Title: Shaping our algorithms before they shape us

Michael Rowe
University of the Western Cape, Cape Town, South Africa
mrowe@uwc.ac.za
+27 21 959 2542

Abstract

A common refrain among teachers is that they cannot be replaced by intelligent machines because of the essentially human element that lies at the centre of teaching and learning and which is resilient to the challenges of brute force computation. While it is true that there are some aspects of the teacher-student relationship that may ultimately present insurmountable obstacles to the complete automation of teaching there are important gaps in practice where artificial intelligence (AI) will inevitably find room to move. Machine learning is the branch of AI research that uses algorithms to find statistical correlations between variables that may or may not be known to the researchers. It is a field of study in which we can get machines to learn something without knowing what it is that we want them to learn. The implications of this are profound and are leading to significant progress made in natural language processing, computer vision, navigation, and planning. But machine learning is not all-powerful and there are important technical limitations that will constrain the extent of its use and promotion in education, provided that teachers are aware of these limitations and are included in the process of shepherding the technology into practice. This has always been important but when a technology has the potential of AI we would do well to ensure that teachers are intentionally included in the design, development, implementation and evaluation of AI-based systems in education.

Keywords

artificial intelligence, machine learning, educational technology

Introduction

It has become commonplace to argue that the practice of teaching with care, ethics and justice cannot be replicated by machines and that the essence of a learner- or relationship-centred pedagogy relies on human values and the nature of the interactions between people. As is often the case we are happy to acknowledge that technological disruption is inevitable elsewhere but are usually able to find good explanations for why our own professions are safe (Susskind & Susskind, 2015). The claim that smart algorithms and humanoid robots are not capable of the emotional connection that drives meaningful, socially constructed learning may well be true but will also not matter. Even though individual human teachers are unlikely to be replaced, teaching as a profession remains vulnerable to automation in the face of increasingly capable machines. In the context of teaching and learning there are three assumptions that constrain our thinking and cloud our judgment when considering the impact of artificial intelligence (AI) on teaching and learning. The first assumption is that all teachers are able to work in a system where care, justice and human connection are present and prioritised. The second is that "teaching" is a single, monolithic practice instead of a collection of isolated and routine tasks. The third assumption is that these tasks – whether routine or not – nonetheless require a level of intelligence that are beyond the capabilities of computer algorithms. By interrogating these assumptions we can disrupt the common sense understanding of teaching and learning and expose the gaps in practice into which AI-based systems will inevitably move.

The first assumption is that only human teachers possess the emotional and personal connections necessary for the meaningful interactions that are central to socially constructed learning. And while this may be true (for now) the reality is that much of the education system remains untouched by concerns of care and justice. The infrastructure of education ensures that control and authority are vested in the teacher who is positioned, both physically and epistemologically, as the only legitimate source of knowledge in the classroom. Students are reminded that their words and personal experiences have no value in their own learning (Freire, 2005) and this lack of power dulls their enthusiasm and cultivates an obedience to a system that generates attitudes of conformity (hooks, 2004). In this paradigm learning is may be seen as a process of moving information from the notes of the teacher to the notes of the student via "communiques" that students receive, memorise and repeat (Freire, 2005). This banking model of education turns the student into a container to be filled and the more passively the students allow themselves to be filled, the "better" they are. Students and teachers are thus both reduced to “cheerful robots” through an instrumental

rationality in which matters of justice, values, ethics, and power are erased (Giroux, 2011). In addition, assessment emphasises the failings of students in a process that seems preoccupied with ranking them rather than creating formative contexts in which learning is prioritised (Morris & Stommel, 2015). The argument against the introduction of AI into educational contexts cannot be that algorithms are unable to offer an ethical and care-full pedagogy in the context of human relationships because – for reasons beyond their control – many teachers may find themselves unable to offer this kind of pedagogy. AI-based systems in education will not need to replicate the best possible version of a human teacher. For economic reasons, they will only need to be a little bit cheaper than the *average* teacher (Brynjolfsson & McAfee, 2014). The perception of cost-effectiveness alone will be enough to drive the implementation of AI in education . In addition, there is evidence that the mere appearance of intelligence is enough to mediate our cognitive and emotional response to machines (Susskind & Susskind, 2015). In other words, we may not need our teaching machines to care; it may be that the appearance of caring is enough for students to form learning relationships with them, further denting the argument that emotional connection is a requirement for learning.

