John W. Jones/johnwesleyjoensjr@gmail.com

# AI in Education: Some Thoughts about Ethics, Equity, and Social Impact

Following the public release of the chatbot, ChatGPT, discussions about the impacts and dangers of highly sophisticated artificial intelligence (AI) applications exploded into the popular awareness. Both social and traditional mainstream media were inundated with commentary about the new technology. However, among specialist and professional communities, the conversations about the implications of this technology are not a new phenomenon. A major component of the recent explosion of interest in AI concerned the impacts of the technology on education and learning. Educators in particular are concerned about plagiarism and cheating (Cotton et al. 2023). In a wider context, educators and curriculum planners are attempting to understand how AI will necessitate changes to curricula and whether and how AI utilities can be integrated into instruction. The technology has the potential to powerfully change education, but just because the potential exists, this does not imply that the potential will be realized. Without proper planning, students might resort to the "low hanging fruit" of using chatbots to generate completed homework assignments, and teachers and administrators may succumb to the instinct to reflexively ban the technology outright.

In 2019 UNESCO organized its first International Conference on Artificial Intelligence and Education. This conference produced the Beijing Consensus on Artificial Intelligence and Education, in which the participants gave their recommendations for strategies to deal with AI in education, and suggested changes to policies and practices for governments and educational practitioners. Importantly, the Consensus recognized, "…the distinctive features of human intelligence" vis-à-vis artificial intelligence, and called for a "humanistic" approach to implementing AI in education, based on the principles of the Universal Declaration of Human Rights (UNESCO 2019).

Consistent with this focus on humanistic approaches, the Consensus, while recognizing the immense opportunity afforded by AI for aiding instruction and enhancing learner outcomes, emphasizes that, "…human interaction and collaboration between teachers and learners must remain at the core of education", because "… teachers cannot be displaced by machines, and... their rights and working conditions…" must be protected (Consensus 2019). The UNESCO Consensus also highlights issues of equity and ethics, stating that AI should be used in education in such a way to promote high quality education for all students, irrespective of class, gender, ethnicity, or nationality.

To its credit, UNESCO appears to be taking the issue of AI and education seriously, and there have been two more international conferences on the issue after the first in 2019. However, according to its own research, most educational institutions are poorly prepared to deal with the impact of AI. A global survey the organization conducted of more than 450 schools and universities found that fewer than 10% had created policies or guidance for the use of AI in an educational setting. The results of the survey reveal that educational institutions are still struggling to come to grips with the disruptive potential of AI, specifically ChatGPT and other such chatbot applications. Most of the responses seem to be of an ad hoc nature, and many of the institutions that have policies for dealing with AI have not formalized these into a specific set of rules, relying instead upon oral instruction and the discretion of individual instructors.

The strategies developed by local educational institutions for incorporating and governing AI are important because these institutions will react much more quickly than regulators/policy makers at the state/provincial and national levels. Much of the experimenting with how to effectively leverage AI to make the most of its powers, while successfully mitigating or countering its downsides, will be performed by individual teachers and groups of colleagues in schools around the world. This chapter will focus on the issues of ethics and equity as they relate to the use of AI in an educational context. These are factors which, if not properly prepared for, could potentially negatively impact many students as schools begin to formulate their plans to integrate AI.

The chapter will begin by discussing some current and possible future uses of Artificial Intelligence in Education, also known as “AIED”. Next, the discussion will focus on the topic of ideology and technology/AI, factors that affect end-user relationships to technology, and the ways that, consciously or unconsciously, human biases and prejudices can become encoded into new technological systems. Following this, the discussion will turn toward real-world examples of bias in technology and the implications for people subject to the regulations of institutions like schools and courts of law. Among the real-world examples that this section will discuss are those that relate to non-standard and minority language use in schools, and how discrimination against minority languages can impact the performance of students.

Linguistic considerations are especially salient here, because the AI intelligence systems depend on massive databases of recorded human linguistic behavior, called corpora, which in themselves are not equitably representative of the languages within a particular society. This part of the discussion is in many ways the crux of the present chapter and demands special attention. In the many hype and anxiety-filled discussions around AI, the basis of this technology in human practices, which are themselves embedded within cultures and ways of viewing the world, is often lost.

