Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

A GENERAL FRAMEWORK FOR HUMAN-IN-THE-LOOP COGNITIVE DIGITAL TWINS

Parisa Niloofar

Mærsk Mc-Kinney Møller Institute
University of Southern Denmark
Campusvej 55
Odense, 5230, DENMARK

Sanja Lazarova-Molnar

Institute AIFB
Karlsruhe Institute of Technology
Kaiserstr. 89
Karlsruhe, 76133, GERMANY

Femi Omitaomu
Haowen Xu

Oak Ridge National Laboratory
1 Bethel Valley Road
Oak Ridge, TN 37831, USA

Xueping Li

University of Tennessee, Knoxville
851 Neyland Drive
Knoxville, TN 37996, USA

ABSTRACT

Modelling and analysis of systems that are equipped with sensors and connected to the Internet are becoming more automated and less human-dependent. However, bringing expert knowledge into the loop along with data obtained from Internet of Thing (IoT) devices minimizes the risk of making poor and unexplainable decisions and helps to assess the impact of different strategies before applying them in reality. While Digital Twins are more of a data-driven simulation of the physical system, Cognitive Digital Twins bring the human dimension into the modelling and simulation. In this paper, we aim to emphasize the crucial role of explainability and the underlying rationale behind automated or interactive decision-making processes. Furthermore, we propose an initial framework that delineates the specific points within the feedback loop of a cognitive digital twin where human involvement can be incorporated.

1 INTRODUCTION

The recent advances in digital transformation are metamorphosing Industry 4.0 for effective decision-making. In many applications, field assets, machines, products, plants, and factories are increasingly being connected to the Internet. This connectedness allows systems to be located, communicated with, analyzed, and controlled via the network. Thus, the recent digital transformations are revolutionizing new types of services and business models. On the computational side, cyber-physical systems (CPSs) have been proposed as a key concept of Industry 4.0 architecture. The term cyber-physical systems refers to a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities (Baheti and Gill 2011). The cyber-model of such a physical system is called a digital twin (DT). Simply stated, a DT concept includes constructing a digital representation or model of an individual product to improve the accuracy of maintenance and performance predictions for individual products (Kobryn 2020). Zhang et al. (2022) argue that DTs are essentially dynamic data-driven models that serve as real-time symbiotic "virtual replicas" of real-world systems. For actionable applications, such a replica must be a *realistic* and *dynamic* representation of the physical system with respect to the predefined goals. The word “twin” in DT implies that the replica system would be linked to the physical system throughout its entire lifecycle. In other words, the digital replica can be treated as an entity on its own. The recent advances in smart sensors, Internet of Things, cloud computing, machine learning (ML), and artificial intelligence (AI) are enabling this transformation. Applying the DT concept to Industry 4.0 will allow the creation of fit-for-purpose digital representations of industrial operations and processes using collected data and information to enable analysis, decision making, and control for a defined objective and scope (Kumbhar et al. 2023; Chaudhari et al. 2022; Mourtzis 2021; Radanliev et al. 2022).

The Cognitive DT (CDT or CT) concept reveals a promising evolution of the current DT concept towards a more intelligent, comprehensive, and full lifecycle representation of complex systems (Zheng et al. 2022). A CDT serves as an extension of the traditional DT by incorporating three additional components, namely the access, analytics, and cognition layers (Ali et al. 2021). The *access layer* is the enhanced communication layer, especially with respect to access to data about the state of the physical system. The *analytics layer* brings advanced ML and AI into the framework to enhance actionable knowledge. The *cognitive layer* enables human cognition to convert the traditional DTs into smart and intelligent systems.

The initial idea for incorporating the human dimension into DTs is proposed by (Agrawal et al. 2023). CDTs demonstrate a close association with both Human-Computer Interaction (HCI) and Human-Machine Interaction (HMI), which focus on establishing seamless interfaces between humans and DTs (Alcaraz and Lopez 2022). The "human-in-the-loop" concept is also known as interactive analytics, in which analytic algorithms occasionally consult human experts for feedback and course correction. This concept is widely adopted by the visual analytics communities to facilitate explorative analysis of complex datasets, where the (a) questions are ill-defined or unknown a priori and training data is not available, and (b) many processes entail tacit experiences and knowledge that are difficult to capture using mathematical representations (Endert et al., 2014).

For human experts to provide feedback and interact with DTs, explainability of the algorithms and results are essential. This particularly holds true for numerous prevalent black-box ML modeling techniques and algorithms. Unexplainable results render human experts incompetent to effectively improve the performance or rectify faults in a system. Nevertheless, the constant-evolving behaviors of dynamic systems often entail new information with partially unknown mechanisms, which are not characterized by the previous training data and the existing body of knowledge. These new details about the (cyber)physical system are often presented to its DT as corner cases, where the existing methods and pre-defined rules are inadequate for facilitating decision-making and process optimizations. In such corner cases, it becomes crucial to integrate human supervision, along with expert knowledge, experience, and justifications, into a DT. This integration aims to enhance the comprehension of the unknowns within (cyber)physical systems and to refine the design of the underlying data-driven methodology.

In this study, we see DTs as Dynamic Data-Driven Simulations (DDDS) where the model updating is autonomous and fully data-driven. In contrast, in the context of CDTs, human cognition is integrated with data in various dynamic data-driven modeling stages. We, furthermore, discuss the role of explainability of the models implemented for model derivation and exemplify our discussions through a reliability analysis example. The structure of the paper is as follows: Section 2 provides a background on dynamic data-driven simulation and modelling, CDT, human-in-the-loop and explainability. In Section 3, we present a categorization of model extraction methodologies, related explainability aspects and their feedback loops. We also illustrate our framework by highlighting the steps in which domain expert can interact with a DT to build a CDT. Finally, in Section 4, we conclude the paper.

