8

PROCEED WITH CAUTION

The Pitfalls and Potential of AI and Education

Kelly Shiohira and Wayne Holmes

8.1 Introduction

The arguments for the integration of AI into education are multiple and multifaceted. AI has permeated everyday life, and there is a growing number of AI-assisted educational technologies being implemented in classrooms worldwide. Like any tool, AI can be used to improve or to harm society. From a humanistic perspective, citizens need to understand their roles and rights with respect to AI, recognize when they are unfairly disadvantaged by AI and know the avenues of recourse, and above all become conscientious users of AI products—particularly AI products designed for education.

At the same time, calls to skill or upskill individuals in topics such as big data, the data value chain (the process of data creation and use from first identifying a need for data to its final use and possible reuse), and cloud storage are coming from academia and industry, with some companies substantially invested in training partnerships with universities and other educational institutions, as well as the development of their own platforms and courses.[1] The preoccupation with data is nearly universal: Customers are themselves the product for a number of big corporations, which include not only social media platforms but also less conspicuous companies that typically exploit data such as supermarkets and health insurers. Accordingly, the demand for data scientists and engineers has risen sharply in recent years and shows no sign of slowing down. Education is expected to deliver this expertise.

AI is also frequently hailed as a "solution" to core problems in education, including a lack of qualified teachers, poor learning outcomes, and the need to manage achievement gaps (Davies et al., 2020; OECD, 2021b; Seldon & Abidoye, 2018). A multitude of products have been developed and billions of

DOI: 10.4324/9781003230762-11

dollars invested in AI for education, and some studies show promising results for content knowledge gains (e.g., Major et al., 2021). However, wide-scale implementation and testing have yet to occur, even as a range of claims are being made without robust evidence (Miao and Holmes, 2021; Bryant et al., 2020).

At the same time as it is being integrated into education, the practicalities of many types of AI conflict with a fundamental education philosophy that education is, as Nelson Mandela noted, "the most powerful weapon that you can use to change the world". The Declaration on the Rights of the Child calls education an "amplifier of all other rights". Education is also often cited as a tool to break cycles of poverty (e.g., Conchas, 2008). While the success of education in these areas may be up for debate, underlying all of these assertions is a belief that education can change the future of an individual, regardless of current circumstances.

However, the AI systems being introduced into education are built on statistical models. Predictions are extrapolations based on trends or patterns derived from previous data. Learners are clustered based on their common attributes and treated commonly. In short, there is a set model to be followed, and each person can be plotted along it, defined by the average experience. At the very least, the dichotomy between education as an enabler of the extraordinary and AI as reliant on the predictable must be acknowledged. This conversation begins by considering the current and potential uses of AI in classrooms and other learning environments; the limitations of learning *with*, *through* and *about* AI; and more broadly what the role of AI *should* be—the types of problems it is best suited to addressing and what exchanges are made in its implementation.

However, first, we need to be clear about what we mean by AI, which is actually notoriously difficult to define. One helpful definition is given by UNICEF (2021):

> Machine-based systems that can, given a set of human-defined objectives, make predictions, recommendations, or decisions that influence real or virtual environments. AI systems interact with us and act on our environment, either directly or indirectly. Often, they appear to operate autonomously, and can adapt their behaviour by learning about the context.

It is important to reiterate that AI is not a machine that thinks independently. AI is merely a set of human-created algorithms that aim to mimic some human thought processes, such as decision-making. They do so to a greater or lesser extent, using a range of database techniques (machine learning approaches such as supervised, unsupervised, reinforcement and deep learning, all of which depend on huge amounts of data) and knowledge-based techniques (e.g., model-based expert systems) (Holmes et al., 2019).

All of these techniques raise ethical issues, about which there is no universal consensus, although many sets of principles (Jobin et al., 2019). Meanwhile, international organizations have begun to publish normative instruments such as the Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021). Drawing on these and related sources, the key terms employed in discussions of AI ethics are:

- **Accuracy**: Potential and actual sources of error or bias must be identified, recorded, and addressed.
- **Bias**: Cases in which an algorithm discriminates based on gender, race, ability, or other characteristics.
- **Transparency**: The possibilities to (i) access, review, monitor and criticize algorithms (*auditability*) and (ii) understand the outputs of an algorithm (*explainability*).
- **Fairness**: Ensuring that algorithms and their outputs do not lead to discrimination, while everyone has equal access to AI and its benefits.
- **Privacy**: The right of individuals to understand and control the use of their data (text, sound, image), and protect it from exposure to risks such as identity theft or misuse.
- **Responsibility** or **human oversight**: Human agents must be able to change or reform an algorithm and make timely redress in the case of adverse effects.
- **Safety** and **security**: AI must not cause harm, such as discrimination, violation of privacy, and bodily harm, or have negative psychological, social, economic, and emotional effects.
- **AI for the Public Good**: AI should be used for public, societal, individual, and/or economic good.

8.2 Implementing AI in Education

AI is being implemented in education in three main ways: Through learning *about* AI, or programs of study concerned with teaching AI as a subject; learning *with* AI, or applications of AI to the learning of other subject areas; and preparing *for* AI, or the education necessary to understand the current and future impact of AI on society, individuals, work and the environment.

8.2.1 Learning with AI

"Learning with AI", or using AI tools to support teaching and learning in other subjects, also known as "AI in Education" or AIED, may be further categorized as system-supporting, student-supporting, and educator-supporting.

