Transportation Engineering 4 (2021) 100064

Contents lists available at ScienceDirect

Transportation Engineering

journal homepage: www.elsevier.com/locate/treng

Assessing whether artificial intelligence is an enabler or an inhibitor of sustainability at indicator level

Shivam Gupta[a], Simone D. Langhans[b], Sami Domisch[c], Francesco Fuso-Nerini[d], Anna Felländer[e], Manuela Battaglini[f], Max Tegmark[g], Ricardo Vinuesa[h,e,*]

[a] Bonn Alliance for Sustainability Research/Innovation Campus Bonn (ICB), University of Bonn, Bonn D-53113, Germany
[b] Basque Centre for Climate Change (BC3), Leioa 48940, Spain
[c] Department of Ecosystem Research, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, Berlin 12587, Germany
[d] Unit of Energy Systems Analysis (dESA), KTH Royal Institute of Technology, Brinellvägen, 68SE-1004 Stockholm, Sweden
[e] AI Sustainability Center, Stockholm SE-114 34, Sweden
[f] Transparent Internet, Tårup Bygade 30 Mesinge, DK-5370, Denmark
[g] Center for Brains, Minds and Machines, Massachusetts Institute of Technology, Cambridge, MA 02139, United States
[h] SimEx/FLOW, Engineering Mechanics, KTH Royal Institute of Technology, Stockholm SE-10044, Sweden

ARTICLE INFO

Keywords:
AI
Sustainability
Machine learning
Transportation system
Climate change

ABSTRACT

Since the early phase of the artificial-intelligence (AI) era expectations towards AI are high, with experts believing that AI paves the way for managing and handling various global challenges. However, the significant enabling and inhibiting influence of AI for sustainable development needs to be assessed carefully, given that the technology diffuses rapidly and affects millions of people worldwide on a day-to-day basis. To address this challenge, a panel discussion was organized by the KTH Royal Institute of Technology, the AI Sustainability Center and MIT Massachusetts Institute of Technology, gathering a wide range of AI experts. This paper summarizes the insights from the panel discussion around the following themes: The role of AI in achieving the Sustainable Development Goals (SDGs); AI for a prosperous 21st century; Transparency, automated decision-making processes, and personal profiling; and Measuring the relevance of Digitalization and Artificial Intelligence (D&AI) at the indicator level of SDGs. The research-backed panel discussion was dedicated to recognize and prioritize the agenda for addressing the pressing research gaps for academic research, funding bodies, professionals, as well as industry with an emphasis on the transportation sector. A common conclusion across these themes was the need to go beyond the development of AI in sectorial silos, so as to understand the impacts AI might have across societal, environmental, and economic outcomes. The recordings of the panel discussion can be found at:

https://www.kth.se/en/2.18487/evenemang/the-role-of-ai-in-achieving-the-sdgs-enabler-or-inhibitor-1.1001364?date=2020–08–20&length=1&orglength=185&orgdate=2020–06–30

Short link: https://bit.ly/2Kap1tE

Motivation

The fast-paced rise of artificial intelligence (AI) impacts a wide range of sectors, hence it is crucial to vigilantly assess the opportunities and challenges AI may pose for sustainable development. For instance, a recent review by Vinuesa et al. [59] reported both positive and negative impacts of AI on the 17 Sustainable Development Goals (SDGs) from the United Nations [55]. The amount of literature regarding the use and applications of AI grows rapidly, reporting on the potential AI holds to provide support in various critical contexts, such as in the medical domain or in disaster management [48,51]. In addition, other areas can significantly benefit from AI, e.g. the transportation sector through AI applications aimed at reducing the drag of a number of vehicles [23,46] as well as the urban sustainability sector through more robust methods to prevent high pollution levels and urban-heat-island effects [53]. Despite these benefits it is crucial to understand the practical implications of deploying AI-based technology and the potential negative impacts on SDGs related for example to equality or climate change. In the fol-

* Corresponding author at: SimEx/FLOW, Engineering Mechanics, KTH Royal Institute of Technology, Stockholm SE-10044, Sweden.
E-mail address: rvinuesa@mech.kth.se (R. Vinuesa).

https://doi.org/10.1016/j.treng.2021.100064
Received 18 December 2020; Received in revised form 5 March 2021; Accepted 7 March 2021

lowing, we discuss the various topics covered in the panel discussion[1] organized by KTH Royal Institute of Technology, the AI Sustainability Center and MIT Massachusetts Institute of Technology. Throughout this paper, we used the terms 'sustainable transition' referring to the radical transformation towards a sustainable society, 'sustainable development' referring to the development which helps in meeting the current requirements without compromising the ability of future generations, and 'sustainable transformation' referring to the shift that may fundamentally alter human and environmental interactions. Emphasis will be placed on (i) discerning which applications can benefit most from AI, (ii) which ones may be most significantly at risk, as well as (iii) on the impact of AI in producing a social divide.

The role of artificial intelligence in achieving the Sustainable Development Goals

The SDGs provide an excellent framework for assessing the impact of AI on different sectors as well as for identifying possible synergies among them. We started by dividing the 17 SDGs into three categories: Society, Economy, and Environment. This division is consistent with others previously reported in the literature [45,54]. Through a detailed expert-elicitation process, we found that 79% of the SDG targets can be positively affected by AI, while 35% may be negatively affected by AI development [59]. When analyzing the three different groups, we observed that the Environment category entails the highest potential with 93% of the targets being positively affected, whereas Society has the largest negative effect with 38% of the targets exhibiting a negative interaction with AI. It is important to note that, when taking into account the type of evidence indicating the connection with AI, we observed that the positive effects on the Environment and on Society were quite robust, whereas additional research is needed to evaluate the possible positive effects of AI on the Economy. Additional details can be found in the work by Vinuesa et al. [59].

