23rd International Conference on Modeling & Applied Simulation, 010
21st International Multidisciplinary Modeling & Simulation Multiconference

2724-0037 © 2024 The Authors.
doi: 10.46354/i3m.2024.mas.010

Safeguarding Critical Infrastructures with Digital Twins and AI

Andrea Tundis[1,*] and Oscar H. Ramírez-Agudelo[2]

[1]Institute for the Protection of Terrestrial Infrastructures (PI), German Aerospace Center (DLR), Rathausallee 12, St. Augustin, 53757, Germany
[2]Institute for AI Safety and Security (KI), German Aerospace Center (DLR), Rathausallee 12, St. Augustin, 53757, Germany

[*]Corresponding author. Email address: andrea.tundis@dlr.de

Abstract

Critical infrastructures (CIs) are complex systems that are integral parts of our daily lives. We use them to access various services related to basic needs, to obtain water, energy and information, to move from one place to another, to work remotely, and so on. It is therefore essential, but challenging, to secure and protect them. On the other hand, digital twins (DTs) and artificial intelligence (AI) represent solid approaches that are well suited to modelling and analyzing complex systems, respectively. In this context, this work can be seen as a white paper that aims first to explore the main characteristics and limitations of DTs and AI when considered in isolation, and then to discuss how their combination as an intelligent entity - which represents a paradigm shift in the protection and resilience of CIs - might be beneficial to overcome such challenges and thus be useful to enhance their protection.

Keywords: Digital Twins; Artificial intelligence; Critical infrastructures protection; Resilience; Cyber security.

1. Introduction

In contemporary society, Critical Infrastructures (CIs) constitute the backbone of modern civilization, underpinning essential services such as energy distribution, transportation networks, communication systems, and healthcare facilities (Abdulova and Kalashnikov, 2021). However, the increasing complexity, interdependence, and susceptibility to diverse threats pose significant challenges to their resilience and protection (Pastorek and Tundis, 2024; Zibulewsky, 2001; Marquard et al., 2024; López Díaz and Tundis, 2023; Sattler et al., 2023). In response to these challenges, the combination of artificial intelligence (AI) and digital twins (DTs) emerges as a promising paradigm for enhancing the safeguarding and sustainability of such critical systems.

DT are virtual representations that replicate the physical attributes and behavior of infrastructural assets in quasi real-time, offer unprecedented insights into system dynamics, performance, and vulnerabilities (Savaglio et al., 2023). By leveraging advanced sensing technologies, IoT devices, data analytics, and simulation capabilities, DTs enable comprehensive monitoring, predictive maintenance, and scenario analysis, empowering operators, as well as reinforcing cybersecurity of CIs (Masi et al., 2023; Buccafurri et al., 2023, 2022), to proactively identify and address potential risks before they escalate into disruptions, by also considering the socio-technical perspective (Herrmann, 2014).

AI models strongly rely on data availability. DTs offer the possibility to create useful (raw) data for the AI approaches to train and transform it into actionable insights. For example, through the implementation of deep learning (DL) algorithms, AI systems can analyze vast amounts of data collected from DTs, discern patterns, detect objects (Wang et al., 2024) or anomalies, forecast time-series (Leppich et al., 2024) and future states with remarkable accuracy (Kharchenko et al., 2020). This cognitive capacity enables AI to serve as an *intelligent guardian*, continuously surveilling CIs, identifying emerging threats, and orchestrating adaptive and timely responses.

However, different research challenges exist, and some of the main open questions that are tackled in this work are the following:

RQ1 What are the current limitations of DTs and of AI approaches when considered in isolation as stand-alone approaches?

RQ2 How can DTs and AI complement each other in order to overcome their limitations.

RQ3 How can AI-driven DTs be effectively utilized to enhance the resilience and reliability of critical infrastructures across diverse sectors?

RQ4 What are the ethical, regulatory, and societal implications of integrating AI and DTs as intelligent guardians for safeguarding critical infrastructures, and how can these concerns be addressed to ensure responsible innovation and deployment?

RQ5 Data privacy and data quality is important in the development of AI algorithms, as data incorrectness might negatively impact in the results. Therefore, granting high quality data is a fundamental aspect to be considered.

On the basis of that, this paper explores the concepts related to the synergistic relationship between AI and DTs in safeguarding CIs across diverse domains so as to enhance their resilience (see Figure 1).

In particular, on the basis of our interdisciplinary experience, which combines industry practices and academic research, we have examined:

- aspects related to DTs and AI by identifying potential limitations or inhibition;
- how AI and DT complement each other through a *AI-driven Digital Twin conceptual framework* that combines typical DT modeling characteristics with those typical of AI analysis;
- scenarios from different sectors to exemplify the potential benefits and impact of such *"intelligent guardian"*, in mitigating risks, enhancing operational efficiency, and ensuring the continuity of CIs services;
- the importance of adhere to ethical guidelines and regulations related to data privacy, along with associated societal implications as well as algorithmic bias when promoting innovation, especially with the deployment of AI-driven DTs in CIs.

