Emerging science, frontier technologies, and the SDGs

Perspectives from the UN system and science and technology communities

IATT Report for the STI Forum 2021

Interagency Task Team on Science, Technology and Innovation for the Sustainable Development Goals

Advance unedited report

Emerging science, frontier technologies, and the SDGs

Perspectives from the UN system and science and technology communities

IATT Report for the Multi-stakeholder Forum on Science, Technology and Innovation for the Sustainable Development Report 2021

New York, May 2021

Advanced unedited copy: This document is an advanced unedited copy. Following editing and layout a final version will be published.

Acknowledgments: The present report was produced by Work Stream 10 of the Interagency Task Team on STI for the SDGs under the leadership of Richard A. Roehrl (DESA) and Bob Bell Jr. (UNCTAD). It was edited by Richard A. Roehrl. Members of the team include: Fouad Mrad (ESCWA), Irmgard Nubler (ILO), Fernando Santiago Rodriguez and Hyunjoo Kim (UNIDO), Clovis Freire Junior and Michael Lim (UNCTAD), Preetam Maloor and Christophe Larouer (ITU), Susan Schneegans (UNESCO), Pierre Henri Boileau and Rachel Kosse (UNEP), Naoto Kanehira (World Bank). The team is also grateful for all anonymous peer-reviewers.

Important inputs for a preparatory expert group meeting were provided by Nebojsa Nakicenovic (TU Vienna), Guo Huadong (Chinese Academy of Sciences), Ada Yonath (Weizmann Institute of Science), Paulo Gadelha (Oswaldo Cruz Foundation, FIOCRUZ), Xiaolan Fu (University of Oxford), José Ramón López-Portillo Romano (Q Element Ltd.), Michiharu Nakamura (Japan Science and Technology Agency), Vaughan Turekian (US National Academies of Sciences, Engineering, and Medicine), Prof. Romain Murenzi (The World Academy of Sciences), E. William Colglazier (American Association for the Advancement of Science), Myrna Cunningham (FILAC), Bob Watson (formerly IPBES), Ivar Baste, Senior (Norwegian Environment Agency), Akira Homma (formerly Bio-Manguinhos, FIOCRUZ), Fadia Homeidan (American University of Beirut), Carolina Boterno Cabrera (Karisma Foundation), Paolo Laterni (WIPO), and Rachel Kosse (UNEP).

The team is very grateful to the high-quality authored contributions that have been included in this report, see table on page 4ff.

Contact/editor: R.A. Roehrl, roehrl@un.org

Suggested citation: IATT (2021). *Emerging science, frontier technologies, and the SDGs - Perspectives from the UN system and science and technology communities*. New York: United Nations Interagency Task Team on Science, Technology and Innovation for the Sustainable Development Goals. ***May 2021.*** http://sdgs.un.org/tfm/

C. Digitalisation, artificial intelligence and robotics

Potential Threats of Human Digital Twins for Digital Sovereignty and the Sustainable Development Goals

Kevin Mallinger, Alexander Schatten, Gerald Sendera, Markus Klemen, and A Min Tjoa (SBA Research)

Abstract

Human digital twins (DT) provide new ways of processing the activities of its user and are able to exert significant influence on the behaviour of individuals. As the technology promises manifold business opportunities, widespread incorporation is highly likely but may lead to a variety of social, environmental and technological challenges. A new regulatory context is needed to face possible downsides and foster an ethical and lean digitisation approach. This brief provides a concrete definition, a description of the implications and the respective policy recommendations.

Digital Twins - a Definition

Digital Twins are a virtual representation of physical objects with a bilateral exchange of information. A *human digital twin* is, therefore, the aggregation of human related data that is supposed to represent its real counterpart in the virtual world. A (human) digital twin continuously controls and monitors its physical twins' status, with the aim of optimizing its performance by triggering self-optimisation and self-healing mechanisms (Barricelli et al., 2020).

Problem Statement

Digital Twins (DT) are a new technological approach combining the capabilities of distributed sensor technologies, artificial intelligence, big data analytics, cloud computing and shared databases to create a digital virtual representation of a physical object. They have first been used in industrial applications such as aviation or production systems (Uhlenkamp et al., 2019), but are increasingly applied to monitor, represent, and influence human behaviour. A human DT is formed by large amounts of data produced by physical and virtual objects we interact with (e.g., fitness trackers, social media tools, smart watches, search engines, etc.) to support behavioural analysis and prediction. As the goal of digital twins is the optimisation of certain behavioural traits (such as health-status, user-activity in a system, response to ads, etc.), its feedback to the physical object is designed to achieve this goal.