Some examples of positive effects due to AI include SDGs 8 and 9, where AI-enabled technology may help create new jobs and enhance productivity, as well as SDG 1 through the use of satellite data to track poverty [25]. Other positive applications of AI target SDGs 7 and 13, specifically in the context of more efficient energy use (through smart-grid applications) and through developing more robust tools to predict and manage pollution in cities [53]. This is of great relevance e.g. in Europe where urban pollution is responsible for approx. 800,000 deaths per year [31]. AI can therefore have a positive impact on SDG 13 (climate action); note that the other two SDGs in the Environment group, i.e. SDGs 14 and 15, can also significantly benefit from AI through improved conservation and management of biodiversity and natural resources.

AI has the potential to threaten the achievement of a number of applications relevant to the SDGs within Economy and Society. First, jobs may be lost [8], although based on alternative studies the net impact may not be negative [1]. This threat is reflected in uneven opportunities to access AI computing resources, which ultimately may increase inequalities. Furthermore, AI algorithms may polarize societies and increase discrimination [15]. These can be reflected in data-driven methods to handle the COVID-19 pandemic [35], and more specifically contact-tracing smartphone applications [60], where the handling is key to avoid discrimination and polarization. Note that polarized and unequal societies can in turn have important impacts on peace and stability [39], i.e. for SDGs 16 and 17. It is important to highlight that AI can have large effects on the global energy demand. The total electricity demand of information and communications technologies (ICT) could require up to 20% of the global electricity by 2030, being at 1% today [26].

[1] https://www.kth.se/en/2.18487/evenemang/the-role-of-ai-in-achieving-the-sdgs-enabler-or-inhibitor-1.1001364?date=2020-08-20&length=1&orglength=185&orgdate=2020-06-30

Technology, individuals, and governments are three agents interacting with each other as well as with the Environment. Thereby, the speed of change in technology is so high that individuals (when it comes to technology adoption) and governments (regarding regulations) significantly lag behind. Consequently, there are large research gaps to be attended to, to manage the transition to an AI-based society. In addition, our results have exposed the significant vulnerability of infrastructures. Substantial work is needed to overcome AI gaps in transparency, safety, and ethical standards, so that everyone can benefit from the large potential of AI (see also [59]).

AI for a prosperous 21st century

AI is a double-edged sword: On the one hand, AI has great potential to contribute immensely to solving existing problems and therewith to people's well-being in the near future. As AI could replace many procedures and routines carried out by humans, the risk of human failure could be reduced and hence numerous lives could be saved by, for instance, (i) self-driving cars based on AI-technology that reduce the risk of accidents [27]; but see [41], (ii) diminishing the amount of mistakes made in hospital care [36]; (iii) improving diagnostics in e.g. lung cancer [42] and blindness [40]; (iv) accelerating medical science given new AI-technologies that also contribute towards the discovery of new drugs [16]. In addition, (v) education is likely to become more customizable and accessible around the world. On the other hand, AI can be purposefully misused causing harm, such as done by AI-driven data mining (see *Cambridge Analytica*, [13] to harvest data about human online behavior with the aim to influence elections. Likewise, AI-technology is used to carry out drone-flown bomb attacks that have killed people in various instances. Other examples include, for instance, the glitch in the AI-trading system at *Knight Capital* which led to a 440 Million USD loss [14], or *Boeing*'s safety-critical systems being potentially vulnerable to being hacked through their entertainment system software [20].

The complexity of AI technologies and the fact that decision makers often do not understand its power, are posing major challenges for a fair, ethical and sustainable use of AI [59]. The key is to obtain as much of the benefits of AI, while concurrently avoiding misuses, one that needs to be addressed urgently since AI technology is advancing steadily and becoming more powerful rapidly. A solution to this challenge is to draw a clear line between the acceptable and unacceptable use of AI. The widespread doubt that such an agreement is unrealistic can be rebutted by the example of the ban of bioweapons agreed upon during the 1975 Asilomar conference on the regulation of biotechnology, more specifically DNA technologies [6] which since then is globally followed. As an analogy, such beneficial use of AI is shown e.g. in terms of DNA technologies that keep playing an essential role in the development of vaccines. In 2017, the *Asilomar Conference on Beneficial AI* created a set of guidelines for AI research – the 23 Asilomar AI Principles (https://futureoflife.org/ai-principles/) which are being endorsed by a growing number of key players in AI research, industry and development.

The ultimate driver of an AI agreement is a globally-shared positive vision based on equal benefits for all. Despite sounding utopian, there are various examples of shared visions that have led to positive changes in the past: the invention of democracy, the scientific, industrial and computer revolutions, free healthcare, free higher education, or peace through interdependence in Europe. Taking the SDGs as our baseline, the global-AI vision may go beyond and aspire prosperity for all, knowledge through AI, AI for science, saving the climate with AI, curing cancer with AI, and health for all.