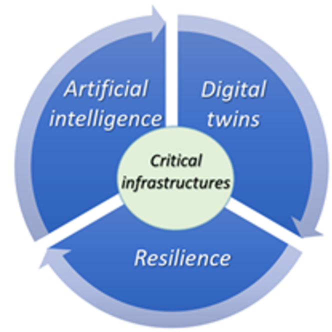

Figure 1. Exploring concepts of Artificial Intelligence and Digital Twins for enhancing the Resilience and the Protection of Critical Infrastructures.

The rest of the paper is structured as follow. Section 2 provides a background on DTs and AI in context of CIs by highlighting their current limitations when considered in isolation. The proposed AI-driven DT conceptual framework is elaborated in Section 3. Section 4 presents a set of application scenarios where the framework can be applied by pointing out the role of DTs and AI including their impact and benefits; whereas implications of the AI are discussed in Section 5. Conclusion and future works are summarized in Section 6.

2. Background and limitations

This section discusses DTs and AI aspects in connection to CIs by pointing out for each of them their current limitations and challenges.

2.1. Digital Twin for Infrastructures

DTs are sophisticated virtual replicas of physical entities, processes, or systems that facilitate real-time monitoring, simulation, and optimization. They are increasingly recognized as a transformative technology for managing CIs such as power grids, transportation networks, water supply systems, and healthcare facilities. These infrastructures are vital for societal functioning and economic stability, and DTs provide a powerful tool for ensuring their security, efficiency, and resilience. Some relevant aspects that can be dealt through DTs in relation to safeguarding Critical Infrastructures (CIs) are:

- *Real-Time Monitoring and Maintenance.* DTs provide continuous, real-time monitoring of infrastructure components, allowing for immediate detection of anomalies and issues. This facilitates proactive maintenance, reducing the likelihood of unexpected failures and ensuring consistent service delivery.

- *Predictive Analytics and Risk Mitigation.* By leveraging historical data and advanced analytics, DTs can forecast potential problems and risks. This predictive capability helps in implementing preventive measures, thereby minimizing the impact of equipment failures, natural disasters, and cyber-attacks.

- *Simulation and Scenario Planning.* DTs enable the simulation of various scenarios, including extreme events such as natural disasters and system failures. These simulations help in developing and testing emergency response plans, ensuring infrastructure resilience and preparedness.

- *Resource Optimization.* DTs optimize the allocation and utilization of resources like energy, water, and transportation. This leads to enhanced operational efficiency and supports sustainable management practices, ensuring the effective functioning of critical infrastructures.

- *Cybersecurity.* DTs can be targets for cyber-attacks, and ensuring their security is critical. They provide a platform for monitoring and responding to cyber threats in real-time, protecting both the virtual models and their physical counterparts from unauthorized access and breaches.

Although the advantages of the above-mentioned aspects, their realization is not an easy task as it requires to face with different challenges. In particular: (i) it is difficult to integrate and ensure accurate data from various sources, moreover large-scale real-time monitoring generates enormous data, requiring significant computational resources; (ii) traditional models may not capture all variables or rare events, in addition large and complex datasets can overwhelm traditional analytics; (iii) the realization of accurate and detailed simulations is typically subjective and high-fidelity simulations require significant computational power and time; (iv) ensuring optimal use of resources across large infrastructures is difficult, and traditional systems may struggle to adapt resource allocation dynamically; (v) finally, DTs can be targets for cyber-attacks, and protecting them is challenging, as traditional security measures may not detect sophisticated threats in real-time. A summary of DT aspects and related challenges is provided in Table 1.

Table 1. DT aspects and related limitations and challenges

DT aspect	Limitations and challenges
Real-time monitoring and maintenance	- Data Integration & Management - Scalability
Predictive analytics and risk mitigation	- Predictive Accuracy - Data Volume & Complexity
Simulation and scenario planning	- Model Complexity - Computational Resources
Resource optimization	- Operational Efficiency - Dynamic Adaptation
Cybersecurity	- Vulnerability to Attacks - Detection & Response

2.2. AI for Critical Infrastructures

This section covers AI for CIs. It means applying AI to enhance the safety and security of infrastructures such as airports, governmental buildings, energy grids, transport systems, etc. As a consequence, AI also helps to improve the resilience of the CIs. In this section, first, an overview of available AI technologies is given, followed by how AI is useful in different domains and its benefits in CIs. Likewise, application examples are provided and the current limitations and challenges are presented.

