

Article

Sustainable AI and Intergenerational Justice

Aurélie Halsband

German Reference Centre for Ethics in the Life Sciences (DRZE), University of Bonn, 53113 Bonn, Germany; ahalsban@uni-bonn.de

Abstract: Recently, attention has been drawn to the sustainability of artificial intelligence (AI) in terms of environmental costs. However, sustainability is not tantamount to the reduction of environmental costs. By shifting the focus to intergenerational justice as one of the constitutive normative pillars of sustainability, the paper identifies a reductionist view on the sustainability of AI and constructively contributes a conceptual extension. It further develops a framework that establishes normative issues of intergenerational justice raised by the uses of AI. The framework reveals how using AI for decision support to policies with long-term impacts can negatively affect future persons. In particular, the analysis demonstrates that uses of AI for decision support to policies of environmental protection or climate mitigation include assumptions about social discounting and future persons' preferences. These assumptions are highly controversial and have a significant influence on the weight assigned to the potentially detrimental impacts of a policy on future persons. Furthermore, these underlying assumptions are seldom transparent within AI. Subsequently, the analysis provides a list of assessment questions that constitutes a guideline for the revision of AI techniques in this regard. In so doing, insights about how AI can be made more sustainable become apparent.

Keywords: artificial intelligence; sustainable AI; intergenerational justice; future generations; policy-making; explainability; transparency



Citation: Halsband, A. Sustainable AI and Intergenerational Justice. *Sustainability* **2022**, *14*, 3922. <https://doi.org/10.3390/su14073922>

Academic Editor: Fabrizio D'Ascenzo

Received: 22 February 2022

Accepted: 23 March 2022

Published: 26 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Within its broad field of application, artificial intelligence (AI) is increasingly framed as a promising tool to enhance sustainable development. The European Commission sees AI as one of the digital technologies that are a “critical enabler for attaining the sustainability goals of the Green deal” i.a. by accelerating and maximizing “the impact of policies to deal with climate change and protect the environment” [1] (p. 9).

Recently, attention has been drawn to the environmental impact of AI itself under the umbrella term of “sustainable AI” [2] (see also [3]), stressing the need to critically assess especially the immense energy consumption of AI. However, sustainability is not tantamount to the reduction of environmental costs. By shifting the focus to intergenerational justice as one of the constitutive normative pillars of sustainability, the paper demonstrates and addresses the threat of a reductionist view on sustainable AI. It identifies the question of whether and, if so, to what extent AI can be sustainable as a major research question necessitating a theoretical underpinning. The ethical analysis contributes to the assessment of AI's long-term impacts on sustainability by revealing major implications of intergenerational justice as the underlying normative component (see [4] (p. 2,4)).

Although “sustainability” is a frequently mentioned standard that institutions and persons commit themselves to, the definition and use of this concept are often inconsistent [5]. While the concept's applicability itself is contested [6], as are different interpretations of its content, there is at least a consensus on its core idea: sustainability is the presupposition of intergenerational equity, implying the obligation to conserve “what matters for future generations” [7] (p. 54) (see also [8], p. 60). It is this shared perspective on obligations towards future persons that I will use as the starting point for my analysis.

That is to say, instead of defending a specific interpretation of sustainability, the goal of my analysis is to focus on intergenerational justice as one of its constitutive normative pillars. In so doing, the encompassing demands that are implied with the objective of creating sustainable AI become apparent: if sustainability is fundamentally about conserving “what matters for future generations” [8] (p. 54), this conservative effort will exceed a mere reduction of environmental costs such as those resulting from high energy consumption. This comprehensive approach to sustainable AI is also reflected in the European Commission’s description of the conditions that AI must satisfy in regard to sustainability: “AI technology must be in line with the human responsibility to ensure the basic preconditions for life on our planet, continued prospering for mankind and preservation of a good environment for future generations” [9] (p. 19).

By addressing the question of whether and, if so, to what extent the development and use of AI can be sustainable from the specific normative angle of intergenerational justice, the analysis contributes to closing two research gaps. Firstly, it depicts the reductionist understanding of sustainability in the context of sustainable AI, which has been focused on the welcome call for emission reductions and carbon footprint assessments of AI [10], yet without reference to the further demands of sustainability. This merely implicit reference to intergenerational justice in spite of its fundamental normative function has also been an issue of criticism [11,12] of the United Nations’ understanding of sustainability that underlies the formulation of its 17 Sustainable Development Goals (SDGs) [13]. Secondly, the integration of the concept of intergenerational justice provides an addendum to previous analyses of justice issues raised by AI. Although the principle of justice has frequently been applied to evaluate the different uses of AI, these have been focused on issues of discrimination resulting from biased algorithms or on broader issues of distributive justice, e.g., arising from exclusive access to AI technologies because of diverging financial means (cf. e.g., [14], p. 699). Within the emerging application of AI to climate mitigation, additional issues of justice have been discussed such as using AI to nudge people into climate-friendly behaviour or the question of who within the global community should bear the costs of using AI to enhance climate mitigation [3]. Yet, intergenerational justice opens the view on “novel forms of ethical challenges” raised by the use of AI in the context of climate change mitigation and the broader field of environmental policies [15] (p. 13). While issues of *intragenerational* justice raised by AI have been addressed before, the intergenerational justice dimension has received little attention up to now [3] (p. 70) and, to my knowledge, there has been no analysis in the context of AI.

