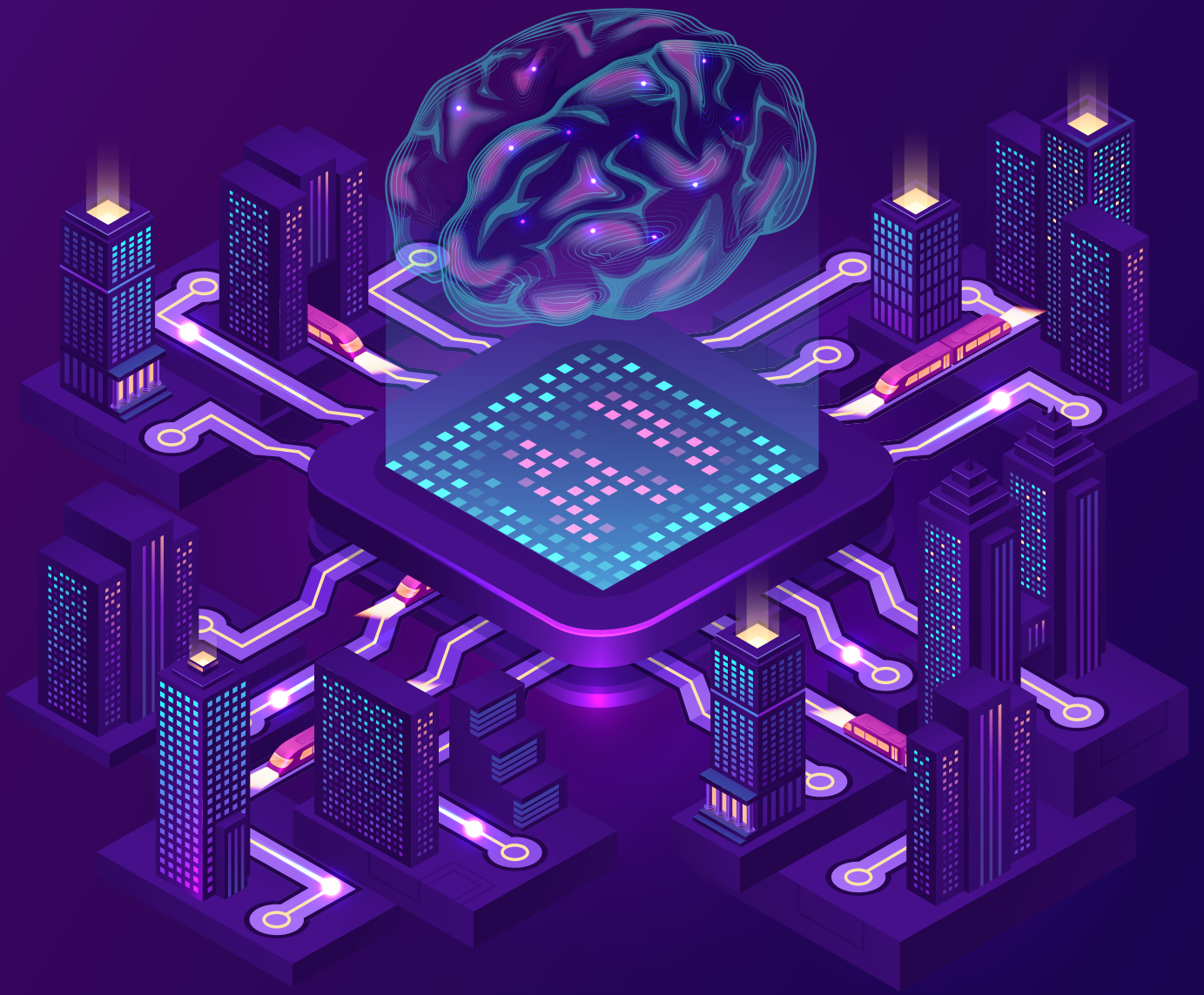
Trustworthy Digital Twins in Intelligent Transport Systems



The world is how we shape it

sopra steria

Contents

1. Introduction

This report has been written by Sopra Steria as part of an ongoing commitment to providing thought leadership in the ethics of Digital Twins. It is a companion report to 'Digital Twins: Ethics and the Gemini Principles' written by Sopra Steria and published by the National Digital Twin programme in December 2021.

The purpose of the report is to provide an ethics by design approach to developing digital twins. Such an approach **"Trustworthy Artificial Intelligence (AI)"** ethics to the development and operation of digital twins. This will help to mitigate a range of ethical issues which may otherwise not have been considered by the developers or users of the digital twin. It does not claim to be comprehensive, as the ethical issues surrounding digital twins will depend in part on the context of development and use. It is, however, a start to help developers, users, and policy makers to think through the ethical issues arising around digital twins.