8.2.1.1 System-Supporting AI

System-supporting AI includes AI tools designed to help with student recruitment, timetabling, finances, and other back-end administrative tasks required of educational institutions. Some innovative system-supporting AI applications are currently being explored, for example, the Skills-OVATE[2] portal in Europe, which uses data scraping of online job adverts and AI to reveal competencies required by businesses in order to enable closer matches between skills offerings and skills demand, and the PSET CLOUD[3] in South Africa, which seeks to leverage interoperability of existing government data systems, big data and AI to support improved decision-making by government, educational institutions and citizens through an analysis of skills supply and demand in the country.

8.2.1.2 Student-Supporting AI

Student-supporting AI has been developed over the last 40 years and is now offered by a wide range of "EdTech" corporations. Covid-19 also triggered a new round of corporate investment in AI EdTech, with the global corporate investment in AI EdTech quintupling between 2019 and 2020, launching education applications into the top three AI investment areas that year (Zhang et al., 2022).

The most prominent type of student-supporting AI is the so-called "intelligent tutoring systems" or "personalized learning platforms". These screen-based systems provide some information, an activity, and possibly a quiz, and as the student engages, they collect back-end information such as performance, time on task, and type of error. These inputs are used to direct further learning, so students follow their own adapted but still mostly rote-learning pathways. Most of these platforms focus on primary and secondary school learners, for example, ASSISTments[4] or Byjus,[5] but there are some examples of government systems focused on adult learning such as the FutureSkills[6] initiative in India. Some platforms such as the ViLLE[7] learning platform from Finland enable learners to choose difficulty levels and engage from multiple locations (UNESCO, 2022).

Other student-supporting AI includes dialogue-based tutoring systems which use a dialogic Socratic approach to teaching and learning; chatbots (an example is Genie, Deakin University, 2018), which leverage natural language processing; augmented reality (e.g., Vulcan[8] uses augmented reality and sensor information to provide immediate feedback on simulated tasks such as welding, carpentry, and painting); automatic writing evaluation tools; and exploratory learning environments.

Automatic writing evaluation tools seek to evaluate long-form text with a similar aim to that achieved by computer-graded multiple-choice tests. They typically evaluate an input such as a paragraph or an essay against a database of graded standardized essays, for example, those submitted for national assessments. These systems can provide immediate feedback but they are not without issues (e.g., Feathers, 2019). For example, automated writing assessment can be completely fooled by nonsensical but lyrical prose, and forces conformity to a majority-based writing style, "unable to accept effective rhetorical and stylistic uses of language from alternative traditions derived from class or race" (Hockly, 2019: 84).

In exploratory learning environments, students engage practically to modify or build and then test components of a computer simulation or model. For example, open-source physics enables students to build computer-based representations of physics models, such as throwing a ball with different amounts of force to see the various results.[9] Similarly, FractionsLab provides virtual manipulatives that allow students to explore foundational fractions concepts and processes (Mavrikis et al., 2022). Despite their flexibility, exploratory learning environments are not unstructured, but have set learning outcomes and intended learning processes (de Jong, 1991). However, unlike intelligent tutoring systems, exploratory learning environments accommodate open-ended tasks and focus on the process of learning as well as content knowledge (Lameras & Arnab, 2021). Critics of exploratory learning environments cite poor learning outcomes due to cognitive overload, but AI-assisted exploratory learning applications may provide automatic guidance and feedback to help learners navigate the intended learning processes, achieve the prespecified outcomes, and avoid becoming stuck in unproductive learning cul-de-sacs (Holmes et al., 2019; Fratamico et al., 2017).

8.2.1.3 Educator-Supporting AI

According to a recent survey (Bryant et al., 2020), teachers work an average of 50 hours per week, with about half of that time allocated to direct student interactions. The rest is allocated to preparation, administration, professional development, and evaluation and feedback. The report speculates that 11 of these administrative hours and two direct contact hours per week could be eliminated through the use of artificial intelligence. In fact, most student-supporting AI tools do replace some teacher functions, and thus presumably do reduce some demand on teacher time, which certainly sounds like a benefit. However, what is lost in the process is also a discussion worth having. What do teachers learn about individual students during their marking and assessment? What does the reduction of this time yield in practice? These are questions that have not yet been properly explored, much less answered.

However, there are glimmers of AI tools designed specifically to directly support teachers, for example, Graide[10] and X5GON.[11] Graide supports teachers while they are grading assignments but, unlike with automatic writing evaluation, it is the teacher who does the grading, not the AI. Meanwhile, X5GON enables teachers to identify open educational resources from across the internet to help them with their specific teaching requirements.

8.2.2 Learning about AI

Students and adult learners "learn about AI" techniques and technologies, through traditional as well as informal and non-formal learning opportunities. In traditional learning environments, AI is treated as a course, subject or part of an existing subject, and a curriculum is developed with accompanying learning outcomes and assessment frameworks. However, a review of 148 UNESCO member states in 2022 found that only 21 governments have developed curricula for AI as a mandatory and/or elective subject, while only 27 have developed AI as a component of an existing ICT or IT curriculum, suggesting that this space is still very much in development. At the primary level, students typically learn to recognize examples of AI and engage with coding; while at the secondary school level, students deal with data, coding, and the integration of AI into society. At both levels, students are expected to know the definitions and features of AI, how to apply AI to solve problems or perform tasks, and how to self-advocate if their rights are violated (Miao and Shiohira, 2022). At the tertiary level, universities internationally offer degrees, modules and short courses in a range of AI subjects, including data analytics, machine learning, neural networks, big data, and so on, culminating in anything from a course certificate to a postgraduate qualification. Meanwhile, non-formal learning opportunities[12] include a wide range of AI courses and tutorials available for students aged 8 and up (though it must be noted a majority require Internet access and so are inaccessible to millions of young people worldwide) as well as "hackathons" or similar initiatives which can be run by government, industry or the third sector. Informal or unstructured learning opportunities include participation in groups such as school or community clubs invested in AI and related topics.

