Author pre-print.
Publication accepted for the 1st International Conference on Engineering Digital Twins. Official publication may differ.

AI Simulation by Digital Twins

Systematic Survey of the State of the Art and a Reference Framework

Xiaoran Liu
McMaster University
Hamilton, Canada
liu2706@mcmaster.ca

Istvan David
McMaster University
Hamilton, Canada
istvan.david@mcmaster.ca

ABSTRACT

Insufficient data volume and quality are particularly pressing challenges in the adoption of modern subsymbolic AI. To alleviate these challenges, AI simulation recommends developing virtual training environments in which AI agents can be safely and efficiently developed. Digital twins open new avenues in AI simulation, as these high-fidelity virtual replicas of physical systems are equipped with state-of-the-art simulators and the ability to further interact with the physical system for additional data collection. In this paper, we report on our systematic survey of digital twin-enabled AI simulation. By analyzing 22 primary studies, we identify technological trends and derive a reference framework to situate digital twins and AI components. Finally, we identify challenges and research opportunities for prospective researchers.

CCS CONCEPTS

• **General and reference** → **Surveys and overviews**; • **Computing methodologies** → *Learning settings*.

KEYWORDS

AI, artificial intelligence, data science, deep neural networks, digital twins, lifecycle model, machine learning, neural networks, reinforcement learning, SLR, subsymbolic AI, survey, training

ACM Reference Format:
Xiaoran Liu and Istvan David. 2024. AI Simulation by Digital Twins: Systematic Survey of the State of the Art and a Reference Framework. In *Proceedings of International Conference on Engineering Digital Twins (EDTConf'24)*. ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Modern artificial intelligence (AI) is enabled by massive volumes of data processed by powerful computational methods [84]. This is a stark contrast with traditional AI, which is supported by symbolic methods and logic [69]. The volume and quality of available data to train AI is the cornerstone of success in modern AI. However, accessing and harvesting real-world data is a substantial barrier due to its scarcity, cost, or difficult accessibility, hindering the development of precise and resilient AI models. For example, in manufacturing, proprietary data, data silos, and sensitive operational procedures complicate the acquisition of data [43]. Data-related barriers, in turn, limit the applicability of otherwise powerful AI methods.

AI simulation is a prime candidate for alleviating these problems. As defined by Gartner recently, AI simulation is the technique of "*the combined application of AI and simulation technologies to jointly develop AI agents and the simulated environments in which they can be trained, tested and sometimes deployed. It includes both the use of AI to make simulations more efficient and useful, and the use of a wide range of simulation models to develop more versatile and adaptive AI systems*" [47]. After modeling the phenomenon or system at hand, a simulation of the model computes the dynamic input/output behavior [77], representative of the system. A simulation produces data, called the simulation trace, that represents the behavior of the simulated system over time. These traces can be used as training data for AI agents, assuming that the simulation is a faithful, valid and detailed representation of the modeled system, and that the simulation can still be executed efficiently and in a timely manner.

With the emergence of digital twins (DT) [54], the quality attributes of simulators have improved as well. Simulators are first-class components of DTs [36] and enablers of sophisticated services, e.g., real-time adaptation [73], predictive analytics [62], and process control in manufacturing [28]. These advanced services require well-performing and high-fidelity simulators—the types of simulators that align well with the goals of AI simulation.

A recent interview study on DTs with nineteen academic and industry participants by Muctadir et al. [58] mentions that "*machine learning and reinforcement learning could possibly be combined with DTs in the future, to help to learn about complex systems (i.e., safety-critical systems) in a virtual environment, when this is difficult to do on the real-world system.*" Similar ambitions have been identified by Mihai et al. [56] as future prospects of DTs. Indeed, the improvements in simulator engineering that have been driven by DTs, are generating interest in DTs for AI simulation. It is plausible to anticipate that the next generation of AI simulation techniques will be heavily influenced by the further advancements of DT technology [51, 66]. Therefore, it is important to understand the state of affairs in digital twinning for AI simulation purposes, prepare for the related challenges, and set targeted research agendas.

This work marks a step towards converging AI simulation and DT technology. We review the state of the art on AI simulation by DTs, derive a framework, identify trends in system organization, AI flavors, and simulation, and outline future avenues of research.

Context and scope. In this work, we focus on **AI simulation *by* digital twins**. We acknowledge the utility of the other direction, i.e., simulators of DTs being enabled by AI [55]; however, we consider such works outside the scope of the current study.

EDTConf'24, September 23-24, 2024, Linz, Austria

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

Contributions. The contributions of this work are the following.

- We design, conduct, and report a **systematic survey** of the state of the art in AI simulation by digital twins.
- Based on the results of our survey, we derive a **conceptual reference framework** to integrate (i) digital twins and (ii) AI components for the purpose of AI simulation.
- We identify **technological trends, key challenges, and research opportunities** in AI simulation by digital twins for prospective researchers.

Replicability. For independent verification, we publish a replication package containing the data and analysis scripts of our study.[1]

2 BACKGROUND AND RELATED WORK

We now review the background concepts and related works.

2.1 Data challenges in AI training

The data-related challenges of modern AI are well-documented. In their review of fifteen key challenges in AI, Hagendorff and Wezel [49, Challenge 13] identify the problem of the acute scarcity of labels despite labeled data being a hard precondition to many AI systems. Obtaining data of sufficient quantity and quality can be challenging. Data quality directly affects the effectiveness of model training. Common data quality issues include missing data, inconsistencies, duplications, and noise. Obtaining high-quality data typically requires data cleaning and pre-processing. Hagendorff and Wezel [49] consider these challenges ephemeral, i.e., technological advancement is expected to solve these challenges in the short run.

Data enhancement techniques—such as rotating, flipping, scaling—can be used to generate more synthetic data to extend the training dataset [76]. The improvement of data quality is mainly realized through data cleaning and preprocessing, including methods such as removing duplicates, handling missing values, and eliminating noise [46]. In addition, automated tools and algorithms can be utilized to assess and monitor data quality, which can detect and fix problems in time [40]. Regarding the optimization of data annotation, in addition to the use of semi-automated tools and algorithms to assist manual annotation methods [42, 60], crowdsourcing platforms (e.g., Amazon's Mechanical Turk [23]) can also be used for large-scale collaborative annotation. Other alternatives are being actively researched currently, such as assisted human labeling [24, 37], and labeling with ChatGPT [61, 71, 72].

In this work, we draw attention to the emerging topic of AI simulation as a potential solution to these problems.

2.2 Simulation

Simulators are programs that encode the probabilistic mechanism that represents the real phenomenon and enact this probabilistic mechanism over a sufficiently long period of time to produce simulation traces describing the real system [82].

From the '60s, computer simulation was employed in select domains by few experts until, in the '80s, it became a key enabler in solving complex engineering problems. In the past decade, advancements in digital technology shifted the typical role of simulators again, this time down to the operational phase of systems [30]. As a prime exemplification of this trend, simulators are first-class components of DTs [36] and enablers of the sophisticated features and services DTs provide, e.g., providing a learning environment for training purposes of human and computer agents [52].

At the core of the simulator, the physical asset is represented by a model, from which complex algorithms calculate the metrics of interest. This model captures the essential properties of the simulated asset in appropriate detail to consider the results of the simulation representative. The execution of a simulation produces a *simulation trace*, that represents the behavior of the simulated system over time [67]. These simulation traces are the *data* that can be used to train and tune AI agents.

2.3 Related work

Although our work marks the first survey on AI simulation by DTs, the benefits of combining DTs and AI have been recognized before. In their review of applying AI in Industry 4.0, Baduge et al. [26] identify the integration potential of AI with DTs to enhance the precision of DT models and iteratively refine these models using continuously gathered data. Emmert-Streib [41] investigate techniques that combine AI and DTs, and identify "generative modeling", roughly analogous to AI simulation, as an opportunity with elevated potential. This underlines the importance of our work.

A related body of knowledge is the one dedicated to the opposite direction of support between AI and DTs, i.e., AI for DTs. Yitmen et al. [81] use AI to improve the creation of DT simulation models by simplifying their structure and functionality. David et al. [35] propose a method for inferring DT simulation models through deep reinforcement learning. Their evaluation shows that DTs augmented with reinforcement learning facilities can efficiently learn from the right signals. Neethirajan [59] investigates the use cases and potentials of generative adversarial networks in the livestock industry to generate simulation data for the development of DTs.

Multiple secondary studies on DT practices relate to our work. Muctadir et al. [58] conduct an interview study focusing on the trends in DT development, maintenance, and operation. Their interviews with 19 experts from industry and academia reveal problematic areas, such as the lack of uniform definitions, tools, techniques, and methodologies, and call for the adoption of more rigorous software engineering practices in support of the DTs' lifecycles. Our study corroborates these findings at many points, as explained later. Mihai et al. [56] survey the enabling technologies, trends, and future prospects of DTs. A key technological prospect they identify is the strong convergence of AI and DTs. Their leads are mostly complementary to our focus as they sample techniques in which machine learning "represents the foundation of a DT". The broader definition of AI simulation is inclusive of this direction as well.

3 STUDY DESIGN

In this section, we design a study to systematically survey the literature on digital twins for AI simulation.

3.1 Goal and research questions

The goal of this study is to analyze the use-cases, technical characteristics, and context of digital twins, used for AI simulation. To this end, we formulate the following research questions.