2 BACKGROUND AND RELATED WORK

Boschert and Rosen (2016) describe DT as a model of a component, product, or system developed by a collection of engineering, operational, and behavioral data which support executable models, where the models evolve over the lifecycle of the system and support the derivation of solutions which assist the real-time optimization of the system or service. Simulation in its classical way, also builds a virtual model of the real-world system but offers much less than DTs. In fact, traditional simulation and modeling enable DTs that are not up to date (or they are "dead"). Dynamic data-driven simulation (DDDS) sustains the vitality of virtual models by offering dynamic feedback, enabling the continuous updating of these models using real-time data sourced from the physical domain (Rokka Chhetri and Al Faruque 2020). In the next subsection, we provide more details on DDDS.

2.1 Dynamic Data-Driven Simulation (DDDS)

DDDS has been successfully applied to broad range of application areas, such as smart manufacturing (Friederich et al. 2022), smart cities (Fujimoto et al. 2016), health care (Gaynor et al. 2005), and security (Faniyi et al. 2012). DDDS uses data assimilation to dynamically incorporate real-time observation data into a running simulation model (Figure *1*). Data assimilation is an analysis technique in which the observed data is assimilated into the model to produce a time sequence of estimated system states (Bouttier and Courtier 1999). The goal of data assimilation is to provide an updated estimate of the "current" system state, which is often hidden and cannot be observed (Hu 2015). The estimated system state is then used to simulate/predict the system dynamics in the future.

In DDDS, new sensor data arrives sequentially, and the simulation system needs to be continuously updated. Often, data arrivals are not consistent with their real times of occurrences. Therefore, time delays should be considered in model building. Since a simulation system includes model design along with its parameters, they need to be updated as the real-time sensor data are streaming. However, many sophisticated simulation models (such as agent-based crowd behavior simulation models) lack the analytic structures from which functional forms of probability distributions can be derived (Hu 2015). The reason is that usually these probability distributions have non-Gaussian behaviors. It means there is a need to implement estimation techniques that are independent of the form of the probability distributions and that means using nonparametric methods.

Providing an updated estimate of the "current" system state continuously, keeps the digital replica of the real-world system alive. The aim of having a live simulated version of the real system, is to be able to act on time and with less hazardous consequences.

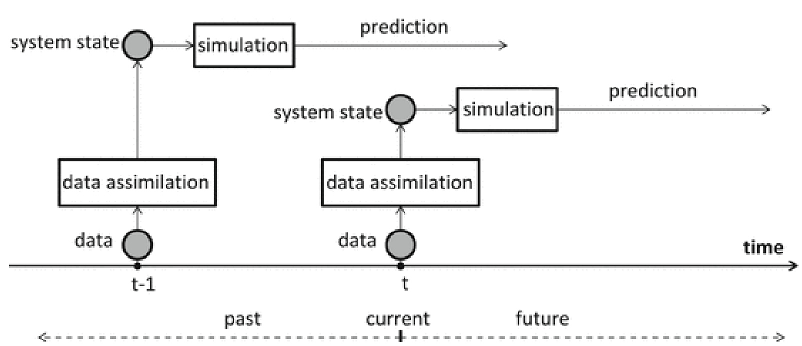


Figure 1: Data assimilation in dynamic data driven simulation (Hu 2015).

2.2 Human-in-the-Loop and Explainability

Some examples of fully data-driven approaches, i.e., reinforcement learning, depending on the context, might require thousands or millions of data samples to converge to a satisfactory policy and are subject to catastrophic failures during training (G. Goecks 2020). Moreover, while data-driven methods hold great promise for automatically extracting relevant features, they can also extract meaningless information (Hagan et al. 2014), especially in the presence of unobserved factors or hidden variables. For instance, data-driven methods extract a relationship between the height of a child and the number of words he/she knows. While a piece of knowledge about the child, like the child's age explains this relationship. Classification, prediction, and clustering of behaviors in today's adaptive systems are becoming increasingly challenging due to the volume, high dimensionality, and multi-modality of data (Drachen et al. 2016). Hence, there exists a need to bring humans into data-driven modelling to support sophisticated decision-makings, where the process complexity combined with potential unknown mechanisms and concerned cases have presented both technical and methodological challenges to the AI and ML techniques.

Humans/experts can provide prior information about the system and integrate it into the model building, or just after the collection of a few data samples, integrate their updated knowledge in model training, validation or optimization. Human intervention can prevent making decisions leading to catastrophic actions. In the context of developing a DT, the term "cognition" could carry two different connotations, which either refer to the DT's cognitive capability that is enabled through the computational methods (e.g., AI and ML) or the DT's ability to facilitate human cognition of complex data, procedure, and problems involved in a manufacturing process.

Humans need to understand the patterns or rationale behind the system's decision-making logic in order to trust and follow the instructions or take further decisions (Zhang et al. 2022). The explainability of computational methods, such as simulation and ML models, is developed to help humans understand, appropriately trust, and effectively manage the emerging generation of AI-based smart systems, enabling humans to recognize, learn, understand, interpret, articulate, supervise an AI model or intelligent system to make optimized decisions (Minh et al. 2022). In the smart manufacturing sector for instance, the interpretability and explainability of complex dynamics, datasets, and black-box models usually rely on the DT's capability to facilitate human cognition. This cognitive capability is often enabled through advanced HCI techniques (Ahmed et al. 2022). At the conceptual level, Ahmed et al., (2022) outlined four major approaches to increase the explainability of AI models and smart systems in the context of Industry 4.0. These approaches include (a) model-specific and model agnostic approaches, (b) local and global approaches, (c) pre-model, in-model, and post-model approaches, and (d) visualization and surrogate approaches. Some studies have combined these approaches to develop efficient DT for solving real-world decision problems.