The second assumption is that teaching as a profession is a single, discrete occupation rather than a collection of often unrelated tasks, many of which are physically or cognitively routine and thus vulnerable to algorithmic manipulation (Susskind & Susskind, 2015). The non-routine cognitive tasks are those that are necessary for the development of empathy and human connection and are implicated in meaningful learning. A teacher in the 21st century cannot simply provide students with facts about the world and then rank them according to their test scores because these are the tasks that intelligent agents will soon perform at negligible expense. The ability of an algorithm to scale from 100 users to 100 000 at a marginal cost of effectively zero is an aspect of education that anyone in control of a budget will find appealing. If teaching is a series of tasks, some of which are routine and therefore vulnerable to automation, then we should ask which are the ones that require insight, care, creativity and connection, because these are less vulnerable to algorithmic control. Teachers who work collaboratively with students to provide them with the tools to engage critically and creatively with the world are more than conduits for specialised knowledge. This relationship-centred process is not amenable to computational solutions and would, in some sense, protect these tasks from the risk presented by AI-based systems. Teachers who , through systemic constraints and overly rigid regulatory requirements, are subject to what Freire called the banking model of education, may be rendered obsolete as emerging technologies outperform them in the narrow tasks of information transfer and student ranking.

The final assumption is that AI will simply not be able to take over even the most basic tasks that teachers must perform because their current performance is so severely constrained by high costs and narrowly defined working parameters. However, human beings have a tendency to focus our attention on very short time horizons at the expense of seeing what it possible over longer periods. While it is true that the performance of machine learning algorithms currently leaves a lot to be desired, it would be a mistake to discount both the current performance and improvement in AI research over the next decade. For example, image classification is both mundane and incredible. Mundane because computer vision has only now reached a point where it is at parity with human performance in a small subset of categories. But incredible because five years ago even this parity was impossible (Frankish & Ramsay, 2017). This is a testament to advances in the data available for algorith training, more powerful computation at much lower cost, and improved algorithm design (Brynjolfsson & McAfee, 2014). Even if all technical progress on machine learning were to stop tomorrow (which seems unlikely given the amount of interest and funding available), the steady increase in computational power and access to data means that the functional intelligence of AI systems will keep getting better. At some point this will bottom out, and in fact, there are already signs that the results produced from the current generation of machine learning systems will soon plateau (Jordan, 2018). But even when that happens we will still see gains in performance speed and associated decreases in cost that will continue driving the introduction of AI-based systems into education.

I believe that these three assumptions underpin many of the arguments against the implementation of AI in education and that when challenged, they open up the possibility that the introduction of AI into the educational context is not only possible but probably inevitable. If the primary role of a teacher is to provide access to specialised knowledge and rank student performance according to a set of standardised competencies, we may soon see the end of the profession. However, if the role of a teacher is to *become with* the student in a pedagogy of liberation in which both teachers and students engage in acts of cognition with the intent of transforming the world (Freire, 2005), then we may see AI used to augment a creative, ethical, care-full and socially just pedagogy. In this latter case it seems unlikely that AI will replace teachers but rather that teachers who use AI appropriately may replace those who do not. The purpose of this chapter is explore the implications of AI in education and to make suggestions for how teachers can be included in the development of AI-based systems in order to ensure that the human values of care and justice in embedded in intelligent teaching machines.

AI in the context of education

The field of artificial intelligence research began in the 1950s with the design of algorithms that could solve well-defined problems in structured, clearly described environments. This meant manually coding all possible routes through a solution space which, even in tightly controlled artificial environments, would cause the algorithms to break down (Frankish & Ramsay, 2017). These early AI systems were not resilient to small changes in even very simple scenarios and could not adapt to the much more complex interactions found in the real world. Because algorithm designers could not predict, and therefore solve, all of the problems arising in uncertain environments, it was soon understood that these brittle systems could not produce results that had much practical or commercial value (ibid.). Developers realised that they needed computational models that could more closely approximate reality and would need to "learn" and adapt in real time by using feedback from the environment. Machine learning (ML) is the branch of AI research that deals with this challenge and is responsible for much of the recent progress being made in AI. Machine learning works by using statistical techniques to identify relationships between variables in very large data sets with varying levels of confidence (Pearl & Mckenzie, 2018). Basically, machine learning uses statistics to solve classification problems and predict missing information (Agrawal, Gans & Goldfarb, 2018).

