1. Introduction

Artificial Intelligence (AI) technologies have been researched in educational contexts for more than thirty years (Woolf 1988; Cumming and McDougall 2000; du Boulay 2016). More recently, commercial AI products have also entered the classroom. However, while many assume that Artificial Intelligence in Education (AIED) means students taught by robot teachers, the reality is more prosaic yet still has the potential to be transformative. This [chapter] introduces AIED, an approach that has so far received little mainstream attention, both as a set of technologies and as a field of inquiry. It discusses AIED's AI foundations, its use of models, its possible future, and the human context. It begins with some brief examples of AIED technologies.

The first example, *Cognitive Tutor*, is a type of AIED known as an *intelligent tutoring system* (ITS, which currently are probably the most common of AIED technologies). It addresses the domain of mathematics for students of primary or secondary school age, and aims to mirror a human tutor by delivering instruction personalised to each individual. *Cognitive Tutor* is also a rare case of an AIED technology that has bridged the gap from university research (at Carnegie Mellon University) to a successful commercial operation (Carnegie Learning[1]) and is also unusual in having robust independent evidence of its effectiveness (Pane et al. 2014). As individual students work through carefully structured mathematics tasks, the system monitors the student's progress (successes and misconceptions), re-phrases questions and re-directs the student along more suitable learning pathways, and provides individualised feedback (explaining not just why the student got something wrong but also how they can get it right). It achieves all this by combining individual student interaction data with the interaction data of the many thousands of students who have already experienced the system, using that data to learn, adapt and improve its models of mathematical skills and student learning.

A second quite different AIED example is *MASELTOV*[2], a research project in which AI was used to support language learning by recent migrants to the UK, using devices that many people carry with them all the time – smartphones (Gaved et al. 2014). The *MASELTOV* smartphone app used GPS data and AI techniques to provide context-sensitive and personalised language-learning support. For example, the app was able to detect when a user entered a doctor's surgery or a supermarket, each of which would trigger it to recommend appropriate English resources personalised to the individual's language skills. In the supermarket, the app would provide vocabulary and phrases to help the user find the items that they wanted to buy; in the doctor's surgery, it would provide appropriate words (such as symptoms, parts of the body and diagnoses) together with information about the available health services.

A final brief example comes from China. *Smart Learning Partner*[3], from Beijing Normal University's Advanced Innovation Center for Future Education, is a mobile app that enables students to connect with tutors using their smartphones. Students can use the app at any time of the day or night to search for a tutor, in order to ask them specific questions about any school topic for which they want some additional support. There are several thousands of tutors available on the app, thanks to local government funding, all of whom have been rated (much like a shopping app or a dating app) by users (in this case, the users are other students). The student chooses their tutor (based on the school topic and the tutor ratings), connects and is given thirty minutes of free one-to-one online tuition (sharing voice and screens but not video). Although the AI is relatively simple, *Smart Learning Partner* uses it to provide a unique student-centred system that enables students to get exactly the support that they want (rather than the instruction that a system such as an ITS might prescribe). Data from all the interactions are then aggregated and made available to the schools, so that trends in student questions can in a virtuous circle be identified and given more attention in the classroom.

2. The AI foundations of AIED

A full understanding of AIED depends on understanding something about AI more generally. The field of AI first emerged from a seminal workshop held at Dartmouth College in the US as long ago as 1956. Over the following decades, AI developed in fits and starts with periods of rapid progress interspaced with periods, known as *AI winters*, where confidence and funding all but evaporated. Most recently, over the past decade, with the advent of faster computer processors, the availability of large amounts of big data, and the development of new computational approaches, AI has entered a period of renaissance.

Nonetheless, what actually constitutes AI still is often disputed (as is the name itself, with some researchers preferring *augmented* rather than *artificial* intelligence). In fact, for many, as has been suggested earlier, AI is synonymous with humanoid robots, which might be because AI and robots seem to feature together in the news and on television almost every day. In fact, while robotics is a core area of AI research, AI is being used in many different and more down to earth ways and is growing exponentially (and the dystopian images of futuristic robots remain firmly in the realm of science fiction). Paradoxically, though, the more that AI is integrated into our daily lives, the less we think of it as AI:

> A lot of cutting edge AI has filtered into general applications, often without being called AI because once something becomes useful enough and common enough it's not labelled AI anymore.[4] (Nick Bostrom, Director of the Future of Humanity Institute, University of Oxford).

