



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## (Re)politicising data-driven education

**Citation for published version:**

Knox, J 2022, '(Re)politicising data-driven education: from ethical principles to radical participation', *Learning, Media and Technology*. <https://doi.org/10.1080/17439884.2022.2158466>

**Digital Object Identifier (DOI):**

[10.1080/17439884.2022.2158466](https://doi.org/10.1080/17439884.2022.2158466)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Learning, Media and Technology

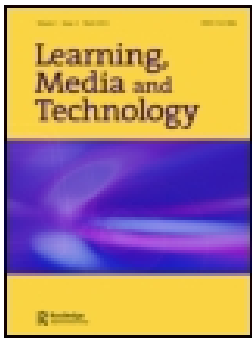
**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.





## (Re)politicising data-driven education: from ethical principles to radical participation

Jeremy Knox

**To cite this article:** Jeremy Knox (2022): (Re)politicising data-driven education: from ethical principles to radical participation, Learning, Media and Technology, DOI: 10.1080/17439884.2022.2158466

**To link to this article:** <https://doi.org/10.1080/17439884.2022.2158466>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 21 Dec 2022.



Submit your article to this journal [↗](#)



Article views: 345



View related articles [↗](#)



View Crossmark data [↗](#)

# (Re)politicising data-driven education: from ethical principles to radical participation

Jeremy Knox 

Centre for Research in Digital Education, The University of Edinburgh, Edinburgh, UK

## ABSTRACT

This paper examines ways in which the ethics of data-driven technologies might be (re)politicised, particularly where educational institutions are involved. The recent proliferation of principles, guidelines, and frameworks for ethical 'AI' (artificial intelligence) have emerged from a plethora of organisations in recent years, and seem poised to impact educational governance. This trend will be firstly shown to align with a narrow form of ethics - deontology - and overlook other potential ways ethical reasoning might contribute to thinking about 'AI'. Secondly, the attention to ethical principles will be suggested to focus excessively on the technology itself, with the effect of masking political concerns for equity and justice. Thirdly and finally, the paper will propose a more radical form of participation in ethical decision-making that not only challenges the assumption of universal consensus, but also draws more authentically on the capacities for debate, contestation, and exchange inherent in the educational institution.

## ARTICLE HISTORY

Received 17 June 2022



Accepted 8 December 2022

## KEYWORDS

AI; artificial intelligence; ethics; social justice; agonism

## Introduction

Following the increasing societal impact of technologies involved in intensive data-driven processing – often obscurely termed 'artificial intelligence', or 'AI', 'machine learning', and 'analytics' – there has been a surge of interest in whether or not such systems might be deemed 'ethical'. Central to this trend has been considerations of the ethics of so-called 'artificial intelligence', or 'AI'; terms which have become a convenient shorthand for a vast array of technologies, mostly focused on using large data sets to train algorithms. Indeed, as Tucker suggests, the term 'now functions in the vernacular primarily to obfuscate, alienate, and glamorize' (2022), pointing out the ways in which 'AI' tends to be used to mask more accurate descriptions of specific data processing, that in their detail, would highlight ethical issues such as transparency, responsibility, and agency. Nevertheless, a plethora of principles, guidelines, and frameworks for the ethical development and use of 'AI' now dominate public discourse, not only establishing specific themes and dimensions, such as privacy, fairness, or accountability, but also prefiguring the domain of ethics as one of standard-setting and protocolisation. All manner of organisations now appear to have ethical principles for 'AI', from supranational bodies such as the EU to luxury car manufacturers such as Rolls Royce,<sup>1</sup> despite being involved in often very different practices of data generation, extraction, and processing, applied to contexts that vary considerably. This proliferation of ethical codes has engendered a substantial research agenda focused on analysing, visualising, and distilling the

**CONTACT** Jeremy Knox  jeremy.knox@ed.ac.uk  Centre for Research in Digital Education, The University of Edinburgh, Edinburgh, EH8 8AQ, UK

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

vast field of published standards into key elements and definitive codes (see Floridi and Cowls 2019; Jobin, Ienca, and Vayena 2019; Fjeld et al. 2020). As such, the emerging understanding of the ethics of intensive data processing technologies seems to increasingly resemble an apolitical space of consensus, where sets of decisive themes supposedly encompass and delimit all ethical eventualities, and any demand for deliberation and debate appears ever more redundant. In other words, any organisation considering the ethical dimensions of their data processing routines need not look very far to encounter an existing set of guiding principles that might be repurposed with little in the way of customisation, or indeed, the need to consult stakeholders.

Nevertheless, the wealth of ethical principles currently in the publication has not avoided criticism, chiefly in the form of questioning the extent to which ethical principles for data-driven technologies are published cynically, often as a way of private companies avoiding regulation or the need to take practical action (Hagendorff 2020; Benkler 2019; Yeung, Howes, and Pogrebna 2020). However, a more profound set of questions remain, not only about the kind of ‘ethics’ being foregrounded through the apparent obsession with principles and protocols, but also the ways in which such frameworks tend to function to foreclose both alternative views of social impact, as well as opportunities to participate in defining ethical issues themselves. Furthermore, in their assumed capacity to explain, rationalise, and even predict circumstances which might qualify as being a matter of ‘ethics’, such principles seem to align with underlying beliefs about the solutionist orientations of data-driven technologies themselves; as a means to remodel ‘all complex social situations either as neatly defined problems with definite, computable solutions or as transparent and self-evident processes that can be easily optimized – if only the right algorithms are in place’ (Morozov 2013, 5). It might be argued, therefore, that ethical principles are too often presented as straightforward ‘solutions’ to the ‘problems’ associated with ethics, where some form of policy rubric is assumed to provide the means to avoid unethical behaviour, often substituting for any empirical understanding of how specific data practices actually impact those involved.

