AI and Ethics
https://doi.org/10.1007/s43681-023-00323-3

ORIGINAL RESEARCH

Challenging AI for Sustainability: what ought it mean?

Sophia Falk[1] · **Aimee van Wynsberghe**[1]

Received: 20 April 2023 / Accepted: 5 July 2023

Abstract

This paper argues that the terms 'Sustainable artificial intelligence (AI)' in general and 'Sustainability of AI' in particular are overused to the extent that they have lost their meaning. The AI for (social) good movement is a manifestation of this trend in which almost any application used in the context of healthcare or agriculture can be classified as AI for good regardless of whether such applications have been evaluated from a broader perspective. In this paper, we aim to create a common understanding of what the 'AI for Sustainability' movement ought to mean. We distinguish between two possible AI for Sustainability applications, namely those that fulfill the necessary conditions and those that fulfill the sufficient conditions. The former are purely predictive systems that serve as information providers. The latter are directly involved in an activity that contributes to a sustainability goal. We argue that taking action is a key element in distinguishing between these two application groups, as inaction is the key bottleneck in effectively tackling climate change. Furthermore, we question how effective the use of AI applications can be for sustainability when the systems themselves are inherently unsustainable. Hence, AI for Sustainability should include both an action that contributes to a sustainable end goal as well as an investigation of the sustainability issues of the AI system itself. Following that, Sustainable AI research can be on a gradient: AI in an application domain, AI towards sustainability, and AI for Sustainability.

Keywords Sustainable AI · Carbon emissions · AI ethics · Sustainable development goals · Climate action

1 Introduction

Increased climate awareness and dedication to combat the climate crisis are shaping the rise of technology applications to solve, or at least mitigate concerns, at the core of the climate crisis. There is an overall trend that the latest technology developments—namely artificial intelligence (AI) and machine learning (ML)—are also increasingly being used to tackle sustainability issues and the global climate crisis [10, 17, 44]. At the same time, growing attention is paid to the environmental costs of making and using AI and ML applications [12–14, 23, 26, 28–30, 34]. It has become known that the training of AI systems with large datasets is associated with high energy-related carbon emissions, while the cooling of data centers is related to high water and again energy consumption levels. Consequently, the environmental footprint of the servers running AI systems can no longer be ignored. Moreover, the production of the hardware needed to run AI systems relies on the unsustainable extraction of critical minerals and metals.

Clearly then, there are two sides to the umbrella term 'Sustainable AI'. The field of Sustainable AI has been proposed by van Wynsberghe [43] to address both the use of AI *for* sustainable ends as well as the Sustainability *of* AI itself [43]. Since the time of this original publication, there have been numerous publications delving deeper into the meaning and manifestation of Sustainable AI. The field includes both understanding AI as physical infrastructure that requires minerals, metals, and energy to operate, as well as applying AI to pursue sustainability, or desirable goals [6]. Indeed, the Sustainability of AI is very difficult to assess, as many different variables and factors play a role, but it seems that the underlying notion is nevertheless easier to grasp—how to measure and make sense of the environmental costs associated with the making and using of AI/ML.

While AI for Sustainability is a concept that has gained traction in recent years, this debate lacks a clear and common understanding of what AI for Sustainability really

✉ Sophia Falk
falk@iwe.uni-bonn.de

1 Bonn Sustainable AI Lab, Universität Bonn, Institut für Wissenschaft und Ethik, Bonner Talweg 57, 53113 Bonn, Germany

means, and further what it ought to mean. It is nothing new that the term 'sustainability' on its own is often used to create a good image of a product or service. The sustainability concept in general can be criticized as a seemingly contentless 'fuzzy notion' or non-committal 'all-purpose glue' [45]. When analyzing the use of the word 'sustainability', it becomes obvious that the term is a great example for a fuzzy concept used as a buzzword in all sorts of contexts in academia and beyond. Because of its lack of definition and its capacity to legitimize 'nearly everything', the term sustainability needs to be questioned [18]. Apparently, its meaning is either unclear or irrelevant to many users, so the term as such seems to gradually lose its meaning.

One might assume that AI for Sustainability can be defined in terms of application domain, i.e., if AI is used for achieving the UN's sustainable development goals (SDGs), then it is AI for Sustainability [32]. Accordingly, if an AI application is used in agriculture or for the reduction of energy consumption, then it may be considered AI for Sustainability. We suggest this is not specific enough. Even when AI is used in certain application domains, it does not guarantee a successful end goal. The problem with this is that there are a variety of ways in which AIs can be used for the SDGs. On the one hand, AI can be used to make predictions about future events, on the other hand AI is used as part of the solution to future events. For example, AI has powerful forecasting capabilities and smart grid systems to manage the supply and demand of renewable energy [1, 27]. More accurate weather forecasts can optimize efficiency, reduce costs, and prevent the generation of unnecessary carbon pollution. Other AIs support biodiversity [16] or inform greener transportation networks [33]. This already shows how differently the systems are used to supposedly fulfill the same purpose. A second problem is that most, if not all, attempts to use AI for the SDGs do not consider the sustainability of the AI system itself. Can this really be considered AI *for* Sustainability then?

In this paper, we will tackle two objectives. First, to create more clarity on the topic of AI for Sustainability, we will propose necessary and sufficient criteria concerning the labeling of research as AI for Sustainability. We suggest that it is not enough to use AI applications for monitoring or to make predictions about future events, but that some form of action connected to it is required for the AI to become a part of the solution to the outcome. This is an important distinction to ensure that we reduce the inflationary use of the term 'AI for Sustainability', and thereby prevent the term from becoming meaningless. Second, we wish to challenge the very notion of AI for Sustainability by questioning whether such a label is ever warranted if the application is not at the same time addressing Sustainability *of* AI issues.

2 Sustainable AI: the next chapter

In recent years, growing attention has been directed toward the sustainability aspects of AI, or more accurately put, the environmental consequences of making and using AI. Leading researchers such as Strubell, Brevini, Coeckelbergh and many more address strong sustainability concerns from different perspectives in their research. Strubell et al. [41], Lacoste et al. [26], and Dodge et al. [14] have focused their investigation on the Sustainability of AI systems by estimating the energy consumption of AI algorithms, especially in ML and natural language processing (NLP) methods. Crawford and Joler [13] take a more general approach by examining different life stages of an AI system, outlining the complex relationships between environment, resources, work force, labor conditions, intellectual human capital, users of an AI system as part of a sustainability analysis.