This report opens with a list of values and sub-values for Trustworthy AI that should be preserved or respected (Section 1). It then proposes a conceptual map based on three phases of a digital twin (Section 2). The map allows for a more thorough examination of the ethical issues related to the use of a digital twin in Intelligent Transportation Systems (ITS). Section 3 identifies ethical issues related to each phase of digital twin development, based on values related to trustworthy AI. The report concludes (Section 4) with recommendations for future discussions on trustworthy digital twins in ITS and transparency by design of user interfaces.

2. Requirements for Ethical Digital Twins

> This report draws on Sopra Steria's seven categories of digital ethics, drawn from academic and industry standards. When applied, these should establish a relationship of trust with Digital Twins and should cultivate trustworthy Digital Twins[1]. Regarding this, this section provides list of values and sub-values that should be promoted in and by Artificial Intelligence in the creation of Digital Twins.

The following is a list of values or requirements for trustworthy Digital Twins:

Sopra Steria's 7 categories of Digital Ethics

- Societal Impact
- Displacement, skills & work
- Fairness, equality, diversity & accessibility
- Privacy
- Transparency
- Environmental sustainability
- Safety

3. Three phases of Digital Twins

To organise the identification of ethical issues that arise from Digital Twins in Intelligent Transport Systems, this report distinguishes three phases of Digital Twin development and use: 1) Resource Phase, 2) Representation Phase, and 3) Action Phase.

Resource Phase

After determination that an AI system is required for use in ITS, data sets are selected, pre-processed, and transformed to feed into an AI model[2].

In this stage, different data sets are selected in regard to the goal of the modelling (source data stage);

Collected data are pre-processed by removing noise, handling missing data fields, aggregating data, and coding data (pre-processed stage);

The features which are useful for achieving the goals of the task are founded with the help of dimensionality reduction or transformation methods (transformed data stage).

Representation Phase

In the second phase, the prepared datasets are used to train the model. Model development proceeds in an iterative cycle whereby different models and training data configurations are tested until a validated model is ready for analysis.

The first stage involves selecting computational modelling tools that match the goals of the process, which can be achieved, for example, via clustering or classification (computational modelling tools stage);

Different training data configurations are tested to identify and describe properties in clusters/classes (training environment stage);

The final stage involves searching for data patterns and relations in large databases using ML algorithms (model repositories stage).

Action Phase

Finally, in the third phase, identified patterns are interpreted, and corresponding actions are determined.

Extracted patterns and models, mostly statistics, are visualised and converted into intelligible information, such as graphs and tables (statistical analysis stage);

The discovered information (knowledge) is documented, implemented though a user interface, and corresponding actions are determined (determination of corresponding action stage);

In the case of autonomous digital twins, the discovered knowledge is communicated to an intelligent system, such as a robot, and then the interventions in the physical twin take place, without human interference (autonomous decision-making stage).

Figure 1 presents the various phases and the relevant stages of each phase.