AI technologies. AI technologies include algorithms in the field of machine learning such as random forest (Louppe, 2015), decision trees, k–nearest neighbors, logistic regression, etc. (Alpaydin, 2020). In the area of deep learning, the most popular algorithms include object detection (Hussain, 2024), image classification, image segmentation (Kirillov et al., 2019). Additionally, Generative Adversarial Networks – GANs (Goodfellow et al., 2014), reinforcement learning (Li, 2018), and natural language processing (NLP, e.g. Large Language Models –*LLMs*; Zou et al., 2023) also become relevant technologies when applying to CIs. The different aforementioned technologies help to analyze vast amounts of data, in order to get insights such as detecting patterns and possible anomalies.

Benefits of AI in CIs. AI enables the optimization of resources, and to automate responses or routine tasks in (quasi) real-time (Yigit et al., 2024). It has the potential to improve risk assessment, prediction, and mitigation, enabling proactive measures to prevent or minimize the impact of attacks in CIs (Koessler and Schuett, 2023). For instance, in case of vulnerabilities or attacks, AI helps to optimise the analysis and as a consequence an informed model-based decision is offered. This leads to address the situation, which in some cases might be severe, with the aim to reduce the risk and increase the safeguarding of CIs. As a consequence AI is capable to significantly improve their situational awareness (Chen et al., 2024).

Application examples. AI can be applied in different domains. Some of the most relevant for CIs are:

- Cybersecurity: AI can help monitoring the safety and securiyt. It has been applied to detect potential cyber-attacks (Ferrag et al., 2024).
- Transport: AI is nowadays being used to surveil and optimize the operations of the public transport such as bus and trains. It is also used to predict the traffic flow (Yan and Li, 2023).
- Distribution of energy and water supply: to predict a failure and to manage and predict the supply (Kumar and Prabhansu, 2023).

Limitation and challenges. AI can be considered a powerful tool as it brings benefits as above explained. However, at the same time it has some challenges and limitations to work in the future. At the same time, in the society might also bring up objection, disagreements which lead to disagreement and objection in using this technology. Here it is mentioned some topics in both aspects:

- *High Performance computing (HPC) facilities*: AI requires most of the time sophisticated and expensive computing technology in order to determine a solution.
- *Data*: it is the bottle neck of AI due to the vast amount of data required in order to determine a solution. When limited and specific data is available is not possible to generalise the solution.
- *Model*: when a model is computed, million of parameters have been fitted. It is difficult to keep track of a good model.
- *Hyper-parameter optimisation*: to choose correctly the parameters is a key factor to land to robust model. Hyper-parameter tuning can be a very computationally expensive process Bischl et al. (2021).
- *Ethical concerns*: related to data and algorithm bias. Moreover, transparency is also an issue, as it is typically considered as black box due to the limitation in its understanding (Mbiazi et al., 2023). Additionally, in applications such as autonomous and health care may face moral dilemmas.
- *Privacy*: AI still fails to protect sensitive data, e.g. for instance by cyber-attack (Mbiazi et al., 2023). Therefore, there is still skepticism in the society when using it. At the same time, people still argue about the security and confidence of using it.

A summary of AI aspects and related challenges is provided in Table 2.

Table 2. AI aspects and related limitations and challenges

AI aspect	Limitations and challenges
HPC	- Expensive computing facilities
Data	- Need vast amount
Model	- Million of parameters modelled
Hyper-parameter	- Optimization is also computacionally expensive
Ethical *Transparency*	- related to data and algorithm bias - AI most of the times seen as a black box

3. AI-driven Digital Twin conceptual framework

As represented in Figure 2 the proposed AI-driven Digital Twin concept consists of four main building blocks that are described in the following: (i) Digital Twin Core Engine, (ii) Hybrid Models, (iii) Geo- and Remote Sensing Data and (iv) Artificial Intelligence (AI).

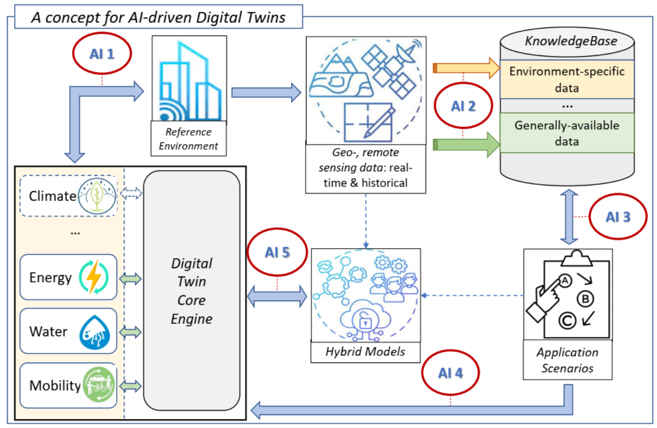

Figure 2. Digital Twin and AI conceptual framework.

