Interfacing Artificial Intelligence for Social Good (AI4SG) and Relational AI Ethics: A Systematic Literature Review

Cheshta Arora[1†], Debarun Sarkar[2*†]

[1] *Independent Researcher, Norway*
[2] *Independent Researcher, India*

Abstract

The paper is a systematic literature review of the emerging field of AI4SG and relational AI ethics. Since these two fields have not interacted with each other, the paper aims to lay the foundation for such an interface. The paper argues that such an interface is exigent because AI4SG is emerging as a normative field of AI ethics. Interfacing with existing literature on relational ethics can help interrogate the normative core of AI4SG and open up both the 'social' and the 'technical' as more than instrumental concerns.

Keywords

AI4SG, AI ethics, relational, systematic literature review, technology

1. Introduction

The paper aims to bring into conversation two fields of AI ethics through a systematic literature review of the two fields, that of AI4SG and the relational ethics approach to AI.

The gold rush in the last few years concerning ethical AI [7] has been a result of an exponential increase in the use of AI across diverse domains. The growth and proliferation of AI systems and machine learning methodologies across various domains and fields have led to the rise of the field of AI ethics. The growth of the field has also been aided by the increase in funding emerging from the big tech [79]. However, despite several ethical frameworks, guidelines and consultations from both private, public as well as civil society organizations, there seems to be little consensus about "what constitutes 'ethical AI' and which ethical requirements, technical standards and best practices are needed for its realization" [62:1]. In an attempt to rescue a conflicted field, scholars have moved toward understanding the normative core of these frameworks and the parameters around which they converge. For example, Fjeld et al. [33] identified the following eight key parameters: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values.

While ethical debates in the field of AI ethics have flourished in recent years at a growing speed, the field of AI4SG is slowly emerging as a justificatory ground for developing AI. As the paper notes in a later section, the AI4SG discourse has been funded and facilitated by powerful actors such as United Nations, international non-governmental organizations, consultancy companies and big tech. Against this backdrop, critical interrogation of the very notion of AI4SG is exigent. While AI4SG privileges normative ethical approaches, the field of AI ethics on the other hand, while moving towards convergences, is still open to contestation, with relational ethics offering the most radical critique to normative ethical approaches that posit a transcendentalist and/or vertical systems of morality and fall back on a problem/solution framework. Relational ethical approaches, on the other hand, pivot around ontological heterogenesis [14], horizontal relations, and situated and partial perspectives [50] in which moral systems are emergent and processual [92].

Tethics2023: Conference on Technology Ethics, October 18–19, 2023, Turku, Finland
*Corresponding author.
†These authors contributed equally.
EMAIL: cheshtaarora@outlook.in (C. Arora); debarun@outlook.com (D. Sarkar)
ORCID: 0000-0003-2470-7783 (C. Arora); 0000-0002-6873-4727 (D. Sarkar)
© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3582/FP_06.pdf

It is in this light that the paper does a systematic literature review of AI4SG and relational AI ethics to understand how and if at all they interface. The key finding of the paper is that there has been no interface between these two fields in major peer-reviewed publications. This review aims to lay the ground for the future interface between the two domains.

Following the introduction, a brief review of existing surveys is provided, and the methodology of the paper is elaborated. The fourth section presents a review and commentary on the AI4SG discourse and the fifth section reviews the field of relational AI ethics. A brief conclusion, limitations, and need for future work are noted.

2. Existing Literature

Previous surveys of AI4SG, while larger in scope, did not conduct a systematic literature review of the field of AI4SG [91]. This can be broadly attributed to two tendencies. First, in such works, a large-scale survey of AI literature was conducted which was then filtered through the lens of 'social good'. Second, these surveys review workshops in conferences or projects to champion such works under the moniker of AI4SG. In such works, the category of AI4SG is not interrogated as constructed and produced but rather the focus is on establishing the field of AI4SG.

We depart from such approaches for a very simple reason. AI4SG is not a category or approach which can be given a hall pass in the debates of AI ethics because of their claims of 'social good'. Instead, unpacking the very category of 'social' and 'good' is necessary. It is important to note that in the discipline of sociology, the very category of 'social' is contestable and no single consensus exists. In recent years, the 'social' has emerged as a domain with a multiplicity of connections that spans human and non-human domains [64]. This fundamentally open view of the 'social' means that political consensus remains a myth that perpetuates hegemonic discourses and ideas [75]. The field of AI4SG risks emerging as a narrow and dumbed-down field of AI ethics which reproduces and accepts United Nations Sustainability Development Goals and various common-sensical ethical notions without rigorous interrogation of their social & political foreclosures.

On the other hand, while relational ethics has emerged in certain subdomains of AI ethics literature more strongly than others, such as in the case of robot ethics and AI in medical domains, relational ethical approaches remain relatively marginal because they are difficult to operationalize within solutionist fields such as engineering, policy-making and governance. At the time of writing, we didn't find any comprehensive review of the field of relational AI ethics.

3. Methodology

To conduct the systematic literature review, we referred to Scopus and Web of Science. Limiting the review to these two indexes has definite shortcomings as they are heavily skewed towards English language publications and don't include a wide variety of open access, pre-print and experimental publications. Yet, limiting ourselves to these two indexes also allows us to map the state of the field in relatively high-impact journals and conference proceedings.

The paper also refers to certain developments in broad fields of social sciences to establish links and conversations which would otherwise not be possible through a traditional systematic literature review. For example, while AI4SG invokes the concept of 'social', the discourse on AI4SG has largely bypassed discussions on 'social' that have transpired over the past two decades in sociology. As anthropologists/sociologists reviewing the field of AI4SG, we cannot help but bring these blatantly neglected discussions to light. The discourse of AI4SG also invokes uncritically the United Nations (UN) Sustainable Development Goals (SDGs). Hence, to situate such arbitrary external relationships it was necessary to situate such invocations alongside pre-existing critiques of such frameworks to provide the reader with a balanced view of a relatively nascent field.

For AI4SG, two keywords '({artificial intelligence for social good} OR {artificial intelligence for good} OR {AI4SG} OR {AI4G})' were queried (see Figure 1 for PRISMA flow [78] and Table 1 for an overview of the documents). Use of the short form was made to precisely map the emerging consensus in the field as terms such as 'good' and 'social good' are often used by engineering and

technology papers in a generic non-descript manner. This remains a limitation of the chosen methodology but also adds much-needed precision as the aim is to narrow down the emerging field of AI4SG and how the field of AI ethics is coagulating around this term. After the retrieval of 29 documents from the indexes, the selected documents were reduced to 17 after the removal of duplicates and irrelevant documents. Through citation searching, 2 non-peer-reviewed documents are also referred to in the narrative of AI4SG due to their significance to the field [47,91].

For the second part of the review, the search phrase '{AI OR {artificial AND intelligence}} AND relational AND ethics' was used to spread the net as wide as possible (see Figure 2 for PRISMA flow and Table 2 for an overview of the documents). While the field of AI ethics has grown over the past few years, relational approaches in AI ethics are still a relatively small field with higher activity in the domains of health and robot rights and ethics. Five important publications in the field that the index searches did not throw up were reviewed manually and are referred to in the narrative of the paper for their significant relevance to the field of relational ethics and artificial intelligence [3,20,22,43,45]. In both cases, the PRISMA chart and the tabular classification of the literature only refer to the documents retrieved from databases for reproducibility. The final database retrieval for the first part was done on 14th May 2023 and the second part was done on 20th January 2023.