Some areas of education appear to be following the broader trend for ethical principles, with the University of Buckingham, a private university in the UK, recently publishing ‘The Ethical Framework for AI in Education’ (examined further below), with the express aim of enabling ‘all learners to benefit optimally from AI in education’ (IEAIE 2021, 2). Whether or not the Buckingham example establishes a broader trend amongst public universities and other educational institutions, there remains a pressing need to examine alternative ways of understanding the ethical implications of data-driven technologies for the sector. Given the potential capacity of universities, in particular, to mobilise multi-, inter-, and trans-disciplinary research and teaching, as well as cross-sectoral collaboration, education seems well-placed to engage in approaches that acknowledge and embrace the inherent politics of technology development. As such, following an examination of the recent trend for the publishing of ethical principles and its potential impact on educational institutions, this paper will outline three ways in which the ethics of data-driven technologies might be approached alternatively.

## The proliferation of principles

Despite a long history of ethical reasoning related to artificial intelligence, Borenstein et al. emphasise a ‘sudden burst of interest’ (2021, 97) in recent years, principally due to the ways in which particular techniques of intensive data processing, such as ‘machine learning’, have been assumed under the umbrella term of ‘AI’. Borenstein et al. (2021) suggest a more than seven-fold increase in references to ‘AI ethics’ in Google Scholar citations, from a mere 45 in 2017 to a substantial 342 in 2020, implying a wide public interest in the ways such technologies might be impacting the workings of government, businesses, and social institutions. Indeed, around this time, a vast array of national and regional governments, supra-national organisations, NGOs, businesses, academic groups, standardisation organisations, and consumer and public interest bodies, began to publish ethical principles, guidelines, and frameworks for ‘AI’. The trend might be traced back

to two reports published in 2016: firstly ‘Ethically Aligned Design’ (IEEE 2016) by the IEEE (The Institute of Electrical and Electronics Engineers) Global Initiative on Ethics of Autonomous and Intelligent Systems; and secondly, ‘Preparing for the Future of Artificial Intelligence’ (EOPNSTCCT 2016) published under the Obama administration in the US. The first was published in draft form by one of the most prominent professional technology organisations globally – the IEEE – and opened for public consultation. It defined four key principles: ‘human benefit’, ‘responsibility’, ‘transparency’, and ‘education and awareness’, accompanied by a number of proposed issues and recommendations (see IEEE 2016). These principles and recommendations were subsequently revised and refined following substantial feedback (see IEEE 2019), and as such, demonstrate a certain level of participative development, specifically through iterative feedback, situated at the very beginnings of the trend for ethical protocols and frameworks. However, there are significant differences here to the ideas of multi-stakeholder engagement that will be examined below. The report from the Obama administration is also important to highlight here, in the sense that it was one of the first examples of high-level government policy on the social impact of data-driven technologies, establishing a tension between supposed economic benefits and risks to national security, public safety, and individual privacy (EOPNSTCCT 2016). In this sense, one might argue the report established a particular view that ‘AI’ would impact society in both advantageous and detrimental ways, and only by avoiding the latter could the technology fulfil its promise in driving the supposed ‘fourth industrial revolution’ (Schwab 2016).

The prominence of these reports incited a broader trend for the establishing of ethical codes, including, in 2018, the ‘Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems’ from the European Commission and the ‘Charlevoix Common Vision for the Future of Artificial Intelligence’ from the leaders of the G7. In 2019, ‘Principles on Artificial Intelligence’ was published by the Organisation for Economic Co-operation and Development (OECD) and the ‘Beijing AI Principles’ were announced by the Beijing Academy for Artificial Intelligence (BAAI). Numerous technology companies appeared to view this as a key moment to participate, with Tencent, Google, Baidu, IBM, Intel, and Microsoft all publishing versions of ethical principles around this time (see Jobin, Ienca, and Vayena 2019; Hagedorff 2020; Fjeld et al. 2020 for more extensive lists). Various research institutes, professional organisations, and civil society groups also joined the trend, notable examples including the ‘Leverhulme Centre for the Future of Intelligence’<sup>2</sup> at the University of Cambridge, the ‘AI Now’ institute<sup>3</sup> at New York University, and the non-profit organisation ‘Data Society’.<sup>4</sup> Additionally notable here are coalitions formed explicitly to produce ethical frameworks, such as the ‘Partnership on AI’<sup>5</sup> and ‘AI4People’.<sup>6</sup> The first attempt at global standard-setting for the ethics of ‘AI’ took place on the 24th of November 2021, with the adoption on the UNESCO ‘Recommendation on the Ethics of Artificial Intelligence’ (UNESCO 2021).

With such a proliferation of standards, codes, and principles, various research projects have sought to document, analyse, map, and visualise the range of publications and policies on ‘ethical AI’. These include the ‘Toolbox: Dynamics of AI Principles’<sup>7</sup> produced by the AI Ethics Lab and the ‘AI Ethics Guidelines Global Inventory’,<sup>8</sup> produced by the non-profit research organisation Algorithm Watch.<sup>9</sup> The latter provides a searchable database of 173 guidelines and is presented as a public resource for those interested in further studying and critiquing the trend. Other initiatives have sought to visualise the broad array of ethical principles, such as the ‘Mapping AI and data ethics’<sup>10</sup> project by the Ada Lovelace Institute, ‘Fluxus Landscape: An Expansive View of AI Ethics and Governance’<sup>11</sup> by Icarus Salon, and the ‘AI Governance Database’<sup>12</sup> from Nesta. Significantly, this practice is elsewhere explicitly framed as a research methodology, specifically aimed at developing the understanding of ethics or providing the means to refine a definitive set of standards. For example, the ‘Linking Artificial Intelligence Principles’ (LAIP)<sup>13</sup> project (also see Zeng, Lu, and Huangfu 2019) documents 89 AI ethics principles and suggests both a ‘coarse’ sorting of published principles into 11 categories, as well as a further 41 ‘finer topics’. Zheng et al. further suggest ‘the necessity of linking and incorporating various AI Principles into a comprehensive framework and focusing on how they can interact and complement each other’ (2019). More explicitly, in a study

from the Berkman Klein Centre, Fjeld et al. document and analyse 36 principles in order to derive eight key themes: ‘privacy; accountability; safety and security; transparency and explainability; fairness and non-discrimination; human control of technology; professional responsibility; and the promotion of human values’ (2020, 4). Fjeld et al. describe their analysis process in terms of the:

desire for a way to compare these documents – and the individual principles they contain – side by side, to assess them and identify trends, and to uncover the hidden momentum in a fractured, global conversation around the future of AI. (Fjeld et al. 2020, 3)

