order to effectively respond to major unconventional emergencies, Dong et al. [2] established a conceptual framework for an integrated simulation prototype of emergency management by integrating social networks and propagation dynamics models and proposed a simulation planning method based on unconventional emergencies. Yin et al. [3] aimed at the behavior of users who may enter another related topic after participating in the discussion of a topic and proposed a multi-information public opinion dissemination model of major public health emergencies to understand the emergency of the public opinion dissemination mode.

The second is the study of network node identification in the spread of public opinion. Studies were mainly based on specific security accidents, natural disasters, and public health events. Social network analysis and database topic graph technology were used to quantitatively test the topological structure of public opinion dissemination and determine the key nodes in the process of public opinion dissemination. For example, Yang et al. [4] considered the role of different users in rumor dissemination, designed the state transition function for each node, proposed the rumor dissemination ILSR model, and verified its validity through the WS network, BA scale-free network, and the real Facebook network; Barbosa et al. [5] considering that individuals have different information dissemination preferences for different types of friends, established a multirelationship social network information dissemination model. Rui et al. [6] used an improved SIR model to explore the dissemination process of public opinion in coupled social networks and found the coupling network will have an impact on the spread of public opinion and could more accurately describe the real network environment.

The third is the study of the process, stage, path, and mechanism of the spread of online public opinion. Most of the studies on different stages of public opinion, such as latency, growth, spread, outbreak, decline, and death stages, are based on specific case events such as security and violence. Methods such as the tag propagation algorithm, the system dynamics theoretical model, the Bayesian network model, the random diffusion model, the content analysis method, or the text mining analysis were used to study the characteristics and paths of public opinion transmission as well as put forward measures and suggestions for public opinion processing, early alarming, and prediction in stages. For example, Chen et al. [7] introduced the evolutionary game theory into the study of emergent online public opinion dissemination; constructed a tripartite evolutionary game model among the netizens, online media, and the government and put forward suggestions for effective control of public opinion from the government's perspective. Chao et al. [8] researched the evaluation system of public opinion crisis management capabilities based on the "Life Cycle Theory" and constructed evaluation index systems for the five capabilities of public opinion early alarming, precontrol, response, handle, and utilization.

In summary, online public opinion studies are usually based on online public data resources; relied on the evolution stage of online public opinion; and supported by content association rule analysis, topic detection and tracking technology, online public opinion evaluation index system, dynamic game model, and numerical simulation and other methods [9]. In terms of theory, based on different perspectives and methods, scholars have carried out relevant research on the risk early alarming problem in the incubation period, the negative information monitoring problem in the diffusion period, and the derivative public opinion monitoring problem in the decay period, but they put more emphasis on the research content and ignored the research mechanism, lacked interdisciplinary knowledge fusion and related research on the intersection of cutting-edge technology and methods. In terms of practical application, the existing public opinion monitoring software technology has been able to real-time and automatically obtain most of the monitoring indicators, but it lacks the intuitive display of early alarming content and the reasonable configuration of risk plan content. Therefore, the theoretical and practical research of online public opinion needs to break through the traditional analysis framework and integrate emerging technology methods to promote the further development of this field, and digital twin technology is a scientific and effective breakthrough.

3. Why Digital Twin of Public Opinion?

International data company IDC said that digital twins have become a solution to manufacture companies to move towards Industry 4.0 [10].

In fact, not only the industrial society but network society can also establish a digital twin system, which is manifested in the digital restoration of the real world and the digital mapping of the virtual world. This paper defines this technology as public opinion digital twin (PODT). PODT can be understood as the digitalization of the physical world and the modeling of cyberspace. It can not only infer the flows of opinions in real time by crawling public data from the Internet but also extract future trends from behavioral

data of simulated IDs. To be specific, the network public data will be mapped into a digital model for simulation activities in cyberspace, so as to restore the actual situation of the public opinion event in the real world in the digital model. At the same time, we will establish the simulation ID of the network agent that affects the public opinion event. Through inputting the behavioral data of the analog ID into the digital model, we could observe the evolution trend of public opinion events in the digital model.

A conceptual framework of this PODT is shown in Figure 1.

From the perspective of the entire life cycle of public opinion, PODT serves different manners during four different periods—emergence and formation, growth and spread, stabilization and climax, and decline and ebb [11, 12].

(1) During the period of emergence and formation, the PODT can restore the event itself and warn someone who is potentially involved in public criticism. Meanwhile, it will conduct feature identification and risk prediction for hidden dangers and signs.

(2) During the period of growth and spread, through continuous machine learning, the PODT can iterate the digital model for a more accurate digital model. It can improve its reliability in risk prediction via estimated factors and problems in the evolution of public opinion.

(3) During the periods of stability and peaks, by comprehensively monitoring and evaluating the virtual IDs of actors that affect public opinion, the PODT continuously explores possible solutions and predicts the possible trend under these solutions. Finally, it can extend the life cycle of positive public opinion and shorten the life cycle of negative public opinion through selecting the best solution.

(4) During the last period, the PODT observes the digital models of offline events and the behavioral parameters and indicators of virtual IDs, summarize the experience, and store them in the digital model and virtual ID for future reference.

At current, PODT is in a stage of conceptual modeling. A big gap exists between the current model and an ideal sandbox system for simulation, evolution, and intelligent decision making.

4. Conceptual Design of Public Opinion Digital Twin

4.1. Definition of Public Opinion Digital Replica.
Considering the related explanations of digital replica in the industrial sphere [13–15], the definition of public opinion digital replica (PODR) can be interpreted as an integrated multiobject, multivariable, hyper-realistic and dynamic probability simulation model, which can be used to simulate, monitor, manage, investigate, verify, and predict the evolving process, status, and impact of public opinion events in offline and online environments. PODR is a simulation model of offline physical objects and online virtual objects that is integrated in the information space as well as the full life-cycle digital archives of observed objects, which could unify and integrate management of the entire life cycle of the targets. The PODR can be used to simulate, supervise, manage, investigate, verify, and predict the formation process and state of both offline and online objects in real and digital environments.

