Technology in Society 68 (2022) 101926

Contents lists available at ScienceDirect

Technology in Society

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

Sustainable AI: An integrated model to guide public sector decision-making

Christopher Wilson, PhD *, Maja van der Velden, PhD

University of Oslo, Norway

ARTICLE INFO

Keywords:
Artificial intelligence
Public administration
Sustainability
Social sustainability
AI governance

ABSTRACT

Ethics, explainability, responsibility, and accountability are important concepts for questioning the societal impacts of artificial intelligence and machine learning (AI), but are insufficient to guide the public sector in regulating and implementing AI. Recent frameworks for AI governance help to operationalize these by identifying the processes and layers of governance in which they must be considered, but do not provide public sector workers with guidance on how they should be pursued or understood. This analysis explores how the concept of sustainable AI can help to fill this gap. It does so by reviewing how the concept has been used by the research community and aligning research on sustainable development with research on public sector AI. Doing so identifies the utility of boundary conditions that have been asserted for social sustainability according to the Framework for Strategic Sustainable Development, and which are here integrated with prominent concepts from the discourse on AI and society. This results in a conceptual model that integrates five boundary conditions to assist public sector decision-making about how to govern AI: Diversity, Capacity for learning, Capacity for self-organization Common meaning, and Trust. These are presented together with practical approaches for their presentation, and guiding questions to aid public sector workers in making the decisions that are required by other operational frameworks for ethical AI.

1. Introduction

The long-term societal impacts of artificial intelligence and machine learning technologies (AI) are widely speculated, but poorly understood. This poses dual governance challenges for the public sector, which must not only regulate how AI interacts with society, but is itself increasingly adopting AI to automate government workflows and deliver services [1,2]. The challenges of regulating and implementing AI are distinct but related, in so far as they both must confront widely discussed, but often unspecified risks related to AI's bias, misuse, and perceptions of illegitimacy.

Efforts to address these risks are confounded by the subtle and complex ways in which novel technologies like AI interact with society. This is in part due to the novel and pervasive nature of the technology, and some applications are so new that we must wait to see what consequences unfold. Other risks are systematic and can be modeled, but a recent review of AI initiatives aiming to improve ecological sustainability notes that even these are generally "poorly elaborated, and more often than not, overlooked" [3, p. 2]. Nor is this simply a problem at scale, as the opacity surrounding human-machine interaction persists at the micro-level of allocating responsibility for algorithm-assisted decision-making [4]. As Cath [5] rightly notes, moreover, the inscrutability of AI's effects on society increases as AI becomes more widespread and normalized: "the more AI matters the less one may be able to realise how much it does" (p. 507). This is particularly challenging for public sector workers mandated to protect the public good, but who may not recognize the societal risks and ethical challenges posed by AI. Research on public administration has demonstrated that recognition of the problems that public policy should solve is dependent not only on individuals' personal values [6], but on the international normative frameworks to which they are exposed and the salience of policy problems within those networks [7]. Even when public sector workers recognize these challenges, however—and they increasingly do [8]—individuals often lack the skills and capacities to address them [9–11]. A recent review of research on public sector AI describes this knowledge gap as "a critical development barrier" for many governments, as the slowly burgeoning thought leadership on AI governance fails to match "the pace with which AI applications are infiltrating government globally" [10, p. 2].

There are limited resources to help public sector workers manage this complexity in governing AI. Though there has been a proliferation of conceptual frameworks and principles, labels such as responsible,

* Corresponding author.
E-mail addresses: christbw@uio.no (C. Wilson), majava@ifi.uio.no (M. van der Velden).

https://doi.org/10.1016/j.techsoc.2022.101926
Received 30 October 2020; Received in revised form 4 February 2022; Accepted 4 February 2022
Available online 7 February 2022

ethical, and explainable AI tend to be highly conceptual and not well suited to guide decision-making. The discourse of ethical AI is instructive in this regard. In applied contexts, *ethics* and ethical responsibility have proved remarkably challenging to define [13–15]. As a discourse, the abstract quality of AI ethics has proved vulnerable to elite capture [16] and hijacking [17–19]. More operational tools, meanwhile, tend to target users with technical expertise, like the Equity Evaluation Corpus or the Tensor Flow Privacy Library (see Ref. [20] for a comparison), and are not easily accessible or operationalized for decision-making by public managers, administrators, or civil servants. The few practical resources designed to assist non-technical public sector workers with AI governance challenges tend to emphasize deliberative and participatory processes in the interest of government accountability (see for example [4,21], and their lack of use may be due to perceptions in the public sector that the burdens of participation and accountability initiatives outweigh any potential benefits [22–24]. This is a problem for public sector governance of AI, which requires frameworks that are both operational and conceptually coherent in order to actually be *ethical* or *responsible*.

The aim of this paper is to explore if the concept of *sustainable* AI, as grounded in the theory and practice of sustainable development, might be better suited to guide decision-making about how to regulate and implement AI. Though also abstract at first glance, alignment to the Sustainable Development Goals (SDGs) framework allows the concept of sustainable AI to draw on the policies and practices associated with the SDGs, which have been elaborated, road tested and integrated into public sector practice for several years now. This can increase the salience and accessibility of the concept and makes it easier to operationalize in public sector decision-making than more diffuse notions of responsibility or ethics.

The notion of sustainable AI is often casually, but increasingly referenced in research on AI governance [25], in policy debate [20], and even in national strategies for dealing with AI [26]. The concept is not clearly or consistently defined, however, and it is not clear how or to what degree it would support public sector decision-making in governing AI. To explore this potential, this research asks*: How can the concept of sustainable AI be defined and operationalized to guide public sector decision making?*

The article proceeds as follows. Section 2 presents background on the context for public sector decision-making about AI governance and about how the concept of sustainability has been conceptualized in the context of sustainable development and the SDGs. Section 3 then presents our research approach, and section 4 presents the results of our review of the literature, with an emphasis on social sustainability and operationalizes social sustainability through boundary conditions necessary to preserve social functions. It then analyses these boundary conditions by integrating the literature on sustainability and public sector AI in order to arrive at a preliminary conceptual model of sustainable AI for the public sector. Section 5 discusses the analysis and presents the integrated model of boundary conditions for sustainable AI, together which operational considerations for how it can be applied in public sector decision-making. The final part of this section presents some concluding remarks as well as outlines the limitations of this study.

2. Background and context

2.1. Public sector decision-making about AI governance

The rapid diffusion of AI technologies presents the public sector with a novel set of regulatory and adoption challenges. Public managers, administrators, and civil servants must make decisions about how to balance the potential benefits of these technologies against their potential harms. We refer to this collectively as AI governance, whether it involves decisions about how AI should be adopted or regulated. Notably, the potential benefits of AI differ across each of these dynamics, with regulatory decision-making oriented towards the maximization of economic and societal benefits [5,27], while decisions about AI adoption emphasize administrative efficiency and improved public service delivery [28,29]. The risks implied in both types of decision making are largely synonymous, however, and are often articulated in relationship to high level concepts such as bias, fairness, and privacy, and the obligation to uphold fundamental democratic principles [10].

Whether deciding on appropriate disclosure requirements, private sector data management, or quality assurance processes for algorithmic decision making in case management, addressing these risks is challenging in the public sector due to a variety of factors, including a fast-moving policy discourse [30], skeptical publics [31,32], and the limited capacities of government workers [9] and institutions [33]. These challenges are salient at all phases of the policy process and "permeate all layers of application," as noted in a recent systematic review of research on AI in the public sector [1, p. 1]. As discussed in this article's introduction, however, most frameworks and concepts for addressing these risks are difficult to operationalize in the public sector and tend either towards high levels of abstraction or technical detail.

One notable exception to this trend is the "Integrated AI Governance Framework for Public Administration" developed by Wirtz et al. [8], which is oriented towards helping the public sector to manage the most salient challenges identified in a previous literature review, and grouped as shown in Table 1. The challenges identified here are dissimilar in many regards, spanning both practical and abstract considerations, but nevertheless provide important clarity in the otherwise sprawling literature on AI and society, by indicating the types of issues and challenges that are most salient in the context of public sector decision-making.

Wirtz et al.'s integrative framework further suggests that public sector workers address these challenges through a combination regulatory, policy, and collaborative efforts, which include specific components such as "hazard identification" and "monitoring of unintended effects" at the regulatory layer, or the development of Standards at the public policy layer. This is also a valuable contribution insofar as it moves past the conceptual ambiguity of notions like "ethical AI" to provide a menu of activities that public sector workers can pursue to manage risks and challenges related to AI governance. The framework does not, however, provide substantive guidance that public sector workers can use to make actual decisions about how to conduct risk/benefit analysis or develop standards.

Though the integrated framework suggested by Wirtz et al. does not provide concrete guidance on *how* to make decisions about AI regulation or implementation, it does highlight the salience of societal and social issues for public sector decision-making, and how these are linked to concepts like safety, justice, and fairness. This provides a frame for narrowing the concept of sustainable AI for public sector decision-making, with a focus on finding the balance between AI's potential societal benefits with AI's potential societal harms.

2.2. Operationalization of sustainability in the context of the SDGs

The concept of sustainable development is anchored in the 1987

Table 1
Main governance challenges for the public sector, adapted from Wirtz et al. [8].