In this context the word "learn" is problematic because we may associate it with other aspects of human experience, including emotions and consciousness. This is confusing because of our tendency to anthropomorthise non-human objects, leading to an assumption that "learning" algorithms are conscious, moral and emotional and that they should also understand the outputs of their computation (Frankish & Ramsay, 2017). When algorithms fail to display these additional characteristics, we may see this as an argument for why they cannot replace teachers. But no serious AI researcher claims that a machine translation algorithm understands language or that a game-playing algorithm understands that it is playing a game (Searl, 2011). An algorithm does not know what a game is. Our ability to understand the world and interact with others is a non-declarative type of knowledge that cannot be captured by any rule-based system (Frankish & Ramsay, 2017). The statistical computation in machine learning is fundamentally different to the phenomenological process of human learning and the conflation of the two leads to misunderstanding and unrealistic expectations of ML algorithms. Indeed, ML systems need not have an understanding of the world, nor of emotion, consciousness or human-relationships in order to outperform humans in a wide variety of the tasks that make up the job description of a teacher. For example, timetable and curriculum planning, lecture transcription, and content preparation and review are all tasks that take

up significant portions of teachers time, are all vulnerable to automation, and do not require an emotional connection to anyone. Algorithms are not conscious and they do not care about us but this says nothing about their ability to outperform us across an increasing variety of tasks.

While high intelligence scores are correlated with an increased probability of success across many domains (Ritchie, 2017), successful human interaction is often dependent on empathy, imagination, tolerance of ambiguity, and the use of metaphor, all of which are resistant to computational solutions (Frankish & Ramsay, 2017). In addition, relevance, or the ability to distinguish the essential from the inessential, and to effortlessly draw on our experience and knowledge in accordance with the demands of the task, is also regarded as a major stumbling block for AI (ibid.). Another important aspect of human interaction is an innate understanding of causality; we can predict what is likely to happen next based on our understanding of what has happened in the past. Indeed, prediction is the aspect of AI that will become increasingly commoditised (i.e. it will soon be very cheap and ubiquitous) driving its integration across and deeply within many sectors of society (Agrawal, Gans & Goldfarb, 2018). However, while much has been made of the predictive utility of ML it cannot say anything about the direction of causality (Pearl & Mckenzie, 2018). An algorithm is able to identify patterns between variables and identify when they are correlated but it cannot say that one variable was a necessary precursor to another.For example, an algorithm may determine that low grades and parental income are positively correlated i.e. as parental income decreases there is a better than random chance that grades will also decrease. However, it is impossible for current AI systems to conclude that the cause of the low grades is the low income.

There are therefore a range of human capabilities that are necessary for meaningful interactions that are central to learning and teaching, and which present significant technical challenges that make them resistant to computational solutions. Simply increasing the functional intelligence of an AI-based system is not enough for them to work around issues of human values and care in the context of learning and teaching. However, we should also be clear that practical AI research is not about trying to accurately model all the components of human cognition that are implicated in effective communication. We can already generate useful inferences and accurate predictions about the world using nothing more than brute force computation and, relative to the human brain, unsophisticated algorithms. The successful implementation of AI-based systems in education will not require consciousness, morality, or an ability to generalise reasoning across contexts. Intelligent reasoning about interactions in the world do not require a complete understanding of it, or even an accurate representational model of human cognition (Frankish & Ramsey, 2017). They will only need to

complete very narrow, routine tasks that can be well-defined, at a lower cost than human beings. And they can already do this very well.

Teachers need not worry about being replaced by humanoid robots but rather that software upgrades to existing systems will gradually take over tasks that were previously theirs to perform. As the success of real world organisations and industries is increasingly premised on their ability to abstract workflows and production into software, we should begin planning for a similar disruption of education (Andreesson, 2011). This disruption will happen when algorithms become the dominant decision-makers in the system, first at the level of isolated apps (e.g. automated essay graders), then integrated across platforms (e.g. student risk assessment via learning management systems), institutions (e.g. admisssions and assessment) and finally, at the level of the industry as a whole (e.g. managing students from entry into the system all the way through to employment). When AI is fully integrated into the education system then every organisation in that system will have to become a software organisation. This is what the disruption of education will look like; less a robot army and more a gradual loss of decision-making autonomy.