In this sense the very proliferation of published ethical principles is framed as a research conundrum itself, requiring analytical methods to distil, refine, and discover an underlying authenticity. Jobin et al. present a similar approach, analysing 84 publications and producing as a result ‘[e]ven overarching ethical values and principles’ of ‘transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity’ (2019, 391). Significantly, they claim an ‘emerging convergence’ around five specific principles (transparency, justice and fairness, non-maleficence, responsibility and privacy), derived from the assertion that the terms were ‘referenced in more than half of all the sources’ (Jobin, Ienca, and Vayena 2019, 391). An additionally notable example here is the suggestion of a ‘unified framework’ of five principles – beneficence, non-maleficence, autonomy, justice, and explicability – offered by Floridi and Cowls (2019).<sup>14</sup> They further suggest, ‘the sheer volume of proposed principles threatens to become overwhelming and confusing’ (Floridi and Cowls 2019, 2), potentially resulting in ‘unnecessary repetition and overlap, if the various sets of principles are similar, or confusion and ambiguity, if they differ’ (Floridi and Cowls 2019, 3).

Across these examples, therefore, developing a consensus around core ethical statements and positions appears to be an explicit goal, while any notion of uncertainty or difference in interpretation and understanding is considered a weakness, and something to be methodologically expunged from the process. As such, there seems to be an underlying commitment to processes of abstraction that assumes a straightforward route to universal relevance. Furthermore, the trend for principles is overwhelmingly deterministic, in the sense that it assumes ‘AI’ technologies to be both the cause and potential solution to various unethical occurrences. Indeed, the entire framing of (un)ethical conditions is premised upon the prevalence of the technology, its ultimate desirability, and its eventual refinement, and is therefore completely unable to address wider questions about the ‘ethics’ of society more generally, especially where data-driven technologies might be involved indirectly, peripherally, or not at all. These are dilemmas which, as this paper argues, necessitate a challenge to the dominance of ethical principle-making, particularly where the practice may become adopted uncritically by educational institutions. Such a challenge is suggested, not necessarily to call for an end to the exercise of producing ethical principles entirely, but rather to highlight both potential limitations in the form of ethics they embody, as well as alternative ways of understanding the societal impacts from intensive data processing, for which educational institutions seem uniquely placed to develop.

As introduced above, the ‘Ethical Framework for AI in Education’ (see IEAIE 2021) offers one key example of policy making in the education sector that appears to align with the wider trend for principles and codes. The report was produced by a short-lived ‘Institute for Ethical AI in Education’, funded by Microsoft and education publishers McGraw Hill and Pearson, with the specific remit to produce an ethical framework over a 36-month period between 2018 and 2021. In a similar vein to the numerous publications outlined previously, the ‘Ethical Framework for AI in Education’ defines nine ‘objectives’: achieving educational goals; forms of assessment; administration and workload; equity; autonomy; privacy; transparency and accountability; informed participation; and ethical design (IEAIE 2021). The accompanying report describes a process of ‘wide consultation designed to listen to and learn from the perspectives of a cross-section of stakeholders’ (IEAIE 2021, 2). This involved roundtable events and a summit held in November 2020, deployed to ‘arrive at a shared understanding of the ethical implications of using AI in education, and to make

recommendations on how AI should be designed and applied ethically in practice’ (IEAIE 2021, 3). As such, while forms of consultation were identified, the orientation of the endeavour appears resolutely geared towards consensus-building and the development of solutions to ease ethical concerns and hence advance the uptake of data-driven technologies in educational settings. The deterministic framing of the report is further clarified in the suggestion:

By utilising AI ethically and with purpose, societies can look forward to addressing previously overwhelming educational inequalities and enabling all learners, from all backgrounds, to achieve their full potential, as long as there is universal and equal access to the necessary hardware, infrastructure and connectivity. (IEAIE 2021, 4)

As such, the ‘Ethical Framework for AI in Education’ seems to establish a concerning precedent for the education sector, not only in formalising narrow sets of universal guidelines and principles as the accepted means of ‘engineering away’ ethical concerns, but also presenting the ensuing propagation of the technology as unquestionably desirable and beneficial, all underpinned by the assumption of a consensual future of ethically-pure education, driven by data. These are the three aspects of the emerging domain of ‘AI ethics’ that subsequent sections of this paper seek to challenge. Firstly, the drastically limited articulation of ethics promoted by the trend for principles and codes, for which alternative philosophical positions will be outlined. Secondly, the problematic focus on technology as both source and solution to ethical dilemmas, for which perspectives from social justice will be examined. And thirdly, the prevailing inclination towards universal consensus and accord, for which a more radical suggestion of political contestation will be suggested.

## Returning to ‘ethics’ as a discipline

One of the most prominent critiques of ethical principles has suggested the practice of ‘ethics washing’, directed principally at private companies in their apparent rush to be seen to be participating in the trend (Benkler 2019; Hagendorff 2020; Yeung, Howes, and Pogrebna 2020). Thus, at least amongst the technology industry, ethics is often seen as a matter of public relations rather than any authentic interest or engagement with the ethics or morals of professional practice (Hagendorff 2020). Indeed, a cynical approach to ethical principles associated with AI seems so commonplace that the MIT Technology Review recently published a humorous guide to proclaiming ethical intent ‘without incriminating yourself’ (see Hao 2021). Bietti further suggests:

The word ‘ethics’ is under siege in technology policy circles. Weaponized in support of deregulation, self-regulation or handsoff governance, ‘ethics’ is increasingly identified with technology companies’ self-regulatory efforts and with shallow appearances of ethical behavior. (Bietti 2021, 210)

Here Bietti exemplifies what might be seen as a movement to reclaim the domain of ethics from what is often perceived to be a superficial engagement with the discipline. Indeed, Vallor contends that ethics has been ‘stripped for parts’ (2021) through the recent attention to ‘AI’ principles and frameworks. As Tasioulas further makes clear, such concerns are often directed at the increasing role of the private technology sector:

Thanks in part to the incursion of big tech into the AI ethics space, ‘ethics’ is often interpreted in an unduly diminished way. For example, as a form of soft, self-regulation lacking legal enforceability. Or, even more strangely, it is identified with a narrow sub-set of ethical values. (Tasioulas 2021)

Drawing on such critical perspectives is vital to correctly situate the practice of ‘principle-making’ within a much broader disciplinary domain of ethics and moral philosophy. The trend toward ethical principles might be aligned most closely with deontology or deontological ethics (Hagendorff 2020). Deontology is often focused, although not exclusively so, on rules for moral conduct, for example, determining the ‘goodness’ or ‘correctness’ of an individual’s actions by examining the extent to which such actions have adhered to underlying laws that define ‘good’ or ‘correct’ behaviour. In this sense, the prominence of ethical principles tends to establish and emphasise a very particular form of ethics – one that gives foundational significance to rules and codes – amongst many

other articulations of morality. For example, in the academic domains of philosophy, deontology is usually contrasted with consequentialism, precisely because consequentialism shifts the focus of ethical judgement to the *consequences* of individual action rather than the extent to which an ethical code has been followed by the individual actor. Indeed, as is often pointed out by consequentialists, a deontological perspective would judge an ethical outcome to be ‘good’, even where the result of the action turned out to be ‘bad’, as long as the actor followed the principle. While it is not the intention of this paper to provide a comprehensive philosophical explanation or critique of ‘rules-based’ ethics, this brief definition signals the very specific and limited form of ethical reasoning elevated through the recent overprovision of ‘AI principles’, and one that has arguably concealed a great variety and diversity of alternative ideas.

Key examples here include Hagendorff’s (2020) contrasting of ‘virtue ethics’ – essentially, ethical decision-making based on ‘good character’ – with deontology and the habitual publishing of principles, drawing on a more fully-developed position of ‘technomoral virtues’ developed by Vallor (2016). As such, Hagendorff proposes a focus ‘on the individual level’ of ‘technologists or software engineers and their social context’ (2020, 122), rather than the technology itself. As such, this presents one striking alternative to the overwhelming focus on ‘AI’ encountered in the previously discussed principles, and one that seems to productively shift attention to individual conduct as the site of ethical reasoning. As Vallor suggests, virtue ethics is ‘a way of thinking about the good life as achievable through specific moral traits and capacities that humans can actively cultivate in themselves’ (Vallor 2016, 10), seeming to offer a more expansive view of ethical possibilities than those directly focused on the specific design or functioning of data-driven technology. This is one important step in foregrounding questions of desirable ethical conditions, in and of themselves, rather than those tied conditionally to the use of the technology. Furthermore, and importantly, the focus on virtue ethics is not necessarily individualistic, nor divorced entirely from ideas of politics. Vallor (2016) articulates this more clearly through reference to Confucianism, as one of three classical approaches to virtue ethics that might inform a more contemporary ‘technomoral’ version. Vallor describes Confucianism in terms of:

the need for persons to cultivate in themselves the kind of moral virtues that enable the flourishing of relationships within the family – virtues that are then gradually extended outward to other relationships to promote broader political flourishing. The Confucian self is not an isolated, autonomous individual but a being defined by relationships and reciprocal obligations to others. (Vallor 2016, 38)

Regardless of what may indeed be the many merits of virtue ethics as a strategy for navigating the datafied society of our time, including its educational institutions, it remains important to emphasise here, not any one particular alternative form ethical reasoning, but rather the broad contribution that might be made through further engagement with the philosophy of ethics. Such an engagement would seem to enrich the considerations of how to achieve desirable outcomes from the increasing incursions of ‘AI’. As Vallor further suggests in response to critical views of the potential role of ethics in the emerging discussions of the societal impact of AI:

there is a full and vital conversation going on that’s part of ethics in the humanities, that’s not remotely politically denatured. The fact that it’s not often present in the AI ethics discourse is not a reason to have less ethics in the discourse, it’s a reason to have richer contributions from the humanities brought in. (Vallor 2021)

Here the deep disciplinary traditions of the university might be seen as a resource, not only to challenge the narrow and reductionist discourses of AI ethics, but also to augment and reinvigorate the current interest in data-driven technologies and their increasing influence over societal values.

### **Situating ‘AI’ *within* social justice**

Alongside concerns over so-called ‘ethics washing’, questions have been raised about the extent to which high-level principles can be straightforwardly transformed into ‘on the ground’ ethical



practices (Jobin, Ienca, and Vayena 2019). Mittelstadt suggests ‘a principled approach may have limited impact on design and governance’ (2019, 501) due to a number of factors, including professional norms, common aims, existing procedures for translating principles into practices, and accountability mechanisms. Furthermore, as McNamara, Smith, and Murphy-Hill (2018) demonstrate, ethical codes have little impact on changing the behaviour of technology professionals. Proposed solutions to the intangibility of ethical principles include suggestions to develop more training on how to operationalise such frameworks (Canca 2020), or to better situate them within existing governance structures (Fjeld et al. 2020). However, such suggestions seem to favour hierarchical approaches to ethics management, and preserve the idea that principles themselves have a particular value and authority, which might be ultimately achieved through better design and engineering; either of the technologies themselves, or of the wider contexts in which they are implemented or governed. Here, more attention might be paid to the ways principles are engaged as modes of prediction, oversight, and control, in ways that mirror the technological solutionism of the data-driven technologies they purport to ameliorate. In their study of ethical principles, Greene et al. find a ‘deterministic vision ... the ethics of which are best addressed through technical and design expertise’, within an approach ‘closer to conventional business ethics than more radical traditions of social and political justice’ (Greene, Lauren Hoffmann, and Stark 2019, 1). Further, Greene et al. also find little engagement with the idea that data-driven technology ‘can be limited or constrained’ (2019, 1), suggesting that ethical principles are often driven by those with a vested interest in the increased circulation of the technology.