Type of challenge	Challenge
AI & Society	Workforce transformation
	Societal acceptance
	Transformation of human-to-machine interaction
AI Law and Regulation	Governance of autonomous intelligence systems
	Responsibility and accountability
	Privacy and safety
AI Ethics	AI-rulemaking for human behavior
	Compatibility of AI vs human value judgement
	Moral dilemmas
	AI discrimination

Brundtland Report as "development that meets the needs of the present without compromising the ability for future generations to meet their own needs" [34], and has since been elaborated in a variety of contexts and has driven global collaboration. This broad notion of sustainability is widely understood to consist three pillars: environmental, social, and economic sustainability, whose relationships and interdependencies be conceptualized in a variety of ways within the sustainable development paradigm [35].

International collaboration for sustainable development has culminated in the Sustainable Development Goals (SDGs), which were adopted by all 193 UN member states 2015 [36]. The SDGs are highly operational, consisting of 17 broad goals, in turn consisting of 169 specific targets and nearly 300 indicators, and countries progress towards achieving these targets is closely monitored by UN Agencies as well as independent organizations, and citizen initiatives [37]. Because country contexts and capacities vary so significantly, the United Nations Development Programme provides support to countries to institutionalize efforts to achieve sustainable development, through the creation of regulatory structures such as National Councils, and processes for coordinating between legislative, executive, and other public service and administration agencies [38]. A recent mapping of sustainable development policy intermediaries found over 120 independent online resources to support countries in this regard [39], and early reporting on countries' self-assessments to the UN suggest that this has supported broad diffusion of the SDG framework across developed and developing country contexts [40].

In parallel with the diffusion of the sustainable development paradigm among national governments, significant work has been done to define sustainability across sectors and in other operational contexts. Most notably, over 25 years of academic collaboration and review led to the development of a Framework for Strategic Sustainable Development (FSSD) [41]. Notably, the FSSD operationalizes the concept of sustainability in terms of defining "boundary conditions" and setting red-lines that must be respected in order to protect "the basic conditions that are necessary to fulfill for the ecological and social systems to not degrade systematically," [41]. As an articulation of sustainability in negative terms of what cannot be compromised, this conceptualization contrasts strongly with most positive conceptualizations of sustainability as an objective in the public sector [42], including the SDGs.

Since its launch, the FSSD has been tested and applied in a variety of contexts and with a variety of actors from the public and non-profit sector, testing its utility to

> give guidance on how any region, organization or project can develop a vision framed by principles for social and ecological sustainability, analyse and assess the current situation in relation to that vision and thus clarify the gap, generate ideas for possible actions that could help to bridge the gap, and prioritize such actions into a step-wise and economically attractive plan [43].

These efforts have highlighted the importance of an iterative approach to simultaneously operationalizing and defining sustainability concepts.

3. Research approach

Despite significant operationalization of the concept of sustainability, and a clear operational need for public sector decision-making about AI governance, the key challenge for this analysis is that the concept of sustainable AI is asserted regularly but inconsistently. It has been casually referenced in regard to the environmental consequences of advanced computing power [44,45], as a corporate strategy [46,47], as measure of the degree to which AI threatens human safety [48], and as a social movement oriented towards social justice [49]. Many of these references are highly casual and tangential, resulting in a scope of literature that is too broad and diffuse to provide guidance. Our research question guiding this study could thus be formulated as follows: How to operationalize the concept of sustainability AI for public sector decision-making.

To answer this question, we adopted a six-step approach to i) reviewing the literature on sustainable AI; ii) establish the applicability of sustainable AI to the target context of public sector decision-making about AI governance; and iii) formulate conceptual and operational definitions of sustainable AI for the public sector. An overview of the six steps is presented below in Table 2.

In the first step, we aimed to capture deliberate conceptualizations of sustainable AI in research. We did this by querying Scopus and Web of Science data bases on precise search terms 'sustainable AI' and 'sustainable artificial intelligence' and results were filtered to include only results that had these terms had to be found in the title, abstract, or keywords of the sources. Google Scholar was queried for articles with the same terms in titles only. This returned 16, 14, and 21 articles from each query respectively. Eight articles from the Scopus search, 5 articles from the Web of Science search, and 12 articles fulfilled our criteria. After eliminating the duplicates and applying filters in step 2, 15 articles remained that contained definitions or descriptions of one of the search terms. These articles also referenced three research institutions with a mandate for exploring sustainable AI and had that term in their name. These were also included in the review, in order assess how sustainable AI is conceptualized in the research community beyond peer reviewed research. This resulted in 7 definitions and 10 descriptions of sustainable AI. Despite their limited number, the results of this search provide a strong and representative baseline of how sustainable AI is conceptualized in the research community and aligns with the lack of consistency and clarity described in reviews of the sustainability literature. In the fifth step we compare the definitions and descriptions with the target context of public sector decision-making in described in the previous section, in order to narrow our review to a subset of the relevant literature focused on the Framework for Strategic Sustainable Development and social sustainability and define a structure for aligning the literatures on sustainability and AI in the public sector. The final step

Table 2
Process for literature review and analysis.

Step		Focus	Results
1	Data base query	Scopus, Web of Science, Google Scholar: "Sustainable AI" or "Sustainable artificial intelligence"	n = 51
2	Filter	1. Removal of duplicates 2. Search terms in title, abstract, or key words 3. Contain descriptions or definitions of sustainable AI	N = 15
3	Expansion	Research centers mentioned in articles that have an explicit nominal focus on sustainable AI	• *The Nordic center for Sustainable and Trustworthy AI Research (Nordstar)* in Oslo • *AI Sustainability Center* in Stockholm • *Sustainable AI Lab* in Berlin
4	Review	Definitions and description of sustainable AI	7 definitions and 10 descriptions presented in Table 3
5	Comparison with target context	Operational context for public sector decision-making about AI governance	Focused review on social sustainability and the FSSD and identification of 5 boundary conditions for sustainable AI
6	Integration of literatures on sustainability and AI	Boundary conditions for maintaining social and environmental sustainability	5 boundary conditions for social sustainability that can be applied in the context of sustainable AI

integrates those literatures according to five boundary conditions and asserts operational and conceptual definitions of sustainable AI for the public sector.

4. Results and analysis

4.1. Mapping conceptualization of sustainable AI in the research community

A review of how research and research institutions have conceptualized sustainable AI results in 7 specific definitions and 10 specific descriptions, which are presented below in Table 3. All of these conceptualizations explicitly reference the conceptual paradigm of sustainable development associated with the SDGs but have a different focus on the different dimensions of sustainability. They also differ according to their contexts of application (specific sectors, industries, or legal contexts), the relationship between AI and sustainability (AI applications that are themselves sustainable vs AI applications that support or promote sustainability). These differences are broadly but unevenly distributed across the conceptualizations returned from the first steps in our literature (see Table 3).

In regard to the dimensions of sustainability, four instances focus only on environmentally sustainable AI [55,57,63,65], while others only focus on socially sustainable AI [52,60,62]. Half of the articles have a holistic understanding of sustainability. Economic sustainability is not featured independently. Rohde et al. [50] and Tsafack Chetsa's [56] definitions are close to the so-called Triple Bottom Line [67] or Three Pillars [68] definitions of sustainability, while others refer to the Brundlandt definition of sustainability [54,58]. Vinuesa et al. [53] define sustainable AI as AI that enables achieving the Sustainable Development Goals (SDGs), while the SDGs and sustainable development are also described as guiding frameworks for the development and evaluation of sustainable AI [58,61,62].

In regard to application contexts, several conceptualizations discuss sustainable AI on a general policy level [49,50,53,62], referring to Agenda 2030 and the EU. The sustainability of AI as a technological product is the focus of three articles [57,63,64]. Several articles discuss sustainable AI in a particular field: consumer autonomy [54], business innovation [55], and human rights [59].

In regard to the relationships between sustainability and AI, eight resources discuss the need for the sustainability of AI [49,50,55,57–59,62–64]. They mention for example the need "to foster change in the lifecycle of AI products (i.e., idea generation, training, re-tuning, implementation, governance) towards greater ecological integrity and social justice" [49] and to "have the minimum carbon footprint" [63]. Two resources focus on how AI can contribute to achieving sustainability [53,54]. The final three resources mention aspects related to both AI for sustainability and the sustainability of AI [51,60,61].

4.2. Narrowing the review towards the target context

Three conclusions can be drawn from the broad review of how sustainable AI has been deliberately conceptualized by the research community. Firstly, consistent reference to the sustainable development paradigm justifies the use of that framework to operationalize the concept of sustainable AI. Secondly, the variety of application contexts and AI sustainability relationships suggests a complex and fragmented operational landscape. This aligns broadly with the vast array of substantive issues at issue in public sector contexts [1], and with efforts by Wirtz et al. [8] to integrate these with the challenges, mechanisms, and layers of AI governance in the public sector. It also confirms the need for more hands-on operational tools, such as those developed under the Framework for Strategic Sustainable Development and suggests that this framework might be used to operationalize the sustainable AI for public sector decision-making.

Lastly, differences in how the literature attends to the different

Table 3
Definitions and descriptions of Sustainable AI.