Shaping our algorithms

We are fond of saying that technology must follow pedagogy and in the case of AI in education it is clear that this is pecisely what has happened. The technology *has* followed the pedagogy. Education has misrepresented itself as objective, quantifiable and apolitical (Morris & Stommel, 2015) and as a result, educational technology companies have positioned AI in education as objective, quantifiable, and politically neutral. They assert that technology is neutral and that they simply need to get the right data and analysis in order to find the ground truth that will allow us to "fix education" (Hendrick, 2018). But every feature of a technology is the result of a series of human decisions that optimise the technology towards a fitness function and as long as people are involved the fitness function cannot be objective or neutral. Before we can decide if a technology is fit for purpose we first have to know the purpose of the technology, and with that, the values that are encoded into it. Likewise, algorithms do not have purposes that are predicted by inviolable laws of nature. They are optimised towards a fitness function that is the result of human choices (Frankish & Ramsay, 2017). The dominant discourse around AI in education is about saving time, reducing costs and increasing efficiency, reflecting a continuation of the neoliberal policies driving austerity and cuts to services that have emerged over the past two decades. But there is nothing inevitable about these values and it is reasonable to consider a different set of values against which the fitness function of algorithms can be optimised. For example, instead of developing an algorithm that

maximises cost-effectiveness, profit or attention, there is nothing preventing us from choosing to maximise human well-being instead. The sense of inevitability associated with technological progress is disempowering because it can make us believe that we cannot change its direction or destination. But the reality is that human decisions informed by human values is what drives technological progress and the same must be true of the decisions that inform the implementation of AI in education.

As the potential of AI to affect society grows, the role of teachers needs to change in order to help students prepare for a critical engagement within that society. Any discussion around AI in education should therefore emphasise the need to understand and shape this increasingly ubiquitous technology, foregrounding the necessity of input from all stakeholders rather than only from those who are technologically literate (Jordan, 2018). If the narrative around AI in education is being driven by venture capital firms and wealthy entrepreneurs (Williamson, 2018) it may be because teachers have been distanced from decision-making and increasingly managed by a regime of performance targets that incites them to perform in narrowly measurable ways. Much like medical doctors are, in some sense, the final arbiters of what technology is allowed to operate in the clinical context, so should teachers, informed by values of care and justice, have make the final decision about what technology should be allowed in schools. If education technology was shaped by regulation and policies developed by students and teachers then edtech startups would not be able to "move fast and break things". There is a narrative among these startups that education is broken, which gives them leave to bring their signifiant resources to bear on the “problem” (Hendrick, 2018). It may therefore be up to teachers to challenge the assumptions of edtech companies by guiding the design, implementation and evaluation of AI-based systems within a contextual framework that includes them. When teachers are absent from the conversation around the use of technology in education, techno-evangelists will position the technology as a form of emancipation, thereby freeing teachers from an outdated model that is not fit for purpose (Hendrick, 2018) In order to shape the space in which AI operates teachers must ensure that the values of a socially just pedagogy are integrated into the development of ML algorithms.

Bostrum and Yudkowsky (2015) ask what human values must be integrated with the computational intelligence of smart machines and suggest that they might include responsibility, transparency, auditability, incorruptibility, and predictability (2015). They suggest that since these are some of the criteria that we apply to humans performing social functions we should therefore look for the same criteria in any algorithm intended to replace human judgement in those functions. There is no reason that the fitness function of ML algorithms could not be optimised towards developing care-

full and socially just pedagogies that privilege student learning and well-being. If we simply accept that AI-based systems incorporate our values, we may become passive or unable to respond when we eventually find that they do not. But if we being by asking whose values are represented in code and what those values are, we may find that we disagree with them. Once we understand that there is nothing inevitable about the path that a technology takes to mature, nor what the final product should be, it becomes easier to see that we are not powerless to influence the design of these systems. The infrastructure and communication channels that are necessary for democratic participation in the design and implementation of AI in education is currently missing, making it difficult for teachers to be included. However, unless teachers are intentionally involved in establishing the guiding values of AI in education, we run the risk that our professional decision making will not be informed by machine intelligence but rather that we are subject to it.