Digital Twin Core Engine (DTCE). The DTCE represent the "core platform" which models areas/critical infrastructures (e.g. Energy, water and mobility/transport) of the Reference Environment (e.g. a urban district, a city, a region). It enables the simulation and analysis of functional and dysfunctional behaviors, such as the occurrence of power outages, water shortages and traffic disruptions. Since the real world is not linear but consists of complex relationships and interactions, a dependency model is part of it, thus making it possible to represent emergent behavior and to investigate cascade effects within and between the areas. This models both structural and behavioral aspects for a precise analysis of the Reference Environmentt. The structural part includes physical elements such as topography, land cover, infrastructure and buildings. The behavioral part includes the interactions between these elements, such as traffic flows, water distribution, energy consumption. Since human behavior influences and is influenced by urban systems, socio-technical aspects are also part of it (e.g. human-in-the-loop). By integrating system structure and behavior modeling as well as modeling socio-technical aspects, emergent behavior and cascade effects are also captured, which are fundamental to informed decision-making processes for safeguarding and protecting critical infrastructures.

Geo- and remote sensing data. Historical and real-time geo- and remote sensing data are needed and to be combined, by distinguishing two main groups of data: (i) Environment specific data: use of data that environment owns and includes data obtained by IoT devices and local sensors, which not only serve as interfaces for data access and exchange, but are also equipped with local intelligence and computing capabilities within the edge intelligence paradigm; (ii) Generally available data: use of generally accessible sources such as online services, satellites, online social media and crowdsourced impact assessment information, remote sensing data, ground-based observations that are dynamically obtained and updated by configuring specific system parameters in relation to the addressed area. Such data categories form the so-called Knowledge Base (KB), which is used to instantiate specific application scenarios of interest and feed the DTCE.

Hybrid Models. Hybrid models are need to obtain insights (e.g. temporal and spatial distribution of threats events) as well as to simulate rare emergency-related phenomena and crises and thus support smart planning and decision-making processes for local measures. Hybrid models can, among other things, simulate threats events in detail, assess their impacts on infrastructure and support decision-makers by optimizing strategies by integrating diverse data sources and modeled features. Examples of such models can include: "Physical models" to analyse physical processes of the reference environment; "Agent-based models" to create a virtual replica to simulate the behaviour of individual entities (including people and vehicles) and their interaction within the urban environment and to study their emergent behaviour in relation to normal and emergency situations; "Scenario-based models" to explore and evaluate possible future outcomes by considering a range of different scenarios or conditions.

Artificial Intelligence (AI). Figure 2 shows *AI* interfaces between the main DT building blocks. Every aspect may include one (or more) application. In the following, the idea on how this interaction happens is presented:

- AI 1: This interface is the interplay between the data collected from the different CIs and the *Reference Environment*. This implies that the data has been collected and the implementation of the AI algorithms has taken place. This offers the possibility of *(quasi) real-time analysis*.
- AI 2: The AI interface between *Geo- and remote sensing data* and *knowledge based* might bring *predictive analysis*, this includes statistical analysis, implementation of machine and deep learning to analyze current and historical data in order to make predictions such as forcasting to make prediction in future events. This connector includes algorithm training and model evaluation.
- AI 3: The AI interface between *knowledge based* and *application scenarios*. As an example, some of these scenarios include: healthcare, transportation and logistics (e.g. autonomous driving), energy, etc. This AI link aims to grant the security of AI implementations to avoid reduction on the functionality. To name some attack: cyber-attack, data poisoning (e.g. label flipping), model stealing, etc.
- AI 4: The AI interface between *application scenarios* and *Hybrid model* is about the integration and combination of, for instance, machine learning with simulation models to improve the accuracy, robustness, and adaptability. All in all, to construct a more generalised model.
- AI 5: The AI interface between *Hybrid model* and *DTCE* constitutes a key component to enhance the Digital Twin model. AI can help automating the data collection, data cleaning, and data fusion between them both. Methods such as transfer learning and Generative AI. Some applications are: object detection, text classification.

4. Enhancing critical infrastructures safeguard through DTs and AI: scenarios, benefits and impact

The integration of AI and DTs offers a robust approach to mitigate risks, enhance operational efficiency, and ensure the continuity of essential services. Digital twins provide real-time virtual replicas of physical assets, enabling continuous monitoring and predictive maintenance. AI algorithms analyze this data to detect anomalies, predict failures, and optimize performance, reducing downtime and preventing catastrophic failures. This proactive approach is crucial for critical infrastructures, where unanticipated disruptions can have widespread consequences. For instance, in power generation, AI-driven digital twins can forecast equipment failures, optimize energy production, and monitoring electricity demand (Gebhard et al., 2022). In transportation management, they can enhance traffic flow, predict maintenance needs for infrastructure, and improve public transportation schedules (Bamunuarachchi et al., 2021). In healthcare, they can monitor hospital equipment, optimize resource allocation, provide clinical decision support or personalised health forecasting (Viceconti et al., 2024).