2.3 Cognitive Digital Twins

As extension of standard DTs, CDTs are known for many unique advantages, which include (a) self-learning capability for the effective detection and response to anomalies and disruptions, (b) situational awareness capability that enables both local and global views of system operations, (c) memory capability for holding information (e.g., working, episodic, and semantic memory) and knowledge during the autonomously control and algorithm/process improvement, and (d) short-, mid- and long-term optimization and reasoning capability (Eirinakis et al. 2020; Al Faruque et al. 2021).

According to Kobryn (2020), the original concept of DT focused on the health management of engineered safety critical systems with stringent reliability and safety requirements (e.g., airplanes). However, the scope of DT applications is rapidly expanding across the entire product/ system lifecycle. Rozanec et al. (2020) developed an actionable CDT to model a shopfloor. At the high-level, Ali et al. (2021) describe a vision of developing CDT from smart manufacturing as an extension of existing DTs with additional capabilities of communication, analytics, and intelligence through a three-layer system design that includes the access, analytics, and cognition layers. This definition of CDT complies with the dynamic data-driven simulation except the fact that it is not only the real-time data that is used to update the model, but also fusion of human knowledge is an important factor in designing, updating, optimizing and validating the model.

3 COGNITIVE DIGITAL TWIN FRAMEWORK ENABLING HUMAN-IN-THE-LOOP

To delve into the involvement of humans and their interaction within CDT, it is crucial to examine the processes involved in conducting or constructing simulation models. Lazarova-Molnar and Li (2019) revisited the conventional steps of classical simulation studies and introduced new or updated steps specifically tailored for data-driven simulation modeling, which serves as a pivotal advancement towards the realization of DTs. These steps provide valuable insights into how, where, and to what extent humans can participate and interact within the context of CDT.

An important accomplishment of data-driven simulation is its capability to (semi)automate the process of simulation modeling. However, relying solely on data poses challenges, as gathering a substantial amount of data to construct an accurate digital replica of the physical twin can be time-consuming and resource-intensive. Additionally, solely relying on data neglects a crucial source of information: the insights and expertise of individuals familiar with the systems being modeled. To address the scarcity of data, one solution is collaborative data-driven modeling, as proposed by (Niloofar and Lazarova-Molnar 2022). Another approach involves leveraging the invaluable knowledge of domain experts. By combining these strategies, we can overcome data limitations and incorporate human expertise to enhance the effectiveness and accuracy of the DT modeling process. However, as was mentioned in Section 2.2 humans need to understand the rationale behind the system's decision-making logic in order to trust the instructions. The explainability of computational methods, can help humans to understand and effectively make optimized decisions (Minh et al. 2022). Now to have an overview of the existing computational methodologies and to better understand their effect on the effectiveness of human's contribution in DTs we first present a categorization of deriving models versus their explainability and type of feedback loops. Next, we delve into CDTs presenting our framework and a case study about reliability assessment, where the model is white-box and explainable. Finally, we explain where in the framework human cognition can help in modelling processes.

3.1 Model Derivation, Explainability and Feedback Loops

Methodologies for modeling and simulation of a system can be broadly classified into three categories: knowledge-driven, data-driven and dynamic data-driven approaches. Traditionally, simulation models are constructed by domain experts, a process referred to as knowledge-driven modeling. In this approach, experts utilize their expertise and knowledge of the field to develop simulation models. On the other hand, data-driven simulation is an alternative approach where the simulation models are parameterized using data. This allows users to create and execute simulation models without the need for explicit modeling.

Dynamic data-driven simulation represents a specific category within data-driven simulation methodologies. It harnesses real-time data to detect and adapt the system model, continuously incorporating the simulation results back into the model. This iterative feedback loop ensures that the models remain accurate and timely, facilitating enhanced precision in the simulation outcomes. The process of feeding the simulation or analysis results back into the model is commonly referred to as a *feedback loop*, while the actions that subsequently impact the real system are known as *feedback actions*. These feedback loops can take on two forms: automated or interactive, depending on the involvement of human interaction.

The primary objectives of feedback loops are twofold. Firstly, they serve to validate and refine the model's accuracy through continuous evaluation against real-world data. Secondly, once the model reaches a satisfactory level of accuracy, feedback actions can be employed to enhance the overall performance of the system. Model validation ensures that the model is an accurate representation of the real system, hence, the explainability of the results might not be necessary in case of feedbacks with no human interaction (autonomous feedback). However, when the simulation model has been validated and the goal is to enhance the system's performance through actions, the absence of explanations can present challenges. Without explanations, humans may face difficulties in discerning whether a perceived mistake by the black-box model is unintentional or deliberate, aiming for long-term benefits (Zhang et al. 2022). Therefore, in scenarios where actions are taken to improve system performance, incorporating explainability becomes vital. Providing transparent justifications and insights into the decision-making process allows humans to better comprehend the reasoning behind suggested actions. This, in turn, mitigates confusion and empowers users to make informed decisions, promoting trust and facilitating effective collaboration between humans and the simulation model.