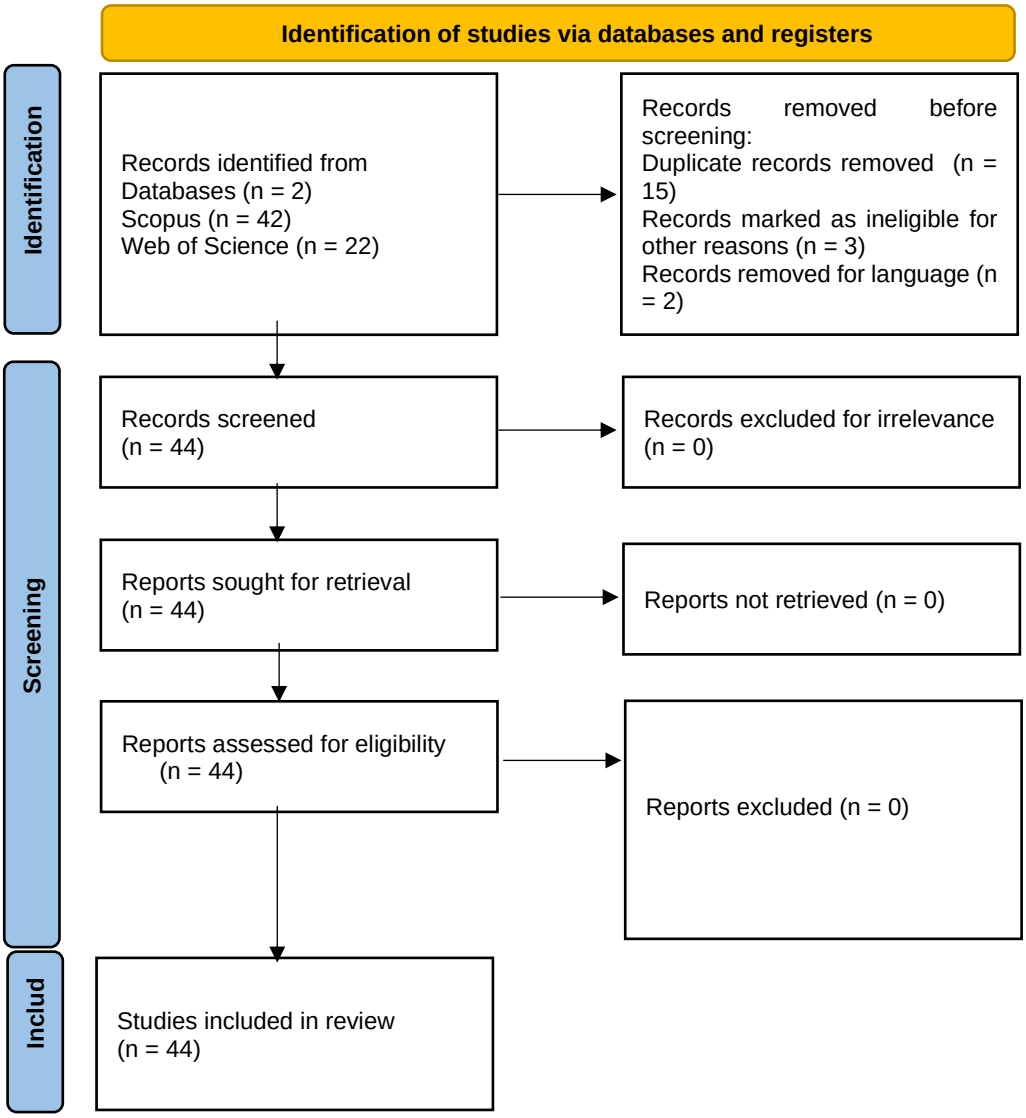

Figure 1: PRISMA flow chart for {artificial intelligence for social good} OR {artificial intelligence for good} OR {AI4SG} OR {AI4G}

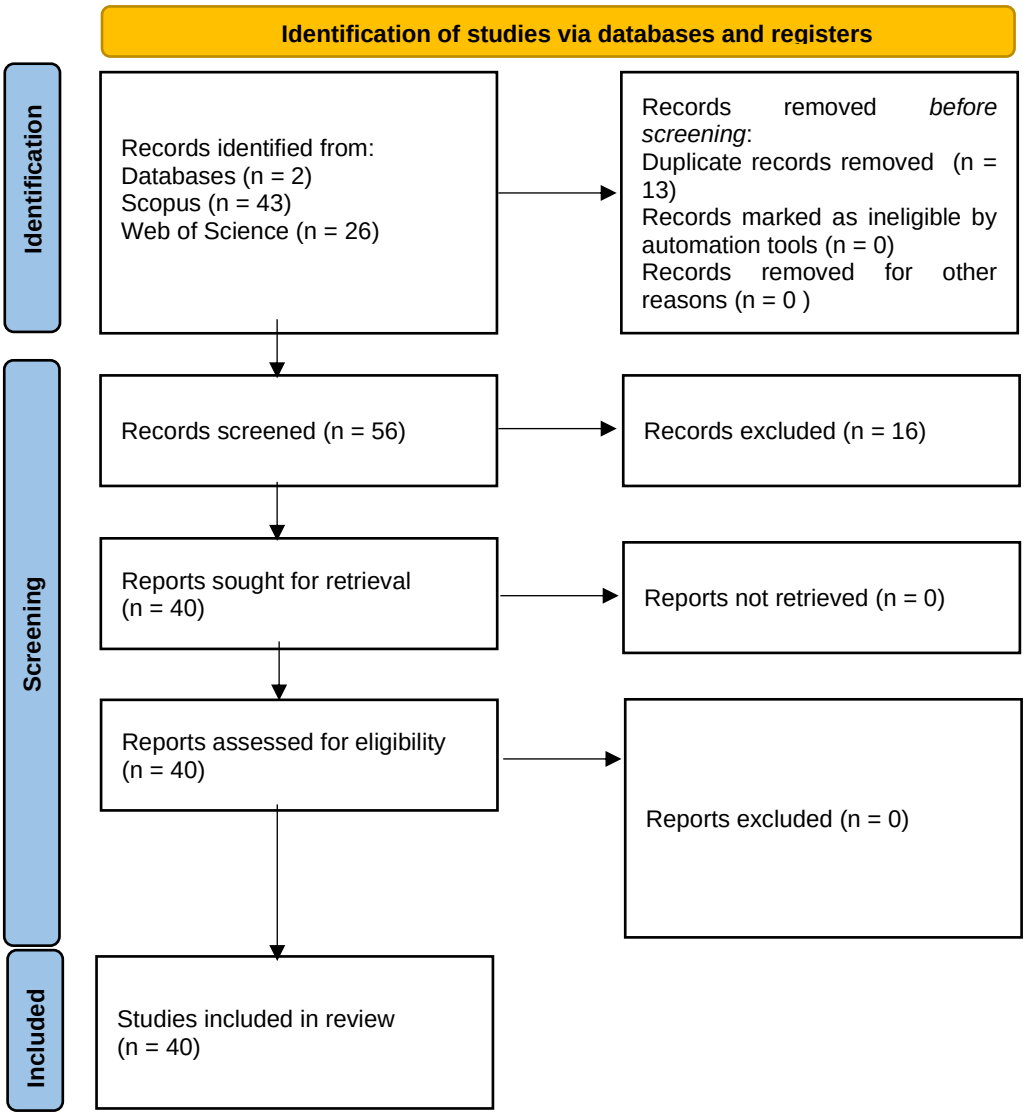

Figure 2: PRISMA flow chart for {AI OR {artificial AND intelligence}} AND relational AND ethics

4. AI4SG

4.1. What is AI4SG?

AI4SG (and other ancillary notions such as AI4G) has emerged in the past few years to distinguish the usage of AI from "malicious use, e.g. DeepNude, or for highly dubious purposes" [91:3]. One of the earliest celebrations of AI4SG was in a workshop report published by Hager et al. [47]. The workshops were co-sponsored by the "Computing Community Consortium (CCC), along with the White House Office of Science and Technology Policy (OSTP), and the Association for the Advancement of Artificial Intelligence" [47:1]. The report began with an explicit announcement that "AI can be a major force for social good" [47:1] and went on to make its case by noting "Social Issues AI can help address" listing out "Justice", "Economic Development", "Workforce Development", "Public Safety", "Policing" and "Education" and also noted "Success Stories and Case Studies" including "Public Health", "Education", "Public Safety" and "Economic Development".

Though Hager et al. [47] is a non-peer-reviewed report, it is important to begin the narrative of AI4SG from it to highlight the already entrenched acceptance of AI4SG discourse in AI discourse. A common-sensical acceptance of the 'social' as a coherent body to be governed, managed and in turn engineered for a hegemonic notion of 'good' is inbuilt into the discourse.

Cowls et al. [25:112] define AI4SG as:

"the design, development and deployment of AI systems in ways that help to (i) prevent, mitigate and/or resolve problems adversely affecting human life and/or the wellbeing of the natural world, and/or (ii) enable socially preferable or environmentally sustainable developments, while (iii) not introducing new forms of harm and/or amplifying existing disparities and inequities."