Further, the fields of AI for Sustainability, and AI for the SDGs, have become an important part of the global AI ethics discussion. AI ethicists Floridi and Coeckelbergh explore the potential of AI methods to be applied to help deal with a wide range of environmental and social issues [10, 17]. While Coeckelbergh calls for responsible use of 'AI for climate' to create a greener, more sustainable world and mitigate climate change [14], the ethical framework by Floridi et al. [17] investigates the potential of AI to solve complex environmental and societal problems to build the foundation of a 'Good AI society'. According to Floridi et al., AI put at the service of human intelligence has the possibility to greatly enhance human agency [17].

In 2021, van Wynsberghe suggested that it is not enough to look at the impacts of AI on the environment and the application of AI for sustainable ends in isolation. Rather, for Sustainable AI, both branches need to be considered, namely one side being AI *for* Sustainability and the other the Sustainability *of* AI [43]. The question now is what does Sustainable AI mean in practice? Is it a combination of the two, meaning AI for Sustainability can only be considered as such if the Sustainability of AI is also considered? This means that at least AI for Sustainability can only be defined in the context of Sustainability of AI. Yet, the Sustainability of AI can also be considered in isolation, without the purpose of the AI application ever having to be defined or known. However, this paper is about addressing the branch AI for Sustainability and to question what that means, and/or what it ought to mean.

Therefore, the scope of what we conceive of as an 'AI system' is relevant to the following discussion. We refer to AI at two different system levels. When we talk about AI for Sustainability, we refer to the use of the trained algorithm itself, i.e., the software level. However, we do not limit our research to convolutional neural networks, deep learning (DL), ML or any specific method for developing AI algorithms. Furthermore, we include as resulting software outputs any expert system, natural language processing, speech recognition or machine vision associated with addressing sustainability issues in some way. At this moment, it is not a question of whether we agree that it is really an AI for Sustainability, but what is currently classified as such in research and development.

When we consider the Sustainability of AI, we refer to both the hardware level of the systems on which the algorithms are trained and on a software level, the energy consumption of this process. One limitation of this is that non-AI-related programs also run on the same hardware. Nevertheless, data centers, where AI, among other programs, is trained, are the closest link that can be made to measure the energy and material consumption-related ecological footprint.

3 AI for Sustainability in the literature

As part of the research in Sustainable AI, the authors of this paper were interested in knowing about the kinds of research in this space, and to do so through an extensive literature review. One of the reasons for this was our belief that there has been an increase in the use of the term ‘sustainability’ in the AI discourse of late. Second, there is no common understanding of the kind of research that can be labeled as AI for Sustainability.

The literature search by title, abstract, and keywords[1] related to Sustainable AI in the Web of Science and Scopus databases resulted in 8756 publications. In the first step, duplicates and certain publication formats, such as non-peer-reviewed articles, conference proceedings, and book chapter, were excluded leaving us with approximately 6700 publications. The fact that AI plays a prominent role in the literature in the context of (environmental) sustainability is confirmed by the sheer number of publication results. However, this does not indicate whether it is a rather new trend. Therefore, the remaining literature was analyzed by publication year. At this stage, no literature is excluded by title, topic, or abstract, as we are interested in all publications associated in the databases with (environmental) sustainability in the context of AI.

Between 2000 and 2016, the number of publications indexed with keywords related to Sustainable AI remained well below 100 publications/year for a decade before slowly increasing and finally reaching the 200 publications/year mark in 2017. Then in 2018, the topic ‘Sustainable AI’ gained traction and the number of publications increased exponentially (Fig. 1). In 2022, we have almost 2000 publications/year. Therefore, the hypothesis that the number of publications on the topic of sustainability in the AI discussion has recently been increasing has firm evidence.

Similar results were reported in the worldwide AI ethics review of 200 guidelines and recommendations for AI governance [25]. Of the 200 government documents reviewed that were published between 2014 and 2022, 30.5% were published in 2018. They reasoned that this was due to historical events during this period, such as the first death from an Uber self-driving car in 2018 and the disclosure of the Cambridge Analytica case, in which personal data were used without consent for personal profiling and targeting advertising for political purposes. These and other incidents may explain the increasing attention to AI ethics and regulation [25]. The review demonstrates that governments made an effort to regulate AI and started to turn it into guidelines. Seemingly, the topic of ethical Sustainable AI has benefited from the increased interest in AI.

However, it was challenging to review the sustainability concept in the context of the two branches of Sustainable AI. The first objective was to divide the publications into two groups: those dealing with AI for Sustainability and those dealing with the Sustainability of AI. Viewing the publication results, it became obvious that there is no common understanding of the kind of research that can be labeled AI for Sustainability. The research field is flooded with different types of publications in which the buzzword ‘sustainability’ is used in an unconscious or non-committal way, leading to the AI for Sustainability movement losing its meaning.

Hence, we did not continue with a further analysis of the publications and instead started questioning, what should be expected of the system to contribute to sustainability, and what do we mean by ‘sustainability’?

3.1 What should be expected of the term sustainability?

The forward-looking nature of ecological sustainability was defined by the World Commission on Environment and Development (Brundtland Report) as ‘meeting today’s needs without compromising the ability of future generations to meet their needs’. This approach is important but poses some difficulties since future generations’ needs are neither easy to define nor determinate [31]. Further, environmental sustainability in the broadest sense aims to protect the earth as a whole in the long-term. This covers, among others,

[1] Using the following keywords (some of them written with a wildcard character, e.g., sustainab*): AI, ML, DL, neural networks, and sustainability, ecological, climate change, climate solutions, environmental cost, climate cost, carbon, energy, natural resources, biodiversity, and Sustainable Development Goals, global efforts, policy, justice, ethic, and moral philosophy. Following several review processes, the search string was finalized into a set of three search strings.

