International Journal of Artificial Intelligence in Education (2022) 32:527–563
https://doi.org/10.1007/s40593-021-00270-2

ARTICLE

Education for AI, *not* AI for Education: The Role of Education and Ethics in National AI Policy Strategies

Daniel Schiff[1]

Accepted: 26 July 2021 / Published online: 2 September 2021

Abstract

As of 2021, more than 30 countries have released national artificial intelligence (AI) policy strategies. These documents articulate plans and expectations regarding how AI will impact policy sectors, including education, and typically discuss the social and ethical implications of AI. This article engages in thematic analysis of 24 such national AI policy strategies, reviewing the role of education in global AI policy discourse. It finds that the use of AI in education (AIED) is largely absent from policy conversations, while the instrumental value of education in supporting an AI-ready workforce and training more AI experts is overwhelmingly prioritized. Further, the ethical implications of AIED receive scant attention despite the prominence of AI ethics discussion generally in these documents. This suggests that AIED and its broader policy and ethical implications—good or bad—have failed to reach mainstream awareness and the agendas of key decision-makers, a concern given that effective policy and careful consideration of ethics are inextricably linked, as this article argues. In light of these findings, the article applies a framework of five AI ethics principles to consider ways in which policymakers can better incorporate AIED's implications. Finally, the article offers recommendations for AIED scholars on strategies for engagement with the policymaking process, and for performing ethics and policy-oriented AIED research to that end, in order to shape policy deliberations on behalf of the public good.

Keywords Artificial intelligence · Ethics · Policy · Social implications of technology

"*Around the world, virtually no research has been undertaken, no guidelines have been provided, no policies have been developed, and no regulations have been enacted to address the specific ethical issues raised by the use of Artificial Intelligence in Education. As a field (while we apply our university research regulations),*

Daniel Schiff
schiff@gatech.edu

[1] School of Public Policy, Georgia Institute of Technology, Atlanta, GA, USA

we are working without any fully-worked out moral groundings specific to the field of AIED."–(Holmes et al. 2018).

A Policy Window for AI in Education

The social, ethical, legal, and economic implications of artificial intelligence (AI) have captured global attention in the last decade, opening up a new window in which to shape AI policy. Researchers, policymakers, and the public have become increasingly aware of issues related to labor displacement (Acemoglu and Restrepo 2019), autonomous vehicles (Bagloee et al. 2016), algorithmic bias (Caliskan et al. 2017), and privacy (Manheim and Kaplan 2018), amongst other issues (Mittelstadt et al. 2016). As a result, new organizations and collaborations, like AI Now, the Partnership on AI, and the IEEE Global Initiative on the Ethics of Autonomous and Intelligent Systems have formed in recent years. New academic conferences like AAAI/ACM's AI, Ethics, and Society, and the ACM's Fairness, Accountability, and Transparency (ACM FAccT) have emerged along with increased research and public funding to address ethical issues like explainability and bias in algorithms (Cath et al. 2018; Chae 2020).

In turn, governments, corporations, and non-governmental actors have published more than 100 AI ethics codes, frameworks, and policy strategies in recent years (Fjeld et al. 2020; Jobin et al. 2019; Schiff et al. 2021). This includes more than 30 countries and regions that have produced national AI policy strategies, while 20 or more have developed preliminary frameworks or task forces with plans to produce formal strategies in the near future (Zhang et al. 2021). This article examines 24 such English-language national AI policy strategies produced between 2016 and the beginning of 2020. These policy documents represent a valuable source of information for understanding how governments around the world are conceiving of AI's impacts, benefits, and risks. In particular, each document discusses the implications and strategies surrounding AI in different policy sectors—such as transportation, healthcare, and education—allowing for evaluation of global discourse on the role of education in AI. This supports an examination of how key policy decision-makers around the world conceive of the relationship between AI and education. For example, do policymakers view AI as a tool to improve education or are they focused on education as a tool to improve AI? Are they concerned about possible ethical harms resulting from improper uses of AI in educational systems, or have these issues not captured their attention? Yet another topic of interest is how policymakers imagine the relationship between ethics and policy, an issue this article turns to next.

The Role of AI Ethics in AI Policy

National AI policy strategies typically discuss a similar set of items: the potential of AI to drive innovation, economic transformation, and growth; the need to capitalize on AI to advance national competitiveness; how AI will or could impact various policy sectors (e.g., agriculture, transportation, finance, manufacturing, etc.); and a review of social and ethical concerns, in some cases addressed by a set of proposed ethical governance principles (Cath et al. 2018; Daly et al. 2019). Governments have already committed at least 10 billion in funding, and as much as 25 billion, over the next few years (Allen 2019; Dutton et al. 2018; Future of Life Institute 2020).[1] This level of investment is understandable, given predictions that AI, as a general-purpose technology, can produce additional short-term economic activity of 13 trillion or more by 2030, approximately 1.2% additional GDP growth annually (McKinsey Global Institute, 2018). As a result, national AI policy strategies are an excellent source to understand how some of the most important global actors make sense of AI-related issues and how they intend to respond to them. Perhaps apart from a handful of large multinational corporations, government actors exert the greatest influence over how society will respond to AI, especially in public domains like education. Their national AI strategies are arguably the clearest reflection of the understanding and priorities of top global decision-makers.

Yet, as Holmes et al. (2018) have observed in the context of AI in education (AIED), "Around the world, virtually no research has been undertaken, no guidelines have been provided, no policies have been developed, and no regulations have been enacted to address the specific ethical issues raised by the use of Artificial Intelligence in Education." It is therefore of great interest that the large majority of national policy strategies include explicit sections focused on the ethics of AI (Schiff et al., 2020a, 2020b). This surprising fact may be explained by significant scholarly and public attention attached to AI ethics in the last decade. For example, in discussing the application of AI to healthcare, national AI policy strategies consider ethical issues such as privacy, informed consent, black-box algorithms, and how physician–patient interaction may be influenced by AI (Char et al. 2018). In the case of transportation and autonomous vehicles, numerous AI policy strategies reflect on the "trolley problem" thought experiment, an ethical dilemma often applied in order to weigh and prioritize harms related to pedestrian versus driver injuries and fatalities (Bagloee et al. 2016). National AI strategies have drawn on ethics principles developed by civil society as well as promoted home-grown ethical principles such as the EU's (2019) Guidelines for Trustworthy AI.

The inclusion of significant ethical discourse in AI policy strategies is promising: it supports the possibility that policymakers will deliberately incorporate ethical principles and concerns into new initiatives, funding mechanisms, and regulatory strategies. For example, in the case of self-driving cars, this includes reflection on ethical issues of accountability, liability, and responsibility, including how these ethical questions will be incorporated into regulatory frameworks (Boeglin, 2015). In the case of healthcare, this includes discussion of data governance, and the trade-off between building better AI models that require more data and preserving patient privacy (Stahl and Coeckelbergh 2016). Simply, awareness of ethical implications and

[1] Australia, Canada, Singapore, Denmark, Denmark, Taiwan, France, EU, UK, and South Korea have committed nearly 8 billion with the US contributing or planning to contribute at least 4 billion and China at least 14 billion. These are moving targets and low-end estimates, especially as private investment constitutes an even greater sum.

risks enhances public and policy discourse. In turn, more thoughtful discourse can support responsible governance, development, and implementation of AI.

The linkage between effective policy and attention to ethics is not surprising if we understand policy as one of the greatest levers a society has to promote its ethical worldview. It is also important to note that ethical implications can be positive in nature, such as a technological tool that reduces inequality, or negative, such as a tool that leads to violations of civil or political rights. In either case, ethical concerns need not be limited to micro-ethical benefits or harms associated with a single product or user, but rather should be understood to address fundamental societal goals like mitigating poverty, improving well-being, and safeguarding human rights. On this viewpoint, policymakers must therefore attend to mixed ethical implications if they wish to promote social goods and prevent policy harms through effective policymaking. Conversely, understanding the broader policy impacts of a set of technologies like AIED should support more careful consideration of possible ethical implications as well.

The remarkable attention to AI ethics generally in national AI policy strategies is thus a good starting point, but it does not imply that there is careful attention to the policy or ethical implications of AIED in particular. And while prior work has assessed the role of ethical principles at a general level in AI policy documents (Cath et al. 2018; Daly et al. 2019; Dutton et al. 2018; Gibert et al. 2018), to my knowledge, no prior work looks intensively at the topic of education, nor at how AI ethics considerations are applied to education. This article therefore considers and provides empirical analysis to answer the following question: *how do national-level AI policy strategies understand the role of education in AI policy?*

Overview

After introducing the methodology, I analyze the aforementioned national AI policy strategies in terms of their discussion of education. In the section on The Role of Education in AI Policy Strategies, I provide evidence surrounding five primary themes: three related to the use of *Education for AI* (with education understood as training), and two related to the use of *AI for (or in) Education* (AIED). The findings indicate that traditional topics associated with AIED, such as the role of intelligent tutoring systems in improving teaching and learning, receive relatively little attention in these key policy documents. Instead, global policymakers are largely concerned with education as an instrumental strategy to prepare an AI-ready workforce for labor disruption and to increase the supply of high-skill workers, especially 'AI talent.' Despite the fact that national AI policy strategies represent significant breadth and depth in their reflection on various social and ethical issues (Fjeld et al. 2020; Jobin et al. 2019), and address a sweeping scope of policy sectors (Schiff et al. 2021), traditional AIED is neglected and consideration of AIED's ethical implications is almost entirely absent. Based on the findings, I suggest that classical topics in AIED and AIED's social and ethical implications may be neglected in policy discourse because they are 'squeezed out' by an instrumental frame of *Education for AI*, an economic logic prioritizing growth and competitiveness.

In light of the 'non-finding' that AI policy strategies are not engaged with AIED's ethical implications, the section Bringing AIED Ethics into Policy proposes ways in which AI policymakers could take these concerns more seriously. Specifically, I adopt a taxonomy of AI ethics principles developed by Floridi and Cowls (2019): beneficence, non-maleficence, autonomy, justice, and explicability. This framework is derived from a review of six high-profile initiatives, bears substantial similarities with other scholarly efforts to AI create ethics typologies (Fjeld et al. 2020; Jobin et al. 2019; Zeng et al. 2018), and is highly familiar given its overlap with traditional bioethics principles (Beauchamp and Childress 1979). For each principle, I discuss relevant ethics issues for AIED and offer considerations for AI policy and regulation.

Finally, to help facilitate the engagement of AIED scholars along these issues, the Recommendations for AIED Researchers section reviews strategies for engagement with the policymaking process. I argue that policy-engaged AIED researchers must not only increase their research on AIED ethics, but also on AIED policy, drawing on topics and methods in traditional education policy research that move beyond classroom and demonstration projects to larger units of analysis, addressing issues like school funding and teacher preparation. While it is a disappointment that national AI policies have not taken sufficient account of the importance of AIED and its social and ethical implications, is not too late to inform discourse and positively shape the direction of emerging AI policy. The window to act remains open.

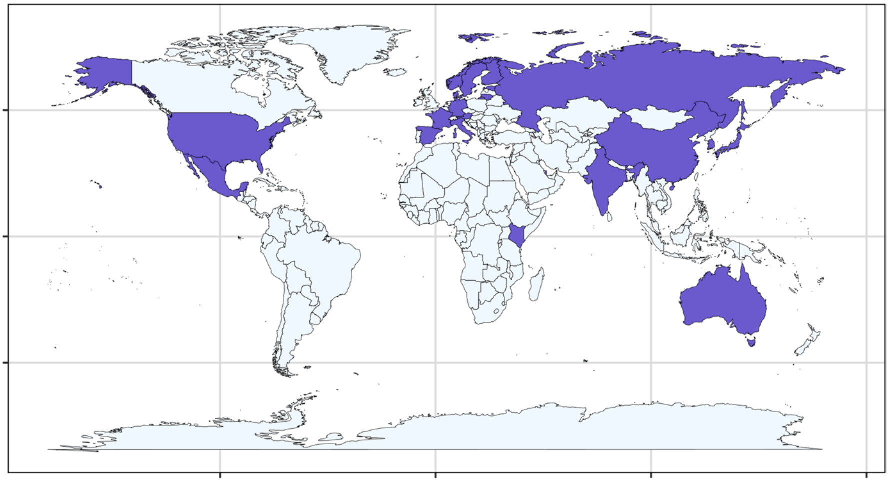
Methodology

Below, I briefly introduced the methodology used to examine national AI policy strategies. For interested readers, more extensive details surrounding data collection, screening, coding, and analysis strategies, as well as limitations, are included in Appendix 1.