While such a definition of AI4SG is deceptively simple, it bypasses and in turn obfuscates the political nature of the existing problems. This obfuscation is particularly evident in Cowls et al. [25] as they deploy the logic which announces that "AI4SG = AI×SDGs" [25:112]. Though they note the limitations of such an equation which links AI4SG to the UN Sustainable Development Goals they argue for such an approach under the assumption that "SDGs offer clear, well defined and shareable boundaries to identify positively what is socially good AI" and that "SDGs are internationally agreed on goals for development and have begun informing relevant policies worldwide, so they raise fewer questions about relativity and cultural dependency of values" [25:112]. Such framing has some glaring shortcomings that Cowls et al. choose to ignore. UN SDGs, like most United Nations consensus exercises, are not legally binding nor represent goals and principles that political groups intra-nationally or trans-nationally agree upon. Secondly, SDGs, as Telleria and Garcia-Arias [95:15] argue, is an "empty signifier that keeps disparate and even contradictory demands united". It does this by constructing "a fantasmatic explanation of international development and sustainability issues that conceals the antagonistic dimension of social, political and economic issues" [95:15]. In a later special issue introduction edited by Cowls, Cowls notes contrary to such a straightforward and simple solution that "The contributions make clear that neither 'AI' nor 'social good' should be thought of as uncontested or incontestable terms, and we should remain wary of the twin dangers of unjustified hype and unseen harm arising from the continued growth of interest in, and application of, AI" [24:54].

Cowls et al. [25] accompanied by Floridi et al. [34] aimed to lay down broad principles of AI4SG grounded on common-sensical notions of AI4SG such as trustworthiness, safeguards, consultation with users, transparency, explainability, consent, fairness and human autonomy which they termed "best practice" [34:175]. This tendency to prescribe ethical guidelines rigorously continues Luciano Floridi's project which can best be termed moral totalitarianism wherein "commandments, moral imperatives, ethical principles, codes of conduct, practical laws . . . these all endeavour to provide a clear set of instructions or patterns of operation that are designed to program and direct human social behaviour and interaction" [45:74].

4.2. AI4SG as a normative ethical field

While the field of AI ethics and its ancillary fields such as robot ethics have flourished over the last decade with a range of viewpoints, AI4SG has emerged as a normative ethical field. Following Cowls et al. [25] and Floridi et al. [34], a growing body of literature today uses the AI4SG approach to evaluate projects, justify AI approaches and projects as well as develop new principles [1,15,16,66,74,90,94,96]. At its core in such works has been a tacit acceptance of the formula that AI4SG = AI×SDGs or a mere commonsensical evocation of AI4SG. This has shaped the emerging domain of AI4SG as a normative field wherein hegemonic and normative ethical principles are being deployed to legitimize the growth of AI.

Another interesting phenomenon is the presence of the AI4SG or AI4G terms in the keywords, funding details or as a stray reference in the abstract with no follow-through in the paper [9,23,89,93,97,98]. These pieces are particularly interesting because they point towards the emergence of this field as a commonsensical domain which facilitates the usage of these terms in a marketable manner. The presence of the phrases in the funding details points towards a deeper entrenchment of these terms and their marketability for fundraising. The growth of policy discussions and references to AI4SG by state bodies such as states in Europe or states such as the United Arab Emirates also points towards a larger interest in the discourse of AI4SG by diverse kinds of states [35,71].

The influence of normative ethics doesn't simply end at an explicit link between AI4SG and the UN SDGs. Mabaso [68] suggests AI4SG be led by virtue ethics, listing out "three features of exemplarism, namely grounding in moral exemplars, meeting community expectations and practical simplicity" [68:58]. Moral exemplarism is still premised on ideals found in exemplar figures such as the 'best teacher' or an 'ideal community leader' etc. wherein the minority views or other heterogenous practices remain unacknowledged.

A contrarian attitude is visible in a minor domain of AI4SG which is concerned with thinking through adversarial strategies such as the inclusion of "well-calibrated noise, imperceptible to humans" [98:1] in datasets to reduce the risk of genomic data being matched to images of human faces. Interestingly, while [98] doesn't use the phrase AI4SG anywhere in the article, one of the project titles of a National Science Foundation-funded project referred explicitly to "Adversarial artificial intelligence for social good" [98:9].

4.3. Critiques, Challenges and Extension

A slow but steady body of work is today emerging from sciences, engineering and other applied disciplines which now deploy the notion of AI4SG to justify and frame their research work [6,12,15,16,23,26,30,55,57,59,63,66,67,73,74,88–90,94,99,101,102]. As such works display, AI4SG can best be identified as a scaffolding which allows the evaluation of an AI project amidst certain a priori, vague, liberal notions which form part of its ethos. As Schelenz & Pawelec [84] note, there is an evident echo of the discourse of Information and Communications Technologies for Development (ICT4D) that flourished for much of the past two decades. They note that both AI4SG and ICT4D display a "lack of a shared conceptual framework, excessive techno-optimism, techno-determinism, modernity bias, unequal power relations, as well as a lack of participatory approaches, sustainability, and ethical reflection" [84:12]. Though with a pinch of scepticism, techno-optimism noting the radical potential of AI in development research, practice and theory is already evident in Bjola [12].

While addressing primarily AI4SG discourse within the domain of health, Holzmeyer [56] levies a wide-ranging set of critiques against AI4SG which largely mirrors various critiques of AI. Holzmeyer anchors their critique of AI4SG around four axes, distraction from root causes, vulnerabilities and risks of big data, politics of data and expertise that frames computational solutions as better and perpetuation of "cycles of (re)validating that environmental and social variables matter to health" leading to "continued rationalization of inaction" [56:109]. A broader radical critique that Holzmeyer levies are grounded on the fact that AI4SG is "led by many of the same corporate actors that are incubating AI systems broadly" in collaboration with hegemonic actors such as the Big Tech, United Nations, established NGOs and consultancies such as McKinsey, Deloitte, Accenture. The involvement of such powerful and hegemonic actors in the AI4SG discourse necessitates a radical political interrogation and is proof that AI4SG is not a mere technical problem to be addressed via mathematical formulas and code but necessitates a larger socio-technical view of the situation. Various actors involved in AI4SG and AI are limited by "their inattention to or blindspots around larger social, political and economic systems often hamper consideration of the fuller range of social justice issues at stake" [56:112].

Madianou's [69] critique of AI4SG proceeds in another direction building on the work of Shi, Wang and Fang [91] which highlighted the global north bias of AI4SG discourse. While discussing chatbots in particular, Madianou's [69] critique seeks to highlight "concerns regarding linguistic or cultural sensitivity" [69:864]. In the process, she highlights the relationship that AI4SG has with "colonial legacies of humanitarianism" [69:864] occluding the relationships of power that are at play. This is echoed by Sapignoli [83], who notes that "new technologically sophisticated practices can reproduce historical inequalities as well as unintentionally create new ones" [83:296]. A scoping review of literature on AI for good health ends on a critical note as well, noting that "AI is being developed and implemented worldwide, and without considering what it means for populations at large, and particularly those who are hardest to reach, we risk leaving behind those who are already the most underserved" [77:13]. The review ends by noting "the dearth of literature on the ethics of AI within public health" [77:13].

Sapignoli stresses the increasing usage of AI in global governance and human rights domains arguing for ethnographic approaches to study sociotechnical systems calling for an anthropology-

inflected "AI-turn in international governance" which "would be an ethics of realism that takes into consideration and reveals the cultural, political, and economic context in which AI programs are embedded and how their applications are translated in diverse cultural contexts and jurisdictions with different consequences" [83:298]. Sapignoli's call stresses on both the theoretical—"will not only generate new insights about the nature of knowledge production and decision-making, but could also bring about a transformation of anthropological theory"—and the applied dimension—"Just as social scientists have engaged in a critical analysis of the rule of law, these scientists should unpack the rule of AI design, and hopefully mitigate its consequences, both intended and unintended" [83:298].
While critiques of AI4SG have grown, so have interfaces with various other theoretical, philosophical and methodological approaches. The call for more meaningful stakeholder participation has emerged as one of the calls for operationalizing AI4SG [63]. A more nuanced elaboration of stakeholder consultation is elaborated by Bondi et al. [13] who argue for an interfacing of the capabilities approach to operationalise AI4SG. The cost-benefit analysis in such an instance is not tied to global SDG goals but is evaluated by considering the capabilities of the various participants noting that "the goals of the project are properly aligned to do so without negatively impacting other capabilities" [13:7]. Berberich et al. [10] call for thinking through AI4SG from the perspective of the notion of harmony in East Asian philosophy. They echo calls for human-centred AI when they note that "intelligent systems should always adapt to the human pace instead of the other way around" [10:636]. At its core, their vision of an AI system acknowledges the affective constitution and limits of human relationalities when they note that "tactful AI must have some capability of approximately inferring the emotional and cognitive states of people with whom they are interacting and a model on the effects that different possible actions might have on humans." [10:636].