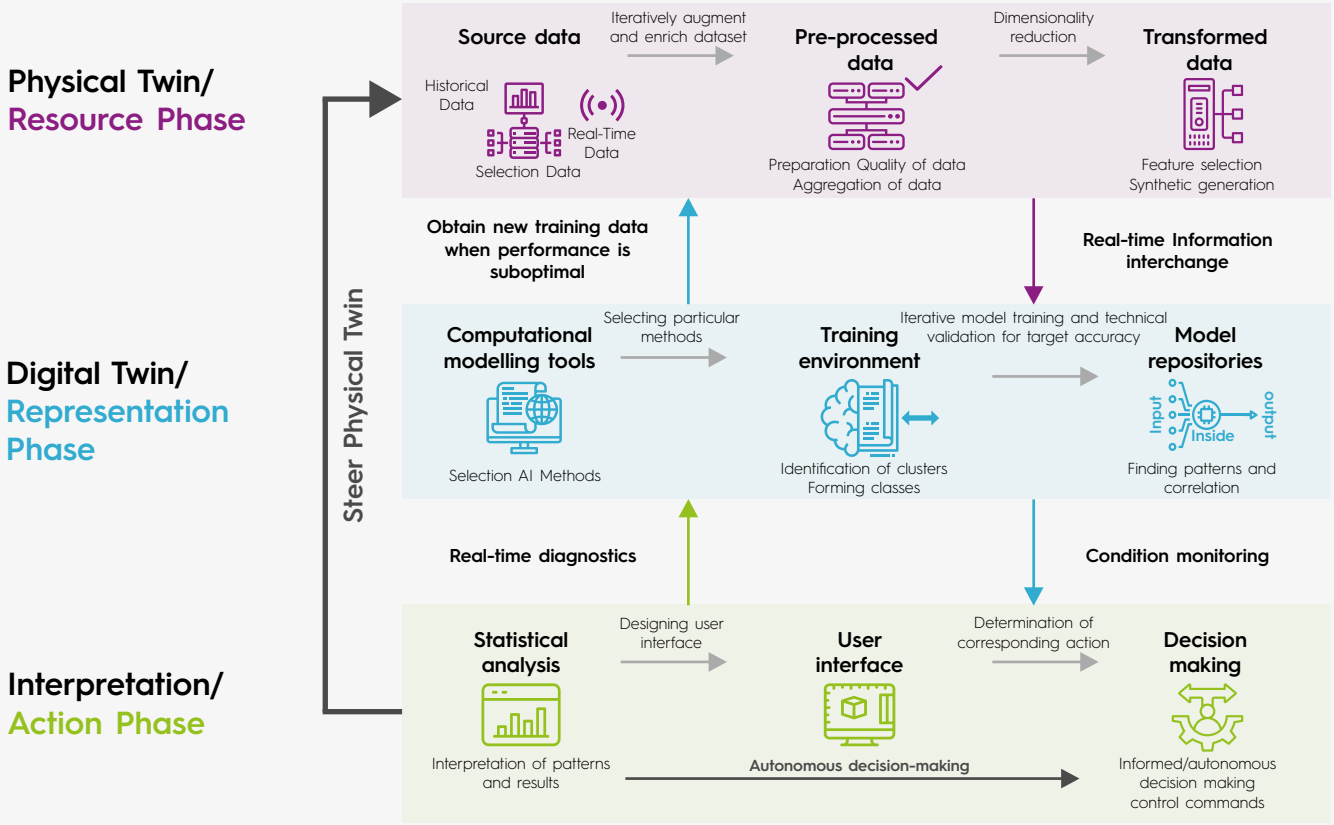


Figure 1. Conceptual diagram covering the whole Digital Twin development cycle

The following presents ethical issues of Digital Twins in ITS in the three phases:

Resource, which labels issues around collecting, selecting, combining and transforming data; Representation, which refers to concerns around how knowledge is created in and with digital twins; Action, for issues that have to do with the use of this knowledge to inform decision-making.

4. Identifying Ethical Issues of Digital Twins

Privacy, bias, security, function creep, data ownership, trust in data, and safety are among the concerns raised about the data (resources) that Digital Twins collect and use.

4.1 Resource Phase

A careful data selection process is required, as this will influence many aspects of the model learning process and performance, and is a major source of bias, privacy, security, function creep, and data ownership concerns. The collected data are then pre-processed by removing noise, handling missing data fields, aggregating, and coding, all of which raise concerns about the safety and quality of the data. In the final stage of the resource phase, issues such as trust in data and security arise.

4.1.1 Source data stage

First, in the source data stage, it is necessary to identify the data sources required for model building. In the case of digital twins, the operations of a physical twin can be digitized by data captured from remote or aerial sensing technologies such as satellites or drones, sensors embedded in vehicles and transport infrastructures, cameras, and mobile phone data generated in mobile communication networks[3]. However, these digital streams are not the only sources of data that can feed a digital twin. In addition to streaming data, accumulated historical data, such as data collected from urban maps, can inform a digital twin. Thus, real-time data and historical data can be selected and extracted from different sources.

Real-time data, particularly data which are generated in mobile communication networks have the potential to improve current travel demand models and, in general, to plan for better urban transport systems. However, people's patterns of movement in space and time are highly dimensional, making mobile phone data a potential proxy identifier for a single person. For instance, data with a temporal resolution of one hour and a spatial resolution equal to the density of cell towers have been shown to require just four spatio-temporal points to isolate and uniquely identify 95% of mobile phone users[4]. As a result, this type of information can reasonably be considered as personal information.

4.1.1.1 Privacy

Digital twins in ITS can raise question about privacy as data gathered tell something about identity, location and behaviour of persons.

a. Identity privacy in relation to ITS refers to the privacy of a driver, traveller, passenger, or pedestrian. This can take the form of their first and last name, driver's license number, car registration number, etc.

b. Behavioural privacy refers to the privacy of a group of individuals and their actions within the ITS. As the ITS collects information on travellers' routing patterns to make the routes safer and more efficient, movement patterns of individual travellers are also recorded in the system, which can provide information about the behaviour of those individuals. For example, the origin and destination points of individual travellers can lead to privacy issues as these can enable a malicious actor to infer the residence or workplace of a traveller.

c. Location privacy within an ITS would be classified as privacy of location and space, or the right of a user to travel or move about the system without concern of their location information being exposed. While precise location information is beneficial for ITS to provide location-aware services, such information can also be used to invade the privacy of individuals[5]. As with behavioural data, locational data can reveal private information about an individual, such as their home, workplace, or place of worship**.**

Different approaches have been developed to preserve privacy, such as anonymisation, pseudonymisation, differential privacy and homomorphic encryption.

a. Anonymisation to preserve identity privacy: anonymisation of data is a process that occurs in-house, which means that the names of data subjects are erased. However, due to its high dimensionality, even if data are anonymised there still exists the possibility of re-identifying users behind mobile phone traces. For example, linking anonymised data with other identified data increases the chance of re-identification. That is, it is possible to identify an individual in a pooled or aggregated dataset. With the increase in number and size of datasets, the risk of re-identification raises privacy risks to travellers whose data are collected and stored[6].

b. Homomorphic encryption is another approach used to protect location privacy. However, it is extremely challenging for GPS-based navigation system to provide services while also preserving the location privacy of users. It is crucial is to find a balance between providing beneficial and accurate services to users while also preserving location privacy[7].

c. Differential privacy can be used to preserve the behavioural privacy of ITS users. The goal of differential privacy is to preserve privacy by furnishing ways to maximize the accuracy of queries from statistical databases while minimizing the probability of identifying its records. However, differential privacy faces challenges when being used across recurrent or time-series data[8].