4.2. Key Issues.
Many researchers and companies have participated in the creation of simulation software and hardware for the industrial digital twin [14]. However, PODT is still in its infancy. Currently, the full-cycle association between offline objects and online objects is hard to realize during the evolution of public opinion [15]. The following issues should be addressed when considering PODT as the analytics tool for online PO solutions:

(1) The core of PODT is to build a virtual subspace that is comprehensively twinned to the online community. Unfamiliar with cycle management in manufacturing that operators fully master the information data of physical equipment, cyberspace is a huge and complex system containing quick iterative processes and multidimensional massive data. There is no doubt that it poses challenges to the PODT in data collection, processing, computing, storage, and management [16].

(2) The challenge to establish a digital twin model that is associated with online public opinion lifecycle management, real-time twinning systems, and decision management systems [17]. After feeding the online opinion data into the system, the PODT on the cloud server needs to generate detailed instructions related to public opinion inference, situation perception, simulation, ideology, social security, black public relations, and other projects. Therefore, the entire process of twinning and decision making should be updated accordingly if there is any change in the public opinion environment.

(3) Driven by digital twin, more comprehensive connections between virtual space and virtual subspace should be established for the integrated transmission of event database, response database, and effect database. The main purpose of the PODT model is to continuously accumulate knowledge in the event database, response database, and continuously reuse and improve solutions to confront public critics to enrich the effect database. When there is not an optimized data transmission mechanism between different database, it also faces the modular challenge of different databases acting on continuous public opinion threat.

(4) In terms of artificial intelligence, big data analytics should be integrated into the PODT model. When real-time network public data is collected directly from cyberspace, it will cover virtual space information on the PODT model. Big data analysis should
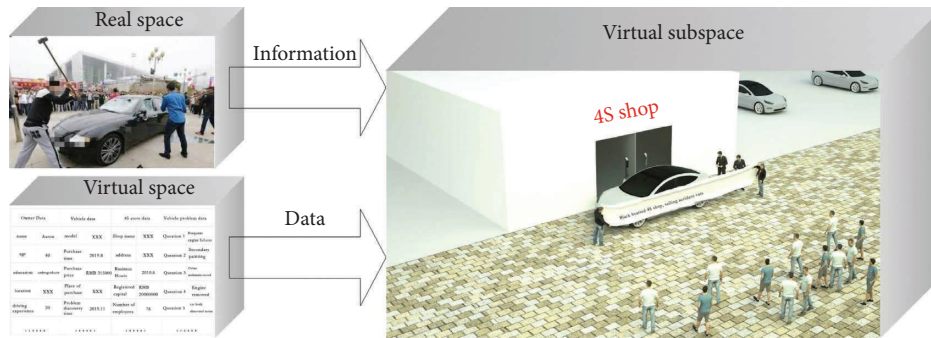
FIGURE 1: Conceptual framework of public opinion digital twin.

determine whether there are differences or not and find out the reasons for the differences when compared with actual similar public opinion cases [10]. In addition, it is desirable to intelligently decouple the problem of virtual space and virtual subspace connection.

The complete PODT will contain a large number of computer models that require high-performance parallel computing to be effectively used. The mechanism for performing these calculations and maintaining the database is not simple.

4.3. Public Opinion Digital Replica Service. PODR establishes a digital simulation on the information platform by integrating data generated from online public discussion, supplemented by artificial intelligence, machine learning, and software analysis. This simulation will automatically make corresponding changes based on data input or deletion. Ideally, a digital replica can support self-learning based on data input and present the real state of the object in the digital world in almost real time [18]. A digital replica learns by itself based on information from major media platforms as well as historical data on the Internet, or integrated network data learning. The latter two often refer to multiple research objects in the same batch performing different operations at the same time and feeding back data to the same information platform. A digital replica can be used to perform deep learning and accurate simulation based on massive information feedback [19].

The public data collected online is associated with the corresponding life-cycle analytics module in the PODR, which creates links between PO events in real space, virtual space, and virtual subspace. The 3D model formed by PODT can not only be displayed on the screen but also interact with multiple mechanisms from multiple dimensions, such as rotation zoom, highlight comparison. It has a good result for anomalous objects when some exception objects exist in public opinion events [20]. Meanwhile, it can bring improvement in efficiency and quality in the online opinion analysis process.

In terms of PODR with complex mechanism, establishing accurate and reliable system-level digital models can be difficult. Therefore, the adoption of an individual analytical model in the target system cannot obtain the best

evaluation results when it evaluates the status of offline and online virtual objects. Combining with a machine learning algorithm, this paper proposes a data-driven method to update and modify the digital model using real-time data of PO events and historical data of similar PO events. Besides, a cloud server can be adopted to manage the massive PO life-cycle data and maintain the stable operation of a digital twin. Figure 2 shows how the data layer, technical layer, and service layer of PODT interact with each other.

Moreover, by fusing multidimensional, multimodal, and heterogeneous data from diverse data platforms on the Internet, PODT services become more comprehensive and precise [21].

4.3.1. PODT Supervision. PODT Supervision can be divided to two service modes—object supervision and spatial supervision, by which the gradual or sudden variations of PODR can be accurately detected. The object supervision compares the whole and real-time public data crawled from Internet with historical data that has been recorded in the cloud server. Although there are tons of open source crawler frameworks on the Internet, it is difficult to accurately capture data from various large websites on a regular scale [22]. The challenges include changing website formats, flexible and scalable architecture, defeating website's antirobot means, and maintaining data quality. Moreover, the amount of information has exploded over time. Self-supervision is employed to capture the gradual deviation between the current PODT data and historical PODT data to reflect the current state of the crawler. The spatial supervision compares real-time PODT and related parameters with simulated mapping data in virtual subspace. Generally, virtual space and virtual subspace share the same configurations, thus the data of the both is always consistent. However, if sudden interference occurs (e.g., cyber-attacks), inconsistencies can also exist. In this case, the data from the virtual subspace will be considered as reference to detect sudden changes in the PODR due to interference in the online space. Because the virtual subspace and the virtual space are iterated together, the reference data will also be updated over time, making the PODT detection more accurate.