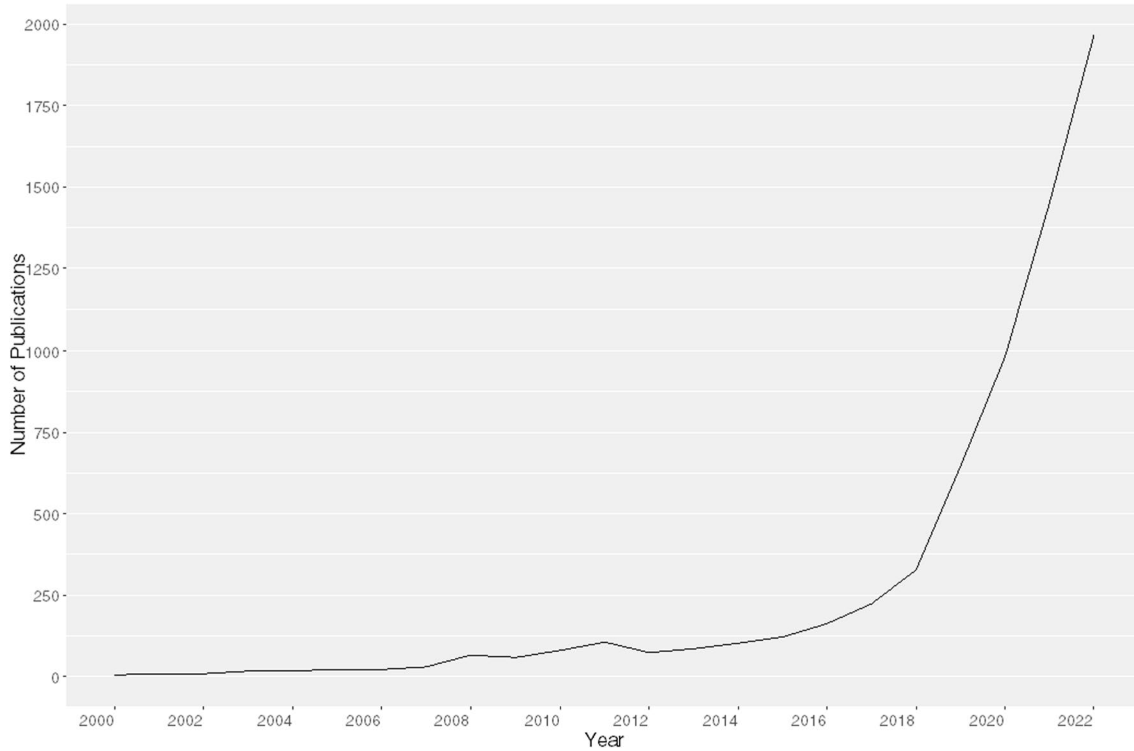

Fig. 1 Increase in publications in the field of 'Sustainable AI' between 2000 and 2022 (number of publications per year) (own illustration)

resource management, environmental protection, preservation of wildlife, habitat restoration, conservation, and preservation of natural resources. Thus, sustainability refers to the ecosystem integrity and carrying capacity of the natural environment. This implies that natural resources must not be harvested faster than they can be regenerated while waste must not be emitted faster than it can be assimilated by the environment [31]. Since we disregard this implication, we are confronted with the climate crisis. The impacts of which provide a convincing argument for the importance of environmental sustainability [31].

Multiple frameworks and indicators have been proposed to track sustainability and sustainable development over time, such as the planetary boundary framework [37], the sustainable development goals from the UN or life cycle assessments that quantify emissions from all life stages of a product or service [15]. In a methodological framework assessment to identify shared environmental sustainability indicators of the previous three frameworks, Dong and Hauschild [15] find a shared focus on seven environmental impact categories, i.e., climate change, acidification, ozone depletion, eutrophication, chemical pollution, freshwater use, and change in biosphere integrity/biodiversity [15]. These indicators assess under what conditions the natural environment remains productively stable and resilient to support human and animal life and development, and if the current environmental state meets these conditions. Overall, environmental sustainability cannot be achieved through isolated initiatives, but through a combination of several actions that address each of these indicators. Thus, the goal of an AI for Sustainability must support at least one of the seven environmental sustainability indicators. In combination and collaboration, multiple AIs for sustainability can then potentially achieve the overall goal of sustainability. From hereafter, when the AI system is referred to achieve or contribute to sustainability or a sustainability goal, at least one of the previously described indicators must be positively impacted by the AI system to support the productive, stable, and resilient natural environment.

3.2 Common attributes of AI for Sustainability

When examining the publication results through this lens, the first distinction that can be drawn is in which industry or sector, e.g., transportation, agriculture, smart cities, energy systems, the AI system is employed to reportedly improve sustainability. Upon closer inspection, we thought it is necessary to identify common attributes of the different systems within their industry group, instead we found two different kinds of systems presented in this research field in general. While numerous papers explore AI systems for monitoring and information generation, applied to prediction and forecasting models, other papers investigate AI systems that have an agentic action component that directly contributes to the end result of sustainability. AI systems can be distinguished by the existence or absence of an artificial agent[2] providing the system with the capacity to exercise an action [35]. In short, there are monitoring/ predictive AI systems and agentic AI systems that perform an action. Examples of the former include predictive modeling and demand forecasting such as solar-irradiation forecasting [2], wind speed forecasting [11], country-specific energy import predictions [9], short-term power load forecasting [5], building energy usage and power consumption predictions [3, 46], extreme weather events forecasting [4, 36], suitable habitat forecasting and conservation planning [16] and so on. Examples of the latter can include pesticide reduction in agriculture [40], optimal (renewable) energy management in smart grids, and power flow optimizations in microgrids [1, 8, 27] or optimal traffic scheduling techniques [33]. However, numerous publications focus on economic efficiency gains, and describe as a side effect a trend toward reduced resource dependency that contributes to environmental sustainability goals, such as reduced energy consumption. Yet, efficiency gains often lead to rebound effects rather than improved sustainability [21, 23, 47].