Data Collection and Screening Process

As the literature here is best understood as "grey literature" (Mahood et al. 2014), the data collection process relied on linkhubs, web searching via Google, and hand searching. An initial search, aided by lists maintained by the Future of Life Institute (2020) and Tim Dutton (2018), produced 30 countries with candidate national AI strategies, all published between the middle of 2016 and beginning the 2020. These strategies are predominantly developed by wealthy G-7 and OECD countries, such as the United States, United Kingdom, France, Germany, Italy, Japan, Finland, Norway, and others. Other important global actors (Cooper and Flemes 2013) like India, China, and Russia, and a set of smaller countries like Singapore and Malta, have also produced AI policy strategies.

The screening process involved identifying criteria for types of documents, publication language, the population and phenomena of interest, and time period (Stern et al. 2014). First, I exclude countries without completed documents, such as



Notes: Author's calculations as of January 2020. Sample includes 24 completed English-language national AI strategies.

Fig. 1 National AI Policy strategies analyzed

countries that have only announced taskforces, general initiatives, created websites, or made financial commitments. Next, I exclude countries where no English-language document is available, though allowing for a high-quality official or unofficial translation. This does not result in a significant reduction of sample size. Finally, for countries that have produced multiple relevant AI policy drafts or documents, I select a single such document, typically the newest and most comprehensive. Of an initial list of 76 public sector documents, the screening process resulted in a final sample size of 24 national AI policy strategies, presented below in Fig. 1.[2]

Coding and Analysis Approach

I manually analyze each document (approximately 1491 pages total) for any discussion of education, construed broadly. This includes youth and adult education, training and re-skilling, investing in educational systems, public education and awareness, the need to develop more AI talent (e.g., computer scientists, engineers), and social and ethical issues related to AI and education. Based on the overarching research question, prior subject matter knowledge, and brief review of the documents, I developed a preliminary codebook (Miles et al. 2014) consisting of 11 topics of interest. Next, I applied the codebook to five randomly-selected documents to test whether the thematic schema could be applied straightforwardly and could

[2] I do not include documents produced by intergovernmental bodies, such as the United Nations or European Union. While these documents are similar in nature, they are less tied to direct national institutions and strategies, and are therefore less analogous to the other documents. For further details and rationales regarding data collection and inclusion/exclusion criteria, please see Appendix 1.

capture the full range of issues in the documents relevant to the article's conceptual scope (Roberts et al. 2019).

Based on this initial test, I modified, removed, and collapsed some codes, resulting in an inductively-iterated final set of seven coding topics surrounding AI and education, presented below (Thomas 2006).[3] Next, I apply a simple form of content analysis, evaluating each document for the presence or absence of each topic (White and Marsh 2006). Documents where the topic is reflected are indicated in green; yellow indicates borderline cases where discussion of the topic is too narrow or ambiguous to safely code as green.[4] Red means the topic is absent. The determination of each code was made both in consideration of how each document treats different topics, as well as how different documents treat the same topic.[5]

While coding each document, I also extracted representative quotes and organized them into a research memo, organized by topic. From this, I synthesized insights from the frequency and character of these topics, applying a thematic analytic approach (Castleberry and Nolen 2018) to identify major findings. This interpretive exercise involves considering second-order meanings or explanations for the patterns identified in the (Miles et al. 2014), including the finding that AIED's ethical implications are neglected. I present results for each topic, along with interpretation of key findings within and across topics, to support a broader discussion of the role of education and ethics in AI policy in the following sections.

One limitation of the approach is that the formation of themes and subsequent coding and analysis were performed by a single researcher, making it not possible to, for example, assess inter-rater reliability. However, a single coder approach can be appropriate and, in cases, even preferrable in qualitative research (Harding and Whitehead 2013) and multiple researchers may not be necessary to provide sufficient consistency and credibility. Relatedly, it is common in qualitative research to consider alternative criteria to traditional quantitative criteria of validity and reliability. Lincoln and Guba (1985) propose credibility, transferability, dependability, and confirmability as alternatives to internal validity, external validity, reliability, and objectivity, respectively. To satisfy these criteria, the analysis here includes triangulation across multiple sources, numerous quotations to provide descriptive evidence, and examines publicly-available data (as opposed to private interview data, for example) which allows for verification (Lincoln and Guba 1986). Further, as compared to other content analysis research projects, the coding strategy here involves a short number of coding categories (seven as opposed to dozens), which

[3] Further details about codebook development and iteration are available in Appendix 1.

[4] For example, Malta's (2019) document initially notes that AI for healthcare may be amongst the highest impact projects and "most pressing domestic challenges" worthy of prioritization, but it does not proceed to include any substantive discussion or a subsection on healthcare. In comparison to the document's discussion of other topics, and in comparison to other countries' AI policy documents that discuss healthcare in more depth, this relatively more narrow treatment of the topic led to coding it as yellow. Similarly, Russia's (2019) brief mention of using AI to "[improve] the quality of education services" does not provide enough detail to be clear about the role of AIED as a potential tool for teaching and learning, and so is considered too ambiguous to code as either green or red.

[5] Any errors are the sole fault of the author.

are not coded at a fine level of detail and are fairly easy to conceptually separate. Nevertheless, the analysis is necessarily subjective.

Analysis Topics

The codebook development process resulted in five primary topics and two additional coding categories. The five primary topics are divided into two domains, which I call *Education for AI* and *AI for Education* (or AIED), respectively:

Education for AI (i.e., training)

- *Training AI Experts*: discussion of developing future AI practitioners, such as computer scientists and engineers.
- *Preparing the Workforce for AI*: discussion of education and training efforts to help workers adapt to labor disruption due to AI.
- *Public AI Literacy*: discussion of the need to educate the broader public about AI.

AI for Education (i.e., AIED)

- *Teaching and Learning*: discussion of AI-based teaching and learning tools such as intelligent tutoring systems, pedagogical agents, and predictive assessments.
- *Administrative Tools*: discussion of AI used to support administration in educational systems, for example, to make admission, promotion, or graduation decisions.

The above classification cannot parse the many AIED applications at a fine level of detail, which reflects that documents often addressed these topics in a more generic way. Yet it is enough to show attention or inattention of the AI policy strategies to AIED. An important note is that, despite an initial attempt to code for discussion of AIED ethics specifically given its importance to this study, discussion of these topics was too rare to justify having as a theme.[6] Finally, I code for two more categories to provide context.

- *Education as Priority Topic*: indication of whether education is a priority focus for the national policy strategy. This can be either stated explicitly or demonstrated by inclusion of a specific section or significant discussion of education.
- *AI for Healthcare*: discussion of the application of AI to healthcare, for example, in medical imaging and diagnosis, drug dosing, or surgical interventions. I include this topic as a comparison point, to ascertain whether national AI strategies treat education and healthcare similarly.

[6] I nevertheless captured these mentions in the research memo and discuss them at the end of this section.

Table 1 The role of education in national AI policy strategies

Country	Document Name	Date	Education for AI			AI for Education		Context	
			Training AI Experts	Preparing Workforce for AI	Public AI Literacy	Teaching and Learning	Admin. Tools	Education as Priority Topic	AI for Health
Australia	Artificial Intelligence: Australia's Ethics Framework (A Discussion Paper)	#########	x	x	x		~		x
Austria	Shaping the Future of Austria with Robotics and Artificial Intelligence	#########	x	x	x			x	x
China	A next-generation artificial intelligence development plan	#########	x		x	x	x	x	x
Denmark	National Strategy for Artificial Intelligence	#########	x	x	x			x	x
Estonia	Estonia's national artificial intelligence strategy 2019-2021	#########	x		x			x	
Finland	Finland's Age of Artificial Intelligence: Turning Finland into a leading country in the application of artificial intelligence	#########	x	x	x			x	x
France	For a Meaningful Artificial Intelligence: Towards a French and European Strategy	#########	x	x				x	x
Germany	National Strategy for Artificial Intelligence: AI Made in Germany	#########	x	x	x			x	x
India	National Strategy for Artificial Intelligence: #AIForAll	#########	x	x	x	x	x	x	x
Italy	Artificial Intelligence at the service of citizens	#########	x	x	x	x		x	x
Japan	Artificial Intelligence Technology Strategy	#########	x	x					x
Kenya	Emerging Digital Technologies for Kenya: Exploration & Analysis	#########		x	x	x	x	x	x
Lithuania	Lithuanian Artificial Intelligence Strategy: A Vision of the Future	#########	x	x	x			x	x
Malta	Malta: Towards An AI Strategy	#########	x	x	x	x	x		~
Mexico	Towards an AI Strategy in Mexico: Harnessing the AI Revolution	#########	x	x	~	~		x	x
Norway	National Strategy for Artificial Intelligence	#########	x	x	x			x	x
Qatar	Blueprint: National Artificial Intelligence Strategy for Qatar	#########	x	x	x			x	x
Russia	National Strategy for the Development of Artificial Intelligence Over the Period Extending up to the Year 2030	#########	x	x	x	~		x	x
Singapore	National Artificial Intelligence Strategy: Advancing Our Smart Nation Journey	#########	x	x	x	x		x	x
South Korea	Mid-to Long-term Master Plan in Preparation for the Intelligent Information Society: Managing the Fourth Industrial Revolution	#########	x	x	x	x		x	x
Spain	Spanish RDI Strategy in Artificial Intelligence	#########	x	x	x	x	x	x	x
Sweden	National approach for artificial intelligence	#########	x	~				x	~
United Kingdom	AI in the UK: Ready, Willing and able?	#########	x	x	x			x	x
United States	The National Artificial Intelligence Research and Development Strategic Plan	#########		x		x		x	x

The Role of Education in AI Policy Strategies

Table 1, below, presents a document-level global analysis of the key AI education topics.

Education for AI, not AI for Education

First, the analysis demonstrates that education is a priority topic internationally, stated explicitly or given significant attention in dedicated sections for 21 of the 24 national AI policy strategies. But what do the documents' authors and policymakers mean when they refer to education in the context of AI?

A key finding of the analysis is that the documents are largely referring to *Education for AI*, substantially surpassing discussion of *AI for Education*. All 24 documents contain discussion of at least one of the three *Education for AI* topics. At least 22 documents discuss two of the three *Education for AI* topics, and at least 16 discuss all three topics. In contrast, only 9 documents discuss *AI for Education* in any depth (three have borderline engagement). Only five documents of 24 discuss both *AI for Education* topics: China, India, Kenya, Malta, and Spain.

There are two prima facie reasons for this result worth briefly reviewing and rejecting. The first is that national AI policy strategies might treat all policy sectors in a similar way, by only considering how social and economic infrastructure can advance AI innovation, rather than by considering how AI can itself be used to promote change in these policy sectors. However, the fact that at least 21 of 24 documents explicitly discuss how AI can be applied to healthcare undermines this case. Simply, there is no mandate that the documents discuss *Education for AI* while neglecting *AI for Education*; it is a choice of the authors which is dissimilar to their treatment of *AI for Healthcare*. In fact, education's bifurcated role appears to be a unique case, as many documents discuss the use of AI not only for healthcare, but also for transportation, agriculture, finance, and many other sectors.

Second, one might claim that there are other sources through which one can examine national approaches to education, such as policy documents and guidance produced by national education agencies. While examining additional sources can no doubt provide a more comprehensive picture, examining national AI policy strategies provides a more direct look at the highest-level priorities—strategic and funding—of top decision-makers.[7] These documents reveal whether global leaders are aware of, invested in, or concerned about AIED and its potential uses as well as ethical implications. This then shapes how education is incorporated into policy agendas, funded, and regulated by additional governmental entities, including in terms of its ethical benefits or risks. How these particular documents frame issues like AI in education and healthcare is therefore of critical relevance. Next, I discuss the three *Education for AI* topics.

Education for AI: Training AI Experts

Developing more expert AI computer scientists and engineers is a dominant priority in most countries. Nearly all countries discuss "a shortage of engineers trained in artificial intelligence" (Villani et al. 2018) which suggests the need to "address the shortfall" (Smart Nation and Digital Government Office, Singapore 2019). In response, for example, Singapore proposes a nine-month AI Apprenticeship Programme, Lithuania its first ever Master's program in artificial intelligence (Ministry of the Economy and Innovation 2019), and Mexico its Wizeline AI Academy, including tuition-free coursework and integrated mentorship (British Embassy in Mexico, Oxford Insights, and C Minds 2018). Emphasizing the stakes of the competition for AI experts, Russia argues that by 2030, its "institutions of higher learning must occupy leading positions the world over" in AI (Office of the President, Russia 2019). This straightforward ambition is articulated by 22 of 24 countries.