Table 1
Typology of AI4SG literature

Category	Description	Papers
Application of AI4SG	Description of application of AI projects claiming to work with an AI4SG or AI4G ethos.	[23,57,59,67,88–90,98,101]
Keyword invocation of AI4SG or funding details	Articles which mention AI4SG or AI4G only in their keywords or their funding clause.	[9,23,88,93,97,98]
Critique of AI4SG and ancillary approaches	Critique of AI4SG or AI4G conceptualization along with critiques of other ancillary notions such as ICT4D	[10,24,56,69,84]
AI4SG principles	Delineating, defining and scaffolding AI4SG principles for operationalization.	[1,10,13,34,63,68,96]
The challenge of principles implementation	Notes the 'gap' between principles and practices	[85]
Database of AI4SG projects	Collation of AI4SG projects	[25]
Domain-specific review	Domain-specific literature of AI4SG	[77]
Policy related work	Policy-related work thinking operationalizing AI4SG	[35,71]
Social scientific intervention	Social scientific methodological interventions to further AI4SG	[83]

Schiff et al. [85] focus on the principles-to-practice gap identifying six barriers, incentive dilemmas, the complexity of AI's impacts, the disciplinary divide, many hands problem, governance of knowledge and overabundance of tools. Despite thinking through the problem of implementation Schiff et al. [85] privilege only certain kinds of disciplines' inclusion in debates around implementation noting

"collaboration between computing, engineering, organizational, business, and other scholars" [85:90]. While "other scholars" are included as a last thought, computing disciplines, business and organizational disciplines are recruited to think through the implementation challenge excluding all core social science disciplines and philosophy. It is precisely this tendency to seek quick fixes and solutions that necessitates a review of the field of relational ethics and artificial intelligence.

5. Relational Ethics and AI

5.1. What is relational ethics?

The existing literature presents relational ethics as an approach that proposes a "fundamental shift—from rational to relational—in thinking about personhood, data, justice and everything in between" [11:1]. It hopes to move away from normative ethical principles that are "often formulated as problem/solution" [11:1] towards a more critical evaluation of the field that calls for "rethinking of justice and ethics as a set of broad, contingent, and fluid concepts and down-to-earth practices that are best viewed as a habit and not a mere methodology of data science" [11:1]. One of the earliest instances, where the power of a relativist stance is vindicated is in a 1998 article by Rössler & Matsuno [82]. They juxtaposed two approaches: deep technology and standard ethics vs standard technology and deep ethics.

This earliest distinction is interesting as these fault lines can be traced in the literature even today. While not mutually exclusive, the literature on AI relational ethics is divided along two axes: First axis concentrates on the 'what is' question, which debunks the property-based approach to the question of bots and their moral status in favour of relational ontology wherein "The partners do not pre-exist their relating; the partners are precisely what comes out of the inter- and intra-relating of fleshly, significant, semioticmaterial being'' (Haraway 2008, p. 165 as cited in [22:10]); second axis addresses the 'how' part of relational ethics which foregrounds ethical approaches/principles that go beyond problem/solution dyad to address power imbalances, inequalities, and relations of injustices inherent in the design and development of new and existing technologies.

While the former hopes to completely redefine the notion of the self and the other when it interrogates the moral status of AI and its relationship to the human—a deep technology view to the question of human, machine or animal—the latter hopes to arrive at relational yet standard approaches to design, develop and evaluate AI systems that minimise algorithmic harms and injustices. Both these axes, however, oppose other normative ethical frameworks that maintain "a strict divide between the creator (spiritual) and the created (non-spiritual)… with humans in the position of creator" [18:57].

These critiques point towards the limits of principalism and approaches to moral standing that assume an "ontological platform onto which morality is mounted" [22:9]. Relational ethics, on the other hand, argues that,

> "the moral-epistemic ground…is not situated in the 'essence' of the [object], but is something that comes into being, is gradually revealed and constructed as we get to know objects in-relation, as human beings who interact with them, watch them, call them by name, etc.—whether as scientists, philosophers, or as lay persons" [22:16].

5.2. Thinking through and operationalising relational ethics

The first axis that is invested in thinking through the deep technology problematic of relational ethics has been gaining slow yet steady ground since the 2010s focusing on theorizing human-bot relations while intervening in the debate on the moral status and rights of robots and AI. Coeckelbergh [18] opened the field by interrogating the anthropocentric bias in thinking about the "technological future" arguing in favour of a "deep relational view on human being and self" that borrowed from ecological and eastern philosophical approaches. Cockelbergh and Gunkel have been the most prominent voices foregrounding a relational view of the question of the moral status of bots and human-bot relations. They reframe the standard ethical question "does the animal have a face" to ask "[W]hat does it take for an animal to supervene and be revealed as having face in the Levinasian sense?" which demands that we learn how to respond and what responding will mean in such a case [22:9]. Thus, in asking the

question of response-ability, they replace the figure of a moral philosopher and observer with an active practitioner and sense-maker who is part of a lifeworld that operates on a certain 'we'.

The question of relations, of how to respond to plants and other non-humans, thus is a political societal and collective question [20]. A property-based approach to the question of the moral/legal status of others privileges ontological properties such as consciousness, sentience, intelligence, suffering etc as valid grounds to decide an object's moral/legal status. Developing his critique of this approach further, Gunkel[46] put forth three philosophical problems with the properties approach: substantive, terminological, and epistemological complications. Similar critiques have been put forth by others against a property-based approach to argue in favour of a deep ecology approach [60], an eco-centric approach to AI [65], establish humans as relational selves [28], and identify dangers of ethical disobligation in our encounter with AI [17]. The field is also influenced by indigenous epistemologies to reframe the notion of intelligence [70] and reconceptualize a notion of the self, privacy, and personhood [61]. It should be noted that the field of human rights for robots is much larger than what we have captured here. We have only focussed on relational approaches to the problematic of robot rights and status but for a more comprehensive view, one can refer to Gordon & Pasvenskiene [43] and Harris and Anthis [51].

The second axis moves away from the question of robot rights and moral/legal status to operationalize relational ethics more concretely. Birhane makes a clear attempt to call for relational approaches to AI ethics, harms and injustice to make a case for reconceptualizing the notion of "personhood, data, justice and everything in between" [11:1]. Birhane responds to the growing trend in the field that approaches AI ethics as a technical problem, and the subsequent flourishing of AI ethics toolkits, to argue for an understanding of AI ethics that goes "beyond technical solutions" [11:1].

While Birhane and others limit themselves to a critical examination of the field rather than offering solutions, a significant effort has been made, both before and after Birhane, to arrive at relational reconceptualizations of specific ethical principles such as transparency, responsibility, fairness, intelligence as well as the need to understand AI as an algorithmic assemblage with deeply relational affordances [29]. Ananny [3:7] puts forth a sociotechnical definition of AI systems, as a unit of AI ethical analysis, that they call networked information algorithm (NIA),

> "as an assemblage …of institutionally situated computational code, human practices, and normative logics that creates, sustains, and signifies relationships among people and data through minimally observable, semiautonomous action. Although code, practices, and norms may be observed individually in other contexts, their full ''meaning and force ... can only be understood in terms of relations with other modular units…''.

Ananny [3:2] identifies three dimensions for scrutinizing AI ethics and holding algorithmic assemblages accountable:

> "The ability to convene people by inferring associations from computational data, the power to judge similarity and suggest probable actions, and the capacity to organize time and influence when action happens"

They too distinguish their approach from "mathematical, mechanistic focus" [3:5] which only asks if a code is biased or not to instead asks whether different assemblages ''help us get into satisfactory relation with other parts of our experience'' [3:7].

Focussing on the use of AI in education, Henry & Oliver [52] make a similar move to argue that "ethics should not be understood as abstract values or design decisions, but as socio-technical achievements" [52:330]. They deploy Puig de la Bellasca's speculative critique to offer a vision of caring that is 'an analytics or provocation' that presupposes heterogeneity as the ontological ground. In exploring the tension between ethics and justice, they foreground questions of not only "for whom" but who cares, what for and why do we care?" [52:334].

Along similar lines, others have focussed on identifying what new demands AI makes of ethical frameworks and how these frameworks and principles might be reconceptualized in light of these demands. Coeckelbergh focuses on the question "of responsibility attribution for artificial intelligence technologies" [21:2052] and introduces another actor in the responsibility relation: moral patients, who are affected by the action of the agent and those to whom moral agents are responsible and answerable. Giovanola and Tiribelli [40,41] rearticulate 'fairness' as a "distributive and socio-relational dimension that comprises of three main components: fair equality of opportunity, equal right to justification, and fair equality of relationship" [41:9]; De Togni et al. [27] review three levels of 'intelligences' and

'robotness' to argue that intelligence is socially relevant, and [49] define AI-mediated communication as interpersonal communication involving intelligent actors.