## The Possibilities of AI in Education

Compared to other technologies such as smart boards, tablets, and the internet, the educational possibilities of AI in Education (AIED) have heretofore been largely unexplored, especially when one considers the very recent developments that have made the technology available to a wide, non-expert user base. Some think that the widespread implementation of AI into the field of education has the power to bring about a “4th educational revolution”, which will change the face of education so radically that most of its features—a single teacher that is the central authority within a room of a few dozen pupils, a school day divided by different subjects—will cease to be (Seldon, A., and Abidoye, O. 2018). Although there is much hype surrounding the possibilities of AI in education, if used correctly, AI can aid in shifting education to adjust to the reality of a society in which information is ubiquitous, in which "expert amateurs" who understand concepts are more needed than specialists, and in which schools are no longer needed primarily for content delivery (Holmes, W. Bialik, M. Fadel, C. 2019).

Some proponents of AIED think that it can help to solve a number of current problems in education not directly related to learning, for example, excessive teacher workload leading to burnout, difficulty sharing insights between schools and colleges, and variation in education quality leading to a lack of social mobility (Baker, T. Smith, L. Anissa, N. 2019). Learner-facing AIED, such as adaptive learning platforms, can produce a number of positive effects for students that otherwise would be difficult to effect within a classroom with many students. These AIED can aid in preparing content, helping students to apply knowledge, engaging students in learning tasks, helping students to learn through feedback, and helping students to become self-regulated learners (Lameras, P. and Arnab, S. 2022).

Connecting the teacher-facing and the learning-facing types of AI, there are decreasing levels of teacher involvement and control. Molenaar has posited six steps, from instruction with the teacher alone, all the way to full automation of the teaching process in which the AIED acts alone (Molenaar 2022). The full automation and replacement of teachers is not likely, however, and some sort of intermediate step is projected to be most common, with AIED's acting as Intelligent Tutoring System's (ITS), giving granular, step-wise feedback on student tasks while the teacher sets the main learning goals, for example (Merikko, J. and Kivimaki, V. 2022).

The future of AI in the classroom most likely will result in AI serving as a kind of labor-saving device and task organizer on the teacher-facing side, and as an adaptive scaffolder and predictive guide on the learner-facing side. Given the shortages of new teachers, high burnout rates, and high workloads that teachers and schools face, the proper implementation of AIED could be of great assistance. For students, the ability to have an in-class and possibly at-home tutor that understands the individual learner's past difficulties and growth and is also able to predict future difficulties and adjust accordingly, holds the promise of helping students who might have difficulty in a one-size-fits-all learning environment to reach success.

The next section will address a crucial factor that will affect whether and how educators and institutions adopt AIED, namely, ideology. Whether institutions and individuals meet the new technology with optimism, or whether they approach it with dread and apprehension, will determine the extent to which the possibilities of AIED can be effectively realized. After discussing the importance of ideology, the discussion will move to the very real negative possibilities inherent in the unthinking, unprincipled adoption and deployment of AIED. Real-world examples will be used to illustrate the necessity of serious ethical thinking if AI is to be used in any serious, widespread way in education.

## Ideology and AIED

Ideological orientations towards technology and the attitudes of individuals towards technological change will be an important factor that influences whether institutions proactively prepare to integrate AI or set up institutional roadblocks to its adoption. Researchers investigating technology adoption and consumer practices often dichotomize people's attitudes toward technology; as either optimistically technophilic or pessimistically technophobic (Best and Kellner 2001, Thompson 2004).

Kozinets (2008) avoids this dichotomy through a model of technology ideology that consists of 4 different ideological "nodes" and their interconnections by means of a semiotic square. The basic ideological nodes are the "techtopian", the "green luddite", the "work machine", and the "techspressive". "Techtopian" describes individuals who hold views of technology that emphasize the positive social benefits of technological progress; these are people who believe that improvements in technology can facilitate such humanistic ends as moral improvement and community building. The "work machine" position describes the ideology of those individuals who also have a positive orientation towards technological progress, but whose main focus are the improvements in economic efficiency and productivity. While techno-positive this position is less concerned with the social outcomes of technological progress. The "green luddite" position describes the ideology of people who emphasize the negative impacts of technological change, particularly the negative impacts on the natural environment and upon individuals and societies. The "techspressive" ideological outlook is also optimistic towards new technology but focuses on individual enjoyment and expression afforded by that technology. The techspressive individual is less concerned about any positive benefits that might accrue to society through a new technology than about how that technology can benefit the techspressive personally.