[1] https://zenodo.org/doi/10.5281/zenodo.13293237

RQ1. *In what **domains and problems** are digital twins used to support AI simulation?*
By addressing this research question, we aim to understand the motivation for employing digital twins for AI simulation.

RQ2. *What are the **technical characteristics of digital twins** used in AI simulation?*
We aim to understand which digital twin styles are used (e.g., twin, shadow, human-in-the-loop), how DTs are architected, and which M&S formalisms are used for AI simulation.

RQ3. *Which **AI/ML techniques** are supported by digital twin-enabled AI simulation?*
In response to this research question, we attempt to categorize and analyze various Artificial Intelligence (AI) and Machine Learning (ML) techniques used in the research, identify the specific algorithms and methods used, and activities along the overall ML development process that are supported by AI simulation (e.g., training, validation, etc.).

RQ4. *What **lifecycle models** are used in support of digital twin-enabled AI simulation?*
By addressing this research question, our goal is to understand the lifecycle of AI simulation with a particular focus on the maintenance of simulators, and whether simulated data is validated in a specific step(s) along the lifecycle.

RQ5. *What are the **open challenges** in DT-enabled AI simulation?*
We aim to identify challenges to which researchers in the DT and model-driven engineering communities can contribute.

3.2 Search and selection

To identify relevant studies, we employ a combination of automated search, manual search and snowballing. In the following, we elaborate on this process. Tab. 1 reports the relevant figures.

3.2.1 Automated search. We construct our initial search string from the topic of interest ("AI simulation") and its explanation ("development or training of AI or ML by digital twins" [47]):

```
("AI simulation") OR
(("digital twin*") AND
 ("train*" OR "develop*") AND
 ("AI" OR "artificial intelligence" OR "ML" OR "machine learning"))
```

Experimentation with different variations of the search string yields a negligible amount of true positives and a substantial amount of false positives. This is likely because AI simulation is a new, emergent field (explains the lack of results from the second, detailed part of the search string), and the term "AI Simulation" might not be widely adopted in academic works just yet (explains the lack of results from the first part of the search string).

To mitigate false positives, we use a high-level search string that finds AI simulation studies explicitly labeled as such; and augment the initial result set by manual search (Sec. 3.2.2) and expert knowledge (Sec. 3.2.3). We use the following search string to scan Scopus, Web of Science, IEEE Xplore, and ACM Digital Library:

```
("AI simulation") AND
("digital twin*" OR "digital shadow*")
```

[2]Of the 499 backward references, 8 were selected for inspection by interpreting their citation context in the data extraction phase.
[3]Of the 618 citations, 192 were selected via a citation-based preliminary screening.
[4]κ calculated from the 8+192=200 studies screened by both authors.
[5]After clustering, only 19 newly included studies remain in this phase.

Table 1: Statistics of the search and selection rounds

Initial search	**All**	**Excluded**	**Included**	κ
Automated search	4	2	2	1.000
Manual search			4	
Expert knowledge			4	
Subtotal			**10**	
Snowballing 1	**All**	**Excluded**	**Included**	κ
Backward	558	553	5 (0.90%)	
Forward	90	86	4 (4.44%)	
Subtotal	648	639	**9 (1.39%)**	0.840
QA 1	**All**	**Excluded**	**Retained**	
Subtotal	19	7	→**12** (1.82%)	
Snowballing 2	**All**	**Excluded**	**Included**	κ
Backward[2]	499	494	5 (1.00%)	
Forward[3]	618	601	17 (2.75%)	
Subtotal[4]	1 117	1 095	**22 (1.97%)**	0.662
QA 2	**All**	**Excluded**	**Retained**	
Subtotal[5]	19	9	→**10** (0.89%)	
Final	1 775	1 753	→**22** (1.24%)	

As reported in Tab. 1, the automated search finds 4 primary studies. The search strings yields 4 primary studies on Scopus, 5 on Web of Science, of which 6 remain after duplicate removal, and 4 after removing two patents. After screening, we tentatively retain **2 primary studies**, subject to further quality assessment.

3.2.2 Manual search. Based on our expert knowledge, we identify key venues (conferences and journals) and search for potentially relevant studies in the past five years (2019–2024). Specifically,

- we scan top AI conferences for studies on DTs (IJCAI, ICML, NeurIPS, AAAI, ICLR)[6]; and
- we scan top conferences and journals in computing, software, and systems, related to DTs for studies about AI or ML (MODELS; SoSyM, JSS, IEEE Software)[7];

When choosing AI venues, we consider the currently top (CORE-A*) conferences in AI. Considering the conference-focused publication trends in AI, we deem this sample sufficient for our purposes. When choosing DT venues, we rely on our expert knowledge and the publication venues of the community-curated list of key publications by the Engineering Digital Twins (EDT) Community [39]. The selected ones are flagship publication outlets for the DT community (including a CORE-A conference and multiple journals).

When scanning conferences, we also consider their satellite events, such as workshops. We scan the past five editions of each conference, given that AI simulation is a relatively new concept that appeared in Gartner's glossary in 2023 for the first time.

We select potentially relevant studies by checking them against the exclusion criteria (Sec. 3.2.5) using adaptive reading depth [64]. That is, we first check the title and abstract of the study, and if deemed relevant, we assess whether the study merits consideration to be included by processing the full text. We tentatively include **4 primary studies**, subject to further quality assessment.

[6]https://ijcai.org, https://icml.cc, https://neurips.cc, https://aaai.org, https://iclr.cc
[7]http://modelsconference.org; https://sosym.org, https://sciencedirect.com/journal/journal-of-systems-and-software, https://computer.org/csdl/magazine/so

3.2.3 Expert knowledge. To round out the initial phase of the search, we add studies that we are familiar with and have not been found by the search string or manual search. Similar to the manual search phase, we again select relevant studies by checking them against the exclusion criteria (Sec. 3.2.5) using adaptive reading depth [64] (first checking the title and abstract of the study, and if relevant, scanning the full-text for details). We tentatively include **4 primary studies**, subject to further quality assessment.

After this phase, the initial set consists of **10 primary studies**, subject to further quality assessment.

3.2.4 Snowballing. We apply two rounds of backward and forward snowballing to enrich the corpus. The studies we include in the second round of snowballing align well with the information from already included studies with minimal new or unexpected findings. Thus, we decide to conclude snowballing after two rounds.

For backward snowballing, we extract references from primary studies manually. For forward snowballing, we follow the recommendations of Wohlin et al. [80] and extract references from Google Scholar. We automate this step through Publish or Perish [50].

In the first snowballing round, we apply an exhaustive snowballing strategy in which both researchers screen every study. We observe a high kappa of 0.84 ("almost perfect agreement"). We assert that the level of agreement allows for a more rapid snowballing style in subsequent snowballing rounds. We tentatively include **9 primary studies**, subject to further quality assessment.

In the second snowballing round, we apply a more agile snowballing strategy. In backward snowballing, we follow Wohlin [79] and mark potentially relevant references as we examine studies in the data extraction phase. Excluding duplicates, we eventually mark 8 references of the total 499 as relevant. These 8 references are screened by both researchers and 5 are included. In forward snowballing, one researcher conducts a preliminary screening in which clearly irrelevant studies are excluded. Of the total 618 studies, 192 are retained for screening by both reviewers. We observe a kappa of 0.662 ("substantial agreement"), which we find satisfactory considering that we mitigated the threat of kappa inflation by excluding a significant number of irrelevant studies. We tentatively include **22 primary studies**, of which 3 are from the same group of authors we already have in our corpus, and on the same topic. Thus, we apply clustering and nominate one study from each cluster as the representative primary study. Eventually, we consider **19 primary studies** after this round, subject to further quality assessment.

After each snowballing phase, newly considered publications go through the same evaluation process as prior studies.

3.2.5 Exclusion criteria. We use the following exclusion criteria to filter works that are not relevant to our study. We use these criteria in the manual search and the snowballing rounds. A study is excluded if it meets at least one exclusion criterion. Exclusion criteria are evaluated based on the *full reference* (title, authors, venue) and the *abstract*, by both authors.

E1. No or unclear DT; or the DT is not used for AI simulation.
E2. No or unclear AI/ML technique.
E3. Not DT for AI – either no link between DT and AI, or the opposite direction (AI for DT).
E4. Other: off-topic; not English; not publicly available; secondary or tertiary studies; full proceedings; short papers (< 5 pages).

3.2.6 Quality assessment. In accordance with the guidelines of Kitchenham and Charters [53], we define a checklist to assess the quality of the corpus. Quality criteria are derived from the research questions. Each question is answered by "yes" (1 point), "partially" (0.5 points), or "no" (0 points), based on the full text. To retain a study, we require a score of at least 2/4 points (50%).

Q1. Digital twinning scenario clearly described.
Q2. Simulation method clearly described.
Q3. AI/ML method clearly identified.
Q4. Acknowledges limitations and challenges.

After the first round of snowballing, we assess studies included in the initial round and first snowballing round. Of the total 19 candidate studies, we exclude 7 and retain **12 primary studies**. After the second round of snowballing, we assess studies included in the second snowballing round. Of the total 19 candidate studies, we exclude 9 and retain **10 primary studies**.

Eventually, our corpus consists of **22 primary studies**.

3.2.7 Threats to validity and quality assessment. Here, we identify the key threats to the validity, elaborate on the mitigation strategies we applied, and assess the quality of the study.

External validity. External validity is concerned with the generalizability of results. Our work is focusing on AI simulation through digital twins and therefore, some takeaways cannot be safely extrapolated to AI simulation in general. We mitigated such threats by being explicit about digital twins and digital shadows in our search strings and the manual search.