Societal and ethical considerations when applying the SDGs to AI

Prior to the SDGs the United Nations established the Millennium Development Goals which prompted the "Information and Communication

Technologies (ICTs) for Development" initiative. The Broadband Commission for Sustainable Development assessed the role that technology could play, not just in enabling the goals, but in creating technology as an accelerator of the goals. The primary intent of the Commission was to address the digital divide by pairing digitalization with sustainability, while expanding the ICT industry to enable access to the newly mobile and digital world for billions of people.

More recently, Melinda Gates and Jack Ma declared *The Age of Digital Interdependence* which included five sets of recommendations on how the international community could work together to optimize the use of digital technologies and to mitigate their risks. The recommendations included: inclusive digital economy; human and institutional capacity; human rights and human agency; trust, security, and stability; and global digital cooperation.

These considerations are especially relevant with the advent of AI given the possibilities for exponential benefits and exponential risks. AI is different from the ICTs of the past, because it is self-scaling, self-learning, and self-propagating. With the technology becoming more advanced, the lack of proper governance, transparency, explainability, and accountability becomes more pronounced. This needs to be addressed. In the context of sustainability, AI-technology results in a new type of digital pollution in the form of privacy intrusion, discrimination, and biases. Unethical outcomes are a result of the application, and not the fault of the technology, since the technology itself can be considered neutral. However, with the intention to drive technology as a force for good comes also a responsibility to implement safeguards and ensure the technology is not misused. The consequences of failing this responsibility are severe, resulting in the amplification of discrimination and inequality, inference of faulty or incorrect conclusions, and violation of privacy and trust. There are four common types of pitfalls that lead to unethical outcomes:

- Immature data and AI (insufficient training of algorithms on datasets as well as lack of representative data could lead to incorrect and unethical recommendations).
- Misuse / overuse of data (the AI application, or solution, could be overly intrusive using too broad or too deep open data or it could be used for unintended purposes by others).
- Bias of the creator (values and bias are intentionally or unintentionally programmed by the creator who may also lack knowledge/skills of how the solution could scale in a broader context).
- Data bias (the available data is not an accurate reflection of reality, or the preferred reality, and may lead to incorrect and unethical recommendations).

To create a sustainable approach to AI and to understand its impact on people and society, AI-technologies must be assessed through a multidisciplinary lens, combining the technical, legal and societal perspectives. With new and unforeseen events such as the COVID-19 pandemic, the Black-Lives-Matter movement, or economic and job-related crises (with detrimental effects on ethical questions and standards of our society), it is more important than ever to deploy data-driven technology in a safe and trustworthy way.

Transparency, automated decision-making processes, and personal profiling

We live in a big-data-analytics era where automated decision making, machine learning and profiling techniques are increasingly able to make assessments and predictions of individuals' lives based on historical data. These techniques can be harmful to individuals, because current laws and their interpretations neither provide a data subject with sufficient control over the assessments made by automated decision-making processes nor with sufficient control over how these profiles are used. The main legal and ethical issues are discrimination, biases, and lack of transparency, and individuals have no access to their profiles. Consequently, people's autonomy, dignity, and freedom are in risk.

There are many examples of assessments being made based on online automated decision-making processes:

An algorithm decides whether one stays in or out of the job market with a simple personality test. Thereby, one would not know that she/he is being discriminated against, because it is the algorithm that decides about the recruitment. Hence, people's personality and job suitability are assessed based on a 30 s video. If the candidate does not pass with a certain score, she/he may even be out of the labor market for good (https://www.8andabove.com/). Despite such algorithmic biases, AI-driven decisions may be easily accepted by humans, since the rejection of AI-driven decisions would be labor-intensive and tedious. Moreover, an AI-driven classification of job seekers as e.g. "hopeless" has the potential to trigger a process where resource deprivation by the public employment service can actually lead to the realization, and later to the (circular) validation, of this particular prediction [3].

Insurance companies collect data from social networks to predict how much users' health care could cost them.[2] Alexa can predict a user's health status through analysing voice and coughing, which is followed by sending advertisements for sore-throat products.[3] Facebook can predict your political views (https://www.citizenme.com/facebook-predict-political-view), your race, religion, and sexual orientation, and even when you are going to die.[4] Is it fair to wonder how such automated decision-making and personal profiling are regulated in the GDPR (General Data Protection Regulation)? Is there a "Right to Explanation" in the GDPR? Data subject rights to transparency are described in Articles 13–15 GDPR. The right to be notified (Articles 13–14 GDPR) is a data controller's duty and covers data provided directly by the data subject, observed data and data from a third party. Also, the right to access (Article 15 GDPR) has to be appealed for by the data subject. Regarding the right to be notified, Article 13.2 (f) GDPR informs about notification requirements when personal data is collected directly from the data subject, and article 14.2 (g) GDPR informs about notification requirements when personal data is obtained from a third-party. It has been suggested that the notification duties outlined in these two articles grant an *ex post* explanation, which means that one has the right to be notified about the existence of the logic involved, as well as the significance and the envisaged consequences of automated decision-making. But this suggestion is wrong for a reason. These notification duties precede decision-making and apply in the moment data is collected for processing and refer only to input data.

Regarding the right to access, article 15.1 (h) GDPR says that individuals have the right to access their personal data and to the following information: whether there will be automated decision-making, including profiling, meaningful information about the logic involved in the decision process, as well as the impact and the envisaged consequences of such processing for the data subject. The reality is that, with a lack of an explicit deadline for appealing, the right of access is limited to explanations of systems' functionalities. This is, again, an *ex-ante* explanation.