This kind of model helps us to make sense of the reactions coming from educators regarding AI. There are people who are ambivalent about the new technology, not simply as some kind of kneejerk reaction to any and all change, but because they are concerned about the effects on tradition, community, and maybe even "natural" processes of learning and teaching. Others look forward to the new technology precisely because they believe that the goals of communal and individual growth can be better fulfilled by incorporating the new technology. The truth is that all these ideological positions are over-accentuating some factors while diminishing others, so it is important for institutions and educators to plan carefully so that they do not miss out on the great opportunities offered by AI technology, but also so that they do not ignore very real risks.

Ideology is a factor on the consumer side of technology uptake, but it is also an all-too-often unacknowledged component of the design, development, and deployment of technology. Technology (and even science) is not "value-free" or neutral but is a manifestation of the weltanschauung, or world view, and the interests of the people and groups who create and develop it. This is as true in the realm of software and computer applications as in the realm of physical tools. Some would even go as far as to say that "software is a functional analog of ideology" (Chun 2004). An ideological system evaluates entities and objects differently, it hierarchizes things and people and makes some things visible by emphasizing them while rendering other things and processes invisible, or even unthinkable. These very qualities require our keen attention and concern, especially when instantiated within computer applications that have decision-making ability that can affect peoples' lives.

There is evidence that, due to an ability to capture semantics through processes such as analyses of word co-occurrence in text corpora, which are large databases of textual or audio examples of human linguistic behavior, AI applications are able to "learn" cultural biases and prejudices. This is possible because language is a repository and manifestation of the largely unexamined ideological assumptions of a group. This ideology shows up in language when words occur in temporal and situational proximity to each other. The statistical properties of word context can be input into AI, which will then learn meanings and even connotations of words, and in the process, pick up biases present in language (Caliskan, Bryson, and Narayanan 2017).

This ability for human biases and prejudices to become coded into (natural) language, and then subsequently coded into machine languages and software highlights the necessity of considering the ethical implications of deploying AI, especially in an educational context. The problem of bias in all kinds of computer systems, not just AI, has been recognized for decades. Friedman and Nissenbaum (1996) propose a framework identifying three categories of bias in computer systems: preexisting, technical, and emergent. Preexisting bias is based in social institutions, practices, and attitudes that exist independently of, and usually prior to the creation of a technological system. Preexisting bias can be individual, resulting from the desires of an individual who has significant control over the design of the system, or societal, arising from general social values and opinions. Technical biases are a result of practical, physical, and technical limitations and constraints of the systems themselves. Technical bias can be based in limitations in hardware, software, and peripherals, algorithms that do not treat all groups fairly in all situations, imperfections in or misuse of pseudorandom number generation, or attempts to formalize human constructs that are not amenable to quantification or discretization. Emergent biases are observable when systems are in use and usually occur as a result of changes in social values or a mismatch between the original design and the expertise or values of a new population in which the system is deployed.

This framework is valuable because it clearly posits some aspects of technological systems which are counterintuitive, especially for those who may hold an optimistic opinion of technology's effects, or those who believe that technology is inherently neutral. First, it posits that technology does not stand apart from the society that creates it. The biases and tendencies of the societies and individuals who create technology are reflected in the technology itself. For example, biases in data that result from preexisting institutional and individual biases in measurement will affect downstream applications that rely on these datasets. This is because "an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit prejudices of prior decision makers" (Barocas and Selbst 2016). The framework also shows that actual technical constraints (in physical size or computing power, for example) are only one way that technology may be biased, though one that can be overlooked if the perspectives of users who may be disadvantaged are not considered. The framework also shows that new cases of inequity can arise when existing systems are used in contexts outside of those cases for which they were originally designed. Thus, even a hypothetically perfectly equitable system might disadvantage some group if used in a novel context.

Also, technological bias can be a result of the workings of algorithms, whose goal is to minimize overall prediction errors. Therefore, systems that rely on these algorithms will prioritize data from categories with majority representation, since the behavior of these categories gives the best probability of identifying statistically "normal" behavior. In the social context, this can lead to a bias in favor of majority groups over minorities (Pessach and Shmueli 2021). Another type of bias, which will be discussed in greater length below, occurs when criteria that are not consciously chosen for the purpose of disadvantaging any group serve as proxies for socially significant attributes because of their correlation or co-occurrence with these other attributes. In these cases, the bias could still result even if the dataset does not directly or consciously include any preexisting individual or societal bias.

