Digital Transformation for Sustainable Development Goals (SDGs) - A Security, Safety and Privacy Perspective on AI

Andreas Holzinger[1,2](✉), Edgar Weippl[3,4], A Min Tjoa[5], and Peter Kieseberg[6]

[1] Human-Centered AI Lab, Medical University Graz, Graz, Austria
andreas.holzinger@medunigraz.at
[2] xAI Lab, Alberta Machine Intelligence Institute, Edmonton, Canada
[3] SBA Research, Vienna, Austria
[4] Research Group Security and Privacy, University of Vienna, Vienna, Austria
[5] Vienna University of Technology, Vienna, Austria
[6] St. Pölten University of Applied Sciences, St. Pölten, Austria

Abstract. The main driver of the digital transformation currently underway is undoubtedly artificial intelligence (AI). The potential of AI to benefit humanity and its environment is undeniably enormous. AI can definitely help find new solutions to the most pressing challenges facing our human society in virtually all areas of life: from agriculture and forest ecosystems that affect our entire planet, to the health of every single human being. However, this article highlights a very different aspect. For all its benefits, the large-scale adoption of AI technologies also holds enormous and unimagined potential for new kinds of unforeseen threats. Therefore, all stakeholders, governments, policy makers, and industry, together with academia, must ensure that AI is developed with these potential threats in mind and that the safety, traceability, transparency, explainability, validity, and verifiability of AI applications in our everyday lives are ensured. It is the responsibility of all stakeholders to ensure the use of trustworthy and ethically reliable AI and to avoid the misuse of AI technologies. Achieving this will require a concerted effort to ensure that AI is always consistent with human values and includes a future that is safe in every way for all people on this planet. In this paper, we describe some of these threats and show that safety, security and explainability are indispensable cross-cutting issues and highlight this with two exemplary selected application areas: smart agriculture and smart health.

Keywords: Artificial intelligence · Digital transformation · Robustness · Resilience · Explainability · Explainable AI · Safety · Security · AI risks · AI threats · Smart agriculture · Smart health

Published by Springer Nature Switzerland AG 2021
A. Holzinger et al. (Eds.): CD-MAKE 2021, LNCS 12844, pp. 1–20, 2021.
https://doi.org/10.1007/978-3-030-84060-0_1

1 Introduction and Motivation

Often referred to as buzzwords, such as AI, Blockchain, Big Data, Internet of Things (IoT), ..., these technology trends of the last decades are the actual drivers of the digital transformation that is actually taking place [71,75]. Thus, these technologies no longer have just an additional support function, rather these technologies are changing complete process chains and permeate almost all our areas of life and work, from Smart Agriculture to Smart Health to name just two application areas. The main driver of digital transformation is undoubtedly the broad field of artificial intelligence (AI).

AI has gained a lot of attraction in the last decade. Many of the basic concepts date back to the middle of the last century, however the right combination and synergies of three approaches has led to a revolution that now brings AI to everyone's attention: (1) powerful, cost-effective, and available hardware (2) successful methods from statistical machine learning (e.g., Deep Learning), and (3) a growing amount of available data. AI-related components now permeate all sorts of labour, industries and applications, e.g.

- *Autonomous AI systems* that automate decisions without *any* human intervention (e.g., fully autonomous self driving cars [44], autonomous medical diagnosis [2], autonomous drones [19], ...).
- *Automated AI systems* that perform labor-intensive tasks requiring certain intelligence, and complete them automatically within a certain domain and given tasks (e.g., industrial robotic process automatization [1], automated medical workflows [48], automated forest management [49], ...).
- *Assisted AI systems* that help humans perform repetitive routine tasks faster and both quantitatively and qualitatively better (e.g., ambient assisted smart living [69], weather forecasting, ...).
- *Augmented AI systems* that help people understand complex and uncertain future events (e.g., Explainable AI in Digital Pathology [28], Simple Augmented Reality applications [56], Augmented AI in agriculture [68], ...).

This widespread adoption also lowers the barrier to entry for other players in the domain, whether they are scientists from entirely different domains (e.g. health, farming, climate research, ...) using AI technology to solve problems or companies by adding intelligent components to existing tools. While this trend undeniably brings tremendous benefits, opportunities, and possibilities in terms of new capabilities and applications, this rising trend can also lead to problems, especially when it comes to the security, trustworthiness, and privacy of these systems.

In recent years, the topic of sustainability has gained a lot of attention, especially with the declaration of the Sustainable Development Goals (SDGs) by the UN [65] with its 17 core goals. As such, several ideas and approaches have been put forward for using AI-related technologies to support these goals. While this is obviously a very valuable approach, we still need to understand the shortcomings of AI to make these approaches inherently sustainable. While many researchers reduce the problems of AI to purely theoretical aspects, we will define some key issues in this paper. Nevertheless, we believe that AI technologies can significantly improve our lives and support the SDGs through improved digital transformation, however much additional work is needed to actually make a difference.

This paper is organized as follows: Sect. 2 provides an overview of the SDGs as well as some background and related work on how AI can support these goals. Here we also present two specific examples that affect virtually everyone in their daily lives: smart health, with the goal of precision medicine, and smart agriculture, with the goal of precision farming. In Sect. 3, we analyze a selection of topics that we believe need to be considered in supporting the goals outlined in the SDGs through AI technology. While we have derived some of these issues from the relevant literature, we have added some new ones based on our many years of practical experience in developing AI-based systems - always with a focus on safety, security, and privacy. In the Conclusion section, we summarize the paper.

2 Background and Related Work

In this section, we provide an overview of the background of the Sustainable Development Goals, as well as an outline of related work that shows how AI-based systems can support them.

2.1 The UN Sustainable Development Goals (SDGs)

The idea of "sustainable development" was already discussed by the United Nations Brundtland Commission in 1987: "Sustainable development is development that meets the needs of the present without com-promising the ability of future generations to meet their own needs" [13].

The fundamental concept of sustainability originated at the United Nations (UN) Conference on Environment and Development (UNCED) in Rio de Janeiro in 1992 (the so-called "Earth Summit" [63]), where a *Declaration of Principles and Desired Action on International Agreements on Climate Change and Biodiversity and a Declaration of Principles on Forests* were presented. Subsequently, in 2002, the commitment to sustainable development was reaffirmed at the World Summit on Sustainable Development in Johannesburg, South Africa.