Definitions
1) "[D]eveloping, implementing, and using AI in a way that minimizes negative social, ecological and economic impacts of the applied algorithms (sustainable AI)" [50].
2) "The AI Sustainability Center supports an approach in which the positive and negative impacts of AI on people and society are as important as the commercial benefits or efficiency gains. We call it Sustainable AI" [51,52].
3) "Sustainable AI is a movement to foster change in the entire lifecycle of AI products (i.e. idea generation, training, re-tuning, implementation, governance) towards greater ecological integrity and social justice" [49].
4) "Sustainable AI is AI that enables reaching the SDGs" [53].
5) "(…) the extent to which AI technology is developed in a direction that meets the needs of the present without compromising the ability of future generations to meet their own needs" [54].
6) "[S]ustainable artificial intelligence that is not harmful but beneficial for human life" [55].
7) "One can think of sustainable AI/DS as AI subjected to organizing principles, including, but not limited to, processes which could be organization specific, regulations, best practices, and definitions/standards for meeting the transformative potential of DS while simultaneously protecting the environment, enabling economic growth, and social equity" [56].
Descriptions
1) "This work explores the environmental impact of AI from a holistic perspective. More specifically, we present the challenges and opportunities to designing sustainable AI computing (…)" [57].
2) "[S]ustainable development (SD) (Brundlandt) should be the guiding framework for research and development of artificial intelligence (AI). Instead of merely focusing on ethics or human rights, scholars and policy makers should acknowledge sustainable AI development (SAID) as guiding framework" [58].
3) "The growth in AI and automation will continue regardless of how the space is regulated and monitored. Whether it is sustainable in the long run, though, and is viewed positively rather than negatively, will depend on taking a responsible, rights-based approach" [59].
4) "Exploring the connections between an AI's technical design and its social implications will be key in ensuring feasible and sustainable AI systems that benefit society and that people want to use" [60].
5) "[T]he SDGs provide an ideal framework to test the desirability of AI solutions" [61].
6) "The objective of an inclusive, sustainable, and human-centered AI in Europe will likely require a normative frame- work at the European level. Financial and regulatory stimuli are required to foster SDG-driven AI and public–private collaboration in the sharing of technology and data (…). Furthermore, a human-centered AI should be human rights-based (…). Although there has been some limited discussion at the European government level of the impact of AI on human rights, especially regarding the right to privacy, the impact on social, economic, and cultural rights has so far received little attention" [62].
7) "For example, time has come to focus on sustainable AI (Pal, 2018b). Here we like to refer to two issues: The first issue is that the development (training) of the AI system should have the minimum carbon footprint. To achieve human-like performance often this important issue is ignored. To illustrate the severity of this issue we consider a recent study which used an evolution-based search to find a better architecture for machine translation and language modeling than the Transformer model (So et al., 2019). The architecture search ran for 979 M training steps requiring about 32,623 h on TPUv2 equivalently 274,120 h on 8 P100 GPUs. This may result in 626,155 lbs of CO2 emission–this is about 5 times the lifetime average emission by an American Car (Strubell et al., 2019). The second point is that the solutions provided by an AI system should be sustainable with the minimum impact on the environment" [63].
8) "From the perspective of AI Ethics, Aimee van Wynsberghe defined the term sustainable AI as ". . . a field of research that applies to the technology of AI (. . .) while addressing issues of AI sustainability and/or sustainable development" [3]. In other words, the term of sustainable AI takes into consideration the entire AI lifecycle, from training to its implementation and use" [64].
9) "First, state-of-the-art algorithms in AI demand massive computing power and energy: to handle the ever-increasing Big Data repositories, AI systems must scale in proportion to all the available data. In other words, future AI must be sustainable" [65].
10) "measuring & assessing the environmental impact of AI, ways of making AI systems more sustainable, and directing AI towards the sustainable development goals [66].

dimensions of sustainability, social, environmental, and economic, highlights the importance of social sustainability for public governance decision-making, and the emphasis on protecting societal values related to justice, fairness, and safety in AI governance, as described in section 2.1. This focus is not exclusive, and interacts significantly with dimensions of sustainability [35], as emphasized in a recent review of AI initiatives for environmental sustainability [3]. The review found that their application introduced systemic social risks "since the application of AI-technologies in combination with globalization processes, are likely to create novel connections between humans, machines and the living planet including ecosystems and the climate system" [3]. As a point of departure, however, this suggests that application of the FSSD to social sustainability provides useful tools for operationalizing the sustainable AI concept for public sector decision-making.

4.3. Social sustainability in the context of strategic sustainable development

Of the three forms of sustainability, social sustainability has often been regarded the most undertheorized [69], and has been differently defined across disciplines [35,70–75]. This lack of clarity is perhaps also why social sustainability has often been overlooked in national frameworks and reporting for sustainable development [35,76], and has prompted efforts to assert a scientifically grounded and operational definition of social sustainability through the FSSD [43,77]. Their effort leveraged a literature review and systems mapping to "better aid more concrete planning and decision-making for sustainable development" across sectors, in a way that is broadly analogous to the current analysis (p. 34). This resulted in a conceptual model and principled definition. Most relevant for public sector decision-making, the authors also applied the boundary conditions articulated in the FSSD as boundaries that any program or initiative must avoid violating in order to preserve social sustainability:

> By clustering a myriad of down-stream impacts into overriding mechanisms of degradation and equipping them with a "not" to serve as exclusion criteria, boundary conditions for redesign are derived. The sustainability of the [social and ecological] systems and the definition of the goal (sustainability) at the principle level then creates the space and opportunity for people to meet their needs in whatever way they chose and for societies to create scenarios to prosper and flourish [42, p. 35].

The authors identified five specific boundary conditions: *diversity*, *capacity for learning*, *capacity for self-organization*, *common meaning*, and *trust*. The remaining subsections will consider each of these in the context of the broader debate on AI, ethics and society, and the operational implications this has for public sector decision-making about AI governance.

4.4. Boundary conditions for sustainable AI

4.4.1. Diversity

Missimer et al. [43] describe three types of diversity as boundary conditions for sustainable development, including a mix of social components "whose history and accumulated experience help cope with change," diverse types of knowledge used to understand systems, and "diversity in governance as a source for resilience" (p.36-37). This resonates strongly with calls to ensure the inclusion of "diversity and inclusion within system development teams and stakeholders, broadening and diversifying the sources of knowledge, expertise, disciplines and perspectives" [78] that inform the design, implementation, and regulation of AI.

Diversity and inclusion in AI processes are particularly important for the public sector, and governments have been prominently encouraged to initiate multi-stakeholder consultative and deliberative processes to develop governance systems for AI [5,13]. The OECD [28] urges the public sector to "provide for multi-disciplinary, diverse, and inclusive perspectives" in shaping national approaches to AI and argues that the inclusion of diverse perspectives is "perhaps the main enabling factor to achieving AI initiatives that are both effective and ethical, both successful and fair" (p. 101). Of particular importance in this regard is representation of the perspectives and lived realities of different social groups who will in some way interact with the AI; particularly historically disadvantaged groups that may not have equal access to AI services or the processes through which they are developed, or that are at risk of further marginalization as a result of the use of AI-based public services [79–81]. Thus UNESCO [82] argues that "anyone or any entity with a legitimate or bona fide interest in an issue brought about by the AI development can be considered as a relevant stakeholder," and that multi-stakeholder participation can help "prevent the domination of the Internet and other new technologies by one constituency at the expense of another. (p. 116).

In the context of public sector decision-making about AI governance, the boundary condition of diversity can thus be understood as avoiding the degradation of social sustainability through elite capture of AI systems or the exclusion of affected stakeholders from AI governance. This is challenging insofar as the FSSD framework implies a starting point where the sustainability of systems has not yet been degraded, but there appears to be broad agreement in much of the literature on AI and society that diversity and inclusion in AI development and implementation are exceedingly rare. We may be starting from a default position of degradation in regard to this boundary condition, because of the inherently opaque and esoteric nature of AI.

Simultaneously, widespread criticism of this state of affairs has produced a wealth of literature providing guidance on how to avoid exclusion and capture of AI systems through the inclusive participation of stakeholders in the development, implementation, monitoring or review of AI platforms. This includes the use of national level multi-stakeholder fora [82], national commissions for regulation [5] or trust-building [83], or citizen assemblies to provide a broad representation of inputs to design and monitoring of AI implementation [84]. At a more operational level, toolkits have been developed to help governments conduct public deliberative processes [85] and inclusive impact assessments for AI [21]. Frameworks for embedding these perspectives in AI systems will be discussed in section 4.4.4 below.

4.4.2. Capacity for learning

The boundary condition of learning capacity is described as the ability "to sense changes and respond to them effectively […] and includes social memory, the capacity to learn from experience, as a mechanism" [43]. The capacity for societies to learn from their interactions with AI platforms and systems begs a number of questions and preconditions. The most fundamental of these is basic awareness of AI, regarding how these technologies work and how they are being implemented. Surveys suggest that awareness in this regard is generally low among the public [86,87], and that this can be closely linked to the public's skepticism with AI [8]. Indeed, there appears to be a fundamental human tendency to blame AI when things go wrong [88].

While AI's technical complexity and opacity has often been suggested as a reason why it is not feasible to engage with the general public on AI issues [89], there has been a dramatic surge of research and advocacy aiming to make algorithms and related processes explainable [90–92]. Though some critics have argued that the notion of algorithmic explainability assumes the existence of an engaged and critically informed audience to whom AI might be explained [93], a number of practical frameworks have been developed to help governments engage non-experts in technical discourse [94]. Balaram et al. [86] note, moreover, that deliberative processes have been demonstrated to be particularly well suited to making sense of contentious or technically complicated topics. Robbins [95] suggests a gradated understanding of knowledge about AI, in which some types of meta-level knowledge, for

example regarding things like "training data, inputs, functions, outputs, and boundaries," can be leveraged for regulation and monitoring without requiring more detailed technical knowledge and know-how (p. 391).