Construct validity. Our observations are artifacts of the sampled studies. Potential selection bias and missed publications may impact our observations and threaten the construct validity of this study. To mitigate this threat, we employed a diverse selection process (automated search, manual search, and input from expert knowledge), as well as snowballing until saturation [48].

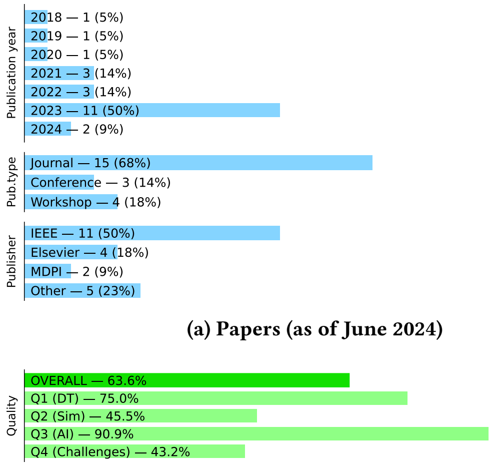
Internal validity. We may have missed some works due to terminology. "AI Simulation" is an emerging concept. Nonetheless, our scope, which is specific to digital twins, narrows our search and provides us with a sufficiently descriptive search term that finds relevant studies. Selection bias may be present in our work due to applying only two rounds of snowballing. However, the low inclusion rate of 0.89% at the end of the snowballing phase suggests that additional snowballing would yield minimal value.

Study quality. Our work scores 72.7% in the particularly rigorous quality checklist of Petersen et al. [65]. (Need for review: 1 point; search strategy: 2 points; evaluation of the search: 2 points; extraction and classification: 2 points; study validity: 1 point.) This quality score is *significantly* higher than the typical values in software engineering. (Petersen et al. [65] reports a median of 33%, with only 25% of their sampled studies having a quality score of above 40%.) We consider our study of exceptionally **high quality**.

3.3 Publication trends and quality

Fig. 1 reports the basic mappings of publication trends in our corpus.

The number of publications shows an increasing trend, with a clear increase in publication output in 2023, constituting half of the



(a) Papers (as of June 2024)

(b) Quality scores

Figure 1: Publication trends

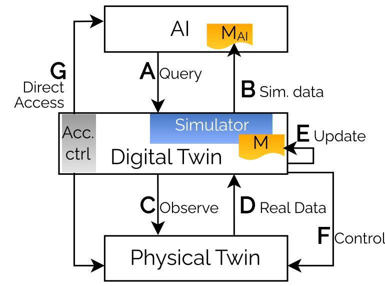


Figure 2: The DT4AI framework

corpus. The relatively low number of studies in 2024 is partly due to our study being conducted in Q2/2024 and possibly due to seasonal variations in area-specific publication trends (e.g., timing of conferences). We observe an increasing interest in AI simulation. About 68% of the sampled studies are journal articles, suggesting mature research our analysis draws from. The high quality of the corpus is further demonstrated by the high number of top publishers.

With that said, the *reporting* quality of publications (Fig. 1b) is moderate, scoring around 63.6% overall. This is score is due to the largely ignored details about simulation formalisms and methods (Q2, 45.5%) and the lack of broad vision about challenges and research recommendations (Q4, 43.2%). However, AI aspects are particularly well-documented (Q3, 90.9%), and the technical details of digital twinning are sufficiently presented (Q1, 75%).

Overall, we judge the corpus to be in a good shape to allow for sound conclusions about digital twinning and AI within reasonable boundaries of validity, but we anticipate limited leads about simulation formalisms and methods.

4 THE DT4AI FRAMEWORK

To integrate DTs, AI, and simulation, we construct a conceptual reference framework from the sampled primary studies. We rely on a mixed sample- and case-based generalization [78]. This approach is particularly useful when constructing middle-range theories that balance generality with practicality, such as engineering sciences. In Sec. 3, we sampled a statistically adequate corpus. Subsequently, we decompose each study individually into architectural units as architectural abstractions allow for better judging of similarity between cases [78]. Finally, we identify recurring patterns.

The resulting **DT4AI framework** is shown in Fig. 2 and defines the following concepts.

AI training. Interplay between the *AI* and the *Digital Twin.*

- **A:** ***Query.*** Represents the request for data issued by the *AI* component to the *Digital Twin.* As shown in Tab. 2, the *Query* can be either explicit (the *AI* agent actively pulling data) or implicit (the *Digital Twin* pushing data).
- **B:** ***Simulated data.*** The result of a simulation is a simulation trace, i.e., the data the *AI* component receives in response to its *Query*. The *Digital Twin* is equipped with a model (or set of models) *M*, which serves the input to the *Simulator*. The **A-B** training cycle can take either a batch or live format. In the former, the *Trace volume* is big data; in the latter, the trace consists of small pieces of data (small data).

Data collection. Interplay between the *Digital* and the *Physical Twin.*

- **C:** ***Observe.*** The *Digital Twin* is connected to the *Physical Twin* through the usual data link and is able to passively observe or actively interrogate the *Physical Twin* (Tab. 2).
- **D:** ***Real data.*** Represents the data collected from the *Physical Twin.* Depending on the type of the *Observation*, *Data* might be of low context, i.e., large volumes with low information value [34] (in case of passive observation); or of high context, i.e., smaller volumes of data in response to active experimentation. In situations when the *Digital Twin* gets detached from the *Physical Twin*, e.g., due to the retirement of the latter, data can be historical as well. As shown in Tab. 2, the **C-D** Observe/Data cycle can be automated (scheduled by the *Digital Twin*) or on-demand (based on the requests of the *AI* or human operators).
- **E:** ***Update.*** After collecting data from the *Physical Twin*, the model (*M*) of the *Digital Twin* needs to be updated in order to reflect the new data in simulations and transitively. This *Update* can be achieved in a synchronous (blocking behavior but easier implementation) or asynchronous fashion (non-blocking behavior but more complex implementation, e.g., timeout and request obsolescence management).

Control and access control. Interplays between the *Digital Twin* and the *Physical Twin.*

- **F:** ***Control.*** As customary, the *Digital Twin* can control the *Physical Twin* through the usual control links. As listed in Tab. 2, control can be achieved *in-place*, e.g., a learned policy on the digital side can govern the behavior of the physical system; or (parts of) the control logic can be *deployed* onto the *Physical Twin* for local control.
- **G:** ***Access control.*** The *AI* component might interact with the *Physical Twin* without the participation of the simulation facilities of the *Digital Twin.* In these situations, the *Digital*

Twin provides *Access control* to the *Physical Twin*. We do not consider this case alone AI simulation by a digital twin; however, as discussed later, direct control with the *Physical Twin* can be used in combination with *AI* simulation, e.g., to adapt the trained agent to a physical setting.

The DT4AI framework enables the systematic comparison and discussion of different AI simulation approaches that enabled by digital twins. In Sec. 5, we organize evidence along the framework by instantiating it for the different flavors of architectures, AI methods, and simulation lifecycles we found in the state of the art.

Table 2: Variation points in the DT4AI framework

	AI training
A Query	{Implicit, Explicit}
B Sim. data volume	{Big data , Small data}
A-B Training fashion	{Batch, Live}
	Data collection
C Observe	{Passive observation, Active experimentation}
D Data	{Stationary historical data, Low-context data update, High-context data update}
C-D Observe/Data trigger	{Automated, On-demand}
E Update synchronicity	{Synchronous, Asynchronous}
	Control
F Control	{In-place control, Deploy-and-Control}

Table 3: Application domains

Domain	#Studies	Studies
Networks	11 (50.0%)	
↳ Wireless	8 (36.4%)	[2, 4, 5, 8, 9, 14, 18, 22]
↳ Edge	2 (9.1%)	[6, 10]
↳ General	1 (4.5%)	[7]
Robotics and AVs	6 (27.3%)	[11, 13, 15, 17, 20, 21]
Manufacturing	2 (9.1%)	[1, 19]
Energy	1 (4.5%)	[16]
Urban	1 (4.5%)	[12]
Agriculture	1 (4.5%)	[3]

Table 4: Architectural choices

Architecture	#Studies	Studies
Digital twin	19 (86.4%)	
↳ Autonomous	16 (72.7%)	[1, 2, 4–8, 10, 11, 14–16, 18, 19, 21, 22]
↳ Human-suprvised	2 (9.1%)	[17, 20]
↳ Human-actuated	1 (4.5%)	[3]
Digital shadow	2 (9.1%)	[9, 13]
Digital model	1 (4.5%)	[12]
Policy DT	1 (4.5%)	[16]

5 STATE OF THE ART

In this section, we report the key results of our empirical inquiry into the state of the art of AI simulation by DTs. Readers are referred to the replication package for the complete data extraction sheet.

5.1 Domains and problems (RQ1)

As shown in Tab. 3, half of the primary studies we sampled focus on a network problem. Wireless networks (8 of 22 – 36.4%) are the most represented, typically focusing on various optimization tasks by machine learning, such as optimization of resource allocation in 5G+ networks [22] and edge computing [5]. Robotics, including the management of automated vehicles (AVs) accounts for 6 of 22 (27.3%) cases, with typical examples of training AI models for the control of ordinary [11] and underwater [20] robot arms, and controlling the flocking motion of unmanned aerial vehicles (UAVs) [13].