Furthermore, the Trade Secrets Directive states in its article 2.1[5] that trade secret is everything that is not known, it is anything that has commercial value and is anything where reasonable steps are taken to keep it secret. As a direct consequence of this definition, an individual does not have the right to be notified, as Article 13.2 (f) and 14.2 (g) of the GDPR considers, nor the right of access established in Article 15.1 (h)

[2] https://choice.npr.org/index.html?origin=https://www.npr.org/sections/health-shots/2018/07/17/629441555/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates?t=1544433231164

[3] https://www.telegraph.co.uk/technology/2018/10/09/amazon-patents-new-alexa-feature-knows-offers-medicine/

[4] https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-patent-predict-die-death-prediction-algorithm-personal-data-privacy-a8417771.html

[5] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L0943

of the GDPR to the profile that the companies have about them and the purpose it is used for. In other words, the GDPR does not contemplate in an explicit way a right to an explanation to the data subject to automated decision-making, which makes the principles of transparency and accountability impossible to apply [4].

Measuring the relevance of Digitalization and artificial intelligence (D&AI) at the indicator level of SDGs: Digitainability

Application of Digitalization and AI (D&AI) for sustainable development is being viewed as a major movement shaping the economy, environment, and society. The union of the two domains (Digitalization and AI), particularly under the framework of SDG and D&AI is less explored yet promising, since it fosters coherent efforts towards sustainable development by providing new value-generating opportunities [18,37]. However, due to the multifaceted nature and underlying complexities of the SDG 2030 Agenda, efforts to explore the meaningfulness of a D&AI application for SDGs are still sparse [65]. Only a few studies have so far investigated the nexus between D&AI and the SDGs (e.g., [17,30,59]), thus limiting the possibility for open debates and a coherent dialog between key-stakeholders .

Considering the numerous hurdles and uncertainties related to D&AI, the Bonn Alliance for Sustainability Research/Innovation Campus Bonn is working on the project "digitainable" to uncover the intricate relationship between D&AI and sustainability. In this context, "digitainability" - a combination word of the "digitalization" and "sustainability" was framed, referring to the cross-fertilization between the processes of digitalization and sustainable development. Research is underway to comprehend the relevance of D&AI at the indicator level of SDGs. Research which explores associations among SDGs and D&AI from a practical viewpoint and fosters the development of various qualitative and quantities methods considering technological as well as social perspectives, are ongoing to develop evidence for a D&AI and SDG nexus (e.g., [58,64]). However, synergies and trade-offs among SDGs and their indicators are critical aspects that need to be carefully assessed. Research on interactions to prioritize coherent actions for sustainable development is rapidly growing (e.g., [2,22,61]). However, there are gaps in terms of viewpoints for identifying the interactions and asserting fragmented outcomes, depending on the focus and context. Therefore, to better understand and generate data-backed evidence, which is yet not well recognized and hardly reflected in debates, further research towards delivering systemic-methodologies is required. Although still evolving, research on a mindful use of D&AI is starting to shape. Further work and funding are needed to continue exploring the conscious and sustainable applications of D&AI technologies for sustainable transformation from both social and technological perspectives.

Transdisciplinarity for sustainability and AI applications

Situations for sustainable transformation require balancing sustainable and unsustainable trends, which vary spatiotemporally and comprehend social and technological advancements. Sustainability as a concept is driven by values, requiring a vision of what needs to be sustained and at what cost. We can observe this by looking at the challenges in integrating the three pillars of sustainability – economy, environment, and society – which in some contexts are not mutually compatible. Considering the need for sustainable transformation within the current global change context, it is essential to understand the synergies and trade-offs between sustainable and economic development, between environment and economic development, economic development and social contexts, etc. The simplistic, reductive, and linear logic behind disciplinary knowledge production is limited in addressing complications beyond specific scope and methods. Interpretation challenges also exist between disciplines, seeking inventions in identifying the key objectives and their subtle drives for using D&AI and the need for cost-effective approaches to address the data gaps [12]. Thus, we should be considering a holistic, systemic, trans-disciplinary, and forward-looking perspective, both for D&AI and sustainability [5,43,57]. Systemically quantifiable and transdisciplinary perspectives are needed for accelerating the progress of SDGs, considering local and global perspectives and vertical and horizontal relations among key stakeholders.

Attention towards sustainability and D&AI applications

Achieving the SDGs is at the risk of a considerable interruption due to emerging global challenges such as climate change, environmental degradation, overuse of minerals, political and financial instability, and recently the COVID-19 pandemic. SDGs could benefit significantly from D&AI support, if such support is deliberated and developed with their capabilities and limitations in mind [66]. To ensure that technological interventions meet the envisaged end requirements, it is crucial to identify and realize vital factors that explain the meaningful application of technologies. In order to generate both a big picture and detailed insights about the mindful use of D&AI, practical and research implications need to be well considered. Particularly, to cope with governance-related D&AI and sustainability action uncertainties and context, novel analytical tools based on practical implications of D&AI, considering actors' perspectives, are highly desired [44,47]. Practical implications also need to demonstrate trade-offs, prioritizations, and negotiations among SDGs considering different contexts and objectives. For example, the prevention of conflicts between SDG indicator targets on certain occasions may also hamper the implementation of other SDGs [63]. Some conflicts may be necessary, as for instance to achieve a long-term integration as the relation between Indicator 9.4.1 (CO_2 emission per unit of value added) and Indicator 9.2.1 (Manufacturing value added as a proportion of GDP and per capita) [34]. Existing literature corroborates the imperative of measuring the progress by monitoring and evaluating the SDGs for achievement [11,67]. However, a significant challenge remains in uncovering suitable taxonomies to measure the dynamics of progress and comprehensive methodologies that can help explore the context-based assessment for practical applications of D&AI at the indicator level of SDGs. Methodologies and analytical tools beyond goals and target levels, i.e., those that are capable of reflecting digitainability's qualitative and quantitative aspects while capturing the complexities of the SDGs, are needed for the mindful application of D&AI for sustainable transformation.