Education for AI: Preparing the Workforce for AI

Next, nearly all of the national policy strategies (at least 21 of 24) also include some discussion of labor displacement risks due to automation, and how skill and task content will change (Autor, 2015), urging a need for new policy around education and training (West 2018). As described in Mexico's AI policy strategy, which itself did a review of other countries' documents:

> Every existing AI strategy contains sections examining capacity, skills, and education. These strategies vary in scope, but tend to recognise the importance of teaching digital skills from an early stage in the national curriculum, as well as emphasising the need for lifelong learning to enable workforces to adapt to new developments in technology. The majority of strategies recognised that developing high quality, in country AI expertise is vital for a country to remain

[7] Note that other policy sectors, like healthcare, also have dedicated agencies and other policy documents, but nevertheless receive more attention than education in AI policy strategies.

at the forefront of the AI revolution (British Embassy in Mexico, Oxford Insights, and C Minds 2018).

This means that "it is not enough that the part of the population involved in technology development is correctly educated and trained. It is essential that people who hold jobs in which artificial intelligence must be utilised understand the possibilities and limitations that artificial intelligence adds to work tasks" (Ministry of Economic Affairs and Employment, Finland 2017). In this sense, the entire workforce must be educated about AI, not just future machine learning experts.

Preparing the workforce again includes focusing on STEM skills, including general digital competencies, as well as non-cognitive and social-emotional skills, like creativity and collaboration. Policymakers imagine a shift to the adult education system as well given a need to "equip, reskill and upskill workers in every sector of society" (Office of the Prime Minister, Malta 2019). This includes developing life-long learning programs, reforming curricula, and "adapting to the changing needs of the labour market" (Ministry of Finance and Ministry of Industry, Business and Financial Affairs, Danish Government 2019). Workforce preparation is thought especially critical "for cases where AI systems provide professional support for skilled staff in their everyday work, for example in the area of education and nursing care" (Federal Ministry of Education and Research, Germany 2018). This reflects a clear economic and instrumental logic for education's role vis-à-vis AI.

Education for AI: Public AI Literacy

Finally, policymakers want to educate the public about AI. Their reasons for doing so are somewhat more complex and include: increasing public awareness and engagement in policy discourse, increasing public trust of AI, and setting the stage for a stronger workforce and for more AI experts (the two topics Education for AI: Public AI Literacy).

Most generally, educating the public about AI is important so that the public can "flourish mentally, emotionally and economically alongside artificial intelligence" (House of Lords, Select Committee on Artificial Intelligence 2018). However, for citizens to meaningfully shape a future with AI as part of policy and public dialogue, they must be able to assess its risks and opportunities (Dawson et al. 2019). In this sense, "educating larger numbers of people on the principles of AI and algorithms as a precondition for an inclusive policy is vital" (Villani et al. 2018). This first purpose for educating the public therefore surrounds civic engagement and public participation.

Next, policymakers are concerned that "a lack of knowledge and acceptance on the part of the general public could impede the development and dissemination of the technology" (Federal Ministry of Education and Research, Germany 2018). This means "promot[ing] AI as an enabler rather than something to fear," which can be enabled through "public talks and discussion forums where stakeholders can share experiences, thoughts and ideas around AI and its impacts on society" (Office of the Prime Minister, Malta 2019). Educating the public is thus also about building public and especially consumer trust and alleviating fears in order to enable AI innovation, again an economic logic.

However, achieving these impacts requires more than public discussion, and features significant transformation from K-12 educational systems through adult formal education. Singapore, through its AI for Everyone, AI for Kids and AI for Students initiative, aims to provide AI literacy courses to 100,000 adults and youth by 2025 (Smart Nation and Digital Government Office, Singapore 2019). Similarly, Finland has rolled out a free series of courses known as Elements of AI, again aimed at the public. Qatar's AI + X paradigm argues that AI "should be an integral part of the curriculum at all educational levels in all disciplines" (Qatar Center for Artificial Intelligence 2019). This strategy implies "firmly establishing basic AI knowledge as a part of syllabi, not just in computer science, but also in other natural, social, cultural, media-related and engineering sciences for initial and further vocational training, where appropriate" (Federal Ministry of Education and Research, Germany 2018). Policymakers imagining scaffolding curricula from kindergarten through adult education, starting with "foundational concepts to spark their interest in AI," followed by "basic AI competencies and literacy" and finally more complex learning during higher education (Smart Nation and Digital Government Office, Singapore 2019).

While also building general public awareness and trust, these proposed strategies often explicitly build towards the other *Education for AI* Topics of workforce development and training of future AI experts, as "it is necessary for [citizens] to have the competence to relate to [AI] and its different fields of action" (Ministry of Science, Innovation and Universities, Spain 2019). Here, digital literacy and STEM skills, as well as social-emotional and non-cognitive skills such as creativity and collaboration are emphasized (Korean Ministry of Science, ICT and Future Planning, 2016). These non-STEM skills are also thought important in supporting the ability of human workers to retain economic value in the workplace, as skills like creativity and emotional intelligence are considered harder to automate (Frey and Osborne, 2017). The quest to educate the public on AI may be in part about public civic engagement, but it is also about ensuring workforce and AI expert development. In combination, these three purposes of *Education for AI* manifest a clear economic logic. This suggests an initial possible explanation for why *AI for Education* (AIED), below, appears to be a significantly lesser priority.

AI for Education: Teaching and Learning

A modest number of countries discuss *AI for Education*, generally in the context of teaching and learning tools such as intelligent tutoring systems, automating homework grading, providing support outside of the classroom, adaptive assessment, new forms of collaborative learning, and companion robots. The countries that do discuss AIED are China, India, Italy, Kenya, Malta, Singapore, South Korea, Spain, and the United States. While several discussions of AIED are extremely brief, more robust discussions are offered by India, Kenya, Singapore, and Spain. All of these countries emphasize personalization as a 'key' aspect of AIED, including support of diverse students, such as those with dyslexia or with functional illiteracy (Department of Public Administration, Italy 2018). In turn, these efforts at personalization, such as through an AI Learning Companion, can support "students' holistic development" (Smart Nation and Digital Government Office, Singapore 2019) and ownership of their own learning, also enabling teachers to better support student learning in the process.

Of note, India and Kenya are two of the few lower-middle income countries[8] analyzed in this study that have produced AI policy strategies, and they offer noticeably different perspectives compared to those offered by higher-income countries. India, which describes education as a priority AI policy sector, notes that current "quality and access issues observed in the Indian education sector" motivate the need for more AI applications in education (NITI Aayog, India 2018). Similarly, policymakers in Kenya argued that personalized learning is "a major problem that normal schooling has not been able to address" that can be solved by implementing intelligent tutoring systems to "identify each individual child's competency and deliver the right lesson at the right time" (Blockchain & Artificial Intelligence Taskforce, Kenya 2019). The discourse in wealthier countries like Malta, Singapore, and Spain is somewhat different. Rather than addressing fundamental gaps in a nation's educational system (a supplement), they imagine AIED tools more as a complement which "can free up capacity for teachers to focus on and other critical tasks by supporting in areas such as automating feedback to delivery to students, student assessments, and virtual teaching assistants" (Office of the Prime Minister, Malta 2019).

These results suggest that a country's economic status in part explains its manner of engagement with AIED, as national wealth can shape key resource constraints or capacities in terms of teacher development, education infrastructure, and relevant surrounding infrastructure (e.g., health, roads, electricity, internet), as well as affecting baseline student needs and governance (Schendel and McCowan, 2016). Yet, while a full review of national education policy (Heyneman 2004) and reasons for technological innovation is beyond the scope of this article, there are other reasons why countries may be innovators in education technology, such as levels of human capital, foreign direct investment, and successful regional agglomeration (Lee 2001; Porter 1996). Some of these factors may help to explain why Kenya, for example, is the "heart of East Africa's technology ecosystem" (Giuliani and Ajadi, 2019). Or, for example, Malta's engagement with AIED may exemplify a small country innovation strategy focused on advanced technology (Roolaht 2012).

AI for Education: Administrative Tools

The full range of AIED tools discussed goes beyond teaching and learning tools, though not in depth, as even fewer countries address the topic of AIED in the context of administrative tools. Here, the most robust discussions are by India, Malta, and Spain. Applications of AIED to educational administration include system-wide

[8] See The World Bank Country and Lending Groups classification for classifications by national income.

uses that go beyond individual teacher support tools. For example, India proposes to explore automation of teacher hiring and transfer systems based on teacher qualifications and preferences from schools, a strategy to improve "plugging of gaps in teacher distribution more effectively" (NITI Aayog, India 2018). India also considers the use of professional development tools for educators "based on their performance, identification of their knowledge and skill gaps" which can be "continuously adapted as teachers' skills and concepts improve" and finally "predictive tools to inform pre-emptive action for students predicted to drop out of school" (NITI Aayog, India 2018). Finally, policymakers in Malta also imagine AIED for more general administrative tasks, such as "timetable preparation, facilities booking, and predicting inspections" (Office of the Prime Minister, Malta 2019).

In summary, while nine countries discuss some version of *AI for Education*, arguably only four or five discuss it in any depth beyond superficial mentions. In contrast, nearly every country emphasizes education for workforce and AI expert development, articulating a range of creative and specific implementation strategies, a number of which are in development or already in place. The lack of engagement with AIED for the large majority of countries suggests not only an absence of tangible planning to implement AIED across educational systems, but even a lack of basic awareness about the policy and ethical implications of AIED.

Where is AIED and Ethics in International Policy Discourse?

The evidence suggests—per the author's interpretation—a compelling possible explanation for why AIED is largely absent from policy discourse, as reflected by AI policy strategies.[9] Namely, in the policy agenda-setting conversations taking place behind the scenes, in the minds of policymakers, and reflected in the national AI policy strategies, education may have a particular (instrumental) framing in the context of AI. Akin to a framing contest, (Boin, Hart, and McConnell 2009; Kaplan, 2008), this instrumental frame may overshadow and marginalize other framings such as *AI for Education*, leading to a lack of imagination regarding AIED's possible transformative implications.

As manifested openly in the recent national AI policy strategies analyzed here, this instrumental framing emphasizes economic (as well as political and military) competition in the context of AI. Such a framing is consistent with longstanding policy and

[9] There are additional explanations to consider as well, though this study does not provide clear evidence to establish these. First, policymakers may simply never have been informed about AIED. AIED has typically been contained within an expert academic domain primarily accessible to computer scientists (Schiff 2021). Relatedly, while some AIED applications are beginning to hit mainstream classrooms, relatively few people, adults or children, have experience with them personally. In contrast, general education and its role in the labor market is something that nearly all members of society experience and can relate to, and a traditional focus of policymakers. Barring basic awareness, policymakers may not realize the transformative potential of AIED. If this is accurate, a clear prescription is to significantly increase efforts to inform policymakers about AIED and its ethical implications. This begs the question of whether the AIED community is prepared to do so, something which this article addresses in its Recommendations for AIED Researchers section.

education discourse emphasizing the role of engineering and, more recently, STEM in national competitiveness and productivity (Alper and U.S. National Academies of Sciences, Engineering, and Medicine 2016; Downey and Lucena, 1997). In the context of AI, the greatest concerns surrounding education appear to be labor disruption caused by automation and AI and the need for highly-skilled computer scientists and engineers to move national innovation forward. The result is that education appears to be viewed instrumentally as a strategy to support national economic viability and competitiveness. In contrast, the actual application of AI to transform education is given much less attention, far less than similarly important policy sectors. For example, sectors like healthcare and transportation are viewed as critical sectors for the application of AI, investments that will lead to positive national transformation, new government services, and economic growth.

Yet, even under an economic logic, there is arguably no reason that education should not be viewed similarly. Education is among the largest sectors and constitutes among the most government expenditures, also employing a large number of people. Indeed, the relatively few countries that discuss AIED in some depth do recognize the potential for its transformative applications. For example, Spain's AI policy strategy discusses how AIED "could completely transform current education," by improving talent retention, identifying student competencies, addressing student diversity, and predicting student failures (Ministry of Science, Innovation and Universities, Spain 2019). Mexico's (2018) strategy notes that AI in education can "enable poorer citizens to access higher-quality and cheaper services" and Kenya's (2019) strategy argues that AIED could make learning more immersive and engaging, thereby improving education in Africa. While AIED still receives relatively little attention compared to other policy sectors, these countries' approaches suggest a starting point for how other nations' policy strategies can broaden their thinking about AI in education.