All the above make the actual design of equitable and fair technological systems of paramount importance. Dignum (2018) identifies three relationships between ethics and design in AI: ethics *by* design, which concerns "integration of ethical reasoning capabilities as part of the behavior of artificial autonomous system [sic]; ethics *in* design, which concerns "the regulatory and engineering methods that support the analysis and evaluation of the ethical implications of AI systems as these integrate or replace traditional social structures"; and ethics *for* design, which entails "the codes of conduct, standards and certification processes that ensure the integrity of developers and users as they research, design, construct, employ and manage artificial intelligent systems" (Dignum 2018, 2). All three of these relationships are important for educators to consider when formulating strategies to deal with AI, but perhaps it is most important for them to think about ethics in design. As end-users of these technologies, educators must understand the non-neutral nature of the design of AI applications and their deployment in the classroom.

Issues of access, fairness, and diversity as these relate to the use of AI in schools have become even more important as of late, as educational institutions, governments, and corporations strive to highlight DEI (Diversity, Equity, and Inclusion) initiatives. However, there is some cause to doubt just how substantive and effective such campaigns can be for addressing deep-seated, systemic issues. Related to the above discussion of encoded ideology, despite our best intentions, our blindness to the encoded bias within technology creates situations which perpetuate discrimination and inequality. For example, sloppy statistical modeling employed by job recruiters, financial institutions, or even the judicial system can prevent people from being hired or getting a loan for higher education or can cause a presiding judge to deliver a harsher sentence. These algorithm-enabled outcomes are what mathematician Cathy O'Neil calls "weapons of math destruction". Echoing Chung (2004), O'Neil states that, "models are opinions embedded in mathematics" (Doctorow 2016). This leads to a situation that Ruha Benjamin (2019) calls "the New Jim Code", or "the employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era" (Benjamin 6).

Next, this chapter will identify some ways that those opinions, within the context of AI in education, can harm various groups, ethnic/racial minorities, women, the poor, formerly incarcerated people, etc…, and to suggest some possible ways that these tendencies can be mitigated so that the possibilities of AI can be leveraged to equitably benefit education.

## AI Pitfalls for Diversity and Fairness

There are several examples of O'Neil's sloppy statistical models creating concrete, negative outcomes for marginalized people. A famous case involves a "criminal risk assessment" tool used by courts of law when determining sentencing called COMPAS, or "Correctional Offender Management Profiling for Alternative Sanctions", which "predicts" the chance of recidivism, or the probability that an offender will commit another crime, based on 137 "features", such as employment status or income level. COMPAS, which assigns a numerical recidivism "risk" score for each defendant, is produced by a company called Equivant (formerly named "Northpointe") and has been used to analyze the records of more than 1 million offenders since its development in 1998. However, research has indicated that COMPAS is not very accurate in predicting recidivism, and although race is not one of the features used to create its predictions, the program consistently judged Black offenders as higher risks of recidivism than Whites (Dressel and Farid 2016).

In May 2016, writing for *ProPublica*, Angwin et al. analyzed the COMPAS outcomes of more than 7000 individuals arrested in Broward County, Florida between 2013 and 2014. They determined that COMPAS's overall accuracy for white defendants was 67.0%, only slightly higher than its accuracy of 63.8% for black defendants, but the mistakes made by COMPAS did not affect Black and White defendants in the same way. COMPAS incorrectly predicted that Black defendants who actually did not go on to reoffend would do so at a rate of 44.9%, nearly twice as high as their white counterparts at 23.5%; while white defendants who did reoffend were incorrectly predicted to not commit further crimes at a rate of 47.7%, nearly twice as high as their black counterparts at 28.0%. Therefore, the COMPAS scores underpredicted recidivism for white defendants, and overpredicted recidivism for black defendants.