Instead, AI is often labelled as a computer program (such as email spam filtering), a mobile phone assistant (such as Siri), or perhaps an app (such as Duolingo).

Nonetheless, many recent developments in AI have been both ground-breaking and transformative. AI techniques such as machine learning, neural networks, evolutionary computation, and supervised, unsupervised and reinforcement learning, have been used in applications as diverse as autonomous vehicles, online shopping, auto-journalism, online dating, stocks and shares dealing, and legal and financial services.

Automatic face recognition, for example, is one area that has fairly recently made a dramatic leap forward while simultaneously becoming almost invisible in daily life (it is the technology used in smartphone cameras to ensure that faces are always in sharp focus and at e-passport gates to identify travellers before allowing them to enter a country). Face recognition was noticeably improved when, in 2012, Google presented a brain-inspired AI *neural network* comprising 16,000 computer processors with 10 million randomly selected YouTube video thumbnails. By using *deep-learning* techniques, and despite not being told how to recognise anything in particular, this *machine learning* system soon *learned* how to detect human faces in photographs. Two years later, Facebook introduced a nine-layer deep AI *neural network*, involving more than 120 million parameters, to identify (not just detect) faces in timeline photographs. It was trained on a dataset of four million images of faces that had previously been labelled by humans (Facebook users), and achieved an accuracy in excess of 97 per cent, which almost matches human-level performance. However, although impressive, these examples also highlight a key difference between AI and human intelligence: a human doesn't need to see ten, or even four, million faces before it can recognise a family member, a friend or a celebrity.

Another area that has seen much AI development is meteorological forecasting, with machine learning being shown to be more accurate at predicting weather than traditional simulation-based forecasting. Meteorologists have long tracked weather data which they enter into complex knowledge-based simulations to make forecasts. However, AI forecasting mines vast amounts of historical weather data, and uses *neural networks* and *deep learning* to identify data patterns (rather than to feed into simulations) in order to make data-based predictions about future weather conditions.

A final brief example is the use of AI in medical diagnosis, with AI techniques being used by radiologists to help them identify anomalies in medical images more quickly and while making fewer mistakes. For example, one system looks for irregularities in X-ray images and, depending on what it finds, assigns it a priority. If it finds nodules on an image of a pair of lungs, it assigns a high-priority status and sends it to a pulmonary radiologist for further checks.

One thing that all these examples demonstrate is that AI is a highly technical area, which is too complex to explore in depth here (two seminal books that do cover much of AI's complexity are Russell and Norvig 2016; and Domingos 2017). In fact, although AI is increasingly being offered as a 'service' (for example, Google's TensorFlow, IBM Watson and Microsoft's Azure), many people involved have advanced degrees in mathematics or physics. Nonetheless, because some have already been mentioned repeatedly and because they play an important role in AIED, some closely interlinked AI topics will be briefly introduced: algorithms, machine learning, deep learning, neural networks and Bayesian networks. The section then concludes with a brief mention of so-called *Artificial General Intelligence*.

Algorithms

AI often involves talk of *algorithms*, which are simply descriptions of the steps needed to solve problems (ordinary computer programs are really nothing more than lengthy algorithms). It is probably fair to say that Google owes its existence to a single algorithm, PageRank (Figure 1), which was developed in 1996 by the Google founders at Stanford University. PageRank (apparently named after the Google founder Larry Page rather than web pages) is an algorithm that ranked the relative importance of a website by counting the number and quality of external links to the website's pages, to determine where the website appeared in a Google search.

$$PageRank\ of\ site = \sum \frac{PageRank\ of\ inbound\ link}{Number\ of\ links\ on\ that\ page}$$

Figure 1. The PageRank algorithm that played a major role in the early years of Google.