4.3.2. PODT Analytics. When data from both virtual space and virtual subspace are available, sufficient data sources can be used for PODT analytics. Among them, multilevel and
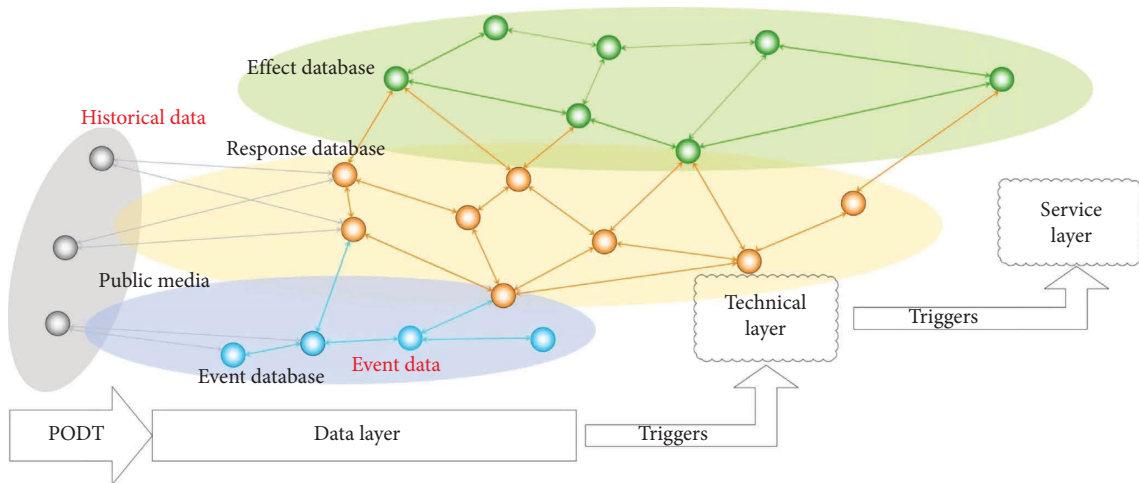
FIGURE 2: Interactive mechanism of the PODT technique.

multistage analysis, statistical and predictive analysis, and behavioral analysis are introduced as follows.

(i) Multilevel and multistage analysis: regarding team building costs, it is impractical to collect all online public data in real time. Fortunately, the simulation module in virtual subspace can serve as a substitute. This module can provide simulated PODT of multiple levels. For example, PODR is applicable to analyze the entire PO event, the context of a single event or a single participant of an event. It can also provide simulated PODT of multiple stages. For example, the PODR can be adopted in multiple stages in single PO event, which includes the emergence and formation, growth and spread, stabilization and climax, and decline and ebb. Based on the simulated data, multilevel and multistage PODT analytics can be performed.

(ii) Statistical and predictive analysis: Since the virtual subspace can record historical PODT data in the virtual subspace, statistical analysis can be performed in different intervals to extract PODT patterns from both long-term and short-term aspects. In addition, the trend of PODT and related parameters can be analyzed based on predictions performed in the virtual subspace.

(iii) Behavioral analysis: This module analyses the participant behaviors, such as friendly behaviors or aggressive behaviors, through methods of event reflection and virtual ID. From the perspective of PO event reflection, the PO contextual features can be extracted from the relevant parameters of real-time PODT data. In terms of virtual ID, the behavioral features of participants can be further analyzed and predicted from the PO event reflection. The combination of contextual features and behavioral features, each of which links to a corresponding solution, contributes to more accurate behavioral analysis.

5. Very First Conceptual Model for the Public Opinion Digital Twin

5.1. Public Opinion Digital Replica Data.

The fundamental PODR data includes event periodic data and event twin data, of which the former can be regarded as a direct data source for PODR and the latter as a supplemental data source. The data structure is detailed in Table 1. These data are used for analytics as follows:

(1) Simulating, supervising, managing, analyzing, and verifying the digital duplication; functioning as early warning, prevention, and prediction in the PO evolution process; providing the fundamentals for follow-up PO supervision and guidance, and thereby proposing an optimized strategy for PO governance.

(2) Collaborating on efficient digital twinning from the physical world to the digital world and facilitating the combination of management and service in the entire cycle of PO evolution, eventually efficient twinning and intelligent interaction between cyberspace and the real world are enabled.

(3) Recording the entire PO evolution process (including initiation, development, and termination) during the digitalization of PODR and enabling information retrieval and 3D visualization functions for historical and real-time data.

5.2. Prototype of Public Opinion Digital Replica

5.2.1. Twinning Approach.
PODT provides a real-time twinning of online public data among real space, virtual space, and virtual subspace, in which a full-cycle evolution is simulated and predicted. The procedure can be divided into five sections:

(i) Context twinning: accurate, comprehensive, and dynamic twin of the offline PO events in the real world.

TABLE 1: Data structure of PODR data.

Event periodical data	Latency period	Risk latency, event trigger, etc.
	Formation period	PO fermentation, amplification and propagation, rumor spread, event definition, etc.
	Fluctuation period	PO popularization, convergence, breakthrough, reversion, counterwork, climax, defocus, etc.
	Receding period	PO ebb, resurgence, verification, rigidity, etc.
Event twin data	Twin design	Design elements, logic, operations, etc.
	Twin duplication	Duplication progress, procedure supervision, optimization, etc.
	Twin management	Management framework, upgrade, operation, etc.
	Twin derivation	Derived event database, response database and effect database

(ii) Participant twinning: formulating user behaviors related to the target events in cyberspace.

(iii) Knowledge twinning: event-evolving patterns that are analyzed and recognized by historical data mining.

(iv) Simulation twinning: PO evolution under simulated scenarios where specific data were input or removed.