There are two key and interrelated assumptions here worth emphasising. Firstly, that principles, in their very abstraction from more concrete ‘on the ground’ practices, are assumed to serve as objective and authoritative representations of reality, and in such a way, align with the positivistic orientations of the computer science, statistics, and psychology disciplines from which data-driven technologies tend to emerge. Secondly, that ‘AI’ is framed as both the source of, and solution to, proposed ethical dilemmas, in such a way that the technology itself is positioned as an undisputable part of the future being envisioned. For example, where ethical principles herald ‘privacy’ as a palpable concern for society in a coming age of ‘AI’, it is usually proposed as a condition brought about by the surveillant functioning of the data-driven technology, but also one that is potentially ‘solved’ by, for example, ensuring that the systems comply with relevant legal frameworks (this is one of the examples offered by the Ethical Framework for AI in Education – see IEAIE 2021, 8). As such, principles seem oriented towards paving the way for further uptake of the technology rather than being concerned, arguably, with ethical dilemmas more broadly.

Rather than focusing on ever-improving technical design or inevitable technological futures, therefore, a more critical approach to data-driven technologies might turn towards the social, political, and justice-related contexts into which such systems are so often situated. Whittaker et al. decry persistent attempts from the technology industry to ‘reframe political questions as technical concerns’ (2018, 32), further suggesting:

historical patterns of discrimination and classification, which often construct harmful representations of people based on perceived differences, are reflected in the assumptions and data that inform AI systems, often resulting in allocative harms. This perspective requires one to move beyond locating biases in an algorithm or dataset. (Whittaker et al. 2018, 25)

In other words, the search for ethical concerns might be reversed; not by identifying something engendered by the technology specifically (and ultimately solvable by it), but rather through focusing on the contexts into which the technologies are deployed. Underpinning this position is a considerable body of research that has long argued for, not only an understanding of technology as ‘embedded in social, political, cultural, and economic worlds, shaped by humans, institutions, and imperatives’ (Crawford 2021, 211), but also a specific conception of data itself as inherently political (Bigo et al. 2019). This position is further developed in research concerned explicitly with ‘data justice’ suggested by Dencik et al. (2019) to be:

a framework for shifting the entry-point and debate on data-related developments in a way that foregrounds social justice concerns and ongoing historical struggles against inequality, oppression and domination. (Dencik et al. 2019, 876)

Such perspectives require a disciplinary shift in thinking that foregrounds already-existing societal concerns when considering the development or deployment of technologies rather than their internal design or technical coherence. In other words, this is a wholesale transformation to thinking first and foremost about the injustices and discrimination that particular communities or social contexts *already experience*, and using this as a basis with which to assess the impact of technology in terms of whether such injustices are exacerbated or eased. In this sense, work in data justice offers a productive re-grounding of the form of ethics seemingly favoured by abstract principles, by situating ethical concerns, not in pre-emptive statements and their accompanying technical solutions, but rather in the intersections between already-established social experiences and the growing influence of data-driven decision-making. For Green, this is an overtly political move, and one that necessitates a rearticulation of the role of the ‘data scientist’ specifically: ‘it is not enough to have good intentions – data scientists must ground their efforts in clear political commitments and rigorous evaluations of the consequences’ (Green 2021, 261). Green goes on to claim:

As a form of political action, data science can no longer be separated from broader analyses of social structures, public policies, and social movements. Instead, the field must debate what impacts are desirable and how to promote those outcomes – thus prompting rigorous evaluations of the issues at hand and openness to the possibility of non-technological alternatives. (Green 2021, 261)

This is precisely the sensitivity that might productively inform ethical considerations in education, particularly where data-driven technologies are increasingly shaping educational roles, responsibilities, and experiences. This might take the form of purposefully (re)orienting questions about the design and ethics of data-driven technologies towards already-existing concerns and experiences from those actively involved in education, as opposed to preconceived ideas about what might be ‘enhanced’ through the application of technology (see Bayne 2015), or indeed what might be understood as ethical through the deriving of core principles. For example, rather than prefiguring ethical concerns by rearticulating established terms, such as ‘privacy’, this might involve posing broad questions, perhaps simply: *what happens to the classroom when data-driven technologies are used? Or, how are teacher and student roles, experiences, and agencies shaped?* Further, this might divert attention from sets of questions about ethics that assume the inevitability of ubiquitous datafication – for example, framing issues of privacy, equity, or transparency and accountability exclusively in terms of data management or technical functioning – and foreground questions that concern education more generally. For example, what might be determined as the meaning and purpose of education (see for example, Biesta 2009; Biesta 2010), and therefore where and how data-driven technologies might be situated in relation. For example, Biesta places significant emphasis on both ‘socialisation’ and ‘subjectification’ as key purposes of education, the former having ‘to do with the many ways in which, through education, we become members of and part of particular social, cultural and political ‘orders’’, while the latter concerns ‘ways of being that hint at independence’ and ‘in which the individual is not simply a ‘specimen’ of a more encompassing order’ (2009, 40). Such purposes, concerned with how young people join society and become individuals, seem fundamental to the functioning of education as a social institution, yet it is difficult to meaningfully align very focused principles and codes to such high-level considerations. Where Biesta suggests that what matters in education ‘is the ‘quality’ of subjectification, i.e., the kind of subjectivity – or kinds of subjectivities – that are made possible as a result of particular educational arrangements and configurations’ (Biesta 2009, 41), one might expect definitive ethical guidelines for ‘AI’ to provide some insight or direction, particularly given that data-driven systems seem likely to substantially shape the kind of subjects that might emerge from the experience. The Ethical Framework for AI in Education (IEAIE 2021), as discussed previously, provides no

discernible acknowledgement of the ways data-driven systems might take part in particular kinds of subjectification, let alone any suggestion of what we *should* think about the technological impact of datafication on the sense of individuality. Suggesting such a diversion is not intended to imply a straightforward solution, or indeed an alternative form of determinism (‘social’ or ‘individual’ rather than ‘technological’). Neither is such a shift in focus aimed at negating the importance of issues related to privacy, equity, or transparency and accountability – all of these concerns have undoubted relevance to questions of the ethics of education, just not exclusively so to questions about the collection, generation, processing, and feedback of data.