Others in the field have focussed on identifying the social impacts of AI systems in domains such as healthcare, education, warfare, and labour to ask for deeper stakeholder engagement to balance considerations of relational care, safety, privacy and transparency [53:582]; Döbler & Bartnik [29] focus on technological affordances that can reveal, manipulate and conceal possible actions. Schoenherr [87] focuses on social-cognitive factors that affect society's attitudes towards surveillance technologies, impending identity crises due to the changing definitions of intelligence [70], ethico-politics of autonomous weapon systems [86], the impact of AI systems on work that might encourage a shift towards a "relational realm rather than a transactional one" [80].

Table 2
Typology of relational AI ethics literature

Category	Description	Papers
Conceptual debates on specific AI ethical parameters	Works that develop or nuance the definition and meaning of specific ethical parameters such as responsibility, transparency, and fairness from a relational ethics approach.	[19,21,27,31,32,40,41]
Human-bot relations	Works that theorise human-bot relations from a relational perspective.	[17,18,27,28,36,65,70]
Social Impacts of AI	Works that focus on the social impacts of AI with an emphasis on the adoption of more relational perspectives to address these impacts.	[8,19,29,44,48,49,53,54,70,72,80,86,87]
Moral Status of robots	Works that intervene in the debate on the moral status of robots while advocating for a relational approach to the problem of moral status.	[2,17,18,37,42,46,51,60,61,65]
Critiques and challenges	Works critiquing the relational approach to the problem of the moral status of AI	[38,76]
Theoretical approaches to AI (relational) ethics	Works that advance the field of (relational) AI ethics more broadly.	[11,32,32,39,46,52,81,100]
Ethical evaluation of AI	Works that focus on using a relational perspective to the problem of governance or evaluation of AI.	[54,58]

5.3. Critiques and Challenges

Relational ethics is often critiqued as a metaphysical approach that says more about the nature of being without generating a clear ethics i.e., the challenge of deriving an 'ought' from an 'is' [38]. Müller [76:582] classifies the relational turn as a "relativist account of moral status with all the problems that come with that: no possibility to be right or wrong (to 'respond' in the right or wrong way), no possibility of better or worse views, no possibility of moral progress" etc. According to Müller, non-human objects deserve 'consideration' where behaviour towards them can be judged due to their relationship with moral agent/patient but not as objects with moral status themselves.

Gibert & Martin [38] offer a twofold interpretation of the relational argument. First, that relational ethics works with a notion of moral status implied in a valuable relationship; second, its emphasis on the role of a community in the development and assignment of social identity, personhood, and moral status. Within this framework, they identify problems of consistency and objectivity wherein the meaning of a valuable relationship is subjective, introduces fluctuations in the moral status, and the inadequate distinction between the token/type. Rejecting grounds of relationality, intelligence or life as inadequate, they make a case for sentience as an adequate ground to grant moral status to AI with a caveat that no AI systems in the contemporary exhibit these qualities of sentience as yet.

Gordon & Gunkel [42] can be read as responding to Gibert & Martin [38] in their aim to demonstrate "the importance of discussing the moral impact and status of intelligent and autonomous machines" [38:3] even in the case of faulty AI systems, we have today like COMPASS etc. By including error-prone AI systems in the discussion, they hope to interrogate or stress-test existing moral theories and practices that are premised on notions of inclusion/exclusion.

This is an important point to consider and is part of the challenge of a relational approach. A relational approach to AI ethics borrows from poststructuralist approaches that primarily operate at the level of critique which can help identify limits of existing knowledge paradigms, ethical frameworks or standard operations. Recognizing the critical gesture inherent in the relational approach to AI ethics, Birhane [11] asked for "a system-wide acceptance of critical work as an essential component of AI ethics, fairness, and justice" [11:1]. A system-wide acceptance can help expand the methodological repertoire of AI ethics scholars that can include more qualitative approaches to help nuance the debate on AI ethics and governance. At the same time however, and as has been noted in the previous section, a relational ethics approach is not averse to redefining specific ethical principles from a relational point of view which has helped expand and redefine standard ethical principles of transparency, responsibility, fairness etc. While more nuanced theoretical approaches to relational ethics are welcoming and needed, we believe that more work is needed that can expand and redefine ethical principles from the relational ethics point of view which make it easier to operationalize the relational approach and make 'doing' relational ethics more accessible.

6. The need for interfacing relational ethics and AI4SG

As noted, AI4SG is slowly emerging as a normative ethical field within the larger domain of AI ethics with minor contrarian or non-normative approaches. This necessitates interrogating and interfacing non-normative ethical approaches with AI4SG. The dominant approach of linking AI4SG with SDGs doesn't guard against AI harms and injustices but instead reduces the problem of 'social' to a technical process of identifying socially relevant AI systems as identified by the international bureaucratic machine.

Relational ethics take AI assemblage as its unit of analysis that opens up the 'social' to the political, societal and collective question that addresses the relations of power, inequality, harms and injustices as more than instrumental concerns. To interface AI4SG with relational ethics is to open up the black box of not just technical concerns but also social relations and see AI ethics as a socio-technical achievement [4,5]. A relational approach can help unpack the construction and production of each of the signifiers of AI4SG as not finished concepts but as works-in-progress.

7. Limitations and future work

The fundamental limitation of the paper is the usage of mainstream indexes such as Scopus and Web of Science and the inclusion criteria of the English language. Future research in all languages across other indexes could help shed light on how and if at all AI4SG and relational ethics are interfacing in the existing literature.

Building on Birhane's [11] work of reinterrogating and reconceptualising all concepts and processes of AI ethics appears to be a fruitful endeavour to relook at AI4SG discourse and its foreclosures.

Acknowledgements

This material is based upon work supported in whole or in part by The Notre Dame-IBM Tech Ethics Lab. Such support does not constitute an endorsement by the sponsor of the views expressed in this publication.

References

[1] Ramya Akula and Ivan Garibay. 2021. Ethical AI for Social Good. In *HCI International 2021 - Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings*, Springer-Verlag, Berlin, Heidelberg, 369–380. DOI:https://doi.org/10.1007/978-3-030-90963-5_28

[2] Marcos Alonso. 2023. Can Robots have Personal Identity? *Int J of Soc Robotics* 15, 2 (February 2023), 211–220. DOI:https://doi.org/10.1007/s12369-022-00958-y

[3] Mike Ananny. 2016. Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology, & Human Values* 41, 1 (January 2016), 93–117. DOI:https://doi.org/10.1177/0162243915606523

[4] Cheshta Arora and Debarun Sarkar. 2023. Destabilizing Auditing: Auditing as 'care-ful socio-analogue/digital relation. In *21st Annual STS Conference Graz 2023 "Critical Issues in Science, Technology and Society Studies*. Graz, Austria.

[5] Cheshta Arora and Debarun Sarkar. 2023. Auditing Artificial Intelligence as a New Layer of Mediation: Introduction of a new black box to address another black box. *Hipertext.net* 26 (May 2023), 65–68. DOI:https://doi.org/10.31009/hipertext.net.2023.i26.10

[6] Aníbal Monasterio Astobiza, Mario Toboso, Manuel Aparicio, and Daniel López. 2021. AI Ethics for Sustainable Development Goals. *IEEE Technology and Society Magazine* 40, 2 (June 2021), 66–71. DOI:https://doi.org/10.1109/MTS.2021.3056294

[7] Jacqui Ayling and Adriane Chapman. 2022. Putting AI ethics to work: are the tools fit for purpose? *AI and Ethics* 2, (August 2022). DOI:https://doi.org/10.1007/s43681-021-00084-x

[8] Cosmin Badea and Gregory Artus. 2022. Morality, Machines, and the Interpretation Problem: A Value-based, Wittgensteinian Approach to Building Moral Agents. In *Artificial Intelligence XXXIX: 42nd SGAI International Conference on Artificial Intelligence, AI 2022, Cambridge, UK, December 13–15, 2022, Proceedings*, Springer-Verlag, Berlin, Heidelberg, 124–137. DOI:https://doi.org/10.1007/978-3-031-21441-7_9

[9] Engineer Bainomugisha, Joel Ssematimba, and Deo Okure. 2023. Design Considerations for a Distributed Low-Cost Air Quality Sensing System for Urban Environments in Low-Resource Settings. *Atmosphere* 14, 2 (February 2023), 354. DOI:https://doi.org/10.3390/atmos14020354

[10] Nicolas Berberich, Toyoaki Nishida, and Shoko Suzuki. 2020. Harmonizing Artificial Intelligence for Social Good. *Philos. Technol.* 33, 4 (December 2020), 613–638. DOI:https://doi.org/10.1007/s13347-020-00421-8

[11] Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. *Patterns* 2, 2 (February 2021), 100205. DOI:https://doi.org/10.1016/j.patter.2021.100205

[12] Corneliu Bjola. 2022. AI for development: implications for theory and practice. *Oxford Development Studies* 50, 1 (January 2022), 78–90. DOI:https://doi.org/10.1080/13600818.2021.1960960

[13] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A. Killian. 2021. Envisioning Communities: A Participatory Approach Towards AI for Social Good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (AIES '21), Association for Computing Machinery, New York, NY, USA, 425–436. DOI:https://doi.org/10.1145/3461702.3462612

[14] Rosi Braidotti. 2006. *Transpositions: on nomadic ethics*. Polity Press, Cambridge, UK ; Malden, MA.