1. Data collection

Successful AI relies on finding patterns in large datasets. However, there are many reasons for why training algorithms with educational data may lead to inconsistent, inaccurate, unreliable or invalid conclusions. Education data is often poorly structured, inauthentic, lacking demonstrable validity and reliability, and consists almost entirely of grades (Lynch, 2018). These proxies for learning are used to train the algorithms embedded in AI systems, not because they are pedagogically meaningful but because they are easy to collect (ibid.). There are other reasons for why the outputs of algorithmic decision-making in education may be wrong. The knowledge base might be biased; the inferences drawn may be incorrect because of errors in the algorithm; the algorithm's reasoning might not be able to adapt to unexpected contingencies; and the decision criteria and outputs may not be universally acceptable (Mittelstadt et al., 2016). In addition to the social biases encoded in training data, it should not be controversial to say that human decision-making is also influenced by subconscious biases and that these biases are so deep that we are blind to them (Kahneman, 2011). Cleaning and transforming the data for algorithm training adds further uncertainty to the process as there are many subjective decisions that will need to be made, each of which adds further opportunities for introducing errors that will have an impact on the algorithm outputs. If these sources of bias remain unchecked algorithms may consolidate and deepen the already systemic inequalities in education and society all while making them harder to notice and challenge (Hart, 2017). It is therefore incumbent on teachers to ensure that ML training data is diverse both in terms of the student voices present in the data, but also diverse in the range of proxies for learning that are gathered. Diversity in student populations means that we can be more confident that AI-based predictions are generalisable across different populations and contexts, regardless of what data it

was trained on. Having diverse teams, including teachers, students and education researchers, will increase the likelihood that our biases are recognised and addressed, rather than becoming encoded within AI-based systems. Of course, the practical challenges of making these changes from within a system that has already disempowered teachers and students is deeply concerning. Teachers not only lack the systems and support for gathering diverse examples of student learning, they lack the time to even think about it. However, without a broader understanding of the data we use to make judgements with respect to student learning, we risk constructing machine intelligence that reflects a narrow and relatively poverty-stricken vision of education.

2. Teaching practice

The computational intelligence of technology is not a substitute for relevant background knowledge in the practice of teaching, and the appropriate use of AI in the classroom will require that computer scientists, teachers, and students work closely together in order to ensure that these systems are fit for purpose. The important question is, whose purpose? In order to be active and informed citizens, students will need a sound understanding of AI, as well as a critical approach to assessing the implications of data collection at very large scales. Thus, teachers will need to use and evaluate AI systems in the classroom so that they are able to contribute to the conversation and play a role in setting the agenda for AI in education. In this way teachers can shape the discourse around AI in education so that it is framed within an approach that prioritises care and human relationships in learning. To this end, teachers will need to engage with AI-based systems in similar ways to how they work with colleagues. They will need to use critical judgement to make decisions about the context in which algorithms produce outputs. Rather than seeing algorithmic decisions as fundamentally "right" or "wrong" teachers will need to understand that algorithms provide probabilistic outputs based on imperfect information and are therefore inherently prone to making mistakes. Just as we will need to decide when those outputs can be trusted, we will need to make choices about when they should be ignored. Unfortunately, there is evidence that we struggle to objectively judge the decisions made by algorithms and that we will often simply follow the instructions we receive (Lyell & Coiera, 2017). For example, teachers may champion the use of recommendation engines to identify personalised content for students, thinking that a more focused collection of information is helpful for learning. But these systems make inferences that result in increasingly deterministic recommendations, which tend to reinforce existing beliefs and practices (Polanski, 2016). If we want students to be exposed to different ideas as part of their learning then recommendation systems that narrow the focus of information may effectively close down students' options for diverse perspectives. Teachers who unquestioningly assume the correctness of the

algorithmic output may inadvertently reinforce stereotypes and systemic biases. If AI-based systems are left to operate solely in the rational domain of cognition and teachers ignore the emotion-laden interactions that drive meaningful learning they may, with the best intentions, lock students into a category of demographically classified content from where it is difficult to see anything else. The use of AI in education has implicit ethical, social, political and pedagogical choices and it is essential that both students and teachers are included in order to develop guidelines and theoretical frameworks that can help minimise the risk of unintended consequences.