Yet even the most attentive countries devote relatively little time to thinking about AIED ethics explicitly. Perhaps the most robust comment comes from Spain, which briefly warns that the "benefits are not risk-free and require proper application." Australia provides a single example of discrimination involved in fees charged for academic thesis review services, and Kenya cautions that for AIED to be successful, it must consider inclusive access. Thus while concerns about bias, privacy, transparency, and a score of other ethical concerns related to AI are pervasive across these documents (Schiff et al. 2021), these conversations rarely intersect with AIED. In fact, the general pro-innovation attitude pervades discussion of AIED even in Spain's document, which argues that intelligent systems in education would lead to "*guaranteeing* inclusive, renewed and adapted training to the needs of students and teachers" (emphasis added). In short, consideration of AIED in terms of its transformative potential and its ethical implications is sorely lacking, even for the countries that devote more attention.

Bringing AIED Ethics into Policy

In light of these findings, why is ethics an appropriate lens through which to approach AIED policy, as this article contends? How should AIED ethics and AIED policy ultimately be related? Foremost as I have argued, ethics and policy are intimately linked. Policy and law are amongst the key instruments societies have for implementing their ethical visions in society, to safeguard rights, provide opportunity, and promote values like justice. Additionally, it is important to reiterate that ethics need not refer purely to restraining harms, like preventing discrimination; it can also refer to proactive efforts to foster inclusivity and improve opportunity, such as for students and teachers. That is, ethics refers to both promoting beneficent social goals as well as mitigating risks and harms. This breadth means that applying an ethics lens to AI policy can support appreciation of AIED's tremendous potential benefits for society, such as improving access and educational quality.[10] A final advantage of approaching AIED policy through a lens of ethics is strategic in nature; because ethics discourse is prominent in AI policy, observing that important ethical benefits and risks associated with AIED are overlooked may be an effective way to bring AIED and its ethical implications to the attention of key policy stakeholders. The absence of AIED and AIED ethics in national AI policy to-date should therefore serve as a clarion call for more thoughtful incorporation.

This article seeks to demonstrate the utility of this approach and perspective by applying an existing AI ethics framework to AIED to see how it can help guide conversations and recommendations for AIED policy. Specifically, I draw on the typology of AI ethics principles developed by Floridi and Cowls (2019): beneficence, non-maleficence, autonomy, justice, and explicability.[11] These principles were synthesized from 47 ethics principles identified through a review of six high-profile AI ethics statements. Such a typology is familiar given its substantial overlap with traditional bioethics principles (Beauchamp and Childress 1979). Moreover, it is similar to work by other scholars who have analyzed AI ethics documents and policy documents. For example, Cath et al. (2018) identify transparency, accountability, and positive impact as key principles; Daly et al. (2019) identify transparency, accountability, and privacy; and Schiff et al. (2021) identify 25 ethical concepts. Most prior research such as work by Fjeld et al. (2020), Hagendorff (2020), Zeng et al. (2018), Gibert et al. (2018), and Jobin et al. (2019) also identifies five to 10 ethical principles that can be easily mapped to the framework provided by Floridi and Cowls.

While the recommendations below are only preliminary and discussed more in depth in other work (Schiff 2021), for each of the five ethics principles, I consider a few possible directions for AIED policy.

[10] The discussions of Spain, Mexico, Kenya, and India demonstrate how such a link between ethics and policy for AIED might be established even though most countries have not yet identified these connections explicitly.

[11] Note that a similar approach has also been adopted by The Institute for Ethical AI in Education (2020), which, through a series of workshops and reports, has explored AIED policy by using a different AI ethics framework, the EU's seven Ethics Guidelines for Trustworthy AI (European Commission 2019). This provides further support to the idea of approaching AI governance through an ethical lens.

Beneficence

Beneficence, in the context of AI, is often described in terms of promoting well-being, public and social good, human values, and sustainability, amongst other concepts (Floridi and Cowls 2019). Policy can better facilitate beneficent uses of AIED by emphasizing the role of AIED beyond instrumental economic goals. For example, AIED has been recognized by countries like Kenya and Mexico to potentially support diverse learners, including students with less access to high-quality education or who have special learning needs (Nye 2015; Pinkwart 2016). Public policy can therefore leverage the unique capacity of AIED technologies to scale and adapt by incentivizing research and subsidizing adoption, especially for students with the greatest needs, a possibility recognized in a few national AI strategies.

Further, while supporting economic growth and successful labor outcomes is important, AIED can also be used to teach a more diverse array of courses. Some countries recognize digital literacy as important to public citizenship, and AIED-driven education can foster civic and moral education in other ways as well, such as by teaching courses focused on social good, sustainability, and other topics. Even coursework focused on STEM education can emphasize social responsibility and impact (Lord et al. 2018). Policymakers can thus help to realize the potential positive impacts of AIED on well-being by beginning to recognize AIED's possible transformative effects beyond labor outcomes, and by building this recognition into the policy agenda-setting process as an expression of complementary ethical and policy goals.

Non-Maleficence

In the context of education, non-maleficent policy should incorporate consideration of privacy and data security, potential biases and resulting harms, and the nudging or manipulation of children, amongst other topics (I discuss some of these below in the subsection on Justice). In the case of privacy, policy needs to weigh the benefit of providing more data for AIED systems against increasing calls to protect personal data and privacy rights (Zarsky 2016). Regulatory efforts like the EU's General Data Protection Regulation (GDPR), frameworks like Privacy by Design (Langheinrich 2001), and developing international standards like the IEEE P7004 Standard for Child and Student Data Governance may help policymakers to strike an appropriate balance for AI in educational systems.

Another potential harm emanates from the fact that intelligent tutoring systems and learning robots are increasingly incorporating the capacity to detect and respond to human emotions, such as through text, audio, individual avatars or agents (Dennis et al. 2016; Graesser et al. 2014; Harley et al. 2017; Woolf et al. 2010). Though there are positive learning outcomes associated with these advances, in the most invasive sense, these tools could also be used to pervasively monitor student attention or other behaviors through eye tracking and measures of other stimuli. Emotionally-aware, privacy-insensitive, and intrusive AIED systems could be used to promote scolding, shaming, or punishing of students, with potentially disproportionate harms against the most vulnerable students (Article 19. 2021). Policy guidelines around AIED should carefully consider the extent to which we want AI systems to measure and manipulate student emotions or behaviors, especially for students in vulnerable groups such as those with disabilities (Borenstein and Arkin 2017).

Autonomy

Floridi and Cowls emphasize how AI systems can potentially replace human decision-making. In the case of AIED, inappropriate delegation of decision-making can potentially undermine the autonomy of students, parents, and teachers. For example, AIED could be imported into low-income settings and countries without adequate deliberation and participation by communities (Schiff 2021). This has happened in the past with education technologies that have been implemented without proper consideration of barriers and supportive structures (see, for example, Ely 1999; Warschauer and Ames 2010). If AIED systems are adopted as large-scale standardized tools, they may ironically undermine the very adaptability they promise. Therefore, public policy should consider how local educational stakeholders (education departments, school districts, schools, communities, teachers, and students) can retain autonomy in decisions about whether and how to adopt AIED.

A related concern is the transformation of the role of teachers and their possible de-professionalization or displacement (Hannafin and Savenye 1993; Noble 1998). There is a long history of teacher resistance to reforms, including educational technology (Cuban 1986; Terhart 2013). Such resistance can be especially strident if AIED tools relegate teachers to roles as mere facilitators (Hao 2019). To preserve a positive, complementary relationship between educators and AIED, policymakers can engage in careful participatory processes with a wide variety of stakeholders, and apply new AIED technologies only after careful pilot projects and understanding of local implementation and policy contexts (Amiel and Reeves 2008).

Justice

An established concern in AI ethics is that of racial (Chouldechova 2017), gender (Caliskan et al. 2017), and other biases in AI (Whittaker et al. 2018). As in other uses of AI, AIED systems are trained on existing data in society which may reflect long-standing historical and cultural biases (Mehrabi et al. 2019). This may mean that predictive algorithms in education, such as those used to guide college admissions decisions, could produce race, gender, or socio-economic biases by directly or indirectly favoring students with certain demographic characteristics (Marcinkowski et al. 2020). A recent controversy over the Office of Qualifications and Examinations Regulation (Ofqual)'s use of an algorithm in college admissions in the United Kingdom exemplifies the risks of unjust or poorly-considered uses of AIED systems in administrative contexts (Hao 2020). As numerous AI policy strategies discuss issues of bias, extending these conversations to AIED should not be difficult.

In the context of teaching and learning tools, AI's reliance on training data may also mean that tools like intelligent tutoring systems reflect learning styles, local customs, vocabulary, visual icons, and gestures relevant to certain student populations while neglecting the context needed to support other students. This becomes problematic when AIED systems developed based on certain considerations are implemented in an a different sociotechnical context, a challenge Selbst et al. (2019) refer to as a "portability trap." For example, if intelligent tutoring systems are trained based on data from high-income schools and universities in Western countries, they may fail to account for learner diversity across socioeconomic, geographic, and cultural divides (Joshi at al. 2018). Indeed, India's AI policy strategy is aware of the need for culturally-sensitive AIED, such as AIED that is available in a plurality of languages and accounts for variance in citizens' digital literacy. However, cultural sensitivity and awareness are currently some of the least prioritized ethical issues in AI policy documents (Schiff et al. 2021). Ongoing research to identify and mitigate biases through best practices or standards (Chatila and Havens 2019) can help to support just policy. It turn, policymakers can incentive this research and build requirements for bias identification and mitigation, along with responses to other ethical concerns, into law.

Critically, lack of attention to biases as well as social context could lead to subsequent exacerbation of inequality. For example, under-resourced schools, regions, or countries may adopt AIED to cut costs, but at the expense of failing to develop indigenous educational infrastructure and capacity, such as a well-prepared and supported teacher workforce (Schiff 2021). In the worst case, these educational systems could be locked into a lower tier as short-term technological solutionism and cost-cutting undermine long-term thinking. To consider issues of justice and fairness, policymakers should keep a longer time horizon in mind. They may benefit from employing tools like policy or regulatory impact assessment (Radaelli 2009) and policy experiments (Bravo-Biosca 2019) to understand the implications of AIED for areas related to justice before instituting major policy reforms.

Explicability

The last principle proposed by Floridi and Cowls, and the only addition to the traditional bioethical principles list, addresses concerns about transparency and accountability. Concerns about black box AI systems (Castelvecch 2016) have led to an explosion of research on transparent and explainable AI (XAI) (Adadi and Berrada 2018). Both technical transparency and more general process and policy transparency should be concerns of policymakers (Zhou and Danks 2020), given the obligation of governments to reflect publicly-held values surrounding transparency and accountability (Bozeman 2002). For example, if transparency and accountability are not promoted, institutions of higher education that employ algorithms in admissions may not be able to explain why certain students were admitted or denied. Parents unhappy with educational outcomes may not receive transparent explanations from school districts about why pedagogical or remediation decisions are made. Teachers may resist opaque tools that evaluate them and shape decisions about their compensation or employment (Morganstein and Wasserstein 2014).

The public outrage over Ofqual's use of an algorithm to shape college admissions decisions in the United Kingdom exemplifies why AIED must be regulated in a way that emphasizes public explicability. As in the case of bias, policymakers can incentivize research and establish regulatory guidelines for transparency and explainability in both the public and private sectors (Katyal 2019). This could include, for example, requiring educational institutions that use algorithms to provide publicly accessible explanations for any decisions made, document these decisions, and establish venues to contest potentially problematic results or processes (Raji et al. 2020).

While this article cannot cover the full set of ethical considerations and implications associated with AIED,[12] the above suggests some areas of concern, a framework of ethical principles from which to draw inspiration, and a variety of possible policymaking levers. Yet, to be effective in promoting these and other policy solutions, AIED researchers will need to step up their engagement with key policymakers, as well as bring more evidence to bear in these conversations. In particular, AIED researchers are well-positioned to expand knowledge on AIED's ethical implications—both negative and positive. Moreover, a critical aspect of understanding these implications in a way that is useful to policymakers is to perform research on larger units of analysis beyond the classroom, such as entire school districts adopting AIED. Researchers of AIED must also increasingly tackle a set of questions familiar to scholars of education policy, such as those related to teacher preparation, school funding, and long-term student outcomes. To help provide some guidance, the next section therefore turns to recommendations for AIED researchers in three areas: Engagement with Policymakers, Research on AIED Ethics, and Research on AIED Policy.