## The politics of participation

Across the previous sections, both the philosophical and justice-related perspectives on ethics have implied an additionally important area of consideration relating to how ethical concerns about data-driven technologies are derived, and specifically, who is included in the process. There are two centrally important aspects to this politics of participation: of the extent of participation, and that of its purpose. The first concerns the fundamental question of *who* gets to decide what (un)ethical issues are where data-driven technology is concerned. As we have seen in the above examples from the IEEE and the temporary Institute for Ethical AI in Education, an engagement with modes of participation has been a core feature of principal-making since the current trend started. It certainly seems that the Institute for Ethical AI in Education were attentive to different stakeholders where, out of the eight roundtable events scheduled to develop the framework, three were ‘dedicated to participation by young people’ (IEAIE 2021, 2). Nevertheless, the extent to which different populations are actively involved, as well as the processes through which contributions are elicited and included remains a methodological as well as a social justice challenge, especially where the resulting principles are promoted as straightforwardly representative of universal concerns. Returning to the social justice theme from the previous section, Fraser’s notion of participatory parity (2008) might usefully inform further considerations here, encouraging questions about distributive (economic), recognitive (cultural), and representational (political) justice. For example, how economic (*re*)distribution impacts the capacity to participate in educational decision-making, how particular cultural groups are *recognised*, or not, as legitimate participants, and how procedures and institutions to limit or expand *representation* and political voice. For Fraser, such participatory parity is central to social justice concerns, constituting ‘social arrangements that permit all to participate as peers in social life’ (2008, 405). A comprehensive engagement with such a framework would seem to necessitate a broad and inclusive approach to participation, as well as an in-depth consideration of the processes and rituals of involvement that go beyond more typical methods of consultation that appear to have informed the principles and frameworks for ‘AI’ ethics discussed previously. Furthermore, participative parity implies a temporal dimension; a requirement for ongoing participation that doesn’t seem to be addressed in the desire to publish definitive principles, that presumably are assumed to remain relevant and authoritative even as contexts might change. It is notable here that the Institute for Ethical AI in Education disbanded after their framework was published, seeming to underscore the idea that ethical principles are permanent, and once they have been discovered, the task is complete.

The second and related aspect of participation, that relating to its purpose, suggests a more profound set of questions about the ultimate intention of ethical reasoning about so-called ‘AI’. As we have seen above, the tendency in principle-making seems overwhelmingly directed towards consensus, where the aim of the endeavour is the production of a universal framework to which all would find agreement and accord. However, the final argument this paper will make is that, in the evident urgency to identify and define harmonious ethical principles for the use of data-driven technologies, some of the value of debate, dialogue, and disagreement, as political processes through which knowledge and understanding is developed, has been lost. Indeed, it is precisely the educational domain, and, in particular, the university, where such activities have been fostered and

held in esteem, and where, therefore, hasty protocolisation might seem somewhat out of place. In Goheen's (1969) discussion of the attributes that most authentically characterise the university, it is the idea of *tensions* that best describe the potentially productive forces that, not only hold the institution together, but also allow it to function. Goheen establishes this analogy by utilising images of the bow and lyre, drawing on the work of the pre-Socratic Greek philosopher Heraclitus:

The tension of the bow, the strain put on its opposite ends, gives the arrow force to carry firmly to the mark. In the playing of a lyre, harmony results only where there is contrast – where there is interplay among tones at variance with one another. (Goheen 1969, 12)

The friction and opposition discernible in these devices, so Goheen (1969) argues, offers the ultimate vision of the university as comprised of, and driven by, the 'cross pull' of tension. The underlying strength of the university, one might therefore argue, and the capacity it might most productively attune to the development of ethical understanding about data-driven technologies is not merely in universal standard-setting, but in the continued political opposition and exchange across its disciplinary diversity. Universities might therefore seek to foster wider processes of open-ended debate, exchange, and dialogue about the impact of 'AI' on education, not only across natural science, engineering, social science, and humanities disciplines, but also with broader educational stakeholders and community organisations. Pertinent examples of university-led research here include the Just AI network<sup>15</sup> and the VirtEU project,<sup>16</sup> both of which have sought to develop specific methodologies for the participative and multi-stakeholder development of ethical understanding. However, neither of these projects has focused specifically on the ways data-driven technologies are being increasingly used in universities themselves.

The emphasis on tensions might also usefully be acknowledged in 'AI' technologies themselves. Rather than being assumed to be contained, precise, and objective instruments, or as Crawford suggests 'bloodless spaces of rational determination' (2016, 79), data-driven technologies are better understood as being designed and developed through contingent practices, predicated upon 'tensions and contests' (Crawford 2016, 79) between the various actors involved. Crawford (2016) draws on Mouffe's (2013) concept of agonism to challenge the idea that data-driven technology is produced through internally-coherent decision-making, instead suggesting a domain of perpetual contestation amongst differing stakeholder agendas. It follows that decision-making about the ethical dimensions of 'AI' might therefore mirror this political analogy concerning the technology itself. In other words, if data-driven systems are themselves comprised of conflict and discord, to suggest that their ethical understanding should unproblematically reach consensus would seem to be disingenuous, or at least inauthentic to the kind of technology being considered. Indeed, Mouffe's agonism offers a way of viewing contestation and the lack of consensus as politically beneficial, and it is this positive rendition of conflict and tension that might be usefully taken forward in the ways educational institutions approach the ethics of 'AI'. For Mouffe, 'the search for consensus without exclusion and the hope for a perfectly reconciled and harmonious society have to be abandoned' (2013, 7), precisely because they are unattainable. As such, the vision of universal principles for ethical 'AI' might be seen as similarly vacuous, given that, at least for Mouffe (2013), exclusion from participation cannot be avoided, and any accord might be equally understood as the dominance of the powerful over the marginalised. Mouffe's response is 'a process of radicalizing democracy – the construction of more democratic, more egalitarian institutions' (2013, 8), forged through the maintenance of political contestation and the deliberate avoidance of ultimate consensus. In this way, Mouffe's agonism, or 'radical negativity', offers a direct challenge to the ostensibly widespread view that definitive principles and universal codes are a fundamental requirement to 'progress' with ethical 'AI'. Instead, universities in particular, grounded as has been claimed in an essential dynamic of productive tension (Goheen 1969), might resist the temptation to close down debate about the ethics of data-driven technologies through the publishing of definitive principles, frameworks, and guidelines, and look to maintain a process of deliberation and exchange as an essential element of a developing ethics-in-practice. This approach might therefore be termed 'radical participation', in

the sense that processes of contribution and dialogue are maintained as fundamental qualities of ethics itself rather than simply a means to an end. Such a commitment, however, ‘requires coming to terms with the lack of a final ground and the undecidability that pervades every order’ (Mouffe 2013, 12), and may therefore present a significant challenge for those that see the increased uptake of data-driven technology itself as an outcome of the principle-making process.