3. Research

Audrey Watters asked, "what will happen when robots grade students' essays? What happens when testing is standardized, automated?...What sorts of signals will we be sending students?" (Watters, 2015). These are empirical questions that can be tested but when we decide the lines of inquiry we want to explore we should ask questions without having already decided the outcome. For example, we might also reasonably ask, What will happen if algorithmic assessment turns out to be more accurate, more consistent and more equitable than human assessment? What if students preferred it? What if algorithmic feedback and instruction improve students' intrinsic motivation? Do students prefer a simple algorithm that is available 24/7 or a friendly teacher who is available on Tuesdays between 14:00 and 16:00? Right now it seems as if the questions are being asked by edtech companies who frame educational problems as problems of efficiency rather than problems of care, relationship and power (Hendrick, 2018). The technology underpinning much of the AI-based progress in education is far from perfect and perhaps more importantly, we do not yet have an agreed upon philosophical foundation upon which to build (Jordan, 2018). Those who are currently developing educational AI may end up making inappropriate recommendations that actually hinder learning, and then attempt to generalise their findings across different contexts. But without a theoretical foundation we may not have good reasons to reject the conclusions that are provided (ibid.). Much like the theory of Connectivism was developed in response to the emergence of networked learning environments (Siemens, 2005), it is likely that we will need a theory of AI in education in response to the emerging challenge of trying to understand the relationship between smart machines and human beings in the context of learning and teaching. Teachers will need to be involved in studies that ask a wide variety of questions around the use of AI in education, none of which begin with assumptions about what works, or what is better. Research that aims to answer empirical questions about student learning should guide decision-making about what projects have merit, and how the outcomes of research should be used to inform education policy.

4. Policy

We lack a language and set of social, professional, ethical, and legal norms that will enable us to appropriately implement AI in education, as well as paradigms and frameworks in which to work. In June 2018 the American Medical Association (2018) released a set of guidelines for the use of augmented intelligence systems in the context of clinical care, highlighting the potential impact of AI in healthcare. The education community needs a similar set of guidelines to ensure that AI in education serves the needs of students, teachers and administrators and is fit for the purpose of enhancing student learning. If edtech startups are moving into the education space, we will need to configure that space to ensure that AI-based systems conform to regulatory frameworks that prioritise pedagogies of care and justice. These frameworks might suggest that teachers help to set priorities for the use of AI in education, identify opportunities to integrate their own perspectives into the development of AI, promote the development of thoughtful, pedagogically sound AI, , and that they are aware of the legal, social, political and ethical implications of these systems in education. All of this would help to develop the regulatory frameworks, discourse and set of norms within which AI-based systems would be required to operate. It might move us towards mandating that AI-based systems must show that they will "first do no harm" and demonstrate evidence that holistic student well-being and learning will not be compromised by the introduction of algorithms in the classroom. Education policymakers and teachers will need to have difficult conversations related to the current landscape of education, including the quality and range of data used to assess student learning, teaching and learning practices, challenges around the replication and generalisation of education research, and a host of other concerns that will emerge with the inevitable movement of AI into the classroom.

Conclusion

The introduction of AI into various aspects of teaching and learning is inevitable and the more we rely on algorithms to make decisions, the more they will shape what is seen, read, discussed, and learned. These systems will continue improving in several narrow domains of practice, eventually outperforming human beings in a wide variety of routine tasks. While these technologies are in their infancy and therefore unlikely to be implemented at scale in the near future, the arguments presented in the opening of the chapter should highlight the problem in ignoring or denigrating the use of AI in education. Machine learning has important technical limitations that will tend to prioritise grades and other easily measurable variables, rather than values like care and justice. It is also clear that teachers and students are limited in what it is that they are practically able to do.

Nonetheless, it is important for all stakeholders to develop strategies for understanding and working with AI-based systems in order to avoid the algorithmic determinism that will otherwise influence our decision-making. We must participate in the conversation around AI development so that the discourse is not framed entirely by software developers and technology entrepreneurs. We must ensure that the voices of students are included, not only in the algorithm training data but in the design, implementation and evaluation of AI-based systems in the classroom. We must refocus our attention on those aspects of teaching and learning that incorporate human values like care, emotional connection, and relationship-building. We should design and conduct education research using AI-based systems with the intention of developing and refining a theoretical framework for AI in education. The introduction of AI into education is not a technology problem; it is a human and social problem. To frame it as a technology problem with a technological solution is to hand the responsibility for stewarding these systems to those who may not have the same pedagogical values as teachers who value student learning. This closes down the opportunities that emerge when a diverse group of people from different disciplines and backgrounds work together on projects with their own unique perspectives. When we see AI as a human problem rather than a technical one it becomes clear that it is incumbent on all of us – teachers, students and software engineers – to develop an equitable and humane pedagogy of AI in education. To shape our algorithms before they shape us.

References

Agrawal, A., Gans, J. & Goldfarb, A. (2018). Prediction machines: The Simple Economics of Artificial Intelligence. Harvard Business Review Press, Boston.