COMPAS derives its scores from a set of 137 questions. The race of the defendant is not one of the questions, rather the questionnaire asks questions like: "'Was one of your parents ever sent to jail or prison?', 'How many of your friends/acquaintances are taking drugs illegally?', and 'How often did you get in fights while at school?'" (J. Angwin, et al. 2016). Equivant/Northpointe challenged *ProPublica*'s study, citing the lack of race as a one of COMPAS's 137 features and arguing that this means that COMPAS's scores are not influenced by any sort of racial bias. However, the lack of race as an explicit criterion does not mean that race is not a factor in how the scores are produced. Although very difficult to define, "race", as a social phenomenon, indicates some kind of caste-like social category, determined almost exclusively by phenotype, particularly skin color, but also usually including as well socio-economic class, education level, occupation, linguistic behavior, criminal history, as determining characteristics. The fact that COMPAS did not directly inquire as to the race of an individual defendant, yet still yielded disparate scores for Blacks and Whites in the aggregate, with the former being penalized more than the latter, implies that COMPAS was somehow outlining the "contours" of race without drawing a clear picture. In the process, the software makes transparent the significance of race in America. To be Black is to be less employed, less educated, more homeless, more in poverty, more likely to have a criminal record, and more likely to be associated with people for whom all or most of the aforementioned apply. To be White is to be less likely to be penalized (or simply to be caught) for possessing these same characteristics. In a way, COMPAS was not wrong because it merely reproduced the circular logic of race in America: A Black defendant is a higher risk for recidivism because that is what the defendant is supposed to do, inner workings of the system of race, expressed as a set of statistical probabilities or scores, make this all but a fait accompli, though the exact mechanisms are all but impossible to ascertain.

Of course, judges consider other factors when deciding sentencing, even when using a COMPAS score, but subconscious influence of a simple, hard, whole number must be difficult to resist. The desire to avoid human biases while simultaneously increasing efficiency is an understandable motivation for using AI applications to classify human populations in this way. But our faith in AI is misplaced, and this is in part a result of how we talk about artificial "intelligence". Part of the creation of AI was the changing of the definition of "intelligence" and the creation of a set of terms with which to talk about AI. We now refer to refrigerators that can adjust their own temperatures as "smart", for example, but this is just a "remarkable feat of discursive reconstruction," that "…resigni[es] what intelligence, agency, and self-awareness are…creat[ing] a new reality in which being smart has an entirely different meaning than it used to (Blikstein et al. 2022).[1]

There is another reason that the faith in the ability of AI to render equitable decisions without human intervention is misplaced is that, as with O'Neil's "Weapons of Math Destruction" and COMPAS's risk assessments, is that the large datasets that AI applications draw from to recognize patterns are themselves embedded and produced by social realities which are not equitable. For example, in a study which explored computational methods for deriving word meanings from co-occurrence (the appearance of words in proximity) statistics drawn from linguistic corpora like the British National Corpus (BNC), the researchers noted that "corpus quality" could affect semantic representations produced by their computational methods (Bullinaria and Levy 2007). The BNC draws textual examples from a variety of sources, so "corpus quality" could be affected by examples drawn from a source in which word frequency

[1] This leads to a subtle way that the use of AI can have a negative impact on education. If the new, artificial intelligence definition of "intelligence" becomes normalized, learners might become compared to the AI tools that should help them learn, with those students who more closely approximate the new definition being the more successful.

distributions are unusual, or by examples drawn from *nonstandard English* (the authors use the example "picture window" vs. "pitcher winder").

Nonstandard English varieties could include forms such as Scottish English, African American Vernacular English (AAVE), or Jamaican Patois. Linguists have long recognized that these nonstandard forms, which occur within all language families, are just as systematically structured and capable of expressing thoughts as standard (or prestige) forms. As sociolinguists point out, however, the factors that make certain forms of speech "standard" and other forms nonstandard or dialect—and thus possibly degrading the quality of text corpora—are related to matters of history, social hierarchy, and status. Most likely, the authors of the above study did not intend to imply that nonstandard forms were inherently mistaken but only wished to convey the idea that, relative to the target language of their study, the nonstandard forms would present issues for their attempts to derive semantics from word co-occurrence data. But this kind of unintended classification of marginalized or nonstandard forms, of any type, linguistic or otherwise, as an error is precisely what must be avoided when introducing AI into education.

This issue has a long and bitter history in the field of American education. The federal court cases *Lau v. Nichols* (1974) and *Martin Luther King Junior Elementary School Children v. Ann Arbor School District Board* (1979) both concerned the ability for students from minority populations to receive instruction that would help them to bridge the gap between their home languages and the Standard American English (SAE) used in school. Lau concerned limited-proficiency Chinese American students in California and was decided before the passage of the 1974 Equal Educational Opportunity Act (EEOA); MLK v Ann Arbor, which made use of the language provision of the EEOA, involved a group of African American students for whom AAVE was their home language.