In the context of public sector decision-making about AI governance, the boundary condition would thus not protect fully transparent AI as lines of code, but the fundamental discoverability of algorithmic processes and systemic interactions, such that it is possible for stakeholders to determine how AI systems are developed and where decisions are made in their interaction with human agents. As with the boundary condition of diversity, the AI's inherent opacity has already led to a status quo of degraded systems in regard to explainability. Public sector decision-making needs to go beyond the preservation as anticipated in the FSSD and proactively seek to make AI systems discoverable and decipherable to the general public. In an operational sense, this most immediately requires that public sector workers not simply accept the idea that AI is an unknowable black box, and instead explore mechanisms for inclusive audits of AI systems [21], participatory technology assessments [96], systematic disclosure systems [97], the appointment of data stewards [98], or the use of external organizations as knowledge brokers to specific groups stakeholders [99].

4.4.3. Capacity for self-organization

As "complex adaptive systems are usually self-organized systems without system-level intent or centralized control," Missimer et al. [43] argue that the capacity for social systems to self-organize "is especially important when confronted with a sudden change in the environment" (p. 37). Without that capacity, social systems are unable to rapidly adapt and respond to disruptive changes, such as those posed by AI to the "informational foundations" of contemporary society [96].

As with the previous boundary conditions, critical research suggests that AI is already degrading society's capacity to self-organize and manage the societal impacts of AI, primarily because "power" to understand, engage, and shape AI is increasingly concentrated in specific societal groups [82], while other groups are increasingly vulnerable to AI bias and discriminatory outcomes, and their capacity to do something about it is increasingly diminished because they lack the means to organize and engage [100,101]. While this is often framed as problem of agency, empowerment, and accountability, it is equally a problem of consent, because AI systems are simultaneously ubiquitous and invisible [102]. These systemic challenges to agency and engagement are further exacerbated by the ways in which specific algorithms have been used to fragment political communication and undermine political agency in the public sphere [82,103].

The preservation of the public sphere falls outside the scope of most public sector decision-making about AI governance, but this boundary condition can be read to obligate the preservation of institutional mechanisms by which social groups engage with AI systems through public institutions and in the public sphere. This understanding resonates strongly with notions of algorithmic accountability [104] that have been prominent in popular debate, but which have failed to find any operational purchase [105].

In seeking to operationalize this condition, public sector decision-making might instead rely on self-sovereign identity frameworks that help users of algorithms and data systems to manage their data ownership and consent [98], alternative platforms and formats for informing the consent of system users [106], toolkits for non-expert engagement in algorithmic design and review [94], or institutional mechanisms for grievances and redress that are familiar from other institutional contexts, but which are easily adapted to AI governance [107].

4.4.4. Common meaning

In explaining the boundary condition of common meaning [43], emphasize

> "the role of common culture and meaning in the creation of social capital, both horizontal and vertical. Particularly in the absence of a long history of reciprocity and the trust which that engenders, stakeholders will often make the decision to enter into the initial reciprocities on the basis of their belief that they share representations, interpretations, and systems of meaning" (p. 37).

These issues are best understood in regard to values that are embedded in AI system, and how they influence AI interactions with society and societal outcomes [108]. Drawing on UNESCO's [109] notion of power differentials in any given society, the question is whose values are pursued and embodied by AI. For the public sector, this implies a boundary condition of ensuring that AI systems do not embody or manifest values that are antithetical to societal values or the values held by affected stakeholder groups.

Protecting this boundary condition requires the public sector to identify key societal values and to ensure that they are embedded in AI platforms whose behavior will to some degree be opaque and unpredictable. The deliberative and consultative processes discussed in regard to diversity are good mechanisms for identifying and defining values, though Dignum [78] notes that the most determinant values in AI systems are often implicit and dependent on socio-cultural context, and so require specific methodologies to make those explicit in AI design and implementation. This call has been met by a host of technical and procedural methods for the value-sensitive algorithmic design [110–112]. Simply embedding values in AI is likely insufficient to protect this boundary condition, however, because AI systems are capable of pursuing conflicting objectives simultaneously [113], and can develop in surprising and opaque ways over time, either through their own learning processes [114,115] or through hidden feedback loops with human actors and other algorithmic processes [116,117].

In line with Neyland's [105] notion of *accountability in action*, protecting the boundary condition of common meaning and societal values requires a process-based approach to embedding values in AI systems, and Rahwan's [118] *society-in-the-loop* (SITL) paradigm describes an architecture with which to do so. This builds on the concept of *human-in-the-loop* systems, whereby individual people interact with AI processes and outcomes to improve their accuracy, identify deviance from desired outcomes, and provide accountability. Because AI systems increasingly serve broad social functions and humans are prone to bias and fallibility, Rahwan proposes a paradigm, in which questions about fundamental rights, ethical values and societal preferences are directly and regularly directed to the human controllers interacting with AI systems. Rahwan identifies several mechanisms for monitoring and enforcing AI compliance with societal values, including reporting, auditing, and oversight programs. The precise mechanisms that are most appropriate in any give context will in turn be defined in regard to societal values and should be determined through the inclusive mechanisms described above.

4.4.5. Trust

Missimer et al. [43] understand trust as a driver of social capital, closely linked to the boundary of common meaning described above. The concept of trust is, however, an exceptionally prominent marker for *good AI* in the debate on AI and society (see the European Commission's Ethics guidelines for trustworthy AI,[1] OECD Principles on Artificial Intelligence,[2] IBM's Principles for Trust and Transparency[3]) and deserves special consideration. As a boundary condition, Missimer et al. [43] define trust as "a quality of connection, which allows the system to remain together despite the level of internal complexity" (p 37), which

[1] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
[2] http://www.oecd.org/going-digital/ai/principles/.
[3] https://www.ibm.com/blogs/policy/trust-principles/.

resonates strongly with the notion that societal trust is a necessary condition for pursuing social good outcomes through AI [13,119], which in turn requires specific investments from the public sector [120].

The over-simplified rhetoric of trust-building in global policy discourse, e.g. Ref. [121], belies the complex and highly contested quality of trust as a concept [122]. In the context of AI and society, several scholars have noted that social trust in AI systems should not be blind, but *appropriate* [123,124], suggesting that the boundary condition in this instance is to preserve the foundation for trust, rather than to preserve trust itself. This in turn requires "allowing people to determine the conditions and parameters under which algorithms operate and to redefine the boundaries between trust and privacy" [8, p. 373]. This is complicated, however, by the fact that most people engaging with AI systems, knowingly or not, will lack the technical and practical expertise to make informed judgements about whether to trust those systems [125]. Often, an individual's trust in AI will be extended through trust in proxies for those systems. In some instances, this can be explicit. For example, Bodó [126] has described trust mediators, which vouch for and explain the conditions of trust, in much the same way as Janssen et al. [98] describe data stewards. In other instances the proxy nature of trust is implicit because trust is systemic in nature, invested in the larger system of public and private actors that are associated with the AI at issue [127].

Understood as such, the boundary condition of trust compels the public sector to not only protect active and explicit trust in AI, but to protect the trust that is implied by use of those AI and related systems and services, knowingly or not. This can be most easily operationalized in terms of protecting the trustworthiness of AI, in other words, to ensure that AI is not designed and used in any way that, if known, would erode trust in that AI or the systems and services in which it is embedded. Google's earlier and now infamous motto: *don't be evil* becomes the histrionic corollary to this boundary condition for public sector decision-making: *don't betray implied trust.*

Operationally, several of the approaches and mechanisms described above can contribute to a hands-on evaluation of whether AI governance decisions risk betraying implied trust in AI, including adapted mechanisms for facilitating informed consent [102]. Publicly visible institutional mechanisms such as third party auditing, sharing of incidents, and *bias bounties* have also been recommended [128]. National dialogues and the development of national institutions mandated to foster public trust in AI can support all of these [5,83]. It is also worth noting the potential backfire of superficial efforts to build public trust through marketing and public relations [129].

5. Discussion and conclusion

5.1. An integrated model of boundary conditions for sustainable AI

The preceding section presented the results of our 6-step literature review and analysis. As described in Table 2, this began with a broad query about how sustainable AI has been deliberately conceptualized by the research community and concluded with a narrower discussion of boundary conditions for social sustainability in the FSSD. The section closed by discussing how each boundary condition is reflected in the broader debate about AI and society, and how those conditions might be operationalized for public sector decision-making about AI.

In doing so, we found that the boundary conditions for strategic social sustainability identified by Missimer et al. [43] align well with contemporary debate on AI ethics and society, and have a good conceptual fit with the notion of sustainable AI mapped at the beginning of our literature review. Importantly, we find that the broader literature on AI and society helps to identify clear criteria and approaches for how each of these boundary conditions can be operationalized in public sector decision-making. This is presented in Table 4, where boundary conditions for sustainable development are aligned with corresponding concepts from the broader literature on AI and society, suggesting operational criteria, approaches, and guiding questions for each (references for operational approaches can be found at the end of each

Table 4
An integrated model of boundary conditions for sustainable AI.

