AI hyped? A horizon scan of discourse on Artificial Intelligence in Education (AIED) and development

Selena Nemorin

University of Oxford

Andreas Vlachidis

University College London

Hayford M. Ayerakwa,

University of Ghana

Panagiotis Andriotis

University of West England

Corresponding author: Selena Nemorin – selena.nemorin@education.ox.ac.uk

Abstract

The study seeks to understand how the AI ecosystem might be implicated in a form of knowledge production which reifies particular kinds of epistemologies over others. Using text mining and thematic analysis, this paper offers a horizon scan of the key themes that have emerged over the past few years during the AIEd debate. We begin with a discussion of the mixed-methods tools we used to experiment with digital methods for data collection and analysis. The paper then examines how AI in education systems are being conceived, hyped, and potentially deployed into global education contexts. Findings are categorised into three themes in the discourse: 1) Geopolitical dominance through education and technological innovation; 2) Creation and expansion of market niches, and 3) Managing narratives, perceptions, and norms.

Introduction

Artificial intelligence (AI) is heralding what is increasingly discussed as the Fourth Industrial Revolution; turning on the extraction and exploitation of data for 'smart' governance through 'data-driven' strategies. AI, or "the theory and development of computer systems able to perform tasks normally requiring human intelligence" (Jobin et al, 2019, p. 389), is also presented as having the capacity to solve some of the world's most pressing problems across domains such as health, finance, housing, transport, and education. The technology has attracted increased interest given its high-profile successes in simulations, games, smart city operating systems, and self-driving vehicles. A range of approaches to AI such as machine learning, deep learning and artificial neural networks have framed the way in which data processing and analysis are now being conducted.

Scholars have claimed that commercial applications of AI in society often have a positive social impact by virtue of increasing availability and accessibility of information; a result of more efficient search tools and language-translation tools, provision of better information communication services, optimised transportation systems, and personalised healthcare and education. Current debates have focussed on AI in education, particularly how the AI ecosystem might be harnessed to progress global education given the need to move teaching and learning online because of the COVID-19 pandemic. As a result, a number of AI-branded products have begun competing for a place in education. Intelligent assistants, for instance, have been touted as having the capacity to take on mundane administrative responsibilities of teachers, such as tracking attendance, and developing lesson plans and classroom activities, thereby freeing up teachers' time to do more productive things (UNESCO, 2021; Baker et al, 2019). As a consequence of the pandemic, companies around the world have scrambled to gain entry into a lucrative educational technology opportunity, spurring calls and demand for AI in education as a solution to accessible remote learning.

While there are increasing critiques of the uses of AI in education, the popular media and a range of online spaces representing AI in education stakeholders show widespread acceptance that AI is the future of education.

Some of the more common claims made about AI is that it offers the potential to accelerate attainment of global education goals by reducing barriers to access education, automating management processes, and optimising methods for improvement of learning outcomes, helping close the digital skills gap and render education more equitable and accessible (Mann & Hilbert, 2018). Positive rhetoric can be found in the discourse of global governance institutions. For instance, Audrey Azoulay, the Director-General of UNESCO (2019a), claims: "Education will be profoundly transformed by AI…Teaching tools, ways of learning, access to knowledge, and teacher training will be revolutionized." Other actors refer to AI as a way of modernising education, providing learning analytics and AI to "accelerate the pace of our understanding," and inform educational strategies (Microsoft, 2021, p. 2). In addition, numerous policy documents have tended to focus on the economic benefits of AI which are expected to improve levels of efficiency, optimisation, and productivity in the labour force. These values promise increased wealth and wellbeing and are said to provide the conditions necessary for human flourishing thus tying the language of AI ideals to well-being. As the EU High-Level Expert Group on AI stated (2019):

> AI is not an end in itself, but rather a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation. (p. 4)

AI has also taken centre stage in debates about how school governance, pedagogy, and learning can be rebooted to fit students with the 21st Century Skills essential for participation in larger

society. In order to realise this vision, public and private partnerships (P3s/PPP) are being established to manage AI initiatives to spur digital transformations in education, innovation, and growth. Policy makers promoting P3s have been guided by the fact that a number of governments in poorer countries have struggled to deliver on Education For All (EFA), Millennium Development Goals (MDG), and now the Sustainable Development Goals (SDG) commitments. However, the justifications for P3s have been guided by issues of governance and rollout instead of agendas addressing round rights or equalities, and how to consider these issues when thinking about local needs (Unterhalter, 2017).

On the rise are initiatives that nurture interdisciplinary and cross-sector partnerships around AI applications as solutions for the SDGs, The AI for Social Good (AI4SG) (Tomašev et al, 2020) initiative, for example, offers guidelines for establishing long-term collaborations between AI researchers and application-domain experts; they would then be connected to existing AI4SG projects with identification of opportunities for collaborations on AI applications targeted towards social good. The kind of philanthrocapitalism evidenced here is also identified in general discourses from groups such as Tech for Good initiatives which see themselves as shaping the global conversation on the development of ethical AI. Many transnational corporations such as PWC, Microsoft, Intel, Squirrel AI, and others have honed their brands on AI and are also pushing the AI for good narrative, as evidenced in their missions and goals statements as well as their work on ethical/responsible AI systems.

The present paper marks first steps in a series of papers that will examine a larger, context-based study of AI in education and development. Using a mixed-methods approach to textual analysis (text-mining and discourse analysis), the paper offers an overview of the discourse around AI in education through analysis of recurring themes and general assumptions. The paper begins with an examination of how AI in education systems are

being conceived, developed, hyped, and potentially deployed into global education contexts. Findings are categorised into three themes in the discourse:

- Geopolitical dominance through education and technological innovation

- Creation and expansion of market niches

- Managing narratives, perceptions, and norms

In general, the present study seeks to understand how AI as a socio-technical system might be implicated in a form of knowledge production which reifies particular kinds of epistemologies over others. These kinds of knowledge, if their assumptions and structures are left unexamined and unchallenged, may result in continued forms of bias against historically oppressed bodies (Eubanks, 2018; Noble, 2018; O'Neil, 2017). For the purposes of the present paper, we will be offering a horizon scan of the themes that have emerged over the past few years of observing the AIEd hype grow. We begin with a discussion of the mixed-methods tools we used to experiment with digital methods for data extraction and analysis. This is followed by a horizon scan of the discourse landscape ending with our concluding thoughts.