To address this gap, the analysis turns to a specific field of application of AI that can significantly impact future persons. Challenges of intergenerational justice are especially raised by the use of AI in those fields of application in which AI provides decision support to issues with long-term impacts, such as environmental protection policies or climate mitigation policies. Other areas where especially policies can have significant impacts on future generations are, e.g., funding strategies of pension schemes or public debt management [16] (p. 62). This paper focuses on the former field of application. Thus, AI, with its specific feature of self-learning (machine learning, ML), is being employed as a tool for climate policy analysis “[. . .] evaluating the outcomes of past policies and assessing future policy alternatives [. . .]. ML can provide data for policy analysis, help improve existing tools for assessing policy options, and provide new tools for evaluating the effects of policies” [17] (p. 52f). In addition, AI has been applied to other environmental issues such as monitoring the extent of deforestation or simulating the effects of climate change [15,17].

As a first step, this analysis provides a normative framework that helps to explore those applications of AI in the context of climate mitigation and environmental protection that raise issues of intergenerational justice, especially those that may have detrimental impacts on future generations. This shall help to contribute to a conceptually informed understanding of sustainability. In a second step, the analysis provides a list of assessment questions that constitutes the first guideline for the revision of AI techniques in this regard.

Overall, the framework offers insights into how sustainable some uses of AI are with the specific normative focus on issues of intergenerational justice.

Although I will mostly refer to ML applications, I use the broader term of AI throughout the paper. The framework and assessment questions will also provide guidance for identifying those types of AI that raise the depicted issues of intergenerational justice.

2. The Normative Framework

2.1. Two Ethical Dimensions of Sustainability

From an ethical perspective, sustainability can be described as a concept with two central dimensions. My analysis starts from the consensus among the debate on defining sustainability that it is majorly based on the obligation to conserve “what matters” for future persons. Obligations to future generations are, in turn, embedded in theories of intergenerational justice which constitute one of the “key components” [16] (p. 62) and the first ethical dimension of sustainability [18] (p. 897).

Note that most concepts of sustainability limit their search for “what matters for future generations” to (parts of) the environment. This specific focus on natural resources as prerequisites for providing future generations with “what matters” constitutes the second ethical dimension of sustainability. An important debate in this context is the dispute between adherents to the concepts of weak sustainability and strong sustainability, where the latter reject, and the former assume that multiple aspects of the “natural capital” currently required to satisfy basic humans needs can prospectively be substituted by technological or other artificial means [18] (p. 904ff). Both ethical dimensions of sustainability heavily rest on normative considerations debated extensively in theories of intergenerational justice. Why we are obligated towards future persons in the first place and if so, to what extent are major subjects of discussion within these theories. For debates within sustainability about the extent or scope of the obligations towards future persons, i.e., if prerequisites for basic needs satisfaction for future persons or a more encompassing perspective on prerequisites for the realisation of different conceptions of the good life should guide the selection of natural resources that ought to be preserved, refer to the general debate about the most convincing distributive principle of intergenerational justice. An example for the former approach is the well-known definition of sustainable development in the World Commission on Environment and Development’s 1987 Brundtland report, in which sustainable development is defined as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs” [19].

For the purpose of my analysis, it is the first ethical dimension of sustainability, i.e., the normative concept of intergenerational justice and its focus on impacts on future persons that will serve as a starting point. This will not constitute an encompassing definition of sustainability. Rather, it is conceptualised as a normative module among both further ethically informed modules (i.a. integrating the second ethical dimension of sustainability) and other evaluative modules informed by the additional constitutive perspectives on sustainability such as those of natural, economic, and social sciences.

2.2. Intergenerational Relations as a Framework

How to evaluate present persons’ actions when they may have negative impacts on future persons is usually understood as an issue of intergenerational justice within ethical theory, presupposing that the concept of justice can be applied to those not yet alive. Within theories of intergenerational justice, the term “future generations” refers as a shorthand to those persons who will come into existence after the presently alive persons’ lifetime, i.e., a group of persons that will have no possibility to directly interact with those presently alive (see [7], p. 43). The need for an ethical assessment of current persons’ actions and their impacts on future persons within a specific theory (of justice) can be justified by the features of the relations between members of different generations. One specific feature of the relation between present persons and future persons is *contingency*, referring to the fact that “future people’s existence, number, and specific identity depend (are contingent)

upon currently living people's decisions and actions" [20]. Most importantly, which and if future persons come into existence depends on the present person's decisions if, and when, to reproduce. Another genuine feature of the relation between members of different generations is the lack of *reciprocity*, stressing the impossibility of direct interaction between persons currently alive and those who are not yet alive. This relation is closely connected to the intergenerational *power-asymmetry*, describing the fact that only present persons can exercise actions affecting—either positively or negatively—future persons during their lifetime. Finally, intergenerational relations are characterised by *uncertainty*, especially about the identities and preferences of future persons.