8.2.3 Preparing for AI

There is no doubt that AI is having a significant impact on the lives of the half of the world's population currently connected to the Internet. It is impossible to go online and not somehow be affected by data harvesting, targeted advertising or personalized recommendations. Even offline, AI is everywhere from setting traffic light intervals to making decisions about financial loans, credit card limits, and visa applications. One type of AI, known as "Generative Adversarial Networks" (GANs), can even be used to create fake but plausible photographs of people. It has also been predicted that half of the current human job tasks could be automated (Manyika et al., 2017; Shiohira, 2021).

There is, therefore, an urgent need for all "learning about AI" to include preparing for life with AI as an integral part of the course of study—and yet, these aspects are all too often ignored (Miao and Shiohira, 2022). Other than one excellent example that invests heavily in understanding the role of AI in everyday life (the MIT "Middle School AI Ethics Curriculum"), most of the courses just mentioned typically invest almost exclusively in the technological dimension of AI, learning what AI is, how it works, and its various capabilities. Few curricula pay significant attention to the human dimension of AI, for example, the positive as well as negative aspects of its integration into everyday life. Ethics, where they are addressed, are primarily based on data challenges, with little reflection on social effects, environmental impact, or the protection of human rights.

However, the latest version of the EU's DigComp Digital Competency Framework (2.2) does specifically look at the impact of AI on humans and the competencies that all citizens should have to enable them to deal with the growing issues, with a focus on misinformation and disinformation in social media and news sites, the exploitation of personal data, data protection, privacy and AI ethics. The Council of Europe is also investigating the impact of AI and education on human rights, democracy and the rule of law (due 2022). In addition, the EU established an expert group[13] that aimed to apply the EU's Ethics Guidelines for Trustworthy AI to education and training. These Guidelines focus on issues such as respect for human autonomy, prevention of harm, fairness, and explicability (European Commission, 2019). However, while these European initiatives are important, they are not well-known. Nonetheless, everyone, from school-aged children to the adult population, requires opportunities to gain the AI knowledge, skills and attitudes recommended by DigComp 2.2, and to become familiar with the issues identified by the EU's expert group.

8.3 Ethics of AI&ED

Readers will have already noticed a number of ethical challenges to the application and teaching of AI in education, such as whether the pedagogical approaches used by a majority of AI are sound, whether the AI applied is effective at improving learning outcomes, and whether sufficient attention is paid to "preparing for AI", the fundamental education needed by everyone to successfully navigate life in an AI world. This section presents a few additional areas in which current practices in AI&ED may be less than ethically robust.

8.3.1 AI&ED and Human Rights

Human rights are universal freedoms and protections that all people are entitled to, regardless of origin, race, gender, ethnicity, or ability. Examples of human rights include the right to life, freedom, and security and the right to a fair trial (UDHR: UN, 1948; ECHR: Council of Europe, 1953). The United Nations Convention on the Rights of the Child (UNCRC: UN, 1989) further details the rights of children, to include, among many others, the right to education, protection from economic exploitation, and privacy.

AI is often declared an instrument to ensure access to quality education, and has been proposed as a "solution" to both the global teacher shortage and the scarcity of teachers in high-risk environments like conflict areas, refugee camps, and remote rural areas.[14] However, AI cannot replace human teachers (Kolchenko, 2018). These contradictory arguments reveal one of two disquieting truths. Either the intention *is* to create AI that replaces teachers, and test this on our most vulnerable populations; or we are content to accept "good enough" for the world's most marginalized and disadvantaged learners by providing technology as a substitute for human teachers.

Another human right is the right to effectively contest decisions, and for decisions with a significant impact on an individual's life not to be made solely by automated processing. The decisions involved in AI programs such as classifying learners by ability certainly have a significant impact on a student's life, and predictive or summative uses of AI such as the now-infamous and disastrous use of algorithms to predict final examination grades in the UK are clearly problematic (Quinn, 2020). The implications are that AI must always be subjected to human oversight, and the final responsibility for decisions made should be held by people. Additionally, avenues of redress or complaint must be defined in the AI&ED space. Due to a lack of transparency around how decisions are made and a conflation of roles between teachers and some AI products, complaints, queries, or other concerns about outcomes remain unresolved.

8.3.2 The Effectiveness of AI in Education

This chapter has mentioned a small number of the AI tools currently in use in education, for all of which there is limited evidence of their effectiveness. And while some metanalyses of "personalized learning" tools show moderate content learning gains (e.g., Zheng et al., 2021), they are based on a small number of studies conducted across a wide range of contexts and technologies, at times sponsored by the companies who developed the technology in question. One unintentional but revealing statistic is the tiny fraction of relevant studies that meet the rigor criteria for a metanalysis. For example, an initial search by Major et al. (2021) returned 198 potential studies after title screening. However, only fifteen met the criteria (appropriate research design, validity, and reliability). Another challenge is the amount of AI research funded by AI companies. An investigation into tenure-track academics at four universities found that over half had received funding from technology companies (Abdalla & Abdalla, 2021).

For the last 50 years, education research has emphasized the importance of learner agency and learner-centric approaches to teaching and learning (Williams, 2017; Hildebrand, 2018). However, there is little evidence of learner-centric approaches in AI in education, despite the frequent references to so-called "personalized-learning". In fact, learners using AI tools may have less control over their learning, certainly have less control over their data contributions to the education system, and may have less ownership of their own learning outcomes than they would be using traditional paper-based learning methods (Lupton & Williamson, 2017).