In fact, the history of AI might be thought of as the history of the development of increasingly sophisticated and increasingly efficient (or elegant) algorithms; and what makes AI algorithms distinct from other types of algorithm is simply that they are applied to areas we might think of as essentially human (such as visual perception, speech recognition, decision-making and learning).

Machine learning

While most computer software (including some AI) involves writing in advance the exact steps that the software will take, or specifying rules that will be followed exactly, machine learning is about getting computers to act without being given explicit steps or rules. Instead of the algorithms being *programmed* what to do, they have the ability to *learn* what to do. Image and speech recognition, self-driving cars, computational biology (for example, using computers to identify tumours), and digital companions (such as Amazon's Alexa), as well as the Google DeepMind AlphaGo program that beat the world's number one player of Go, have all been made possible thanks to machine learning. In fact, machine learning is so widespread today (almost everyone has experienced some form of machine learning usually without being aware of it), that for some researchers and developers it has become synonymous with AI.

There are three headline approaches to machine learning: supervised, unsupervised and reinforcement. In **supervised learning**, the AI is first trained with data for which the output is already known. For example, the AI might be trained with many thousands of photographs of people that have already been labelled by humans (this is broadly speaking the approach, mentioned earlier, used by Facebook to *identify* people in photographs). The AI can then be used to label automatically new data (in this example, to identify and label automatically the same Facebook users in new photographs). In **unsupervised learning**, on the other hand, the program is provided with even larger amounts of unlabelled data, which it uses to find patterns that enable it to classify new data (this is broadly the approach, mentioned earlier, used by Google to *detect* faces in photographs). Finally, in **reinforcement learning** the program is provided with some initial data from which it derives an outcome that is assessed as correct or incorrect, and rewarded accordingly (for example in an AI-driven computer game, the score is increased) or punished (the score is reduced). The program uses this to update itself and then it tries again, thus developing iteratively (evolving) over time.

Deep learning

An extension of machine learning is known as *deep learning*. This involves automatic iterative analysis that clusters and classifies data and makes predictions. For example, once a deep learning algorithm determines that a picture contains a particular shape, it cycles again to find other shapes, and then cycles again to identify the connections between those shapes, iterating repeatedly until it has recognised what it is looking at (for example, a face). Deep learning is the headline approach used by AlphaGo, to learn how to win at the game of Go.

Neural networks

Machine learning often uses *neural networks*, so named because they are inspired by how neurons work and are connected in animal brains. However, although AI neural networks have been trained to do some incredible things, they are primitive in comparison to most higher-order animal brains. They usually involve only a few thousand *neurons* (in some exceptional cases, a few million) compared to the human brain, which has around 100 billion neurons and trillions of connections. In any case, AI neural networks comprise several layers of neurons (Figure 2): typically an input layer (that takes stimuli from the environment), one or more hidden computational layers, and an output layer (that delivers the result of the computation). All the neurons are interconnected, with each connection having a weighting to determine whether one neuron excites or inhibits the next neuron (again in a process inspired by synapses in animal brains). During the machine learning process, it is these weightings that are adjusted, usually by reinforcement learning, and that allow the AI subsequently to compute outputs for new stimuli. Neural networks have been shown to be particularly effective in many different AI systems, for example for image recognition (identifying people) and natural language processing.

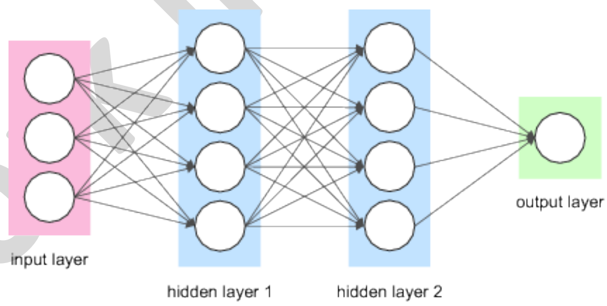

Figure 2. A representation of a typical simple neural network.