As with every innovation, developing and using AI will affect who, how many, and which persons will come into existence (*contingency*). Therefore, I will not treat 'intergenerational contingency' as a genuine normative challenge in the context of AI. Moral implications of the intergenerational relation of contingency have been prominently discussed within the still ongoing debate about the "non-identity problem" [21] (pp. 351–441). In contrast, the focus on the power-asymmetry, as well as the intergenerational relations of non-reciprocity and uncertainty, will help to explore specific uses of AI that raise issues of intergenerational justice. These relations are being used as a framework to break down the encompassing concern of intergenerational justice as to how different entitlements of different persons living at different times—i.e., different generations—should be specified and weighed when they are in conflict.

More generally speaking, the framework supports a continuous ethical assessment of AI as a set of emerging technologies with the specific focus on potentially detrimental impacts that directly result from the use of these technologies in the present but primarily will affect future persons. It rests on past experiences with detrimental side-effects of emerging technologies such as nuclear energy generation and the issue of radioactive waste or high carbon-emitting industries and climatic changes, which both predominantly will affect future generations.

3. Power-Asymmetry and Intertemporal Discounting

With AI's strong potential in the evaluation of large sets of data, it is successively being used to improve policy addresses to the complex phenomenon of climate change and its interdependent causes. Integrated assessment models (IAMs) play an important role in predicting and evaluating the interaction of socioeconomic and climate-related factors [17] (p. 53). The goal of IAMs is "to project alternative future climates with and without various types of climate change policies in place in order to give policymakers at all levels of government and industry an idea of the stakes involved in deciding whether or not to implement various policies" [22] (p. 116). Due to the complexity of the involved models, as well as the amount of data, AI and especially ML are being applied to various sub-models which, together, form the IAMs [17] (p. 53). AI has thus been used to support policy-making in domains with a multitude of factors and stakeholders interacting, such as policies on sustainable development [23] (pp. 22,27) or agricultural public policy [24].

However, this support of policy-making with the help of AI is also confronted with some of the criticism brought forward against features of these policy models in general. One branch of models that are part of IAMs and have important implications regarding intergenerational justice is cost-benefit analyses of climate policies. These models assess the costs and benefits of climate mitigation across a long period of time, surpassing the lifetime of presently alive persons. They assess how costs and benefits are being distributed between different people (i.e., different generations) across different times. How to weigh costs and benefits between persons living at different times within cost-benefit analysis is usually addressed by the inclusion of a social discount rate. Setting the discount rate high involves assigning a significantly smaller value to benefits that accrue in the distant future. This has important normative implications which can be illustrated regarding carbon emission reduction policies:

“[. . .] intertemporal equity is extremely important in determining the appropriate rate of implementation of policies designed to reduce carbon emissions [. . .]. Low discount rates generally make rapid implementation of such policies much more urgent than high discount rates because damages are projected to grow steadily over time at a much more rapid rate than mitigation costs” [22] (p. 126f).

Against this background, the practice of discounting within cost–benefit analyses with large time horizons—such as those on climate mitigation policies—are faced with considerable objections. On the practical level, it may lead to an underestimation of potentially severe costs for future persons and underplay the urgency of action required in the present to reduce these costs. This is because mitigation policies in the context of climate change imply costs (of climate mitigation) that predominantly accrue to present persons and their losses in consumption. The benefits, however, are reduced risks of climate change which most importantly benefit future persons [25] (p. 401). Present persons thus face potentially higher burdens and are consequently tempted to include an elevated discount rate to reduce these burdens. On a more general level, if and at which rate to discount refers to a disputed field of normative assumptions. Different justifications for discounting the future have been discussed, for example, that it may be justified to give less weight to benefits for future persons as they will overall be better off under the assumption of an overall steadily increasing wealth [26] (p. 48f). Whether there are legitimate reasons to discount benefits for future persons has been subject to an extensive discussion within philosophy and between philosophers and economists (see e.g., [21,27]). With respect to applying AI to this domain of policy evaluation, it suffices to state in a first step that the integration of a social discount rate in those contexts with large time horizons needs to be accessible for a normative evaluation. Among other considerations, strongly discounting benefits for future persons can bear the risk of assigning excessively high costs to them. This may then equal a negative manifestation of the intergenerational power-asymmetry.