## Conclusions

Rather than simply being a distraction in comparison to more critical engagement with the social impacts of data-driven technologies, the recent proliferation of policies for ‘AI ethics’ has served to elevate discussions of (un)desirable technologized futures into the public domain, in ways academic research has perhaps failed to achieve. In this sense, the overabundance of ethical principles, guidelines, and frameworks deserves attention. However, in the apparent rush to ‘solve’ ethical dilemmas related to data-driven technologies, the protocolisation of ethics has overshadowed the rich diversity of approaches to moral philosophy, and impoverished the understanding of how society might address the increasing influence of intensive data processing.

Rather than simply following the trend, institutions of education might draw further on their own distinctive capacities and resources to better contend with the significant implications of the growing deployment of intensive data processing, especially where such systems are being directed at the educational activity itself. As has been well documented elsewhere (e.g., Williamson and Hogan 2020; Williamson 2021), the recent pandemic has greatly accelerated the uptake of data-driven technologies, such as educational software platforms, across the sector, where private companies offering such services have seen the public education system as ripe for the expansion of new commercial arrangements. This increasing private involvement in public education, given the differing agendas across educational and commercial domains, might be seen as necessitating an in-depth engagement with ethics. This may be particularly the case where so-called ‘AI’ systems – specifically those involved in not only collecting and processing significant volumes of educational data, but also feeding the resulting analysis back in ways designed to intervene in educational activity, or ‘personalise’ learning trajectories – are becoming more commonplace. As has been argued across the previous sections of this paper, the trend for ethical principles appears, despite what are undoubtedly good intentions, to ultimately serve an agenda of continued, if not amplified, technology acceptance. By portraying ethics as largely a matter of technical engineering, the emphasis on principles seems to offer a smooth path to solutionism but ends up masking much more established ethical, moral, and justice-oriented discussions, which are undoubtedly prevalent in the domains of education, and seem likely to be exacerbated by an uncritical adoption of data-driven decision-making. Further, by advancing principles as definitive, all-encompassing, and universal codes that can bring about the certainty of an ‘ethical AI’, the route to increasing datafication is rendered trouble-free. But in so doing, further debate, exchange, contestation, and indeed development of knowledge about the ethics of data-driven systems are rendered unnecessary. Yet these are the very functions, arguably, that ground the university and furnish it with the capacity to enrich the wider public understanding of our (un)ethical future with ‘AI’.

## Notes

1. See: <https://www.rolls-royce.com/sustainability/ethics-and-compliance/the-aletheia-framework.aspx>.
2. See: <http://lcfi.ac.uk/about/>.
3. See: <https://ainowinstitute.org/>.
4. See: <https://datasociety.net/>.
5. See: <https://partnershiponai.org/about/>.
6. See: <https://www.eismd.eu/ai4people/>.
7. See: <https://aiethicslab.com/big-picture/>.
8. See: <https://inventory.algorithmwatch.org/>.

9. See: <https://algorithmwatch.org/en/>.
10. See: <https://www.adalovelaceinstitute.org/project/mapping-ai-and-data-ethics/>.
11. See: <https://icarus.kumu.io/fluxus-landscape>.
12. See: <https://www.nesta.org.uk/data-visualisation-and-interactive/ai-governance-database/>.
13. See: <https://www.linking-ai-principles.org/>.
14. Floridi and Cowls (2019) suggest this process had substantial impact on the principles and guidelines developed by the European Commission and OECD, through the work of the AI4People coalition.
15. See: <https://www.just-ai.net/>.
16. See: <https://blogit.itu.dk/virteuproject/>.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Jeremy Knox  <http://orcid.org/0000-0003-3486-9487>