Boundary conditions	Corresponding concepts	Criteria for preservation	Operational approaches[a]	Guiding questions
Diversity	Inclusive participation	• All stakeholders affected by or interacting with the AI system • Emphasis on stakeholders with traditionally limited access	• National level multi-stakeholder dialogues and assemblies • National commissions for regulation or trust-building • Deliberative processes • Inclusive impact assessments for AI	Does the design and implementation of AI incorporate the views and needs of all affected stakeholder groups?
Capacity for learning	Transparency and Explainability	• Discoverability and unknowability of process	• Inclusive audits of AI systems • Participatory technology assessments • Systematic disclosure systems • External knowledge brokers	Do people who are affected understand how it works and what the outcomes are?
Capacity for self-organization	Agency, Consent, and Accountability	• AI are subject to democratic principles and institutions • In regard to design, implementation, and monitoring phases	• Grievance and complaint mechanisms • Informed consent • Toolkits for non-expert engagement • Systematic disclosure systems	Do people know they are using it? Are affected stakeholders invited to engage in design and review of AI is implemented? Is there a clear complaint mechanism for affected stakeholders?
Common meaning	Embedded values	• Understanding which values are represented by AI systems? • Identifying the values held by society and by affected stakeholder groups	• Society in the loop o Facilitated social debate on values • Human controllers that oversee and update AI systems	Does the implementation of AI match the values held by affected stakeholder groups and society in general?
Trust	Appropriate and systemic trust	• Trust is defined by those who actively choose to trust AI systems. • Trust is deserved	• Alternative platforms for informed consent • Trust mediators • National institutions for trust-building	Should people trust the AI at issue?

[a] = The approaches listed in this column are also described, with references, at the end of each subsection 4.4.1 - 4.4.5.

sub-section on boundary conditions in the previous section).

Collectively, we present these conditions as an integrated model for public sector decision-making that is both holistic and preliminary. Holistic, because the boundary conditions are interdependent and regularly redundant. This is evident both in the operational approaches, many of which may well help to protect other boundary conditions than those with which they are associated in Table 4, and dependent of many of the corresponding concepts (i.e., neither trust or agency are obviously feasible without transparency and explainability).

The model is also preliminary; it is intended to be refined and applied iteratively, because it is premised on protecting a status quo which does not in fact seem to exist. As mentioned in the discussion of several boundary conditions, the inherent opacity and inaccessibility of AI technologies has led to a situation in which many of these boundary conditions are already degraded. AI technologies and systems are by default neither inclusive nor transparent. As a result, public sector decision-making must take its own limited mandate and the obligation to do no further harm as its point of departure in applying this integrated model to decision-making. Discrete decisions about how algorithms and machine learning are used in managing social welfare cases will not be able to protect the public's capacity to self-organize in any grand sense. It will, however, be able to assess how the public could organize in response to this specific initiative, and that in turn will have knock-on effects in regard to the other boundary conditions and beyond the specific AI implementation at issue.

This model is in keeping with how the FSSD was intended to be leveraged [43]. It also provides a clear complement to discrete tools that are intended to support the public sector in specific aspects of AI governance, and to the integrated framework asserted by Wirtz et al. [8], insofar as it provides a decision-making framework for actual implementation of its various components.

5.2. Concluding remarks and limitations

This paper aimed to explore how the concept of sustainable AI can be defined and operationalized to guide public sector decision-making, in order to support efforts to operationalize ethical AI in the public sector. In doing so it found that there is no widely held definition of sustainable AI, despite its increasing use in the research context. Conceptualizations of sustainable AI nevertheless did consistently reference the sustainable development framework associated with the SDGs, however, validating the use of that paradigm to elaborate the concept in the context of public sector governance. In addition, aligning a more narrow review of that literature with the public sector decision-making identified five boundary conditions for sustainable AI, which the public sector should aim to preserve. Considering these in the context of AI governance, resulted in the following boundary conditions for sustainable AI in the public sector: 1) diversity and inclusion; 2) capacity for learning, transparency and explainability; 3) capacity for self-organization, agency, and accountability; 4) common meaning and embedded values; and 5) systemic and implied trust. These five conditions were presented together as an integrated model, together with operational conditions and approaches that can be leveraged to inform public sector decision-making.

By proposing this integrated model, this paper makes several contributions to both theory and practice of sustainable AI. Most immediately, this involves contributing clarity and rigor to what is in danger of becoming a buzzword in both policy and research discourse about AI and society. Most notably, the boundary conditions here are presented in a manner that facilitates the elaboration of empirical indicators in-line with concept explications that have been advanced in the communications theory [130], or Goertz's seminal method for defining social science concepts [131]. Careful applications of these methods would provide a framework for elaborating some of the theoretical premises implied above and the types of contexts in which they are theoretically sound [132]. This is a crucial first step before empirically assessing whether the theories implicit in these boundary conditions do in fact hold (for example, that inclusive participation does indeed strengthen and protect social sustainability).

In terms of policy and practice, this paper contributes towards operationalizing vague discourses about how AI should be considered in the public sector. Building on important work by Wirtz et al. and others [8,21,86,133], the integrated model for sustainable AI presented here provides a series of practical tests that can be applied by public sector workers at varying degrees of detail when making decisions about how to use or regulate AI. This does not in itself solve the inherent knowledge and capacity gaps that are manifest around AI in the public sector [12]. The integrated model, and the approaches it references, require further explorations, consideration, and contextual analysis to determine if and how they should be applied. It is not certain that this model is immediately applicable across contexts, and implementation of the model may well suggest adjustments to the boundary conditions or associated concepts outlined here. As with Missimer et al.'s [43] model of social sustainability more generally, this is "a starting point, expandable and condensable if necessary" (p. 38).

We nevertheless propose that this model provides a useful and accessible starting point for non-technical or substantive experts, and that its alignment with the SDG policy framework can significantly strengthen recognition and policy salience in a public sector context. In particular, we believe the conceptual framework, suggested approaches, and in particular the guiding questions, can make a significant contribution to helping public sector workers understand, anticipate, and manage AI's societal impact. It does so by building on and complementing recent efforts towards operationalization of ethical AI, and particularly the integrative framework asserted by Wirtz et al. which elaborates the processes and layers at which public sector workers must engage to avoid harms caused by AI. To this *what*, the current model provides guidance on *how* public sector should make decisions about AI governance, by elaborating red lines that cannot be crossed if values related to fairness and safety are to be preserved, conceptualized here as social sustainability. This is one of many preliminary steps towards ensuring that the public sector governs AI sustainably.

However, several limitations should be noted. Firstly, and as described above, the literature considering AI and society as relevant to public sector decision making is vast and diffuse, and our review has been deliberate, but not comprehensive. There is much literature which we have not considered and our efforts to narrow the conceptual focus of sustainable AI have closed some doors which might be worth keeping open in other contexts. Notions of AI sustainability more closely linked to environmental or economic concerns, for example, may be important for other policy or research pursuits. We do not see these as mutually exclusive, however, and are convinced that despite these conceptual limitations, this analysis makes an important conceptual and theoretical contribution by providing the foundation for rigorously explicating the notion of sustainable AI in the public sector.

CRediT author statement

Christopher Wilson: conceptualization, methodology, formal analysis, investigation, writing – original draft, writing - review & editing. **Maja Van Der Velden:** conceptualization, methodology, formal analysis, investigation, writing – original draft, writing - review & editing.

References

[1] W.G. de Sousa, E.R.P. de Melo, P.H.D.S. Bermejo, R.A.S. Farias, A.O. Gomes, How and where is artificial intelligence in the public sector going? A literature review and research agenda, Govern. Inf. Q. 36 (2019) 101392, https://doi.org/10.1016/j.giq.2019.07.004.

[2] C. Wilson, Public engagement and AI: a values analysis of national strategies, Govern. Inf. Q. 39 (2022) 101652, https://doi.org/10.1016/j.giq.2021.101652.

[3] V. Galaz, M.A. Centeno, P.W. Callahan, A. Causevic, T. Patterson, I. Brass, S. Baum, D. Farber, J. Fischer, D. Garcia, T. McPhearson, D. Jimenez, B. King, P. Larcey, K. Levy, Artificial intelligence, systemic risks, and sustainability, Technol. Soc. 67 (2021) 101741, https://doi.org/10.1016/j.techsoc.2021.101741.

[4] R.K.E. Bellamy, K. Dey, M. Hind, S.C. Hoffman, S. Houde, K. Kannan, P. Lohia, S. Mehta, A. Mojsilovic, S. Nagar, K.N. Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K.R. Varshney, Y. Zhang, Think your artificial intelligence software is fair? Think again, IEEE Softw. 36 (2019) 76–80, https://doi.org/10.1109/MS.2019.2908514.

[5] C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, L. Floridi, Artificial intelligence and the 'good society': the US, EU, and UK approach, Sci. Eng. Ethics 24 (2018) 505–528, https://doi.org/10.1007/s11948-017-9901-7.

[6] E.M. Witesman, L.C. Walters, Modeling public decision preferences using context-specific value hierarchies, Am. Rev. Publ. Adm. 45 (2015) 86–105, https://doi.org/10.1177/0275074014536603.

[7] L. Reardon, Networks and problem recognition: advancing the multiple streams approach, Pol. Sci. 51 (2018) 457–476, https://doi.org/10.1007/s11077-018-9330-8.

[8] B.W. Wirtz, J.C. Weyerer, B.J. Sturm, The dark sides of artificial intelligence: an integrated AI governance framework for public administration, Int. J. Publ. Adm. 43 (2020) 818–829, https://doi.org/10.1080/01900692.2020.1749851.

[9] A.A. Guenduez, T. Mettler, K. Schedler, Technological frames in public administration: what do public managers think of big data? Govern. Inf. Q. 37 (2020) 101406, https://doi.org/10.1016/j.giq.2019.101406.

[10] M. Janssen, G. Kuk, The challenges and limits of big data algorithms in technocratic governance, Govern. Inf. Q. 33 (2016) 371–377, https://doi.org/10.1016/j.giq.2016.08.011.

[11] D. Kolkman, The usefulness of algorithmic models in policy making, Govern. Inf. Q. 37 (2020) 101488, https://doi.org/10.1016/j.giq.2020.101488.