At the same time, most student-supporting AI tools adopt a behaviorist or instructional approach to teaching and learning that involves direct instruction or "spoon feeding" information. This type of learning prioritizes remembering over thinking, facts over critical engagement, and undermines what is known today about robust learning as well *as* some of the most cited "transversal" or "soft" skills required for life and work in the 21st century, sometimes abbreviated to the "4Cs": Creativity, Collaboration, Communication, and Critical Thinking.[15] And in so doing, these AI applications can also disempower educators, turning them into mere technology facilitators. It is also unclear whether AI could ever provide the type of time-sensitive responsive education that considers not just academic needs but socio-emotional context, personal interests, and the interpersonal dynamics of teaching and learning. It is important to recognize the amazing skills that human educators bring to the classroom, which no AI tool can replicate.

While the link between AI and personalized learning is often overstated, a link between AI and competency-based education (CBE) is often understated. CBE is a model of education increasingly making its way from the Technical and Vocational Education and Training (TVET) sector into higher education and primary and secondary curriculum design, and in fact CBE is now the dominant form of education globally (Tan et al., 2018). CBE seeks a "fixed learning, flexible time" approach, and requires students to demonstrate the application of knowledge, skills, and values gained during programs to applicable contexts through assessment. In TVET, the assessment criteria are closely linked to industry standards, either through the co-creation of the curriculum and learning outcomes or through the development of criteria based on industry research (Keevy et al., 2021; Johnstone & Leasure, 2015). Whether industry standards should be similarly used in the education of children, which is increasingly happening, is a contentious issue that warrants much more debate.

8.3.3 Personalized Learning

The idea of giving each individual student exactly what they need when they need it is appealing. However, partly because "personalisation" has emerged from the marketing industry and Silicon Valley, there are questions around the extent to which AI can actually achieve personalization and whether it is as good as it might at first appear (Watters, 2021). The first confounding factor is that learning and education are primarily social activities, while the pedagogy adopted by most AI applications is individual and even isolated. The second point is that, even if personal, AI-assisted educational tools typically only provide learners with their own individual pathway through a predefined set of materials, while leading to the same fixed outcomes as everyone else (Holmes et al., 2019, 2022).

The third point is that while a teacher can frame learning for individual students based on their shared experiences and knowledge of the student, AI tools can only provide content based on a common cluster. In other words, AI will provide exactly what the learner needs—as long as exactly what they need is aligned to the common or average needs of other learners deemed similar by whatever parameters are built into the algorithm. In any case, the vast majority of these tools ultimately undermine student agency rather than contribute to it. The student has no choice but to do what the AI requires, meaning that there's no opportunity for them to develop self-regulation skills or to self-actualize in the classroom.

8.3.4 Data Ownership

Education has always been in the business of collecting and analyzing student data, generally with no more than a consent inferred from enrolment. What data is collected and how it is used is not always a transparent process even before AI gets involved, and very little control over their data is granted to students. In fact, databases of student examination data have been used to develop some automated writing evaluation systems (e.g., Letrus[16]), likely unbeknownst to and certainly of no direct benefit to those who have already completed their examinations.

What distinguishes the use of AI in classrooms is scale. Using an AI platform, a single student can generate more than 5 million digital traces in a single day (Ferreira, 2012). This can include both intentionally generated data, such as responses to questions on a platform, and unintentionally generated data, such as facial expression, number of clicks, and seconds per page. Digital traces are one type of data exhaust, or data generated as a by-product of people's online engagements (Digital Element, 2018), the type of data most open to exploitation or resale by companies because users often don't even know they are generating it. This sort of "liquid surveillance" (Bauman and Lyon, 2013) or "surveillance capitalism" (Zuboff, 2019) have only recently begun to be challenged by, for example, the EU General Data Protection Regulation and the US Federal Trade Commission (FTC). The FTC recently announced it would closely monitor online education companies to ensure that children's rights to online privacy are not violated, important given that digital education is dominated by companies operating on corporate surveillance business models (FTC, 2022).

In education, the academic field known as Learning Analytics seeks to use education data to better understand the learning process and improve learning outcomes and environments (du Boulay et al., 2018), as well as to create new AI products for education. However, a critical question educators and education researchers must ask themselves is the same question currently being asked of corporates: To what extent can good intentions for outcomes justify the intentional violation of individual privacy? If we hold such high standards for ethical research, particularly with children, why should students and their parents not have the right to informed consent before this data is used for anything other than their individual report cards?

In higher education, the potential for exploitation is just as great. AI technologies have enabled everything from location-based tracking to engagement determinations based on facial expressions. During the 2020 Virtual Conference on AI in Education and Training (Shiohira and Keevy, 2020), one institution proudly proclaimed that they used AI to track everything from what meals students ate to how long they spent in their dorm rooms. The institution aimed to improve services, a laudable intention, but the methods eliminated any semblance of privacy students may have had. The idea of a "smart campus" in which all interactions are online, recorded, and used for decision-making seems as inevitable as it is disquieting (Stokel-Walker, 2020).

8.3.5 Proprietary Content and Transparency

A vast majority of AIED operating today is corporate or "social enterprise" owned. While these two types of organizations may be fundamentally different (the former prioritizing maximum profits for shareholders, the latter social change), they usually share one important common feature: Proprietary and protected intellectual property. This inevitably raises tensions between the modus operandi of the company and the need for transparency, including auditability or the ability of third parties to review, monitor and criticize algorithms, and explainability or the ability of people engaging with the algorithm to understand its determinations. As noted, transparency is tied to a fundamental human right, the right of individuals to contest decisions. As systems grapple with these sorts of fundamental questions different mechanisms are beginning to emerge. For example, China has outlawed proprietary and closed AI systems from operating in its schools (McMorrow et al., 2021).