4.1.1.2 Bias

Digital twins raise questions about which data sets are generated or chosen as inputs for the twin, which data sets are combined, and how they are made available to generate new knowledge or considerations about actions. A careful data selection process is required because it influences many aspects of the model learning process and subsequent performance while also being the primary source of bias. It is crucial to understand how the data have been gathered because there are sources of imprecision or bias associated with data collection and selection[9].

Bias in data collection, which can present itself as an under - or over-representation of specific groups in the dataset, can cause additional issues in the training data. While under-representation might result in unfair or unequal treatment, over-representation might result in a **"disproportioned attention to a protected class group, and the increased scrutiny may lead to a higher probability of observing a target transgression"[10].** Those who select data should ensure that those data are properly representative, relevant, accurate, and used to form generalisable datasets.

There are at least two sources of potential bias in a Digital Twin used in ITS:

a. **Bias in mobile phone data: there may be a bias in mobile data samples for two reasons:**

   1. Mobile phone data are among the most applied and researched types of data in transport planning. Since mobile phone use is significant but not ubiquitous across any population, the data collected from mobile phones could bias against groups who do not use a cell phone. Current mobile phone penetration in the UK is 95%, meaning that approximately 3.5m people do not use a mobile phone[11].

   2. Different phone companies and owners have different rates of mobile phone use, which affects who is covered in a mobile phone-based sampling network.

b. **Bias in data collected from different modes of transport:** it is more likely that data are collected from motorised vehicles than bicycles and pedestrians. Data can be collected easily from vehicles using sensors built into the vehicle, such as GPS, toll tags, RFID tags, camera-readable license plates, and sensors carried by the driver, such as smartphones. A lack of equipment to collect data from other modes of transport, such as cycling or walking, other than mobile phone data, can lead to bias in transport data collected by different sources and agencies[12].

4.1.1.3 Security

In the source data stage, several security threats may arise. Some of the most common threats include the following:

a. **Data and identity theft:** unprotected infrastructures, such as surveillance systems, can be used to provide attackers with a large amount of data that can be used to steal personal information and identity theft.

b. **Insecure hardware:** sensors are the starting point for many attacks. If they are not tested appropriately, or sufficiently protected, they may pose major threats to the entire system[13].

4.1.1.4 Function creep

Concerns about function creep are raised by data misuse, particularly in the context of mobile traces. Function creep is a shift from a legitimate, justified use to one which is either illegitimate or has not been fully justified. In the case of surveillance, such a shift can involve a move from care to control.

Governments are interested in monitoring mobility, particularly in predicting, tracking, and preventing unauthorised migration flows to their borders. Mobile phone data used as real-time surveillance can help governments control migration, thus moving along the spectrum from care to control[14].

4.1.1.5 Data ownership/stewardship

The re-use of existing data is an important concern that is related to ownership and/or stewardship issues. As data that a digital twin uses may have been collected in the past for purposes other than the creation of that digital twin, questions may be raised about who is entitled to control this re-use of datasets, and who may benefit from the creation of the digital twin[15].

4.1.2 Pre-processed data stage

As stated, collected data are pre-processed by handling missing data fields or coding data. At this stage, there are concerns about the quality and safety of data in this stage.

4.1.2.1 Quality and safety of data

Following the collection and combination of data, data preparation is required for further refinements to improve the quality of data, such as normalization, filtering, missing value imputation, outlier detection, and/or harmonization[16]. Data preparation is a critical process for ensuring the quality[17] and integrity of data; without this, there is a risk that the results of the digital twin will not be well informed, which may lead to safety issues. Analysis of the data is required to ensure that quality is sufficient before beginning any model construction.

4.1.3 Transformed data stage

The features that are useful for achieving the goals of a digital twin are founded in the transformed data stage, using dimensionality reduction or transformation methods. Following that, the prepared data is transferred to the next phase, which is the representation phase. At this point, there are concerns about trust in data and security.

4.1.3.1 Trust in data

Transport use cases can benefit from synthetic training data[18]. Synthetic data generation employs novel AI technology to learn how to generate new data from existing real-world samples, with the goal of preserving the statistical and structural properties of the original data without risking a revelation of personal data, so helping to preserve privacy. However, the use of synthetic data raises potential trust concerns regarding the degree to which a data owner/steward can reasonably trust the synthetic data.

4.1.3.2 Security

Transformed data need to be transferred to the team creating the model. Because of this, security concerns may arise as attackers may inject malicious nodes between two communicating nodes/parties to steal or poison the data.