Terror Defence. In the terror defense sector, AI-driven digital twins are being utilized to enhance security measures, optimize response strategies, and protect critical infrastructures against terrorist threats. By creating virtual replicas of physical environments, security systems, and operational procedures, digital twins enable real-time monitoring, predictive analytics, and adaptive responses. By integrating data from surveillance cameras, social media, and intelligence reports, the digital twin create dynamic model of key areas in the city, to enhancing counter-terrorism efforts (Tundis et al., 2023, 2021a,b). Such approach can be, in turn applied for protecting different critical infrastructures, including airports, seaports, and government buildings, by customizing it on the basis of data from specific surveillance systems, integrating AI-based cyber intelligence, and other operational procedures. All in all such integration can support proactive threats detection, efficient resource allocation, and strategic decision-making to ensure the safety and security of key assets and populations (see Table 3).

Climate Change and Adaptation. In the face of escalating climate change challenges, AI-driven digital twins have emerged as powerful tools for enhancing climate resilience and adaptation. By creating highly detailed virtual replicas of natural environments and urban areas, digital twins can simulate and predict the impact of various climate scenarios, enabling proactive and informed decision-making to mitigate risks and adapt to changes. Copenhagen has implemented a comprehensive digital twin to enhance its climate resilience, particularly in managing flood risks. The DT integrates data from various sources, including weather forecasts, topographic maps, and drainage systems, to create a dynamic model of the city (see Table 4).

Urban Planning. In urban area, a comprehensive digital twin can be used to manage and optimize urban planning. In this case, the integration of data from various sources, including IoT devices, can allow the simulation and prediction of the impact of new developments, infrastructure changes, and population growth. Moreover, the use of AI can enhance the analyze of this data, in order to make informed decisions thus enhancing urban resilience and sustainability (see Table 5).

Healthcare. In the healthcare sector, AI-driven digital twins are revolutionizing patient care, hospital management, and medical research. By creating virtual replicas of patients, medical devices, and healthcare facilities, digital twins enable personalized medicine, optimize resource allocation, and enhance clinical outcomes. This integration supports real-time monitoring, predictive analytics, and adaptive responses, ensuring more efficient and effective healthcare delivery (see Table 6).

Transportation Management. Urban transportation networks can be supported by digital twins to optimize traffic flow and reduce congestion. Furthermore, by simulating various traffic scenarios and using AI to analyze data from sensors and cameras, traffic management can be improved, reduced travel times can be reduced, as well as emissions can be decreased. This has not only enhanced the efficiency of transportation systems but also improved the quality of life for urban residents (see Table 7).

Power Generation. In the energy sector, AI-driven DTs can be implemented for their turbines. These DTs monitor the performance of turbines in real-time, predict maintenance needs, and optimize operations, leading to a significant reduction in unplanned outages and maintenance costs. This can result in increasing efficiency and reliability of power generation systems (see Table 8).

These case studies from terror defence, climate change adaption, urban planning, healthcare, transportation management, and power generation illustrate the tangible benefits and transformative potential impacts of integrating AI and digital twins across various sectors. By enabling predictive maintenance, optimizing operations, and enhancing efficiency, AI-driven digital twins serve as intelligent guardians of critical infrastructures and citizens. They not only mitigate risks and improve dependability but also contribute to sustainability and better quality of life. These examples highlight how the strategic deployment of these technologies can revolutionize the management and protection of essential services, ensuring their resilience in an increasingly complex and dynamic world.

5. Discussion on AI implications in DTs for CIs

Ethical, regulatory, data quality, and societal implications are important aspects to be taken into account, especially when associated with the deployment of AI-driven digital twins in critical infrastructures. Some of them include bias and transparency as discussed in Sect. 2.2 (see also Mbiazi et al., 2023).

Trustworthiness is another critical factor, as it ensures that the information and simulations provided by DTs are accurate and can be depended upon for making crucial decisions. This involves safeguarding the data collected from various sensors against tampering and ensuring its accuracy and authenticity. Moreover, robust cybersecurity measures are essential to protect DT systems from unauthorized access and cyberattacks, which could compromise their functionality and the physical infrastructure they represent. The reliability of the DT's predictive analytics and real-time monitoring capabilities is also vital, as stakeholders rely on these aspects to maintain and optimize infrastructure operations efficiently. Not fulfilling such factors might have a negative impact on the performance, reliability and quality of the AI and its application to DTs

The data used to feed the AI model has a decisive impact on the quality of the AI systems. Degradation of the *data quality* might have a high impact safeguarding critical systems. *Data governance* is another key factor to consider. It aims to provide a formal structure for data management, ensuring that data is treated as a valuable asset and used effectively (Jernite et al., 2022). Ensuring data provenance is important for the integrity of digital systems. All in all, these aspects in turn lay the foundation for responsible AI in the society and thus of the related DTs.