The issue of discounting is, however, not a normative issue genuinely raised by the application of AI. Instead, applying AI to this domain can only be justified if the already discussed limitations of these models are adequately considered. Yet a specific challenge genuine to some of the AI techniques is the issue of providing an explanation for generated decisions. As has been shown, the setting of a social discount rate can have important normative implications regarding future persons. To address these limitations, cost–benefit analyses conducted by AI need to be explainable and transparent regarding the setting of the discount rate, thus leaving the possibility for later revisions of the settings. I will come back to the aspect of explainable AI below. With regard to the limitations of the integrated models, constructive insights for potential revision can be gained from general critical assessments of these models [22] (pp. 124,128f) and from objections to the practice of discounting, e.g., in climate mitigation [25] (pp. 401,405).

Regarding the use of AI to support assessments with large time frames such as climate mitigation policies, another aspect under dispute, which has important implications regarding intergenerational justice, is the underlying calculation of costs. A focus on static costs has been shown to neglect the long-term aspect of climate change by neglecting the dynamics between potentially slightly higher costs in the present that may, however, reduce mitigation costs in the distant and near future [28] (p. 54), thus generating an overall improved cost–benefit ratio. Hence, the calculation of costs represents another aspect that must be accessible for potential revision within assessments that are being conducted or supported by AI.

Finally, policies with long-term impacts will only be able to represent potentially detrimental consequences for future persons if the time frames are set in a way that includes those persons. This illustrates a third aspect that needs to be accessible for potential revision not only within cost–benefit analyses conducted or supported by AI but for all types of policy assessments that may include AI. For example, policy-making regarding energy management relies among others on electricity demand forecasting which is increasingly being supported by AI. Within these forecasts, time horizons for long-term projections range

from a couple of years to projections about the next 50 years [29] (p. 15ff). Consequently, insights about the time frames and thus implicitly about the representation of potential impacts affecting persons in the distant future need to be made accessible within AI-based policy support assessments.

Using AI on contexts and decisions affecting different persons and different times—especially future generations—thus adds to the general challenge of creating AI that is transparent and explainable. Explainability is addressing “the need to understand and hold to account the decision-making processes of AI” [14] (p. 700). The principle of explainability has been established as a genuine principle for the normative evaluation of AI along with the established bioethical principles of beneficence, non-maleficence, autonomy, and justice. Impacts on future persons constitute a yet-underestimated societal area that ought to be assessed using this principle. This will also contribute to the critical assessment of using AI within policy-making that has importantly been focused on issues of acceptance and trust [23] (p. 33f).

4. Uncertain Preferences and “Intergenerational Transfer Bias”

Intergenerational relations are characterised by uncertainty in important domains, such as uncertainty about the preferences of future persons. Consequently, there is no data or only fragmentary data that AI can use in this regard. Using AI for assessments with large time frames will accordingly involve assumptions about preferences that future persons will have and how these can be ‘translated’ into opportunities that present persons should leave open for them. For example, the implications for the use of IAMs in the context of climate mitigation can be described like this:

“People making decisions today on behalf of those not yet alive need to make collective ethical choices about what kind of opportunities (usually characterized as a particular state of the climate system measured by global mean temperature, GHG concentration, or maximum climate damages allowable by some future date) they want to leave future inhabitants of planet Earth [. . .]” [22] (p. 126f).

It is these choices that have normative implications. Take for example a study [30] forecasting both CO₂ emission and energy demand that will arise from the transportation sector in Turkey until 2050 based on machine learning algorithms. Such a forecast necessarily includes assumptions about preferences that persons living in the time frame from 2022–2050 will pursue that are tied to emissions, energy use, and choice of transportation means. However, the longer the time frame of the forecast, the more difficult it will be to anticipate the preferences. A longer time frame of the forecast will also complicate the task of anticipating what the pursuit of these preferences will require, e.g., regarding the use of energy, the emission of greenhouse gases, or the choice of transportation means. This is because the use of these—broadly understood—resources such as the use of energy are tied to the pursuit of preferences but do not represent preferences in themselves. People usually do not enjoy emitting CO₂ but partake in activities that can stand in a causal relation to emissions, such as living in adequately heated buildings when the outside temperature is low. Over longer periods of time, both these causal relations, as well as the preferences, can change.