Although these cases were decided in favor of the plaintiffs, recognition of the validity of nonstandard linguistic forms and the utility of using the home languages of non-native English-speaking students has remained weak within American society and schools, especially when those linguistic forms are related to marginalized and subordinated groups. This issue became a heated national controversy in the 1990s when the Oakland Unified School District (OUSD) in Oakland, California, following the recommendations of a task force convened to find ways to overcome the disparities in educational attainment between Black and White students, decided that recognition and understanding of AAVE would enhance Black students' learning of SAE. In pursuing this decision, the school district consulted with linguistics experts who explained the benefits of using students' nonstandard home languages to help those students master the standard form (Rickford 2005). Although the OUSD did not intend to teach AAVE to Black students, only to use it to teach SAE, the public and the media misunderstood their aims, and decried even the presence of AAVE (called "Ebonics" by some) within schools, equating it with street slang or unintelligent, ungrammatical English (Wheeler 2016).

The "Ebonics" controversy is a glaring example of the socio-political aspect of language as a behavior, and of how attempts to improve education for marginalized students have been stymied by the inability of those aligned with the dominant culture in a society to recognize the validity of nonstandard languages as means of expression and cognition. The controversy also highlights an area for caution in the development of AI and its use in education. The difficulties for development are not so much technological but require awareness of ideology and some principles of sociolinguistics. AI can be trained on corpora and datasets to recognize and to produce text from nonstandard languages, but only if these languages are explicitly considered, and of course, assuming that corpora for these forms exist. In an educational context, if AI are used to rate the reading comprehension of students, for example, the AI will have to recognize when nonstandard speakers are translating English into their home language and not penalize them for this (Wheeler 2016). This might happen when a speaker of AAVE drops the final "-ed" from past tense and past participle forms of verbs, turning "He dropped the ball" into "He drop the ball", for example (Brennan 2018).[2]

The work of Justine Cassell provides some examples for how designers and educators can implement AI in education in a way that acknowledges and respects the varieties of home languages that students bring to the classroom. Cassell, a professor in Carnegie Mellon University's School of Computer Science, studies the use of virtual peers and conversational agents (called "embodied conversational agents") to enhance education. These agents are multimodal interfaces which can have conversations with humans and understand and produce speech, gesture, and facial expressions (Paranjape, B., Bai, Z., Cassell, J. 2018). In her research with children, Cassell uses virtual avatars that children of different backgrounds can identify with visually. She also adjusts the language for children of different backgrounds, and so designed a system that spoke AAVE to African American children, which she claims helped the children to achieve better results when studying science (Hardy 2016). However, when a student must work on a class presentation, Cassell's agent uses SAE, because SAE is still "the code of power" and it is crucial that African American students be able to use it well.

Cassell, who has a background in linguistics, is using the exact same method that the Oakland Unified School District wanted to implement. This kind of sociological and sociolinguistic knowledge is exactly what must inform any educational AI tools, and Cassell's work serves as a good example to educators and designers who want to begin to implement AI in classrooms. A very important aspect of what Cassell is doing is the awareness of power dynamics in society, and the attempt to help students to deal with these from a position of strength. One of the ethical problems related to AI is that the decision-making processes are black boxes. As in the example of COMPAS, the lives of real people may be massively affected by a decision made by an AI, but the processes that lead up to that decision are occluded, locked in a "black box" of proprietary restrictions. The same process actually take place when humans make these decisions, but with AI, since these processes are deliberate choices, we can make some kind of intervention to improve fairness and equity. At least, humans can intervene at the beginning stages of design. Once an AI agent begins to work on its own, any biases it may develop will become more deeply entrenched. This is exacerbated by the ability of AI agents to develop processes which are opaque to even their human creators. This is a strong argument for designing

AI with these ethical considerations from the start, and constantly monitoring them to make sure that the outcomes align with broader ethical goals.