4.2 Representation phase

The issues that arise in the second (Representation) phase are related to whether, how, and what knowledge is created in and with digital twins. In addition to these epistemic concerns, model repositories raise concerns about uncertainty, explainability, accuracy, safety, and traceability. This phase consists of three stages: computational modelling, training environment, and model repositories.

Digital twins provide knowledge to help guide action. This may raise concerns about how digital twins mediate our knowledge-relationships with the world[19]. For example, a digital twin in an ITS may influence traffic engineers' perceptions of safety and efficiency in the transport system because roadways and vehicles are monitored over time. The ITS thus becomes available and accessible for engineers to think about the locations of congestion hotspots and peak hours for various routes. Furthermore, the digital twin may mediate researchers' knowledge about future travel demand and how this can be addressed. Interactions with real entities are thus replaced by interactions with their digital representations when digital twins are used. As a digital twin can represent actual states of objects and processes, engineers and researchers can monitor these objects and processes remotely rather than through direct observation. What matters is that when digital twins are available to investigate road safety, for example, there will be less interaction and engagement between traffic engineers, urban planners, and researchers, and the world around them. The substitution of the real world for the digital twin may result in a lower perception of travellers' or citizens' (social) well-being in favour of efficiency or other perceived values.

Moreover, we might question the type of knowledge digital twins provide about the world. How does a digital twin represent the real object or process it is twinning? Should a digital twin of a transport system be considered as a "representation" of the real system? How representative is the digital twin? What aspects does the digital twin represent? How should we make a distinction between "good" and "bad" digital twin representations of transport systems? Developers of digital twins make decisions about what the digital twin is to represent, and what aspects it should minimally cover, but it is not always transparent how they take these decisions and what vision of "good" and "bad" digital twins are assumed in their choices[20]. What aspects of the world do digital twins enhance and make visible about transport systems, and which disappear? These questions deserve attention when digital twins play a role in ITS.

4.2.1 Computational modelling stage

In general, machine learning algorithms are categorized based on their learning style: supervised or unsupervised. Supervised learning utilizes labelled data and classifies samples into two or more classes (classification) before determining the class to which any new samples belong. Unsupervised learning, by contrast, utilizes unlabelled data clusters (data clustering) to find structures in the data.

4.2.2 Training environment stage

Model training employs machine learning and should be thought of as an iterative process in which various model configurations, training data/variables, and training/validation schemes are tested and compared in order to achieve optimal results on the validation of (unseen) data.

4.2.3 Model development stage

Various issues arise during model development, including[21]:

A. Related to the input data:

4.2.3.1 Epistemic uncertainty

Uncertainty refers to the ability of the model to accurately classify or predict new observations after being trained on a limited set of data. The type of uncertainty that arises in this stage is epistemic uncertainty (uncertainty caused by a lack of knowledge (data) about the best model). It refers to the ignorance of the decision-maker and the use of an imperfect model of the problem. Additional information can help to reduce uncertainty caused by ignorance[22].

B. Related to the model internal structure:

4.2.3.2 Explainability

The interior of a model is usually considered to be a black box[23]. It is important to improve the explainability of models, particularly where these have a bearing on the public, safety considerations or on intellectual property. Explainability is the extent to which the internal mechanisms of a machine system can be explained in human terms. It is the ability to explain what is happening and why it is happening[24]. Human users can better comprehend and trust the output created by machine learning algorithms, and analyse the sensitivity of predictions by understanding how AI systems behave, when the system is explainable.

C. Related to the model outputs:

4.2.3.3 Aleatoric uncertainty

Uncertainty is inherently linked to predictive machine learning. Consider the case when a weather forecaster (probabilistic) component of a model can only provide probabilistic answers for two possible outcomes, but no definite answers. It is similar to what happens when a person flips a coin. If there is aleatoric uncertainty, then the outcome of the model will be necessarily uncertain: heads or tails. Additional information will not reduce this type of uncertainty[25].

4.2.3.4 Accuracy and safety

The accuracy or performance of the model is usually discussed in the output of the model and has an impact on safety. As a result, model outputs must avoid potentially dangerous situations by establishing additional recovery or safety mechanisms.

4.2.3.5 Reliability

Digital twins designed with the aim of future prediction, may struggle to be reliable. This is particularly true when a digital twin interacts with human-generated data. Past and current data may be used to train a digital twin. However, historical data may fail to account for changes in social behaviour, law, and institutions or governments, which can radically alter the model. As a result, a digital twin may be limited to representing the future based primarily on historical data. Hence the future presented by the digital twin will be a repeat of the past[26].

D. Related to the entire model development lifecycle:

4.2.3.6 Traceability of model configuration and training data in model lifecycle

To ensure the success of machine learning in representing reality and predicting the future, a mechanism that facilitates traceability of machine learning processes and outcomes across the entire AI/machine learning lifecycle should be established[27]. This will aid with explainability, trust and transparency of the system.