The concept was intentionally not clearly defined to allow a way to address the very different challenges: from planning sustainable cities to sustainable livelihoods, from sustainable agriculture to smart health, and the efforts to develop common business standards in the UN Global Compact and the World Business Council for Sustainable Development [61].

The adoption of the Sustainable Development Goals (SDGs) in 2015 [70] signaled the commitment of world leaders to a more sustainable path to inclusive and equitable growth. Also known as the 2030 Agenda, the 17 SDGs cover a wide range of development-related issues and include 169 targets and 304 indicators [16], see Fig. 1 for an overview on the 17 SDGs.

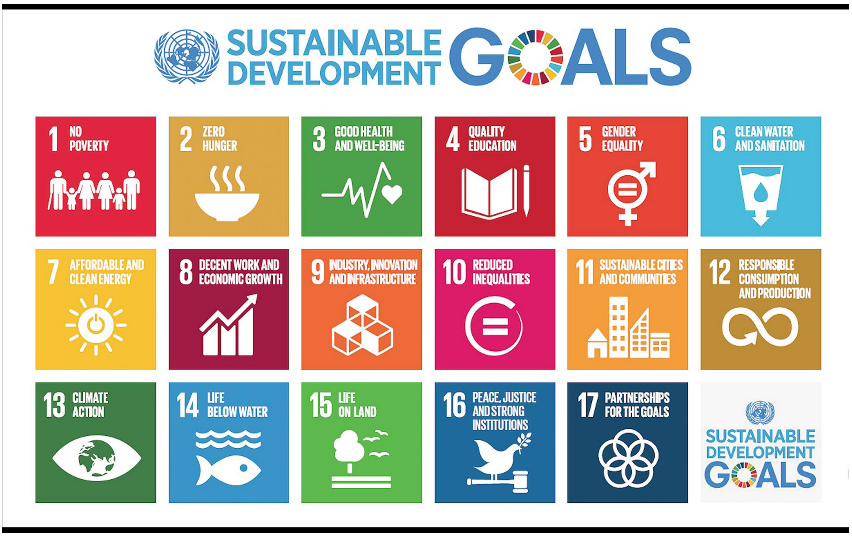
SUSTAINABLE DEVELOPMENT GOALS
1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS
SUSTAINABLE DEVELOPMENT GOALS

Fig. 1. An overview on the UN Sustainable Development Goals (SDGs) [65]

2.2 Smart Farming and Precision Agriculture - An Example for Supporting SDGs Through AI

As a first application example, we choose a topic that concerns every inhabitant of our planet: Smart Farming, i.e. the use of AI in cyber-physical applications for versatile support of process chains in agriculture. Cyber-physical systems (CPS) have been established for some time [8] and although CPS are very versatile, their engineering is most challenging due to the high degree of heterogeneity. Moreover, the importance of CPS for smart farming is often underestimated, but they form the basis for future precision farming, for better crop management and resource use. In this context, massive amounts of data are already being generated in great variety, which can be collected, analyzed and used for decision making. The goal is to develop smart agriculture that will help address the current major socio-economic challenges worldwide (Goal: "zero hunger"). This domain is essential as it extends from agriculture itself far beyond primary production to virtually the entire food supply chain. Here, AI can help in many ways to gain predictive insights into agricultural operations, make real-time operational decisions, and redesign business processes. This has to incorporate often conflicting economic interests, leading to new business models which are developed this way. Clearly, there are all sorts of challenges of a technical nature here, however there are also tensions and shifts in roles and power relationships between different players in the current food supply chain networks. One example is the need of an evidence-based and critical investigation on the observed shift, from individual small farmers to powerful high-tech "agriculture factories". At the same time, there are public institutions that are now publishing data and where, of course, individual privacy must be ensured. The future of smart agriculture could play out on a continuum of two extreme scenarios: 1) closed, proprietary systems in which the farmer is part of a highly integrated food supply chain, or 2) open, collaborative systems in which the farmer and all other actors and stakeholders in the chain's network need the flexibility in choosing business partners for both technology and food production [17,79]. Closely related to smart agriculture is sustainable forest management. The management of forest ecosystems is underestimated, but of eminent importance for the survival of our planet [24] as the restoration of forested land could help to capture carbon and thus mitigate climate change [6]. Computer-based support tools have been in use in this domain for some time [66]. However, monitoring of forest areas is far from trivial, facing major challenges such as high forest density, complexity and diversity of forest structure, complex topography and climatic conditions, and difficult access for human researchers. Here, unmanned aerial vehicles are already making a valuable contribution by enabling the classification of forest types based on Convolutional Neural Networks [49].

2.3 Smart Health and Precision Medicine - An Example for Supporting SDGs Through AI

As a second example, we choose a topic that also affects each and every one of us individually and is also accompanied by a large number of non-trivial problems: health, and here in particular the emerging domain of smart health. The trend toward higher life expectancy together with the increasing complexity of medicine and health services is causing healthcare costs to rise dramatically worldwide. As a result, enormously high expectations are being placed on AI for health worldwide.

The concept of smart health [30] has huge potential to support a future P4 medicine (preventive, participative, predictive and personalized) or in short *personalized medicine* [23]. The goal of this a.k.a. future *precision medicine* [14] is in modeling the complexity of patients health in order to tailor medical decisions, health practices and therapies to the individual patient - for example, a drug precisely designed for a patient's individual needs and specific background in a given context. This trend towards personalized medicine produces huge amounts of data, which makes manual analysis difficult and almost impossible for a human being [26]. For example massive amounts of sensors produce large amounts of high-dimensional, weakly structured data sets and massive amounts of unstructured information. In the medical domain, many different modalities contribute to an outcome. Consequently, the smart health principle makes medicine a truly data-intensive science. To keep up with these growing volumes of complex data AI approaches are mandatory, however, in the medical domain there is always a need for a human-in-the-loop [27], at least the human-in-control - because of legal aspects [64]. However, the synergies between AI and precision medicine promises to revolutionize healthcare: whilst precision medicine methods help identify phenotypes of patients with less frequent treatment response or special healthcare needs, AI is being used to support clinician decision making through augmented intelligence. Translational research exploring this convergence can help solve precision medicine's most difficult challenges, particularly those where non-genomic and genomic determinants combined with information from patients' symptoms, clinical history, and lifestyle habits will enable personalized diagnosis and prognosis [35].

By deploying complex AI systems in physical-digital ecosystems, future physicians will be supported by their AI assistants in managing their flood of data from different modalities, which requires explainability and causability [29]. At the same time, patients will be supported by their on-line health assistants, and moreover, these technologies will support the *preventive medicine* nature to enable healthier living, wellness and well-being, which will also lead to enormous amounts of private data.