(v) Strategy twinning: most recommended solution in response to real input data, which is selected by combining big data and artificial intelligence techniques.

5.2.2. Precise Replication of Physical Objects. PO digital twinning based on discrete event simulation (DES) can be considered as a predictive decision-support model, which enables an accurate replica of physical objects and estimates decision effects in virtual subspace without interrupting the evolution of events (Senington, Baumeister, Ng, & Oscarsson, 2018). As a conceptual design for a digital twin, it can properly schedule social resources and discover issues before their emergence.

This paper seeks a common and extensive framework for PO event simulation. To inspect and evaluate the usage efficiency of social resources more precisely, the proposed approach simulates four periods of PO events. This approach is based on three elements: (1) Dynamic variables such as participants. The participants are generally classified as positive participants, neutral participants, and negative participants, which are represented by different trajectory instances. (2) Static resources, such as the circumstances of a specific PO event. (3) Affected results, such as the participant behaviors after an event. The simulation model is designed by following these steps:

(i) Collecting data information, which includes both offline and online events, and the corresponding influencing factors.

(ii) Generating flow charts that depict various participant movement paths under different administrative actions.

(iii) Monitoring the evolving status and estimating coping strategies by factor input and variable tracing.

To test the approach, five doctor-patient (D-P) scenarios with simulated data are analyzed, each of which shares a duration of 10 hours (9 : 00 to 19 : 00) (Jimenez, Jahankhani,

& Kendzierskyj, 2020; Karakra, Fontanili, Lamine, Lamothe, & Taweel, 2018). Algorithm 1 of Table 2 shows the simulation model for the precise replication of physical objects. Table 3 shows the participant count analyzed in each scenario.

Figure 3(a) is a flow chart that illustrates the participant path in a D-P event. A patient first arrived in the main hall, then entered a director's office, asking for a relevant response. The patient may claim disagreement with the response, and disagreement can rise to argument. As the dispute intensified the patient called family dependents for help and waited in a waiting room. After 2 hours, they met together in the main hall and went back to the office to demand further explanation. They assaulted the officer after the mediation failed. Subsequently, the security guards arrived, and the conflict eventually moved to the exit area. Figure 3(b) depicts a simulated 3D scenario, where the occupied resources are shown in Table 2.

5.2.3. Online Participant Simulation. The approach also collects user data from those involved in PO events. The simulated users can be classified according to real features and their corresponding simulated features, with records of real data that are interpreted into structural information afterward for comparison. By configuring user tags, these "virtual users" are created to simulate real-user operations such as browsing, clicking, and blogging.

The technical framework can be divided into a data preservation layer (DPL), a modeling calculation layer (MCL), and a modeling representation layer (MRL). Real-time data is captured in DPL and delivered to a data distribution service, the Kafka cluster. Then the Kafka cluster distributes the data to an OSS file system and a NLP module, which store and analyze the raw data. After the data is fed into MCL, a data-driven mathematical model is used for multilevel modeling of PODR. The derived model should match and synchronize with user features. By changing the data inputs, the model supports status prediction and robustness evaluation for event evolution. Based on the analytics in MCL, MRL visualizes the analyzed results in a virtual subspace (Zheng, Yang, & Cheng, 2019). Figure 4 shows the layered framework structure.

The system records account data including registration information and browsing information and simulates user behaviors at an online scenario. The main inputs involve variables in three dimensions: (1) Content features: for example, text, images, and videos. (2) Analytics features: including static tags, dynamic tags, and hidden tags depicted

TABLE 2: Precise replication of physical objects.

Algorithm 1 precise replication of physical objects simulation model
Input: dynamic variables such as participants DYN_VAR, Static resources that affect variables STA_RES.
Output: affected results, such as the participant behaviors after an event AFF_RES.
1: //Step 1. Collecting data information
for each both offline events and online events, corresponding influencing factors **do**
2:    Prepare input variables;
3:    Start a crawler using input variables;
4:    Crawl data and save to temporary variables;
5:    Assignment of temporary variables to global variables DYN_VAR, STA_RES;
6: **end for**
7: //Step 2. Generating flow charts
for each different administrative actions **do**
8:    **for** each various participant movement paths **do**
9:       Design flowcharts to describe the relevant paths and the environment in which they occur;
10:    **end for**
11: **end for**
12: //Step 3. Get results
if successfully track variables **then**
13:    Input impact factor;
14:    Monitoring the evolving status;
15:    Estimating coping strategies;
16:    Store the affected entity objects AFF_RES;
17: **end if**

TABLE 3: Resource usage during the simulation.

	Resource counts				
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Receptionists	1	1	1	1	1
Physicians	1	2	2	8	12
Volunteers	1	1	2	2	5
Offices	1	2	1	1	1
Patient dependents	0	0	7	7	7
Triage	2	1	4	5	1
Wheelchairs	1	1	1	1	1
Waiting rooms	1	1	1	1	1
Queuing lines	1	1	1	1	1
Communication tools	1	1	3	4	8
Security guards	2	2	2	4	4
Entrances	1	3	3	2	1
Exits	1	2	2	2	1
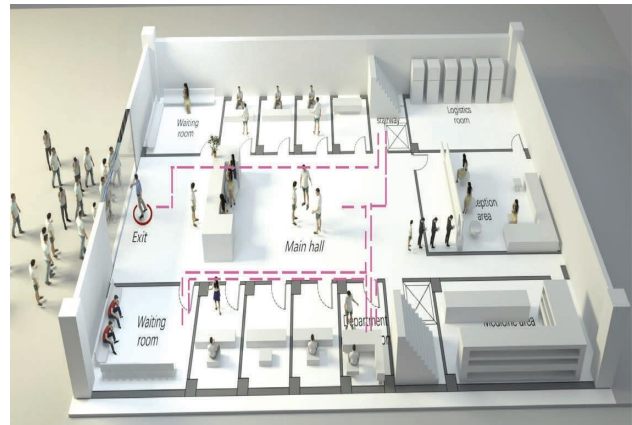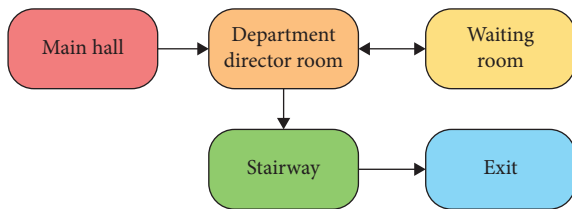
(a)