Research methods

Discourse analysis

Discourse reflects and creates the social world. Analysis of discourse allows us to identify how social actors contextualise practices in texts which not only gives them legitimacy but also allows the perpetuation of power dynamics. Through controlling the discourse dominant knowledge is created. This paper reports on a discourse analysis of a corpus of 143 publicly available documents specific to AI and international education and development. The corpus includes policy documents, manifestos, guidelines, articles, reports, conference agendas, and brochures published by global governance institutions, governments, non-governmental

organisations, industry and academic groups. Texts were selected to represent the range of actors contributing to the discourse around AI and education, with a view to obtaining a general overview of how AI has been discussed in the space of education and international development. The analysis was conducted between January 2019 - August 2021. Each document ranged from 3-818 pages (Figure 1). The discourse analysis involved two phases: 1) horizon scanning and 2) thematic analysis.
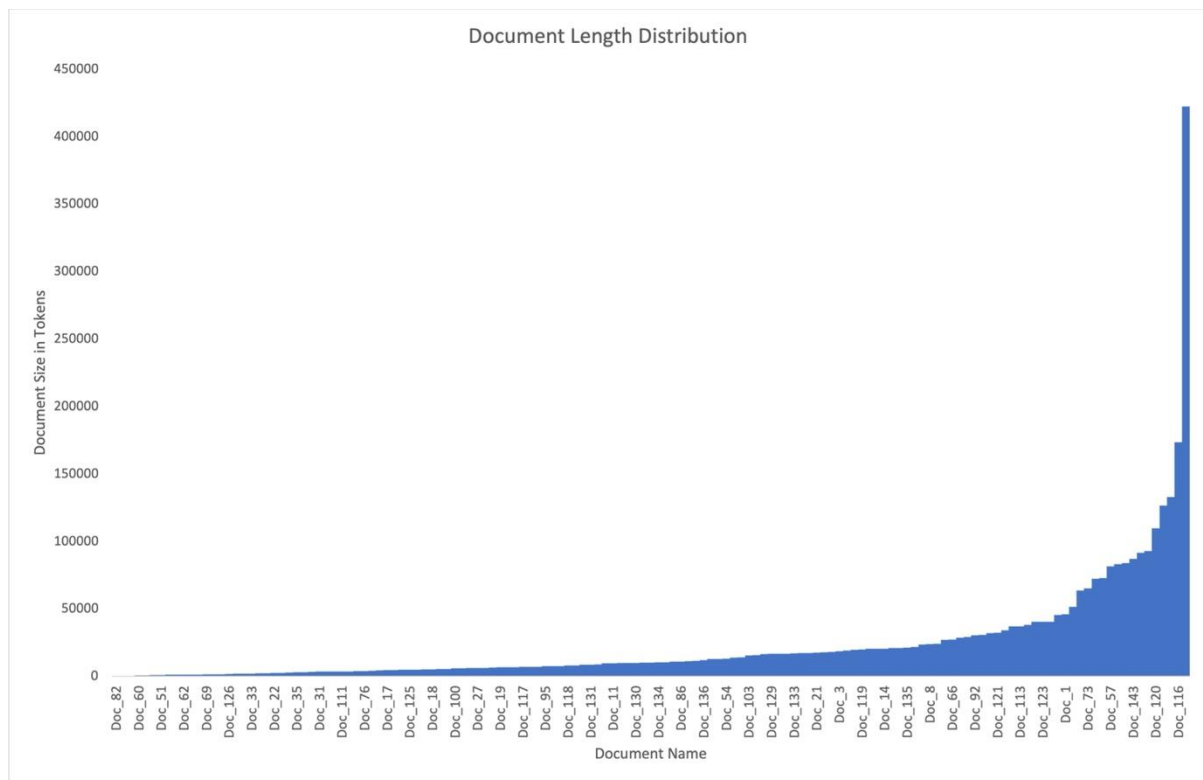
Figure 1: Document length in Tokens (ie., single words, symbols and punctuation).

PHASE I

Horizon scanning: Natural Language Processing

Phase 1 of the study began with categorising the topic into concepts and questions that could be managed and dissected for database use. Associative keywords were determined that could be used in the horizon scan. The study produced a list of 26 keyword combinations using "AI/Artificial Intelligence" as the base-keyword and a range of other keywords acting as contextual triggers. Keywords were selected according to what were perceived to be recurring themes in the general debates around AI at the time. These terms were divided into three primary groups for searches. Multiple searches were conducted using the Keyword Group 1 in combination with all of the keywords in Groups 2 and 3 (e.g. 'artificial intelligence' AND 'education' AND 'sustainable development'). Upon identification, relevant texts were captured in a database according to date, document type, institution, author(s), and country.

When screening the corpus, titles and keyword were used to understand the primary information the resource might offer. For example, if a document appeared to be relevant to the role of industry in education and development, the citation was highlighted. The document would then be retrieved from the documents library and read closely. The context of the highlighted results were then read more closely to begin formulating codes and themes (Srivastava & Oh, 2010).