A simple approach to these assumptions about future preferences within AI-supported assessments could be to presuppose that the preferences of persons in the distant future, including future persons, are broadly overlapping with those of current persons. However, this way to proceed may raise the challenge of a so-called transfer of data bias [31] (p. 4), a challenge especially important in machine learning and its reliance on historic data for training purposes [32] (p. 6f). Simply ‘transferring’ present preferences may bear the risk of providing insufficiently for opportunities that should be left open for future persons because either future persons’ preferences change significantly or the circumstances in which these preferences can be satisfied change. Most importantly, the satisfaction of preferences such as mobility may rely on very different sets of resources in differing circumstances, thus leaving future persons with different opportunities. The fact that resources

may provide different individuals in different circumstances with highly heterogeneous opportunities has been extensively discussed as the issue of “conversion factors” within the literature on the Capabilities Approach [33]. Besides the potentially differing individual conversion of resources, it is even unclear from a philosophical point of view if future persons should be provided *with the same* opportunities. This has been an issue of debate between the adherents of the four most discussed intergenerational “principles” of justice of either equality, proportionality, priority, or sufficiency [5] (p. 7448). To date, there neither emerged a consensus within this philosophical debate nor is AI technology suited to integrate all (theoretical) facets of the debate. However, this specific type of transfer bias, which I have framed as *intergenerational transfer bias*, as well as encompassing questions regarding the choice and extent of opportunities that should be left open for future persons, requires AI applied in these contexts once again to be open for revision. Similar solutions have been proposed for the difficulty of including AI’s potential impacts on non-human animals [31] (p. 6). This way, potentially adapted preferences or changed circumstances may be added to the algorithms. In other cases, considering the uncertainty about future persons’ preferences may require present persons to provide for broader “choice options” that leave the realisation of different preferences in the distant future open (see [7] (p. 53) and [34] (p. 206ff). How this can be realised within AI-based assessments will constitute a challenge for those involved in the design and implementation of these systems.



5. Non-Reciprocity and Indirect Involvement

Unlike with other issues of fairness or justice raised by using AI [3] (p. 71f), the involvement of stakeholders cannot contribute solutions to the presented issues of intergenerational justice. As future persons are yet unborn, there is no reciprocity between future and present persons. An involvement of future persons can thus only be accomplished indirectly.

The success of indirectly involving future persons by present persons’ concern for the well-being of the former can, however, be rather limited [35] (p. 19). A more promising way to take aspects of intergenerational justice into account when using AI is to develop a set of evaluative criteria. As a result of the normative challenges described before, a list of questions guiding the potential revision of AI used in context with long-term impacts emerges (cf. Table 1). The first category of questions is targeted at shaping AI in a way that makes especially those features accessible for potential revision that can have negative impacts on future persons. This way, the threat of having no data on potential detrimental impacts [36] (p. 9) ought to be avoided. Further aspects and data will have to be added. Thus, in the environmental context, a specific focus on irreversible costs such as the acceleration of biodiversity loss or the generation of hazardous waste may have to be added to the evaluation.

The second category of questions supporting the use and assessment of AI in contexts with long-term impacts is targeted at assessing whether the use of AI *itself* negatively impacts future persons. Whereas most of the questions raised above reveal the necessity to revise tools of assessments that are also being operated without AI, the use of AI may itself raise additional challenges to the realisation of intergenerational justice. Here, it is the threat of overseeing insights into potentially detrimental impacts [36] (p. 9) on future persons from available data, as well as the occurrence of unintended adverse impacts [32] (p. 8), that is being targeted. The environmental costs of running AI are an example of a negative impact that refers to AI *itself*, i.e., a genuine impact on future persons caused by using AI.

Table 1. Artificial Intelligence (AI) and Intergenerational Justice: Assessment Questions.

Time Frame	What is the time horizon of the assessment that is supported or entirely conducted by the AI? Is the scope of evaluation surpassing ≈ 20 years, thus making the anticipation of future preferences of both the yet-unborn and those already alive more difficult? If yes, issues of intergenerational justice may be affected by this specific use of AI.	
	 Cost–benefit analysis	<p>Does the analysis involve weighing benefits for different people at different times? If yes, a series of follow-up questions guides the further evaluation:</p> <p>Discount rate How has the discount rate been set? For what reasons? How are potential costs of a project being distributed between different people at different times? Does the assessment assign excessively high burdens to a particular sub-group? Is there an intergenerational transfer bias? How are costs as a backside of the benefits being defined (e.g., static or dynamic) and assessed? On what assumptions about the preferences of potentially affected persons have the benefits been defined?</p>
		<p>Burden distribution</p> <p>Cost definition</p> <p>Benefit definition</p>
AI itself as impact	 Environmental impact	<p>Does the use of AI have negative impacts on future persons that are directly linked to methods and infrastructure of AI itself?</p> <p>Is the environmental impact of AI in proportion to its potential positive impact?</p>

Overall, this list of assessment questions will have to be adapted and revised on a regular basis as it serves to ethically accompany nascent technologies [31] (p. 8). The hope is to provide a normatively informed standard for using AI “properly”, i.e. in accordance with intergenerational justice:

“If AI is underutilised or misused, it may undermine existing environmental policies, slow down efforts to foster sustainability, and impose severe environmental costs on current and future generations. However, if used properly, AI can be a powerful tool to develop effective responses to the climate emergency. Policymakers and the research community must act urgently to ensure that this impact is as positive as possible, in the interest of an equitable and sustainable future” [37] (p. 779).