Recommendations for AIED Researchers

Engagement with Policymakers

Policymaking processes can be complex, contentious, opaque, and can feel quite inaccessible from the perspective of subject experts who wish to provide evidence to policymakers. In turn, even policymakers invested in seeking evidence from these expert 'knowledge brokers' may be impeded by a lack of timely, relevant research specific to their needs (Oliver et al. 2014). Fortunately, a growing paradigm of evidence-based policymaking and associated research suggest some key facilitators of improved researcher-policymaker collaboration (Malin and Brown 2019). As an introduction to research around evidence-based policymaking and policy influence

[12] See Schiff (2021), The Institute for Ethical AI in Education (2020), Holmes et al. (2021), and other articles in this issue for more detailed reviews of AIED ethics.

of researchers, I adapt four recommendations from Phoenix et al. (2019) regarding strategies for AIED researchers to impact policymaking processes.

First, researchers should aim to make their research relevant to policymakers' needs (Parkhurst 2017). Because policymakers are often looking for concrete answers to specific problems, early and regular engagement with policymakers it can help to define key policy questions that need to be answered. Timing is critical as well, as lack of timely access is one of the most prominent barriers to policymaker use of evidence (Oliver et al. 2014). AIED researchers can address timing considerations by enhancing their willingness to produce interim research outputs such as preliminary reports, rather than waiting for the peer review process to reach fruition. While this approach to identifying research questions and producing evidence can be unfamiliar, it will help to reduce ambiguity (Cairney et al. 2016) and provide actionable information for policymaking.

Second, researchers should present their findings in clear, accessible formats (Bucchi 2013). This may mean developing simplified and aesthetically attractive reports that highlight key themes, insights, and even quotes and stories. While researchers can enhance their efforts at science communication even as individuals, educational institutions and professional associations of AIED researchers such as the International Artificial Intelligence in Education Society (IAIED) and The Institute for Ethical AI in Education can also play a key role. These organizations can produce white papers, reports, and other materials that have significant credibility as authoritative expressions of the AIED community's ideas (Lencucha et al. 2010). These kinds of materials often serve as key references for policymakers as they deliberate over policy agendas and look for supportive evidence.

Third, AIED researchers can foster relationships with policymakers as well as intermediary organizations (Phoenix et al. 2019), such as education and teaching-focused research organizations, education funders, advocacy groups, and teachers' unions. These intermediary organizations often have established relationships with policymakers and policymaking institutions, and understand possible venues through which to influence decision-making (Scott et al. 2014). They can therefore serve as important assets for policy-engaged researchers (Sin 2008) to both exert influence as well as gain knowledge and know-how towards building their own capacity. Moreover, through these efforts, it is possible to establish long-term relationships with policymakers, such as public officials who have special interests in education or technology.

Fourth and related, AIED researchers should engage in the identification of 'policy windows' (Kingdon 1995). Policy windows refer to periods when public problems rise to the forefront of the decision agenda. At these points, entrepreneurial advocates can couple possible policy solutions to policy problems and promote meaningful action. AIED researchers acting as policy entrepreneurs can engage in a process of venue shopping (Guiraudon 2000), identifying receptive politicians, decision-making agencies, committees and so on, such as when Rose Luckin, then president of the IAIED, testified before the United Kingdom House of Lords as they were developing their AI policy strategy (House of Lords, Select Committee on Artificial Intelligence 2018). When obvious venues do not exist, an option is to promote the establishment of permanent committees or task forces of political officials with a special focus on education technology or AI. These can then become durable venues through which to inform policymaking and build relationships. Given the substantial interest in AI policy, the next few years may very well constitute a key policy window to place issues of AIED and AIED ethics on the agenda. This article has also argued that the prominent framing of AI ethics in AI policy provides a key issue frame that policy-engaged AIED researchers can take hold of as they seek to enter the current window and inform AIED policy.

Research on AIED Ethics

AIED scholars who wish to better incorporate ethical thinking into their work, such as by reference to the ethical framework of Floridi et al. (2019) explored in this article, have many sources on which to draw. A large body of work has been developed by the broader AI ethics community to address algorithmic bias and transparency, key issues involved in ensuring just and non-maleficent policy. For example, proposed strategies of creating datasheets for datasets (Gebru et al. 2020) and model cards for model reporting (Mitchell et al. 2019) can help to promote transparency and address explicability concerns, while suites of tools such as IBM's AI Fairness 360 (Bellamy et al. 2018) or the Scoping, Mapping, Artifact Collection, Testing, and Reflection (SMACTR) audit framework proposed by Raji et al. (2020) can help researchers establish end-to-end processes for ethical AIED research. Researchers can consult Mehrabi et al. (2019), Guidotti et al. (2018), and Turner-Lee et al. (2019) for reviews of emerging best practices.

There are additional frameworks for thinking about ethics in AI design and research, some of which were developed specifically for AI, and others which were developed for broader contexts but can be useful. Some promising approaches developed specifically for AI include the IEEE 7010 (2020) Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-being, AI Now's Algorithmic Impact Assessment (Reisman et al., 2018); ALGO-CARE (Oswald et al., 2018), and Women Leading in AI's PARETS framework (Women Leading in AI 2019). Impact assessments such as IEEE 7010 can involve a broad effort to understand possible impacts of AI systems, to foster participation in design and consideration of AI systems, to measure impacts in real-world settings, and to improve the development and implementation of AI systems (Schiff, Ayesh, et al. 2020a). These approaches to ethics in design and implementation may be valuable for AIED researchers to consider. Importantly, some impact assessments such as IEEE 7010 consider positive as well as negative impacts, allowing researchers to observe beneficent implications of AIED policy as well as areas where principles like autonomy and justice could be safeguarded or threatened.

Additional promising frameworks that were not developed specifically for AI include Responsible Research and Innovation (Schomberg, 2013), Privacy by Design (Oetzel and Spiekermann 2014), Ethics by Design (Dignum, 2018), Value Sensitive Design (Friedman, Hendry, and Boning 2017), and others. Interested researchers should consult Morley et al. (2019) for a sweeping review of over 100 AI ethics tools and frameworks. Finally, it will be important to consider forthcoming international standards such as the IEEE's Ethics Certification Program for Autonomous and Intelligent Systems and standards projects that address bias (P7003), transparency (P7001), privacy (P7002, P7004, P7005), nudging (P7008), and well-being (P7010) or similar standards in development from ISO/IEC (Winfield 2019) and other international and national standards entities. Standards can provide clear actionable guidance and may have legal or regulatory force in the future, providing ethical frameworks for AI such as the one presented in this article with additional monitoring and enforcement capacity. However, as most of these tools were developed for more general AI systems or computing and design contexts, there is arguably a need for AIED scholars to develop or customize impact assessments and design frameworks for particular use cases within AIED. In sum, AIED researchers have many resources to draw on as they develop their own toolkits individually and collectively for advancing research on AIED ethics.

Research on AIED Policy

As part of advancing the understanding of ethical implications of AIED, AIED researchers should increase efforts to understand the social, economic, and policy impacts of AIED outside of individual teaching and learning tools or single classrooms. This is important as policymakers are especially concerned with large-scale social and economic impacts in education policy, such as whether new education technologies or reforms are likely to foster economic growth, national competitiveness, and long-term student outcomes. For AIED researchers currently focused on technical advances and demonstration projects, this means expanding the scope and scale of attention to large-scale implementation projects at the level of multiple schools or school districts, and to classic questions in the remit of education policy. Notably, these issues are becoming increasingly relevant as AIED applications now reach millions of students (Hao 2019) and online and remote learning have become necessary for educational institutions and governments alike (Ali 2020), opening new inroads for implementing and studying AIED (Schiff, 2021).

Such research could address, for example, the impact of AIED adoption not only on learning outcomes, but also on social-emotional outcomes, promotion and graduation rates, long-term life outcomes, school culture, teaching and learning attitudes, teacher development and professionalism, school system finances, teachers' unions, educational equity, and more. Answering associated research questions considers AIED as a set of tools both for teaching and learning applications as well as for as administrative goals like teacher preparation and dropout prevention. It thereby responds to and can build on how national AI policy strategies in the most engaged countries have discussed AI for education to-date. This line of work can help to identify implementation barriers (Lowther et al. 2008; Nye 2015) as well as expected and unexpected (positive or negative) ethical and policy consequences associated with AIED. For instance, does AIED at scale foster improved differentiation for learners with special needs and students living in poverty, or does it exacerbate inequitable outcomes by reproducing biased teaching and learning approaches? Is AIED likely to improve the quality of educators' working lives such as through new professional development tools, or is it more likely to be invoked to justify spending cuts and de-professionalization of teachers? These questions exemplify how research on AIED's ethical and policy implications is complementary and arguably essential if policy-engaged researchers wish to speak to policymakers' concerns.

While this research agenda is ambitious, there are fortunately existing paradigms in education research that address these and other topics. AIED researchers should consider increased partnership with program evaluation (Dickard 2003) and policy experts (Buchanan 2020) in the field of education as well as with other social scientists. For example, AIED researchers could co-develop research proposals with peers to understand AIED's impacts across multiple levels of educational systems, and over time. Two frameworks of special note are design-based research (Amiel and Reeves, 2008) and design-based implementation framework (DBIR) (Fishman et al., 2013). Both methods are drawn from educational technology research, promote participatory design (Williamson and Eynon, 2020), and recognize the need to move beyond basic research and pilot projects to implementation in real-world settings. DBIR in particular examines the conditions needed for sustainable change in complex environments—the kind of knowledge needed to guide AIED policymaking. Of course, is not necessary that the entire research community shift its focus, but an increased willingness to attend to these issues and consider new research projects will lead to more policy-engaged research output and capacity.

Conclusion

This article engaged in thematic analysis of 24 national AI policy strategies from around the world, identifying key topics that describe how leading policymakers envision the relationship between AI and education. The results as interpreted here indicate that policymakers view education largely as an instrumental tool to support workforce development and training of AI experts. This framing and focus on *Education for AI* comes at the expense of attention to the role of *AI for Education* itself. If such a trend continues, policymakers may fail to realize AIED's transformative potential and may fail to sufficiently fund, regulate, and consider AIED's ethical implications—both positive and negative.

The research design and findings are certainly subject to limitations. Foremost, the documents selected for study, national AI policy strategies, only reflect a portion of discourse around education, AI, and ethics. National and local agencies focused specifically on education as well as non-governmental stakeholders may have more detailed and thoughtful considerations about AIED than are reflected in AI policy strategies alone. However, AI policy strategies are indeed reflective of major and urgent policy priorities of top national leaders, and as such what these documents discuss or fail to discuss is worthy of study. Another limitation of this research is that the identification of themes, coding, and analysis was done by a single researcher. The findings are thus necessarily subjective and reflect a particular perspective. Additional research on AIED policy and ethics that draws on different approaches, such as case studies or interviews, could help to inform a more robust understanding of the issues reviewed here.

Nevertheless, this article's findings raise concerns about how policymakers are currently approaching AIED, concerns that should inspire both those interested in effective policymaking as well as those concerned with AI's ethical impacts. In particular, this article argued that policy and ethics are intimately related, that policymaking is fundamentally a reflection of society's underlying ethical values, and that the focus of AI policymakers on ethics is a valuable lens for policy-engaged AIED researchers to draw on for both principled and strategic reasons. As such, policy that does not take account of AIED's promises for positive impact in addition to associated possible harms risks not only blindly encouraging mishaps, but also foreclosing on valuable opportunities.

In response to these concerns, this article drew on a framework of five AI ethics principles—beneficence, non-maleficence, autonomy, justice, and explicability—and considered how policymakers could generate well-considered AIED policy by attending to these ethical domains. As AIED researchers have a key role to play, this article also offers actionable recommendations for members of the AIED community interested in informing policy, including strategies for engagement with the policymaking process and proposed lines of research for both AIED's ethical and policy implications. Armed with a better understanding of these complementary issues, researchers will be better positioned to inform policymaking during the current policy window and in the future. Thus, while the lack of attention to AIED and its ethical implications to-date in most national AI policy strategies is disappointing, the AI community still has the opportunity to increase engagement with policymakers and shape policy discourse, regulation, and the ethics of AIED on behalf of the public good.

Appendix 1: Methodology

This appendix provides more extensive details surrounding the data collection, screening, coding, and analysis strategy, as well as associated limitations, than is available in the main body of the article.

Data Collection

The data collection process differs from that typically employed in meta-analysis and review papers, such as recommended by the PRISMA Statement (Moher et al. 2009), largely because the documents evaluated here are so-called "gray literature" not found in traditional databases (Mahood et al. 2014; Rothstein and Hopewell 2009). Given the type of documents and substantive focus of the study, the data collection process relied on linkhubs, Google searches, and manual searching of certain documents and countries: e.g., "country name + AI policy strategy." The search process is connected to ongoing research assessing AI ethics and policy (Schiff et al. 2021, 2020a, 2020b). The first author and colleagues have maintained a database of AI policy documents and additional AI ethics documents (available at https://dx.doi.org/10.21227/fcdb-pa48). The database was created in Fall 2018 and updated regularly until early 2020. Data collection focused on policy documents benefited especially from lists maintained by the Future of Life Institute (2020) and Tim Dutton (2018). These linkhubs contain updates about AI policy developments in dozens of countries, including how far along countries are in development of task forces, funding proposals, and formal AI policy strategy documents.