The common trait of addressed problems is their high complexity (e.g., control in dense fluid dynamics [20]) and sparse data from real observations (e.g., in slowly changing settings of agriculture [3]).

RQ1: Domains and problems

AI simulation is primarily used in problems with *high complexity* and *sparse or inaccessible data* from real observations. Networks and robotics are the most prominent adoption domains, accounting for over three-quarters of sampled studies.

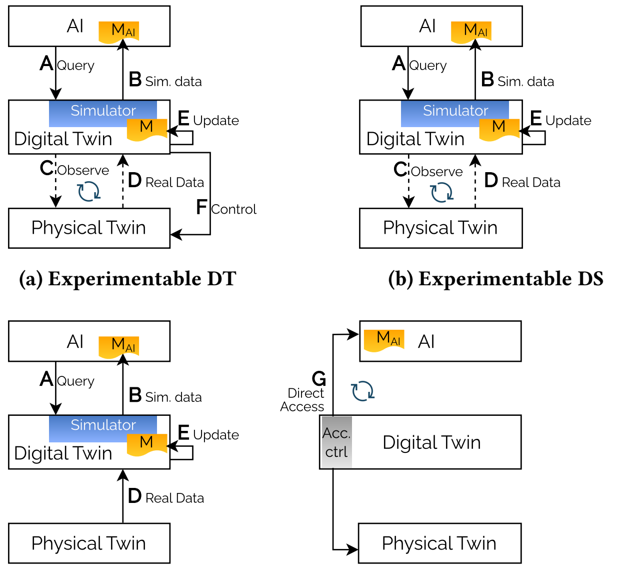
5.2 Technical characteristics of DTs (RQ2)

To analyze the technical characteristics of digital twins used in AI simulation, we rely on the superset of taxonomies by Kritzinger et al. [54] and David and Bork [34] as our initial values. The former defines the foundational classes of digital model, digital shadow, and digital twin; the latter extends this classification by defining human-supervised and human-actuated digital twins situated between fully autonomous digital twins and non-autonomous digital shadows.

As shown in Tab. 4, the majority of the sampled AI simulation techniques (19 of 22 – 86.4%) implement a digital twin. The corresponding architecture is shown in Fig. 3a as an instantiation of the DT4AI framework. Most of these techniques (16 of 22 – 72.7% overall) implement fully autonomous digital twins, and only a fraction relies on human supervision [17] or human actuation [3]. The rest of the architectural patterns in Fig. 3 are seldom encountered. Digital shadows and models account for only 3 of 22 (13.6%) studies.

The instantiation of the DT classes of Kritzinger et al. [54] is shown in Fig. 3. *Experimentable Digital Twins* (Fig. 3a) and *Experimentable Digital Shadows* (Fig. 3b) implement the **C-D** observation loop in an asynchronous way (dashed arrows). This is contrasted with the synchronous input in *Experimentable Models* (Fig. 3b). We observed one case in which a DT is used as a proxy to the physical system for the AI agent to interact with [16]. In this setup, the DT acts as a policy enforcer, hence the name *Policy Digital Twin*. However, this pattern only appears in combination with a full DT.

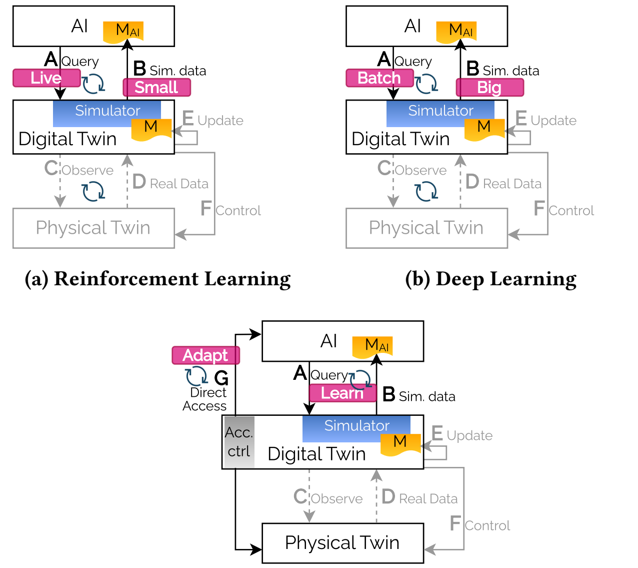
Tab. 5 summarizes the simulation formalisms in the sampled studies. We mostly observe network models, e.g., channel state information [4] and topology models [10] (7 of 22 – 31.8%); models of physics [8, 20] (5 of 22 – 22.7%); and models of geometry and CAx models, e.g., CAD [21] and CAM/CAE [1] (5 of 22 – 22.7%). This aligns with the high representation of network problems (Sec. 5.1). Works that use AI models to encode the simulation model generally do not report the modeling formalism the AI model encodes.



(a) Experimentable DT (b) Experimentable DS

(c) Experimentable Model (d) Policy DT

Figure 3: RQ2: Architectural patterns



(a) Reinforcement Learning (b) Deep Learning

(c) Transfer Learning

Figure 4: RQ3: AI patterns (relevant components highlighted)

Table 5: Modeling and simulation formalisms

Formalism	#Studies	Studies
Network models	7 (31.8%)	[2, 4–6, 10, 14, 18]
Physics	5 (22.7%)	[3, 8, 19–21]
CAD, Geometry	5 (22.7%)	[1, 13, 17, 19, 21]
Process models	3 (13.6%)	[5, 7, 16]
DEVS	1 (4.5%)	[3]
Unclear (DNNs, etc)	5 (22.7%)	[9, 11, 12, 15, 22]

Table 6: DT architecture standards or reference frameworks

Standard	#Studies	Studies
No standard	21 (95.5%)	[2–22]
RAMI4.0	1 (4.5%)	[1]

As shown in Tab. 6, DT architectural standards or reference frameworks are seldom used. We found one study with a standardized architecture (RAMI4.0 by Alexopoulos et al. [1]).

RQ2: Digital Twins

AI simulation chiefly runs through *genuine digital twins* of the *autonomous* kind, but standardization is lagging behind.

5.3 AI and ML techniques (RQ3)

As shown in Tab. 7, the majority of the sampled AI simulation techniques (18 of 22 – 81.8%) rely on some form of reinforcement learning. Deep Reinforcement Learning (DRL, 13 of 22 – 59.1%) is a heavily favored choice, with more value-based methods (8 of 22 – 36.4%) than policy-based (5 of 22 – 22.7%) ones. A deeper look into the details reveals state-of-the-art AI algorithms. Among value-based deep reinforcement learning, we typically find variants of Deep Q Networks [15]; in policy-based methods, we find algorithms such as proximal policy opitmization [11] and deep deterministic policy gradient [8]. The choice of AI methods is rounded out by some approaches adopting deep learning (DL, 4 of 22 – 18.2%; e.g., [12]) and one case of transfer learning (TL, 1 of 22 – 4.5%; e.g., [16]).

Table 7: AI methods

AI	#Studies	Studies
RL	18 (81.8%)	
↳ DRL	13 (59.1%)	
↳ Value	8 (36.4%)	[2, 10, 14, 15, 18, 19, 21, 22]
↳ Policy	5 (22.7%)	[6, 8, 9, 11, 13]
↳ Vanilla	5 (22.7%)	[4, 7, 16, 17, 20]
DL	4 (18.2%)	[1, 3, 5, 12]
TL	1 (4.5%)	[16]

The corresponding instantiations of the DT4AI framework are shown in Fig. 4. Structurally, *Reinforcement learning* (Fig. 4a) and *Deep learning* (Fig. 4a) are identical. However, there are important differences in the interactions within the **A-B** learning cycle. *Reinforcement learning* establishes a *live* interaction, where the *AI* issues frequent, short queries for *small* amounts of simulated data. In contrast, in *Deep learning*, infrequent, often a singular query is issued to which the *Digital twin* responds with *big* data. *Transfer learning* makes use of the *Physical twin*, for which the AI agent uses the *Policy DT* pattern discussed in Sec. 5.2. After the *learning* phase, the *AI* interacts with the *Physical twin* to *adapt* the previously learned knowledge—either to adopt the knowledge to a changing environment or to mitigate sim-to-real threats [83]. In support of this process, the *Digital Twin* ensures the necessary reliability, safety, and security measures [16].

RQ3: AI/ML techniques

AI simulation is predominantly used for *training* purposes of *reinforcement learning* agents, especially in combination with *deep learning* (deep reinforcement learning).

5.4 Simulator lifecycle models (RQ4)

We aim to understand the lifecycle models along which digital twins and, in particular, simulator components are being used and maintained. Unfortunately, the low attention to detail in the simulations aspect (see Fig. 1) makes it challenging to derive in-depth insights.

In general, we observe that AI simulation is provided as a service by the digital twin, and there is *no need to detach* the digital twin from the physical twin when AI simulation takes place.

When it comes to *updating and maintaining* the simulators, we find the patterns reported in Tab. 8. 11 of 22 (50.0%) sampled approaches implement a continuous, automated update mechanism. As shown in the corresponding architecture in Fig. 5a, an automated update mechanism implements the **C-D** loop using *Passive* observation, to which the response is voluminous *Low context* data which the *Digital Twin* has to sift through before *Updating* the model.