8.4 Ethics by Design

The purpose of this chapter is to outline the current state of play of AI and Education, and to raise the ethical and humanistic challenges which at times are overlooked in the enthusiasm for a new approach. However, this does not mean we don't think the ethical and pedagogically sound application of AI to education is impossible. This section outlines some key recommendations to help ensure that going forward AI&ED is ethical and effective.

8.4.1 Don't Ignore "Preparing for AI"

Learning about the impact of AI in society and every day should receive proper weight and attention, rather than being considered an add-on to the technical skills required to develop AI. In fact, in K-12 the ethics and social impact of AI are allocated fewer hours than either the foundations of AI (data literacy, coding, algorithms) or developing AI (Miao and Shiohira, 2022). One way of addressing this is by ensuring that AI curricula and courses focus on both the human and technological dimensions of AI, intertwined throughout. For example, when learning about AI-assisted facial recognition students should be exposed both to how facial recognition systems work and what their potential impact is on society.

8.4.2 Innovate Around Data Privacy

The European Commission has identified AI systems that are used for student assessment, to deliver so-called personalized education, to perform evaluations, or to potentially impact cognitive or emotional development as "high risk", emphasizing the importance of privacy and data governance, transparency, human oversight, accuracy and security (European Commission, 2021). Even those tools that do not fit this description are often using data for purposes that may include comparison to stereotypes or average profiles (Chrysafiadi & Virvou, 2013), and anyone with a basic understanding of statistics (which certainly should include AI developers) can tell immediately how flawed and dangerous that approach can be. Even the best statistical model is open to some possibility of type 1 error (in layman's terms drawing the wrong conclusion from the evidence). The algorithms that underpin AI are also subject to these sorts of mathematical limitations, and this approach can lead to discrimination, particularly in underrepresented populations (Sapiezynski et al., 2019).

All is not lost, however, and ensuring that ethical principles are embedded by design, particularly transparency and human oversight might mitigate the risk. While society and governments seem to be a long way from developing the type of oversight agencies that could ensure the safety and reliability of AIED, researchers are beginning to pay attention (e.g., Miao et al. 2021, Holmes & Porayska-Pomsta, 2022) and there is at least one international commercially backed association, EdSafe AI,[17] invested in guiding institutions, individuals and governments how to implement accountable, fair and ethical AI in education.

The second good practice is to prioritize human oversight and promote teacher agency. AI must be a tool of the teacher and not the other way around. Teachers must be given opportunities to train in the use of AI technologies for teaching and learning, both in initial teacher education and in-service training, and should have ultimate decision-making authority on which AI is used and how it is integrated into their classrooms. Teachers are the fail-safe that will catch the errors AI makes.

A final point is that data must be collected and used ethically. In 2021, the OECD published a series of ethical principles for data in government (OECD, 2021a), but they universally apply. Paraphrased, they are:

- Act with integrity: Do not access, share, or use data for personal profit or goals that do not serve the public interest.
- Ensure trustworthy data access, sharing, and use.
- Incorporate data ethics into planning, funding, and contracts related to data and its management.
- Monitor and retain control over data inputs. Human oversight must always be preserved.
- Ensure that there is a legitimate reason for collecting and using data. being specific about the purpose.
- Place the needs of users (i.e., teachers and learners) at the center of data activities.
- Ensure data is representative and fit for purpose.
- Collect only the minimum amount of data necessary for the defined purpose.
- Define boundaries for data collection, access, sharing, and use.
- Be transparent about what data is collected, when and how it is collected, and for what purpose.
- Ensure data literacy among the (users) so that they may understand the implications of data use.
- Make data and source code open, to support socio-economic benefits, foster citizen engagement and ensure transparency, accountability, and public scrutiny of decisions and outcomes.
- Ensure no personally identifiable information is made public, but recognize that anonymized data can be de-anonymized.
- Broaden human control over their data, including the right to withdraw consent for its use.

This last point in particular links to the principles of self-sovereign identity, a growing movement toward ensuring that the creators of data are the controllers of that data. For example, advocates have recently challenged universities' control over the qualifications earned and paid for by individuals (Dale-Jones and Keevy, 2021), with some innovative solutions, such as DigiLocker in India, being developed (Molokwane & Shiohira, in press). The developers and managers of AIED tools should be equally concerned with the ownership of the data they are collecting and exploiting, and recognize that data does not, as some seem to believe, belong to the developer of the system that collected it, but rather to the individual who generated it.

8.4.3 Facilitate Robust Research

Teaching is fundamentally a research-based science, and when it comes to the use of AI in the classroom the fundamentals do not shift. If an AI application does not have *sufficient independent evidence of its efficacy*, ideally through independent research undertaken by qualified personnel using experimental or robust quasi-experimental designs, it should not be integrated into schools or other educational settings. Governments, businesses and social enterprises invested in AIED should make the additional investment to support independent research from trusted education research institutions. While a few have done so, mostly in the USA, most have not, and few robust large-scale studies are available (Bryant et al., 2020).