For each country noted in these linkhubs, I searched for and accessed documents and initiatives mentioned, and performed additional Google searches and manual searches for each country to ensure that the set of national AI policy strategies was as complete as possible. While it is possible that some countries were omitted, perhaps due to lack of language familiarity, the two key sources are invested in tracking national AI policy developments. All such candidate documents were thus captured in the database managed by the first author and colleagues. From this larger database, I extracted only documents produced by public sector organizations (e.g., countries, not corporations or non-governmental organizations). This resulted in 76 public sector AI documents that formed the candidate pool of national AI policy documents.

Screening Process

The screening process involved identifying criteria for types of documents, publication language, the population and phenomena of interest, and time period (Stern et al. 2014). The purpose of this study was to assess national AI policy strategies as they relate to education. Therefore, the study applied the following inclusion/exclusion criteria:

- Documents needed to be complete and resemble a national AI policy strategy. Some such documents describe themselves as preliminary reports or blueprints, working towards more robust or formalized policy strategies. Nevertheless, many were sufficiently robust so as to be considered as policy strategies. On the other hand, countries that had only announced task forces, funding initiatives, created websites, or otherwise did not have a well-developed document analogous to that of other countries were not included. This ensures countries can be compared fairly and speak to sufficiently detailed items of policy, with a sizable average of 62 pages per document.
- Documents needed to be in English, due to the author's limited language proficiency. However, in a number of cases, governments had produced official English-language translations (e.g., Finland, Italy). While automated translation of non-English documents (e.g., Google Translate) may not be of sufficient quality, there was one unofficial but high-quality translation included in the final sample, of China's AI policy strategy, performed by the Foundation for Law and International Affairs.
- The study also excluded documents produced by inter-governmental organizations, such as the United Nations, the Organization for Economic Cooperation and Development, and the European Union. While these documents are no doubt important, they address a different scope, as they are relatively distant from

national-level institutions, funding activities, and other policy activities, such as those involving education policy. This makes these documents less comparable to national-level AI policy strategies.

- Finally, in a number of cases, countries produced multiple documents that were potentially relevant to AI policy. Only one document was selected per country. The chosen document was typically the most robust and the most recent, at times an evolution of a previous draft or more preliminary document. Further, some candidate documents were not representative of an overarching national AI policy strategy. For example, documents from Germany addressing autonomous vehicle policy and from Finland addressing work in the age of AI were excluded in preference of Germany's National Strategy for AI and Finland's Age of AI. This screening criteria helped to ensure that individual countries were not overrepresented, that information analyzed was not redundant, and that the most robust, high-quality, and comparable policy strategies were selected in each case.

Of the 76 candidate documents, eight did not resemble a complete national policy strategy document, one was not available in English, 13 were inter-governmental, and 30 were excluded in favor of more representative documents. Screening resulted in a final sample of 24 national AI policy strategies.

Codebook Development

After identifying the final sample, the analytical strategy began with the development of a preliminary set of topics, in the form of a codebook (Miles et al. 2014; Thomas 2006). These topics in the codebook were chosen based on the study's conceptual scope and framework, the author's subject matter knowledge, and previous exposure to AI policy strategies. The scope of interest was any discussion of education, construed as broadly as possible, such as youth and adult education, training and re-skilling, investing in educational systems, public education and awareness, the need to develop more AI talent (e.g., computer scientists, engineers), and social and ethical issues related to AI and education.

The initial codebook included 11 categories: Education as Priority Topic, K-12 Education, Post-Secondary Education, Adult Education and Training, General AI Literacy, Training AI Experts, Preparing Workforce, Intelligent Tutoring Systems, Pedagogical Agents and Learning Robots, Predictive Educational Tools, and AI for Healthcare. A best practice in qualitative research is to iterate and refine the codebook through testing on a small subset of the data (Roberts et al. 2019). Therefore, I randomly selected five documents—aiming for a meaningfully-sized and somewhat representative subset—and applied the thematic schema to them. This involved reading the documents to determine whether the coding schema could validly and straightforwardly reflect the way education was discussed in the documents, and to identify if the coding schema captured the full range of issues in the documents relevant to the article's conceptual scope.

Based on this initial test, several categories were modified, removed, and collapsed as follows:

- Education as Priority Topic and AI for Healthcare were retained, as they were easy to apply. Either topic might be explicitly noted as a priority topic in a document, for example, if a list of priority policy sectors was mentioned and education was among that list. Alternatively, education/healthcare were coded as priority topics if a significant subsection was dedicated to them, or if there was a similar amount of discussion relative to the length of the document as compared to other documents that *did* identify education (or healthcare) as an explicit priority.
- K-12 Education, Post-Secondary Education, and Adult Education and Training were removed. These categories were originally designed to separate discussion of education by target age/population group. However, the test documents often did not identify the target age/group when discussing AI and education, making this distinction difficult to code accurately. Moreover, these population differences were deemed less relevant for the overall purpose of the article. For example, that documents emphasized the need to develop more AI researchers seemed more pressing to the document authors than whether this development happened in secondary or postsecondary educational institutions.
- Training AI Experts and Preparing Workforce for AI were straightforward and were retained.
- General AI Literacy was renamed to Public AI Literacy. The former was originally defined to emphasize development of general digital, STEM, and other skills in educational settings. The theme was relabeled and redefined to incorporate AI literacy in both educational (classroom) and ‘public’ settings, because both settings were discussed and justified to pertain to similar policy purposes.
- The revised codebook collapsed Intelligent Tutoring Systems and Pedagogical Agents and Learning Robots into Teaching and Learning. Too few documents addressed these issues at the level of detail of individual AIED technologies or tools to allow for reliable identification, as the documents generally employed more abstracted terms and discussions.
- The revised codebook also abstracted Predictive Educational Tools into Administrative Tools, as there were several examples of AIED tools mentioned that were better captured by the latter, broader terminology, such as the use of AI for inspection or assigning teachers to schools.

These adjustments resulted in a revised codebook with seven categories (a reasonable number for inductive studies) (Thomas 2006), described in the main body. The final coding categories were straightforward to apply to the data and captured relevant concepts within the study’s scope well.

An important note is that, despite an initial attempt to code for discussion of AIED ethics specifically given its importance to this study, discussion of these topics was too rare to justify having as a theme. Most discussion addressing ethics and education was focused on *Education for AI* purposes, such as training future machine learning experts to develop ethical design skills, rather than addressing ethical implications emanating from AIED. Nevertheless, I captured all mentions of ethics in the context of both *Education for AI* and *AI for Education* in my memos, and considered the presence and absence of these topics as part of the interpretive work.

Coding Approach

Next, I applied the codebook to the 24 documents in the sample (approximately 1491 pages total). Each document was read closely and assessed manually along the seven topics using a simple form of content analysis. This consisted of evaluating each document for the presence or absence of each theme (White and Marsh 2006), largely a binary exercise, though some documents were coded as borderline cases. In Table 1, a country is marked as green when a theme was reflected, red when absent, and yellow when the case was sufficiently ambiguous or borderline.

For example, Malta's (2019) document initially notes that AI for healthcare may be amongst the highest impact projects and "most pressing domestic challenges" worthy of prioritization, but it does not proceed to include any substantive discussion or a subsection on healthcare. In comparison to the document's discussion of other topics, and in comparison to other countries' AI policy documents that discuss healthcare in more depth, this relatively more narrow treatment of the topic led to coding it as yellow. Similarly, Russia's (2019) discussion of using AI to "[improve] the quality of education services" does not provide enough detail to be clear about the role of AIED as a potential tool for teaching and learning, and so is considered to be too ambiguous to code as either green or red.

Analysis Approach

Relevant quotes from the documents were captured in a research memo and organized under the seven categories (Thomas 2006) to support higher-order conceptual and thematic interpretation. Additional quotes of interest and minor categories were included here as well, such as any mentions of ethics related to education. From this, I synthesized insights from the frequency and character of these topics, applying a thematic analytic approach (Castleberry and Nolen 2018) to identify major findings. This interpretive exercise involves considering second-order meanings or explanations for the patterns identified in the data (Miles et al. 2014), including the finding that AIED's ethical implications are neglected. I present results for each topic in the main article, along with interpretation of key findings within and across topics, to support a broader discussion of the role of education and ethics in AI policy in the subsequent sections.

Limitations

Because the documents were coded by a single researcher, it is not possible to, for example, assess inter-rater reliability. Further, the conceptualization of the study, codebook development, and interpretation were not subject to the perspectives of other researchers or experts outside of the peer review process. However, quantitative measures of reliability are only sometimes considered essential in qualitative research (Castleberry and Nolen 2018), and a single coder approach can be appropriate and, in cases, even preferrable (Harding and Whitehead 2013). Multiple researchers may not be necessary to provide sufficient consistency and credibility, as single researchers can provide a unitary and consistent perspective, albeit one dependent on that author's subjective assessments. For example, research using semi-structured interviews with dozens of coding categories and many degrees of detail (e.g., scoring attributes from 1–10) benefit especially from the assessment of interrater reliability, particularly if the codes are challenging to conceptually separate or define. In this study, however, the number of topics is small, the level of detail simple, and the concepts are fairly easy to conceptually separate.

Moreover, in qualitative research, there are common criteria of research rigor used as alternatives to traditional quantitative criteria of validity and reliability. For example, one widely used set of criteria comes from Lincoln and Guba (1985), who propose credibility as an alternative to internal validity, transferability as an alternative to external validity, dependability as an alternative to reliability, and confirmability as an alternative to objectivity. To satisfy these criteria, the analysis employed several recommended strategies (Lincoln and Guba 1986). Within-method triangulation across multiple documents (Jonsen and Jehn 2009) and the use of direct quotes as descriptive evidence provide rich support to demonstrate claims, supporting their credibility, dependability, and transferability. Further, because the data are publicly available, as opposed to privately held interview data, for example, they are open to scrutiny and confirmation or disconfirmation. However, in part because of researchers' individual positions and biases (Castleberry and Nolen 2018), it is possible that other researchers would identify different coding categories or identify different salient themes. As such, the single researcher approach is a limitation of this study, discussed in the study's limitations section. Future research examining the role of education in AI policy would be welcome in assessing the extent to which the findings presented here are indeed credible, dependable, confirmable, and transferable.

Funding The author declares no funding.

Availability of data and material Not applicable

Code availability Not applicable

Declarations

Conflicts of interest The author declares no conflicts of interest or competing interests.

References

Acemoglu, D., & Restrepo, P. (2019). The wrong kind of AI? Artificial intelligence and the future of labor demand. Working Paper 25682. National Bureau of Economic Research. https://doi.org/10.3386/w25682.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6,* 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052.

Ali, W. (2020). Online and remote learning in higher education institutes: A necessity in light of COVID-19 pandemic. *Higher Education Studies, 10*(3), 16–25.

Allen, G. C. (2019). Understanding China's AI strategy. *Center for a New American Security* (blog). February 6, 2019. https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy.

Alper, J. and U.S. National Academies of Sciences, Engineering, and Medicine, eds. 2016. *Developing a national STEM workforce strategy: A workshop summary*. Washington, DC: The National Academies Press.

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Journal of Educational Technology and Society, 11*(4), 29–40.

Article 19. 2021. "Emotional Entanglement: China's Emotion Recognition Market and Its Implications for Human Rights." A19/DIG/2021/001. London, UK: Article 19. https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf.

Autor, D. H. (2015). Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives, 29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Bagloee, S. A., Tavana, M., Asadi, M., & Oliver, T. (2016). Autonomous vehicles: Challenges, opportunities, and future implications for transportation policies. *Journal of Modern Transportation, 24*(4), 284–303. https://doi.org/10.1007/s40534-016-0117-3

Beauchamp, T. L., & Childress, J. F. (1979). *Principles of biomedical ethics*. Oxford University Press.

Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P. et al. (2018). AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *[Cs]*, October. http://arxiv.org/abs/1810.01943.

Blockchain & Artificial Intelligence Taskforce, Kenya. 2019. "Kenya: Emerging Digital Technologies for Kenya: Exploration & Analysis." Nairobi, Kenya: Blockchain & Artificial Intelligence taskforce. https://www.ict.go.ke/blockchain/blockchain.pdf.