Possible instances of this technology are widespread and include but are not limited to professional (Nikolakis et al., 2018), medical (Martinez-Velazques et al., 2019) and leisure activities (Barricelli et al., 2020). The usage of personal data has been a sensitive topic of legislation and was also discussed during the pandemic of COVID-19, when population monitoring and *nudging* of behaviour (Leonard et al., 2008) was suggested as political means to mitigate the crisis.

However, the advent of digital twins marks an evolutionary step in surveillance and profiling intensity. Reports indicate that by 2023 individual activities will be tracked digitally by an *Internet of Behaviour* to influence benefit and service eligibility for 40% of people worldwide, which might rise to 50% in 2025 (Plummer et al., 2020).

As this trend continues, an *extensive digital replica of one's personality* (also called *human digital twin*) poses significant threats to the digital sovereignty of individuals, democracy and nations as whole; but also for achieving the sustainable development goals (SDGs) in a variety of fields, such as:

Automated processing of profile data—often of unclear quality assurance: Existing societal biases can be reinforced by unfit designs of human digital twins, algorithmic feedback loops and simple statistical correlations. Without transparent and ethical use of DTs, predominant inequalities (e.g., SDG 5 – Gender equality, SDG 10 – Reduce inequalities) can be enhanced and damage well-being and amplify societal fractures (World Economic Forum, 2021). This risk is especially salient when machine learning ("artificial intelligence") is used for decision making, as these techniques are often non-transparent by design (Birhane, 2021).

Unregulated access to personal data and the digital twin itself: Personal data and the DT itself might be accessed by or sold to third parties which could restrict access of vulnerable groups to economic and social resources (e.g., application for insurances, loans, welfare, schools, jobs). Without reasonable restrictions and democratic oversight, this increased transparency poses a threat for achieving inclusive education (SDG 4), inclusive economic growth (SDG 8), equal opportunities and the reduction of inequalities of outcomes (SDG 10). In this regard, human digital twins also pose serious risks for the sustainable development of identification systems (SDG 16) (World Bank Group, 2021), as the (unregulated) use of such could be detrimental to an inclusive and privacy endorsing design.

Foreign data ownership: The accumulation and computing of personal data is often outsourced to servers or cloud services in foreign jurisdictions. The increased attack surface within such hyper-connected cloud environments and the creation of single points of failure foster significant privacy and security risks (Allianz Global Corporate & Specialty, 2021). As a consequence, data sovereignty of individuals, companies and nations as well as the resilience of national infrastructures (SDG 9 – develop quality, reliable, sustainable and resilient infrastructures) will be harmed as the dependence on services hosted by foreign providers increases.

Shadow human digital twins: Human digital twins can be created of people who did not consent to provide personal data. However, the accumulation and application of data from other people can be sufficient to indirectly infer a DT for individuals who are not even registered in the specific service (e.g., friends "tag" photos with names or location, email communication or calendar). Consequently, data leaks could expose people and organisations without their knowledge and undermine their activities (SDG 16 – peace, justice and strong institutions).

External influence of one's behaviour ("targeting"): The received feedback from human digital twins is designed to change the behaviour of the targeted object. Non-transparent and unethical designs of such intended behavioural changes bare manifold societal, psychological, and ethical risks (see for instance the attempts to change the outcome of elections).

Self-reinforcement of existing opinions, worldviews, and prejudices: Individual human digital twins can be designed to maximise interaction (e.g., "engagement" in social media) or to influence one's behaviour to increase prediction accuracy with less user variance. This can lead to reduced variety and complexity of presented information, increased information segmentation, and self-reinforcement of certain behavioural patterns (Kaakinen et al., 2020; Bhargava et al., 2015). As this imbalance of information will be increased, it poses a severe threat to democratic processes, the situational capacity to tackle existing threats (e.g., mitigation of climate change) and further aggravates the ongoing erosion of social cohesion and global cooperation (World Economic Forum, 2021). This type of reinforcement is in stark contrast to SDG 4 (knowledge to promote sustainable development), SDG 12 (ensure that people have the relevant information for sustainable development), SDG 13 (improve education and awareness on climate change) and SDG 16 (ensure public access to information).