[12] A. Zuiderwijk, Y.-C. Chen, F. Salem, Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda, Govern. Inf. Q. 38 (2021) 101577, https://doi.org/10.1016/j.giq.2021.101577.

[13] S.J. Mikhaylov, M. Esteve, A. Campion, Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration, Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 376 (2018) 20170357, https://doi.org/10.1098/rsta.2017.0357.

[14] W. Orr, J.L. Davis, Attributions of ethical responsibility by Artificial Intelligence practitioners, Inf. Commun. Soc. 23 (2020) 719–735, https://doi.org/10.1080/1369118X.2020.1713842.

[15] P. 6, Ethics, regulation and the new artificial Intelligence, Part I: accountability and power, Inf. Commun. Soc. 4 (2001) 199–229, https://doi.org/10.1080/713768525.

[16] A. Gupta, V. Heath, AI Ethics Groups Are Repeating One of Society's Classic Mistakes, MIT Technol. Rev., 2020. https://www.technologyreview.com/2020/09/14/1008323/ai-ethics-representation-artificial-intelligence-opinion/. (Accessed 28 January 2022).

[17] T. Metzinger, Ethics Washing Made in Europe, Tagesspiegel Online, 2019. https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html. (Accessed 28 January 2022).

[18] B. Rossi, Why the Government's Data Science Ethical Framework Is a Recipe for Disaster, Inf. Age, 2016. https://www.information-age.com/why-governments-data-science-ethical-framework-recipe-disaster-123461541/. (Accessed 28 January 2022).

[19] L. Vesnic-Alujevic, S. Nascimento, A. Pólvora, Societal and ethical impacts of artificial intelligence: critical notes on European policy frameworks, Telecommun. Pol. 44 (2020) 101961, https://doi.org/10.1016/j.telpol.2020.101961.

[20] J. Cussins Newman, Decision Points in AI Governance: Three Case Studies Explore Efforts to Operationalize AI Principles, CLTC UC Berkeley Center for Long-Term Cybersecurity, Berkeley, CA, 2020. https://cltc.berkeley.edu/ai-decision-points/. (Accessed 28 January 2022).

[21] D. Reisman, S. Schultz, K. Crawford, M. Whittaker, Algorithmic Impact Assessment: A Practical Framework for Public Agency Accountability, AI Now Institute, 2018. https://ainowinstitute.org/aiareport2018.pdf.

[22] R.A. Irvin, J. Stansbury, Citizen participation in decision making: is it worth the effort? Publ. Adm. Rev. 64 (2004) 55–65, https://doi.org/10.1111/j.1540-6210.2004.00346.x.

[23] J. Haas, K.M. Vogt, Ignorance and investigation, in: Routledge Int. Handb. Ignorance Stud., Routledge, 2015.

[24] B.W. Wirtz, R. Piehler, M.-J. Thomas, P. Daiser, Resistance of Public Personnel to Open Government: a cognitive theory view of implementation barriers towards open government data, Publ. Manag. Rev. 18 (2016) 1335–1364, https://doi.org/10.1080/14719037.2015.1103889.

[25] E. Thelisson, J.-H. Morin, J. Rochel, AI governance: digital responsibility as a building block, Delphi - interdiscip, Rev. Emerg. Technol. 2 (2020) 167–178, https://doi.org/10.21552/delphi/2019/4/6.

[26] S.C. Robinson, Trust, transparency, and openness: how inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI), Technol. Soc. (2020) 101421, https://doi.org/10.1016/j.techsoc.2020.101421.

[27] O.J. Erdelyi, J. Goldsmith, Regulating Artificial Intelligence: Proposal for a Global Solution, Social Science Research Network, Rochester, NY, 2018. https://papers.ssrn.com/abstract=3263992. (Accessed 1 February 2022).

[28] J. Berryhill, K.K. Heang, R. Clogher, K. McBride, Hello, World: Artificial Intelligence and its Use in the Public Sector, OECD, 2019, https://doi.org/10.1787/726fd39d-en.

[29] J. Reis, P.E. Santo, N. Melão, Artificial intelligence in government services: a systematic literature review, in: Á. Rocha, H. Adeli, L.P. Reis, S. Costanzo (Eds.), New Knowl. Inf. Syst. Technol., Springer International Publishing, Cham, 2019, pp. 241–252, https://doi.org/10.1007/978-3-030-16181-1_23.

[30] J. Eager, M. Whittle, J. Smit, G. Cacciaguerra, E. Lale-Demoz, Opportunities of Artificial Intelligence, Think Thank European Parliament, Brussels, 2020. https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2020)652713. (Accessed 1 February 2022).

[31] A. Ingrams, W. Kaufmann, D. Jacobs, In AI we trust? Citizen perceptions of AI in government decision making, Pol. Internet (2021) 1–20, https://doi.org/10.1002/poi3.276.

[32] P.D. König, G. Wenzelburger, The legitimacy gap of algorithmic decision-making in the public sector: why it arises and how to address it, Technol. Soc. 67 (2021), https://doi.org/10.1016/j.techsoc.2021.101688.

[33] K. Yeung, M. Lodge, Algorithmic regulation, in: K. Yeung, M. Lodge (Eds.), Algorithmic Regul., Oxford University Press, Oxford, 2019, pp. 1–18, https://doi.org/10.1093/oso/9780198838494.003.0001.

[34] World Commission on Environment and Development, Report of the World Commission on Environment and Development: Our Common Future, 1987, https://doi.org/10.1080/07488008808408783.

[35] S. McKenzie, Social Sustainability: towards some definitions, Hawke Res. Inst. Work. Pap. Ser. (2004) 31.

[36] United Nations, Transforming Our World: the 2030 Agenda for Sustainable Development, 2015, https://doi.org/10.1201/b20466-7.

[37] R. Saner, L. Yiu, M. Nguyen, Monitoring the SDGs: digital and social technologies to ensure citizen participation, inclusiveness and transparency, Dev. Pol. Rev. (2019) 1–18, https://doi.org/10.1111/dpr.12433.

[38] United Nations Development Programme, Institutional and Coordination Mechanisms: Guidance Note on Facilitating Integration and Coherence for SDG Implementation, 2017.

[39] O.M. Van Den Broek, R. Klingler-vidra, The UN Sustainable Development Goals as a North Star : How an Intermediary Network Makes , Takes , and Retro Fi Ts the Meaning of the Sustainable Development Goals, Regulation and, 2021, https://doi.org/10.1111/rego.12415.

[40] United Nations Department for Economic and Social Affairs, Compendium of National Institutional Arrangements for Implementing the 2030 Agenda for Sustainable Development, United Nations Department for Economic and Social Affairs, New York, 2019. https://sustainabledevelopment.un.org/content/documents/22008UNPAN99132.pdf.

[41] G.I. Broman, K.H. Robèrt, A framework for strategic sustainable development, J. Clean. Prod. 140 (2017) 17–31, https://doi.org/10.1016/j.jclepro.2015.10.121.

[42] E.S. Zeemering, Sustainability management , strategy and reform in local government, Publ. Manag. Rev. 20 (2018) 136–153, https://doi.org/10.1080/14719037.2017.1293148.

[43] M. Missimer, K.-H. Robèrt, G. Broman, A strategic approach to social sustainability – Part 1: exploring the social system, J. Clean. Prod. 140 (2017) 32–41, https://doi.org/10.1016/j.jclepro.2016.03.170.

[44] R. Messerschmidt, S. Ullrich, A European Way towards Sustainable AI, Soc. Eur., 2020. https://www.socialeurope.eu/a-european-way-towards-sustainable-ai. (Accessed 5 June 2020).

[45] A. Gupta, The Imperative for Sustainable AI Systems, the Gradient, 2021. https://thegradient.pub/sustainable-ai/. (Accessed 31 December 2021).

[46] M. Chavosh Nejad, S. Mansour, A. Karamipour, An AHP-based multi-criteria model for assessment of the social sustainability of technology management process: a case study in banking industry, Technol. Soc. 65 (2021) 101602, https://doi.org/10.1016/j.techsoc.2021.101602.

[47] G. Myers, K. Nejkov, Developing Artificial Intelligence Sustainably: toward a Practical Code of Conduct for Disruptive Technologies, International Finance Corporation, Washington, DC, 2020, https://doi.org/10.1596/33613.

[48] N. Aliman, L. Kester, P. Werkhoven, Sustainable AI safety? Delphi - interdiscip. Rev. Emerg. Technol. 2 (2020) 226–233, https://doi.org/10.21552/delphi/2019/4/12.

[49] A. van Wynsberghe, Sustainable AI: AI for sustainability and the sustainability of AI, AI Ethics 1 (2021) 213–218, https://doi.org/10.1007/s43681-021-00043-6.

[50] F. Rohde, M. Gossen, J. Wagner, T. Santarius, Sustainability challenges of artificial intelligence and policy implications, Ökol. Wirtsch. - Fachz. 36 (2021) 36–40, https://doi.org/10.14512/OEWO360136.

[51] S. Larsson, M. Anneroth, A. Felländer, F. Heintz, R.C. Ångström, Sustainable AI: an Inventory of the State of Knowledge of Ethical, Social, and Legal Challenges Related to Artificial Intelligence, AI Sustainability Center, Stockholm, 2019.

[52] AI Sustainability Center, AI Sustainability Center, 2021. https://aisustainability.org.

[53] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S.D. Langhans, M. Tegmark, F. Fuso Nerini, The role of artificial intelligence in achieving the Sustainable Development Goals, Nat. Commun. 11 (2020) 233, https://doi.org/10.1038/s41467-019-14108-y.