The project "Digitainable" is dedicated to several of these aspects to address gaps identified at the intersection of D&AI and sustainability research. Notably, the research is focused on addressing the following overarching themes: 1. Identifying the role D&AI could play for SDGs' indicators; 2. Developing methods that help access the responsible deployment of D&AI, considering both social and technological perspectives; 3. Scalable D&AI interventions for SDGs; 4. Addressing the gaps between research and practice to utilize D&AI for sustainable transformation meaningfully. This research is designed to foster the context-aware, inclusive, and accountable application of D&AI in the long-term, particularly by enabling collaboration, trust, and knowledge sharing with key stakeholders. Thereby, the Digitalisation Sustainability Matrix (DSM; [24]), a participatory research approach, serves as a means in collaborative methods such as participatory action research (PAR) for the transdisciplinary knowledge curation process. PAR serves as a framework to juxtapose interdisciplinary stakeholders in the knowledge production process to co-develop research, education, and action interlinkages between science and practice [10]. The DSM connects D&AI and sustainability at the SDG indicator level. It is a two-dimensional matrix using D&AI themes and respective technologies connected to indicators of a particular SDG to seek the perspective on positive, negative, and non/unknown relevance for the diverse stakeholders who are part of the PAR. The DSM has been tested in interactive events with experts from diverse sectors and backgrounds, with the objective to capture transdisciplinary knowledge in the action-oriented dialogues concerning the

use of D&AI technologies for the indicators of SDGs. DSM can effectively trigger discussions on crucial aspects that need to be considered for D&AI's practices, which is a step towards deep-rooting the transdisciplinary perspectives for meaningful use of D&AI for SDGs (Gupta et al. [24]).

Designed by all for all

A world which heavily relies on AI technology will have to deal with, in addition to those aforementioned, two pronounced challenges specific to gender in order to provide equal outcomes and opportunities for all. The first one is fixing the gender data gap. The gender data gap is the phenomenon whereby the vast majority of information that has so far been collected globally is on men, ranging from economic data to urban planning data to medical data [38]. When such data is used as a source to help create ideal and functioning living environments, these environments only work optimally for a selected part of the population – the men. For example, until recently clinical trials have not adequately enrolled women or analyzed sex-specific differences in the collected data [62]. This is despite sex differences can be observed in various disease states in prevalence, diagnosis, severity, and outcomes as well as the different physiologies of the sexes may translate into differences in pharmacokinetics or pharmacodynamics for specific drugs [33]. The most famous example is the one of Ambien, a drug treating insomnia, which was the first one to be prescribed in sex-specific doses after discovering that women metabolized it much slower and, therefore, exhibited next-day psychomotor impairment with the original dosage [21]. Another famous example is car safety. Cars are still mostly tested only using a 50th-percentile male dummy [32], resulting in women drivers being 47% and 71% more likely to be seriously and moderately injured, respectively, in a car crash [7]. Hence, women have been disadvantaged for millennia. Now, consequently, any algorithm trained on such male-dominated datasets is unlikely to predict accurate risks or give appropriate results for everyone. Examples for that are voice and speech recognition systems, which perform worse for women than for men [49,50] or face recognition systems which provide more errors with female faces [9].

Second, we need to reach gender equality in AI professionals. Women are highly underrepresented in STEM disciplines (Science, Technology, Engineering, and Mathematics) and the AI field is also a male-dominated one. In 2016, women accounted for less than a third (29.3%) of those employed in scientific research and development across the world [56] and only 19.5% of those held manager positions in the software technology industry [28]. Systems and algorithms created by male designers and engineers might entail biases which are introduced unconsciously during various development stages e.g. the sampling of data, their annotation, the selection of algorithms, or the evaluation of metrics [52]. The consequence are algorithms which process data in a gender-biased way. Clearly, data gaps and algorithm biases do not only hinder reaching gender equality, but also apply to culture and race. One of the few studies on this topic found large racial disparities in the performance of five commercial automatic speech recognition systems, when tested by white and African American speakers [29].

Conclusion and outlook

This article comprehensively presents the multidisciplinary perspectives shared in the panel discussion by experts on the intersection of AI and sustainable development. We shed light into the various critical aspects about whether, how, and what necessary facets need to be considered for responsively harnessing the convenience offered by AI for SDGs. Overall, we provide a general overview of identifying and prioritizing the agenda for addressing the pressing research gaps concerning the advancement of AI for sustainable transformation. The significance of the nexus between the application of AI and the 2030 Agenda could be sensed as an "enduring component", which requires a thorough consideration of the threats, opportunities, synergies, and trade-offs. Thus, a converging transformation is desired from multidisciplinary perspectives taking the SDGs as our baseline. The disruption caused by AI goes beyond technology, representing an interplay between technology, society, environment, and policy. A better understanding of the socio-technical enablers and their impact on society are equally important factors. Given the exceptional prospects that AI may bring in, we suggest that in-depth exploration of the role of AI to provide support and overcome specific limitations hindering the progress of the SDG is crucial.