4.3 Action phase

Issues arising from the collection, combination, and use of data to make representations of the present and future world have been discussed above. This representation is supposed to provide knowledge about reality that can be used to make decisions, and so the knowledge should be actionable. As a result, the development of digital twins raises questions about the desirability and acceptability of the actions that they afford, recommend, or even take themselves[28]. Furthermore, the translation of model outputs into actionable insights raises questions about autonomy, accountability, responsibility, and transparency. This phase consists of three stages: statistical analysis, determination of corresponding action, and autonomous decision-making.

4.3.1 Statistical analysis stage

The patterns identified by machine learning are interpreted at this stage. Extracted patterns and models, mainly statistics, are visualised and converted into understandable information in the form of graphs and tables.

4.3.2 Determination of corresponding action stage

Corresponding actions are determined in this stage, based on the knowledge documented in the previous stage. Digital twins can be used for monitoring and predicting current and future states, but they can also prescribe or suggest certain actions to steer their physical counterpart, depending on their design.

4.3.2.1 Autonomy

Digital twins have an impact on transport managers, municipal councils, governors, and urban planners' actions and decisions. These technologies assist governments, municipalities, transport managers, and urban planners in making decisions about, for example, system performance, congestion and peak-hour locations, and travel demand. As a result, it is critical to investigate the effects of digital twins on the decisions and actions of decision-makers with regard to their autonomy and how digital twins change their (inter)actions. For example, the problem of automation bias is well documented with automated systems. Even when a digital twin makes a recommendation rather than a decision, and a human is kept in the loop, it is important that the person is aware of tendencies to trust the automated system over their own trained intuitions. Furthermore, an over-reliance on automated decision-support/decision-making can lead to a de-skilling of trained engineers, rendering them reliant on the automated system.

4.3.2.2 Transparency

A significant promise of digital twins is that they support or enhance the agency of human actors, allowing users to make better-informed decisions. Prescriptive digital twins recommend certain actions but give human actors the freedom to consider and choose which to take. A digital twin that prescribes actions for human users raises questions as to how it communicates and makes transparent its workings and the underlying processes that lead to a particular recommended action. In this regard, the design of the user interface that displays the output of the digital twin is critical. While some user interfaces make the steps and processes that lead to a recommended decision available for users to inspect and possibly disagree with, others do not. If digital twins are opaque and their workings are not transparent to their users, they risk becoming opaque black boxes[29] with a negative impact on trust and accountability.

4.3.2.3 Accountability

Designers of technologies, in this case user interfaces, should foster accountability by allowing systems to report, explain, and justify their decisions to users and other relevant actors. Being able to rely on a safe and sound design process that accounts for and reports on options, choices, and constraints related to the system's goals and assumptions is an important aspect of accountability[30]. Transparency by Design[31] of the user interface, through which the output of the digital twin is displayed to ensure accountable development of the system, is one solution to this challenge. Justifiability and interpretability are two aspects of accountability that can be covered by the transparency principles[32].

4.3.3 Autonomous decision-making stage

Digital twins can be used to monitor and predict current and future states, but they can also automatically intervene in their physical counterparts, depending on their design. Without human intervention, an autonomous digital twin will intervene in the physical object with which it is twinned. The most important issue that arises from autonomous digital twins is in relation to distributed responsibility[33].

4.3.3.1 Responsibility

An autonomous digital twin which is linked to its physical counterpart by smart devices that can intervene to automatically execute interventions in the physical twin requires no human intervention. Questions arise as to which kinds of decisions and actions should be delegated to a digital twin and the digital systems (such as robots) that surround them, and what role should human actors (continue to) play[34].

Delegating decisions and actions to digital twins raises concerns about the degree of autonomy of digital twins and the locus of responsibility for consequences. Who or what is responsible (and liable) for the actions and interventions that the digital twin supports or undertakes? If these actions have harmful consequences, who is responsible for the eventual damage and can be held accountable? Depending on the design of the digital twin, the designers and end-users may all be (partially) responsible. However, it is unclear who should carry what degree of responsibility and whether and how shared (or distributed) responsibility for the same consequences can be achieved[35].

5. Future discussions: Trustworthy Digital Twins in Intelligent Transport Systems

To develop trustworthy Digital Twins in ITS, future discussions should focus on how to address ethical issues and how to preserve and promote ethical values throughout the Digital Twin development cycle (Figure 2).