To explore the possibilities of human interaction with CDTs, it is crucial to establish a categorization framework based on how models are derived or extracted, the level of explainability they offer, and the presence of feedback mechanisms within these approaches. By categorizing the approaches, we can gain insights into the different ways humans can engage with CDTs. The categorization encompasses relevant dimensions, such as model derivation/extraction methods (e.g., knowledge-driven, data-driven, dynamic data-driven), the degree of model explainability (e.g., explainable, black-box), and the availability of feedback loops (e.g., autonomous, expert-interactive). This framework enables us to identify specific contexts and points within the CDT's lifecycle where human interaction can be integrated. It provides a structured approach to discuss and analyze how humans can contribute to model development, validation, interpretation, and decision-making processes in a CDT ecosystem. In Table 1, different approaches for model derivation, their explainability and the feedback loops are displayed:

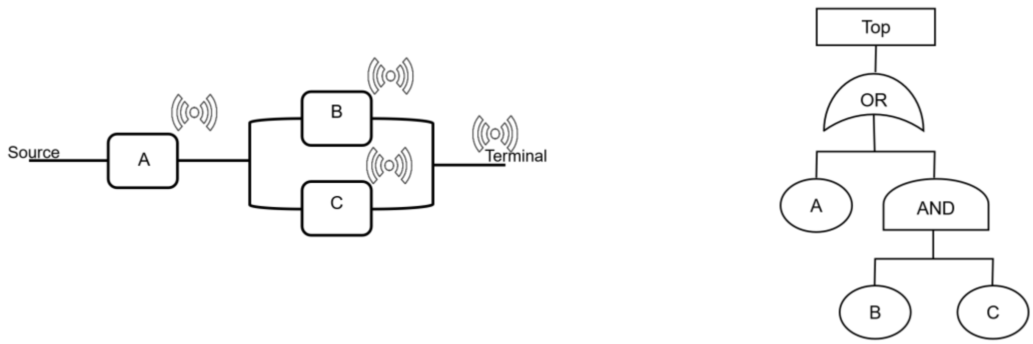
- Knowledge-driven approaches, like classical simulation and modelling, are always explainable.
    - For feedback loops humans are engaged in all the steps.
- Data-driven approaches can be both explainable and black-box. Examples of explainable data-driven approaches are classical statistical models like time series analysis, regression or principal component analysis; and data-driven modeling and simulations. Most of the Machine Learning methodologies, i.e., Neural Networks, deep-learning or reinforcement learning are black-box models.
    - Feedback loops for statistical analysis primarily focus on enhancing the model's performance, rather than improving the system itself. This involves increasing the model's accuracy and involves human participation. While it is possible to automate this process through programming, traditionally it has not been performed that way. On the other hand, feedback loops in data-driven simulations are automated and primarily aim to validate the model. In the case of black-box models where results lack explanation, human involvement often leads to a trial-and-error approach, making it predominantly an automated process.
- Dynamic data-driven approaches' steps are automated and depend on whether the data-driven modelling part is based on black-box models or not, they can be black-box or explainable.
    - Feedback loops in dynamic data-driven approaches continuously update the simulation model using real-time data from the physical domain. Hence, as long as new data is received, model gets refined and improved to reflect the behavior of the physical system. In case human cognition as well as data is involved in feeding back the analysis results to the model or in taking actions then the dynamic data-driven model of the real system is termed a CDT, otherwise it is a DT. The purpose of feedback loops here is mostly for improving the system.

Table 1: Deriving models vs. explainability.

Model Derivation		Type of Model	
		Explainable	Blackbox
Knowledge-driven		Classical Simulations	----
	Feedback loop	• None • Experts	----
Data-driven		Classical Statistical models/More recent simulations	Some Machine Learning methods
	Feedback loop	• None • Experts	• None • Autonomous
Dynamic Data-driven (Digital Twins)		Based on explainable data-driven models	Based on blackbox data-driven models
	Feedback loop	• Autonomous • Experts (Cognitive DT)	• Autonomous

3.2 Illustrative Case Study of Cognitive Digital Twins for Reliability Assessment

To exemplify the framework and illustrate the concept of a CDT, we present a case study involving reliability assessment of a system. Our objective is to develop a CDT for reliability assessment of the system displayed in Figure 2(a). This system fails when there is no flow from source to terminal. To facilitate the assessment, sensors are installed on each element of the system (A, B, and C), enabling the recording of failure and repair times. Additionally, a sensor is placed at the terminal to capture system-level failure and repair events. It is assumed that there is always a flow from the source. In this study, we employ fault trees as the chosen tool for reliability analysis. Fault trees are renowned for their explanatory power and effectiveness in analyzing the reliability of physical and cyber-physical systems. The fault tree representation of the system depicted in Figure 2(a) is shown in Figure 2(b).



(a) (b)

Figure 2: (a) Case study system and (b) its corresponding fault tree.

The workflow of building a CDT of the reliability of the system displayed in Figure 2, can be seen in Figure 3. This framework has four main components: 1) a real system that can evolve through time, 2) CDT of the real system which includes the following: 3) DT of the real system and 4) information source(s). The CDT receives real-time data (with possible time delays) from the real system (illustrated by light blue arrow), and sends back recommended actions that can be based only on the results from: the CDT (purple arrow), or directly from the experts (purple dashed arrow).