In the medical field, the issues of transparency, accountability, and trust are prerequisites for the integration of AI into daily practice. As the importance of medical AI will certainly continue to grow strongly in the coming years, it is imperative that legal and ethical issues are always considered together [53,54].

2.4 Impact of AI on SDGs

There has been ample publications indicating the benefits of certain AI-based systems on SDGs, as e.g. The academic literature also includes analysis on related subjects, like in [73], where the authors give a comprehensive study on the utilization of artificial intelligence in the development of sustainable business models. In another paper [3], the authors discuss the impact of the ongoing Covid pandemic on the SDGs and the efforts taken into reaching them by 2030. They provide an interesting approach on looking for synergies between different targets and approaches in order to prioritize them. The publication [25] on the other hand gives a very good overview on important topics in this area and also provides a first insight into the challenges associated with using AI for supporting SDGs. To the best of our knowledge the most comprehensive study can be found in [77]. In this work, the authors analyzed all 17 SDGS with their 169 targets with respect to whether AI is beneficial or detrimental for a target, more specifically, they scored, how beneficial and how detrimental AI is for any given target. The measurement was made by conducting a consensus-based expert elicitation process based on results from previous studies. While the paper added a short discussion on security-related issues of utilizing AI for achieving the SGDs, the discussion was rather short and did not go into details.

3 Open Issues on Using AI for SDGs

In this section, we discuss security related issues of intelligent systems, some of them being typically overlooked and having received little attention in the academic literature. In this discussion, we use the term *security* rather loosely for any technical issue that has a detrimental effect on users. Still, we typically limit ourselves to the purely technical IT domain and did not include social and economic issues like increasing job loss, even though these also might have an impact on the overall security of a nation. We have loosely grouped these topics in terms of overarching themes to make the work more stringent, even though several topics might belong to more than one of these themes.

3.1 Data and Models

Many AI techniques, especially advanced techniques of statistical machine learning, including neural networks (deep learning) that gathered a lot of attention during the past years, heavily rely on a key resource: Data. Data is used to train these networks, i.e. the whole AI system is not only defined by the pure mathematical/algorithmic side, but to a large extent by the data that was used for training the system. Thus, a lot of problems in the adoption of intelligent systems lie in the trustworthiness of data sets, ranging from unintentional bias that leads to discrimination of people to malicious attacks trying to interfere with the system and manipulate the decision making process in subtle ways. In the following we have gathered a selection of issues that we see as key concerns when trying to solve SGDs through AI-based systems.

Acquiring Training Data. Many aspects of AI hinge on training data, especially related to issues of trust and quality. Still, acquiring such data is a big problem for many applications and due to various reasons ranging from Privacy related issues in the existing medical data to simply no useful data being available at all. The acquisition of data is especially important, as many other issues can be related to it, e.g. the problem of bias, backdoors and providing a good ground truth. Open data might be a solution for this problem in some applications, e.g. forestation related issues, but it will be hard to find widespread acceptance for such a bold step in the medical domain.

Providing a Ground Truth. Related to the issue of generating suitable sets of training data lies the issue of finding a ground truth. This is especially critical in highly complex research questions, where either it is impossible to gather an oversight of the complete data available or the interpretation of the data is depending on other parameters that are hard to impossible to objectify, like e.g. political opinions. Gathering an oversight on a topic can be difficult either due to the sheer amount of information available, but also due to the information being stored in secluded data vaults. e.g. sensitive medical information. There are an ample number of projects that aim at connecting and joining these data faults in a secure and privacy respecting way. As an example for the latter, finding ground truths in fake news detection is especially difficult [10], as news can never cover a complex topic with a lot of subjective decisions and opinions as a whole, thus making it hard to decide, when information was cut maliciously. Furthermore, many legit new outlets also put some kind of spin at the news they are reporting, consciously or unconsciously, by using different terms for the same people (e.g. “freedom fighters” versus “terrorists”) or things. Furthermore, when looking at long-standing conflicts, reports have to cut somewhere in time in order to not become history books - thus often removing important background information. Still, finding a ground truth can also be very challenging in purely technical applications for AI, e.g. in intrusion detection systems (IDS) [15] based on collecting "normal behavior" in order to later detect suspicious traffic.

Bias and Data Quality. Often it is difficult to even find enough data on a subject in order to train a neural network, so the topic of data quality is currently overlooked in many cases [33]. Still, in the recent past, several issues surrounding deficient training data have emerged [52], most prominently regarding racial bias in sensitive applications like predictive policing. Bias is an especially important issue to tackle, as bias in AI applications, e.g. decision support, can become self-reinforcing: If e.g. a certain population is over-proportionally included in a training set the algorithm could advise to look deeper in said population - thus to find even more suitable examples reinforcing the original bias resulting in a spiral of bias reinforcement [57]. There have been several discussions regarding racial bias in predictive policing (see e.g. [57,67]), even though a structured study painted a more complex situation when looking at arrest rates [11]. Still, even besides issues of bias, assessing data quality is a difficult task which requires a lot of further research [5].

Data Preparation and Cleansing. A topic that is typically overlooked in purely theoretical papers, is that data often needs to be prepared in order to be useful. For example, data is often incomplete or contains erroneous records [50]. This is not only true for training data, but also for the processing data. Thus data cleansing is typically applied to the data streams which works on different levels with different techniques [5], by e.g.deleting faulty records, assigning defaults or trying to guess the most probable correct version. This, of course, introduces changes into the data and, in case of training data, into the subsequently calculated model, resulting in different models, i.e. a direct impact of the data cleansing process into the very definition of the neural network at worst. Currently, there is not much discussion on this issue in the scientific community, especially not regarding the legal and organisational implications. It is also a very hard question, which model is more correct when facing two different cleansing strategies that result into different models. The same problems arise in other data preparation steps, e.g. reduction of noise in audio files [74] or pictures, which of course need to be done with respect to the state of the art, but often use heuristic algorithms that sport different results depending on various side parameters (see e.g. [36]). The impact of these data preparation steps need to be analyzed carefully regarding their impact. Selecting an AI technique that is sufficiently stable against the expected level of instability in the data sets is an absolute must, unfortunately, stability has often been pushed into the background of the tool development process.