Nevertheless, there is some concrete research on the impact AI has on sustainable development and how AI methods could be used to further support the SDGs. Vinuesa et al. [44] find that AI can enable the accomplishment of 134 out of 169 targets of the SDGs. They group the SDGs according to the three pillars of sustainability (economy, society, and environment) and identify in the category AI and the environment 25 targets (93%) for which AI could act as an enabler [44]. Sætra [38] already pointed out that the main problem of the study by Vinuesa et al. [44] is that it is very quantitative and empirical, aiming to describe all goals and subgoals in the context of a very condensed article. As a result, it is an article with many 'bold conclusions and attractive numbers and percentages' that need more comprehensive explanations and deeper analysis [38]. Adding to the shortcomings identified by Sætra [38], we suggest that Vinuesa et al.'s positive result is reinforced by the way the authors define AI and what capabilities they expect from a software technology in order for the technology to be included in their assessment as AI for achieving the SDGs. Based on their definition of an AI for Sustainability, any software technology that can make predictions has the capabilities of perception, decision-making, automatic knowledge extraction and pattern recognition from data, interactive communication and logical reasoning is included [44]. This generous inclusion of different sub-areas contributes to such a positive outcome that demonstrably sheds a positive light on the use of AI for achieving sustainability goals, especially around environmental sustainability. Statements and research results like this contribute to the hype cycle of technologies as a quick fix, i.e., technological solutionism.

4 What should AI for Sustainability mean?

In our approach, we aim to create a stricter definition and expectation on how AI systems can contribute to the achievement or approximation of sustainability. Here we can draw a comparison to Green's [19] critique of the lack of a common understanding in computer science of what 'good' in the sense of 'social good' means. His statement 'when the movement encompasses everything, it stands for nothing' [19] perfectly describes the situation of the AI for Sustainability movement. Therefore, we need to find a way to limit the notion to prevent the misappropriation of the term and/ or the risk of 'ethics washing' with the term.

4.1 The need for action

When researching the literature for Sustainable AI, the paper by Atmaja and Fukushi [4] discussing coastal flooding predictions appears. To be fair, the authors did not index their framework as 'AI for Sustainability' themselves; however, the Elsevier data science teams have built extensive keyword queries, supplemented with ML, to map documents to SDGs, and mapped the following 4 to that article; Sustainable cities and communities, climate action, life below water, partnership for the goals. As a result, this paper and many others appear in a literature search in the category Sustainable AI. The paper states that coastal flooding predictions utilizing spatial ML could aid climate-related disaster risk analysis and contribute to risk reduction and policy suggestions to improve disaster resilience. The research aims to archive recent studies on the application of geospatial science empowering AI, notably ML in coastal flood risk assessment [4].

To be sure, this is a noble effort and research that ought to be continued. But such a prediction does not say anything about how to prevent future flooding or how to protect the individuals at risk. Instead, it measures the consequences of previously unsustainable practices in society. Further, this information could be used against those communities, or it could be hidden from them. Such an event happened in Germany in July 2021 where the information, i.e., the high

[2] We are conscious about the fact that the term 'agent' in philosophy of mind has a different approach to its definitions. The nature of agency in this context is detached from debates on free will, practical rationality, and moral responsibility [39]. In the computer sciences, the agent character is rather defined by the standard conception of action, where an agent is a being with the capacity to act and agency purely denotes the exercise or manifestation of this capacity [39]. Depending on the human–system interaction, the action can be executed by an agent, a human, or a combination of both [20].

likelihood of flood danger in certain areas, was known to policy makers but was not widely shared with the affected local communities. Local authorities had not informed people on the night of the flood [22]. Consequently, intense rainfall caused severe floods and 184 fatalities occurred in the German federal states of Rhineland-Palatinate (RP) and North Rhine-Westphalia (NW). According to data from an online survey ($n = 1315$), 35% of respondents from NW and 29% from RP did not receive any warning. Of those who received warnings, 85% did not anticipate particularly severe flooding, and 46% reported lacking situational knowledge regarding appropriate protective behavior [42]. So clearly, having the information does not necessarily change the outcome of the situation. Thus, can the above-mentioned study by Atmaja and Fukushi [4] be considered research within the field of AI for Sustainability?

4.2 Illustrating the many ways in which AI for Sustainability appears in the literature

For a better understanding of how we arrive at our definition, four concrete examples are given in which applications are categorized according to necessary and sufficient conditions. A necessary condition is a condition that must be met for a certain event to occur. A sufficient condition is a condition that will produce the event. The necessary condition must be present but standing on its own is not sufficient to ensure the existence of the event [7]. In this paper, the desired event is achieving sustainability, i.e., contributing to the positive influence on one of the seven environmental impact categories listed above. Not all AI systems meet the sufficient condition to ensure that the event 'sustainability' will occur by performing an action. Therefore, we distinguish AI systems by determining whether they meet the necessary or sufficient condition. The necessary condition is met if the AI system contributes to information generation (e.g., based on monitoring), which is then used to perform an informed action fulfilling the sufficient condition. We do not classify the publications or the research per se, but the AI application discussed, or the possible application of an AI system proposed in the respective publication. In the following examples, we will illustrate the variety of applications that have so far fallen under the umbrella of AI for Sustainability and highlight some applications that meet the sufficient condition, and can, therefore, be labeled AI *for* Sustainability, while other applications meet the necessary condition, and should, thus, be labeled AI *towards* Sustainability instead. In short, we argue that to be considered AI for Sustainability both the necessary and sufficient conditions ought to be met.

Example 1. AI in the energy sector Short-term prediction of the electricity load of individual households is a challenge in the research areas of smart grid management/planning, feasible energy use, energy conservation and electricity market bidding system design. The rationale for this is the unpredictability and uncertainty in the electricity consumption pattern of individual households [5]. The user's electricity consumption profile varies hourly, daily, weekly, and seasonally due to the different environmental and seasonal influences. In this example, the focus is on exploring and evaluating ML models to accurately predict the user's electricity consumption profile for energy management in a smart community [5].

The predictions from this model can be used as a basis for the subsequent implementation of optimal decision rules and therefore this application meets the necessary criteria and should be labeled AI towards Sustainability. In general, energy demand forecasts are essential to help, e.g., policymakers, to identify changes in demand and supply under certain conditions and are crucial for energy planning. Nonetheless, they do not necessarily encourage decision-makers to invest in additional renewable energy sources. Instead, policymakers can use the forecasts to ensure fossil fuel imports from trading partners to meet future energy demand. Hence, any prediction by AI systems can be as accurate as possible but will only meet the necessary conditions.