The hidden layers are key to the power of neural networks but they also bring an important problem. It isn't possible (or at the very least it isn't easy) to interrogate a neural network to find out how it came up with its solution (for example, how did it identify a particular person?). In other words, neural networks can lead to decision making for which the rationalisation is hidden and unknowable, and possibly unjust (O'Neil 2017).

Bayesian networks

Bayesian networks are a type of statistical model employed by some AI algorithms that enable, in uncertain domains, computational tasks such as prediction, anomaly detection and diagnostics. They combine principles from graph theory, probability theory and statistics. Drawn graphically, a Bayes net comprises various lines (also known as edges) which intersect at nodes, with the nodes representing variables and the lines representing interdependencies between those variables.

To give a simple example, using a Bayes net approach, an AI system might be designed to predict (calculate the probability of) the flavour of ice-cream that a customer might buy depending on the weather and temperature of the day. Here, the nodes represent the known data (whether it is sunny, whether it is hot, and choices of ice cream flavour made by previous customers) and an uncertain outcome (what ice cream flavour will be chosen). The Bayes net computation begins with probabilities given in each node that have been derived from training data (comprising records of weather, temperature and customers' choice of ice-cream flavours) to derive the probabilities of various outcomes (the ice-cream flavours that will be chosen by customers in a combination of weather and temperature circumstances). In fact, a typical AI Bayesian network might comprise tens (or hundreds) of variables (nodes) with intricate interdependencies. However, the Bayesian computational approach makes it possible to infer precise probabilities in such complex environments in order to inform usable predictions (to continue with the example, to help the ice-cream seller decide how much of each ice-cream flavour to make).

Artificial General Intelligence

All the examples of AI mentioned so far are domain-specific, which means that they are tightly constrained and very limited. For example, the AI used to win at Go cannot play a game of chess, the AI used to predict the weather cannot predict movements in the stock market, and the AI used to drive a car cannot be used to fly an aeroplane. So-called Artificial General AI, AI that like human intelligence can be used in any circumstances, does not yet exist. And, despite the rapid developments in AI and the concerns expressed by many leading scientists (Hawking et al., 2014), it is unlikely to exist for decades (even for leading AI advocates, General AI appears to be due to arrive at some ever-receding future date, usually around thirty years from the time of writing, Müller and Bostrom 2016). In fact, currently, rather than general applications (AI that can be used in any context, Domingos 2017), the focus for most AI research continues to be on domain-specific areas – such as autonomous vehicles, health, weather forecasting and stocks trading, and education.

3. Introducing AI in education

AI in education research (AIED) has considered a variety of ways in which AI systems might be used to support both formal and informal learning. It has involved the development of many online tools that aim to support learning while being flexible, inclusive, personalised, engaging and effective (Holmes et al. 2018). AIED brings together AI and the learning sciences, and thus involves two main complementary strands: developing AI-based tools to support learning, and using these tools to help understand learning (how learning happens).

> In addition to being the engine behind much "smart" ed tech, AIED is also a powerful tool to open up what is sometimes called the "black box of learning", giving us deeper, and more fine-grained understandings of how learning actually happens. (Luckin et al., 2016, p. 18)

In other words, AIED research can have an important impact both on classroom tools (such as Cognitive Tutor) and on learning theories applicable in classrooms where there is no AI. For example, by modelling how students go about solving an arithmetic problem and, for example, identifying misconceptions that might have been previously unknown to educators, researchers and teachers can begin to understand much more about the process of learning itself which can then be applied to classroom practices.

AIED models

AIED often involves computational *models* (in AI, a model is a highly simplified computational representation of something in the real world, just like a model car is a simplified representation of a real car). In particular, *intelligent tutoring systems* (ITS such as Cognitive Tutor) are often built around three core models: *pedagogy*, *domain* and *learner*, all of which interact in complex ways and are combined to adapt a sequence of learning activities for each individual student (Figure 3). A fourth AIED model is the *open learner* model.