## References

- Bayne, S. 2015. "What's the Matter with 'Technology-Enhanced Learning'?" *Learning, Media and Technology* 40 (1): 5–20. doi:10.1080/17439884.2014.915851.
- Benkler, Y. 2019. "Don't Let Industry Write the Rules for AI." *Nature* 569: 161. doi:10.1038/d41586-019-01413-1.
- Biesta, G. J. J. 2009. "Good Education in an age of Measurement: On the Need to Reconnect with the Question of Purpose in Education." *Educational Assessment, Evaluation and Accountability* 21: 33–46. doi:10.1007/s11092-008-9064-9.
- Biesta, G. J. J. 2010. *Good Education in an Age of Measurement: Ethics, Politics, Democracy*. Boulder, CO: Paradigm.
- Bietti, E. 2021. "Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, January 2020, 210–219. doi:10.1145/3351095.3372860.
- Bigo, D., E. Isin, and E. Ruppert. 2019. *Data Politics: Worlds, Subjects, Rights*. London: Routledge.
- Borenstein, J., F. Grodzinsky, A. Howard, K. Miller, and M. Wolf. 2021. "AI Ethics: A Long History and a Recent Burst of Attention." *Computer* 54 (01): 96–102. doi:10.1109/MC.2020.3034950.
- Canca, C. 2020. "Operationalizing AI Ethics Principles." *Communications of the ACM* 63 (12): 1–6. [http://aiethicslab.com/wp-content/uploads/2020/08/Canca\\_AI-Principles\\_ACM.pdf](http://aiethicslab.com/wp-content/uploads/2020/08/Canca_AI-Principles_ACM.pdf). doi:10.1145/3430368.
- Crawford, K. 2016. "Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics." *Science, Technology, & Human Values* 41 (1): 77–92. doi:10.1177/0162243915589635
- Crawford, K. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.
- Dencik, L., A. Hintz, J. Redden, and E. Treré. 2019. "Exploring Data Justice: Conceptions, Applications and Directions." *Information, Communication & Society* 2(27): 873–881. doi: 10.1080/1369118X.2019.1606268.
- Fjeld, J., N. Achten, H. Hilligoss, A. Nagy, and M. Srikumar. 2020. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI." Berkman Klein Center Research Publication No. 2020-1, January 15. Available at SSRN: <https://ssrn.com/abstract=3518482> or doi:10.2139/ssrn.3518482.
- Floridi, L., and J. Cowls. 2019. "A Unified Framework of Five Principles for AI in Society." *Harvard Data Science Review Issue 1* (1): 1–14. doi:10.1162/99608f92.8cd550d1.
- Goheen, R. F. 1969. *The Human Nature of a University*. Princeton: Princeton University Press.
- Green, B. 2021. "Data Science as Political Action: Grounding Data Science in a Politics of Justice." *Journal of Social Computing* 2 (3): 249–265. doi:10.23919/JSC.2021.0029.
- Greene, D., A. Lauren Hoffmann, and L. Stark. 2019. "Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning." Proceedings from the Hawaii International Conference on System Sciences, January. doi:10.24251/HICSS.2019.258.
- Hagendorff, T. 2020. "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds & Machines* 30: 99–120. doi:10.1007/s11023-020-09517-8.
- Hao, K. 2021. "Big Tech's Guide to Talking About AI Ethics." *MIT Technology Review*, April 13. <https://www.technologyreview.com/2021/04/13/1022568/big-tech-ai-ethics-guide/>.
- IEAIE (Institute for the Ethical AI in Education). 2021. "Report." <https://www.buckingham.ac.uk/wp-content/uploads/2021/03/The-Institute-for-Ethical-AI-in-Education-The-Ethical-Framework-for-AI-in-Education.pdf>.

- IEEE. 2016. “The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. Ethically Aligned Design: A Vision for Prioritizing Wellbeing with Artificial Intelligence and Autonomous Systems, Version 1.” [http://standards.ieee.org/develop/indconn/ec/autonomous\\_systems.html](http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html).
- IEEE. 2019. “The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition.” <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html>.
- EOPNSTCCT (Executive Office of the President National Science and Technology Council Committee on Technology). 2016. “Preparing for the Future of Artificial Intelligence.” [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf).
- Fraser, N. 2008. “Reframing Justice in a Globalizing World.” In *Adding Insult to Injury: Nancy Fraser Debates Her Critics*, edited by K. Olson, 273–294. London: Verso.
- Jobin, A., M. Ienca, and E. Vayena. 2019. “The Global Landscape of AI Ethics Guidelines.” *Nature Machine Intelligence* 1: 389–399. doi:10.1038/s42256-019-0088-2.
- McNamara, A., J. Smith, and E. Murphy-Hill. 2018. “Does ACM’s Code of Ethics Change Ethical Decision Making in Software Development?” ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE).
- Mittelstadt, B. 2019. “Principles Alone Cannot Guarantee Ethical AI.” *Nature Machine Intelligence* 1 (11): 501–507.
- Morozov, E. 2013. *To Save Everything Click Here: Technology, Solutionism and the Urge to Fix Problems that Don’t Exist*. London: Penguin.
- Mouffe, C. 2013. *Agonistics: Thinking The World Politically*. London: Verso.
- Schwab, K. 2016. *The Fourth Industrial Revolution*. London: Portfolio Penguin.
- Tasioulas, J. 2021. “Reclaiming a Broad and Foundational Understanding of Ethics in the AI Domain, with Radical Implications for the Re-ordering of Social Power.” Ada Lovelace blog. <https://www.adalovelaceinstitute.org/blog/role-arts-humanities-thinking-artificial-intelligence-ai/>.
- Tucker, E. 2022. “Artifice and Intelligence in Tech Policy Press.” March 17. <https://techpolicy.press/artifice-and-intelligence/>.
- UNESCO. 2021. “Draft Text of the Recommendation on the Ethics of Artificial Intelligence.” <https://unesdoc.unesco.org/ark:/48223/pf0000377897>.
- Vallor, S. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press.
- Vallor, S. 2021. “Challenging and Redrawing Framings of Technology to Serve Human Flourishing and Justice.” Edinburgh Innovations. <https://edinburgh-innovations.ed.ac.uk/news/mobilising-the-intellectual-resources-of-the-arts-and-humanities>.
- Whittaker, M., K. Crawford, R. Dobbe, G. Fried, E. Kazianas, V. Mathur, S. M. West, R. Richardson, J. Schultz, and O. Schwartz. 2018. “AI Now Report.” [https://ainowinstitute.org/AI\\_Now\\_2018\\_Report.pdf](https://ainowinstitute.org/AI_Now_2018_Report.pdf).
- Williamson, B., and A. Hogan. 2020. “Commercialisation and Privatisation in/of Education in the Context of COVID-19.” Education International. [https://issuu.com/educationinternational/docs/2020\\_eiresearch\\_gr\\_commercialisation\\_privatisation?fr=sZDjkYjE1ODA2MTQ](https://issuu.com/educationinternational/docs/2020_eiresearch_gr_commercialisation_privatisation?fr=sZDjkYjE1ODA2MTQ).
- Williamson, B. 2021. “Education Technology Seizes a Pandemic Opening.” *Current History* 120 (822): 15–20. doi:10.1525/curh.2021.120.822.15.
- Yeung, K., A. Howes, and G. Pogrebná. 2020. “AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing.” In *The Oxford Handbook of AI Ethics*, Oxford University Press, edited by M. Dubber and F. Pasquale, 76–106. Oxford: Oxford University Press. Available at SSRN: <https://ssrn.com/abstract=3435011> or doi:10.2139/ssrn.3435011.
- Zeng, Y., E. Lu, and C. Huangfu. 2019. “Linking Artificial Intelligence Principles.” Proceedings of the AAAI Workshop on Artificial Intelligence Safety. <https://arxiv.org/pdf/1812.04814.pdf>.