Example 2. AI in renewable energy The use of advanced ML techniques to optimize power flow in a community microgrid is an illustration of an application that satisfies the sufficient condition. The increasing penetration of distributed renewable energies regularly causes considerable, and rapid fluctuations in the power and voltage profiles on the electrical grid. Since renewable energy loads are nonlinear, the penetration of distributed renewable energy sources forms a challenge for power system efficiency. Real-time control strategies that are swift and precise have become essential for ensuring that the power system operates at its peak performance [1].

By resolving difficulties in real-time to improve power flow while taking into account the operational restrictions of the community microgrid, ML is seen as a promising tool for regulating the fluctuations of renewable sources and loads [1]. Aldahmashi and Ma's [1] proposed ML algorithms can make fast decisions even with highly uncertain variables in the power systems. Thereby, the ML system reduces power loss and increases power flow from renewable sources. The algorithms assessed have agentic decision-making capacities, helping to reduce dependence on fossil fuels, reduce energy loss, and thus directly contribute to a more sustainable consumption pattern and a reduction in energy-related emissions [1]. This application satisfies the necessary conditions since the ML method employs the capability to learn the best approaches to arrive at optimal solutions by monitoring and extracting important information from past data. The sufficient criterion is subsequently met by the generated information to carry out an action that has a beneficial impact on climate change.

Example 3. AI and climate change Concerning climate change in general, the detection and identification of extreme weather events in large-scale climate simulations is an important concern for risk management, government policy decisions, and improving our fundamental understanding of the climate system. Supervised convolutional neural networks (CNNs) can achieve acceptable accuracy in classifying well-known types of extreme weather events when large amounts of labeled data are available [36]. The system is a smart monitor that helps better understand and mitigate the effects of climate change. However, the value proposition is not enough to meet the sufficient condition. The system does not trigger any effect to influence sustainability and can be labeled AI towards Sustainability.

Example 4. AI in agriculture Looking at sustainable agriculture, the use of an AI system (xarvio) for digital farming solutions in crop protection optimization translated in a European case study to a 30% decrease in fungicide usage on field trial cereal crops and a 72% decrease in tank leftovers reducing environmental pollution. In Brazil, the use of the AI application (using computer vision) resulted in a 61% average savings in weed spraying, cutting back on almost two-third of herbicide and water consumption [40].

This AI system meets the necessary and sufficient conditions as informed decisions are based on its smart monitoring and its collected information on sustainable agriculture. The system contains multiple annotated data points for all possible scenarios, such as weather data, light, season, weed locations, and so on. Based on this, the system decides in milliseconds how many milligrams of fungicide or herbicide to spray. The agentic AI immediately triggers an action that has a significant impact on the natural environment by reducing the agrochemical load on the field. The use of this AI application directly leads to more sustainable consumption and production patterns [40], thus meeting the sufficient condition.

5 AI for Sustainability is not possible without the Sustainability of AI

When analyzing the effectiveness of an AI for Sustainability and positive impact on the climate crisis, it is important to also examine the sustainability of the AI itself. How useful can the impact of an AI system be towards sustainable ends if its own development and use defeats the purpose of its existence in the first place?

Since the pioneering study by Strubell et al. [41] on the vast energy-related carbon emissions from training an AI model, an increasing number of researchers [14, 26, 28, 29] investigate the carbon emissions that result from training a variety of freely available NLP and other AI methods. All are coming to the same conclusion: energy consumption remains a relevant sustainability issue. Large language models are among the biggest ML models, encompassing up to hundreds of billions of parameters, requiring several weeks of GPU hours to train, while emitting carbon in the process. And the trend in recent years suggests that model sizes will continue to grow [24, 29]. For example, training BLOOM, an open-access multilingual language model, consumed 433 MWh resulting in 25 tons of CO_2 equivalent emissions [29, 30]. In other words, BLOOM's training consumed enough energy to power the average US American home for 41 years [30].

Besides, only one training round is considered in these calculations, whereas numerous intermediate models are trained before the final model is also trained repeatedly. In addition, after the models have been published, they are frequently refined by additional inference training to eliminate any deficiencies that are discovered (especially in commercial applications such as ChatGPT). To make the scale even more evident, the environmental impact of one training round for GPT-3 was estimated to be 1,287 MWh resulting in 502 tons of CO_{2eq} emissions [30].

In addition, the focus so far has been on estimating the energy consumption and associated carbon emissions of energy used to run specialized hardware such as GPUs [26, 28, 29, 41]. However, it is important to consider that the greater infrastructure that maintains and connects the hardware, such as power consumption of networking systems, datacenter maintenance and cooling systems, also consumes significant quantities of energy [14, 29]. If these factors are taken into account in the dynamic energy used for BLOOM training, the total carbon footprint of the model training increases by 14.6 tons CO_{2eq}. Note, the estimated BLOOM carbon emissions are the result of one single use case of many, since the estimates depend on hardware used for deployment, the batch size of inferences, and the region where the model is running [29].

Selecting the appropriate geographic location of a data center can have the biggest operational emissions reduction benefit [14]. This is due to renewable energy availability or colder climates drastically reducing energy consumption for cooling the data centers. A previous study by Lacoste et al. [26] confirms these findings. A single choice such as data center location can make the direct emissions of an algorithm vary by a factor of 40, from 20 g CO_2eq/kWh in a location that uses renewable energy sources to 820 g CO_2eq/kWh in a location that solely relies on fossil fuels. For a model such as BERT, which is trained on multiple GPUs for several weeks, this can correspond to avoiding emitting several hundreds of kilograms of CO_2eq by training on a server powered by hydroelectricity instead of fossil fuels [26]. While looking at these numbers, the concern remains that increased energy efficiency and installation of green data centers lose the race against exponential increase in usage.

Still, to reduce the environmental impacts caused by the use and development of large-scale AI systems, the impact of data centers is only a starting point. A footprint calculated in this way for AI falls short when it comes to measuring the actual impact of AI on climate change. The CO_2 footprint of training and using the hardware is only part of the calculation as it only takes the direct or computer-specific effects into account. But at least as important are the indirect effects of AI outside its usage. Not embodied emissions outside training and using AI, such as emissions created during hardware and infrastructure production, transportation, and end-of-life need to be taken into consideration as well [13, 34, 43]. There is no other way to put it than that AI is (to this date) inherently unsustainable given the need for humans to curb energy usage and carbon emissions.