[15] Morgan Briggs. 2021. SatDash: An Interactive Dashboard for Assessing Land Damage in Nigeria and Mali. In *ACM SIGCAS Conference on Computing and Sustainable Societies* (COMPASS '21), Association for Computing Machinery, New York, NY, USA, 100–114. DOI:https://doi.org/10.1145/3460112.3471949

[16] Marianna Capasso and Steven Umbrello. 2022. Responsible nudging for social good: new healthcare skills for AI-driven digital personal assistants. *Med Health Care and Philos* 25, 1 (March 2022), 11–22. DOI:https://doi.org/10.1007/s11019-021-10062-z

[17] William Clapton and Laura J. Shepherd. 2019. Ethics *Ex Machina* : popular culture and the plural futures of politics. *Australian Journal of Political Science* 54, 4 (October 2019), 531–542. DOI:https://doi.org/10.1080/10361146.2019.1663400

[18] Mark Coeckelbergh. 2013. Pervasion of what? Techno–human ecologies and their ubiquitous spirits. *AI & Soc* 28, 1 (February 2013), 55–63. DOI:https://doi.org/10.1007/s00146-012-0418-y

[19] Mark Coeckelbergh. 2016. Responsibility and the Moral Phenomenology of Using Self-Driving Cars. *Applied Artificial Intelligence* 30, 8 (September 2016), 748–757. DOI:https://doi.org/10.1080/08839514.2016.1229759

[20] Mark Coeckelbergh. 2018. What do we mean by a relational ethics?: Growing a relational approach to the moral standing of plants, robots, and other non-humans. In *Plant Ethics*, Angela Kallhoff, Marcello Di Paola and Maria Schörgenhumer (eds.). Routledge.

[21] Mark Coeckelbergh. 2020. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Sci Eng Ethics* 26, 4 (August 2020), 2051–2068. DOI:https://doi.org/10.1007/s11948-019-00146-8

[22] Mark Coeckelbergh and David J. Gunkel. 2014. Facing Animals: A Relational, Other-Oriented Approach to Moral Standing. *J Agric Environ Ethics* 27, 5 (October 2014), 715–733. DOI:https://doi.org/10.1007/s10806-013-9486-3

[23] Marta R. Costa-jussà, Esther González, Asuncion Moreno, and Eudald Cumalat. 2020. Abusive language in Spanish children and young teenager's conversations: data preparation and short text classification with contextual word embeddings. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, 1533–1537. Retrieved June 2, 2023 from https://aclanthology.org/2020.lrec-1.191

[24] Josh Cowls. 2021. 'AI for Social Good': Whose Good and Who's Good? Introduction to the Special Issue on Artificial Intelligence for Social Good. *Philos. Technol.* 34, 1 (November 2021), 1–5. DOI:https://doi.org/10.1007/s13347-021-00466-3

[25] Josh Cowls, Andreas Tsamados, Mariarosaria Taddeo, and Luciano Floridi. 2021. A definition, benchmark and database of AI for social good initiatives. *Nat Mach Intell* 3, 2 (February 2021), 111–115. DOI:https://doi.org/10.1038/s42256-021-00296-0

[26] Veronica Dahl and Juan José Moreno-Navarro. 2022. Doughnut Computing in City Planning for Achieving Human and Planetary Rights. In *Bio-inspired Systems and Applications: from Robotics to Ambient Intelligence* (Lecture Notes in Computer Science), Springer International Publishing, Cham, 562–572. DOI:https://doi.org/10.1007/978-3-031-06527-9_56

[27] Giulia De Togni, Sonja Erikainen, Sarah Chan, and Sarah Cunningham-Burley. 2021. What makes AI 'intelligent' and 'caring'? Exploring affect and relationality across three sites of intelligence and care. *Social Science & Medicine* 277, (May 2021), 113874. DOI:https://doi.org/10.1016/j.socscimed.2021.113874

[28] Nicole Dewandre. 2019. Humans as relational selves. *AI & Soc* 34, 1 (March 2019), 95–98. DOI:https://doi.org/10.1007/s00146-017-0700-0

[29] Niklas Döbler and Clemens Bartnik. 2022. Normative Affordances Through and By Technology: Technological Mediation and Human Enhancement. *International Journal of Interactive Multimedia and Artificial Intelligence* 7, (September 2022), 14–23. DOI:https://doi.org/10.9781/ijimai.2022.09.006

[30] Fei Fang, Milind Tambe, Bistra Dilkina, and Andrew J. Plumptre (Eds.). 2019. *Artificial Intelligence and Conservation*. Cambridge University Press, Cambridge, United Kingdom ; New York, NY.
[31] Heike Felzmann, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2019. Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society* 6, 1 (January 2019), 205395171986054. DOI:https://doi.org/10.1177/2053951719860542
[32] Mark Findlay and Josephine Seah. 2020. An Ecosystem Approach to Ethical AI and Data Use: Experimental Reflections. In *2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G)*, IEEE, Geneva, Switzerland, 192–197. DOI:https://doi.org/10.1109/AI4G50087.2020.9311069
[33] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. DOI:https://doi.org/10.2139/ssrn.3518482
[34] Luciano Floridi, Josh Cowls, Thomas C. King, and Mariarosaria Taddeo. 2020. How to Design AI for Social Good: Seven Essential Factors. *Sci Eng Ethics* 26, 3 (June 2020), 1771–1796. DOI:https://doi.org/10.1007/s11948-020-00213-5
[35] Francesca Foffano, Teresa Scantamburlo, and Atia Cortés. 2023. Investing in AI for social good: an analysis of European national strategies. *AI & Soc* 38, 2 (April 2023), 479–500. DOI:https://doi.org/10.1007/s00146-022-01445-8
[36] Paul Formosa. 2021. Robot Autonomy vs. Human Autonomy: Social Robots, Artificial Intelligence (AI), and the Nature of Autonomy. *Minds & Machines* 31, 4 (December 2021), 595–616. DOI:https://doi.org/10.1007/s11023-021-09579-2
[37] Patrick Gamez, Daniel B. Shank, Carson Arnold, and Mallory North. 2020. Artificial virtue: the machine question and perceptions of moral character in artificial moral agents. *AI & Soc* 35, 4 (December 2020), 795–809. DOI:https://doi.org/10.1007/s00146-020-00977-1
[38] Martin Gibert and Dominic Martin. 2022. In search of the moral status of AI: why sentience is a strong argument. *AI & Soc* 37, 1 (March 2022), 319–330. DOI:https://doi.org/10.1007/s00146-021-01179-z
[39] Karamjit S. Gill. 2016. Data Driven Wave of Certainty- a question of ethical sustainability. *IFAC-PapersOnLine* 49, 29 (2016), 117–122. DOI:https://doi.org/10.1016/j.ifacol.2016.11.068
[40] Benedetta Giovanola and Simona Tiribelli. 2022. Weapons of moral construction? On the value of fairness in algorithmic decision-making. *Ethics and Inf. Technol.* 24, 1 (March 2022). DOI:https://doi.org/10.1007/s10676-022-09622-5
[41] Benedetta Giovanola and Simona Tiribelli. 2022. Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. *AI & Soc* (May 2022). DOI:https://doi.org/10.1007/s00146-022-01455-6
[42] John-Stewart Gordon and David J. Gunkel. 2022. Moral Status and Intelligent Robots. *The Southern J of Philosophy* 60, 1 (March 2022), 88–117. DOI:https://doi.org/10.1111/sjp.12450
[43] John-Stewart Gordon and Ausrine Pasvenskiene. 2021. Human rights for robots? A literature review. *AI Ethics* 1, 4 (November 2021), 579–591. DOI:https://doi.org/10.1007/s43681-021-00050-7
[44] Gabriel Grill and Nazanin Andalibi. 2022. Attitudes and Folk Theories of Data Subjects on Transparency and Accuracy in Emotion Recognition. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1 (avril 2022), 78:1-78:35. DOI:https://doi.org/10.1145/3512925
[45] David Gunkel. 2022. The Symptom of Ethics; Rethinking Ethics in the Face of the Machine. *HMC* 4, (2022), 67–83. DOI:https://doi.org/10.30658/hmc.4.4
[46] David J Gunkel. 2019. No Brainer: Why Consciousness is Neither a Necessary nor Sufficient Condition for AI Ethics. In *Papers of the Towards Conscious AI Systems Symposium*, CEUR-WS, United States. Retrieved January 27, 2023 from https://ceur-ws.org/Vol-2287/paper9.pdf
[47] Gregory D. Hager, Ann Drobnis, Fei Fang, Rayid Ghani, Amy Greenwald, Terah Lyons, David C. Parkes, Jason Schultz, Suchi Saria, Stephen F. Smith, and Milind Tambe. 2019. Artificial Intelligence for Social Good. DOI:https://doi.org/10.48550/arXiv.1901.05406
[48] Hallowell N, Badger S, Sauerbrei A, Nellåker C, and Kerasidou A. 2022. "I don't think people are ready to trust these algorithms at face value": trust and the use of machine learning algorithms