Boeglin, J. (2015). The costs of self-driving cars: reconciling freedom and privacy with tort liability in autonomous vehicle regulation. *Yale JL & Tech., 17*, 171.

Boin, A. 't Hart, P., & McConnell, A. (2009). Crisis exploitation: Political and policy impacts of framing contests. *Journal of European Public Policy, 16*, 81-106. https://doi.org/10.1080/13501760802453221

Borenstein, J., & Arkin, R. C. (2017). Nudging for good: Robots and the ethical appropriateness of nurturing empathy and charitable behavior. *AI & SOCIETY, 32*(4), 499–507.

Bozeman, B. (2002). Public-value failure: When efficient markets may not do. *Public Administration Review, 62*(2), 145–161. https://doi.org/10.1111/0033-3352.00165

Bravo-Biosca, A. (2019). Experimental Innovation Policy. Working Paper 26273. National Bureau of Economic Research. https://doi.org/10.3386/w26273.

British Embassy in Mexico, Oxford Insights, and C Minds. 2018. "Mexico: Towards an AI Strategy in Mexico: Harnessing the AI Revolution." Mexico City, Mexico: British Embassy in Mexico, Oxford Insights, and C Minds. http://go.wizeline.com/rs/571-SRN-279/images/Towards-an-AI-strategy-in-Mexico.pdf.

Bucchi, M. (2013). Style in science communication. *Public Understanding of Science, 22*(8), 904–915. https://doi.org/10.1177/0963662513498202

Buchanan, R. (2020). Through growth to achievement: Examining Edtech as a solution to australia's declining educational achievement. *Policy Futures in Education, 18*, 1026–1043. https://doi.org/10.1177/1478210320910293

Cairney, P., Oliver, K., & Wellstead, A. (2016). To bridge the divide between evidence and policy: reduce ambiguity as much as uncertainty. *Public Administration Review, 76*(3), 399–402. https://doi.org/10.1111/puar.12555

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News, 538*(7623), 20. https://doi.org/10.1038/538020a

Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning, 10*(6), 807–815. https://doi.org/10.1016/j.cptl.2018.03.019

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial Intelligence and the 'Good Society': The US, EU, and UK approach. *Science and Engineering Ethics, 24*(2), 505–528. https://doi.org/10.1007/s11948-017-9901-7

Chae, Y. (2020). U.S. AI regulation guide: Legislative overview and practical considerations. *The Journal of Robotics, Artificial Intelligence and Law, 3*(1), 17–40.

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *The New England Journal of Medicine, 378*(11), 981–983. https://doi.org/10.1056/NEJMp1714229

Chatila, R., & Havens, J. C. (2019). The IEEE global initiative on ethics of autonomous and intelligent systems. In M. I. A. Ferreira, J. S. Sequeira, G. S. Virk, M. O. Tokhi, & E. E. Kadar (Eds.), *Robotics and well-being* (pp. 11–16). Cham: Springer.

Chouldechova, A. (2017). Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. *[Cs, Stat]*, February. http://arxiv.org/abs/1703.00056.

Cooper, A. F., & Flemes, D. (2013). Foreign policy strategies of emerging powers in a multipolar world: An introductory review. *Third World Quarterly, 34*(6), 943–962. https://doi.org/10.1080/01436597.2013.802501

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. Teachers College Press.

Daly, A., Hagendorff, T., Li, H., Mann, M., Marda, V., Wagner, B., Wang, W., & Witteborn, S. (2019). Artificial Intelligence, Governance and Ethics: Global Perspectives. SSRN Scholarly Paper ID 3414805. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=3414805.

Dawson, D., E. Schleiger, J. Horton, J. McLaughlin, C. Robinson, G. Quezada, J. Scowcroft, & S. Hajkowicz. 2019. "Australia: Artificial Intelligence: Australia's Ethics Framework." Canberra, Australia: Data 61, CSIRO.

Dennis, M., Masthoff, J., & Mellish, C. (2016). Adapting progress feedback and emotional support to learner personality. *International Journal of Artificial Intelligence in Education, 26*(3), 877–931. https://doi.org/10.1007/s40593-015-0059-7

Department of Public Administration, Italy. 2018. "Italy: Artificial Intelligence at the Service of Citizens." Rome, Italy: Task Force on Artificial Intelligence of the Agency for Digital Italy and Department for Public Administration, Italy. https://ia.italia.it/assets/whitepaper.pdf.

Dickard, N. (2003). *The sustainability challenge: Taking EdTech to the next level*. For full text: https://eric.ed.gov/?id=ED477720.

Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology, 20*(1): 1–3 https://doi.org/10.1007/s10676-018-9450-z

Downey, G. L., & Lucena, J. C. (1997). Engineering selves: Hiring in to a contested field of education. In N. Sakellariou & R. Milleron (Eds.), *Ethics, politics, and whistleblowing in engineering* (pp. 25–43). CRC Press.

Dutton, T. (2018). An overview of national AI strategies. *Politics + AI* (blog). June 28, 2018. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.

Dutton, T., Barron, B., & Boskovic, G. (2018). Building an AI world: Report on national and regional AI strategies. Toronto, Canada: CIFAR. https://www.cifar.ca/docs/default-source/ai-society/buildinganaiworld_eng.pdf.

Ely, D. P. (1999). Conditions that facilitate the implementation of educational technology innovations. *Educational Technology, 39*(6), 23–27.

European Commission. (2019). Ethical guidelines for trustworthy AI. Brussels, Belgium: European Commission, High-Level Expert Group on Artificial Intelligence (AI HLEG). https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477.

Federal Ministry of Education and Research, Germany. (2018). "Germany: National Strategy for Artificial Intelligence: AI Made in Germany." Bonn, Germany: Federal Ministry of Education and Research, the Federal Ministry for Economic Affairs and Energy, and Federal Ministry of Labour and Social Affairs, Germany. https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf.

Fishman, B. J., Penuel, W. R., Allen, A. R., Cheng, B. R., & Sabelli, N. (2013). Design-based implementation research: an emerging model for transforming the relationship of research and practice. p. 21.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. SSRN Scholarly Paper ID 3518482. Rochester, NY: Berkman Klein Center for Internet & Society. https://papers.ssrn.com/abstract=3518482.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.8cd550d1

Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*(January), 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Friedman, B., Hendry, D. G. & Borning, A. (2017). A survey of value sensitive design methods. *Foundations and Trends® in Human–Computer Interaction,* 11(2): 63–125. https://doi.org/10.1561/1100000015.

Future of Life Institute. (2020). AI Policy—United States. Future of Life Institute. 2020. https://futureoflife.org/ai-policy-united-states/.

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2020). Datasheets for Datasets. *[Cs]*, January. http://arxiv.org/abs/1803.09010.

Gibert, M., Mondin, C., & Chicoisne, G. (2018). Montréal declaration of responsible AI: 2018 Overview of international recommendations for AI Ethics. University of Montréal. https://www.montrealdeclaration-responsibleai.com/reports-of-montreal-declaration.

Giuliani, D., & Ajadi, S. (2019). 618 Active Tech Hubs: The Backbone of Africa's Tech Ecosystem. *GSMA* (blog). July 10, 2019. https://www.gsma.com/mobilefordevelopment/blog/618-active-tech-hubs-the-backbone-of-africas-tech-ecosystem/.

Graesser, A. C., D'Mello, S. K., & Strain, A. C. (2014). Emotions in advanced learning technologies. In L. Linnenbrink-Garcia & R. Pekrun (Eds.), *International handbook of emotions in education* (pp. 473–93). New York: Routledge.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), 1–42. https://doi.org/10.1145/3236009

Guidaudon, V. (2000). European intergration and migration policy. *Journal of Common Market Studies, 38*(2), 251–71. https://doi.org/10.1111/1468-5965.00219

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*. https://doi.org/10.1007/s11023-020-09517-8

Hannafin, R. D., & Savenye, W. C. (1993). Technology in the classroom: The teacher's new role and resistance to it. *Educational Technology, 33*(6), 26–31.

Hao, K. (2019). China has started a grand experiment in AI education. it could reshape how the world learns. *MIT Technology Review*, August 2, 2019. https://www.technologyreview.com/2019/08/02/131198/china-squirrel-has-started-a-grand-experiment-in-ai-education-it-could-reshape-how-the/.

Hao, K. (2020). The UK exam debacle reminds us that algorithms can't fix broken systems. *MIT Technology Review*, August 20, 2020. https://www.technologyreview.com/2020/08/20/1007502/uk-exam-algorithm-cant-fix-broken-system/.

Harding, T., & Whitehead, D. (2013). Analysing data in qualitative research. In Z. Schneider & D. Whitehead (Eds.), *Nursing and midwifery research: Methods and appraisal for evidence-based practice* (4th ed., pp. 127–42). Sydney: Mosby Elsevier.

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: a taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education, 27*(2), 268–297. https://doi.org/10.1007/s40593-016-0126-8

Heyneman, S. P. (2004). International education quality. *Economics of Education Review, 23*(4), 441–452. https://doi.org/10.1016/j.econedurev.2003.10.002

Holmes, W., Bektik, D., Whitelock, D., & Woolf, B. P. (2018). Ethics in AIED: Who Cares?. In *Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part II*, Rosé, C. P., Martínez-Maldonado, R., Hoppe, H. U., Luckin, R., Mavrikis, M., Porayska-Pomsta, K., McLaren, B. and Du Boulay, B (ed) Vol. 10948. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-93846-2.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2021). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-021-00239-1.

House of Lords, Select Committee on Artificial Intelligence. (2018). United Kingdom: AI in the UK: Ready, Willing and Able?. London, UK: House of Lords, Select Committee on Artificial Intelligence. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf.

IEEE. (2020). *7010–2020 - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being*. S.l.: IEEE. https://ieeexplore.ieee.org/servlet/opac?punumber=9084217.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of ai ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Jonsen, K., & Jehn, K. A. (2009). Using triangulation to validate themes in qualitative studies. *Qualitative Research in Organizations and Management: An International Journal, 4*(2), 123–150. https://doi.org/10.1108/17465640910978391

Joshi, Meghna, T. J., & Rangaswamy, N. (2018). Scaling Classroom IT Skill Tutoring: A Case Study from India. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 630:1–630:12. CHI '18. New York, NY, USA: ACM. https://doi.org/10.1145/3173574.3174204.

Kaplan, S. (2008). Framing contests: Strategy making under uncertainty. *Organization Science, 19*(5), 729–752. https://doi.org/10.1287/orsc.1070.0340

Katyal, S. K. (2019). Private accountability in the age of artificial intelligence. *UCLA Law Review, 66*(1), 54–141.

Kingdon, J. W. (1995). *Agendas, alternatives, and public policies* (2nd ed.). HarperCollins College Publishers.

Korean Ministry of Science, ICT and Future Planning. (2016). South Korea: Mid-to long-term master plan in preparation for the intelligent information society: Managing the fourth industrial revolution. South Korea: Korean Ministry of Science, ICT and Future Planning (MSIP). http://english.msit.go.kr/cms/english/pl/policies2/__icsFiles/afieldfile/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf; https://medium.com/syncedreview/south-korea-aims-high-on-ai-pumps-2-billion-into-r-d-de8e5c0c8ac5.

Langheinrich, M. (2001). Privacy by design principles of privacy-aware ubiquitous systems. In G. D. Abowd, B. Brumitt, & S. Shafer (Eds.), *Ubicomp 2001: Ubiquitous computing* (pp. 273–91). Berlin, Heidelberg: Springer.

Lee, J.-W. (2001). Education for technology readiness: Prospects for developing countries. *Journal of Human Development, 2*(1), 115–151. https://doi.org/10.1080/14649880120050

Lencucha, R., Kothari, A., & Hamel, N. (2010). Extending collaborations for knowledge translation: Lessons from the community-based participatory research literature. *Evidence and Policy: A Journal of Research, Debate and Practice, 6*(1), 61–75. https://doi.org/10.1332/174426410X483006

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry* (Vol. 75). Sage.

Lincoln, Y. S., & Guba, E. G. (1986). but is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Program Evaluation, 1986*(30), 73–84. https://doi.org/10.1002/ev.1427

Lord, S. M., Mejia, J. A., Hoople, G., Chen, D., Dalrymple, O., Reddy, E., Przestrzelski, B., & Choi-Fitzpatrick, A. (2018). Creative Curricula for Changemaking Engineers. In *2018 World Engineering Education Forum—Global Engineering Deans Council (WEEF-GEDC)*, 1–5. https://doi.org/10.1109/WEEF-GEDC.2018.8629612.