[54] L. Bjørlo, Ø. Moen, M. Pasquine, The role of consumer autonomy in developing sustainable AI: a conceptual framework, Sustainability 13 (2021) 2332, https://doi.org/10.3390/su13042332.

[55] J.J. Yun, D. Lee, H. Ahn, K. Park, T. Yigitcanlar, Not deep learning but autonomous learning of open innovation for sustainable artificial intelligence, Sustain. Switz. 8 (2016), https://doi.org/10.3390/su8080797.

[56] G.L. Tsafack Chetsa, Towards Sustainable Artificial Intelligence: A Framework to Create Value and Understand Risk, Apress, Berkeley, CA, 2021, https://doi.org/10.1007/978-1-4842-7214-5.
[57] C.-J. Wu, R. Raghavendra, U. Gupta, B. Acun, N. Ardalani, K. Maeng, G. Chang, F. A. Behram, J. Huang, C. Bai, M. Gschwind, A. Gupta, M. Ott, A. Melnikov, S. Candido, D. Brooks, G. Chauhan, B. Lee, H.-H.S. Lee, B. Akyildiz, M. Balandat, J. Spisak, R. Jain, M. Rabbat, K. Hazelwood, Sustainable AI: Environmental Implications, Challenges and Opportunities, ArXiv 211100364 Cs, 2021. http://arxiv.org/abs/2111.00364. (Accessed 19 December 2021).
[58] C. Djeffal, Sustainable AI Development (SAID): on the Road to More Access to Justice, Social Science Research Network, Rochester, NY, 2018, https://doi.org/10.2139/ssrn.3298980.
[59] K. Porter, Shaping the future of sustainable AI and automation: why human rights still matter, Hum. Rights Def. 28 (2019) 33–35.
[60] E. Dahlin, Mind the gap! on the future of AI research, Humanit. Soc. Sci. Commun. 8 (2021) 1–4, https://doi.org/10.1057/s41599-021-00750-9.
[61] S.-C. Yeh, A.-W. Wu, H.-C. Yu, H.C. Wu, Y.-P. Kuo, P.-X. Chen, Public perception of artificial intelligence and its connections to the sustainable development goals, Sustainability 13 (2021) 9165, https://doi.org/10.3390/su13169165.
[62] C. Fernández-Aller, A.F. de Velasco, Á. Manjarrés, D. Pastor-Escuredo, S. Pickin, J.S. Criado, T. Ausín, An inclusive and sustainable artificial intelligence strategy for Europe based on human rights, IEEE Technol. Soc. Mag. 40 (2021) 46–54, https://doi.org/10.1109/MTS.2021.3056283.
[63] N.R. Pal, In search of trustworthy and transparent intelligent systems with human-like cognitive and reasoning capabilities, Front. Robot. AI. 7 (2020), https://doi.org/10.3389/frobt.2020.00076.
[64] I. Kindylidi, T.S. Cabral, Sustainability of AI: the case of provision of information to consumers, Sustainability 13 (2021) 12064, https://doi.org/10.3390/su132112064.
[65] OsloMet, Nordic Center for Sustainable and Trustworthy AI Research (NordSTAR), 2021. https://www.oslomet.no/nordstar. (Accessed 28 January 2022).
[66] University of Bonn, Sustainable AI Lab, Sustain. AI Lab, 2021. https://sustainable-ai.eu/. (Accessed 28 January 2022).
[67] J. Elkington, Cannibals with Forks: the Triple Bottom Line of 21st Century Business, Capstone, Oxford, 1997.
[68] E.B. Barbier, The concept of sustainable economic development, Environ. Conserv. 14 (1987) 101–110, https://doi.org/10.1017/S0376892900011449.
[69] A. Colantonio, Social Sustainability: an Exploratory Analysis of its Definition, Assessment Methods Metrics and Tools, Oxford Brooks University, Oxford, UK, 2007. http://www.brookes.ac.uk/schools/be/oisd/sustainable_communities/. (Accessed 12 June 2020).
[70] Ş.Y. Balaman, Chapter 4 - sustainability issues in biomass-based production chains, in: Ş.Y. Balaman (Ed.), Decis.-Mak. Biomass-Based Prod. Chains, Academic Press, 2019, pp. 77–112, https://doi.org/10.1016/B978-0-12-814278-3.00004-2.
[71] L. Karbasi, Social Sustainability | UN Global Compact, (n.d.). https://www.unglobalcompact.org/what-is-gc/our-work/social (accessed June 14, 2020).
[72] S. Woodcraft, Design for Social Sustainability: A Framework for Creating Thriving New Communities, The Young Foundation, London, UK, 2012.
[73] A. Widok, Social Sustainability: Theories, Concepts, Practicability, in: Berlin, 2009, p. 9.
[74] G. Assefa, B. Frostell, Social sustainability and social acceptance in technology assessment: a case study of energy technologies, Technol. Soc. 29 (2007) 63–78, https://doi.org/10.1016/j.techsoc.2006.10.007.
[75] K. De Fine Licht, A. Folland, Defining "social sustainability": towards a sustainable solution to the conceptual confusion, etikk praksis - nord, J. Appl. Ethics. (2019) 21–39, https://doi.org/10.5324/eip.v13i2.2913.
[76] M. Cuthill, Strengthening the 'social' in sustainable development: developing a conceptual framework for social sustainability in a rapid urban growth region in Australia, Sustain. Dev. 18 (2010) 362–373, https://doi.org/10.1002/sd.397.
[77] M. Missimer, K.-H. Robèrt, G. Broman, A strategic approach to social sustainability – Part 2: a principle-based definition, J. Clean. Prod. 140 (2017) 42–52, https://doi.org/10.1016/j.jclepro.2016.04.059.
[78] V. Dignum, Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way, Springer Nature, 2019.
[79] P. Alston, Digital Technology, Social Protection and Human Rights, OHCHR, Geneva, 2019. https://www.ohchr.org/EN/Issues/Poverty/Pages/DigitalTechnology.aspx. (Accessed 21 June 2020).
[80] J. Niklas, Conceptualizing Socio-Economic Rights in the Discussion on Artificial Intelligence, Social Science Research Network, Rochester, NY, 2019, https://doi.org/10.2139/ssrn.3569780.
[81] J. von Braun, AI and Robotics Implications for the Poor, Social Science Research Network, Rochester, NY, 2019, https://doi.org/10.2139/ssrn.3497591.
[82] UNESCO, Steering AI and Advanced ICTs for Knowledge Societies: a Rights, Openness, Access, and Multi-Stakeholder Perspective, UNESCO Digital Library, 2018. https://unesdoc.unesco.org/ark:/48223/pf0000372132. (Accessed 22 June 2020).
[83] G. Mulgan, A Machine Intelligence Commission for the UK: How to Grow Informed Public Trust and Maximise the Positive Impact of Smart Machines, 2016. https://media.nesta.org.uk/documents/a_machine_intelligence_commission_for_the_uk_-_geoff_mulgan.pdf.
[84] A. Hintz, Towards Civic Participation in the Datafied Society: can citizen assemblies democratize algorithmic governance? AoIR Sel. Pap. Internet Res. (2021) https://doi.org/10.5210/spir.v2021i0.11943.
[85] The Forum for Ethical AI, Democratising Decisions about Technology: A Toolkit, RSA, London, 2019.
[86] B. Balaram, T. Greenham, J. Leonard, Engaging Citizens in the Ethical Use of AI for Automated Decision-Making, RSA, London, 2018. https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial- intelligence—real-public-engagement.pdf.
[87] J. Anderson, L. Rainie, Artificial intelligence and the future of humans, Pew Res. Cent. Internet Sci. Tech. (2018). https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/. (Accessed 28 January 2022).
[88] D.B. Shank, A. DeSanti, T. Maninger, When are artificial intelligence versus human agents faulted for wrongdoing? Moral attributions after individual and joint decisions, Inf. Commun. Soc. 22 (2019) 648–663, https://doi.org/10.1080/1369118X.2019.1568515.
[89] T. Bucher, Neither black nor box: ways of knowing algorithms, in: S. Kubitschko, A. Kaun (Eds.), Innov. Methods Media Commun. Res., Springer International Publishing, Cham, 2016, pp. 81–98, https://doi.org/10.1007/978-3-319-40700-5_5.
[90] B. Goodman, S. Flaxman, European union regulations on algorithmic decision-making and a "right to explanation, AI Mag. 38 (2017) 50–57, https://doi.org/10.1609/aimag.v38i3.2741.
[91] A. Abdul, J. Vermeulen, D. Wang, B.Y. Lim, M. Kankanhalli, Trends and trajectories for explainable, accountable and intelligible systems: an HCI research agenda, in: Proc. 2018 CHI Conf. Hum. Factors Comput. Syst., Association for Computing Machinery, New York, NY, USA, 2018, pp. 1–18, https://doi.org/10.1145/3173574.3174156. (Accessed 28 January 2022).
[92] C.T. Wolf, K.E. Ringland, Designing accessible, explainable AI (XAI) experiences, ACM SIGACCESS Access, Comput. Times 6 (2020) 1, https://doi.org/10.1145/3386296.3386302.
[93] J. Kemper, D. Kolkman, Transparent to whom? No algorithmic accountability without a critical audience, Inf. Commun. Soc. 22 (2019) 2081–2096, https://doi.org/10.1080/1369118X.2018.1477967.
[94] A. Vestby, J. Vestby, Machine learning and the police: asking the right questions, Polic. J. Pol. Pract. 15 (2021) 44–58, https://doi.org/10.1093/police/paz035.
[95] S. Robbins, AI and the path to envelopment: knowledge as a first step towards the responsible regulation and use of AI-powered machines, AI Soc. 35 (2020) 391–400, https://doi.org/10.1007/s00146-019-00891-1.
[96] P.D. König, G. Wenzelburger, Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics, Govern. Inf. Q. 37 (2020) 101489, https://doi.org/10.1016/j.giq.2020.101489.
[97] J. Berscheid, F. Roewer-Despres, Beyond transparency, AI Matters 5 (2019) 13–22, https://doi.org/10.1145/3340470.3340476.
[98] M. Janssen, P. Brous, E. Estevez, L.S. Barbosa, T. Janowski, Data governance: organizing data for trustworthy artificial intelligence, Govern. Inf. Q. 37 (2020) 101493, https://doi.org/10.1016/j.giq.2020.101493.
[99] R. Mcgee, R. Carlitz, Learning Study on the Users in Technology for Transparency and Accountability Initiatives: Assumptions and Realities, 2013.
[100] V. Chiao, Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice, Int. J. Law Context 15 (2019) 126–139, https://doi.org/10.1017/S1744552319000077.
[101] V. Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor, 2018.
[102] M.L. Jones, E. Edenberg, Troubleshooting AI and consent, in: M.D. Dubber, F. Pasquale, S. Das (Eds.), Oxf. Handb. Ethics AI, Oxford University Press, Oxford, UK, 2020, pp. 357–374, https://doi.org/10.1093/oxfordhb/9780190067397.013.23.
[103] M. Latonero, Governing Artificial Intelligence: Upholding Human Rights Dignity, Data & Society, 2018. https://apo.org.au/sites/default/files/resource-files/2018-10/apo-nid196716.pdf.
[104] J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson, H. Yu, Accountable Algorithms, Social Science Research Network, Rochester, NY, 2016. https://papers.ssrn.com/abstract=2765268. (Accessed 28 January 2022).
[105] D. Neyland, Accountability and the algorithm, in: D. Neyland (Ed.), Everyday Life Algorithm, Springer International Publishing, Cham, 2019, pp. 45–71, https://doi.org/10.1007/978-3-030-00578-8_3.
[106] A.J. Andreotta, N. Kirkham, M. Rizzi, AI, Big Data, and the Future of Consent, AI Soc., 2021, pp. 1–14, https://doi.org/10.1007/s00146-021-01262-5.
[107] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations, Minds Mach. 28 (2018) 689–707, https://doi.org/10.1007/s11023-018-9482-5.
[108] H. Surden, Values Embedded in Legal Artificial Intelligence, Social Science Research Network, Rochester, NY, 2017, https://doi.org/10.2139/ssrn.2932333.
[109] UNESCO, Humanistic Futures of Learning: Perspectives from UNESCO Chairs and UNITWIN Networks, UNESCO, Paris, France, 2020. https://unesdoc.unesco.org/ark:/48223/pf0000372577. (Accessed 27 January 2022).
[110] Q.V. Liao, M. Muller, Enabling value sensitive AI systems through participatory design fictions, Preprint (2019) 7.
[111] H. Zhu, B. Yu, A. Halfaker, L. Terveen, Value-sensitive algorithm design: method, case study, and lessons, Proc. ACM Hum.-Comput. Interact. 2 (2018), https://doi.org/10.1145/3274463, 194:1-194:23.
[112] D. Loi, T. Lodato, C.T. Wolf, R. Arar, J. Blomberg, PD manifesto for AI futures, in: Proc. 15th Particip. Des. Conf. Short Pap. Situated Actions Workshop Tutor, vol. 2, Association for Computing Machinery, New York, NY, USA, 2018, pp. 1–4, https://doi.org/10.1145/3210604.3210614.