(b)

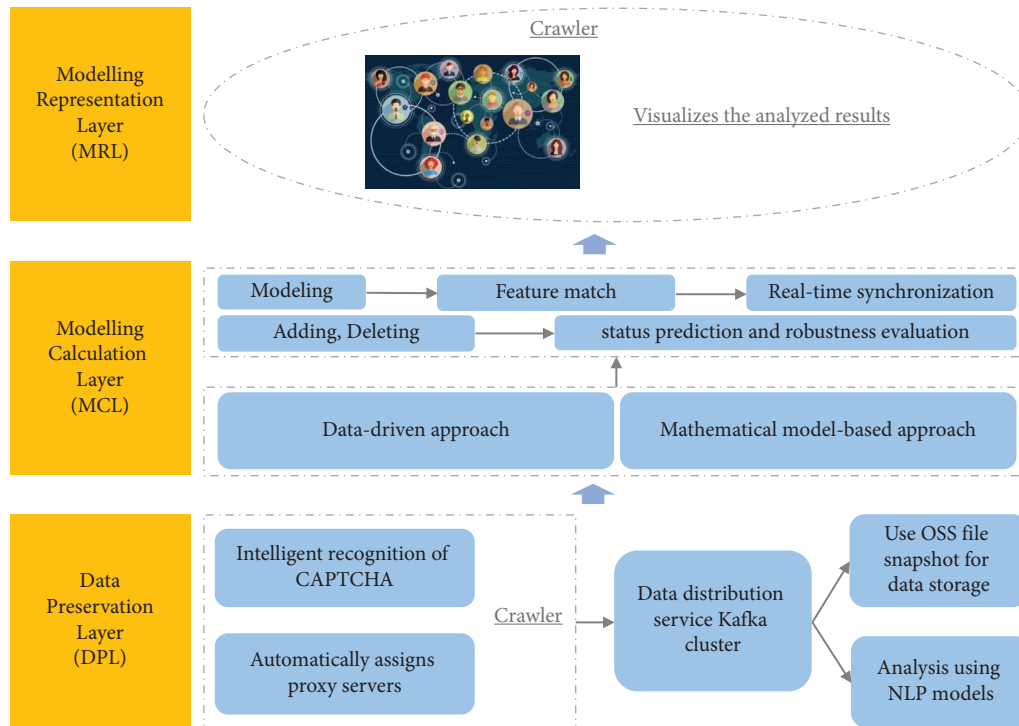FIGURE 3: (a) Tracking flow chart and (b) simulated 3D scenario of a D-P event.

FIGURE 4: Technical framework using layered presentation.

by several models. (3) Contextual features: online behaviors in different spatiotemporal scenarios. A virtual ID digitalization model is established to collect and quantify user behavior data such as views, likes, comments, and reposts. During this process, a noise filter is employed to remove clicks with rather short intervals, and a mechanism of time-varying weight is applied so that behavior feature weights decay over time while new behaviors dominate. By simulating a series of user clicking behaviors, virtual user IDs are established for PO analytics. Moreover, a data feed mechanism is built to enable bidirectional information feedback on virtual space and virtual subspace. To meet the simulation standard for data comparison between real participants and virtual user IDs, interactive behaviors are simulated for feature data analytics. When the interactive behaviors of virtual users are consistent with those of real ones, a certain amount of virtual ID can be added as one of the multimodal data sources. Figure 5 depicts the overall simulation process for online participants.

5.2.4. Elaborated Case Database Construction. The case database in virtual subspace includes three correlative subdatabases: event databases, response databases, and effect databases. The event database is an interpreted PO event set after capturing civil opinions and standpoints from the online public data. The response database provides common reactions and strategies to tackle these PO events. The effect database derives indicators of the probable performance of each event-response pair. The coping effect can be estimated according to the PO feedback comparison before and after the reactions are taken.

5.2.5. Public Opinion Inference Framework. The PO event evolution is typically based on the similarities between and periods of events, which includes the propagation forecast, coping measures and effect estimation. Specifically, by determining event similarity based on instances recorded in the event database, this framework performs dynamic prediction and visualization of follow-up event propagation. Meantime, the module also combines real-time data with big data and AI techniques to automatically recommend the optimized coping strategy. The inference approach can be shows in Figure 6.

5.2.6. Full-Cycle Connection. To establish full-cycle connections among real space, virtual space, and virtual subspace, the system conduct a simulated "social system" or "social experiment," where the real-time data in the virtual space are intelligently twinned to virtual subspace and the derivation in the virtual subspace is restored in virtual space.

5.2.7. Experiment. The synchronized operation of the virtual space is driven by the correlation and integration of the physical world and the digitalized models.