Real-time data from the sensors and the knowledge from the experts can be the sources of information. Reliability model of the real system can be built using:

- Only data: by applying data-driven modelling techniques (Niloofar and Lazarova-Molnar 2021a; Lazarova-Molnar et al. 2020a). In Figure 3 this is indicated by a yellow dotted arrow ( )
- Both data and expert knowledge: fusing data and statements from experts (hybrid modelling) is studied by Niloofar and Lazarova-Molnar (2021b), however this area has a great potential for future work. In Figure 3 this is indicated by a yellow dotted arrow and a blue arrow ( )

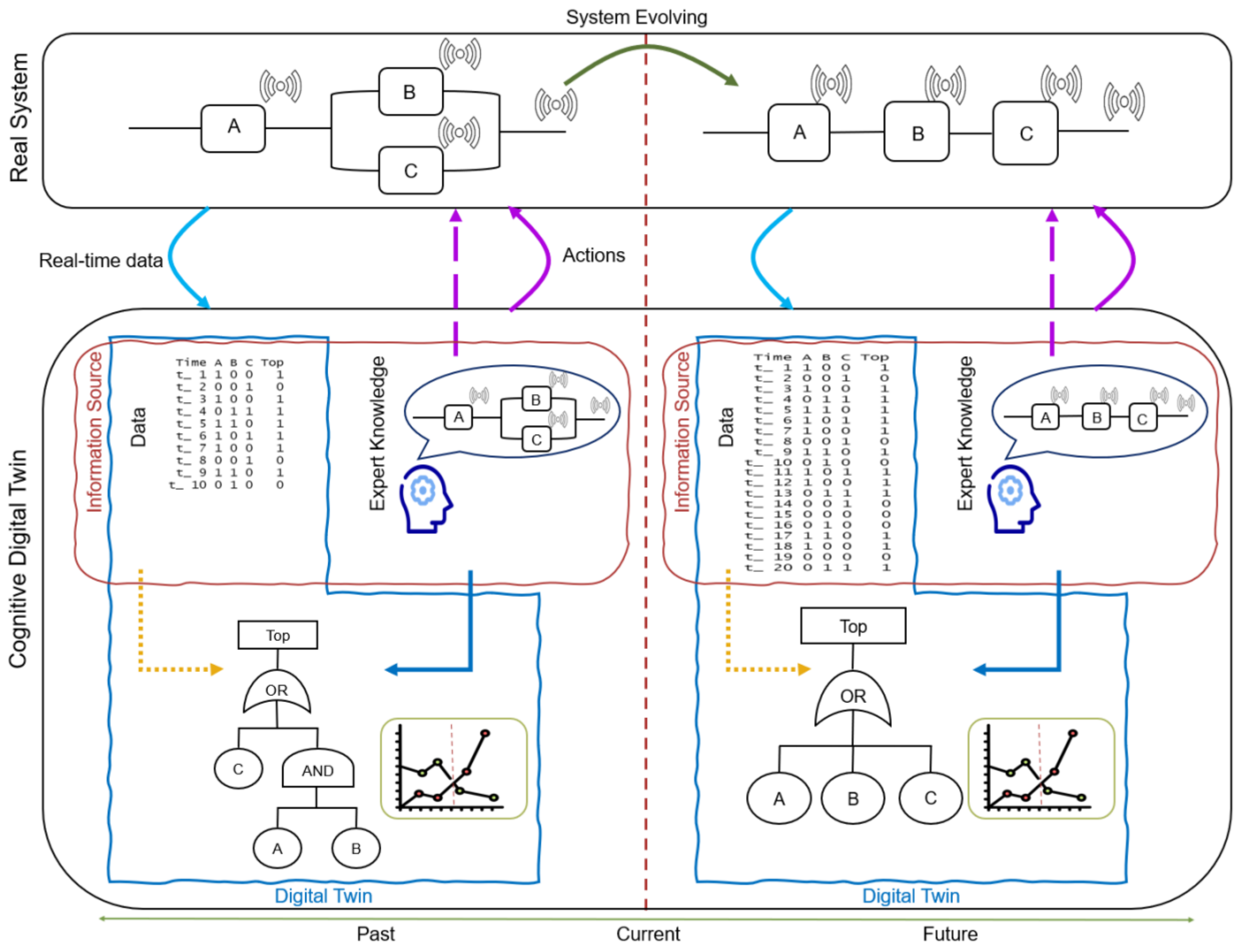


Figure 3: CDT of a system's reliability as a fault tree.

The presented framework enables the system to evolve through time by leveraging the principles of dynamic data-driven modelling and simulation. Dynamic data-driven approaches, often referred to as DTs, take on the characteristics of CDTs when human interaction is introduced at different stages or steps, such as model construction or decision management. In the next subsection we explain how human interactions can be implemented for reliability assessment purposes.

3.3 Where in the Loop?

In this study, we utilize a reliability model in the form of a fault tree to exemplify our framework. However, it is important to note that our framework is not restricted solely to reliability assessments or fault trees. It encompasses a broader scope that can be applied to diverse domains and systems. For instance, the framework can be extended to a smart factory environment, where the objective is to build a CDT using process mining techniques. Alternatively, it can be employed to optimize the energy efficiency of a robot arm or develop decision support systems for healthcare applications involving human subjects.

Assuming real-time data is received from the different component of the real system, the first step is to build a fault tree from this data that is representative of the real system so that in the next step we can apply this model to improve the system or to forecast its future behavior. Data preprocessing is crucial for model building and human knowledge can be applied here, but this is not the focus of this study. Assuming valid data, human cognitive ability can be utilized along with the data in the following stages:

a) Model Design:

For the system illustrated in Figure 2(a), there will be no current from the source to the terminal if either component A fails or both components B and C fail. Failures of these components follow a predefined distribution function, meaning that they each have a failure rate, and in case they are repairable, they also have a repair rate. Hence, to build the model, two types of information need to be extracted from the information source (data and experts): information to build the structure (what we see as the fault tree illustrated in Figure 2(b)) and the information to build the parameters (failure and repair rates). Both are explained in the following.