The AIED **pedagogy model** represents knowledge about effective teaching and learning approaches that have been elicited from teaching experts (and that constitute the learning sciences). This includes, for example, knowledge of instructional approaches (Bereiter and Scardamalia 1989), productive failure (Kapur 2008), guided discovery learning (Bruner 1961), collaborative learning (Dillenbourg 1999), the zone of proximal development (Vygotsky 1978), deliberate practice (Ericsson et al. 1993), interleaved practice (Rohrer and Taylor 2007), cognitive overload (Mayer and Moreno 2003), formative feedback (Shute 2008), uncertain rewards (Fiorillo 2003), and assessment for learning (Black 1986).

The AIED **domain model**, on the other hand, represents knowledge about the subject that the system aims to help the students learn. This might, for example, be knowledge about mathematical procedures, genetic inheritance or the causes of World War I. In fact, over the years, mathematics for primary and secondary school students has dominated AIED (mathematics, along with physics and computer science, are AIED's low-hanging fruits because they are, at least at school and undergraduate level, well-structured and clearly delineated), although recent AIED research has investigated AI to support learning in less well-defined areas (such as essay writing across the humanities, Landauer et al. 2009; Whitelock et al. 2015). Finally, the AIED **learner model** represents knowledge about the students (for example, about student interactions, achievements, challenges, misconceptions, responses and emotional states while using the system), both for all the students who have used the system so far and for the individual student using the system right now.

Figure 3 shows how these three models might be connected in a typical AIED intelligent tutoring system.

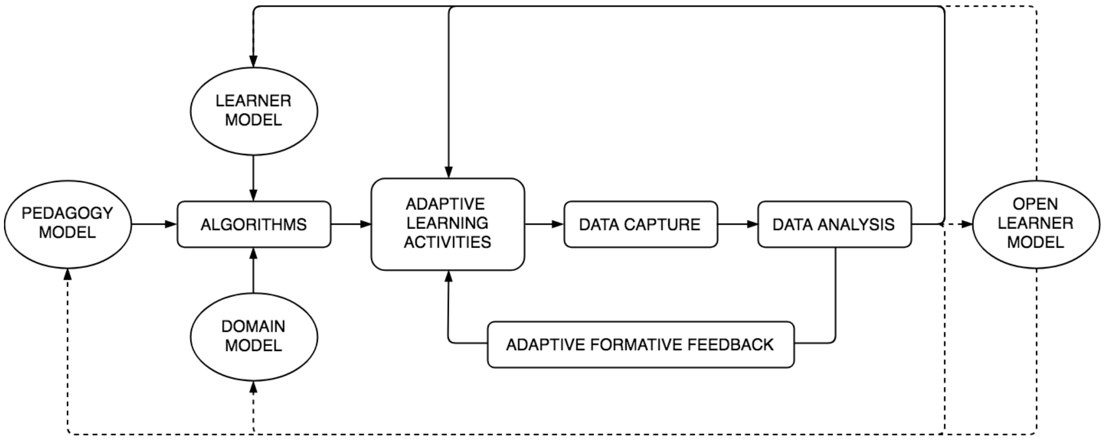

Figure 3. Flowchart representing a typical AIED intelligent tutoring system architecture, including the pedagogy, domain, learner and open learner models.

In this exemplar architecture, algorithms draw on the pedagogy, domain and learner models to determine what specific learning activity (for example, some textual content or a collaborative learning activity) should be presented to the individual student and how it should be adapted to that student's needs and capabilities (over time, this means that individual students might experience their own unique personalised learning pathways). Then, while the student engages with this adaptive learning activity, the system automatically captures thousands of data points representing each individual interaction, the student's achievements and any misconceptions that they have demonstrated. Some systems also capture other data such as the student's speech and an indication of their affective (emotional) state.

All of this data is then analysed (possibly using machine learning or Bayesian network techniques), both to provide the student with individualised formative feedback (to support their learning according to their individual needs) and to update the learner model (to inform the system's next adaptive learning experience). The analysis might also, in some circumstances, update the pedagogy model (with those approaches to pedagogy used by the system that have been shown to support student learning most effectively) and domain models (perhaps with previously unknown but apparently not uncommon misconceptions).