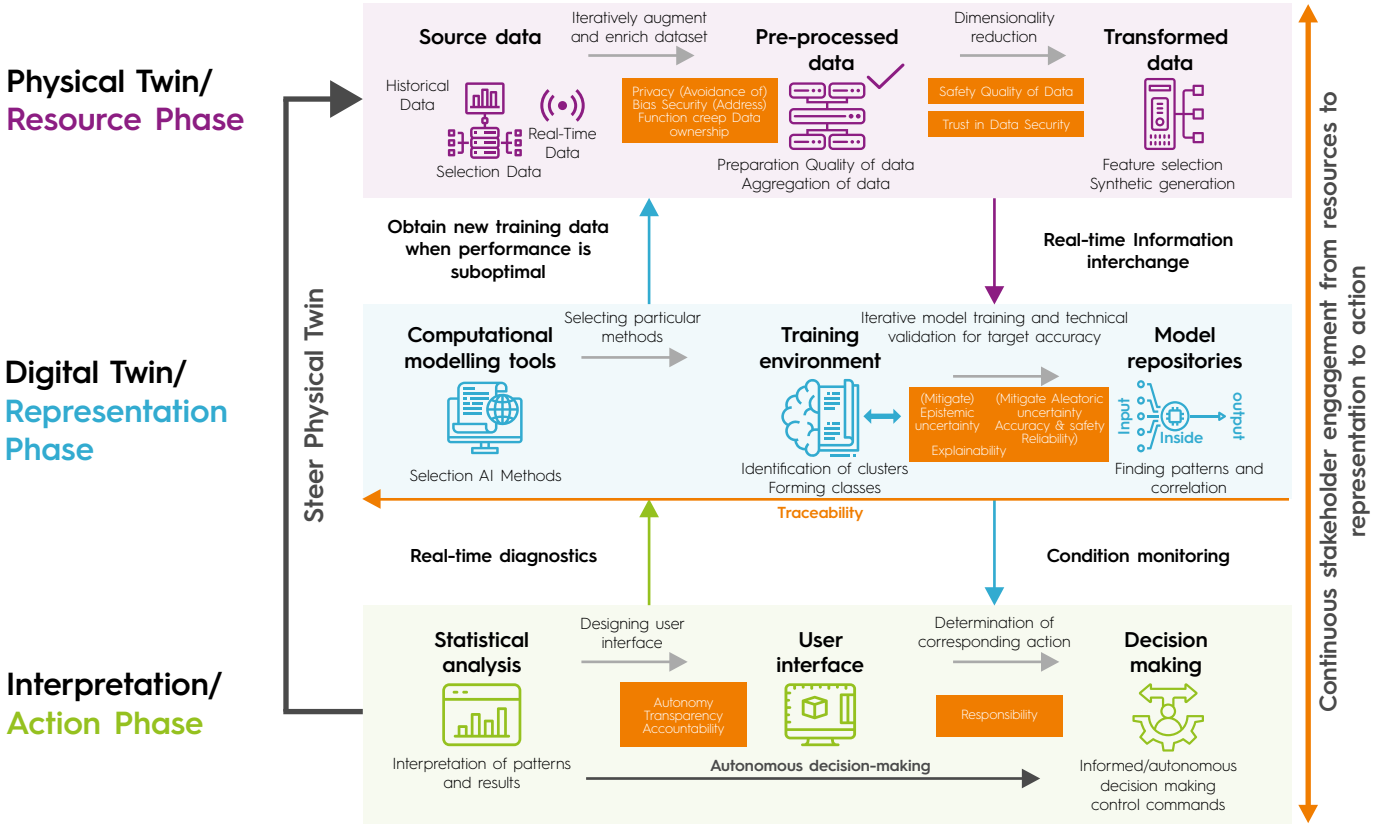


Figure 2. Requirements for Trustworthy Digital Twins in Intelligent Transport System

Since a digital twin will frequently have a societal impact, often by design, it is critical to investigate the needs of various types of end-users and (direct/indirect) stakeholders ahead of time to develop thought-out visions of where and how a digital twin can contribute in a way that stakeholders truly consider an improvement. To tailor a digital twin to stakeholder values and achieve a desired and valued digital twin, stakeholders should be involved in the development of the digital twin at all phases, from resource to representation to action. As a result, it is important to examine and analyse various approaches to engaging stakeholders in a project and find an appropriate way to involve (direct and indirect) stakeholders in the digital twin development cycle.

6. Conclusion

This report has set out an ethics by design approach to digital twins. This approach has been created to mitigate a number of potential ethical issues arising through the design, implementation and use of digital twins. The report has broken down the development of digital twins into three stages: resource, representation, and action. The most pressing ethical issues have been highlighted here for each stage. Finally, a call has been made for stakeholder engagement in the design and use of digital twins if they are to be accepted and trusted by the public.

7. Authors

This report was written for Sopra Steria by Haleh Asgarinia, PhD candidate at the University of Twente, with the support of Dr Kevin Macnish, Digital Ethics Consulting Manager at Sopra Steria.

Haleh's project is part of the Marie Skłodowska-Curie Innovative Training Network 'PROTECT - Protecting Personal Data Amidst Big Data Innovation', funded by the European Commission's Horizon 2020 programme. She was working with Sopra Steria on a three-month internship as a part of that programme.

Kevin has 12 years' experience as an academic and consultant in digital ethics and was a lead contributor to the EU SHERPA project on the ethics and human rights implications of AI. He has published over 40 academic articles, chapters and books in this sphere.

8. References

AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf. (2019). HLEG AI Ethics guidelines for trustworthy AI.