Sharing Models and Training Data. Sharing data and even trained models is an interesting approach in order to battle the problems of acquiring training data (see [80] for a solution platform). When sharing training data in order to enable other parties to train their models, questions of privacy, but also regarding the intrinsic value of the data, need to be kept in mind [81]. Thus, it might be advisable to use fingerprinting technologies in order to be able to detect data leaks [9,39]. There might also be laws and regulations that need to be complied with in the sharing process. Furthermore, there is always the problem of assessing the quality of the shared data with respect to e.g. sample selection, data preparation and overall data quality (see e.g. [34] for a novel blockchain based solution). Still, pooling data and sharing them with other player might be a viable strategy in many application areas related to SDGs in order to circumvent training data shortage. Another related strategy lies in sharing the models, also called pre-trained systems. Here, the original training data is not given away, but solely the trained model, which, of course, hast to adhere to certain prerequisites specific to the system it is later used in. Pre-trained models are often claimed to solve many of the issues associated with the sharing of training data, especially to those related to the GDPR [60] and privacy, still, this needs to be taken with a grain of salt: On the one hand, there have been attacks against pre-trained models [42,46] devised in the past that had some success in recreating information on the original training data of a model, depending on the models complexity and the availability of side information. Still, what is even more a problem, is the amount of trust that has to be placed into the generator of the model in question:

- It is typically impossible to infer anything on the raw training data used for building the model like using sanity checks or verifying data records. The user must therefore trust the generator that the training data had sufficient quality and is unbiased.
- It is very hard for the user to control, whether backdoors were included into the model [22], e.g. a model trained on the impact of emissions could lead to correct results in all cases, except when a special emission pattern typical for e.g. a specific car brand is encountered. In this special case, the model suddenly calculates a far better result.

Especially the backdoor issue is a major trust problem, especially when a lot of risk and investments are involved. The generator of the model must be extremely trustworthy, still another problem derives from the closed nature of a pre-trained model: It is extremely difficult to expand this model with new data in a controlled way, i.e. even in cases of algorithms that feed new information back into the model and thus expand on the pre-trained model, a lot of transparency is lost due to the unknown nature of the original model training data set. As a result, pre-trained models result in a lot of additional issues, still, for several applications they will be the only feasible solution in the near future.

3.2 Providing Trustworthy Systems

Trust is one of the major issues when it comes to IT-systems that are deployed in critical environments, resulting in the notion of trustworthy artificial intelligence [78]. While this is not related to the SDGs outlined in Sect. 2.1 at first glance, it must be taken into account that (i) some of them relate to people and their data making them sensible data driven applications and (ii) others relate to barriers for big industrial corporations, again, requiring measures to be put in place in order to make people trust them not to be manipulated. On a side note, the topic of explainability is of high importance for providing trust, as outlined in [31].

Security Testing. Testing systems for security is a major aspect in finding vulnerabilities in existing systems and providing remedies in the form of patches. Several testing methods are employed, ranging from code-reviews, where the analyst is in the possession of the systems source code, over architecture analysis to black-box testing methods like digital twins or penetration testing. All of these techniques become increasingly difficult in the presence of AI (see [45] for a survey): For the architecture review, while it certainly stays useful in order to find fundamental flaws like problematic access control, insecure system design and so on, the AI component is typically a black box: The training data virtually defines large parts of the AI behavior, while not being represented in the architecture. The same holds true for source code reviews, since the training data is not within the source code, the very aspects that define the AI are not included. But even when including the training data and the trained model in the review, in general not much useful information can be deduced from them due to the explainability problem [21]. As for penetration testing, AI components add an additional layer of complexity as outlined in [72], where the authors also provided some first attempts for a methodological approach towards the topic. Also with respect to digital twins, side-effects and internal workings of such complexity typically cannot be simulated within reasonable time and financial limits. When using pre-trained models, security testing becomes even more cumbersome, as already outlined in the paragraphs on pre-trained models and bias in Sect. 3.1.

Privacy and Profiling. When utilizing AI for enhancing the targets of SDGs it can be tempting to use as fine-granular data as possible in order to gather good results. This can be a problem with respect to end user privacy [37], especially when dealing with personal data, e.g. when analysing issues of gender equality (SDG 5) or health (SDG 3). While certainly beneficial for the results of the AI component [51], this can be detrimental to end user interests, especially in cases, where the users in question are within a suppressed minority, i.e. it must be made sure that the means that are planned to support the SDGs are not misused in order to hurt them. User profiling [18] can be seen as an extreme form of privacy violation, as digital models of users and their interests, as well as preferences are generated and subsequently exploited, e.g. in order to well-placed targeted marketing. Still, the big companies like Facebook and Google can derive much more from this data. Keeping privacy in mind is thus a fundamental step for designing any AI-based application supporting SDGs.

Data Manipulation Detection. Manipulation can take place at many steps inside a data driven workflow (without guarantee of completeness):

- At the collection phase by only collecting data suitable to the attacker.
- Whenever data resides in a data store, e.g. a database.
- Through the introduction of faulty data into the data streams (see also adversarial machine learning).
- During the calculation and processing stages, especially within complex enrichment workflows, often including external enrichment data of varying volatility.
- When the results are sent to the human decision makers (in case they are still foreseen in the system).
- In case of feedback loops in expert systems in the mechanisms that report the feedback from the (human) expert to the machine learning entity.

Detecting such manipulations, which can often be carried out trivially, especially by an internal attacker like a disgruntled database administrator, a strict integrity providing process like a chain of custody [20] has to be put in place in order to mitigate these threats. It must be kept in mind though that these mechanisms must be secure against very potent internal attackers, like e.g. put forward in [38].

Adversarial Machine Learning. In adversarial machine learning [32], an attacker tries to change the underlying decision model of an AI component by feeding it with specifically crafted data. Often, this feeding needs to be done slowly in order to go undetected. While current attacks are rather quite low-level in nature, their effects can be stunning and might even allow attackers to introduce backdoors into existing systems. Due to the explainability problem, the resulting changed models are often hard to detect, and detection typically focuses on the feeding process though. See publication [62] for an in-depth description of this issue.

Resilience and Stability. When using AI-based systems in order to tackle targets derived from the SDGs, a certain level of resilience [47] is direly needed, i.e. the system needs to be able to adapt to successful attacks and maybe even change. This is especially important, as an unreliable system will lose acceptance rather quickly. The same holds true for the topic of stability, where we use two different meanings for the term *stability*: (i) A system that is running stable and uninterrupted and (ii) The utilization of algorithms that do not behave chaotic, i.e. the output should not change too drastically when making small changes to the input. While the reason for the importance of the first meaning is rather straightforward and can be seen as a part of resilience, the second one is required in order to deal correctly with rounded and/or inaccurate inputs in a correct manner: Since input data, especially concerning natural processes, can never exceed a certain, sometimes quite low, margin for accuracy, an algorithm that reacts very strong on such differences might be useless in a practical context.