A range of Natural Language Processing (NLP) techniques were used for revealing common information patterns and textual evidence in support of the study. In particular, our research employed two distinct methodological approaches for the automatic processing of the document collection. A bottom-up approach which employed corpus analysis techniques for investigating and revealing contextual examples of discussion in relation to the role and implications of AI in education. A top-down approach targeted at extracting particular pieces

of information about entities of interest in an attempt to retrieve insights into the relations between people, places, and organisations that may influence the policies and practices of AI in the educational context. The following sections discuss the results of two separate processes applied during the course of our research: *Concordance and Thematic Analysis.*

(i) Concordance

In the context of our research we refer to Concordance to describe a range of word-based techniques that have been employed for finding frequently occurring contextual evidence in the corpus of investigation. Our word-based investigation used word frequency, collocation, and keyword-in-context (KWIC) techniques for producing a range of outputs which revealed commonly occurring words and phrases in the document collection. Initially, a simple word-frequency analysis was conducted to reveal the most frequently occurring textual instances in the corpus. The frequency-based technique known as Collocation was then followed for revealing the most frequently occurring word sequences. The analysis addressed co-occurring word instances in form of pairs (bigrams) and of groups of three words (trigrams).

KWIC introduces top-down characteristics for extracting concordance lines from text. The technique extracts pieces of text and reveals the contextual evidence surrounding such targeted keywords. The KWIC task extracted 150-character long pieces of text which contained the base-keyword and any contextual keyword located five places to the left or to the right from the base-keyword (Figure 2). The extracted pieces were then casted into HTML pages which compiled the extracted information of each document in the collection. Figure 1 presents an example of such an extract that shows the keyword phrase, the immediate context of the keyword phrase and the larger section in which the phrase has been extracted from. Extracts were used to support quick inspection of documents and draw

attention to passages of information which potentially carry interesting and useful elements for discussion.

AI promises to enhance education

Though quality education will always require active engagement by human teachers, **AI promises to enhance education** at all levels, especially by providing personalization at scale.
Enabling more fluid interactions between people and promising AI technologies also remains a critical challenge in Education , which has seen considerable progress in the same period. Though quality education will always require active engagement by human teachers, AI promises to enhance education at all levels, especially by providing personalization at scale.

Figure 2: Keyword in Context extract for the combination AI + education

(ii) Thematic analysis

Phase II comprised coding, layering, and thematising. Textual points of interest were identified in text mining data. Relevant documents were retrieved for more in-depth reading. Using an iterative process of inductive coding recurring ideas were annotated and used to develop refined themes (Van Leeuwen, 2008; Braun & Clarke, 2006). Thematic analysis was employed as a process comprised of five main stages: *Familiarity with the data*; *Generating initial codes; Looking for themes*; *Reviewing themes*; and *Naming and defining themes* (Tracy, 2010).

Thematising revealed a number of distinct ideas embedded within the corpus. For the purpose of the present work, we have focused on the most common themes that, when taken together, highlight the orders of discourse that prioritise geopolitical and economic initiatives while presenting AI as a transformative and modernising solution to the state of education. We acknowledge that many potential areas of focus emerged and will continue to emerge from the data in light of the largely qualitative nature of this study.

#1 Geopolitical dominance through education and technological innovation

As the OECD highlights, AI research and development is being driven by private investments primarily situated in the West and in China. Desire for geopolitical dominance - through AI and education and development – is evidenced in the general discourse of the World Economic Forum (WEF), to offer one example. The discourse speaks of fear for the potential dominance of China. Innovation at scale is proposed as a solution for the West's domination of the *other*. The WEF (2020) states this clearly in its report:

> China's surge into Africa in the 1990s, in search of food, mineral and energy resources to power its growth, helped to pull more than a dozen African nations into middle-income status. But with global economic reach comes global interests and the temptation to project global power; now China has moved into a new phase of expansion – into a global network of ports, technology plays and infrastructure assets that in some theatres seem purposefully designed to challenge the West. (p. 14)

The author continues with discussion about the need for innovation in particular technologies which could affect geopolitical power with a focus on competition instead of collaboration:

> Added to this are powerful changes in the technological sphere: not only the deeper and now pervasive integration of cyber networks into military technology but also the wide penetration of social networks – and above all qualitative leaps in the effectiveness and power of supercomputing, artificial intelligence and biotechnology. Any one of these technologies could amplify shifts in the balance of geopolitical power. (WEF, 2020, p. 15)

Staying with the World Economic Forum, we see how forms of colonialism also emerge in light of the need for geopolitical power. Given China is strong, the West must draw on its assets in order to be competitive:

> Cooperation with other democracies would strengthen the West's hand: in the realm of data and technology, the West should strengthen ties with India, whose data sets and tech entrepreneurs will be valuable assets in the coming competition, as well as with Mexico, whose technology and infrastructure grids can either be the soft underbelly or the strategic reserve of the West. (p. 15)

In the case of the excerpt above, Mexico is viewed as strategic reserve for the West, a resource that can be exploited by transnational techno-industry actors. This geopolitical climate also allows for transnational mobility when admitting and retaining immigrants who are highly skilled to the West, for example (NSC, 2021). In this capacity the WEF asserts the authority of the United States as the driver of technological advancements that have been the "backbone" of international order. The threat is explicitly named as "non-Western powers" and "non-state actors" who would use any opportunity to weaken the Western alliance.

The assumption of Western liberalism institutions as directing the way forward for other countries sets up a discursive colonial mentality, one where other allies such as those from developing contexts are seen as *standing reserve* (Heidegger, 1982). Here, human beings are used strategically, as commodities. In this kind of world, nature – including human beings - become resources for technical applications such as artificial intelligence, which is then taken as a resource for further use, thus tying human beings to technology in a way that reduces the individual to instrumentality.

The WEF is not alone in framing its discussions around AI education and societal imperatives from an economic perspective. At the heart of many OECD policies specific to AI is an economic position, however, unlike the above chapter from the WEF report, the OECD speaks more of collaboration across governments when considering principles of AI in society as summed up by the OECD AI Policy Observatory (2021). Governments, public research, and industry play a key role in leveraging AI to help tackle global challenges such as the SDGs with similar thinking from the UN (ITU, 2018).