However, process and outcome evaluations are as necessary as impact evaluations. An AI product may be effective at, for example, improving learner outcomes in mathematics, but under what conditions, with what additional supports, and with what trade-offs? The unforeseen consequences of AI implementation need investigation as much as the learning outcomes do (Holmes et al., 2021). The obvious effects to be investigated center on students; for example, the long-term effects of learner classification on learners of different ability levels, or the effects of AI use on transversal skills such as communication or collaboration—but there are likely other effects on teacher quality, motivation, and engagement. AI tools may potentially affect even the system level, for example, funding decisions and allocation or curriculum design.

8.4.4 Create AI Programs That Support Innovative Pedagogies

Much has been written about "21st Century Skills", the definition of which shifts as rapidly as does technology. The "knowledge economy" is giving way to the "innovation economy", an age characterized by rapid innovation, shifting technologies, lifelong learning, and the need for meta-skills such as "learning to learn" and the application of skills in new contexts. It is unlikely that traditional rote learning as implemented in most current AIED applications (despite their claims to the contrary) can address these needs. While mathematics and language are a foundation for advanced skills such as communication and critical thinking, the role of memorization and drilling is limited in an education system built for the innovation economy. More robust pedagogies focus on socio-emotional learning, self-actualization, collaboration, and critical thinking.

While few current AI-assisted applications support these pedagogies, some individuals and organizations have looked at how AI might drive new pedagogical approaches better suited to life and work in the innovation age. Some existing AI curricula do attempt to leverage pedagogical theories such as constructivism (Piaget, 1972) and/or constructionism (Papert and Harel, 1991), computational thinking, and design thinking (Razzouk and Shute, 2012), mainly in the form of project-based learning. In some examples, students engage in identifying a social problem or market need and developing an AI-assisted solution, often through the construction of an app. AI concepts are introduced along the way. A few curricula engaging in these sorts of methods are government-led, but far more are industry-created, including offerings from IBM, Intel and Microsoft, or driven by the civil society sector, for example, the offerings of Technovation Girls[18] or the International Society for Technology in Education.[19]

Exploratory learning environments, in play-based learning, activity-based learning, and collaborative group work, offer another avenue for AI to be used to meaningfully enhance pedagogy. In these environments, AI might be used to help students achieve the intended learning outcomes, offering guidance and feedback based on open-ended activities. In this case, the AI essentially functions similarly to a non-player character in a video game, an adaptive agent that points the way toward challenges and offers support when the student is not progressing toward the activity goals. In this context, AI offers benefits such as immediate and adaptive feedback, while open-ended activities mean that the learning process as well as content outcomes are supported. The variability in the program or simulation can give learners AI-supported choices about what and how to engage, preserving more of their autonomy.

8.5 Concluding Comments

As this chapter has discussed, the connections between AI and education (AIED) are growing rapidly. On the one hand, we have the teaching of AI, what has been called "AI literacy", which in turn comprises its technological dimension (how it works and how to create it) and its human dimension (its impact on all people, society and the planet). We have argued that, in any serious attempt to teach about AI, the technological and human dimensions of AI should be given broadly equal emphasis and should be intertwined throughout, rather than a token discussion of ethics being tagged on at the end of a course. We have also noted that, so far, this rarely happens. In fact, exploring the human impact should always have been integral to the teaching of technology. However, with AI, we are talking about a technology with an impact that is more hidden, potentially more powerful and impactful on humans, and at a massively greater scale than anything before. Given the increasingly integral role that AI is playing in everyday life, we argue that the teaching of AI's human and technological dimensions to all citizens, young and old, in developed and developing countries alike, is now critical.

On the other hand, we have taught with AI-assisted applications, in order to support students, educators, and institutions (AIED). Again, as we have noted, many of the existing AI tools that have been developed for use in classrooms to support the teaching and learning of other subjects, raise important challenges that are only now beginning to be considered. For example, to date, few of the available applications have been independently or robustly evaluated, and those have only demonstrated their efficacy in limited contexts and for narrow learning outcomes. Second, most of the existing AI-assisted educational technologies perpetuate poor pedagogic practices, with a tendency to embody an outmoded behaviorist pedagogy of drill-and-practice, albeit sometimes with a few extras.

These applications of AI ultimately undermine both teacher and student agency, as choices are limited and machine-made decisions on learning pathways and content difficult to understand, let alone challenge. Third, the most easily available AIED tools have been developed by commercial organizations, which raises issues of both the commercialization of education and, at times, questionable business practices without clear consent or boundaries. Some business models even depend on capturing, analyzing, and exploiting or selling the data generated by thousands of students whose schools are paying for their access to the platform, usually without genuinely informed consent. Apart from the impact on data privacy, security and ownership, building education's dependence on commercial offerings can be risky. What happens when the company chooses to stop making a particular proprietary tool available?

We could continue to identify multiple other ways in which the application of AI in education should be considered carefully. However, to be clear, we are not arguing that AI has no place in education, but rather that we need to identify the *right kind* of AI and apply it in the *right way* (in particular, with an eye on human rights) if we are to leverage for the common good what AI technologies make possible. Robust debate over the content of AI curricula and the role of AI-assisted applications in classrooms is necessary. On the curriculum side, stakeholders should be consulted to ensure human, social and economic needs are all met, and that the technological and humanistic dimensions are equally valued. In classrooms, rather than starting from the technologies, we should start with the genuine education grand challenges, which educators are usually best placed to identify. Which of these is AI suited to address, and which are social problems that require social solutions? What exchanges and compromises will be made in the implementation of AI-assisted technologies?

Artificial Intelligence, currently in the guise of data-hungry machine learning, is the most powerful technology of our day. It has the potential to inform new approaches to education and new pedagogies, but can also perpetuate or even exacerbate current limitations. Unfortunately, most of the uses of AI in education today do the latter. Hopefully, this chapter will encourage more scholars and educators to consider both the potential and the challenges of AI, to think more seriously about how the benefits of AI can be safely, ethically, and effectively leveraged in classrooms, and ultimately to help ensure that AI&ED is implemented in ways that are genuinely and demonstrably beneficial for the whole of humanity.