The list of normative questions adds to this endeavour of realising AI that is sustainable, where intergenerational justice as one of the two ethical dimensions of sustainability provides a central normative standard to assess AI’s sustainability. Starting with the question of whether and, if so, to what extent AI can be sustainable, the presented research developed a normative framework that attempts to integrate major aspects of intergenerational justice which, in turn, can be applied to assess different uses of AI. The application of this framework to specific uses of AI with potentially significant long-term impacts, namely, decision support for climate mitigation and environmental protection policies, resulted in the list of assessment questions presented above. A major implication that has been deduced is the necessity to make AI transparent and open for revision, especially with regard to the setting of a social discount rate and the assumptions about future persons’ preferences whenever it is used in this context.

6. Discussion and Outlook: Towards the Sustainability of AI

Measuring the use of AI against the standard of intergenerational justice may overburden the involved technologies. If current decision-making procedures, especially about policies with important impacts on future persons, do not fulfil this standard, why should AI? For instance, the German Federal Constitutional Court ruled in March 2021 that the provisions of the Federal Climate Change Act and its governing national climate targets are insufficient regarding the emission regulations because it shifts an excessively large part of the mitigation burden to future persons [38]. The standard of intergenerational justice is thus already presenting severe challenges to policy-making in general. In addition, the normative approaches to intergenerational justice are highly debated and “[. . .] fall astonishingly short of expectations in attempting to deal with the normative issues raised by environmental and resource depletion problems” [16] (p. 61). This may impede the attempt to use them as guidelines for AI design.

Two replies are in order. First, even if intergenerational justice is a contested issue, this does not rule out normative guidance. It rather urges to reveal the choice and reasons for the selection of specific normative premises regarding future persons (see for a similar point regarding sustainability [7] p. 50). The presented list of guideline questions constitutes a framework that supports this endeavour. Impacts on future persons and their normative evaluation thus constitute a further application context for the criteria of transparency and explainability within the debate about AI.

Second, AI technology may even facilitate the application of intergenerational justice as a normative standard. AI's potential to reduce institutional inefficiency in the context of environmental degradation, climate mitigation, or sustainability policies has already been noted (see e.g., [3] p. 69 and [32]). Regarding the intergenerational impact of policies, AI that has been designed and developed in accordance with normative criteria such as those described above may even be employed as a corrective tool by disclosing settings that refer to contested issues of intergenerational justice.

For the time being, however, the use of AI is faced with several constraints regarding intergenerational justice: “[. . .] AI system adoption practices are heavily technologically determined and reductionist in nature, and do not envisage and develop long-term, ethical, responsible and sustainable solutions” [39] (p. 3) (see also [32]). One such reduction is the reduction of the standard of sustainability to the attempt of reducing environmental costs. Unsurprisingly, AI will thus not be able to realise sustainability in itself and instead needs to be included in an encompassing vision as “[. . .] many of our current sustainability interventions via IT are measures to reduce unsustainability instead of creating sustainability, which means that we have to significantly shift our thinking towards a transformation mindset for a joint sustainable vision of the future” [4] (p. 11).

The elaborated normative framework provides a list of assessment questions that explore normative issues regarding impacts on future persons and subsequently the potential need for revision of AI techniques within such a technological approach to a sustainable future. In so doing, insights about how AI can be made more sustainable become apparent. This way, AI may contribute to the pervasive political effort of promoting sustainable development.

To this end, topics for future research are distributed between different scientific disciplines. As an addendum to the ethically informed analysis, the future AI-based support for policies on climate mitigation and environmental protection, and its conformity with the concept of sustainability ought to be assessed from the perspective of policy research. The above-developed framework and assessment guide is conceptualised as a normative module that can be complemented by further normative modules. These would have to represent, for example, issues of *intragenerational* justice and the use of natural resources as the second ethical dimension of sustainability. Furthermore, they would have to be interlinked with more empirically oriented sustainability assessments of AI to form an encompassing standard assessing the sustainability of AI. Attempts for more encompassing evaluations of AI and its impacts on sustainability have been conducted

against the UN's sustainable development goals (SDGs) [40–42], however, not representing issues of intergenerational justice. Also, future research topics include the question of how the policy decisions support provided by AI can be designed to be open for revision in the relevant way described above.