- Structure: statements from know-how, experience, and previous interactions with systems alike that can be either very specific, or expressed using fuzzy terminology are very useful and can be systematically applied to design the model (Niloofar and Lazarova-Molnar 2021b). Here, blacklists, whitelists and Bayesian statistics can be applied (Jensen and Nielsen 2007; Scutari 2010). For instance, assume from data we observe that when A is failed there is no current in the terminal although B and C are working (we imagine the full configuration of the real system is not known to us and we are modelling it only from the information sources that we have access to). If this is the only information we have from the data, then the system is the component A itself and we cannot proceed any further until we receive new information from data. A statement from an expert saying "I know if component B fails, but the other two components work, there will still be a current in the terminal" is a new information that has not been provided by the data yet, maybe because the component B has a very low failure rate and is highly reliable. Receiving this statement from an expert helps in designing the model, and if this is ignored, we need to wait and spend resources for a longer time to gain the same knowledge that we could have used earlier.
- Parameters: failure and repair times of the components and the system can be used to estimate the rates or even the distribution functions (Niloofar and Lazarova-Molnar 2021a; Lazarova-Molnar et al. 2020b). Meanwhile, the companies producing the components (experts) can provide more detailed information of the failure rates from their quality control processes. This information can be implemented as prior knowledge and then Bayesian probability updating methodologies can be applied to update this prior knowledge from the experts with the ones we receive from data.

b) Model Training and Testing

Data-driven approaches are biased toward data seen during the training steps. Challenging point in data-driven modelling of faults is the imbalanced proportion of classes as faults are rarely observed, especially for highly reliable systems (Niloofar and Lazarova-Molnar 2023). Hence, we are troubled with an imbalanced classification where one class of the dependent (response) variable (here, working state) outnumbers the other class (failed state) by a large proportion. There are many ways to combat this issue, where one is to accumulate more data. This, however, is not always possible and can be costly. Another approach is to manually balance the classes. One common method of doing this is to upsample/oversample the minority class or undersample the majority class using resampling (bootstrapping) techniques (Niloofar and Lazarova-Molnar 2023).

Another way to address the issue is to fuse the expert knowledge in the training step. A prior knowledge of the system can help in designing the model. However, this prior knowledge can also be valuable while the model is being trained, because observing new data might provide some insights of the real system that is only visible to human cognition and not the automated process. Again, similar to the methodologies in model design, Bayesian probability updating approaches can be applied here. When it comes to model testing, the fact that the testing data is not used for training the model and is unseen by the model introduces the possibility of inconsistent results compared to those from the training set. The validation of testing results by human knowledge can go both ways: either affirming them or raising awareness and prompting updates to human's understanding of the system.

c) System improvement

Actions result as feedback from the virtual replica or the DT of the real system. In Figure 3, these actions are shown with purple arrows (dashed and simple). Actions are taken with the aim of improving the real system performance. These actions can be based only on the results from: the CDT (purple arrow), or directly from the experts (purple dashed arrow). For our reliability analysis model, an example of an action based on the analytics results of the CDT can be to change the repair policy to increase the reliability of the system. An example of an action directly from the experts can be to change a specific supplier based on experience. As the precision of the CDT in depicting the real system increases the need for direct actions from experts declines.

4 DISCUSSION AND CONCLUSION

Cognitive digital twins add a human cognitive dimension to the concept of dynamic data-driven modelling and simulation, hence enabling humans in the loop. In this study, we presented a framework for human-in-the-loop CDT, focusing on where and how human cognition can be involved to make the simulation-related process more informed and benefitting from human/expert knowledge. Providing explanations to humans (e.g., operators, developers or users) is a key element to facilitate their understanding of the rationale behind the decisions and their rightfulness for a given context. Explanations can also provide assurance for humans to trust the autonomous adaptation by DT and the underlying DDDS (Zhang et al. 2022). Thus, we highlight explainability of models as an important feature when humans interact at different stages of a CDT. In accordance with this, we categorized model derivation approaches into: knowledge-driven, data-driven and dynamic data-driven versus their explainability and the type of feedback loop: none, autonomous (using only data), and experts.

Through an illustrative example in reliability analysis, we highlight the different steps where humans' knowledge of the system can be used to design, train/test the CDT and improve the real system. The initial effort is inspired by simulation studies which is a general process and can be generalized to many (cyber)physical systems. Hence, the reliability assessment case study is just used for illustration purposes and the presented framework can be extended to other fields like manufacturing and process mining case studies. Basically, humans in this sense leave traces about the problem they are focusing on, their understanding (or a lack of understanding) of the problem, their knowledge, and their performance in helping with the prior information they provide. However, one limitation of this work is that to benefit the most from human's cognitive abilities it is best to apply explainable models. Also, in the presented framework, fewer data remains for validation purposes, but Bayesian methods can be helpful by allowing incorporation of prior beliefs and update them based on observed data. In this context, Bayesian modeling can help estimate the posterior probabilities of different outcomes given the validation set.

One future research direction in this space will include the ability to generate new datasets from experts that could then be used to train AI/ML models for automated information integration to enhance the design and development of CDT. One concern, though, is the degree of richness of the contributed information. For experienced experts, there is a high probability that the degree of richness of the contributed information will be high compared to the degree of richness of the contributed data from less-experienced experts. Hence, another possible future research direction may include how to train AI/ML algorithms using data with different degrees of experts' richness.

REFERENCES

Agrawal, A., R. Thiel, P. Jain, V. Singh, and M. Fischer. 2023. "Digital Twin: Where Do Humans Fit In?". *Automation in Construction* 148:104749.

Ahmed, I., G. Jeon, and F. Piccialli. 2022. "From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where". *IEEE Transactions on Industrial Informatics* 18 (8):5031-5042.