The expansion of geopolitics can be seen across a number of areas – economic, environmental, technological, and where education is claimed to be the ground upon which citizens are built and through which human flourishing can be attained. In a collaboration between UNESCO (2015), the Global Education First Initiative, and the Education For All Global Monitoring report, Ban Ki-Moon states:

> Prosperous countries depend on skilled and educated workers. The challenges of conquering poverty, combatting climate change and achieving truly sustainable development in the coming decades compel us to work together. With partnership, leadership and wise investments in education, we can transform individual lives, national economies and our world.

The majority of the documents collected for our preliminary scan demonstrate themes and objectives largely grounded on a socio-economic rationale. These factors tend to cut across global governance documents. While education for development is noble, there are dominant discursive frames that appear in the texts which speak to an overall aim of potentially transforming and reskilling the workforce of the future with the STEM capabilities necessary for productive integration into an AI-driven labour market. This AI-driven market is one that

is pioneered by industry. The technology industry leaders stand to gain windfall amounts from investments in AI-powered technologies (O'Keefe et al, 2020).

Four primary trends have tended to guide international investment in education for development. First of all is the design and development of resources to use education as a tool to contribute to poverty reduction. This trend hinges on the importance of education for national economic growth and development. One example of this tendency is evidenced in the report by the OECD (2021). The recently published OECD Education *Outlook 2021: Pushing the Frontiers of AI, Blockchain, and Robots* report is largely about how to be ready for these new technologies being used in the classroom as a "smart" school space—the language contained therein is attached to economic growth as a driver, and how deploying particular kinds of AI infused technologies in the classroom might work for innovation in education, not least inclusion and equity. Technologies discussed include social robots as teachers, blockchain for education, personalisation of learning using hybrid human-AI learning technologies, and adaptive technologies that can be used at scale.

The second trend guiding international investment in education is the use of standardised measurement and testing of student learning. Actors include the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment, PISA; International Association for the Evaluation of Educational Achievement (IEA); Trends in International Mathematics and Science Study (TIMSS); and the Progress in International Reading Literacy Study (PIRLS).

Third, is the emergence of international policy declarations approved by a majority of world states. These declarations are driven by policies and initiatives developed by international organisations under the umbrella of the United Nations (UN), such as the United Nations Educational, Scientific and Cultural Organization (UNESCO), the United Nations

Children's Emergency Fund (UNICEF), as well as the World Bank, the World Economic Forum (WEF), and the Organisations for Economic and Cultural Development (OECD).

Fourth, is the growing role of the private sector in design and the provision of education (Draxler, 2014). In this context, education is not only seen as a good that nurtures human flourishing; it is also a sector from which significant profits can be made, especially in light of innovations in AI educational technologies. Through AI in education is a path to a global market share valued at US$1.1 billion in 2019, and it is expected to reach US$6 billion in 2024 and US$25.7 billion by 2030 (Holmes et.al, 2021).

#2 Creation and expansion of market niches

The World Bank defines P3s as a "long-term contract between a private party and a government entity, for providing a public asset or service, in which the private party bears significant risk and management responsibility, and remuneration is linked to performance" (World Bank, 2019). Since the 1990s, P3s have pushed for the enhanced provision of education in both the Global North and South; they have become a major part of the infrastructure of high, middle-, and low income countries since mid-2000. These partnerships have been lauded as a promising way of financing and delivering quality educational content to developing countries. Much of this is rooted in philanthropy and policy advocacy; thereby allowing P3s access to the governance processes specific to educational policy making. P3s have been promoted by a range of cross-sector actors including United Nations agencies, the World Bank Group, the World Economic Forum, industry, governments, NGOs, and philanthropic organisations. The SDGs specifically refer to the use of P3s as "means of implementation" for other development goals, in relation to poverty, health, education, and environmental protection (Gideon & Unterhalter, 2021).

P3s are increasingly being established to roll out AI in education initiatives in developing countries to spur transformations in education: modernisation, innovation, and growth. For example, partnering with policy makers, start-ups, technology partners, and civil society groups, Microsoft launched its initiative 4Afrika in an attempt to solve challenges that impact Sub-Saharan Africa and drive growth and development in core sectors such as healthcare, agriculture, and public sector areas such as financial services and education. Microsoft Education is one example of a stakeholder in P3 initiatives in education for development, as are Google, Facebook, ALiPay and Amazon to name a few. Suggesting its predictions for the widespread use of AI, Microsoft has rebranded itself as an AI company.

Microsoft has played a substantial role in education and development for some time. The business is becoming increasingly involved with learning analytics as a solution for education post-pandemic. In light of their recent publication, Microsoft Education views its role as an AI solutions provider, where AI and learner analytics - using continuous real-time data – can offer numerous benefits for school governance. (Microsoft, 2021, p. 5). The following excerpt highlights the way Microsoft potentially understands education as an industry that can be modelled on the advertising template:

> At the same time, other industries are making use of real-time data to assess progress against goals and strategies continuously. Retail stores using AI can determine within 24 hours the impact on purchasing behaviors in response to moving the shelf location of a product. Education systems should be able to see the impact on student well-being and learning of changes in teaching strategies, tools, and programs just as quickly, so they can become more proactive in diagnosing and acting in ways that truly support every learner. (p. 5)

Here is one example of the view that it is acceptable to nudge students into desirable behaviours, as in the role of advertising and selective product placement which alters behaviours. This form of niche creation can be viewed in light of previous work on niche creation in advertising where the individual's phenomenal world is analysed to create advertisements that connect to customer affectively (Nemorin, 2018). In the case of education: designing learning content to engage and monitor students through affective digital immersion. The general rationale that has tended to drive the celebratory discourse of AI in education is that vast amounts of data can be used to improve management of student populations, including student behaviours and dispositions, and academic performance. Behavioural and skill-based changes are enacted as personalisation and informed by data obtained through neuro- and biometric trackers. Numerous other bodily monitoring instruments are currently being developed, tested, and deployed to connect learners to machines through multimodal and multi-sensory techniques (OECD, 2021).