This mechanism is contrasted with on-demand techniques that account for 2 of 22 (9.1%) cases in our sample. As shown in Fig. 5, on-demand mechanisms respond to situations in which the digital twin cannot provide sufficient simulated data, e.g., due to the *Query* of the *AI* being outside the validity range of the simulator. For example, in [18], the reinforcement learning agent asks for the simulation of a state that the simulator has limited or no data about. In these situations, the *Digital Twin* needs to sample from the *Physical Twin*, either in an asynchronous (Fig. 5b) or synchronous fashion (Fig. 5c). In both cases, *identifying missing data* is the first step, from which an *Active* implementation of the **C-D** loop follows. *Active* observation is achieved by *Controlling* the *Physical Twin* appropriately. In response, the observation provides *High context* data, which is more related to the particular action the *Digital Twin* has taken to sample the behavior of its physical surroundings. As customary in synchronous modes of operation, the execution of AI training might be *Blocked* until the update is complete.

RQ4: Simulator lifecycle models

Only about 60% of digital twin-driven AI simulation techniques support the *maintainance* of the simulator's quality and fidelity. Most techniques implement *automated, passive data collection* from the physical twin for this purpose.

5.5 Open Challenges (RQ5)

We now discuss key challenges mentioned in the primary studies. We warn that challenges and limitations are sporadically reported (see Fig. 1). To mitigate threats to validity, we avoid interpretation at this point as much as possible and report only factual information.

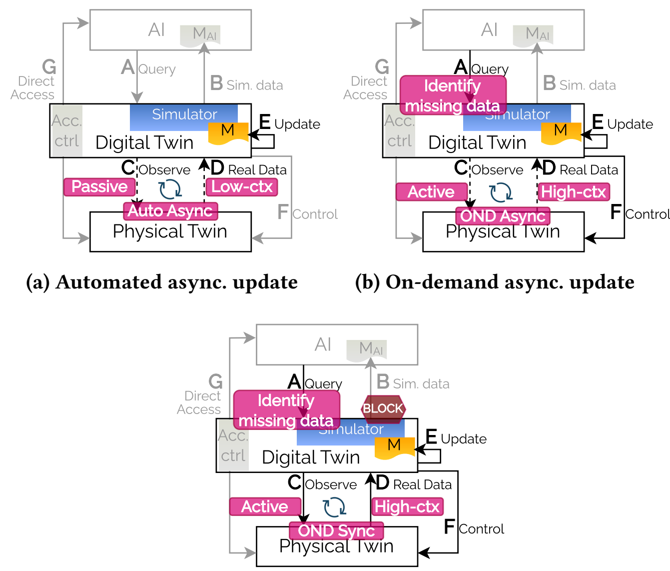
5.5.1 Fidelity and other extra-functional properties. Obviously, fidelity is a key property of simulated data, directly linked to the fidelity and accuracy of the digital twin's simulation model [1]; but **fidelity is hard to assess and ensure** [14]. Shen et al. [13] highlight the challenges involved in achieving accurate virtual replications. Sim-to-real transfer is particularly challenging, as noted by Li et al. [8] who state that "*the gap between simulation and reality greatly limits the application of deep reinforcement learning in the path planning problem of multi-UAV*." Pun et al. [12] recognize sim-to-real discrepancies, particularly when encoding simulation models in generative adversarial networks (GANs). Among other extra-functional properties, **safety** [21], **reliability** [5], and **security** [16] are mentioned. For example, reliability is a particular concern in multi-access edge computing (also known as mobile edge computing) [5] due to its ultra-low latency guarantees.

5.5.2 Interactions with the physical system. Shui et al. [14] warn that the **frequency of interactions** between the digital and physical twin might be limited and thus, inadequate for acquiring sufficient amounts of real data. Similar problems have been voiced by Shen et al. [13]. In some cases, data might be provided by human stakeholders [3], naturally limiting the update frequency of the model and the quality of collected data.

5.5.3 Process aspects. In general, the **transition from concept to practical implementation** of digital twins is recognized as a complex process. Matulis and Harvey [11] voice concerns over the complexity of real-life manufacturing settings which challenges

Table 8: Simulator maintenance patterns

Update	#Studies	Studies
Automated	11 (50.0%)	[2, 4–6, 10, 11, 13–15, 21, 22]
On-demand	2 (9.1%)	[1, 18]
No update	9 (40.9%)	[3, 7–9, 12, 16, 17, 19, 20]



(a) Automated async. update **(b) On-demand async. update**

(c) On-demand sync. update

Figure 5: RQ4: Simulator lifecycle patterns (relevant components highlighted)

deployment. Tubeuf et al. [16] mention the challenges of deploying overly sophisticated models into real settings. Among the frequent organizational challenges, implementation expenses and organizational maturity levels are cited. Alexopoulos et al. [1] identify development and integration as the key cost factors in their DT-enabled AI simulation approach. David et al. [3] report mismatches between twinning ambitions and low levels of operational maturity from underdigitalized domains, such as cyber-biophysical systems.

5.5.4 Challenges as boundary conditions. There are challenges that are outside of the expertise of DT and MDE experts. These challenges are to be treated as boundary conditions in prospective projects. Multiple studies cite the **elevated computational and hardware demands** of DT-enabled AI simulation. Hardware constraints and inadequate computing power substantially impacts AI training, as noted by Matulis and Harvey [11]. Storage space might be a limitation as well, especially in solutions running through external cloud-hosted services, as discussed, e.g., by Deng et al. [4]. Multiple studies mention the **tuning challenges of AI algorithms**, with typical examples of finding the trade-off between exploration and exploitation in reinforcement learning [19] and fine-tuning the hyperparameters of deep learning [16].

RQ5: Open challenges

Challenges in DT-enabled AI simulation include both *technical* (e.g., assessing and ensuring fidelity and establishing sufficient interactions with the physical twin) and *organizational* kinds (e.g., managing development processes).

6 DISCUSSION

We now discuss the key takeaways and lessons learned.

6.1 Key takeaways

6.1.1 Digital twinning brings unique benefits (and challenges) to AI simulation. Digital twinning seems to be a useful instrument in implementing AI simulation. As a key benefit, digital twins provide mature system organization principles and architectures in which the key components of AI simulation be situated—simulators as clearly defined functional entities [68] and AI as a service through well-defined end-points [3]. Another benefit of digital twins is their uniquely tight coupling with the underlying physical systems, which digital twins can observe (Fig. 5a – e.g., [2, 4]) and interrogate upon request (Fig. 5c–5b – e.g., [1, 18]), allowing for evolutionary strategies of simulators. On the negative side (Sec. 5.5), fidelity and proper lifecycle models for digital twins remain a challenge.

These demonstrated contributions to the surging AI market suggest a likely increased adoption rate of digital twin technology. We anticipate digitally adept domains to follow suit with networking and robotics (Sec. 5.1) and adopt digital twins for AI simulation and traditional control and governance-related purposes. Thus, the link between digital twins and AI is shaping up to be one of the impactful directions for digital twin researchers.

6.1.2 (Wireless) networks and robotics paving the way for DT-enabled AI simulation. There is a clear trend in the application domains of DT-enabled AI simulation, with networks and robotics accounting for 17 of 22 (77.3%) of the sampled approaches (see Sec. 5.1). These numbers are rather unexpected after the recent cross-domain systematic mapping study on software engineering for digital twins by Dalibor et al. [32], who do not mention these fields as frequent adopters of digital twins [32, Fig. 5]. Granted, robotics might fit the "manufacturing" in that classification, but the emergence of the networking domain as a top adopter suggests a shift in tone-setters as DT-enabled AI simulation might be growing out of domains different from traditional digital twinning. The strong showing of robotics might be explained by digital twinning being an already adopted technology. The high research activity in networking seems to be a transformative tendency, potentially due to the relative lack of digital twinning impediments [74, Sec. 3.3] in the domain.

6.1.3 Genuine digital twins dominate AI simulation. One of the unexpected observations of this study is the strong alignment of the notion of a digital twin with the classical definitions Kritzinger et al. [54]. 19 of 22 (86.4%) studies (Sec. 5.2) report a digital twin that (i) collects real-time from a physical system and (ii) exerts control on the physical system. This number is much higher than in traditionally considered digital twinning domains, such as manufacturing, where "digital shadows" are quite often encountered. We hypothesize that the recent surge (2022–2024) of digital twinning in the network domain benefited from mature technologies in an already highly digitalized domain, allowing for advanced digital twin solutions.

6.1.4 Deep learning proliferates – and that, in different flavors. The main observation in regard to RQ3 (Sec. 5.3) is that reinforcement learning is particularly highly utilized (e.g., [17, 20]). We see reinforcement learning as a naturally good fit with twinned setups. Reinforcement learning relies on a trial-and-error learning Markovian process [70], in which digital twins can act as the supporting technology for safe, reliable, and reproducible experiments. This role is in line with the "risk-free experimentation aid" role of DTs envisioned by Barat et al. [27] in techno-socio-economic systems.

Within reinforcement learning, we find a high number of *deep* reinforcement learning methods (e.g., [6, 8]), that is, reinforcement learning that encodes the policy as a deep neural network. This number, 13 of 22 (59.1%), together with other deep learning techniques (e.g., [5, 12]) amounting to 4 of 22 (18.2%) studies, means that a total of 17 of 22 (77.3%) methods rely on deep neural networks. Thus, DTs have to be able to provide large amounts of data, either in small batches through rapid interactions (Fig. 4a) or as big data at once (Fig. 4b). Both scenarios challenge extra-functional quality metrics of DT, such as performance, reliability, and availability [16].