3.3 Control

Control in this context means: Who runs the system, who is responsible for the code, for the data, who can change the software - all these elements are vital to clarify when tackling the grand challenges put forward by the SDGs, especially as a lot of financial and political impact is caused by many of them.

Control over Data. Perhaps one of the most important aspects often overlooked in supporting the SDGs with AI-based systems is the issue of control over the data that is processed. This does not only refer to the training data sets, but also the actual processing data that is analyzed. For example, car companies have been found out to change their motor software in order to detect test settings and adjust the exhaust accordingly. Large companies trying to game such systems needs to be taken into account in many measures, especially regarding climate related SDGs.

Control over Systems. What has been said about the data can also be put forward for the system - the one who controls the system can exert a lot of power over the important topics put forward in the SDGs and the methods used to support them.

Control over Rules. Even more overlooked, control over measures to support SDGs can be achieved quite elegantly and simple by being in charge of making the rules: By being able to specify side or target parameters, big companies might try to seemingly fulfill targets set out through the SGDs, but rather than changing root causes just working around them.

3.4 Transparency

Transparency is a very problematic topic in AI [43], especially due to the problem of providing explainable artificial intelligence for systems exceeding quite a low level of complexity. Thus, while this issue has been the root cause for many of the problems already outlined before, we want to discuss some issues very specific to different notions of transparency in data driven systems, ranging from transparency regarding the internal workings of the AI system to issues of reproducibility.

Functional Testing. Functional testing typically involves testing a system for its proper working, i.e. identifying that all features have been implemented, the correct results are calculated and so forth. This is typically done by providing test cases, but also incorporates more advanced techniques like fuzzying or combinatorial testing [55]. For an AI system, it might be hard to determine the actual test cases, i.e. it might not be simple to define the correct function set of the system, especially when thinking of systems in the area of decision support: Did the IDS not report because of an active decision, or because it simply did not call the respective analysis routines at all (a slightly exaggerated example).

Process Transparency. When using data in cascades of intelligent systems, and especially when training neural networks with said data, transparency becomes rather difficult, especially being able to answer questions on the actual sources of information particles that were later on aggregated. This can be a very problematic attack vector for processes that aggregate data from sources of different sensibility [12], be it patient data or vital information on secret business processes of a company. Thus, in order to mitigate a large amount of attacks, transparency over the whole process chain needs to be provided. In case of utilizing sensible personal information, this is also a requirement derived from the GDPR, but related issues can also be encountered in other applications domain, especially within (military) intelligence. In case of AI-supported SDGs, such information can e.g. incorporate sensitive internal technical details of machines, where a successful attack going for extraction of (parts of) this information can cause great damage to the original data owner. Process transparency is also vital in case of re-processing of data.

Reproducibility. In many cases it can be important to exactly reproduce a result (i) in order to proof that it was actually produced the first time or (ii) in order to learn from the calculation process and maybe challenge and adapt it:

- Reproducing a state of knowledge of the human decision maker: AI systems are currently often seen in a supporting role for human decision makers, i.e. the AI analysis the data for patterns and provides a human with the results who is then in charge of the decision. Especially in sensitive and time critical environments, the human decision maker has to take a decision under a lot of stress based on incomplete information. If the decision was wrong, amply time will be dedicated to the subsequent blame game. Thus, it is vital for the human decision maker to be able to reproduce the exact state of knowledge at the point of decision making [41]. This is also very important in order to learn from mistakes and improve on the decision making process as a whole, including the human, as well as the AI component.
- Re-processing data: In many applications, data needs to be re-processed, i.e. the analytical workflows have to be redone on data that has already been processed once. This can have implications in case the process data is fed back into the model, as it would increase the impact of re-processed data, since it would be included into the model again as often as it is re-processed.
- Post-processing data: Sometimes time-sensitive data (e.g. call detail records in telecommunications) arrives late in the analysis process, but still needs to be processed as if it had arrived on the correct time. This is very difficult with respect to the versioning of models and enrichment data, especially in case of feeding back results into the model. The difference between post-processing and re-processing is that re-processing uses the models and enrichment data current at the time of re-processing, while in post-processing the original state at the time the processing should have originally happened needs to be provided.

There are a set of problems surrounding the topic of reproducibility, with the following being most important from our perspective with respect to security:

- Changes in the model: Especially in algorithms that continuously change the model (e.g. self learning algorithms [4]), it can be very hard to (i) track the impact of changes to the model on the decision making process and (ii) go back in time for post-processing. Here, a very fine-granular and still manageable solution for model versioning needs to be provided.

- Heuristic approaches: In case of e.g. random values introduced into the algorithm, running the same algorithm on the same data set using the same model and enrichment information can (and typically will) result in differences in the results. Thus, in order to provide a high level of reproducibility and transparency, all internal values need to be logged in order to be able to redo the whole process.
- Volatile enrichment data: Also enrichment information can be tricky, especially when it is not hosted by the AI system but externally and only invoked through limited interfaces. Managing and versioning this information is vital in order to provide a decent level of reproducibility in many data driven systems.

Deletion and Rectification. Sometimes it becomes necessary to delete or rectify data inside a workflow. Reasons can be different, but especially within the legal domain of the GDPR, persons have the right to revoke their consent to voluntary data processing and having their data deleted from the databases. Changing data in AI processes can be hard, especially when the information had been used in order to train a neural network [7]. While removing the deleted data from the actual trained network might not be required from a legal point of view [76], it can become important in cases where the data is wrong and has an impact on the decision making process, e.g. by introducing a class of cases not actually existing, by introducing bias into the model or a backdoor. Thus, mechanisms on a technical and also on an organizational level must be put into place in order to be able to deal with such requests in an ordered and timely manner.

3.5 Other Issues

In this section, we have gathered some other issues that need to be discussed when planning to support a SDG-target with an intelligent system.

Liability. There is still a much debate inside the legal academic world, as to who is going to be held responsible for damages caused by AI systems [59], especially in expert in the loop systems [58]. In the case of using AI for SDG targets, the topics addressed can be very complex (e.g. climate models) when compared to end user apps, with a high impact of the resulting recommendations on society, the economy and other fields directly affecting millions to billions of citizens in the world. Thus, since these big questions ought to be solved on an international level headed by the UN, liability needs to be solved on an international basis too.