## Broader Concerns

Outside of questions of fairness and ethics within schools and classrooms, we must also look to broader systemic structures. AI can be leveraged to enhance education, but we must be aware of the practical realities of differential valuation relative to educational experiences and pedigrees. Even if the introduction of AI alone into the classrooms of poor, urban, or rural districts were enough help students from these contexts to equal or even exceed the performance of their more privileged counterparts, increased ability alone is not enough to place these students on an equal footing. While educational performance, students might be equal, in the area of prestige students will still not be treated the same way. Just as a community colleges and regional universities contain many brilliant students who are nevertheless judged unfairly compared to their peers who graduate from Ivy League institutions, we should guard against the primary and secondary educational experiences of poor, minority, and rural students becoming a kind of digital assembly line experience.

We know that not only what students learn, but how they learn it, is an important part of modern education. This "hidden curriculum" prepares students from different class backgrounds for life in the adult world (Anyon 1980). Even if the inclusion of AI can raise educational attainment, there is still the very real possibility that the way AI is used in different school settings may "…facilitat[e] the perpetuation of an unequal division of labor in U.S. society, where some… will plan and others… will have jobs that entail carrying out the plans"; or, to update to the present situation, some will be managed by people and AI, and other will manage people and AI (Anyon 1981, p. 35). Unfortunately, such stratification is already taking place in other areas of education.

Cyber charter schools, variously called virtual charters or online charter schools, have been increasing in size and number over the past two decades. An online variation of the charter school concept, receive charters from the states in which they operate and can receive public funds. Many states have turned to online charter schools as a solution for educating students from challenging, under-performing urban areas and students in rural areas where schools may be scarce. There is some evidence that charters seriously underperform when compared to traditional in-person instruction at “brick-and-mortar” public schools (Lafee 2016). Despite the dubious benefit for at-risk learners, there has been rapid growth of these schools in areas like the overwhelmingly (80%) Black and Latino Philadelphia-area school system, even though 100% of the students enrolled in cyber charters in that city failed state achievement tests between 2011 and 2014 (Rooks 2017). If AI is to make a positive impact on the educational attainment of students from all backgrounds, it must not be simply one more digital solution on top of a pile of already failing digital/online strategies. The poorest school districts with the students who most need good educational resources could, due to budget cuts and staffing difficulties, be left with overworked teachers leading overstuffed online classrooms largely “taught” by poorly calibrated AI agents. The differences in the “hidden curricula” of students graduating from this kind of scenario and those who attend public or private schools wealthy enough to keep class sizes small and to hire highly qualified teachers who can make good use of any AI technology would be very stark.

## Conclusion

This paper has attempted to address the eventuality of an AI-enhanced educational future in a balanced way. Following a discussion of the possible uses of AI in education (AIED), the paper discusses the importance of ideology in shaping how educational institutions and individual educators react to AIED. Ideology is important because a technology-negative outlook can prevent the schools from benefiting from the AIED, while an overly optimistic approach will lead adopters to miss serious concerns that must be addressed if AIED is to benefit and not harm institutions, communities, and learners.

The ideology of the individual school or teacher is not the only important factor. These new technologies are being developed and implemented within a larger social, cultural, political, and economic context. Whatever the desires of the teachers and schools, there are external processes beyond their control. One group of processes relates to the actual design of AIED utilities and agents. If the designers pay no attention to questions of social justice or ethics, then the most careful use of the technology will still result in unjust outcomes. This article only briefly mentioned other considerations, like the labor and teacher pay, that also must be figured into any ethical and responsible use of AIED. These are questions that must be thought about seriously, but they seem to be missed in the larger sensational public discussions about plagiarism.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. ProPublica, 23 May 2016; https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Anyon, J. (1981). Social Class and School Knowledge. Curriculum Inquiry, 11(1), pp. 3-42.

Baker, T., Smith, L., & Anissa, N. (2019). Educ-AI-tion rebooted? Exploring the Future of Artificial Intelligence in Schools and Colleges. https://www.nesta.org.uk/report/education-rebooted/.

Barocas, S. and Selbst, A. (2016). Big Data's Disparate Impact. California Law Review, 104(3), pp. 671-732.

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code. Cambridge: Polity Press.

Best, S., & Kellner, D. (2001). The Postmodern Adventure: Science, Technology, and Cultural Studies at the Third Millennium. London: Guilford.