Notes

1 For example, see IBM SkillsBuild at https://skillsbuild.org
2 See https://www.cedefop.europa.eu/en/tools/skills-online-vacancies
3 See https://psetcloud.org.za
4 See https://new.assistments.org

5 See https://byjus.com/global
6 See https://futureskillsprime.in
7 See https://www.oppimisanalytiikka.fi/ville
8 See https://www.vulcan-edu.com/en/home/
9 See https://www.compadre.org/osp
10 See https://www.graide.co.uk
11 See https://www.x5gon.org
12 Non-formal learning opportunities are structured learning programmes delivered by trained personnel that take place outside of the traditional classroom or the formal education system.
13 See https://www.pubaffairsbruxelles.eu/education-commission-launches-expert-group-to-develop-ethical-guidelines-on-artificial-intelligence-and-data-for-educators-eu-commission-press
14 For example, see "Can't Wait to Learn": https://www.warchildholland.org/news/artificial-intelligence-learning/
15 The "4Cs" terminology was coined by the Partnership for 21st Century Skills. https://www.battelleforkids.org/networks/p21
16 See https://www.letrus.com
17 See https://www.edsafeai.org
18 See https://technovationchallenge.org
19 See https://www.iste.org

References

Abdalla, M., & Abdalla, M. (2021). The Grey Hoodie Project: Big Tobacco, Big Tech, and the Threat on Academic Integrity. *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society July 2021*, 287–297. https://doi.org/10.1145/3461702.3462563

Bauman, Z., & Lyon, D. (2013). *Liquid surveillance: A conversation*. Polity Press. ISBN: 978-0-745-66282-4

Bryant, J., Heitz, C., Sanghvi, S., & Wagle, D. (2020). *Artificial intelligence in education: How will it impact K-12 teachers*. McKinsey Global Institute. https://www.mckinsey.com/industries/education/our-insights/how-artificial-intelligence-will-impact-k-12-teachers

Chrysafiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, *40*(11), 4715–4729.

Conchas, G. (2008). Cited in Teachers College Columbia University. Education as a Tool for Breaking the Cycles of Poverty. https://www.tc.columbia.edu/articles/2008/november/education-as-a-tool-for-breaking-the-cycles-of-poverty

Dale-Jones, B., & Keevy, J. (2021). Digital credentials: Discussions on fluency, data privacy and the recognition of learning in higher education beyond COVID-19. In M. Venter, and S. Hattingh (Eds.), *Learning for a better future: Perspectives on higher education, cities, business & civil society*. AOSIS Publishing.

Davies, H. C., Eynon, R., & Salveson, C. (2020). The mobilisation of AI in education: A bourdieusean field analysis. *Sociology*, *55*(3), 539–560. https://doi.org/10.1177/0038038520967888

De Jong, T. (1991). Learning and instruction with computer simulations. *Education and Computing*, *6*(3–4), 217–229.

Deakin University. (2018). Meet genie, the new smartphone app for deakin students. *Deakin Life*. https://blogs.deakin.edu.au/deakinlife/2018/07/13/meet-genie-the-new-smartphone-app-for-deakin-students/

Digital Element. (2018). *Digital data exhaust report 2018*. LBMA. https://www.digitalelement.com/wp-content/uploads/2018/10/Digital-Data-Exhaust-Survey-Report-2018.pdf

du Boulay, B., Poulovassilis, A., Holmes, W., & Mavrikis, M. (2018). What does the research say about how artificial intelligence and big data can close the achievement gap? In R. Luckin (Ed.), *Enhancing learning and teaching with technology* (pp. 256–285). Institute of Education Press.

European Commission. (2021). Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN

European Commission. Directorate General for Communications Networks, Content and Technology. & High Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. Publications Office. https://data.europa.eu/doi/10.2759/346720

Feathers, T. (2019). Flawed algorithms are grading millions of students' essays. *Vice*. https://www.vice.com/en_us/article/pa7dj9/flawed-algorithms-are-grading-millions-of-students-essays

Ferreira, J. (2012). Video: CEO of Knewton, an ITS company, talking at the Office of Ed Tech at the White House Education Datapalooza Event. https://www.youtube.com/watch?v=GeajedxpWJA

Fratamico, L., Conati, C., Kardan, S. et al. (2017). Applying a framework for student modeling in exploratory learning environments: Comparing data representation granularity to handle environment complexity. *International Journal of Artificial Intelligence in Education*, *27*, 320–352. https://doi.org/10.1007/s40593-016-0131-y

FTC. (2022). *FTC to Crack Down on Companies that Illegally Surveil Children Learning Online* [Press release]. https://www.ftc.gov/news-events/news/press-releases/2022/05/ftc-crack-down-companies-illegally-surveil-children-learning-online (Accessed 30 May 2022).

Hildebrand, D. (2018). Experience is not the whole story: The integral role of the situation in Dewey's democracy and education. *Journal of Philosophy of Education*, *52*(2), 287–300.

Hockly, N. (2019). Automated writing evaluation. *ELT Journal 73*, 82–88. https://doi.org/10.1093/elt/ccy044

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education. Promises and implications for teaching and learning*. Boston, MA: Center for Curriculum Redesign. https://drive.google.com/file/d/1lmzlbhKvYyRB6J0USCndqXitmVgsfTbI/view

Holmes, W., Persson, J., Chounta, I-A., Wasson, B. & Dimitrova, V., (2022). *Artificial intelligence and education. A critical view through the lens of human rights, democracy, and the rule of law*. Strasbourg, France: The Council of Europe.