Anda, C., Ordonez Medina, S. A., & Axhausen, K. W. (2021). Synthesising digital twin travellers: Individual travel demand from aggregated mobile phone data. Transportation Research Part C: Emerging Technologies, 128, 103118. https://doi.org/10.1016/j.trc.2021.103118

Barrett, H., & Rose, D. C. (2020). Perceptions of the Fourth Agricultural Revolution: What's In, What's Out, and What Consequences are Anticipated? Sociologia Ruralis, n/a(n/a). https://doi.org/10.1111/soru.12324

d'Alessandro, B., O'Neil, C., and LaGatta, T. (2017). Conscientious classification: a data scientist's guide to discrimination-aware classification. Big Data 5(2), pp. 120–34.

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. Scientific Reports, 3, 1376. https://doi.org/10/msd

Dignum, V. (2019). Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer International Publishing. https://doi.org/10.1007/978-3-030-30371-6

Farsi, M., Daneshkhah, A., Hosseinian-Far, A., & Jahankhani, H. (Eds.). (2020). digital twin Technologies and Smart Cities. Springer International Publishing. https://doi.org/10.1007/978-3-030-18732-3

Griffin, G. P., Mulhall, M., Simek, C., & Riggs, W. W. (2020). Mitigating Bias in Big Data for Transportation. Journal of Big Data Analytics in Transportation, 2(1), 49–59. https://doi.org/10.1007/s42421-020-00013-0

Hahn, D., Munir, A., & Behzadan, V. (2019). Security and Privacy Issues in Intelligent Transportation Systems: Classification and Challenges. IEEE Intelligent Transportation Systems Magazine, PP, 1–1. https://doi.org/10.1109/MITS.2019.2898973

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Machine Learning, 110(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

Kargl, F., Friedman, A., & Boreli, R. (2013). Differential privacy in intelligent transportation systems. Proceedings of the Sixth ACM Conference on Security and Privacy in Wireless and Mobile Networks, 107–112. https://doi.org/10.1145/2462096.2462114

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. Zenodo. https://doi.org/10.5281/ZENODO.3240529

Lyon, D. (2008). Surveillance Society. http://www.festivaldeldiritto.it/2008/pdf/interventi/david_lyon.pdf

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate. Big Data & Society, In press. https://doi.org/10.1177/2053951716679679

Pasquale, F. (2016). The Black Box Society: The Secret Algorithms That Control Money and Information.

Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. Science and Engineering Ethics, 26(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

Taylor, L. (2016). No place to hide? The ethics and analytics of tracking mobility using mobile phone data. https://journals.sagepub.com/doi/abs/10.1177/0263775815608851?journalCode=epda

van der Burg, S., Kloppenburg, S., Kok, E. J., & van der Voort, M. (2021). Digital twins in agri-food: Societal and ethical themes and questions for further research. NJAS: Impact in Agricultural and Life Sciences, 93(1), 98–125. https://doi.org/10.1080/27685241.2021.1989269

Zhou, Y., & Zhang, D. (2019). Double Mix-Zone for Location Privacy in VANET. Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City, 322–327. https://doi.org/10.1145/3377170.3377250

References

1. https://soprasteria.turtl.co/story/digital-ethics/page/4?sfvrsn=544f38dc_2
2. (Fayyad et al., 1996)
3. (van der Burg et al., 2021)
4. (de Montjoye et al., 2013)
5. (Hahn et al., 2019)
6. (Anda et al., 2021)
7. (Zhou & Zhang, 2019)
8. (Kargl et al., 2013)
9. (SCALPEL-AI, HORIZON-CL4-2021-HUMAN-01-01 Type of Action: HORIZON-RIA, 2021)
10. (d'Alessandro et al., 2017)
11. https://www.statista.com/statistics/300378/mobile-phone-usage-in-the-uk/
12. (Griffin et al., 2020)
13. (Farsi et al., 2020)
14. (Taylor, 2016)
15. (van der Burg et al., 2020)
16. (SCALPEL-AI, HORIZON-CL4-2021-HUMAN-01-01 Type of Action: HORIZON-RIA, 2021)
17. (Epelde et al., 2020)
18. (Anda et al., 2021)
19. (van der Burg et al., 2020)
20. (van der Burg et al., 2020)
21. From the SCALPEL-AI proposal, I use the idea of approaching an AI model from three different perspectives: input, inside, and output.
22. (Hüllermeier & Waegeman, 2021)
23. (Belle & Papantonis, 2021)
24. (Richard Gall, 2021)
25. (Hüllermeier & Waegeman, 2021)
26. (Barrett & Rose, 2020)
27. (AIHLEG, 2019)
28. (van der Burg et al., 2020)
29. (Pasquale, 2016)
30. (Dignum, 2019)
31. assuming that AI is used to design user interfaces (UIs)
32. (Leslie, 2019)
33. (Mittelstadt et al., 2016)
34. (van der Burg et al., 2021)
35. (Mittelstadt et al., 2016; Ryan, 2020)