Over-Engineering Due to Ubiquitous AI. This is a very new issue that we did not encounter in any literature, still, we believe it is a big issue with respect to security: In many recent technologies it can be seen that the development path leads from the technology being exotic and expensive first to a fast decline of costs, thus making the technology available to virtually anyone at low costs. This can also be seen in the realm of AI, where AI-based applications become increasingly ubiquitous with many new applications targeting ever increasing customer basis. In addition it can be seen that implementations tend to gravitate towards the use of standard technologies and frameworks, i.e. many implementations go back to the same basic technologies. In order for such a basic technology to survive, it must provide the capabilities required by most implementations, else other frameworks will take its place. This means that even for implementations only requiring a small an primitive subset of technique, typically rather powerful frameworks are used. We can also see this trend in the hardware world, where even for very primitive sensors, standard chips are deployed that run a full UNIX with many advanced features. This is rather problematic from a security standpoint. While it can certainly be argued in many other technologies that using the same fundamental frameworks is in contrary beneficial to the overall security of the system due to the high amount of security analysis received by a single framework, powerful systems also allow for sophisticated attack vectors and typically result in bigger attack surfaces [40]. With respect to AI, additionally the explainability problem must be considered. Using powerful off-the-shelf frameworks can thus result in the utilization of very powerful and highly complex systems for very simple tasks, e.g. (as an exaggerated example) using a trained deep neural network instead of a simple rule set in a decision support system. This problem, combined with the trend of providing AI almost anywhere results in huge amounts of (critical) systems that can only be tested for security at a very high price, thus introducing a huge uncharted attack surface.

4 Conclusion

Artificial intelligence permeates almost all areas of life and work. In this paper, we have developed a brief overview on the topic of supporting the targets of the UN Sustainable Development Goals (SDGs) from a security, safety, and privacy perspective. To this end, we have identified potential problems and threats that AI-based systems are causing now and will cause in the future - in particular, novel threats that are not even thought of in the initial euphoria of planning, developing, or even deploying AI. We discuss this using two selected application areas, Smart Agriculture and Smart Health, both of which are of eminent importance to each and everyone of us. These findings are for scientists, developers and policy makers when considering the impact such solutions to the SDGs will have on industry and society. This inevitably leads to many conflicting interests and strong attacker motivation by powerful entities. Moreover, the goal of this paper is to provide starting points for future work. While we strongly believe that artificial intelligence will play an important role in supporting the goals articulated by the SDGs, implementation must be done carefully to reduce collateral damage and/or not to undermine the original intent by creating tools to the detriment of the supported goals.

Acknowledgements. The authors declare that there are no conflict of interests. This work does not raise any ethical issues. Parts of this work have been funded by the Austrian Science Fund (FWF), Project: P-32554, and the Austrian Research Promotion Agency through grant 866880 (Big Data Analytics).

References

1. Van der Aalst, W.M., Bichler, M., Heinzl, A.: Robotic process automation. Bus. Inf. Syst. Eng. **60**, 269–272 (2018). https://doi.org/10.1007/s12599-018-0542-4
2. Abràmoff, M.D., Lavin, P.T., Birch, M., Shah, N., Folk, J.C.: Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit. Med. **1**(1), 1–8 (2018). https://doi.org/10.1038/s41746-018-0040-6
3. Asadikia, A., Rajabifard, A., Kalantari, M.: Systematic prioritisation of SDGs: Machine learning approach. World Dev. **140**, 105269 (2021)
4. Auer, P., Cesa-Bianchi, N., Gentile, C.: Adaptive and self-confident on-line learning algorithms. Journal of Computer and System Sciences 64(1), 48–75 (2002)
5. Azeroual, O., Saake, G., Abuosba, M.: Data quality measures and data cleansing for research information systems. arXiv preprint arXiv:1901.06208 (2019)
6. Bastin, J.-F., et al.: The global tree restoration potential. Sci. **365**(6448), 76–79 (2019). https://doi.org/10.1126/science.aax0848
7. Baumhauer, T., Schöttle, P., Zeppelzauer, M.: Machine unlearning: linear filtration for logit-based classifiers. arXiv preprint arXiv:2002.02730 (2020)
8. Bennaceur, A., et al.: Modelling and analysing resilient cyber-physical systems. In: IEEE/ACM 14th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS). IEEE (2019)
9. Boneh, D., Shaw, J.: Collusion-secure fingerprinting for digital data. IEEE Transactions on Information Theory 44(5), 1897–1905 (1998)
10. Bozarth, L., Saraf, A., Budak, C.: Higher ground? How groundtruth labeling impacts our understanding of fake news about the 2016 us presidential nominees. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, pp. 48–59 (2020)
11. Brantingham, P.J., Valasik, M., Mohler, G.O.: Does predictive policing lead to biased arrests? results from a randomized controlled trial. Stat. Public Policy **5**(1), 1–6 (2018)
12. Böhm, C., et al.: GovWILD: integrating open government data for transparency. In: Proceedings of the 21st International Conference on World Wide Web, pp. 321–324 (2012)
13. Bundtland, G.H.: Report of the World Commission on Environment and Development: Our common future. Uni. Nations Gen. Assembly Doc. A **42**(427), 1–300 (1987)
14. Chen, R., Snyder, M.: Promise of personalized omics to precision medicine. Wiley Interdisciplinary Reviews: Systems Biology and Medicine **5**(1), 73–82 (2013). https://doi.org/10.1002/wsbm.1198
15. Du, P., Sun, Z., Chen, H., Cho, J.H., Xu, S.: Statistical estimation of malware detection metrics in the absence of ground truth. IEEE Trans. Inf. Forensics Secur. **13**(12), 2965–2980 (2018)
16. ElMassah, S., Mohieldin, M.: Digital transformation and localizing the sustainable development goals (sdgs). Ecol. Econ. **169**, 106490 (2020). https://doi.org/10.1016/j.ecolecon.2019.106490