[113] P. Vamplew, R. Dazeley, C. Foale, S. Firmin, J. Mummery, Human-aligned artificial intelligence is a multiobjective problem, Ethics Inf. Technol. 20 (2018) 27–40, https://doi.org/10.1007/s10676-017-9440-6.
[114] S. Das, A. Dey, A. Pal, N. Roy, Applications of artificial intelligence in machine learning: review and prospect, Int. J. Comput. Appl. 115 (2015) 31–41.
[115] Z. Ghahramani, Probabilistic machine learning and artificial intelligence, Nature 521 (2015) 452–459, https://doi.org/10.1038/nature14541.
[116] D. Ensign, S.A. Friedler, S. Neville, C. Scheidegger, S. Venkatasubramanian, Runaway Feedback Loops in Predictive Policing, ArXiv170609847 Cs Stat, 2017. http://arxiv.org/abs/1706.09847. (Accessed 28 January 2022).
[117] S. Milano, M. Taddeo, L. Floridi, Recommender systems and their ethical challenges, AI Soc. 35 (2020) 957–967, https://doi.org/10.1007/s00146-020-00950-y.
[118] I. Rahwan, Society-in-the-loop: programming the algorithmic social contract, Ethics Inf. Technol. 20 (2018) 5–14, https://doi.org/10.1007/s10676-017-9430-8.
[119] N. Tomašev, J. Cornebise, F. Hutter, S. Mohamed, A. Picciariello, B. Connelly, D. C.M. Belgrave, D. Ezer, F.C. van der Haert, F. Mugisha, G. Abila, H. Arai, H. Almiraat, J. Proskurnia, K. Snyder, M. Otake-Matsuura, M. Othman, T. Glasmachers, W. de Wever, Y.W. Teh, M.E. Khan, R.D. Winne, T. Schaul, C. Clopath, AI for social good: unlocking the opportunity for positive impact, Nat. Commun. 11 (2020) 2468, https://doi.org/10.1038/s41467-020-15871-z.
[120] T. Harrison, L.F. Luna-Reyes, T. Pardo, N. De Paula, M. Najafabadi, J. Palmer, The data firehose and AI in government: why data management is a key to value and ethics, in: Proc. 20th Annu. Int. Conf. Digit. Gov. Res., Association for Computing Machinery, New York, NY, USA, 2019, pp. 171–176, https://doi.org/10.1145/3325112.3325245.
[121] European Commission, White Paper on Artificial Intelligence: a European Approach to Excellence and Trust, European Commission, Brussels, 2020. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en. (Accessed 28 January 2022).
[122] F. Bannister, R. Connolly, Trust and transformational government: a proposed framework for research, Govern. Inf. Q. 28 (2011) 137–147, https://doi.org/10.1016/j.giq.2010.06.010.
[123] M. Chui, M. Harryson, J. Manyika, R. Roberts |, R. Chung, A. van Heteren, P. Nel, Applying AI for Social Good, McKinsey, San Fransisco, 2018. https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good. (Accessed 28 January 2022).
[124] P. Vassilakopoulou, Sociotechnical Approach for Accountability by Design in AI Systems, 2020. ECIS 2020 Res.–Prog. Pap, https://aisel.aisnet.org/ecis2020_rip/12.
[125] K.S. Gill, AI&Society: editorial volume 35.2: the trappings of AI Agency, AI Soc. 35 (2020) 289–296, https://doi.org/10.1007/s00146-020-00961-9.
[126] B. Bodó, Mediated Trust: A Theoretical Framework to Address the Trustworthiness of Technological Trust Mediators, vol. 23, New Media Soc, 2021, pp. 2668–2690, https://doi.org/10.1177/1461444820939922.
[127] R. Steedman, H. Kennedy, R. Jones, Complex ecologies of trust in data practices and data-driven systems, Inf. Commun. Soc. 4462 (2020), https://doi.org/10.1080/1369118X.2020.1748090.
[128] M. Brundage, S. Avin, J. Wang, H. Belfield, G. Krueger, G. Hadfield, H. Khlaaf, J. Yang, H. Toner, R. Fong, T. Maharaj, P.W. Koh, S. Hooker, J. Leung, A. Trask, E. Bluemke, J. Lebensold, C. O'Keefe, M. Koren, T. Ryffel, J.B. Rubinovitz, T. Besiroglu, F. Carugati, J. Clark, P. Eckersley, S. de Haas, M. Johnson, B. Laurie, A. Ingerman, I. Krawczuk, A. Askell, R. Cammarota, A. Lohn, D. Krueger, C. Stix, P. Henderson, L. Graham, C. Prunkl, B. Martin, E. Seger, N. Zilberman, S.Ó. hÉigeartaigh, F. Kroeger, G. Sastry, R. Kagan, A. Weller, B. Tse, E. Barnes, A. Dafoe, P. Scharre, A. Herbert-Voss, M. Rasser, S. Sodhani, C. Flynn, T. K. Gilbert, L. Dyer, S. Khan, Y. Bengio, M. Anderljung, Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims, 2020. ArXiv200407213 Cs, http://arxiv.org/abs/2004.07213. (Accessed 1 February 2022).
[129] C. Bourne, AI cheerleaders: public relations, neoliberalism and artificial intelligence, Publ. Relat. Inq. 8 (2019) 109–125, https://doi.org/10.1177/2046147X19835250.
[130] S.H. Chaffee, Explication (Communication Concepts), Sage Publications, London, 1991.
[131] G. Goertz, Social Science Concepts: A User's Guide, Princeton University Press, 2006, https://doi.org/10.2307/j.ctvcm4gmg.
[132] A.L. George, A. Bennett, Case Studies and Theory Development in the Social Sciences, MIT Press, Cambridge, MA, USA, 2005.
[133] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations, Minds Mach. 28 (2018) 689–707, https://doi.org/10.1007/s11023-018-9482-5.