Case 1. Public Opinion Event of the "Unfinished Building."
    Taking the PO event of an unfinished building as an example, we will briefly explain the simulation process of PODT. Figures 7 and 8 present two snapshots of the precise replication. As depicted in Figure 7, a series of negative factors have resulted in a broken funding chain and the stalled construction of real estate. The owners who have pre-purchased these estates cannot check in on the agreed date,
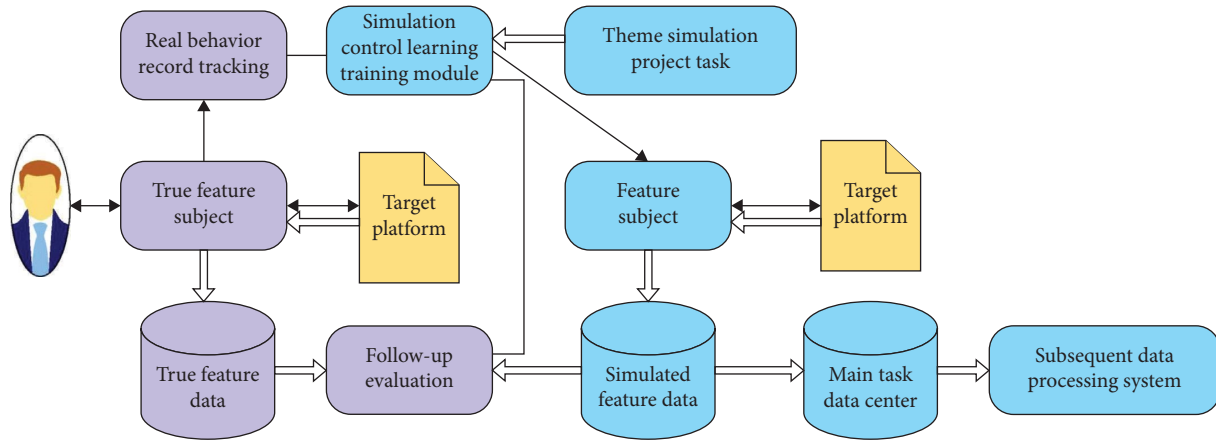
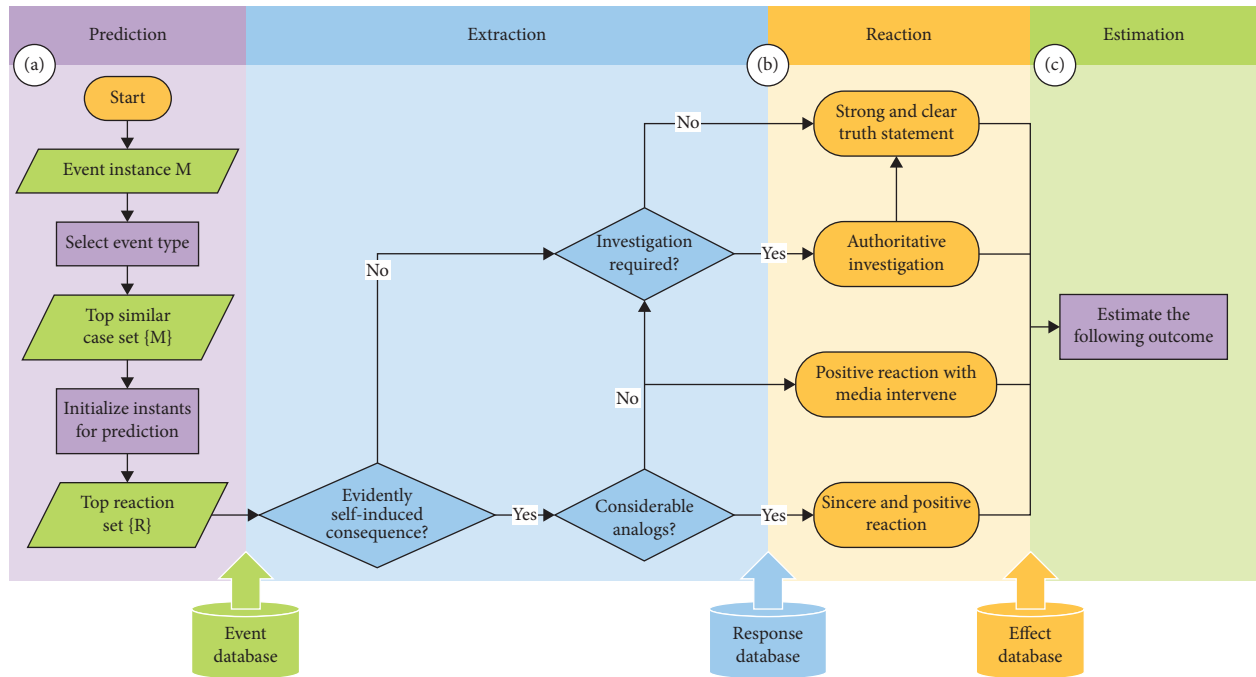Figure 5: Schematic diagram of online user simulation.

Figure 6: Illustration of the public opinion inference framework. (a) The target event M is fed into the pipeline, and the event type (e.g., civil case ⟶ unnatural death) and timeline are normalized to derive a related case set {M} and a recommended solution set {R}. (b) After a series of determinations, proper reactions are extracted according to various circumstances. (c) Finally, the system will estimate the performance of each reaction.

hence the owners' economic loss and collective right-protection activities. Figure 8 shows the further evolution of the event where the disagreement on compensation issues has triggered conflicts between buyers and developers. The rights defenders require the development companies to come up with reasonable follow-up plans.

Figures 9 and 10 present the user simulation of online commenters. Due to the failure of raising funds and reaching an agreement with the owners, the developers hired a large number of astroturfers (paid commenters to postfalse propaganda) to stabilize confidence of the remaining owners and distract the focus of right defenders. The system simulates potential evolutions of the PO event in the context of different postcounts and poster emotions. Figure 10 shows the real estate trend one year later, when the development enterprise has raised a certain amount of funds through development bonds, bank credit, investment trusts, and funding allocation, and a 60% funding gap has been filled. In this case, the developer has restarted the estate construction, most of the renovation has been completed, and the check-in agreement has been renegotiated. Although some owners still participate in right defense, the total number of defenders has decreased.

Case 2. Public Opinion Event of "Surrogacy and Abandonment."

FIGURE 7: The developer's funding chain has broken, and the properties cannot be delivered on time. So the owners are collectively defending their rights.

FIGURE 9: The development company hires astroturfers to maintain positive public relations.