in the diagnosis of rare disease. *BMC medical ethics* 23, 1 (November 2022). DOI:https://doi.org/10.1186/s12910-022-00842-4

[49] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication* 25, 1 (March 2020), 89–100. DOI:https://doi.org/10.1093/jcmc/zmz022

[50] Donna Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14, 3 (1988), 575–599. DOI:https://doi.org/10.2307/3178066

[51] Jamie Harris and Jacy Anthis. 2021. The Moral Consideration of Artificial Entities: A Literature Review. *Science and Engineering Ethics* 27, (August 2021). DOI:https://doi.org/10.1007/s11948-021-00331-8

[52] Jade Vu Henry and Martin Oliver. 2022. Who Will Watch the Watchmen? The Ethico-political Arrangements of Algorithmic Proctoring for Academic Integrity. *Postdigit Sci Educ* 4, 2 (April 2022), 330–353. DOI:https://doi.org/10.1007/s42438-021-00273-1

[53] Anita Ho. 2020. Are we ready for artificial intelligence health monitoring in elder care? *BMC Geriatr* 20, 1 (December 2020), 358. DOI:https://doi.org/10.1186/s12877-020-01764-9

[54] C. W. L. Ho, D. Soon, K. Caals, and J. Kapur. 2019. Governance of automated image analysis and artificial intelligence analytics in healthcare. *Clin Radiol* 74, 5 (May 2019), 329–337. DOI:https://doi.org/10.1016/j.crad.2019.02.005

[55] Andreas Holzinger, Anna Saranti, Alessa Angerschmid, Carl Orge Retzlaff, Andreas Gronauer, Vladimir Pejakovic, Francisco Medel-Jimenez, Theresa Krexner, Christoph Gollob, and Karl Stampfer. 2022. Digital Transformation in Smart Farm and Forest Operations Needs Human-Centered AI: Challenges and Future Directions. *Sensors* 22, 8 (January 2022), 3043. DOI:https://doi.org/10.3390/s22083043

[56] Cheryl Holzmeyer. 2021. Beyond 'AI for Social Good' (AI4SG): social transformations—not tech-fixes—for health equity. *Interdisciplinary Science Reviews* 46, 1–2 (April 2021), 94–125. DOI:https://doi.org/10.1080/03080188.2020.1840221

[57] Meng-Leong How, Sin-Mei Cheah, Aik Cheow Khor, and Yong Jiet Chan. 2020. Artificial Intelligence-Enhanced Predictive Insights for Advancing Financial Inclusion: A Human-Centric AI-Thinking Approach. *Big Data and Cognitive Computing* 4, 2 (June 2020), 8. DOI:https://doi.org/10.3390/bdcc4020008

[58] Liza Ireni-Saban and Maya Sherman. 2021. *Ethical Governance of Artificial Intelligence in the Public Sector*. Routledge, London. DOI:https://doi.org/10.4324/9781003106678

[59] Lakshmi Shankar Iyer and Sirish Venkatagiri. 2019. Is Digital GOOD for providing water security?- A case study of India. *International Journal of Advanced Science and Technology* 28, 16 (November 2019), 1074–1092.

[60] Nancy S. Jecker. 2021. Can we wrong a robot? *AI & Soc* (November 2021). DOI:https://doi.org/10.1007/s00146-021-01278-x

[61] Nancy S. Jecker, Caesar A. Atiure, and Martin Odei Ajei. 2022. The Moral Standing of Social Robots: Untapped Insights from Africa. *Philos. Technol.* 35, 2 (April 2022), 34. DOI:https://doi.org/10.1007/s13347-022-00531-5

[62] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nat Mach Intell* 1, 9 (September 2019), 389–399. DOI:https://doi.org/10.1038/s42256-019-0088-2

[63] Meghana Kshirsagar, Caleb Robinson, Siyu Yang, Shahrzad Gholami, Ivan Klyuzhin, Sumit Mukherjee, Md Nasir, Anthony Ortiz, Felipe Oviedo, Darren Tanner, Anusua Trivedi, Yixi Xu, Ming Zhong, Bistra Dilkina, Rahul Dodhia, and Juan M. Lavista Ferres. 2021. Becoming Good at AI for Good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (AIES '21), Association for Computing Machinery, New York, NY, USA, 664–673. DOI:https://doi.org/10.1145/3461702.3462599

[64] Bruno Latour. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. OUP Oxford.

[65] Migle Laukyte. 2019. Against Human Exceptionalism: Environmental Ethics and the Machine Question. In *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence*, Don

Berkich and Matteo Vincenzo d'Alfonso (eds.). Springer International Publishing, Cham, 325–339. DOI:https://doi.org/10.1007/978-3-030-01800-9_18

[66] Ying-Tung Lin, Tzu-Wei Hung, and Linus Ta-Lun Huang. 2021. Engineering Equity: How AI Can Help Reduce the Harm of Implicit Bias. *Philos. Technol.* 34, S1 (November 2021), 65–90. DOI:https://doi.org/10.1007/s13347-020-00406-7

[67] Yanbiao Ma, Yuxin Li, Kexin Feng, Yu Xia, Qi Huang, Hongyan Zhang, Colin Prieur, Giorgio Licciardi, Hana Malha, Jocelyn Chanussot, Pedram Ghamisi, Ronny Hänsch, and Naoto Yokoya. 2021. The Outcome of the 2021 IEEE GRSS Data Fusion Contest - Track DSE: Detection of Settlements Without Electricity. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, (2021), 12375–12385. DOI:https://doi.org/10.1109/JSTARS.2021.3130446

[68] Bongani Andy Mabaso. 2021. Artificial Moral Agents Within an Ethos of AI4SG. *Philos. Technol.* 34, 1 (November 2021), 7–21. DOI:https://doi.org/10.1007/s13347-020-00400-z

[69] Mirca Madianou. 2021. Nonhuman humanitarianism: when "AI for good" can be harmful. *Information, Communication & Society* 24, 6 (April 2021), 850–868. DOI:https://doi.org/10.1080/1369118X.2021.1909100

[70] Suvradip Maitra. 2020. Artificial Intelligence and Indigenous Perspectives: Protecting and Empowering Intelligent Human Beings. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (AIES '20), Association for Computing Machinery, New York, NY, USA, 320–326. DOI:https://doi.org/10.1145/3375627.3375845

[71] Laura Margheri. 2016. Women in Engineering, Science, and Technology in the United Arab Emirates [Women in Engineering]. *IEEE Robot. Automat. Mag.* 23, 2 (June 2016), 102–104. DOI:https://doi.org/10.1109/MRA.2016.2558299

[72] Domènec Melé. 2021. Ethics at the workplace in the fourth industrial revolution: A Catholic social teaching perspective. *Business Ethics, the Environment & Responsibility* 30, 4 (2021), 772–783. DOI:https://doi.org/10.1111/beer.12368

[73] Omar Mohammed, Kris Manohar, Kimberley Gillette, Brandon Murphy, and Patrick Hosein. 2022. Data4Good: An Established Framework for supporting Civil Society Organizations. In *2022 IEEE International Humanitarian Technology Conference (IHTC)*, 73–78. DOI:https://doi.org/10.1109/IHTC56573.2022.9998401

[74] Michal Monselise and Christopher C. Yang. 2022. AI for Social Good in Healthcare: Moving Towards a Clear Framework and Evaluating Applications. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*, IEEE, Rochester, MN, USA, 470–471. DOI:https://doi.org/10.1109/ICHI54592.2022.00072

[75] Chantal Mouffe. 2013. *Agonistics: Thinking the World Politically*. Verso Books.