How can it then be employed with good consciousness to solve sustainability issues? Is an AI for Sustainability at all possible to overcompensate its own negative influence? Similar concerns have been raised by Cowls et al. [12] who note that the carbon footprint of AI research can be significant and emphasize that more evidence is needed to weigh up the greenhouse gas emissions generated by AI research against the energy and resource efficiency gains that AI can offer [12].

Therefore, we suggest that AI for Sustainability applications should only be considered as such when they satisfy the condition that they also take into account the environmental damages and mitigate said damages through reduction of energy consumption, etc [3].

As mentioned above, it is difficult to assess the upstream and downstream emissions of AI development, apart from the energy-related emissions generated by the training and use of AI models. In addition, to this date, it is impossible to calculate whether the net environmental impact of AI for Sustainability is positive or how the positive impact on the respective sustainability goal can outweigh the very different negative impact of the models' development on sustainability. However, the aim should be to establish a transparency status quo at least when it comes to easy-to-implement sustainability queries. The question, therefore, arises as to which sustainability reporting standards AI development should meet to fulfill the necessary condition.

As a first step, we propose to calculate the carbon emissions that result from the training process. This can be done by monitoring the training of the models with tools like Code Carbon, which captures the energy consumption of the computation with its tracker function [14]. The calculated carbon footprint should be published in an emissions report as a transparency tool. In addition, it might incentivise developers to reduce emissions by optimizing their code accordingly or hosting their cloud infrastructure in geographical regions that use renewable energy sources. This is, of course, at the level of the software developer. Data center hardware manufacturers would have completely different reporting standards; e.g., graphic processing units (GPUs) manufacturers should consider certification schemes for metals and minerals (and other factors of production) and include them in a sustainability report to contribute to transparency. Ideally, in the long term, software developers and datacentre operators will then be able to choose more sustainably produced computer hardware over others.

In addition, if data centers were required to track their ecological footprint (e.g., water consumption for cooling, energy consumption for individual computations and the operation of the data center, etc.) and make this report publicly available, responsible software developers could make a more informed decision about where to train their models. In this way, sustainable practices would be implemented more and more in the industry at different levels.

More sustainable industry practices must be achieved with regulations such as corporate social responsibility or the shift in the market where consumer awareness is gradually changing. In any case, pressure to act must come from several sides at once. From regulators, customers, and investors until sustainability becomes a virtue or mainstream.

6 What does this mean for the AI community?

Since the volume and complexity of novel data available for decision-making exceed existing analytics capabilities, organizations are being pushed to automate and extend data-driven analytics using algorithmic technologies [20]. We are by no means rejecting this but simply seeking a clearer, more common, understanding of the terms related to algorithmic technologies. While some AIs are able to support the fight against the climate crisis, there are too many free riders who use the term AI for Sustainability only for marketing reasons to create a good image of their product or service, while probably achieving the opposite of what they claim as the development of the system contributes negatively to climate change.

[3] The authors do not restrict themselves to AIs for Sustainability applications; rather, all AI applications should be held accountable for their environmental externalities like other emission intensive industries

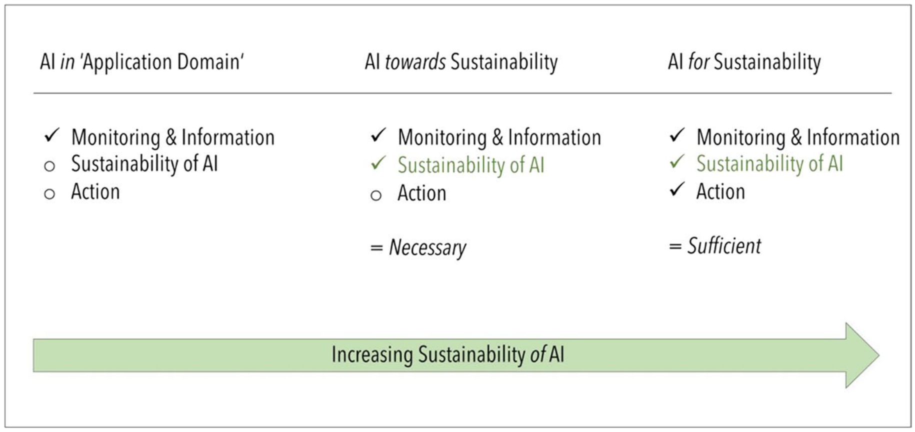

Fig. 2 Gradient describing the spectrum of research conducted in the field of Sustainable AI (own illustration)

It seems necessary to raise awareness about the different degrees to which AI may or may not contribute to its stated and marketed goal. There is a need for a more differentiated perception and critical reflection of AI systems, which are often inconsiderately or for the wrong reasons referred to in the discourse as AI for Sustainability.

Having described the sustainability issues especially related to the development and use of AI systems, we add another necessary criterion, i.e., the sustainability analysis of the AI system. As previously stated, we want to challenge the notion of AI for Sustainability by questioning whether such a designation is even justified if the application does not simultaneously address issues of AI sustainability. Since investigating the Sustainability of AI applications is rare and not yet part of the status quo, we suggest this should become common practice.

Therefore, based on all the information discussed in this paper, we propose a definition that adopts the necessary and sufficient condition for assigning the labels AI in 'application domain', AI towards Sustainability, and AI for Sustainability with each category increasing the sustainability inquiry of the AI application (Fig. 2). The highest achievable level is the AI for Sustainability, which fulfills three conditions: monitoring and information provision, a sustainability analysis of the application, and an action component that contributes to a sustainability goal.

We propose to use this gradient in the future to categorize AI applications and Sustainable AI research. Although desirable, the Sustainability of AI is not currently analyzed frequently and only to a certain extent. On this basis, we generally refuse to call a system AI for Sustainability for the time being. If a system is then labeled AI for Sustainability in accordance with this gradient, the label indeed has a meaning.

7 Conclusion

AI for Sustainability is not reduced to a certain industry, sector, or research area and each AI system is designed to solve a different problem. Moreover, there is currently a discrepancy in which AI for Sustainability applications are referring to—sustainable ends, sustainable methods, sustainable…? To prevent the oversaturation of the term AI for Sustainability, we have presented three criteria that any application labeled as such must satisfy, namely monitoring and information provision, a sustainability analysis of the application, and an action component that contributes to a sustainability goal.