Blikstein, P., Zheng, Y., & Zhou, K. Z. (2022). Ceci n'est pas une école: Discourses of Artificial Intelligence in Education through the lens of Semiotic Analytics. European Journal of Education, 57, pp. 571-583. https://doi.org/10.1111/ejed.12528.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting Semantic Representations from Word Co-occurrence Statistics: A Computational Study. Behavior Research Methods, 39(3), pp. 510-526.

Brennan, W. (2018). THE CODE-SWITCHER: Julie Washington's lifelong quest to change the way we teach young speakers of African-American English. The Atlantic, 321(3).

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), pp. 183-186.

Chun, W. H. K. (2004). On Software, or the Persistence of Visual Knowledge. Grey Room, 18, pp. 26-51.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023): Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT. Innovations in Education and Teaching International. https://doi.org/10.1080/14703297.2023.2190148

Dignum, V. (2018). Ethics in Artificial Intelligence: Introduction to the Special Issue. Ethics and Information Technology, 20, pp. 1-3. https://doi.org/10.1007/s10676-018-9450-z

Doctorow, C. (2016). Weapons of Math Destruction: Invisible, Ubiquitous Algorithms are Ruining Millions of Lives. https://boingboing.net/2016/09/06/weapons-of-math-destruction-i.html Accessed, 5 July 2023.

Dressel, J., & Farid, H. (2018). The Accuracy, Fairness, and Limits of Predicting Recidivism. Science Advances, 4(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580

Friedman, B. and Nissenbaum, H. (1996). Bias in Computer Systems. ACM Transactions on Information Systems, 14(3), pp. 330-347.

Hardy, Q. (2016). Seeking a Choice of Voices in Conversational Computing. New York Times, 9 October. https://www.nytimes.com/2016/10/10/technology/looking-for-a-choice-of-voices-in-ai-technology. html. Accessed 9 July 2023.

Holmes, W., Bialik, M., Fadel, C. (2019). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston, MA: Center for Curriculum Redesign

Kozinets, R. V. (2008). Technology/Ideology: How Ideological Fields Influence Consumers' Technology Narratives. Journal of Consumer Research, 34(6), pp. 865-881.

Lameras, P., & Arnab, S. (2022). Power to the Teachers: An Exploratory Review on Artificial Intelligence in Education. Information, 13(14). DOI: 10.3390/info13010014.

Merikko, J., & Kivimäki, V. (2022). "Replacing teachers? Doubt it." Practitioners' views on Adaptive Learning Technologies' Impact on the Teaching Profession. Frontiers in Education 7. https://doi.org/10.3389/feduc.2022.1010255

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. European Journal of Education, 57, 632–645. https://doi.org/10.1111/ejed.12527.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York, NY: Crown Books.

Paranjape, B., Bai, Z., Cassell, J. (2018). Predicting the Temporal and Social Dynamics of Curiosity in Small Group Learning. In: Penstein Rosé, C., et al. (Eds.) Artificial Intelligence in Education. AIED 2018. Lecture Notes in Computer Science, 10947. Cham: Springer. https://doi.org/10.1007/978-3-319-93843-1_31

Rickford, J.R. (2005). Using the Vernacular to Teach the Standard. In J. D. Ramirez, T. G. Wiley, G. de Klerk, E. Lee, & W.E. Wright (Eds.), Ebonics: The Urban Education Debate (pp. 18-40). Tonawanda, NY: Multilingual Matters.

Rooks, N. (2017). Cutting School: Privatization, Segregation, and the End of Public Education. New York, NY: The New Press

Seldon, A., and Abidoye, O. (2018). The Fourth Education Revolution: Will Artificial Intelligence Liberate or Infantilise Humanity?. Buckingham: University of Buckingham.

Thompson, C. J. (2004). Marketplace Mythology and Discourses of Power. Journal of Consumer Research, 31 (June), 162–80.

UNESCO (2023). UNESCO Survey: Less than 10% of Schools and Universities have Formal Guidance on AI. https://www.unesco.org/en/articles/unesco-survey-less-10-schools-and-universities-have-formal-guidance-ai. Accessed 5 July 2023.

UNESCO (2019). Beijing Consensus on Artificial Intelligence and Education. https://unesdoc.unesco.org/ark:/48223/pf0000368303. Accessed 5 July 2023.

Wheeler, R. (2016). "So Much Research, So Little Change": Teaching Standard English in African American Classrooms. Annual Review of Linguistics, 2, pp. 367-390. https://doi.org/10.1146/annurev-linguistics-011415-040434