When examined closely, at the core of many current AI-driven educational initiatives lies a computational understanding of education and learning that reduces student and teacher lifeworlds to sets of data logics that can be managed and understood. Underpinned by an instrumental rationality and a desire for mechanisation and control of bare life (*zoe*), these processes of datafication (van Dijck, 2014) seek to make schooling as a lived practice knowable, predictable, and thereby governable. The datafication of education in discourse is aligned with positivist thinking and reductionist impulses. In fact, AI and datafication of education seems to be an inevitability in the discourse and explicitly outline in the UNESCO (2019a) report *The challenges and opportunities of Artificial Intelligence in education*:

> Developing quality and inclusive data systems: If the world is headed towards the datafication of education, the quality of data should be the main chief concern. It´s essential to develop state capabilities to improve data collection and systematization.

16

AI developments should be an opportunity to increase the importance of data in educational system management. (p. 7)

As well as positioning AI in school as an inevitability, AI is also presented as a necessity for optimised "data-driven" and "evidence-based" educational governance. This view of AI in education rests on the assumption that no space in the human body is sacred enough to be protected from the creep of AI's attention. This social imaginary suggests that every aspect of bare life is and should be thrown open for measurement and behavioural management via timely nudges, for example.

As part of a culture of measurement, such findings that are used by governance bodies are justified as efforts to raising standards (Biesta, 2009), but this form of automation of schooling can lead to educational governance being enacted in ways that reproduce and amplify forms of exclusion and discrimination, and assimilate difference and the less measurable ways of being-in-the-world into a totalising mono-structure of power through knowledge. As such, the consequence is the potential reproduction of inequalities, specifically "social power and control being reinforced, or perhaps reconstituted, through data-driven processes" (Selwyn, 2015, p.71).

The ways in which students potentially become extractable resources to generate [public] data can be understood as viewing subjects as objects, of seeing students as data nodes for extraction and exploitation of valuable data. Students can be seen in this sense as having limited agency and no control over their personal data, including having a say in the decisions made using their data. (Couldry & Mejias, 2019). These kinds of moves pose social and ethical implications. For example, mining the emotional lives of students is, as McStay (2020) argues, normatively wrong, particularly when the value extraction is not serving the best interests of students. Some of these AI-neurological methods of connecting students to techniques of measurement and addictive technologies are also pushing boundaries of bodily

17

sovereignty in the sense that surveillant technologies are increasingly intruding into the integrity of the human body and fundamental human rights to cognitive freedom— the right to mental self-determination. All aspects of the child being measured to analyse and know its niche behaviours. Aspects of the child and then represented as educational and marketing materials, and confined to a learning world that is decided upon by an artificial intelligence through strategies of personalisation.

Also located in a number of other documents is the explicit desire to regulate and know the student's phenomenal environment which would seem at odds with the kind of regulatory discourse emerging from EU discussions which suggest that systems trained on and exploiting affect ought to be prohibited. In April, the European Commission (2021) presented its Proposal for a Regulation AI, including harmonising rules for AI. In their response to the proposal, the European Data Protection Board and the European Data Protection Supervisor welcomed the concerns of the legislating body in addressing these issues, highlighting the importance of the proposal with regards to data protection implications. EDPB and the EDPS considered that intrusive forms of AI such as the kind that may affect human dignity ought to be deemed as as high-risk, The EDPS and EDPS also viewed "the use of AI to infer emotions of a natural person is highly undesirable and should be prohibited" (p. 3).

Taking a similar stance, the AI Now Institute (2020) as part of the consultation process on the European Commission's White Paper on Artificial Intelligence - A European Approach submitted as a key recommendation:

> Given its contested scientific foundations and evidence of amplifying racial and gender bias, affect recognition technology should be banned for all important decisions that impact people's lives and access to opportunities. (p. 2)

Indeed, challenges include the problem of categorising, sorting, and affective modulation or 'nudging' of students (Nemorin, 2015) through intrusive surveillance practices attached to extracting data on life, or biological data such as eyeball tracking mood surveillance, and overall student surveillance using more mundane technologies as well as the AI powered systems that have started emerging over the past few years, resulting in a systems-informed learning process which has become increasingly the norm post COVID pandemic. What is problematic here is that some of these ways of approaching students in simplified terms can be viewed, and reasonably so, as experimentation that violates a student's right to: [digital] privacy, 2) cognitive freedom, and, overall, 3) dignity of self. Also at question is the way in which personalisation of learning by AI and learning analytics potentially creates a behavioural market niche, a filter bubble (Parisier, 2011) profile so to speak which the student cannot escape while progressing through an educational system increasingly delimited by algorithms.

#3 Managing narratives, norms, and perception

There is some form of consensus around the vital values that ought to guide the promotion and use of AI. Key among the list are fairness, accountability, transparency, and explainability. Notwithstanding the importance of these principles as critical for the development of AI, the world is far from harmonising on what will constitute universally acceptable norms for the development and use of AI (Ray, 2019). At the moment, the search for universally acceptable norms has in itself resulted in a race with governments, international organisations, and private sector actors churning out policies and documents that seek to promote AI norms. For example, the Artificial Intelligence Strategy for the German Federal Government (2018) underscores the importance of promoting responsible AI in an ethically and legally sound environments.