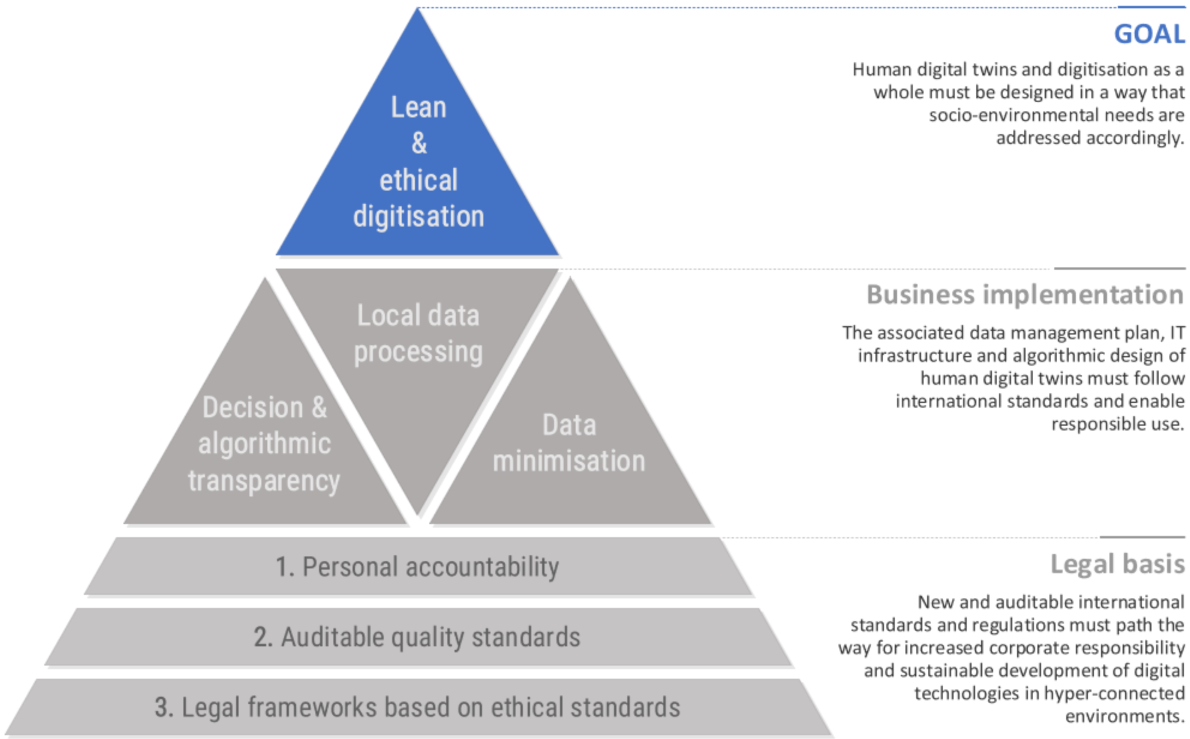
Policy Recommendations

We call for a *lean and ethical digitisation approach.* This is based on a principle that has already been described within Art. 5 GDPR as "data minimisation", which can be perceived as "an attitude to how we capture and store data, stating that we should only handle data that we really need." This principle forces us to figure out the minimum amount of data required to achieve the defined goal of a DT but should be extended in a way that requires incorporation of ethical standards and open communication of what we really intend to do with the digitisation approach.

Digital twins are designed to "make decisions" for their human counterpart. Next to *data transparency, decision transparency* is required to keep these processes fair and democratic. It should be critically reviewed how "black box" algorithms (like machine learning, artificial intelligence algorithms) that operate on behalf of people or make autonomous decisions that cannot be explained are applicable to transparency requirements. Such algorithmic transparency is necessary for reviews regarding mutually accepted ethical standards and further legal regulation.

A subsequent aspect is the *processing location of data.* Currently we see two paradigms at play, whereas the first and increasingly dominating principle—for digital twins but also in a more general way—is cloud/centralised computing. This has the distinct disadvantage that is inherent with centralised architectures: security and reliability issues affect large numbers of customers at the same time, and whoever has access to the processed data could exploit it (e.g., correlating data between customers, selling data). The second approach is to keep processing and computation (e.g., facial recognition, email indexing) on the customer device and only exchange data to central processing or for sharing with other users that were explicitly selected by the user.