6.1.5 Simulation: "using the most appropriate formalisms". The choice of modeling and simulation formalisms aligns with the distribution of domains (Sec. 5.1). We see a number of network models describing topologies and channel dynamics (e.g., [10, 14]), used in network-themed studies. We see a number of physics and CAD 3D geometry models in robotics and manufacturing-themed studies (e.g., [13, 17]), which is in line with the observations of Dalibor et al., who report a high number of CAD 3D models and mathematical physical models in their systematic mapping study [32, Fig. 11].

In some cases, however, the exact simulation formalism is hard to identify. These are the cases in which the simulation model itself is encoded in a neural network, such as a deep neural network (e.g., [9]) or a generative adversarial network (GAN) (e.g. [12]).

6.2 Lessons learned for the DT and MDE Communities

6.2.1 Architectural concerns. Perhaps the most important lesson learned for the DT and MDE communities is the complete lack of digital twin standards, architectural blueprints, and reporting guidelines in the primary studies we sampled. As reported in Sec. 5.2, we found only one paper (4.5%) that relies on the Reference Architectural Model Industrie 4.0 (RAMI4.0) [44], but even in this sole case [1], the work failed to make a connection with the Asset Administration Shell (AAS) [45], the standardized digital representation of assets within RAMI for digital twinning purposes. The lack of architectural standardization is particularly concerning in cases when legacy systems are retrofitted to accommodate digital twins, and the ramifications of system evolution are not being investigated at the architectural level. Standards, such as the ISO 23247 Digital Twin Framework for Manufacturing [68], hold particular potential in this aspect and should be considered by prospective researchers. We recommend the DT community to focus efforts on **making DT architectural standards more accessible to AI researchers** for the sake of scalable, reliable, and sustainable AI simulation.

6.2.2 Towards better technical sustainability of AI simulation by digital twins. Technical sustainability is the ability of a system to be used over an extended lifetime [63]. In terms of AI simulation, prolonged usability boils down primarily to maintaining the faithfulness and validity of simulators. The general notion of AI simulation does not consider this longitudinal dimension [47]. Digital twins improve the technical sustainability outlooks of AI simulation by construction. It is thanks to the tight coupling with its physical environment that digital twins can support various modes of maintaining their simulators' faithfulness, e.g., through observing or experimenting with the physical environment (Sec. 5.4). Recent developments in digital twin evolution [33] provide an additional dimension of sustainable AI simulation.

In this respect, we note a low number of techniques that implement on-demand simulator maintenance (Tab. 8), as only about 18% of the sampled studies do so. We **recommend researching sophisticated maintenance mechanisms and architectures** (Fig. 5b–5c) in response to the anticipated demand for such features. We warn that these efforts might be challenged by the lack of standards which we observed in Sec. 5.2, and which is an acute issue in digital twin engineering in general [58, Sec 6.3.3].

6.2.3 Validity and sim2real. Increasing efforts have been dedicated to transferring the knowledge obtained in a simulated environment to real-world applications, known as sim-to-real [38]—a potential problem in critical systems, such as autonomous vehicles [51]. Our investigation of the state of the art reveals that little attention is dedicated to the sim-to-real problem in digital twin-based AI simulation currently. (See the replication package for more detail.)

The validity of models has been of a particular interest in the modeling and simulation community. Especially in recent years, the traditional and vague notion of a simulation frame by Zeigler et al. [82] has been clarified by a series of works. Biglari and Denil [29], Mittal et al. [57], and Van Acker et al. [75] situate validity at the digital-to-physical boundary of digital twins, reflecting on the validity of simulation models w.r.t. to environmental conditions, engineering assumptions, etc. We recommend **modeling and simulation experts to adopt research results on validity frames in support of AI simulation** to allow for better sim-to-real transfer.

6.2.4 Human factors in AI simulation. We observe an overall ignorance of human factors. This holds both for human experts in the AI simulation loop and for human stakeholders in the development and operation of digital twins serving AI simulation. These trends are best exemplified in Sec. 5.2 and, specifically, in the breakdown of system organization patterns in Tab. 4. These trends are not entirely surprising as socio-technical views on digital twins are in their early phase [34]. The role of the human in the loop is fully expected to grow, e.g., in training the virtual replicas [58, Fig 7], and guiding AI agents in their learning phase [31]. With that, we recommend digital twin experts to **adopt more human-centered views on digital twins, both in terms of the human as an interactive user of AI simulation and as a stakeholder in the development and operation process of DTs**.

6.2.5 Reporting quality and recommendations. Finally, we remark on some quality-related trends in the primary studies in our corpus. First of all, we notice a low level of detail in discussing the simulation aspects of AI simulation (Sec. 3.3). This is a severe shortcoming, considering the central role of simulation in these approaches. The lack of detail about simulation is especially concerning, given that the validity of simulation models is the primary factor that determines the validity of data that is generated for training AI agents. We recommend prospective researchers to **be more detailed and transparent about simulation formalisms, methods, algorithms, and tools** when reporting their work. This will allow for independent validation and reproduction of results.

We also note the staggering lack of support for the reproducibility and independent validation of results. We have not found any data supplements or replication packages despite data being the central artifact in AI simulation. We urge **methodologists in simulation and AI to develop joint standards**, and **conference organizers to introduce artifact evaluation practices**, such as the ACM Artifact Review and Badging procedure [25].

7 CONCLUSION

In this paper, we analyzed the trends in digital twin-enabled AI simulation, and derived a conceptual reference framework to situate digital twins and AI with respect to each other. Our inquiry into the state of the art suggests that AI simulation by digital twins is a rapidly emerging field with demonstrated benefits in specific problem domains. At the same time, AI simulation by digital twins is still in its infancy, marked by limited usage of digital twin capabilities, simple lifecycle models, and lacking architectural guidelines—challenges that require active involvement from the digital twin engineering community. To foster involvement, we identify challenges and research opportunities for prospective researchers.

PRIMARY STUDIES

[1] Kosmas Alexopoulos, Nikolaos Nikolakis, and George Chryssolouris. 2020. Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing. *Int J Comput Integr Manuf* 33, 5 (2020), 429–439. https://doi.org/10.1080/0951192X.2020.1747642

[2] Yingping Cui, Tiejun Lv, Wei Ni, and Abbas Jamalipour. 2023. Digital Twin-Aided Learning for Managing Reconfigurable Intelligent Surface-Assisted, Uplink, User-Centric Cell-Free Systems. *IEEE Journal on Selected Areas in Communications* 41, 10 (2023), 3175–3190. https://doi.org/10.1109/JSAC.2023.3310050

[3] Istvan David, Pascal Archambault, Quentin Wolak, Cong Vinh Vu, Timothé Lalonde, Kashif Riaz, Eugene Syriani, and Houari Sahraoui. 2023. Digital Twins for Cyber-Biophysical Systems: Challenges and Lessons Learned. In *2023 ACM/IEEE 26th International Conference on Model Driven Engineering Languages and Systems (MODELS)*. 1–12. https://doi.org/10.1109/MODELS58315.2023.00014

[4] Juan Deng, Qingbi Zheng, Guangyi Liu, Jielin Bai, Kaicong Tian, Changhao Sun, Yujie Yan, and Yitong Liu. 2021. A Digital Twin Approach for Self-optimization of Mobile Networks. In *2021 IEEE Wireless Communications and Networking Conference Workshops*. 1–6. https://doi.org/10.1109/WCNCW49093.2021.9420037

[5] Rui Dong, Changyang She, Wibowo Hardjawana, Yonghui Li, and Branka Vucetic. 2019. Deep Learning for Hybrid 5G Services in Mobile Edge Computing Systems: Learn From a Digital Twin. *IEEE Transactions on Wireless Communications* 18, 10 (2019), 4692–4707. https://doi.org/10.1109/TWC.2019.2927312

[6] Hongzhi Guo, Xiaoyi Zhou, Jiadai Wang, Jiajia Liu, and Abderrahim Benslimane. 2023. Intelligent Task Offloading and Resource Allocation in Digital Twin Based Aerial Computing Networks. *IEEE Journal on Selected Areas in Communications* 41, 10 (2023), 3095–3110. https://doi.org/10.1109/JSAC.2023.3310067

[7] Kim Hammar and Rolf Stadler. 2023. Digital Twins for Security Automation. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*. 1–6. https://doi.org/10.1109/NOMS56928.2023.10154288

[8] Siyuan Li, Xi Lin, Jun Wu, Ali Kashif Bashir, and Raheel Nawaz. 2022. When digital twin meets deep reinforcement learning in multi-UAV path planning. In *Proceedings of the 5th International ACM Mobicom Workshop on Drone Assisted Wireless Communications for 5G and Beyond (DroneCom '22)*. ACM, 61–66. https://doi.org/10.1145/3555661.3560865

[9] Yuxuan Li, Cheng Zhang, Yongming Huang, and Qingbi Zheng. 2023. Optimization of Broadcast Beams in Massive MIMO: Learning from A Digital Twin. In *2023 IEEE Globecom Workshops (GC Wkshps)*. 2043–2048. https://doi.org/10.1109/GCWkshps58843.2023.10464807

[10] Tong Liu, Lun Tang, Weili Wang, Qianbin Chen, and Xiaoping Zeng. 2022. Digital-Twin-Assisted Task Offloading Based on Edge Collaboration in the Digital Twin Edge Network. *IEEE Internet of Things Journal* 9, 2 (2022), 1427–1444. https://doi.org/10.1109/JIOT.2021.3086961