17. Eyhorn, F., Muller, A., Reganold, J.P., Frison, E., Herren, H.R., Luttikholt, L., Mueller, A., Sanders, J., Scialabba, N.E.H., Seufert, V.: Sustainability in global agriculture driven by organic farming. Nature Sustainability **2**(4), 253–255 (2019). https://doi.org/10.1038/s41893-019-0266-6
18. Fawcett, T., Provost, F.J.: Combining data mining and machine learning for effective user profiling. In: KDD, pp. 8–13 (1996)
19. Floreano, D., Wood, R.J.: Science, technology and the future of small autonomous drones. Nature **521**(7553), 460–466 (2015). https://doi.org/10.1038/nature14542
20. Giannelli, P.C.: Chain of custody and the handling of real evidence. Am. Crim. L. Rev. 20, 527 (1982)
21. Goebel, Randy, Chander, Ajay, Holzinger, Katharina, Lecue, Freddy, Akata, Zeynep, Stumpf, Simone, Kieseberg, Peter, Holzinger, Andreas: Explainable AI: the new 42? In: Holzinger, Andreas, Kieseberg, Peter, Tjoa, A Min, Weippl, Edgar (eds.) CD-MAKE 2018. LNCS, vol. 11015, pp. 295–303. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99740-7_21
22. Gu, T., Liu, K., Dolan-Gavitt, B., Garg, S.: Badnets: Evaluating backdooring attacks on deep neural networks. IEEE Access **7**, 47230–47244 (2019)
23. Hamburg, M.A., Collins, F.S.: The path to personalized medicine. New England Journal of Medicine 363(4), 301–304 (2010). doi: 10.1056/NEJMp1006304
24. Hasenauer, H.E.: Sustainable forest management: growth models for Europe. Springer, Heidelberg (2006)
25. Herweijer, C., Waughray, D.: Harnessing Artificial Intelligence for the Earth. Fourth Industrial Revolution for the Earth Series. World Economic Forum (January 2018), p. 52 (2018)
26. Holzinger, A.: Trends in interactive knowledge discovery for personalized medicine: Cognitive science meets machine learning. IEEE Intelligent Informatics Bulletin **15**(1), 6–14 (2014)
27. Holzinger, A.: Interactive machine learning for health informatics: When do we need the human-in-the-loop? Brain Inf. **3**(2), 119–131 (2016). https://doi.org/10.1007/s40708-016-0042-6
28. Holzinger, A., et al.: Towards the augmented pathologist: challenges of explainable-AI in digital pathology. arXiv:1712.06657 (2017)
29. Holzinger, A., Malle, B., Saranti, A., Pfeifer, B.: Towards multi-modal causability with graph neural networks enabling information fusion for explainable ai. Information Fusion **71**(7), 28–37 (2021). https://doi.org/10.1016/j.inffus.2021.01.008
30. Holzinger, A., Röcker, C., Ziefle, M.: From smart health to smart hospitals. In: Smart Health: State-of-the-Art and Beyond. Springer Lecture Notes in Computer Science, LNCS 8700, pp. 1–20. Springer, Heidelberg, Berlin (2015)
31. Holzinger, K., Mak, K., Kieseberg, P., Holzinger, A.: Can we trust machine learning results? Artificial intelligence in safety-critical decision support. Ercim News **2018**, 42-43 (2018)
32. Huang, L., Joseph, A.D., Nelson, B., Rubinstein, B.I., Tygar, J.D.: Adversarial machine learning. In: Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence, pp. 43–58 (2011)
33. Jain, A., et al.: Overview and importance of data quality for machine learning tasks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3561–3562 (2020)
34. Jiang, X., Yu, F.R., Song, T., Ma, Z., Song, Y., Zhu, D.: Blockchain-enabled cross-domain object detection for autonomous driving: A model sharing approach. IEEE Internet Things J. **7**(5), 3681–3692 (2020)

35. Johnson, K.B., Wei, W., Weeraratne, D., Frisse, M.E., Misulis, K., Rhee, K., Zhao, J., Snowdon, J.L.: Precision medicine, AI, and the future of personalized health care. Clin. Transl. Sci. **14**(1), 86–93 (2021). https://doi.org/10.1111/cts.12884
36. Khan, W.U., Ye, Z., Altaf, F., Chaudhary, N.I., Raja, M.A.Z.: A novel application of fireworks heuristic paradigms for reliable treatment of nonlinear active noise control. Appl. Acoust. **146**, 246–260 (2019)
37. Kieseberg, P., Hobel, H., Schrittwieser, S., Weippl, E., Holzinger, A.: Protecting anonymity in data-driven biomedical science. In: Interactive knowledge discovery and data mining in biomedical informatics, pp. 301–316. Springer (2014)
38. Kieseberg, P., Schantl, J., Frühwirt, P., Weippl, E., Holzinger, A.: Witnesses for the doctor in the loop. In: International Conference on Brain Informatics and Health. pp. 369–378. Springer (2015)
39. Kieseberg, P., Schrittwieser, S., Mulazzani, M., Echizen, I., Weippl, E.R.: An algorithm for collusion-resistant anonymization and fingerprinting of sensitive microdata. Electron. Mark. **24**(2), 113–124 (2014)
40. Kieseberg, P., Weippl, E.R.: Security challenges in cyber-physical production systems. In: International Conference on Software Quality, pp. 3–16 (2018)
41. Kieseberg, P., Weippl, E.R., Holzinger, A.: Trust for the doctor-in-the-loop. Ercim News **2016**(1), 32–33 (2016)
42. Kurita, K., Michel, P., Neubig, G.: Weight poisoning attacks on pre-trained models. arXiv preprint arXiv:2004.06660 (2020)
43. Larsson, S., Heintz, F.: Transparency in artificial intelligence. Internet Policy Review **9**(2), 1–16 (2020)
44. Levinson, J., et al.: Towards fully autonomous driving: systems and algorithms. In: 2011 IEEE Intelligent Vehicles Symposium (IV). IEEE (2011). https://doi.org/10.1109/IVS.2011.5940562
45. Li, J.H.: Cyber security meets artificial intelligence: a survey. Front. Inf. Technol. 19(12), 1462–1474 (2018)
46. Li, Q., Guo, Y., Chen, H.: Practical no-box adversarial attacks against DNNs. arXiv preprint arXiv:2012.02525 (2020)
47. Linkov, I., Kott, A.: Fundamental concepts of cyber resilience: Introduction and overview. In: Cyber resilience of systems and networks, pp. 1–25. Springer (2019)
48. Liu, C., Xiong, H., Papadimitriou, S., Ge, Y., Xiao, K.: A proactive workflow model for healthcare operation and management. IEEE transactions on knowledge and data engineering 29(3), 586–598 (2016). doi: 10.1109/TKDE.2016.2631537
49. Liu, T., et al.: Unmanned aerial vehicle and artificial intelligence revolutionizing efficient and precision sustainable forest management. J. Clean. Prod. 127546 (2021). https://doi.org/10.1016/j.jclepro.2021.127546
50. Maletic, J.I., Marcus, A.: Data cleansing: A prelude to knowledge discovery. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 19–32. Springer, Heidelberg (2009). https://doi.org/10.1007/978-0-387-09823-4_2
51. Malle, B., Kieseberg, P., Holzinger, A.: Do not disturb? classifier behavior on perturbed datasets. In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction. pp. 155–173. Springer (2017)
52. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. arXiv preprint arXiv:1908.09635 (2019)
53. Mueller, H., Mayrhofer, M.T., Veen, E.-B.V., Holzinger, A.: The ten commandments of ethical medical AI. IEEE COMPUT. **54**(7), 119–123 (2021). https://doi.org/10.1109/MC.2021.3074263