American Medical Association (2018). AMA passes first policy recommendations on augmented intelligence. Available from https://www.ama-assn.org/ama-passes-first-policy-recommendations-augmented-intelligence.

Andreesson, M. (2011). Why software is eating the world. Available at https://a16z.com/2016/08/20/why-software-is-eating-the-world/.

Bostrom, N. & Yudkowsky, E. (2017). The ethics of artificial intelligence. In Frankish, K. & Ramsay, W.M. (2017). The Cambridge Handbook of Artificial Intelligence, 316–334.

Brynjolfsson, E. & McAfee, A. (2014). The second machine-age: Work, progress, and prosperity in a time of brilliant technologies. W.W. Norton & Company, New York.

Frankish, K. & Ramsay, W.M. (2017). The Cambridge Handbook of Artificial Intelligence. Cambridge: Cambridge University Press.

Freire, P. (2005). Pedagogy of the Oppressed. 30th Anniversary Edition. Continuum. The Continuum International Publishing Group Ltd., The Tower Building, 11 York Road, London SE1 7NX.

Giroux, H. (2011). On Critical Pedagogy. Continuum. The Continuum International Publishing Group Ltd., The Tower Building, 11 York Road, London SE1 7NX.

Hart, R.D. (2017). If you're not a white male, artificial intelligence's use in healthcare could be dangerous. Quartz. Available at https://qz.com/1023448/if-youre-not-a-white-male-artificial-intelligences-use-in-healthcare-could-be-dangerous/.

Hendrick, C. (2018). Challenging the 'education is broken' and Silicon Valley narratives. ResearchED. Available at https://researched.org.uk/challenging-the-education-is-broken-and-silicon-valley-narratives/.

hooks, bell (1994). Teaching to Transgress. Education as the Practice of Freedom. Routledge, Taylor & Francis Group, 711 Third Avenue, New York, NY 10017.

Jordan, M. (2018). Artificial Intelligence — The Revolution Hasn't Happened Yet. Medium. Available at https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7.

Kahneman, D. (2011). Thinking fast, and slow. Farrar, Straus and Giroux.

Keynes, J.M. (1936). The General Theory of Employment, Interest and Money. Palgrave McMillan.

Lynch, (2017). How AI will destroy education. Medium. Available at https://buzzrobot.com/how-ai-will-destroy-education-20053b7b88a6.

Mittelstadt, B.D., Allo., P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate. Big Data & Society, December: 1–21

Obermeyer, Z. & Thomas, H.L. (2017). Lost in Thought - The Limits of the Human Mind and the Future of Medicine. New England Journal of Medicine, 1209–11.

Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: A systematic review. *Journal of the American Medical Informatics Association*, *24*(2), 423–431.

Lynch, J. (2018). How AI will destroy education. Medium. Available at https://buzzrobot.com/how-ai-will-destroy-education-20053b7b88a6.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data and Society (December):1–21.

Pearl, J. & Mckenzie, D. (2018). The book of Why: The new science of cause and effect. New York: Basic Books.

Polanski, V. (2016). Would you let an algorithm choose the next US president? World Economic Forum. Available at https://www.weforum.org/agenda/2016/11/would-you-let-an-algorithm-choose-the-next-us-president/.

Ritchie, S. (2016). Intelligence: All that matters. Teach yourself. United Kingdom: Hachette

Searl, J. (2011). Watson Doesn't Know It Won on 'Jeopardy!' Wall Street Journal. Retrieved on 27 September 2018 from https://www.wsj.com/articles/SB10001424052748703407304576154313126987674.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning, 2(1), 3-10.

Susskind, R. & Susskind, D. (2015). The future of the professions: How technology will transform the work of human experts. Oxford Univesity Press, Oxford.

Morris, S.M. & Stommel, J. (2017). Open Education as Resistance: MOOCs as Critical Digital Pedagogy. In, Losh, E. (Ed.) MOOCs and their Afterlives: Experiments in Scale and Access in Higher Education. University of Chicago Press, London.

Susskind, R. & Susskind, D. (2015). The Future of the Professions: How Technology will Transform the Work of Human Experts. Oxford; Oxford University Press.

Watters, A. (2015). Teaching Machines and Turing Machines: The History of the Future of Labor and Learning. Available at http://hackeducation.com/2015/08/10/digpedlab.

Williamson, B. (2018). The tech elite is making a power-grab for public education. Code Acts in Education. Available at https://codeactsineducation.wordpress.com/2018/09/14/new-tech-power-elite-education/.