[11] Marius Matulis and Carlo Harvey. 2021. A robot arm digital twin utilising reinforcement learning. *Computers & Graphics* 95 (2021), 106–114. https://doi.org/10.1016/j.cag.2021.01.011

[12] Ava Pun, Gary Sun, Jingkang Wang, Yun Chen, Ze Yang, Sivabalan Manivasagam, Wei-Chiu Ma, and Raquel Urtasun. 2023. Neural Lighting Simulation for Urban Scenes. In *Advances in Neural Information Processing Systems*, Vol. 36. Curran Associates, Inc., 19291–19326. https://openreview.net/pdf?id=mcx8IGneYw

[13] Gaoqing Shen, Lei Lei, Zhilin Li, Shengsuo Cai, Lijuan Zhang, Pan Cao, and Xiaojiao Liu. 2022. Deep Reinforcement Learning for Flocking Motion of Multi-UAV Systems: Learn From a Digital Twin. *IEEE Internet of Things Journal* 9, 13 (2022), 11141–11153. https://doi.org/10.1109/JIOT.2021.3127873

[14] Tingyu Shui, Jie Hu, Kun Yang, Honghui Kang, Hua Rui, and Bo Wang. 2023. Cell-Free Networking for Integrated Data and Energy Transfer: Digital Twin Based Double Parameterized DQN for Energy Sustainability. *IEEE Trans Wirel Commun* 22, 11 (2023), 8035–8049. https://doi.org/10.1109/TWC.2023.3258143

[15] Xin Tang, Xiaohuan Li, Rong Yu, Yuan Wu, Jin Ye, Fengzhu Tang, and Qian Chen. 2023. Digital-Twin-Assisted Task Assignment in Multi-UAV Systems: A Deep Reinforcement Learning Approach. *IEEE Internet of Things Journal* 10, 17 (2023), 15362–15375. https://doi.org/10.1109/JIOT.2023.3263574

[16] Carlotta Tubeuf, Felix Birkelbach, Anton Maly, and Rene Hofmann. 2023. Increasing the Flexibility of Hydropower with Reinforcement Learning on a Digital Twin Platform. *Energies* 16, 4 (2023). https://doi.org/10.3390/en16041796

[17] Igor Verner, Dan Cuperman, Amy Fang, Michael Reitman, Tal Romm, and Gali Balikin. 2018. Robot Online Learning Through Digital Twin Experiments: A Weightlifting Project. In *Online Engineering & Internet of Things*. Springer International Publishing, Cham, 307–314.

[18] Irene Vilà, Oriol Sallent, and Jordi Pérez-Romero. 2023. On the Design of a Network Digital Twin for the Radio Access Network in 5G and Beyond. *Sensors* 23, 3 (2023). https://doi.org/10.3390/s23031197

[19] Kaishu Xia, Christopher Sacco, Max Kirkpatrick, Clint Saidy, Lam Nguyen, Anil Kircaliali, and Ramy Harik. 2021. A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence. *Journal of Manufacturing Systems* 58 (2021), 210–230. https://doi.org/10.1016/j.jmsy.2020.06.012

[20] Xubo Yang, Jian Gao, Peng Wang, Wenyi Long, and Chongbo Fu. 2024. Social Learning with Actor–Critic for dynamic grasping of underwater robots via digital twins. *Ocean Engineering* 306 (2024), 118070. https://doi.org/10.1016/j.oceaneng.2024.118070

[21] Lixiang Zhang, Chen Yang, Yan Yan, Ze Cai, and Yaoguang Hu. 2024. Automated guided vehicle dispatching and routing integration via digital twin with deep reinforcement learning. *Journal of Manufacturing Systems* 72 (2024), 492–503. https://doi.org/10.1016/j.jmsy.2023.12.008

[22] Yining Zhang, Hengsheng Zhang, Xiaokai Liu, Chenglin Zhao, and Fangmin Xu. 2023. Digital twin-enabled deep reinforcement learning for joint scheduling of ultra-reliable low latency communication and enhanced mobile broad band: A reliability-guaranteed approach. *Transactions on Emerging Telecommunications Technologies* 34, 3 (2023), e4705. https://doi.org/10.1002/ett.4705

REFERENCES

[23] Amazon, Inc. 2012. Amazon Mechanical Turk. https://www.mturk.com/. Accessed: 2024-06-24.

[24] Zahra Ashktorab, Michael Desmond, Josh Andres, Michael Muller, Narendra Nath Joshi, Michelle Brachman, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Christine T. Wolf, Evelyn Duesterwald, Casey Dugan, Werner Geyer, and Darrell Reimer. 2021. AI-Assisted Human Labeling: Batching for Efficiency without Overreliance. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 89 (apr 2021), 27 pages. https://doi.org/10.1145/3449163

[25] Association for Computing Machinery. 2020. Artifact Review and Badging Version 1.1. https://www.acm.org/publications/policies/artifact-review-and-badging-current. Accessed: 2024-06-24.

[26] Shanaka Kristombu Baduge, Sadeep Thilakarathna, Jude Shalitha Perera, Mehrdad Arashpour, Pejman Sharafi, Bertrand Teodosio, Ankit Shringi, and Priyan Mendis. 2022. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction* 141 (2022), 104440. https://doi.org/10.1016/j.autcon.2022.104440

[27] Souvik Barat, Vinay Kulkarni, Tony Clark, and Balbir Barn. 2022. Digital twin as risk-free experimentation aid for techno-socio-economic systems. In *Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems (MODELS '22)*. ACM, 66–75. https://doi.org/10.1145/3550355.3552409

[28] Pascal Bibow, Manuela Dalibor, Christian Hopmann, Ben Mainz, Bernhard Rumpe, David Schmalzing, Mauritius Schmitz, and Andreas Wortmann. 2020. Model-Driven Development of a Digital Twin for Injection Molding. In *Advanced Information Systems Engineering*. Springer, 85–100.

[29] Raheleh Biglari and Joachim Denil. 2022. Model validity and tolerance quantification for real-time adaptive approximation. In *Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings (MODELS '22)*. ACM, 668–676. https://doi.org/10.1145/3550356.3561604

[30] Stefan Boschert and Roland Rosen. 2016. Digital twin–the simulation aspect. In *Mechatronic futures*. Springer, 59–74.

[31] Kyanna Dagenais and Istvan David. 2024. *Opinion-Guided Reinforcement Learning*. Technical Report. arXiv:2405.17287 [cs.LG]

[32] Manuela Dalibor, Nico Jansen, Bernhard Rumpe, David Schmalzing, Louis Wachtmeister, Manuel Wimmer, and Andreas Wortmann. 2022. A Cross-Domain Systematic Mapping Study on Software Engineering for Digital Twins. *J Syst Softw* 193 (2022), 111361. https://doi.org/10.1016/j.jss.2022.111361

[33] Istvan David and Dominik Bork. 2023. Towards a Taxonomy of Digital Twin Evolution for Technical Sustainability. In *ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion, MODELS-C*. IEEE, 934–938. https://doi.org/10.1109/MODELS-C59198.2023.00147

[34] Istvan David and Dominik Bork. 2024. Infonomics of Autonomous Digital Twins. In *Advanced Information Systems Engineering (Lecture Notes in Computer Science, Vol. 14663)*. Springer, 563–578. https://doi.org/10.1007/978-3-031-61057-8_33

[35] Istvan David, Jessie Galasso, and Eugene Syriani. 2021. Inference of Simulation Models in Digital Twins by Reinforcement Learning. In *2021 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*. 221–224. https://doi.org/10.1109/MODELS-C53483.2021.00038

[36] Istvan David and Eugene Syriani. 2024. *Automated Inference of Simulators in Digital Twins*. CRC Press, Chapter 8, 122–148. https://doi.org/10.1201/9781003425724-11

[37] Michael Desmond, Michael Muller, Zahra Ashktorab, Casey Dugan, Evelyn Duesterwald, Kristina Brimijoin, Catherine Finegan-Dollak, Michelle Brachman, Aabhas Sharma, Narendra Nath Joshi, and Qian Pan. 2021. Increasing the Speed and Accuracy of Data Labeling Through an AI Assisted Interface. In *Proc. of the 26th International Conference on Intelligent User Interfaces*. ACM, 392–401.

[38] Carl Doersch and Andrew Zisserman. 2019. Sim2real transfer learning for 3D human pose estimation: motion to the rescue. In *Advances in Neural Information Processing Systems*, Vol. 32. Curran Associates, Inc.

[39] EDT Community. 2024. Awesome Digital Twins – Interesting Papers. https://github.com/edt-community/awesome-digital-twins?tab=readme-ov-file#interesting-papers. Accessed: 2024-06-24.

[40] Lisa Ehrlinger and Wolfram Wöß. 2022. A survey of data quality measurement and monitoring tools. *Frontiers in big data* 5 (2022), 850611.

[41] Frank Emmert-Streib. 2023. What Is the Role of AI for Digital Twins? *AI* 4, 3 (2023), 721–728. https://doi.org/10.3390/ai4030038

[42] Michael Erdmann, Alexander Maedche, Hans-Peter Schnurr, and Steffen Staab. 2000. From manual to semi-automatic semantic annotation: About ontology-based text annotation tools. In *Proceedings of the COLING-2000 workshop on semantic annotation and intelligent content*. 79–85.

[43] Bahar Farahani and Amin Karimi Monsefi. 2023. Smart and collaborative industrial IoT: A federated learning and data space approach. *Digital Communications and Networks* 9, 2 (2023), 436–447.