7. Conclusions

The analysis developed a normative framework to assess whether and, if so, to what extent the development and use of AI can be sustainable from the specific normative angle of intergenerational justice. Starting from the observation that recent calls for more sustainable AI are based on a narrow understanding of sustainability, it instructed a return to intergenerational justice as a central ethical dimension of sustainability. This contributed to a conceptually informed understanding of sustainability, moving beyond an equation of sustainability with the reduction of environmental costs. The normative framework used intergenerational power asymmetries, as well as the intergenerational relations of non-reciprocity and uncertainty to explore specific uses of AI that raise issues of intergenerational justice. Due to its long-term impacts, the policy decisions support provided by AI in the context of climate mitigation and environmental protection was identified as a significant application field in need of a normative assessment. More specifically, the setting of a social discount rate and the assumptions about future persons' preferences within AI-supported policy assessments were presented as potentially having detrimental impacts on future generations. A major implication has thus been the insight that AI must be made transparent and open for revision, especially with regard to social discounting and assumed preferences over large time horizons. To instruct the implementation of these insights, the analysis provided a list of assessment questions that constitute a first guideline for the revision of AI techniques. It operationalises key aspects of intergenerational justice as one of the constitutive concepts of sustainability and thus contributes a normative module for an ethically informed assessment of the sustainability of AI.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. European Commission (EC). *The European Green Deal*. COM (2019) 640 Final; European Commission: Geneva, Switzerland, 2019; Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1576150542719&uri=COM%3A2019%3A640%3AFIN> (accessed on 21 February 2022).
2. Van Wynsberghe, A. Sustainable AI: AI for sustainability and the sustainability of AI. *AI Ethics* **2021**, *1*, 213–218. [CrossRef]
3. Coeckelbergh, M. AI for climate: Freedom, justice, and other ethical and political challenges. *AI Ethics* **2021**, *1*, 67–72. [CrossRef]
4. Khakurel, J.; Penzenstadler, B.; Porras, J.; Knutas, A.; Zhang, W. The rise of artificial intelligence under the lens of sustainability. *Technologies* **2018**, *6*, 100. [CrossRef]
5. Stumpf, K.H.; Baumgärtner, S.; Becker, C.U.; Sievers-Glotzbach, S. The Justice Dimension of Sustainability. A Systematic and General Conceptual Framework. *Sustainability* **2015**, *7*, 7438–7472. [CrossRef]
6. Beckerman, W. 'Sustainable Development': Is it a Useful Concept? *Environ. Value* **1994**, *3*, 191–209. [CrossRef]
7. Barry, B. Sustainability and intergenerational justice. *Theoria* **1997**, *44*, 43–64. [CrossRef]
8. Ott, K. The case for strong sustainability. In *Greifswald's Environmental Ethics. From the Work of the Michael Otto Professorship at Ernst Moritz Arndt University. 1997–2002*; Ott, K., Thapa, P.P., Eds.; Steinbecker: Greifswald, Germany, 2003; pp. 59–64.
9. European Commission. European Commission. European Group on Ethics in Science and New Technologies (EGE). In *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*; European Commission: Brussels, Belgium, 2018; Available online: <https://data.europa.eu/doi/10.2777/786515> (accessed on 21 February 2022).
10. Strubell, E.; Ganesh, A.; McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. *arXiv* **2019**, arXiv:1906.02243.
11. Vasconcellos Oliveira, R. Back to the Future: The Potential of Intergenerational Justice for the Achievement of the Sustainable Development Goals. *Sustainability* **2018**, *10*, 427. [CrossRef]
12. Spijkers, O. Intergenerational Equity and the Sustainable Development Goals. *Sustainability* **2018**, *10*, 3836. [CrossRef]