[76] Vincent Müller. 2021. Is it time for robot rights? Moral status in artificial entities. *Ethics and Information Technology* 23, (December 2021). DOI:https://doi.org/10.1007/s10676-021-09596-w

[77] Kathleen Murphy, Erica Di Ruggiero, Ross Upshur, Donald J. Willison, Neha Malhotra, Jia Ce Cai, Nakul Malhotra, Vincci Lui, and Jennifer Gibson. 2021. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Medical Ethics* 22, 1 (February 2021), 14. DOI:https://doi.org/10.1186/s12910-021-00577-8

[78] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews* 10, 1 (March 2021), 89. DOI:https://doi.org/10.1186/s13643-021-01626-4

[79] Thao Phan, Jake Goldfein, Declan Kuch, and Monique Mann. 2022. Introduction: Economies of Virtue. In *Economics of Virtue – The Circulation of 'Ethics" in AI*, Thao Phan, Jake Goldfein, Declan Kuch and Monique Mann (eds.). Institute of Network Cultures, Amsterdam, 6–22. Retrieved from https://networkcultures.org/wp-content/uploads/2022/12/EconomiesofVirtueINC2022TOD46-2.pdf

[80] Carlos Rodriguez-Lluesma, Pablo García-Ruiz, and Javier Pinto-Garay. 2021. The digital transformation of work: A relational view. *Business Ethics, the Environment & Responsibility* 30, 1 (2021), 157–167. DOI:https://doi.org/10.1111/beer.12323
[81] O. E. Rössler and K. Matsuno. 1998. The relativist stance. *Biosystems* 46, 1–2 (April 1998), 213–216. DOI:https://doi.org/10.1016/s0303-2647(97)00101-9
[82] Otto E Rössler and Koichiro Matsuno. 1998. The relativist stance. *Biosystems* 46, 1–2 (April 1998), 213–216. DOI:https://doi.org/10.1016/S0303-2647(97)00101-9
[83] Maria Sapignoli. 2021. Anthropology and the AI-Turn in Global Governance. 115, (ed 2021), 294–298. DOI:https://doi.org/10.1017/aju.2021.38
[84] Laura Schelenz and Maria Pawelec. 2022. Information and Communication Technologies for Development (ICT4D) critique. *Information Technology for Development* 28, 1 (January 2022), 165–188. DOI:https://doi.org/10.1080/02681102.2021.1937473
[85] Daniel Schiff, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, and Michael Lennon. 2021. Explaining the Principles to Practices Gap in AI. *IEEE Technol. Soc. Mag.* 40, 2 (June 2021), 81–94. DOI:https://doi.org/10.1109/MTS.2021.3056286
[86] Birgit Schippers. 2020. Autonomous weapons systems and ethics in International Relations. In *The Routledge Handbook to Rethinking Ethics in International Relations*, Birgit Schippers (ed.). Oxon, 312–325. DOI:https://doi.org/10.4324/9781315613529
[87] Jordan Richard Schoenherr. 2020. Understanding Surveillance Societies: Social Cognition and the Adoption of Surveillance Technologies. In *2020 IEEE International Symposium on Technology and Society (ISTAS)*, 346–357. DOI:https://doi.org/10.1109/ISTAS50296.2020.9462205
[88] Navid Shaghaghi, Andres Calle, and George Kouretas. 2020. Expanding eVision's Scope of Influenza Forecasting. In *2020 IEEE Global Humanitarian Technology Conference (GHTC)*, IEEE, Seattle, WA, USA, 1–10. DOI:https://doi.org/10.1109/GHTC46280.2020.9342864
[89] Navid Shaghaghi, Andres Calle, George Kouretas, Supriya Karishetti, and Tanmay Wagh. 2021. Expanding eVision's Granularity of Influenza Forecasting. In *Wireless Mobile Communication and Healthcare* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), Springer International Publishing, Cham, 227–243. DOI:https://doi.org/10.1007/978-3-030-70569-5_14
[90] Zheyuan Ryan Shi. 2021. AI for Social Good: Between My Research and the Real World. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 18 (May 2021), 15732–15733. DOI:https://doi.org/10.1609/aaai.v35i18.17863
[91] Zheyuan Ryan Shi, Claire Wang, and Fei Fang. 2020. Artificial Intelligence for Social Good: A Survey. DOI:https://doi.org/10.48550/arXiv.2001.01818
[92] Benedictus de Spinoza. 1901. *Improvement of the understanding, Ethics and Correspondence of Benedict de Spinoza*. London [England] ; Toronto [Canada] : L. Beling Tetens. Retrieved December 19, 2018 from http://archive.org/details/improvementofund00spinuoft
[93] Sanja Štajner, Daniel Ferrés, Matthew Shardlow, Kai North, Marcos Zampieri, and Horacio Saggion. 2022. Lexical simplification benchmarks for English, Portuguese, and Spanish. *Frontiers in Artificial Intelligence* 5, (2022). Retrieved May 28, 2023 from https://www.frontiersin.org/articles/10.3389/frai.2022.991242
[94] David Steingard, Marcello Balduccini, and Akanksha Sinha. 2022. Applying AI for social good: Aligning academic journal ratings with the United Nations Sustainable Development Goals (SDGs). *AI & Soc* (June 2022). DOI:https://doi.org/10.1007/s00146-022-01459-2
[95] Juan Telleria and Jorge Garcia-Arias. 2022. The fantasmatic narrative of 'sustainable development'. A political analysis of the 2030 Global Development Agenda. *Environment and Planning C: Politics and Space* 40, 1 (February 2022), 241–259. DOI:https://doi.org/10.1177/23996544211018214
[96] Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle C. M. Belgrave, Daphne Ezer, Fanny Cachat van der Haert, Frank Mugisha, Gerald Abila, Hiromi Arai, Hisham Almiraat, Julia Proskurnia, Kyle Snyder, Mihoko Otake-Matsuura, Mustafa Othman, Tobias Glasmachers, Wilfried de Wever, Yee Whye Teh, Mohammad Emtiyaz Khan, Ruben De Winne, Tom Schaul, and Claudia Clopath. 2020. AI for

social good: unlocking the opportunity for positive impact. *Nat Commun* 11, 1 (May 2020), 2468. DOI:https://doi.org/10.1038/s41467-020-15871-z

[97] Sapdo Utomo, A John, Ayush Pratap, Zhi-Sheng Jiang, P. Karthikeyan, and Pao-Ann Hsiung. 2023. AIX Implementation in Image-Based PM2.5 Estimation: Toward an AI Model for Better Understanding. In *2023 15th International Conference on Knowledge and Smart Technology (KST)*, 1–6. DOI:https://doi.org/10.1109/KST57286.2023.10086917

[98] Rajagopal Venkatesaramani, Bradley A. Malin, and Yevgeniy Vorobeychik. 2021. Re-identification of individuals in genomic datasets using public face images. *Science Advances* 7, 47 (November 2021), eabg3296. DOI:https://doi.org/10.1126/sciadv.abg3296

[99] Vasiliki Voukelatou, Lorenzo Gabrielli, Ioanna Miliou, Stefano Cresci, Rajesh Sharma, Maurizio Tesconi, and Luca Pappalardo. 2021. Measuring objective and subjective well-being: dimensions and data sources. *Int J Data Sci Anal* 11, 4 (May 2021), 279–309. DOI:https://doi.org/10.1007/s41060-020-00224-2

[100] Rosalie Waelen. 2022. Why AI Ethics Is a Critical Theory. *Philos. Technol.* 35, 1 (February 2022), 9. DOI:https://doi.org/10.1007/s13347-022-00507-5

[101] Hua Wang, Sneha Gupta, Arvind Singhal, Poonam Muttreja, Sanghamitra Singh, Poorva Sharma, and Alice Piterova. 2022. An Artificial Intelligence Chatbot for Young People's Sexual and Reproductive Health in India (SnehAI): Instrumental Case Study. *J Med Internet Res* 24, 1 (January 2022), e29969. DOI:https://doi.org/10.2196/29969

[102] Ruitao Xie, Connor Qiu, and Guoping Qiu. 2022. Finding Beautiful and Happy Images for Mental Health and Well-Being Applications. In *Pattern Recognition and Computer Vision* (Lecture Notes in Computer Science), Springer Nature Switzerland, Cham, 704–717. DOI:https://doi.org/10.1007/978-3-031-18913-5_54