[44] Federal Ministry for Economic Affairs and Germany Climate Action. 2018. Reference Architectural Model Industrie 4.0 (RAMI4.0) – An Introduction. https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/rami40-an-introduction.html. Accessed: 2024-06-24.

[45] OPC Foundation. 2021. Industry 4.0 Asset Administration Shell. https://reference.opcfoundation.org/I4AAS/v100/docs. Accessed: 2024-06-24.

[46] Salvador García, Sergio Ramírez-Gallego, Julián Luengo, José Manuel Benítez, and Francisco Herrera. 2016. Big data preprocessing: methods and prospects. *Big data analytics* 1 (2016), 1–22.

[47] Gartner Research. 2023. Innovation Insight: AI Simulation. https://www.gartner.com/en/documents/4037399. Accessed: 2024-06-24.

[48] Trisha Greenhalgh and Richard Peacock. 2005. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ* 331, 7524 (2005), 1064–1065.

[49] Thilo Hagendorff and Katharina Wezel. 2020. 15 challenges for AI: or what AI (currently) can't do. *AI & SOCIETY* 35, 2 (01 Jun 2020), 355–365. https://doi.org/10.1007/s00146-019-00886-y

[50] Anne-Wil Harzing. 2016. Publish or Perish. https://harzing.com/resources/publish-or-perish. Accessed: 2024-06-24.

[51] Xuemin Hu, Shen Li, Tingyu Huang, Bo Tang, Rouxing Huai, and Long Chen. 2024. How Simulation Helps Autonomous Driving: A Survey of Sim2real, Digital Twins, and Parallel Intelligence. *IEEE Transactions on Intelligent Vehicles* 9, 1 (2024), 593–612. https://doi.org/10.1109/TIV.2023.3312777

[52] Florian Jaensch, Akos Csiszar, Christian Scheifele, and Alexander Verl. 2018. Digital Twins of Manufacturing Systems as a Base for Machine Learning. In *25th International Conference on Mechatronics and Machine Vision in Practice*. 1–6.

[53] Barbara Kitchenham and Stuart Charters. 2007. *Guidelines for performing systematic literature reviews in software engineering*. Technical Report EBSE-2007-01. Keele, UK, Keele University. 1–65 pages.

[54] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. 2018. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine* 51, 11 (2018), 1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018.

[55] Christian Legaard, Thomas Schranz, Gerald Schweiger, Ján Drgoňa, Basak Falay, Cláudio Gomes, Alexandros Iosifidis, Mahdi Abkar, and Peter Larsen. 2023. Constructing Neural Network Based Models for Simulating Dynamical Systems. *ACM Comput. Surv.* 55, 11, Article 236 (feb 2023), 34 pages. https://doi.org/10.1145/3567591

[56] Stefan Mihai, Mahnoor Yaqoob, Dang V. Hung, William Davis, Praveer Towakel, Mohsin Raza, Mehmet Karamanoglu, Balbir Barn, Dattaprasad Shetve, Raja V. Prasad, Hrishikesh Venkataraman, Ramona Trestian, and Huan X. Nguyen. 2022. Digital Twins: A Survey on Enabling Technologies, Challenges, Trends and Future Prospects. *IEEE Communications Surveys & Tutorials* 24, 4 (2022), 2255–2291. https://doi.org/10.1109/COMST.2022.3208773

[57] Rakshit Mittal, Raheleh Eslampanah, Lucas Lima, Hans Vangheluwe, and Dominique Blouin. 2023. Towards an Ontological Framework for Validity Frames. In *2023 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*. 801–805. https://doi.org/10.1109/MODELS-C59198.2023.00128

[58] Hossain M. Muctadir, David A. Manrique Negrin, Raghavendran Gunasekaran, Loek Cleophas, Mark van den Brand, and Boudewijn R. Haverkort. 2024. Current trends in digital twin development, maintenance, and operation: an interview study. *Soft. Sys. Mod.* (2024). https://doi.org/10.1007/s10270-024-01167-z

[59] Suresh Neethirajan. 2021. Is Seeing Still Believing? Leveraging Deepfake Technology for Livestock Farming. *Frontiers in Veterinary Science* 8 (2021). https://doi.org/10.3389/fvets.2021.740253

[60] Aurélie Névéol, Rezarta Islamaj Doğan, and Zhiyong Lu. 2011. Semi-automatic semantic annotation of PubMed queries: a study on quality, efficiency, satisfaction. *Journal of biomedical informatics* 44, 2 (2011), 310–318.

[61] Thi Huyen Nguyen and Koustav Rudra. 2024. Human vs ChatGPT: Effect of Data Annotation in Interpretable Crisis-Related Microblog Classification. In *Proceedings of the ACM on Web Conference 2024 (WWW '24)*. ACM, 4534–4543. https://doi.org/10.1145/3589334.3648141

[62] Randy Paredis, Hans Vangheluwe, and Pamela Adelino Ramos Albertins. 2024. COOCK project Smart Port 2025 D3.1: "To Twin Or Not To Twin". arXiv:2401.12747

[63] Birgit Penzenstadler and Henning Femmer. 2013. A generic model for sustainability with process- and product-specific instances. In *Proceedings of the 2013 Workshop on Green in/by Software Engineering* (Fukuoka, Japan) *(GIBSE '13)*. ACM, 3–8. https://doi.org/10.1145/2451605.2451609

[64] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. Systematic mapping studies in software engineering. In *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering* (Italy) *(EASE'08)*. BCS Learning & Development Ltd., Swindon, GBR, 68–77.

[65] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Inf Softw Technol* 64 (2015), 1–18. https://doi.org/10.1016/j.infsof.2015.03.007

[66] Qinglin Qi and Fei Tao. 2018. Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. *IEEE Access* 6 (2018), 3585–3593. https://doi.org/10.1109/ACCESS.2018.2793265

[67] Sheldon M Ross. 2022. *Simulation* (6 ed.). Academic Press, San Diego, CA.

[68] Guodong Shao. 2021. *Use case scenarios for digital twin implementation based on ISO 23247*. Technical Report. https://doi.org/10.6028/nist.ams.400-2

[69] P. Smolensky. 1987. Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review* 1, 2 (01 Jun 1987), 95–109. https://doi.org/10.1007/BF00130011

[70] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

[71] Eugene Syriani, Istvan David, and Gauransh Kumar. 2023. Assessing the Ability of ChatGPT to Screen Articles for Systematic Reviews. arXiv:2307.06464

[72] Eugene Syriani, Istvan David, and Gauransh Kumar. 2024. Screening articles for systematic reviews with ChatGPT. *Journal of Computer Languages* 80 (2024), 101287. https://doi.org/10.1016/j.cola.2024.101287

[73] Nikita Tomin, Victor Kurbatsky, Vadim Borisov, and Sergey Musalev. 2020. Development of Digital Twin for Load Center on the Example of Distribution Network of an Urban District. *E3S Web of Conferences* 209 (2020), 02029. https://doi.org/10.1051/e3sconf/202020902029

[74] Jakob Trauer, Michael Mutschler, Markus Mörtl, and Markus Zimmermann. 2022. Challenges in Implementing Digital Twins – A Survey *(Intl Design Engineering Technical Conf and Computers and Information in Engineering Conf, Vol. 42nd Computers and Information in Eng Conf)*.

[75] Bert Van Acker, Paul De Meulenaere, Hans Vangheluwe, and Joachim Denil. 2024. Validity Frame–enabled model-based engineering processes. *SIMULATION* 100, 2 (2024), 185–226. https://doi.org/10.1177/00375497231205035

[76] David A Van Dyk and Xiao-Li Meng. 2001. The art of data augmentation. *Journal of Computational and Graphical Statistics* 10, 1 (2001), 1–50.

[77] Hans Vangheluwe, Juan de Lara, and Pieter J Mosterman. 2002. An introduction to multi-paradigm modelling and simulation. In *Proceedings of the AIS2002 Conference (AI, Simulation and Planning in High Autonomy Systems)*. 9–20.

[78] Roel J. Wieringa and Maya Daneva. 2015. Six strategies for generalizing software engineering theories. *Sci. Comput. Program.* 101 (2015), 136–152. https://doi.org/10.1016/j.scico.2014.11.013

[79] Claes Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE '14)*. ACM, Article 38, 10 pages. https://doi.org/10.1145/2601248.2601268

[80] Claes Wohlin, Emilia Mendes, Katia Romero Felizardo, and Marcos Kalinowski. 2020. Guidelines for the search strategy to update systematic literature reviews in software engineering. *Information and Software Technology* 127 (2020), 106366.

[81] Ibrahim Yitmen, Sepehr Alizadehsalehi, İlknur Akıner, and Muhammed Ernur Akıner. 2021. An Adapted Model of Cognitive Digital Twins for Building Lifecycle Management. *Applied Sciences* 11, 9 (2021). https://doi.org/10.3390/app11094276

[82] Bernard P Zeigler, Alexandre Muzy, and Ernesto Kofman. 2018. *Theory of modeling and simulation: discrete event & iterative system computational foundations*. Academic press.

[83] Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. 2020. Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: a Survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. 737–744. https://doi.org/10.1109/SSCI47803.2020.9308468

[84] Lina Zhou, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos. 2017. Machine learning on big data: Opportunities and challenges. *Neurocomputing* 237 (2017), 350–361. https://doi.org/10.1016/j.neucom.2017.01.026