Lowther, D. L., Inan, F. A., Daniel Strahl, J., & Ross, S. M. (2008). Does technology integration 'work' when key barriers are removed? *Educational Media International, 45*(3), 195–213. https://doi.org/10.1080/09523980802284317

Mahood, Q., Van Eerd, D., & Irvin, E. (2014). Searching for grey literature for systematic reviews: Challenges and benefits. *Research Synthesis Methods, 5*(3), 221–234. https://doi.org/10.1002/jrsm.1106

Malin, J., & Brown, C. (Eds.). (2019). *The role of knowledge brokers in education: Connecting the dots between research and practice*. Routledge.

Manheim, K. M., & Kaplan, L. (2018). Artificial intelligence: Risks to privacy and democracy. SSRN Scholarly Paper ID 3273016. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=3273016.

Marcinkowski, F., Kieslich, K., Starke, C., & Lünich, M. (2020). Implications of AI (Un-)Fairness in Higher Education Admissions: The Effects of Perceived AI (Un-)Fairness on Exit, Voice and Organizational Reputation. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 122–30. Barcelona Spain: ACM. https://doi.org/10.1145/3351095.3372867.

McKinsey Global Institute. (2018). Notes From the AI Frontier: Modeling the Impact of AI on the World Economy. McKinsey Global Institute. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx.

Mehrabi, N., Morstatter, F., Saxena, F., Lerman, K., & Galstyan, A. (2019). A survey on bias and fairness in machine learning. *[Cs]*, September. http://arxiv.org/abs/1908.09635.
Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications Inc.
Ministry of Economic Affairs and Employment, Finland. (2017). Finland: Finland's Age of Artificial Intelligence Turning Finland into a Leading Country in the Application of Artificial Intelligence. 47/2017. Helsinki, Finland: Ministry of Economic Affairs and Employment. http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf.
Ministry of Finance and Ministry of Industry, Business and Financial Affairs, Danish Government. (2019). Denmark: National Strategy for Artificial Intelligence. Copenhagen, Denmark: Ministry of Finance and Ministry of Industry, Business and Financial Affairs, Danish Government. https://investindk.com/-/media/invest-in-denmark/files/danish_national_strategy_for_ai2019.ashx?la=en.
Ministry of Science, Innovation and Universities, Spain. (2019). Spanish RDI Strategy in Artificial Intelligence. Madrid, Spain: Ministry of Science, Innovation and Universities. http://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_EN.PDF.
Ministry of the Economy and Innovation. (2019). Lithuanian Artificial Intelligence Strategy: A Vision of the Future. Vilnius, Lithuania: Ministry of the Economy and Innovation. http://kurklt.lt/wp-content/uploads/2018/09/StrategyIndesignpdf.pdf.
Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D. & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, 220–29. https://doi.org/10.1145/3287560.3287596.
Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society, 3*(2), 205395171667967. https://doi.org/10.1177/2053951716679679
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine, 6* (7): e1000097.https://doi.org/10.1371/journal.pmed.1000097
Morganstein, D., & Wasserstein, R. (2014). ASA Statement on value-added models. *Statistics and Public Policy, 1*(1), 108–110. https://doi.org/10.1080/2330443X.2014.956906
Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-019-00165-5
NITI Aayog, India. (2018). India: National Strategy for Artificial Intelligence #AIFORALL. Delhi, India: NITI Aayog. http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf.
Noble, D. F. (1998). Digital diploma mills: The automation of higher education. *Science as Culture, 7*(3), 355–368.
Nye, B. D. (2015). Intelligent tutoring systems by and for the developing world: A review of trends and approaches for educational technology in a global context. *International Journal of Artificial Intelligence in Education, 25*(2), 177–203. https://doi.org/10.1007/s40593-014-0028-6
Oetzel, M. C., & Spiekermann, S. (2014). A systematic methodology for privacy impact assessments: A design science approach. *European Journal of Information Systems, 2*3(2), 126–150. https://doi.org/10.1057/ejis.2013.18.
Office of the President, Russia. (2019). Russia: National Strategy for the Development of Artificial Intelligence Over the Period Extending up to the Year 2030. Moscow, Russia. http://kremlin.ru/events/president/news/60630.
Office of the Prime Minister, Malta. (2019). Malta: Towards An AI Strategy. Valletta, Malta: Parliamentary Secretariat for Financial Services, Digital Economy and Innovation, Office of the Prime Minister, Malta. http://malta.ai/wp-content/uploads/2019/04/Draft_Policy_document_-_online_version.pdf.
Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Services Research, 14*(1), 2. https://doi.org/10.1186/1472-6963-14-2
Oswald, M., Grace, J., Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: Lessons from the Durham hart model and 'experimental' proportionality. *Information and Communications Technology Law, 27*(2), 223–250. https://doi.org/10.1080/13600834.2018.1458455
Parkhurst, J. (2017). *The politics of evidence :From evidence -based policy to the good governance of evidence*. Taylor & Francis. https://library.oapen.org/handle/20.500.12657/31002.

Phoenix, J. H., Atkinson, L. G., & Baker, H. (2019). Creating and communicating social research for policymakers in government. *Palgrave Communications, 5*(1), 98. https://doi.org/10.1057/s41599-019-0310-1
Pinkwart, N. (2016). Another 25 years of AIED? Challenges and opportunities for intelligent educational technologies of the future. *International Journal of Artificial Intelligence in Education, 26*(2), 771–783. https://doi.org/10.1007/s40593-016-0099-7
Porter, M. E. (1996). Competitive advantage, agglomeration economies, and regional policy. *International Regional Science Review, 19*(1–2), 85–90. https://doi.org/10.1177/016001769601900208
Qatar Center for Artificial Intelligence. (2019). "Qatar: Blueprint: National Artificial Intelligence Strategy for Qatar." Ar-Rayyan, Qatar: Qatar: Qatar Center for Artificial Intelligence (QCAI), Qatar Computing Research Institute (QCRI), Hamad Bin Khalifa University. https://qcai.qcri.org/wp-content/uploads/2019/02/National_AI_Strategy_for_Qatar-Blueprint_30Jan2019.pdf.
Radaelli, C. M. (2009). Measuring policy learning: Regulatory impact assessment in Europe. *Journal of European Public Policy, 16*(8), 1145–1164. https://doi.org/10.1080/13501760903332647
Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44. FAT* '20. Barcelona, Spain: Association for Computing Machinery. https://doi.org/10.1145/3351095.3372873.
Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic impact assessments: A practical framework for public agency accountability. New York, NY: AI Now. https://ainowinstitute.org/aiareport2018.pdf.
Roberts, K., Dowell, A., & Nie, J.-B. (2019). Attempting rigour and replicability in thematic analysis of qualitative research data: A case study of codebook development. *BMC Medical Research Methodology, 19*(1), 66. https://doi.org/10.1186/s12874-019-0707-y
Roolaht, T. (2012). The characteristics of small country national innovation systems. In E. G. Carayannis, U. Varblane, & T. Roolaht (Eds.)*In novation Systems in Small Catching-Up Economies: New Perspectives on Practice and Policy* (pp. 21–37). Springer. https://doi.org/10.1007/978-1-4614-1548-0_2.
Rothstein, H. R., & Hopewell, S. (2009). Grey literature. *The handbook of research synthesis and meta-analysis* (pp. 103–25). Russell Sage Foundation.
Schendel, R., & McCowan, T. (2016). Expanding higher education systems in low- and middle-income countries: The challenges of equity and quality. *Higher Education, 72*(4), 407–411. https://doi.org/10.1007/s10734-016-0028-6
Schiff, D. (2021). Out of the laboratory and into the classroom: The future of artificial intelligence in education. *AI & Society, 36*(1), 331–348. https://doi.org/10.1007/s00146-020-01033-8
Schiff, D., Ayesh, A., Musikanski, L., & Havens, J. C. (2020a). IEEE 7010: A new standard for assessing the well-being implications of artificial intelligence. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2746–53. https://doi.org/10.1109/SMC42975.2020.9283454.
Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020b). What's next for AI ethics, policy, and governance? A global overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 153–58. AIES '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3375627.3375804.
Schiff, D., Borenstein, J., Laas, K., & Biddle, J. (2021). AI ethics in the public, private, and NGO sectors: A review of a global document collection. *IEEE Transactions on Technology and Society, 2*, 31–42. https://doi.org/10.1109/TTS.2021.3052127
Schomberg, R. von. (2013). A vision of responsible research and innovation. *In Responsible Innovation* (pp. 51–74). Wiley-Blackwell. https://doi.org/10.1002/9781118551424.ch3.
Scott, J., Lubienski, C., DeBray, E., & Jabbar, H. (2014). The intermediary function in evidence production, promotion, and utilization: The case of educational incentives. In K. S. Finnigan & A. J. Daly (Eds.), *Using research evidence in education* (pp. 69–89). Cham: Springer.
Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. FAT* '19. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3287560.3287598.
Sin, C. H. (2008). The role of intermediaries in getting evidence into policy and practice: some useful lessons from examining consultancy-client relationships. *Evidence and Policy: A Journal of Research, Debate and Practice, 4*(1), 85–103. https://doi.org/10.1332/174426408783477828
Smart Nation and Digital Government Office, Singapore. (2019). Singapore: National Artificial Intelligence Strategy: Advancing Our Smart Nation Journey. Singapore: Smart Nation and Digital Government

Office. https://www.smartnation.sg/docs/default-source/default-document-library/national-ai-strategy.pdf.

Stahl, B. C., & Coeckelbergh, M. (2016). Ethics of healthcare robotics: towards responsible research and innovation. *Robotics and Autonomous Systems, 86*(December), 152–161. https://doi.org/10.1016/j.robot.2016.08.018

Stern, C., Jordan, Z., & McArthur, A. (2014). Developing the review question and inclusion criteria. *AJN the American Journal of Nursing, 114*(4), 53–56. https://doi.org/10.1097/01.NAJ.0000445689.67800.86

Terhart, E. (2013). Teacher resistance against school reform: Reflecting an inconvenient truth. *School Leadership and Management, 33*(5), 486–500. https://doi.org/10.1080/13632434.2013.793494

The Institute for Ethical AI in Education. (2020). Interim report: Towards a shared vision of ethical AI in education. The institute for ethical AI in education. https://www.buckingham.ac.uk/wp-content/uploads/2020/02/The-Institute-for-Ethical-AI-in-Educations-Interim-Report-Towards-a-Shared-Vision-of-Ethical-AI-in-Education.pdf.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237–246. https://doi.org/10.1177/1098214005283748

Lee, N. T., Resnick, P., & Barton, G. (2019). Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms. Brookings Institition. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/.

Villani, C., Marc, S., Bonnet, Y., Charly, B., Anne-Charlotte Cornut, François Levin, & Rondepierre, B. (2018). "France: For a Meaningful Artificial Intelligence: Towards a French and European Strategy." French Parliament. https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.

Warschauer, M., & Ames, M. (2010). Can One Laptop per Child save the world's poor?. *Journal of International Affairs*, *64*(1): 33-51.

West, D. M. (2018). *The future of work: Robots, AI, and automation*. Brookings Institution Press.

White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library Trends, 55*(1), 22–45. https://doi.org/10.1353/lib.2006.0053

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S.M., Richardson, R., Schultz, J., & Schwartz, O. (2018). AI Now Report 2018. New York, NY: New York University. https://ainowinstitute.org/AI_Now_2018_Report.pdf.

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics, 2*(2), 46–48. https://doi.org/10.1038/s41928-019-0213-6

Women Leading in AI. (2019). Women leading in AI: 10 Principles of responsible AI. Women Leading in AI. https://womenleadinginai.org/wp-content/uploads/2019/02/WLiAI-Report-2019.pdf.

Woolf, B. P., Ivon, A., David, C., Winslow, B., & Kasia, M. (2010). Affective tutors: Automatic detection of and response to student emotion. In N. Roger, B. Jacqueline, & M. Riichiro (Eds.), *Advances in intelligent tutoring systems* (pp. 207–27). Berlin, Heidelberg: Springer.

Zarsky, T. Z. (2016). Incompatible: The GDPR in the age of big data. *Seton Hall Law Review*, 47, 995–1020.

Zeng, Y., Enmeng L., & Cunqing H. (2018). Linking artificial intelligence principles. *[Cs]*, December. http://arxiv.org/abs/1812.04814.

Zhang, D., Saurabh, M., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J. C., Sellitto, M., Shoham, Y., Clark, J., & Perrault, R. (2021). Artificial Intelligence Index Report 2021. AI Index Steering Committee, Human-Centered AI Institute, Stanford University. https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf

Zhou, Y., & Danks, D. (2020). Different 'intelligibility' for different folks. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 194–99. AIES '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3375627.3375810.