Various experts in the panel discussion highlighted an urgent need to address gaps, which helps in managing the transition to an AI-assisted society. Accelerating progress is urgently required to manage the significant vulnerability of infrastructures pertaining to AI. Notably, the vulnerabilities around transparency, safety, and ethical standards need to be addressed so that everyone can contribute with sureness and can benefit from the potential advantages offered by AI. Further research is required for a profound understanding of the identified challenges taking the SDGs as our baseline. SDGs can play the role of a framework to guide the advancements in AI as a key facilitator for data-driven goals and enrich further opportunities. Considering the new and unforeseen situations that impacted society and ethical effects such as COVID-19, it is more crucial than ever to deploy safe and trustworthy data-driven technologies for sustainable development. Furthermore, research is required to profoundly explore the extent to which AI can contribute to the agenda for sustainable development as well as to tackle potential tradeoffs.

Taking into account the unintended social, ethical, equity-related, and environmental concerns, there is an urgent need to assess and curb social as well as technical feasibilities for enabling AI-driven solutions for sustainable transformation. Methodologies and analytical tools to access the relevance of AI for sustainable development need to be pursued from a systematic and ethical perspective, beyond current techno-centric viewpoints. In-depth understanding beyond goals and target levels that are capable of reflecting digitainability's qualitative and quantitative aspects, while capturing the complexities of the SDGs, are highly desired to facilitate AI for sustainable transformation responsively. As can be observed from the aforementioned sections, AI discussions have been mainly focused on ethical and social considerations. However, since ultimately only staying within the planetary boundaries will allow for a transition into a prosperous and equitable future for all, environmental sustainability needs a prominent place at the forefront of developing and implementing AI-driven technologies. In particular, the role of AI in enabling climate action should be further investigated, as climate change is widely considered as the most prominent crisis to address in the 21st century, and has wide ripple effects across all SDGs (Nerini [19]).

For practitioners of AI, current research on the impacts of AI suggests the need to go beyond the simple understanding of what might be the impacts of an AI application within its field, but also understanding its potential positive and negative effects across societal, environmental and economic outcomes. The SDGs could be used to guide that evaluation, as a multi-disciplinary lens to guide the dialog [59]. For that to happen, the research and international community should develop easy-to-use standard tools to guide AI practitioners in such assessment.

There is a need to foster dialog and collaboration across disciplines to draw insights that can accelerate responsible, ethics-driven, inclusive, and context-aware AI applications for sustainable development. Hence, we believe this article provides a step towards a profound understanding of the challenges and critical aspects of utilizing AI for a prosperous 21st century.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors acknowledge the KTH Sustainability Office and the KTH Digitalization Platform for their provided funding, which enabled the organization of this panel discussion. SG acknowledges the funding provided by the German Federal Ministry for Education and Research (BMBF) for the project "digitainable". SDL acknowledges support through the Spanish Government's María de Maeztu excellence accreditation 2018–2022 (Ref. MDM-2017–0714). SD acknowledges funding by the Leibniz Competition (J45/2018).