Moreover, promoting innovation in the field of AI and DTs also requires to consider other factors such as data privacy, regulatory frameworks and stakeholder engagement. For example:

- *Improving Data Privacy* help the relevant parties and stakeholders to gain trust in using the technology. This concept is known as *Trust building* (Yang et al., 2024).
- *Law-compliant technologies* are essential. For instance by guaranteeing GDPR compliant technologies for the protection of personal data, as well as the control access to authorised personnel, would promote its diffusion. Not doing so, might also bring to break the law impacting *regulatory frameworks* (Hacker, 2023).
- *Stakesholder perspective.* As for stakeholders, a collaborative approach from the different involved parties is important. A good communication and transparency in the collaboration leads to a more robust outcome.

Table 3. Benefits and impacts of integrating AI-driven digital twins in Terror Defence

Benefits	Impact	Role of Digital Twins & Role of AI
Continuous Monitoring	- Real-time situational awareness - Quick identification of suspicious activities	Integrate data from various sources for real-time monitoring & Process data from cameras, sensors, networks to detect threats.
Proactive Threat Detection	- Prevention of terrorist attacks - Increased public safety	Create dynamic models of critical areas & Analyze data from surveillance and social media to identify threats.
Enhanced Situational Awareness	- Timely detection of threats - Quicker intervention and mitigation	Monitor critical infrastructures in real-time & Analyze data from surveillance and cyber intelligence to detect threats.
Improved Crisis Management	- Effective coordination during incidents - Reduced response times and damage	Simulate crisis scenarios for strategy development & Refine response strategies and ensure readiness.
Strengthened Infrastructure	- Reduced vulnerabilities - Increased resilience against attacks	Assess vulnerabilities and suggest improvements & Identify weak points and recommend fortifications.

Table 4. Benefits and impacts of integrating AI-driven digital twins in Climate Change

Benefits	Impact	Role of Digital Twins & Role of AI
Flood management	- Better flood prevention and response strategies - Reduced damage from flooding events	Model flood scenarios and drainage system & Analyze weather and environmental data to predict floods optimize responses.
Heatwave adaptation	- Mitigation of urban heat island effect - Improved public health outcomes during heatwaves	Model urban heat distribution and hotspots & Analyze data to recommend cooling strategies and predict health impacts.
Water resource management	- Sustainable water usage practices - Improved resilience to water scarceness conditions	Model water systems and usage patterns & Predict water availability and optimize management practices.

Table 5. Benefits and impacts of integrating AI-driven digital twins in Urban Planning

Benefits	Impact	Role of Digital Twins & Role of AI
Urban planning and development	- Informed decisions on land use and infrastructure - Sustainable urban growth and development	Create detailed models of the urban environment & Analyze data to predict and plan for future growth.
Disaster management	- Effective emergency response strategies - Enhanced resilience to natural disasters	Simulate various disaster scenarios & Analyze data to develop and optimize response strategies.
Resource management	- Efficient use of water, energy, and waste resources - Reduced operational costs and environmental impact	Model resource usage and infrastructure systems & resource allocation and usage through predictive analytics.

Table 6. Benefits and impacts of integrating AI-driven digital twins in Healthcare

Benefits	Impact	Role of Digital Twins & Role of AI
Enhanced Decision Support	- Improved accuracy of clinical decisions - Faster response times in critical situations	Integrate data from EHRs, wearable devices, and hospital systems & provide real-time decision support and predictive analytics.
Reduced Equipment Downtime	- Continuous availability of critical diagnostic tools - Reduced operational costs	Create digital twins of medical imaging equipment & predict maintenance needs and potential failures.
Timely Interventions	- Improved patient outcomes - Reduced hospital admissions	Model patient monitoring systems with real-time data & Detection of early signs of health deterioration and recommend interventions.
Enhanced Remote Monitoring	- Improved management of chronic diseases - Reduced need for hospital visits	Create digital twins of patients for continuous remote monitoring & analyze data from wearable devices and sensors.

Table 7. Benefits and impacts of integrating AI-driven digital twins in Traffic Management

Benefits	Impact	Role of Digital Twins & Role of AI
Traffic optimization	- Reduced congestion and travel times - Improved traffic flow and reduced emissions	Model traffic networks and simulate various scenarios & Analyze traffic data to optimize signal timings and flow.
Predictive analytics	- Better management of road closures and accidents - Minimization of disruptions in traffic systems	Simulate impact of accidents and closures on traffic & Predict traffic patterns and suggest management strategies.
Public transportation efficiency	- Enhanced scheduling and routing - Increased reliability of public transport services	Model public transportation systems and demand & Optimize routes and schedules based on real-time data.