The Federal Government sees it as its duty to drive forward the responsible use of AI which serves the good of society. Here, we are adhering to ethical and legal principles consistent with our liberal democratic constitutional system throughout the process of developing and using AI. (p. 9)

The OECD emphasises the importance of having a common framework that can be used to asses and compare trustworthy AI. The framework for trustworthy AI aims to identify tools for development, using and deploying trustworthy AI that accounts for the different contexts and also upholds the principle of human rights, fairness, transparency, robustness etc. (OECD, 2021). Similarly, a UK report titled *AI in UK: ready, willing and able* gives a strong indication of what the UK Government is thinking and planning. The report acknowledges the potentials and risks associated with AI and calls for the development of a framework and mechanisms such as data portability and data trust. It proposes the need to put a framework that will guide the production, deployment and use of AI in the UK as it seeks to lead the way as an example to the international community.

It is clear that the promotion of responsible AI is an attempt to court the public's support and trust in the government and among private sector actors. Despite consensus that AI should be ethical, there is also disagreement on what comprises ethical AI, and the ethical criteria, technical standards and protocols necessary for its implementation. To be sure, many of these initiatives have shown to be helpful when thinking about the practicalities of enacting AI in education. Yet, as observers have noted the creation of common norms as ethical principles, while mostly of benefit, are also problematic as they tend to be high level and are susceptible to manipulation, especially by industry (Resseguier & Rodrigues, 2020). This resonates with Ochigame's (2019) claims that the rhetoric of ethical AI "was aligned strategically with a Silicon Valley effort seeking to avoid legally enforceable restrictions of controversial

technologies." What should be of interest is whether these policies address the concerns of the very group they seek to represent and whether universally acceptable norms is a possibility in the current geopolitical landscape?

In addition, while artificially intelligent systems can undoubtedly help social progress, the values attached to many of these systems are deeply intertwined with a form of Eurocentric philosophy and cosmology that does not necessarily translate to indigenous and/or local contexts and traditions. The organisation Algorithm Watch has compiled an online database of ethics guidelines. As of now, the inventory comprises 173 guidelines. A small number of guidelines comprise oversight / enforcement mechanisms with the vast majority of these guidelines emerging from Europe and the United States (Algorithim Watch, 2021). In terms of geographic distribution, the data show high representation of the more economically developed countries (Jobin et al, 2019). Another case in view is the formation of the Advanced Technology External Advisory Council (ATEAC) by Google with the mandate to "develop responsible AI" (Walker, 2019).. However, concerns raised by different actors about some of the members led to the dissolution of the ATEAC (Statt, 2019).

Having regard for the different principles proposed across the board, we find some commonalities in the 5P framework which is a consolidation of the six most up to date and prominent proposals for ethical AI. They are the Asilomar AI Principles (Future of Life Institute, 2017), the Montreal Declaration for Responsible AI (University of Montreal, 2017), the General Principles offered in the second version of Ethically Aligned Design: *A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems,* (IEEE General Principles, 2017), the Ethical Principles offered in the statement *Artificial Intelligence, Robotics and 'Autonomous' Systems*, published by the European Commission's European Group on Ethics in Science and New Technologies (EGE Principles for Ethical AI, 2018), the five overarching principles for an AI code in the UK House of Lords Artificial

Intelligence Committee's report: *AI in the UK: Ready, willing and able?* (House of Lords, 2018), and the *Tenets of the Partnership on AI* (2018), (a multi stakeholder organisation comprising of researchers, civil society organisations, P3 companies building and utilising AI technology, etc.). These principles according to Floridi et al. (2018) yield 45 principles when put together. The 5P thus is an important step aimed at harmonising the different principles that should guide ethical AI. This view is helped by organisations such as UNESCO which will also work with AI designers and professional associations to promote relevant guidelines and ethical codes, as well as to ensure an approach of ethics and human rights by design for AI without stifling innovation. (UN Activities on AI, 2018)

That said, as intercultural information ethics scholar Pak-Hang Wong warns, there is a danger of dominance of "Western" ethics in AI design, particularly the appropriation of ethics by liberal democratic values to the exclusion of other value systems, a claim supported by findings of the present study. While these same liberal democratic values were created to accommodate such differences in populations, there remains a strong bias towards an established Western canon in the practice of developing norms and codes for AI in education–especially important when thinking about education as a space where ideology is passed down (IEEE, 2018). With this in mind, when AI systems designed external to the context within which they are to be introduced, whose knowledge becomes privileged when said systems are fully integrated into education through policy and practice? A key challenge is ambiguous in the framing of norms, coupled with a certain hypocrisy practice, especially among influential countries (Ray, 2019). This level of hypocrisy weakens the core principles underlining AI norms—it is not the issuance of Westernised statements and one-sided declarations that will ensure that AI norms and principles are implemented. An example is India's *National Strategy for Artificial Intelligence* which in one breath seeks to encourage responsible AI,

yet lacks robust data laws that will enhance the development and use of AI. Such vague laws only remain an opportunity for abuse (Marda & Kodali, 2019).

The different union, country and corporate specific AI strategies are aimed at winning the trust of consumers of the AI technology while making effort to have a first mover advantage. In the past few years, a number of corporations including Microsoft and Google have published their own responsible and ethical AI frameworks and initiatives. Similar moves have been made by the Organisation for Economic Cooperation and Development (OECD) and a range of governments. For example, the Federal German Government's AI strategy aims at positioning Germany as the AI hub for the EU with a plan to strategically partner organisations across the world. With most tech giants located in the Global West, this position like many other countries only entrenches the Eurocentric and Westernised design, development, and deployment of AI:

> We want to make Germany and Europe a leading centre for AI and thus help safeguard Germany's competitiveness in the future. (German Federal Government, 2018)

As a result of the frequent data breaches and scandals by both government and tech giants, user trust has deteriorated in recent years. Opinion surveys conducted in the US in 2017 suggest a low trust ratio for Facebook, Google and twitter users (10%; 13%, and 8% respectively). Similar results are reported for India where a chasm report suggest more than half of Indians (57%) did not trust in social media platforms to safeguard their data (Sinha, 2018). Development and use of Responsible AI and user trust cannot be separated from each other. Regulations that seek to harmonise the deployment of AI must work at repairing the already damaged public perception while ensuring greater oversight over tech giants in a culturally acceptable form.