We conclude this by arguing that the existence of information alone does not necessarily impact the outcome of a situation. This was demonstrated by concrete examples, such as the terrible flood disaster in Germany in 2021. Therefore, the sufficient argument is only met when the AI system is connected to an action directly. Of equal importance, we question whether an inherently unsustainable system can contribute to sustainability goals at all. Any positive influence of the system is significantly set back when both the creation of the system and its use are extremely unsustainable. The question remains whether this can be balanced out in the long term, or whether it will only contribute negatively to the climate crisis overall. Hence, when approaching AI for Sustainability, it seems crucial to us that the sustainability of the AI system itself is assessed. At this point, the importance of a forward-looking responsibility in AI development can be emphasized, where initial harm is prevented from occurring in the first place. Therefore, we propose the idea of a spectrum or gradient, in which Sustainable AI can be arranged on three levels: AI in the *application domain*, AI *towards* Sustainability, and AI *for* Sustainability.

So, we ask policymakers, researchers, and industry: where does your AI application fit on the gradient? Further, where should it fit?

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [SF] and [AW]. The first draft of the manuscript was written by [SF] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Funding for this research was provided by the Alexander von Humboldt Foundation in the framework of the Alexander von Humboldt Professorship for Artificial Intelligence and endowed by the Federal Ministry of Research to Prof. Dr. Aimee van Wynsberghe.

Data availability Not applicable.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Aldahmashi, J. and Ma, X.: Advanced machine learning approach of power flow optimization in community microgrid. 2022 27th International Conference on Automation and Computing: Smart Systems and Manufacturing, ICAC (2022). https://doi.org/10.1109/ICAC55051.2022.9911103
2. Alhebshi, F., Alnabilsi, H., Bensenouci, A. and Brahimi, T.: Using artificial intelligence techniques for solar irradiation forecasting: The case of Saudi Arabia. Proceedings of the International Conference on Industrial Engineering and Operations Management, 926–927 (2019).
3. Ali, U., Shamsi, M.H., Nabeel, M., Hoare, C., Alshehri, F., Mangina, E., et al.: Comparative analysis of prediction algorithms for building energy usage prediction at an urban scale. J. Phys: Conf. Ser. (2019). https://doi.org/10.1088/1742-6596/1343/1/012001
4. Atmaja, T., Fukushi, K.: Empowering geo-based AI algorithm to aid coastal flood risk analysis: a review and framework development. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. **3**, 517–523 (2022)
5. Aurangzeb, K.: Short term power load forecasting using machine learning models for energy management in a smart community. 2019 International Conference on Computer and Information Sciences, ICCIS (2019). https://doi.org/10.1109/ICCISci.2019.8716475
6. Bolte, L., Vandemeulebroucke, T., van Wynsberghe, A.: From an ethics of carefulness to an ethics of desirability: going beyond current ethics approaches to sustainable AI. Sustainability **14**, 4472 (2022). https://doi.org/10.3390/su14084472
7. Brennan, A.: Necessary and sufficient conditions. The Stanford Encyclopedia of Philosophy (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds). https://plato.stanford.edu/archives/fall2022/entries/necessary-sufficient/ (2022). Accessed 6 Mar 2023
8. Bousnina, D. and Guerassimoff, G.: Deep reinforcement learning for optimal energy management of multi-energy smart grids. 13164 LNCS, 15–30 (2022). https://doi.org/10.1007/978-3-030-95470-3_2
9. Cinar, D., Kayakutlu, G. and Nahit Serarslan, M.: A hybrid forecasting model for energy imports. 37th International Conference on Computers and Industrial Engineering, 1947–1954 (2007).
10. Coeckelbergh, M.: AI for climate: freedom, justice, and other ethical and political challenges. AI Ethics **1**, 67–72 (2021). https://doi.org/10.1007/s43681-020-00007-2
11. Corne, D.W., Reynolds, A.P., Galloway, S., Owens, E.H. and Peacock, A. D.: Short term wind speed forecasting with evolved neural networks. GECCO 2013 - Proceedings of the 2013 Genetic and Evolutionary Computation Conference Companion, 1521–1527 (2013). https://doi.org/10.1145/2464576.2482731
12. Cowls, J., Tsamados, A., Taddeo, M., et al.: The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. AI & Soc. **38**, 283–307 (2023). https://doi.org/10.1007/s00146-021-01294-x
13. Crawford, K. and Joler, V.: Anatomy of an AI system. The Amazon Echo as an anatomical map of human labor, data and planetary resources. https://anatomyof.ai/ (2018). Accessed 8 Mar 2023
14. Dodge, J., Prewitt, T., Des Combes, R.T., Odmark, E., Schwartz, R., Strubell, E., Luccioni, A.S., Smith, N.A., DeCario, N., Buchanan, W.: Measuring the carbon intensity of AI in cloud instances (2022). arXiv:2206.05229. Accessed 8 Mar 2023
15. Dong, Y., Hauschild, M.Z.: Indicators for environmental sustainability. Procedia. CIRP **61**, 697–702 (2017). https://doi.org/10.1016/j.procir.2016.11.173
16. Fernandes, A.C.M., Gonzalez, R.Q., Lenihan-Clarke, M.A., Trotter, E.F.L., Arsanjani, J.J.: Machine learning for conservation planning in a changing climate. Sustainability **12**, 18 (2020). https://doi.org/10.3390/su12187657
17. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E.: AI4People-an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. (Dordr) **28**(4), 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5
18. Grindsted, T.S.: Geoscience and sustainability: in between keywords and buzzwords. Geoforum **91**, 57–60 (2018). https://doi.org/10.1016/j.geoforum.2018.02.029
19. Green, B.: "Good" isn't good enough. https://www.benzevgreen.com/wp-content/uploads/2019/11/19-ai4sg.pdf (2019). Accessed 8 Mar 2023
20. Grønsund, T., Aanestad, M.: Augmenting the algorithm: emerging human-in-the-loop work configurations. J. Strateg. Inf. Syst. **29**, 2 (2020). https://doi.org/10.1016/j.jsis.2020.101614
21. Hilty, L.M., Aebischer, B.: ICT for Sustainability: an emerging research field. In: Hilty, L.M., Aebischer, B. (eds.) ICT