Table 8. Benefits and impacts of integrating AI-driven digital twins in Power Generation

Benefits	Impact	Role of Digital Twins & Role of AI
Predictive maintenance	- Reduced unplanned outages and maintenance costs - Improved reliability of power generation equipment	Create virtual models of for real-time monitoring and simulation & Analyze data to predict maintenance needs and detect anomalies.
Operational efficiency	- Increased energy production - Cost savings due to optimized operations	Simulate different operating conditions to optimize performance & Optimize operations through data-driven insights.
Risk mitigation	- Early detection of potential issues - Prevention of catastrophic failures	Continuously monitor equipment health and performance & Detect potential failures and suggest preventative actions.

Therefore, to continue helping AI the grow, it is relevant to be transparent and to be ethically correct. As for the future, some strategies to achieve these goals should include to incorporate legal aspects and ethics *by-design*. For instance, the team could be constituted by people with different expertise, and so including lawyers which could take of legal actions. The trend is that AI research institutes and companies are having a team: ethical, legal and social aspects research (Eliot, 2020, ELSA). Moreover, the working groups should try to improve the performance of AI systems not only by monitoring the AI components but also by providing feedback. This can help to improve the performance and robustness.

6. Conclusion and future works

The paper has focused on Critical Infrastructures and the increase of their protection and safeguarding through Digital Twins-based models and Artificial Intelligence-driven approaches.

At first, the fundamental concepts and main limitations of digital twins and artificial intelligence regarding critical infrastructures when considered in isolation were introduced. After that, it has been discussed how DTs enable real-time monitoring, predictive maintenance, risk mitigation, and resource optimization across various sectors such as defence, climate change, energy, transportation, healthcare, and related infrastructures. Furthermore, it has been argued how the integration of AI could enhance these capabilities and their resilience by improving data integration and management, increasing predictive accuracy, optimizing simulation and scenario planning, and providing robust cyber-security measures.

Based on this, a conceptual framework based on four building blocks - *Digital Twin Core Engine*, *Hybrid Models*, *Geo- and Remote Sensing Data* and *Artificial Intelligence* - has been proposed, and specific scenarios have been described by pointing out the combination of DTs and AI, as well as their role, benefits and impact. Moreover, data governance and transparency to protect sensitive information topics, as well as ensuring data provenance, which is crucial for the integrity and trustworthiness of data in digital systems centered on AI have been briefly discussed.

Future work could focus (i) on creating standardized frameworks and interoperability protocols for DTs and AI systems to improve data integration and collaboration across sectors; as well as (ii) on exploring and deepening ethical, regulatory, and societal implications of these technologies, with an emphasis on data privacy, bias on algorithms, and decision-making transparency, to ensure responsible and ethical deployment.

Funding

The research activities related to this work are being conducted in the context of "urbanModel", "MoDa" and "Digitaler Atlas 2.0" projects that are funded from the German Aerospace Center (DLR) e.V.

Acknowledgements

This activity has been carried out in cooperation with the LOEWE Zentrum emergenCITY.

References

Abdulova, E. and Kalashnikov, A. (2021). Some aspects of critical information infrastructure risk management. In *2021 14th International Conference Management of large-scale system development (MLSD)*, pages 1–5.

Alpaydin, E. (2020). *Introduction to Machine Learning, fourth edition*. Adaptive Computation and Machine Learning series. MIT Press.

Bamunuarachchi, D., Georgakopoulos, D., Banerjee, A., and Jayaraman, P. P. (2021). Digital twins supporting efficient digital industrial transformation. *Sensors*, 21(20).

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., and Lindauer, M. (2021). Hyperparameter optimization: Foundations, algorithms, best practices and open challenges.

Buccafurri, F., De Angelis, V., Idone, M. F., Labrini, C., and Lazzaro, S. (2022). Achieving sender anonymity in tor against the global passive adversary. *Applied Sciences*, 12(1).

Buccafurri, F., De Angelis, V., and Lazzaro, S. (2023). Mqtt-a: A broker-bridging p2p architecture to achieve anonymity in mqtt. *IEEE Internet of Things Journal*, 10(17):15443–15463.

Chen, J., Seng, K. P., Smith, J., and Ang, L.-M. (2024). Situation awareness in ai-based technologies and multimodal systems: Architectures, challenges and applications. *IEEE Access*, 12:88779–88818.

Eliot, L. (2020). Ai and legal argumentation: Aligning the autonomous levels of ai legal reasoning.

Ferrag, M. A., Alwahedi, F., Battah, A., Cherif, B., Mechri, A., and Tihanyi, N. (2024). Generative ai and large language models for cyber security: All insights you need.

Gebhard, T., Brucherseifer, E., and Steinke, F. (2022). Monitoring electricity demand synchronization using copulas. In *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks.

Hacker, P. (2023). Ai regulation in europe: From the ai act to future regulatory challenges.

Herrmann, T. (2014). Revisiting socio-technical system design. In *Proceedings of the 2014 European Conference on Cognitive Ergonomics*, ECCE '14, page 1, New York, NY, USA. Association for Computing Machinery.