References

[1] D. Acemoglu, P. Restrepo, Artificial Intelligence, Automation, and Work.", NBER National Bereau of Economic Research, 2018 Working Paper No. 24196.
[2] C. Allen, G. Metternicht, T. Wiedmann, M. Pedercini, Greater gains for Australia by tackling all SDGs but the last steps will be the most challenging, Nat. Sustain. 2 (11) (2019) 1041–1050.
[3] D. Allhutter, F. Cech, F. Fischer, G. Grill, A. Mager, Algorithmic profiling of job seekers in Austria: how austerity politics are made effective, Front. Big Data 3 (2020) 5, doi:10.3389/fdata.
[4] M. Battaglini, S. Rasmussen, Transparency, automated decision-making processes and personal profiling, J. Data Prot. Priv. 2 (2019) 331–341.
[5] M.U. Ben-Eli, Sustainability: definition and five core principles, a systems perspective, Sustain. Sci. 13 (5) (2018) 1337–1343.
[6] P. Berg, D. Baltimore, S. Brenner, R.O. Roblin, M.F. Singer, Summary statement of the Asilomar conference on recombinant DNA molecules, PNAS 72 (6) (1975) 1981–1984 June.
[7] D. Bose, M. Segui-Gomez, J.R. Crandall, Vulnerability of female drivers involved in motor vehicle crashes: an analysis of US population at risk, Am. J. Public Health 101 (12) (2011) 2368–2373.
[8] E. Brynjolfsson, A. McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, W. W. Norton & Company, 2014.
[9] J. Buolamwini, T. Gebru, Gender shades: intersectional accuracy disparities in commercial gender classification, Proc. Mach. Learn. Res. 81 (2018) 1–15.
[10] J.M. Chevalier, D.J. Buckles, Participatory Action Research: Theory and Methods for Engaged Inquiry, Routledge, N.p, 2019.
[11] S. Dalby, S. Horton, R. Mahon, D. Thomaz, Achieving the Sustainable Development Goals: Global Governance Challenges, Routledge, N.p, 2019.
[12] Hai-Anh H. Dang, U. Serajuddin, Tracking the Sustainable Development goals: Emerging measurement Challenges and Further Reflections, The World Bank, N.p., 2019.
[13] European Union. 2019. "European Parliament at a glance. artificial intelligence, data protection and elections." https://www.europarl.europa.eu/RegData/etudes/ATAG/2019/637952/EPRS_ATA(2019)637952_EN.p.
[14] M. Fleury, How artificial intelligence is transforming the financial industry, BBC News (2015) (New York), September 16, 2015 https://www.bbc.com/news/business-34264380 .
[15] D. Francescato, Globalization, artificial intelligence, social networks and political polarization: new challenges for community psychologists, Commun. Psychol. Glob. Perspect. 4 (2018) 20–41.
[16] D.H. Freedman, Hunting for new drugs with AI, Nature 576 (2019) S49–S53.
[17] Global Enabling Sustainability Initiative. 2019. "Digital with Purpose." 2019.
[18] R. Gouvea, D. Kapelianis, S. Kassicieh, Assessing the nexus of sustainability and information & communications technology, Technol. Forecast Soc. Change 130 (2018) 39–44.
[19] F. Nerini, et al., Connecting climate action with other Sustainable Development Goals, Nat. Sustain. 2 (2019) 674–680.
[20] A. Greenberg, A boeing code leak exposes security flaws deep in a 787's guts, WIRED (2019) July 8, 2019 https://www.wired.com/story/boeing-787-code-leak-security-flaws/ .
[21] D.J. Greenblatt, J.S. Harmatz, N.N. Singh, F. Steinberg, T. Roth, M.L. Moline, S.C. Harris, R.P. Kapil, Gender differences in pharmacokinetics and pharmacodynamics of zolpidem following sublingual administration, J. Clin. Pharmacol. 54 (2014) 282–290, doi:10.1002/jcph.220.
[22] D.J. Griggs, M. Nilsson, A. Stevance, M. David, et al., A Guide to SDG interactions: from Science to Implementation, International Council for Science, N.p.Paris, 2017.
[23] Guastoni, L., Güemes, A., Ianiro, A., Discetti, S., Schlatter, P., Azizpour, H., and Vinuesa, R.. 2020. "Convolutional-network models to predict wall-bounded turbulence from wall quantities." *arXiv preprint* arXiv:2006.12483.
[24] S. Gupta, M. Motlagh, J. Rhyner, The digitalization sustainability matrix: a participatory research tool for investigating digitainability, Sustainability 12 (21) (2020) 9283 (November), doi:10.3390/su12219283.
[25] N. Jean, et al., Combining satellite imagery and machine learning to predict poverty, Science 353 (2016) 790–794.
[26] N. Jones, How to stop data centres from gobbling up the world's electricity, Nature 561 (2018) 163–166.
[27] N. Kalra, S.M. Paddock, Driving to safety: How Many Miles of Driving Would it Take to Demonstrate Vehicle Reliability?, RAND Corporation, Santa Monica, CA, 2016 https://www.rand.org/content/dam/rand/pubs/research_reports/RR1400/RR1478/RAND_RR1478.pdf.
[28] Kersley, R., Klerk, E., Boussie, A., Longworth, B.S., Natzkoff, J.A., and Ramji, D.. 2019. "The CS gender 3000 in 2019: the changing face of companies." p. 15. https://www.credit-suisse.com/about-us-news/en/articles/news-and-expertise/cs-gender-3000-report-2019-201910.html.
[29] A. Koenecke, A. Nam, E. Lake, J. Nudell, M. Quartey, Z. Mengesha, C. Toups, J.R. Rickford, D. Jurafsky, S. Goel, Racial disparities in automated speech recognition, PNAS 117 (14) (2020) 7684–7689, doi:10.1073/pnas.1915768117.
[30] O. Kostoska, L. Kocarev, A novel ICT framework for Sustainable Development Goals, Sustainability 11 (7) (2019) 1961.
[31] J. Lelieveld, et al., Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions, Eur. Heart J. 40 (2019) 1590–1596.
[32] A. Linder, W. Svedberg, Review of average sized male and female occupant models in European regulatory safety assessment tests and European laws: gaps and bridging suggestions, Accid. Anal. Prev. 127 (2019) 156–162, doi:10.1016/j.aap.2019.02.030.
[33] K.A. Liu, N.A. Dipietro Mager, Women's involvement in clinical trials: historical perspective and future implications, Pharm. Pract. (Granada) 14 (1) (2016) 708 10.18549/PharmPract.2016.01.708.