Al Faruque, M. A., D. Muthirayan, S.-Y. Yu, and P. P. Khargonekar. 2021. "Cognitive Digital Twin for Manufacturing Systems". In *2021 Design, Automation & Test in Europe Conference & Exhibition,* February 1st-5th, Virtual, 440-445.

Alcaraz, C., and J. Lopez. 2022. "Digital Twin: A Comprehensive Survey of Security Threats". *IEEE Communications Surveys & Tutorials* 24 (3):1475-1503.

Ali, M. I., P. Patel, J. G. Breslin, R. Harik, and A. Sheth. 2021. "Cognitive Digital Twins for Smart Manufacturing". *IEEE Intelligent Systems* 36 (2):96-100.

Baheti, R., and H. Gill. 2011. "Cyber-Physical Systems". *The Impact of Control Technology* 12 (1):161-166.

Boschert, S., and R. Rosen. 2016. "Digital Twin—the Simulation Aspect". *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and Their Designers* :59-74.

Bouttier, F., and P. Courtier. 1999. "Data Assimilation Concepts and Methods". *Meteorological Training Course Lecture Series. ECMWF* 718:59.

Chaudhari, P., C. Gangane, and A. Lahe. 2022. "Digital Twin in Industry 4.0 a Real-Time Virtual Replica of Objects Improves Digital Health Monitoring System". In *Information Systems and Management Science: Conference Proceedings of 4th International Conference on Information Systems and Management Science 2021,* December 14th-15th, 506-517.

Drachen, A., J. Green, C. Gray, E. Harik, P. Lu, R. Sifa, and D. Klabjan. 2016. "Guns and Guardians: Comparative Cluster Analysis and Behavioral Profiling in Destiny". In *2016 IEEE Conference on Computational Intelligence and Games,* September 20th-23rd, Santorini, Greece, 1-8.

Eirinakis, P., K. Kalaboukas, S. Lounis, I. Mourtos, J. M. Rožanec, N. Stojanovic, and G. Zois. 2020. "Enhancing Cognition for Digital Twins". In *2020 IEEE International Conference on Engineering, Technology and Innovation,* June 15th-17th, Cardiff, United Kingdom, 1-7.

Faniyi, F., R. Bahsoon, and G. Theodoropoulos. 2012. "A Dynamic Data-Driven Simulation Approach for Preventing Service Level Agreement Violations in Cloud Federation". *Procedia Computer Science* 9:1167-1176.

Friederich, J., G. Lugaresi, S. Lazarova-Molnar, and A. Matta. 2022. "Process Mining for Dynamic Modeling of Smart Manufacturing Systems: Data Requirements". *Procedia CIRP, The International Academy for Production Engineering* 107:546-551.

Fujimoto, R. M., N. Celik, H. Damgacioglu, M. Hunter, D. Jin, Y.-J. Son, and J. Xu. 2016. "Dynamic Data Driven Application Systems for Smart Cities and Urban Infrastructures". In *Proceedings of the 2016 Winter Simulation Conference,* edited by Theresa M.K. Roeder, Peter I. Frazier, Robert Szechtman, Enlu Zhou, Todd Huschka, and Stephen E. Chick, 1143-1157. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

G. Goecks, V. 2020. Human-in-the-Loop Methods for Data-Driven and Reinforcement Learning Systems, Aerospace Engineering, Cornell University. https://doi.org/10.48550/arXiv.2008.13221, accessed 5th September 2023.

Gaynor, M., M. Seltzer, S. Moulton, and J. Freedman. 2005. "A Dynamic, Data-Driven, Decision Support System for Emergency Medical Services". In *Computational Science–ICCS 2005: 5th International Conference,* May 22nd-25th, GA, USA, 703-711.

Hagan, M. T., H. B. Demuth, and M. H. Beale. 2014. *Neural Network Design.* United Kingdom: PWS Publishing Co.

Hu, X. 2015. "Dynamic Data-Driven Simulation: Connecting Real-Time Data with Simulation". In *Concepts and Methodologies for Modeling and Simulation: A Tribute to Tuncer Ören,* edited by L. Yilmaz, 67-84, New York: Springer.

Jensen, F. V., and T. D. Nielsen. 2007. *Bayesian Networks and Decision Graphs.* New York: Springer.

Kobryn, P. A. 2020. "The Digital Twin Concept". In *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2019 Symposium,* September 25th-27th, North Charleston, South Carolina, 17-24.

Kumbhar, M., A. H. Ng, and S. Bandaru. 2023. "A Digital Twin Based Framework for Detection, Diagnosis, and Improvement of Throughput Bottlenecks". *Journal of manufacturing systems* 66:92-106.

Lazarova-Molnar, S., and X. Li. 2019. "Deriving Simulation Models from Data: Steps of Simulation Studies Revisited". In *Proceedings of the 2019 Winter Simulation Conference,* edited by Navonil Mustafee, Ki-Hwan G. Bae, Sanja Lazarova-Molnar, Markus Rabe, C. Szabo, Peter Haas, and Young-Jun Son, 2771-2782. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lazarova-Molnar, S., P. Niloofar, and G. K. Barta. 2020a. "Automating Reliability Analysis: Data-Driven Learning and Analysis of Multi-State Fault Trees". In *30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference,* November 1st-5th, Venice, Italy, 1805-1812.

Lazarova-Molnar, S., P. Niloofar, and G. K. Barta. 2020b. "Automating Reliability Analysis: Data-Driven Learning and Analysis of Multi-State Fault Trees". In *30th European Safety and Reliability Conference and15th Probabilistic Safety Assessment and Management Conference,* November 1st-5th, Venice, Italy, 1805-1812.

Minh, D., H. X. Wang, Y. F. Li, and T. N. Nguyen. 2022. "Explainable Artificial Intelligence: A Comprehensive Review". *Artificial Intelligence Review*:1-66.