Figure 1: Lean & ethical digitisation approach



If implemented carefully, this leads to more robust application performance and tends to support data minimisation, digital sovereignty and avoid single points of failure. Switching from centralized data centres to edge devices may reduce energy consumption due to less data transmission and decreased cooling of data centres. Furthermore, the closeness of edge-devices to distributed renewable energy sources may enable efficient energy demand management in smart grid systems (Digital Future Society, 2020).

Currently, *business models* dominate certain sectors of IT services (such as social networks, news sites, communication tools) that have their foundation in surveillance or tracking of customers, collecting and trading personal data, and manipulating their users on behalf of their business customers (Zuboff, 2019). Human digital twins extend the capacities of such business models extensively and their incorporation in such should therefore especially be reviewed critically. *Legal frameworks* that aim for *data minimisation, algorithmic transparency* and *local over central processing (Figure 1)* would mark a first step to limit associated societal threats. Regulators should be wary of businesses influencing drafting processes through lobbying efforts.

Existing frameworks (e.g., Sarbanes-Oxley-Act, GDPR) sanction failure in compliance, but prosecution can be cumbersome. As the sense of urgency in organisations to enforce and follow regulations may be lacking, novel legal frameworks should foster an increase in *organisational responsibility and personal accountability* of individuals in leadership to implement adequate measures towards accountability, resilience and the protection of data subjects. Furthermore, circumvention of these frameworks must not be possible by "freely given consent" (e.g., GDPR Art. 7 & Recital 43), as this requirement will often remain heavily imbalanced towards the interests of large companies as data collectors and controllers.

Finally, quality and scaling issues in increasingly complex IT infrastructure pose high risks for the resilience, security and privacy of our digital ecosystems. Therefore, auditable *international quality standards* for human digital twins and their respective software, architectural, process and security designs must be defined.

Whereas the threats of human digital twins are diverse, responsible incorporation could also support ongoing endeavours (e.g., profiled data for energy demand management within smart grid designs) in achieving the SDGs.

References

Allianz Global Corporate & Specialty (2021): Allianz Risk Barometer. Identifying the Major Business Risks for 2021.

Barriceli R. B., Casiraghi E., Gliozzo J., Petrini A. & Valtolina S. (2020): Human Digital Twin for Fitness Management. - IEEE Access 8: 26637-26664.

Bhargava P., Brdiczka O. & Roberts M. (2015): Unsupervised Modeling of Users' Interests from their Facebook Profiles and Activities. - Proceedings of the 20th International Conference on Intelligent User Interfaces, pp. 191–201. ACM 2015.

Birhane A. (2021): Algorithmic injustice: a relational ethics approach. - Patterns 2 (2): 1-9.

Digital Future Society (2020): Risks and opportunities of emerging tech in the climate decade. Barcelona, Spain.

Kaakinen M., Sirola A., Savolainen I. & Oksanen A. (2020): Shared identity and shared information in social media: development and validation of the identity bubble reinforcement scale. - Media Pyschology 23 (1): 25-51.

Martinez-Velazques R., Gamez R. & Saddikk A. E. (2019): Cardio Twin: A Digital Twin of the human heart running on the edge. - IEEE International Symposium on Medical Measurements and Applications: 1-6.

Nikolakis N., Alexopoulos K., Xanthakis E. & Chryssolouris G. (2018): The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor. - International Journal of Computer Integrated Manufacturing 32 (1): 1-12.

Plummer D., Hill J., Sallam R., Karamouzis F., Alvarez G., Litan A., Daigler A., Resnick M., Hunter R., LaRocca-Cerrone K., Canters J., Bai X., McRae E.-R., Zitmoer J., Cerio G. & Prentice B. (2020): Top Strategic Predictions For 2020 and Beyond.

Leonard T. C., Thaler R. H, Cass R. (2008): Nudge: Improving decisions about health, wealth, and happiness. - Constitutional Political Economy 19: 356-360.

Uhlenkamp J.-F., Hribernik K., Wellsandt S. and Thoben K.-D. (2019): Digital Twin Applications: A first systemization of their dimensions. - *IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, Valbonne Sophia-Antipolis, France, 2019: 1-8.

World Bank Group (2021): Principles on Identification for Sustainable Development: Toward the Digital Age - Second Edition (English). - Washington.

World Economic Forum (2021): The Global Risks Report 2021, 16th Edition. - Geneva.

Zuboff S. (2019): The Age of Surveillance Capitalism. - New York.