[34] S. Liu, Interlinkages between indicators of Sustainable Development Goals: evidence from seven low income and lower middle-income countries, Sustain. Dev. Res. 2 (2020) 1June, doi:10.30560/sdr.v2n1p58.
[35] W. Naudé, R. Vinuesa, Data, Global development, and COVID-19: Lessons and Consequences, World Institute for Development Economic Research (UNU-WIDER), 2020 *No. wp-2020-109*.
[36] D.B. Neill, H.J. Heinz, Using artificial intelligence to improve hospital inpatient care, IEEE 28 (2) (2013) 92–95.
[37] T. Osburg, C. Lohrmann, Sustainability in a Digital World, Springer, N.p., 2017.
[38] C.C. Perez, Invisible Women, Penguin Random House, London, 2019.
[39] M. Petit, Towards a critique of algorithmic reason. A state-of-the-art review of artificial intelligence, its influence on politics and its regulation, Quad. Del CAC 44 (2018).
[40] S. Ravindran, How artificial intelligence is helping to prevent blindness.", Nat. Outlook The Eye. (2019), doi:10.1038/d41586-019-01111-y.
[41] L.P. Robert, Are automated vehicles safer than manually driven cars? AI Soc. 34 (2019) 687–688.
[42] N. Savage, Another set of eyes for cancer diagnostics, Nat. Outlook 579 (2020) S14–S16.
[43] R.W. Scholz, E.J. Bartelsman, S. Diefenbach, L. Franke, A. Grunwald, D. Helbing, R. Hill, et al., Unintended side effects of the digital transition: European scientists' messages from a proposition-based expert round table, Sustainability 10 (6) (2018) 2001.
[44] M.S. Smith, C. Cook, Y. Sokona, T. Elmqvist, K. Fukushi, W. Broadgate, M.P. Jarzebski, Advancing sustainability science for the SDGs, Sustain. Sci. 13 (6) (2018) 1483–1487.
[45] SRC, Stockholm Resilience Centre (SRC)Contribution to the 2016 Swedish 2030 Agenda HLPF Report, Stockholm University, 2017.
[46] P.A. Srinivasan, L. Guastoni, H. Azizpour, P. Schlatter, R. Vinuesa, Predictions of turbulent shear flows using deep neural networks, Phys. Rev. Fluids 4 (2019) 054603.
[47] K. Sullivan, S. Thomas, M. Rosano, Using industrial ecology and strategic management concepts to pursue the Sustainable Development Goals, J. Clean. Prod. 174 (2018) 237–246.
[48] W. Sun, P. Bocchini, B.D. Davison, Applications of artificial intelligence for disaster management, Nat. Hazards 2020 (3) (2020) 1–59.
[49] R. Tatman, Gender and dialect bias in YouTube's automatic captions, in: Proceedings of the First Workshop on Ethics in Natural Language Processing, 2017, pp. 53–59.
[50] R. Tatman, C. Kasten, Effects of talker dialect, gender & race on accuracy of Bing Speech and YouTube automatic captions, Interspeech (2017) 934–938, doi:10.21437/Interspeech.2017-1746.
[51] D.S. Ting, Y. Liu, P. Burlina, X. Xu, N.M. Bressler, T.Y. Wong, AI for medical imaging goes deep, Nat. Med. 24 (5) (2018) 539–540.
[52] S. Tolan, Fair and unbiased algorithmic decision making: current state and future challenges, JRC Tech. Rep. (2018).
[53] P. Torres, S.Le Clainche, R.. Vinuesa, On the experimental, numerical and data–driven methods to study urban flows, Energies 14 (2021) 1310.
[54] UN, United Nations Economic and Social Council. 2019. "Sustainable Development."
[55] UN General Assembly (UNGA). 2015. *A/RES/70/1Transforming our world: the 2030 Agenda for sustainable development*. N.p.: Resolut. 25, 1–35.
[56] UNESCO Institute for Statistics. 2019. "Fact sheet No. 55: women in science." p. 2. http://uis.unesco.org/sites/default/files/documents/fs55-women-in-science-2019-en.pdf.
[57] S.C. Valencia, D. Simon, S. Croese, J. Nordqvist, M. Oloko, T. Sharma, N.T. Buck, I. Versace, Adapting the Sustainable Development Goals and the new urban agenda to the city level: initial reflections from a comparative research project, Int. J. Urb. Sustain. Dev. 11 (1) (2019) 4–23.
[58] M. van der Velden, ICT and sustainability: looking beyond the anthropocene, in: Proceedings of the IFIP International Conference on Human Choice and Computers, Springer, 2018, pp. 166–180. N.p.
[59] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S.D. Langhans, M. Tegmark, F.F. Nerini, The role of artificial intelligence in achieving the Sustainable Development Goals, Nat. Commun. 11 (233) (2020) org/, doi:10.1038/s41467-019-14108.
[60] R. Vinuesa, A. Theodorou, M. Battaglini, V. Dignum, A socio-technical framework for digital contact tracing, Results Eng. 8 (2020) 100163.
[61] N. Weitz, Towards systemic and contextual priority setting for implementing the 2030 agenda, Sustain. Sci. 13 (2) (2018) 531–548 *(Springer)*.
[62] H.P. Whitley, W. Lindsey, Sex-based differences in drug activity, Am. Fam. Physician 80 (11) (2009) 1254–1258.

[63] R. Wong, J. van der Heijden, Avoidance of conflicts and trade-offs: a challenge for the policy integration of the United Nations Sustainable Development Goals, Sustain. Dev. 27 (5) (2019) 838–845.
[64] J. Wu, S. Guo, H. Huang, W. Liu, Y. Xiang, Information and communications technologies for sustainable development goals: state-of-the-art, needs and perspectives, IEEE Commun. Surv. Tutor. 20 (3) (2018) 2389–2406.
[65] T. Yigitcanlar, F. Cugurullo, The sustainability of artificial intelligence: an urbanistic viewpoint from the lens of smart and sustainable cities, Sustainability 12 (20) (2020) 1–24 (Switzerland), doi:10.3390/su12208548.
[66] T. Yigitcanlar, K.C. Desouza, L. Butler, F. Roozkhosh, Contributions and risks of artificial intelligence (AI) in building smarter cities: insights from a systematic review of the literature, Energies 13 (6) (2020), doi:10.3390/en13061473.
[67] A. Yonehara, O. Saito, K. Hayashi, M. Nagao, R. Yanagisawa, K. Matsuyama, The role of evaluation in achieving the SDGs, Sustain. Sci. 12 (6) (2017) 969–973, doi:10.1007/s11625-017-0479-4.