Concluding thoughts

Rigorous and sustained evidence that supports practical outcomes of AI implementation in education is limited. Some indicators are examined such as development of strategies for evaluating the impacts of new technologies on numeracy, literacy and science, but there is no sustained research that investigates the overall impact that AI holds for education let alone for making contributions to achieving quality education for all. In addition, despite the hype, there is sparse evidence to support the benefits of P3s in addressing inequalities with regards to provision and access to public services, one of which is education as a public good.

Many countries in the West have set out strategies to take full advantage of the artificial intelligence climate. The underlining motivation in most of these strategies is for countries to gain or expand the geopolitical dominance through AI. One of the ways they have advanced this course has been to build alliance(s) with countries that allow AI-tech companies from the West to advance their desire, including putting packages to attract highly skilled human resources to the West. These activities ultimately disadvantage poor countries and further entrench P3 relationships in developed countries.

Information about AI tools and policies which can impact AI and education actors remain scattered and uncoordinated making it difficult to implement trustworthy AI. At the heart of these AI tools and principles are how Western ideas are being superimposed on countries in the Global South. It remains a daunting task to find a simple document that seeks to bring the ideas of countries in the Global South into the Global AI discourse. How fair can the West be when opportunity is not created for inputs from countries in the South in deciding on what principles should guide responsible development and use of AI technologies? The bigger question remains thus: whose responsibility is it to bring the different actors together and to harmonise the principles into an acceptable and responsible AI?

Acknowledgements

Conflicts of Interest

There are no conflicts of interest associated with this work.

References

AI Now Institute (2020). *Submission to the European Commission on White Paper on AI - A European Approach*. Retrieved July 10, 2021 from https://ainowinstitute.org/ai-now-comments-to-eu-whitepaper-on-ai.pdf

Algorithm Watch. (2021). *AI Ethics Guidelines Global Inventory.* Retrieved August 4, 2021 from https://inventory.algorithmwatch.org/

Asilomar AI Principles. (2017). *Principles developed in conjunction with the 2017 Asilomar conference [Benevolent AI 2017]*. Retrieved August 13, 2021 from https://futureoflife.org/ai-principles.

Baker, T., Smith, L. and Anissa, N. (2019). Educ-AI-tion Rebooted? Exploring the future of artificial intelligence in schools and colleges. *Nesta*. Retrieved February 12, 2020 from

Biesta, G. J. J. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability, 21*(1), 33-46.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3(2)*, 77–101.

Couldry, N. & Mejias, U. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press.

Draxler, A. (2015). Public private partnerships in education. In McGrath, S., & Gu, Q. (Eds.). *Routledge Handbook of International Education and Development*. London, UK: Routledge.

EU High-Level Expert Group on AI. (2019). *Ethics Guidelines for Trustworthy AI*. Retrieved Dec 21, 2019 from https://www.ccdcoe.org/uploads/2019/06/EC-190408-AI-HLEG-Guidelines.pdf

European Commission (2021). *Proposal for a Regulation AI, including harmonising rules for AI*. Retrieved August 5, 2021 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

European Data Protection Board & European Data Protection Supervisor. (2021). *EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Retrieved June 28, 2021 from https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en

European Group on Ethics in Science and New Technologies. (2018). *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*. Retrieved August, 13, 2021 from

https://ec.europ a.eu/info/news/ethics-artificial-intelligence-statement-ege-relea sed-2018-apr-24_en.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St Martin's Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Gideon J. and Unterhalter, E. (2020). *Critical Reflections on Public Private Partnerships.* London, UK: Routledge.

German Federal Government. (2018). *Artificial Intelligence Strategy*. Retrieved August 6, 2021 from https://knowledge4policy.ec.europa.eu/ai-watch/germany-ai-strategy-report_en

Heidegger, M. (1982). *The Question Concerning Technology and Other Essays*. Trans William Levy. New York, NY: Harper Collins.

Holmes, W., Hui, Z., Miao, F., & Ronghuai, H. (2021). *AI and education: A guidance for policymakers*. UNESCO Publishing. Retrieved April 12, 2021 from

IEEE. (2018). Ethically Aligned Design. Retrieved June 20, 2019 from

https://standards.ieee.org/content/dam/ieee-

standards/standards/web/documents/other/ead_v2.pdf

ITU. (2018). *United Nations Activities on Artificial Intelligence*. Retrieved September 30,

2019 from https://www.itu.int/pub/S-GEN-UNACT-2018-1

Jobin, A, Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines.

Nature: Machine Intelligence, 1, 389-99.

Lords, H. O. (2018). AI in the UK: ready, willing and able? Retrieved August 13, 2021 from

https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Mann, S. and Hilbert, M. (2018). *AI4D: Artificial Intelligence for Development.* Retrieved

July 17, 2021 from https://ssrn.com/abstract=3197383

Marda, V. & Kodali S. (2019). The Privacy Cost of Digi Yatra's Seamless Travel

Promise, *Huffpost*. *Retrieved August 5, 2021 from*

https://www.huffpost.com/archive/in/entry/digi-yatra-face-recognition-hyderabad-

airport_in_5d2d9725e4b085eda5a15c28

McStay (2020). Emotional AI and EdTech: serving the public good? *Learning, Media and*

Technology, 45(3), 270-283.

Microsoft. (2021). Understanding the new learning landscape: Accelerating Learning Analytics and AI in Education. Retrieved April 17, 2021 from https://edudownloads.azureedge.net/msdownloads/Microsoft-Accelerating-Learning-Analytics-and-AI-in-Education.pdf

Montreal Declaration for a Responsible Development of Artificial Intelligence. (2017). *Announced at the conclusion of the Forum on the Socially Responsible Development of AI*. Retrieved August 13, 2021 from https://www.montrealdeclaration-responsibl eai.com/the-declaration.