Mourtzis, D. 2021. *Design and Operation of Production Networks for Mass Personalization in the Era of Cloud Technology.* The Netherlands: Elsevier.

Niloofar, P., and S. Lazarova-Molnar. 2021a. "Data-Driven Modelling of Repairable Fault Trees from Time Series Data with Missing Information". In *Proceedings of the 2021 Winter Simulation Conference*, edited by Sojung Kim, Ben Feng, Katy Smith, Sara Masoud, Zeyu Zheng, Claudia Szabo, and Margaret Loper, 1-12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Niloofar, P., and S. Lazarova-Molnar. 2021b. "Fusion of Data and Expert Knowledge for Fault Tree Reliability Analysis of Cyber-Physical Systems". In *2021 5th International Conference on System Reliability and Safety,* November 24th-26th, Palermo, Italy, 92-97.

Niloofar, P., and S. Lazarova-Molnar. 2022. "Collaborative Data-Driven Reliability Analysis of Multi-State Fault Trees". *Proceedings of the Institution of Mechanical Engineers Part O: Journal of Risk and Reliability*: 1748006X221076290.

Niloofar, P., and S. Lazarova-Molnar. 2023. "Data-Driven Extraction and Analysis of Repairable Fault Trees from Time Series Data". *Expert Systems with Applications* 215:119345.

Radanliev, P., D. De Roure, R. Nicolescu, M. Huth, and O. Santos. 2022. "Digital Twins: Artificial Intelligence and the Iot Cyber-Physical Systems in Industry 4.0". *International Journal of Intelligent Robotics and Applications* 6 (1):171-185.

Rokka Chhetri, S., and M. A. Al Faruque. 2020. "Dynamic Data-Driven Digital Twin Modeling". *Data-Driven Modeling of Cyber-Physical Systems using Side-Channel Analysis*:129-153.

Rozanec, J. M., J. Lu, A. Kosmerlj, K. Kenda, K. Dimitris, V. Jovanoski, J. Rupnik, M. Karlovcec, and B. Fortuna. 2020. "Towards Actionable Cognitive Digital Twins for Manufacturing". In *International Workshop On Semantic Digital Twins*, June 3rd, Heraklion, Greece, 1-12.

Scutari, M. 2010. "Learning Bayesian Networks with the Bnlearn R Package". *Journal of Statistical Software* 35 (3):1-22.

Zhang, N., R. Bahsoon, N. Tziritas, and G. Theodoropoulos. 2022. Explainable Human-in-the-Loop Dynamic Data-Driven Digital Twins. https://arxiv.org/abs/2207.09106, accessed 5th September 2023.

Zheng, X., J. Lu, and D. Kiritsis. 2022. "The Emergence of Cognitive Digital Twin: Vision, Challenges and Opportunities". *International Journal of Production Research* 60 (24):7610-7632.

AUTHOR BIOGRAPHIES

PARISA NILOOFAR is an Assistant Professor with the Faculty of Engineering at the University of Southern Denmark. Her current research interests include Bayesian networks, simulation and modelling and reliability modeling. Parisa Niloofar obtained her PhD in Statistics, in 2013, specializing in the area of Graphical modeling. Her current research on (Hybrid)data-driven modelling of cyber-physical systems has added value to the literature. Her email address is parni@mmmi.sdu.dk.

SANJA LAZAROVA-MOLNAR is a Professor at the Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology. She is also a Professor at the University of Southern Denmark, where she leads the research group Modelling, Simulation and Data Analytics. She is a Senior Member of The Institute of Electrical and Electronics Engineers (IEEE), and currently serving as Director-at-Large on the Board of Directors of The Society for Modeling and Simulation International (SCS). Furthermore, she is Chair of IEEE Denmark and Vice-Chair of IEEE Denmark Women in Engineering Affinity Group. Her email address is sanja.lazarova-molnar@kit.edu.

OLUFEMI A. OMITAOMU is a Senior R&D Staff and Group Leader in Computational Urban Sciences group at Oak Ridge National Laboratory, Tennessee, U.S.A. He is also an Adjunct Professor in the Department of Industrial and Systems Engineering at the University of Tennessee, Knoxville, Tennessee, U.S.A. His research expertise includes urban artificial intelligence, urban digital twin, and anomaly detection in complex system. He received his Ph.D. in Industrial Engineering from the University of Tennessee, Knoxville. He is a senior member of IEEE and IISE; member of INFORMS, ACM, and AAAI. His email address is omitaomuoa@ornl.gov.

HAOWEN XU is a research scientist in the Computational Urban Sciences group. He has a research background in urban informatics, visual analytics, and environmental science. His research interests include developing cyberinfrastructure, digital twins, visual dashboards, mobile apps, and cyber-physical systems to facilitate the management, experimental research, and data and model-driven analytics in the urban science sector. He has authored and co-authored over 15 journal publications, including several premier journals such as IEEE T-ITS, STOTEN, EMS, and Journal of Hydrology, as well as received best poster awards from IEEE-MASS and IBPSA. His email address is xuh4@ornl.gov.

XUEPING LI is a Professor of Industrial and Systems Engineering and the Director of the Ideation Laboratory (iLab) and co-Director of the Health Innovation Technology and Simulation (HITS) Lab at the University of Tennessee - Knoxville. He holds a Ph.D. from Arizona State University. His research areas include complex system modeling, simulation, and optimization, with broad applications in supply chain logistics, healthcare, and energy systems. He is an IISE Fellow and a member of IEEE, ASEE and INFORMS. His e-mail address is Xueping.Li@utk.edu.