54. Nebeker, C., Torous, J., Ellis, R.J.B.: Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC med. **17**(1), 1–7 (2019). https://doi.org/10.1186/s12916-019-1377-7
55. Nie, C., Leung, H.: A survey of combinatorial testing. ACM Computing Surveys (CSUR) **43**(2), 1–29 (2011)
56. Nischelwitzer, A., Lenz, F.J., Searle, G., Holzinger, A.: Some aspects of the development of low-cost augmented reality learning environments as examples for future interfaces in technology enhanced learning. In: Stephanidis, C. (ed.) Universal Access to Applications and Services. Lecture Notes in Computer Science (LNCS, vol. 4556), pp. 728–737. Springer, Berlin, Heidelberg, New York (2007)
57. O'Donnell, R.M.: Challenging racist predictive policing algorithms under the equal protection clause. NYUL Rev. **94**, 544 (2019)
58. Price, W.N., Gerke, S., Cohen, I.G.: Potential liability for physicians using artificial intelligence. JAMA 322(18), 1765–1766 (2019)
59. Reed, C., Kennedy, E.J., Silva, S.N.: Responsibility, autonomy and accountability: Legal liability for machine learning. Soc. Sci. Res. Netw. **243**, 1–31 (2016)
60. Regulation, G.D.P.: Regulation EU 2016/679 of the european parliament and of the council of 27 April 2016. Off. J. Eur. Union (2016)
61. Robert, K.W., Parris, T.M., Leiserowitz, A.A.: What is sustainable development? Goals, indicators, values, and practice. Environ. Sci. Policy Sustain. Dev. 47(3), 8–21 (2005)
62. Salem, A., Wen, R., Backes, M., Ma, S., Zhang, Y.: Dynamic backdoor attacks against machine learning models. arXiv preprint arXiv:2003.03675 (2020)
63. Schlosser, P., Pfirman, S.: Earth science for sustainability. Nature Geoscience 5(9), 587–588 (2012). doi: 10.1038/ngeo1567
64. Schneeberger, D., Stoeger, K., Holzinger, A.: The european legal framework for medical ai. In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction, Springer LNCS 12279, pp. 209–226. Springer, Cham (2020). DOI: https://doi.org/10.1007/978-3-030-57321-8-12
65. SDG, U.: Sustainable development goals (2018)
66. Shao, G., Reynolds, K.M., Shao, G.: Computer applications in sustainable forest management. Springer, London (2006)
67. Shapiro, A.: Reform predictive policing. Nat. news **541**(7638), 458 (2017)
68. Silva, S., Duarte, D., Valente, A., Soares, S., Soares, J., Pinto, F.C.: Augmented intelligent distributed sensing system model for precision agriculture. In: 2021 Telecoms Conference (ConfTELE). IEEE (2021). https://doi.org/10.1109/ConfTELE50222.2021.9435498
69. Singh, D., Merdivan, E., Hanke, S., Kropf, J., Geist, M., Holzinger, A.: Convolutional and recurrent neural networks for activity recognition in smart environment. In: Holzinger, A., Goebel, R., Ferri, M., Palade, V. (eds.) Towards Integrative Machine Learning and Knowledge Extraction: BIRS Workshop, Banff, AB, Canada, July 24–26, 2015, Revised Selected Papers, pp. 194–205. Springer International Publishing, Cham (2017)
70. Stafford-Smith, M., Griggs, D., Gaffney, O., Ullah, F., Reyers, B., Kanie, N., Stigson, B., Shrivastava, P., Leach, M., O'Connell, D.: Integration: the key to implementing the sustainable development goals. Sustain. Sci. **12**(6), 911–919 (2017). https://doi.org/10.1007/s11625-016-0383-3
71. Tang, D.: What is digital transformation? EDPACS - The EDP Audit, Control, and Security Newsletter **64**(1), 9–13 (2021). https://doi.org/10.1080/07366981.2020.1847813

72. Tjoa, S., Buttinger, C., Holzinger, K., Kieseberg, P.: Penetration testing artificial intelligence. ERCIM News **2020**(123), 36–37 (2020)
73. Vaio, A.D., Palladino, R., Hassan, R., Escobar, O.: Artificial intelligence and business models in the sustainable development goals perspective: a systematic literature review. Journal of Business Research **121**, 283–314 (2020)
74. Vaseghi, S.V.: Advanced digital signal processing and noise reduction. Wiley, Hoboken (2008)
75. Verhoef, P.C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J.Q., Fabian, N., Haenlein, M.: Digital transformation: a multidisciplinary reflection and research agenda. J. Bus. Res. **122**, 889–901 (2021). https://doi.org/10.1016/j.jbusres.2019.09.022
76. Villaronga, E.F., Kieseberg, P., Li, T.: Humans forget, machines remember: Artificial intelligence and the right to be forgotten. Computer Law & Security Review **34**(2), 304–313 (2017)
77. Vinuesa, R., et al.: The role of artificial intelligence in achieving the sustainable development goals. Nat. Commun. **11**(1), 1–10 (2020)
78. Wing, J.M.: Trustworthy AI. arXiv preprint arXiv:2002.06276 (2020)
79. Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M.J.: Big data in smart farming-a review. Agricultural systems 153, 69–80 (2017)
80. Zhao, S., Talasila, M., Jacobson, G., Borcea, C., Aftab, S.A., Murray, J.F.: Packaging and sharing machine learning models via the acumos AI open platform. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 841–846. IEEE (2018)
81. Zhou, X., et al.: A secure and privacy-preserving machine learning model sharing scheme for edge-enabled iot. IEEE Access **9**, 17256–17265 (2021)