Morning Consult. (2017). Morning Consult National Tracking Poll #170401, *Morning Consult/Politico. Retrieved August 5, 2021 from* https://morningconsult.com/wp-content/uploads/2017/04/170401_crosstabs_Politico_v3_AG.pdf

Ochigame, R. (2019). The invention of "ethical AI": How Big Tech Manipulates Academia to Avoid Regulation. *The Intercept*. Retrieved July 29, 2021 from https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/

Nemorin, S. (2021). *Fair-AIEd*. Retrieved June 20, 2021 from www.fair-ai.com

Nemorin, S. (2018). *Biosurveillance in New Media Marketing: World, Discourse, Representation*. London, UK: Palgrave.

Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: NYU Press.

NSC. (2021). *National Security Commission on Artificial Intelligence*. Retrieved June 29,

2021 from https://www.nscai.gov/

OECD. (2021). AI Policy Observatory. Retrieved June 24, 2021 from

https://oecd.ai/dashboards/ai-principles/P11

OECD. (2021). *Digital Education Outlook 2021: Pushing the Frontiers with Artificial

Intelligence, Blockchain and Robots*. Retrieved July 7, 2021 from https://read.oecd-

ilibrary.org/education/oecd-digital-education-outlook-2021_589b283f-en#page1

O'Keefe, C., Cihon, P., Garfinkel, N., Flynn, C., Leung, J., and Dafoe, A. (2020). *The

windfall clause: Distributing the benefits of AI for the common good.* Retrieved September

12, 2020 from

https://www.fhi.ox.ac.uk/windfallclause/#:~:text=The%20Windfall%20Clause%20is%20an,t

ransformative%20breakthroughs%20in%20AI%20capabilities.

O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and

Threatens Democracy*. New York, NY: Penguin.

Partnership on AI. (2018). *Tenets*. Retrieved August 13, 2018 from

https://www.partnershiponai.org/tenets/.

Ray T. (2019). Formulating AI Norms: Intelligent Systems and Human Values. Observer

Research Foundation. *ORF Issue Brief No. 313.* Retrieved August 6, 2021 from

https://www.orfonline.org/research/formulating-ai-norms-intelligent-systems-and-human-values-55290/

Resseguier, A. and Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society* 1-5.

Selwyn, N. (2015). Data entry: towards the critical study of digital data and education. *Learning, Media and Technology*, 40(1), 64-82.

Sinha, S. (2018). Annual Consumer Survey on Data Privacy in India. *Analytics India. Retrieved August 5, 2021 from* https://analyticsindiamag.com/annual-consumer-survey-on-data-privacy-in-india-2018/

Srivastava, Prachi & Oh, Su-Ann. (2010). Private foundations, philanthropy, and partnership in education and development: mapping the terrain. *International Journal of Educational Development*, 30*(5)*, 460-47.

Statt N. (2019). Google dissolves AI ethics board just one week after forming it. *The Verge.* Retrieved August 6, 2021 from https://www.theverge.com/2019/4/4/18296113/google-ai-ethics-board-endscontroversy-kay-coles-james-heritage-foundation

The IEEE Initiative on Ethics of Autonomous and Intelligent Systems. (2017). *Ethically Aligned Design, v2*. Retrieved August 13, 2021 from https://ethicsinaction.ieee.org.

Tomašev, N., Cornebise, J., Hutter, F. et al. (2020). AI for social good: unlocking the

opportunity for positive impact. *Nature, 11*, 2468. Retrieved July 25 2021 from

https://www.nature.com/articles/s41467-020-15871-z

Tracy, S. J. (2010). Qualitative quality: Eight 'big-tent' criteria for excellent qualitative

research, *Qualitative Inquiry, 16(10)*, 837–851.

UNESCO. (2021). Artificial Intelligence in Education: Challenges and Opportunities for

Sustainable Development. Retrieved June 30, 2021 from

https://www.gcedclearinghouse.org/sites/default/files/resources/190175eng.pdf

UNESCO. (2019a). The challenges and opportunities of Artificial Intelligence in education.

The UNESCO Courier. Retrieved November 10, 2019 from

https://en.unesco.org/courier/2018-3/audrey-azoulay-making-most-artificial-intelligence

UNESCO. (2019b). Artificial Intelligence in Education: Challenges and Opportunities for

Sustainable Development. Retrieved December 14, 2020 from

https://en.unesco.org/news/challenges-and-opportunities-artificial-intelligence-education

UNESCO. (2015). Sustainable development begins with education: How education can

contribute to the proposed post-2015 goals. Retrieved March 23, 2019 from

https://sdgs.un.org/sites/default/files/publications/2275sdbeginswitheducation.pdf

Unterhalter, E. (2017). A Review of public private partnerships around girls' education in developing countries: flicking gender equality on and off. *Journal of International and Comparative Social Policy*, 33*(2),* 181-199.

Van Dijck, J (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance and Society, 12*(2), 197-208. doi.org/10.24908/ss.v12i2.4776

Van Leeuwen, T. (2008). *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford, UK: Oxford University Press.

Walker K. (2019). An external advisory council to help advance the responsible development of AI. Retrieved August 6, 2021 from https://blog.google/technology/ai/external-advisory-council-help-advance-responsible-development-ai/

World Bank. (2019). *What are private public partnerships?* Retrieved May 14, 2021 from https://ppp.worldbank.org/public-private-partnership/overview/what-are-public-private-partnerships

World Bank. (2018). Learning to Realize Education's Promise. *Global Education Monitoring Report (GEMR)*. Retrieved December 12, 2018 from https://www.worldbank.org/en/publication/wdr2018

World Economic Forum. (2020). Shaping a Multiconceptual World. Retrieved February 12,

2021 from

http://www3.weforum.org/docs/WEF_Shaping_a_Multiconceptual_World_2020.pdf