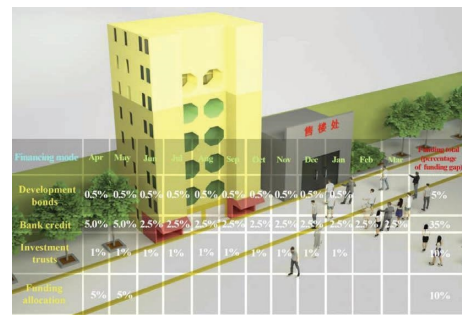FIGURE 8: On-site disputes between buyers and developers in the sales office.

FIGURE 10: Real estate situation after the developer has raised relatively sufficient funds.

Figures 11 and 12 are the accurate reductions of the physical objects in the virtual subspace. Figure 11 restores the scene of the emotional breakdown of the female star and her ex-boyfriend based on the recorded information. Both the male and female, and both parents, discuss how to deal with the child after the surrogacy is seven months old. During the period, the man performed the recording act. During the period, the man carried out the recording behavior. Figure 12 shows the scene of the man with the baby stranded in another country due to the epidemic and litigation issues after the man and woman failed to reach an agreement on the child rearing issue, and the scene was made public through social networking sites, forming a public opinion event of "surrogacy and abandonment."

Figures 13 and 14 simulate the different network behaviors of the female star in the virtual subspace, and the evolution effects of related public opinion events are formed. Figure 13 shows that the female star's first social speech method, to avoid the important and dwell on the trivial, aroused the negative emotions of a large number of netizens. There have been endless discussions on topics such as "surrogacy," "oocyte cryopreservation," and "abandonment." Then, the second and third times when the female star posts her statements on social media, they distract the focus of netizens, triggering the mainstream media to speak sternly, further triggering netizens to resist comments, and eliciting topics such as women's rights and deprivation of liberty. Figure 14 shows that in the first speech, the female

star adopted the method of "proactively admitting wrong surrogacy, emotionally attracting netizens, and actively raising children" in response to the two hot spots of the netizens' attention: Netizens' negative emotions continue to slow down as the event progresses. Compared with the first online behavior, it can effectively reduce the negative emotions of netizens by about 40%.

Both processes are presented by adding the female star's text data to the corresponding 3D effect model, which simulates the possible evolution of public opinion events when the female star adopts different network behaviors.

5.3. Technical Strengths.
As an emerging intelligent technology, PODT has promising prospects in the management, simulation, supervision, verification, and prediction of public opinion. It is expected that in ten years, more than half of the public media platforms will be able to apply this technique to optimize resource allocation in the real world. This design exhibits several strengths in public opinion analytics.

(1) PODT creates a twin connection between the offline/ physical objects and online/virtual objects by linking various features into virtual subspace through a digitalization module. The digitalized twin enables repeatable operations such as copy, transfer, update, and delete, which inspires exploration of novel approaches to optimize the supervision, guidance, and control of PO events.

FIGURE 11: The scene of the discussion of surrogacy issues between the man and woman and their parents.

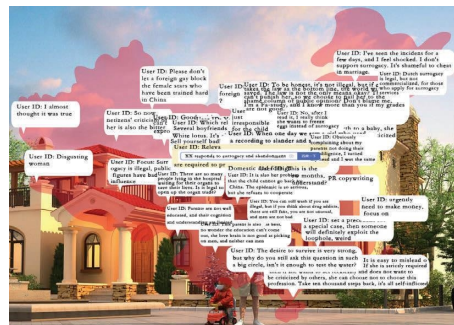FIGURE 12: The scene of a man with a baby stranded in another country.

FIGURE 13: The evolution of public opinion of female star avoiding "surrogacy and abandonment."

FIGURE 14: The future scene of a female star after admitting her wrong surrogacy and raising a child.

TABLE 4: List of abbreviation.

Explanation	Abbreviation
PO	Public opinions
PODT	Public opinion digital twin
SNS	Social networking service
PODR	Public opinion digital replica
DES	Discrete event simulation
DPL	Data preservation layer
MCL	Modeling calculation layer
MRL	Modeling representation layer
OSS	Object storage service
NLP	Natural language processing

(2) Based on big data and machine learning technologies, PODT can infer several indicators that cannot be directly measured by traditional methods. Machine learning helps discover different PO evolution patterns in historical data, which enables the propagation forecast and coping effect prediction.

(3) Existing PO cycle management methods are rarely able to accurately explore and predict the underlying issues behind the events. The PODT technique combines the data collection of the Internet of Things, the data processing of big data, and the data modeling of artificial intelligence to analyze the current state, past problems, and future trends. All possible cases are simulated to provide more comprehensive decision support.

(4) In the field of traditional PO governance, empirical analysis is rather vague for studying public opinion and cannot be regarded as a basis for precise determination. A key advancement of PODT is the ability to digitize expert knowledge that could not be stored previously, which enables intelligent analytics of public opinion. For example, for various risk characteristics that are exposed during the evolution of large-scale PO events, the historical data can be fed into a machine learning model to train digital features of various risk phenomena. Combined with the records of expert suggestions, the fundamentals of risk assessment are digitally established to cope with tricky PO events in the future. The system also continuously trains and updates features for new forms of risk, which is expected to eventually enable autonomous diagnosis and recommendations.

(5) Compared with traditional methods of PO supervision, performing practical operations on virtual models lends fault tolerance to research subjects, which is essential for social research.

6. Conclusion and Discussion

The paper proposed new conceptual modeling, which performs simulation processes by remapping virtual numbers into digital models in cyberspace, mapping network public data into a digital space, combining data, algorithms, and decision-making analysis, and reappearing the event scenes so that we can monitor the possible occurrence of the research object in the digital model. By clicking the mouse and entering or deleting variable data, the corresponding public opinion event effect model will be presented while simulating any possibilities of future scenarios and obtaining the optimal public opinion event solution. Table 4 shows the abbreviations associated with this concept in the text.