Holmes, W. & Porayska-Pomsta, K. (2022). *The ethics of artificial intelligence in education. Practices, challenges, and debates*. New York: Routledge.

Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*, 389–399.

Johnstone, S., & Leasure, D. (2015). How competency-based education can fulfill the promise of educational technology. *Lecture Notes in Computer Science*, *9177*(3), 127–136. Springer, 2015.

Keevy, J., Shiohira, K., Matlala, R., & Molokwane, P. (2021). *New qualifications and competencies for future-oriented TVET: TVET delivery, Providing innovative solutions Volume 3*. UNEVOC.

Kolchenko, V. (2018). Can modern AI replace teachers? Not so fast! Artificial intelligence and adaptive learning: Personalized education in the AI age. *HAPS Educator*, *22*(3), 249–252. doi:https://doi.org/10.21692/haps.2018.032

Lameras, P., & Arnab, S. (2021). Power to the teachers: An exploratory review on artificial intelligence in education. *Information*, *13*(1), 14.

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, *19*(5), 780–794. https://doi.org/10.1177/1461444816686328

Major, L., Francis, G., &Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low- and middle income countries: A meta-analysis. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13116

Manyika, J., Lund, S., Chui, M., et.al. (2017). *Jobs lost, jobs gained: Workforce transitions in a time of automation*. McKinsey Global Institute.

Mavrikis, M., Rummel, N., Wiedmann, M., Loibl, K., & Holmes, W. (2022). *Combining exploratory learning with structured practice educational technologies to foster both conceptual and procedural fractions knowledge*. Educational Technology Research and Development. https://doi.org/10.1007/s11423-022-10104-0

McMorrow, R., Yu, S., Kinder, T., & Hale, T. (2021, July 26). China's education sector crackdown hits foreign investors. *Financial Times*. https://www.ft.com/content/dfae3282-e14e-4fea-aa5f-c2e914444fb8

Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: A guidance for policymakers*. UNESCO Publishing.

Miao, F., & Shiohira, K. (2022). *K-12 AI curricula. A mapping of government-endorsed AI curricula*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380602

Molokwane, P., & Shiohira, K. (in press). *The african continental qualification framework: Innovation and technology training module*. Addis Ababa: African Union.

OECD. (2021a). *Good practice principles for data governance in the public sector*. Paris: OECD. https://www.oecd.org/gov/digital-government/good-practice-principles-for-data-ethics-in-the-public-sector.pdf

OECD. (2021b). *OECD digital education outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots*. OECD. https://doi.org/10.1787/589b283f-en

Papert, S., & Harel, I. (1991). *Constructionism*. Ablex Publishing Corporation.

Piaget, J. (1972). *The principles of genetic epistemology*. Routledge & Kegan Paul.

Quinn, B. (2020). UK exams debacle: How did this year's results end up in chaos? *The Guardian*. http://www.theguardian.com/education/2020/aug/17/uk-exams-debacle-how-did-results-end-up-chaos

Razzouk, R., & Shute, V. (2012). What is design thinking and why is it important?. *Review of Educational Research*, *82*(3), 330–348.

Sapiezynski, P., Ghosh, A., Kaplan, L., Mislove, A., & Rieke, A. (2019). Algorithms that "Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. https://doi.org/10.1145/3514094.3534135

Seldon, A., & Abidoye, O. (2018). *The fourth education revolution: Will artificial intelligence liberate or infantilise humanity?* The University of Buckingham Press.

Shiohira, K. (2021). *Understanding the impact of artificial intelligence on skills development*. UNESCO-UNEVOC International Centre for Technical and Vocational Education and Training. https://unevoc.unesco.org/pub/understanding_the_impact_of_ai_on_skills_development.pdf

Shiohira, K., & Keevy, J. (2020). *Virtual conference on artificial intelligence in education and training: Virtual conference report*. UNESCO-UNEVOC TVeT Forum, 11 to 15 November 2019. https://unevoc.unesco.org/home/UNEVOC+Publications/lang=en/akt=detail/qs=6295

Stokel-Walker, C. (2020). Universities are using surveillance software to spy on students. *Wired UK*. https://www.wired.co.uk/article/university-covid-learning-student-monitoring

Tan, K., Chong, M. C., Subramaniam, P., & Wong, L. P. (2018). The effectiveness of outcome based education on the competencies of nursing students: A systematic review. *Nurse Education Today*, *64*, 180–189.

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNESCO. (2022). *UNESCO international literacy prizes 2021: Inclusive distance and digital literacy learning, trend analysis of nominated applications*. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000380601/PDF/380601eng.pdf.multi (Accessed 16 May 2022).

UNICEF. (2021). Policy guidance on AI for children. https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf.pdf

Watters, A. (2021). *Teaching machines*. Boston, MA: The MIT Press.

Williams, M. (2017). John Dewey in the 21st Century. *Journal of Inquiry & Action in Education*, *9*(1). 91–102. https://files.eric.ed.gov/fulltext/EJ1158258.pdf

Zhang, D., Maslej, N., Brynjolfsson, E., et al. (2022). *The AI index 2022 annual report*. Stanford, CA: The Stanford Institute for Human-Centered AI. https://hai-annual-report.stanford.edu

Zheng, L., Niu, J., Zhong, L., & Gyasi, J. (2021). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments*. Online. https://doi.org/10.1080/10494820.2021.2015693

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. London: Profile Books.