Hussain, M. (2024). Yolov5, yolov8 and yolov10: The go-to detectors for real-time vision.

Jernite, Y., Nguyen, H., Biderman, S., Rogers, A., Masoud, M., Danchev, V., Tan, S., Luccioni, A. S., Subramani, N., Johnson, I., Dupont, G., Dodge, J., Lo, K., Talat, Z., Radev, D., Gokaslan, A., Nikpoor, S., Henderson, P., Bommasani, R., and Mitchell, M. (2022). Data governance in the age of large-scale data-driven language technology. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM.

Kharchenko, V., Illiashenko, O., Morozova, O., and Sokolov, S. (2020). Combination of digital twin and artificial intelligence in manufacturing using industrial iot. In *2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, pages 196–201.

Kirillov, A., He, K., Girshick, R., Rother, C., and Dollár, P. (2019). Panoptic segmentation.

Koessler, L. and Schuett, J. (2023). Risk assessment at agi companies: A review of popular risk assessment techniques from other safety-critical industries.

Kumar, N. and Prabhansu (2023). 1 - artificial intelligence and machine learning applications in energy storage system: technology overview and perspectives. In Prabhansu and Kumar, N., editors, *Emerging Trends in Energy Storage Systems and Industrial Applications*, pages 1–26. Academic Press.

Leppich, R., Borst, V., Lesch, V., and Kounev, S. (2024). Time series representation models.

Li, Y. (2018). Deep reinforcement learning: An overview.

López Díaz, M. and Tundis, A. (2023). Combining tapas and sumo towards crises management based on traffic data. In *22nd International Conference on Modeling and Applied Simulation, MAS 2023*, pages :01–:08.

Louppe, G. (2015). Understanding random forests: From theory to practice.

Marquard, M., López Díaz, M., and Tundis, A. (2024). Using travel demand modelling and traffic simulation for crisis management. *Transport Transitions: Advancing Sustainable and Inclusive Mobility – Proc. of the 10th Transport Research Arena (TRA) Conference, Dublin (Ireland), 15-18 April*.

Masi, M., Sellitto, G., Aranha, H., and Pavleska, T. (2023). Securing critical infrastructures with a cybersecurity digital twin. *Software and Systems Modeling*, 22:1–19.

Mbiazi, D., Bhange, M., Babaei, M., Sheth, I., and Kenfack, P. J. (2023). Survey on ai ethics: A socio-technical perspective.

Pastorek, A. and Tundis, A. (2024). Navigating the landscape of iot security and associated risks in critical infrastructures. In *Proceedings of the 19th International Conference on Availability, Reliability and Security*, ARES '24, New York, NY, USA. Association for Computing Machinery.

Sattler, B. J., Stadler, J., Tundis, A., Friesen, J., and Pelz, P. F. (2023). A framework for the simulation-based selection of social models for socio-technical models of infrastructures using technical requirements analysis. In *Proceedings of the International Conference on Modeling and Applied Simulation, MAS 2023*.

Savaglio, C., Barbuto, V., Awan, F. M., Minerva, R., Crespi, N., and Fortino, G. (2023). Opportunistic digital twin: an edge intelligence enabler for smart city. *ACM Trans. Sen. Netw.*

Tundis, A., Melnik, M., Naveed, H., and Mühlhäuser, M. (2021a). A social media-based over layer on the edge for handling emergency-related events. *Computers Electrical Engineering*, 96:107570.

Tundis, A., Shams, A. A., and Mühlhäuser, M. (2023). From the detection towards a pyramidal classification of terrorist propaganda. *Journal of Information Security and Applications*, 79:103646.

Tundis, A., Uzair, M., and Mühlhäuser, M. (2021b). An iot-based context-aware model for danger situations detection. *Computers Electrical Engineering*, 96:107571.

Viceconti, M., De Vos, M., Mellone, S., and Geris, L. (2024). Position paper from the digital twins in healthcare to the virtual human twin: A moon-shot project for digital health research. *IEEE Journal of Biomedical and Health Informatics*, 28(1):491–501.

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., and Ding, G. (2024). Yolov10: Real-time end-to-end object detection.

Yan, H. and Li, Y. (2023). A survey of generative ai for intelligent transportation systems.

Yang, L., Tian, M., Xin, D., Cheng, Q., and Zheng, J. (2024). Ai-driven anonymization: Protecting personal data privacy while leveraging machine learning.

Yigit, Y., Ferrag, M. A., Sarker, I. H., Maglaras, L. A., Chrysoulas, C., Moradpoor, N., and Janicke, H. (2024). Critical infrastructure protection: Generative ai, challenges, and opportunities.

Zibulewsky, J. (2001). Defining disaster: The emergency department perspective. *Baylor University Medical Center Proceedings*, 14(2):144–149. PMID: 16369605.

Zou, H., Kim, Z. M., and Kang, D. (2023). A survey of diffusion models in natural language processing.