13. United Nations General Assembly. *Transforming our World: The 2030 Agenda for Sustainable Development, Resolution 70/1, Adopted 25 September 2015*; United Nations: New York, NY, USA, 2015.
14. Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [CrossRef]
15. Cows, J.; Tsamados, A.; Taddeo, M.; Floridi, L. The AI Gambit—Leveraging artificial intelligence to combat climate change: Opportunities, challenges, and recommendations. *AI SoC* **2021**. [CrossRef] [PubMed]
16. Gosseries, A. Theories of intergenerational justice: A synopsis. *SAPIENS* **2008**, *1*, 61–71. [CrossRef]
17. Rolnick, D.; Donti, P.L.; Kaack, L.H.; Kochanski, K.; Lacoste, A.; Sankaran, K.; Ross, A.S.; Milojevic-Dupont, N.; Jaques, N.; Waldman-Brown, A.; et al. Tackling climate change with machine learning. *arXiv* **2019**, arXiv:1906.05433.
18. Ott, K. Institutionalizing Strong Sustainability: A Rawlsian Perspective. *Sustainability* **2014**, *6*, 894–912. [CrossRef]
19. United Nations (UN). Report of the world commission on environment and development. In *Our Common Future*; Oxford University Press: Oxford, UK, 1987; Available online: <http://www.un-documents.net/wced-ocf.htm> (accessed on 21 February 2022).
20. Meyer, L. Intergenerational Justice. In *The Stanford Encyclopedia of Philosophy*; Stanford University: Stanford, CA, USA, 2021; Available online: <https://plato.stanford.edu/archives/sum2021/entries/justice-intergenerational> (accessed on 21 February 2022).
21. Parfit, D. *Reasons and Persons*, 3rd ed.; Oxford University Press: Oxford, UK, 1987.
22. Weyant, J. Some contributions of integrated assessment models of global climate change. *Rev. Environ. Econ. Policy* **2017**, *11*, 115–137. [CrossRef]
23. Milano, M.; O’Sullivan, B.; Gavanelli, M. Sustainable policy making: A strategic challenge for artificial intelligence. *AI Mag.* **2014**, *35*, 22–35. [CrossRef]
24. Sánchez, J.M.; Rodríguez, J.P.; Espitia, H.E. Review of artificial intelligence applied in decision-making processes in agricultural public policy. *Processes* **2020**, *8*, 1374. [CrossRef]
25. Davidson, M.D. Climate change and the ethics of discounting. *WIREs Clim Change* **2015**, *6*, 401–412. [CrossRef]
26. O’Neill, J. *Ecology, Policy and Politics: Human Well-Being and the Natural World*; Routledge: London, UK; New York, NY, USA, 2002.
27. Broome, J. Discounting the Future. *Philos. Public Aff.* **1994**, *23*, 128–156. [CrossRef]
28. Gillingham, K.; Stock, J.H. The cost of reducing greenhouse gas emissions. *J. Econ. Perspect.* **2018**, *32*, 53–72. [CrossRef]
29. Mir, A.A.; Alghassab, M.; Ullah, K.; Khan, Z.A.; Lu, Y.; Imran, M. A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons. *Sustainability* **2020**, *12*, 5931. [CrossRef]
30. Ağbulut, Ü. Forecasting of transportation-related energy demand and CO₂ emissions in Turkey with different machine learning algorithms. *Sustain. Prod. Consum.* **2022**, *29*, 141–157. [CrossRef]
31. Galaz, V.; Centeno, M.A.; Callahan, P.W.; Causevic, A.; Patterson, T.; Brass, I.; Baum, S.; Farber, D.; Fischer, J.; Garcia, D.; et al. Artificial intelligence, systemic risks, and sustainability. *Technol. Soc.* **2021**, *67*, 101741. [CrossRef]
32. Nishant, R.; Kennedy, M.; Corbett, J. Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *Int. J. Inf. Manag.* **2020**, *53*, 102104. [CrossRef]
33. Robeyns, I.; Byskov, M.F. The Capability Approach. In *The Stanford Encyclopedia of Philosophy*; Stanford University: Stanford, CA, USA, 2021; Available online: <https://plato.stanford.edu/archives/win2021/entries/capability-approach> (accessed on 21 February 2022).
34. Halsband, A. *Konkrete Nachhaltigkeit. Welche Natur wir für künftige Generationen erhalten sollten*; Baden-Baden: Nomos, Germany, 2016.
35. Klockmann, V.; Von Schenk, A.; Villeval, M.C. Artificial Intelligence, Ethics, and Intergenerational Responsibility. *GATE Work Pap.* **2021**. [CrossRef]
36. Walsh, T.; Evatt, A.; de Witt, C.S. Artificial Intelligence & Climate Change: Supplementary Impact Report. 2020. Available online: <https://www.semanticscholar.org/paper/Artificial-Intelligence-%26-Climate-Change-%3A-Impact-a-Walsh-Evatt/a840e7c4af0f10b3fac136ddc99e31c6c7d58507> (accessed on 21 February 2022).
37. Taddeo, M.; Tsamados, A.; Cows, J.; Floridi, L. Artificial intelligence and the climate emergency: Opportunities, challenges, and recommendations. *One Earth* **2021**, *4*, 776–779. [CrossRef]
38. German Federal Constitutional Court. Constitutional Complaints against the Federal Climate Change Act Partially Successful. Press Release No. 31/2021 of 29 April 2021. Order of 24 March 2021. 1 BvR 2656/18, 1 BvR 288/20, 1 BvR 96/20, 1 BvR 78/20. Available online: <https://www.bundesverfassungsgericht.de/SharedDocs/Pressemitteilungen/EN/2021/bvg21-031.html> (accessed on 21 February 2022).
39. Yigitcanlar, T.; Mehmood, R.; Corchado, J.M. Green artificial intelligence: Towards an efficient, sustainable and equitable technology for smart cities and futures. *Sustainability* **2021**, *13*, 8952. [CrossRef]
40. Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Felländer, A.; Langhans, S.D.; Tegmark, M.; Fuso Nerini, F. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat. Commun.* **2020**, *11*, 233. [CrossRef]
41. Truby, J. Governing Artificial Intelligence to benefit the UN Sustainable Development Goals. *Sustain. Dev.* **2020**, *28*, 946–959. [CrossRef]
42. Sætra, H.S. AI in Context and the Sustainable Development Goals: Factoring in the Unsustainability of the Sociotechnical System. *Sustainability* **2021**, *13*, 1738. [CrossRef]