The most significant limitation of this method is that the simulated user information is quite limited and may involve false information. For example, the user attributes are only available from the users' registration information. But actually, the users might provide false information because of privacy issues. Of course, security facilities will also be a long-term common issue facing the landing of digital twins in the field of public opinion information. Once a vicious attack occurs, it will seriously endanger social stability. This is a dynamic and long-term process. The focus of the future work of Public Opinion Digital Twin Technology can combine multiple information security application scenarios, such as anti-attack intelligent identification and active defense, to explore and establish an AI pre-training model that meets the requirements of multi-scenario applications. During this process, a large amount of involved contextual data is also recorded in the virtual space, which helps to depict the interaction between the objects and their circumstances. The following challenges should be further addressed when dealing with more complicated PO situations:

(1) Richer environments should be developed through real-time data links between virtual space and virtual subspace. In terms of factors that lead to system vulnerability, such as hack attacks or privacy issues, proper security techniques are especially essential for confidentiality protection.

(2) The virtual subspace should be developed to support the offline generation of the simulation scheme, and the adopted model should be transferred to real-time environments. Evaluation methods need to be considered for the robustness assessment of the developed models.

Data Availability

The [.csv] data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there no conflicts of interest.

Acknowledgments

References

[1] D. J. Daley and D. G. Kendall, "Epidemics and rumours," *Nature*, vol. 204, no. 4963, p. 1118, 1964.

[2] S. Y. T. Dong, F. H. Fan, and Y. C. Huang, "Studies on the population dynamics of a rumor-spreading model in online social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 492, pp. 10–20, 2018.

[3] F. Yin, J. Lv, X. Zhang, X. Xia, and J. Wu, "COVID-19 information propagation dynamics in the Chinese Sina-microblog," *Mathematical Biosciences and Engineering*, vol. 17, no. 3, pp. 2676–2692, 2020.

[4] L. Yang, Z. Li, and A. Giua, "Rumor containment by spreading correct information in social networks," in *Proceedings of the 2019 American Control Conference (ACC)*, IEEE, Philadelphia, PA, USA, July 2019.

[5] C. Barbosa, L. Félix, V. Vieira, and X. Carolina, "Sara - a semi-automatic framework for social network analysis," in *Proceedings of the XXV Simpósio Brasileiro de Sistemas Multimídia e Web*, Florianópolis, 2019.

[6] X. Rui, F. Meng, Z. Wang, G. Yuan, and C. Du, "SPIR: the potential spreaders involved SIR model for information diffusion in social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 506, pp. 254–269, 2018.

[7] C. Chen, H. Liu, and X. Guan, "A game theory based model for internet public opinion's embryonic stage," *Cybernetics and Information Technologies*, vol. 16, no. 6, pp. 43–59, 2016.

[8] X. Chao, G. Kou, Y. Peng, and EH. Viedma, "Large-scale group decision-making with non-cooperative behaviors and heterogeneous preferences: an application in financial inclusion," *European Journal of Operational Research*, vol. 288, no. 1, pp. 271–293, 2021.

[9] B. Ashtari Talkhestani and M. Weyrich, "Digital Twin of manufacturing systems: a case study on increasing the efficiency of reconfiguration," *Automatisierungstechnik*, vol. 68, no. 6, pp. 435–444, 2020.

[10] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9-12, pp. 3563–3576, 2018.

[11] C. Zhuang, J. Liu, and H. Xiong, "Digital twin-based smart production management and control framework for the complex product assembly shop-floor," *International Journal of Advanced Manufacturing Technology*, vol. 96, no. 1-4, pp. 1149–1163, 2018.

[12] K. Wang, D. Liu, Z. Liu, Q. Wang, and J. Tan, "An assembly precision analysis method based on a general part digital twin model," *Robotics and Computer-Integrated Manufacturing*, vol. 68, Article ID 102089, 2021.

[13] X. Tong, Q. Liu, S. Pi, and Y. Xiao, "Real-time machining data application and service based on IMT digital twin," *Journal of Intelligent Manufacturing*, vol. 31, no. 5, pp. 1113–1132, 2020.

[14] N. Nikolakis, K. Alexopoulos, E. Xanthakis, and G. Chryssolouris, "The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor," *International Journal of Computer Integrated Manufacturing*, vol. 32, no. 1, pp. 1–12, 2019.

[15] C. Stary, "Digital twin generation: Re-conceptualizing agent systems for behavior-centered cyber-physical system development," *Sensors*, vol. 21, no. 4, p. 1096, 2021.

[16] Z. Wang, W. Feng, J. Ye, J. Yang, and C. Liu, "A study on intelligent manufacturing industrial internet for injection molding industry based on digital twin," *Complexity*, vol. 2021, Article ID 8838914, 16 pages, 2021.

[17] C. Zhuang, T. Miao, J. Liu, and H. Xiong, "The connotation of digital twin, and the construction and application method of shop-floor digital twin," *Robotics and Computer-Integrated Manufacturing*, vol. 68, no. 4, Article ID 102075, 2021.

[18] H. Wang, H. Li, X. Wen, and G. Luo, "Unified modeling for digital twin of a knowledge-based system design," *Robotics and Computer-Integrated Manufacturing*, vol. 68, Article ID 102074, 2021.

[19] A. Ladj, Z. Wang, O. Meski, F. Belkadi, M. Ritou, and C. Da Cunha, "A knowledge-based Digital Shadow for machining industry in a Digital Twin perspective," *Journal of Manufacturing Systems*, vol. 58, pp. 168–179, 2021.

[20] C. Boje, A. Guerriero, S. Kubicki, and Y. Rezgui, "Towards a semantic construction digital twin: directions for future research," *Automation in Construction*, vol. 114, Article ID 103179, 2020.

[21] V. A. Dolgov, P. A. Nikishechkin, and V. E. Arkhangelskii, "Models for managing production systems of machine-building enterprises based on the development and using of their digital twins," *The European Physical Journal Conferences*, vol. 248, no. 4, p. 04015, 2021.

[22] C. Star, C. Kaar, and M. Jahn, "Featuring dual learning experiences in tangible CPS education: a synchronized internet-of-things – digital-twin system," in *Proceedings of the EICS '21: ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, 2021.