Innovations for Sustainability. Springer, Cham, Switzerland (2015)
22. John, A.: Flut im Ahrtal: "Es ging um Minuten". Online: https://www.tagesschau.de/inland/gesellschaft/katastrophenschutz-flut-ahrtal-101.html (2022). Accessed 8 Mar 2023
23. Kaack, L.H., Donti, P.L., Strubell, E., et al.: Aligning artificial intelligence with climate change mitigation. Nat. Clim. Chang. **12**, 518–527 (2022). https://doi.org/10.1038/s41558-022-01377-7
24. Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Amodei, D. et al.: Scaling laws for neural language models (2020). https://arxiv.org/pdf/2001.08361.pdf. Accessed 5 Apr 2023
25. Kluge Corrêa, N., Galvão, C., Santos, J.W., Del Pino, C., Pontes Pinto, E., Barbosa, C., Massman, D., Mambrini, R., Galvão, L., Terem, E.: Worldwide AI Ethics: a review of 200 guidelines and reccomendations for AI governance (2023). https://arxiv.org/pdf/2206.11922.pdf. Accessed 5 Apr 2023
26. Lacoste, A., Luccioni, A., Schmidt, V., Dandres, T.: Quantifying the carbon emissions of machine learning (2019). arXiv:1910.09700. Accessed 13 Mar 2023
27. Li, C., Liu, C., Yu, X., Deng, K., Huang, T. and Liu, L.: Integrating demand response and renewable energy in wholesale market. IJCAI International Joint Conference on Artificial Intelligence 2018, 382–388 (2018). https://doi.org/10.24963/ijcai.2018/53
28. Ligozat, A., Lefèvre, J., Bugeau, A., and Combaz, J.: Unraveling the hidden environmental impacts of AI solutions for environment (2022). arXiv:2110.11822. Accessed 13 Mar 2023
29. Luccioni, A.S., Viguier, S., Ligozat, A.: Estimating the carbon footprint of Bloom, a 176 parameter language model (2022). https://arxiv.org/pdf/2211.02001.pdf. Accessed 11 Apr 2023
30. Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Parli, V., Shoham, Y., Wald, R., Clark, J., and Perrault, R.: The AI index 2023 annual report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, (2023)
31. Mensah, J.: Sustainable development: meaning, history, principles, pillars, and implications for human action: literature review. Cogent. Soc. Sci. (2019). https://doi.org/10.1080/23311886.2019.1653531
32. Moore, J.: AI for not bad. Front. Big Data **2**, 32 (2019). https://doi.org/10.3389/fdata.2019.00032
33. Nama, M., Nath, A., Bechra, N., Bhatia, J., Tanwar, S., Chaturvedi, M., et al.: Machine learning-based traffic scheduling techniques for intelligent transportation system: opportunities and challenges. Int. J. Commun. Syst. **34**, 9 (2021). https://doi.org/10.1002/dac.4814
34. OECD: measuring the environmental impacts of artificial intelligence compute and applications. The AI footprint. OECD digital economy papers, November 2022, No. 341 (2022).
35. Orseau, L., McGregor McGill, S., Legg, S.: Agents and devices: a relative definition of agency (2018). https://arxiv.org/pdf/1805.12387.pdf. Accessed 6 Feb 2023
36. Racah, E., Beckham, C., Maharaj, T., Kahou, S.E., Prabhat and Pal, C.: Extreme weather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. Advances in Neural Information Processing Systems 2017, Vol. 2017, pp.3403–3414 (2017).
37. Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., III., Lambin, E., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H., Nykvist, B., De Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., et al.: Planetary boundaries: exploring the safe operating space for humanity. Ecol. Soc. **14**(2), 32 (2009). https://doi.org/10.1038/461472a
38. Sætra, H.S.: AI in context and the sustainable development goals: factoring in the unsustainability of the sociotechnical system. Sustainability **13**(4), 1738 (2021). https://doi.org/10.3390/su13041738
39. Schlosser, M.: Agency. The Stanford Encyclopedia of Philosophy (Winter 2019 Edition), Edward N. Zalta (ed.). https://plato.stanford.edu/archives/win2019/entries/agency/ (2019). Accessed 8 Mar 2023
40. Shankar, P., Werner, N., Selinger, S. and Janssen, O.: Artificial intelligence driven crop protection optimization for sustainable agriculture. 2020 IEEE/ITU International Conference on Artificial Intelligence for Good, 1–6. (2020). https://doi.org/10.1109/AI4G50087.2020.9311082
41. Strubell, E., Ganesh, A., McCallum, A.: Energy and policy considerations for deep learning in NLP. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650 (2019). https://doi.org/10.18653/v1/P19-1355
42. Thieken, A.H., Bubeck, P., Heidenreich, A., von Keyserlingk, J., Dillenardt, L., Otto, A.: Performance of the flood warning system in Germany in July 2021–insights from affected residents. Nat. Hazards Earth Syst. Sci. **23**, 973–990 (2023). https://doi.org/10.5194/nhess-23-973-2023
43. van Wynsberghe, A.: Sustainable AI: AI for sustainability and the sustainability of AI. AI Ethics **1**, 213–218 (2021). https://doi.org/10.1007/s43681-021-00043-6
44. Vinuesa, R., Azizpour, H., Leite, I., et al.: The role of artificial intelligence in achieving the sustainable development goals. Nat. Commun. **11**, 233 (2020). https://doi.org/10.1038/s41467-019-14108-y
45. Vogt, M., Weber, C.: Current challenges to the concept of sustainability. Global Sustain. **2**, E4 (2019). https://doi.org/10.1017/sus.2019.1
46. Yaqing, T.: Prediction of building power consumption based on GAWNN. IOP Conf. Ser.: Mater. Sci. Eng. (2019). https://doi.org/10.1088/1757-899X/563/4/042081
47. Widdicks, K., Lucivero, F., Samuel, G., Somavilla Croxatto, L., Smith, M.T., Ten Holter, C., Penzenstadler, B., et al.: Systems thinking and efficiency under emissions constraints: addressing rebound effects in digital innovation and policy. Patterns (2023). https://doi.org/10.1016/j.patter.2023.100679