Some AIED ITS also feature a fourth model, the open learner model shown in Figure 3 (Dimitrova et al. 2007). Open learner models aim to make visible (explicit), for the learners and teachers to inspect, both the teaching and learning that has taken place and the decisions that have been taken by the system (which is especially important if the system uses a neural network approach, where it can be otherwise difficult to decipher how a decision has been made). This enables learners to monitor their achievements and personal challenges, supporting their metacognition, and enables teachers to better understand each individual learner's learning (their approach, any misconceptions and their learning trajectories) in the context of the whole class.

AIED and Learning Analytics

AIED is sometimes linked to another developing field of research in education known as Learning Analytics (LA) or Educational Data Mining (EDM). LA, to focus on just one of these approaches, involves *"the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs"* (Siemens 2011). It applies statistical techniques from *big data* research (Mayer-Schonberger and Cukier 2013) to digital traces in educational contexts. In many ways, there are clear overlaps between LA and AIED (Figure 4 and Figure 5). In both LA and AIED, student interaction and outcomes data are analysed, and the results may be shown in visualisations (for example, in student dashboards). However, although the distinction is becoming increasingly blurred, while LA typically uses the data and analysis to provide insights to inform *human intervention* (by, for example, teachers), AIED uses the data and analysis to initiate some kind of *automatic intervention* (such as personalised feedback or learning pathways for students, or automatic student forum post aggregation for teachers).

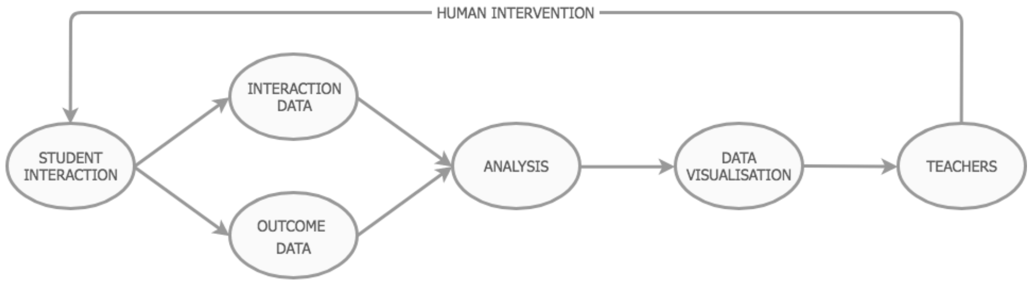

Figure 4. A simplified overview of Learning Analytics

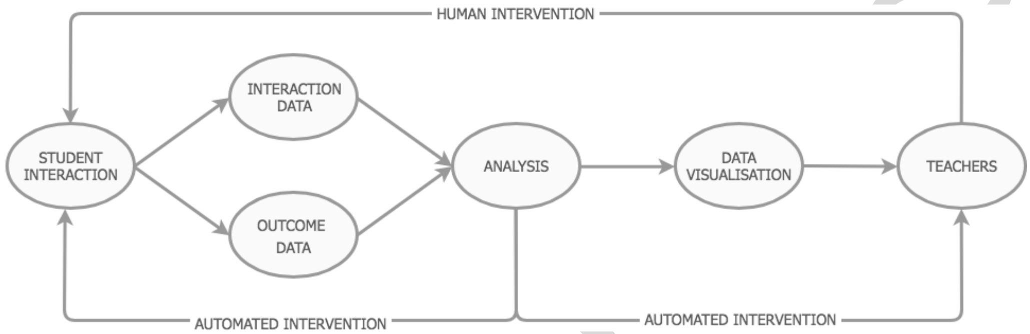

Figure 5. A simplified overview of AIED

4. AIED applications

There are many AIED-driven applications being used in schools and universities. Here, building on the examples mentioned earlier, an illustrative sample are surveyed.

As mentioned earlier, the most common types of AIED are so-called *intelligent tutoring systems* (ITS), with Cognitive Tutor being a leading example (for a comprehensive history and discussion of ITS see Woolf 2008). ITS aim to simulate one-to-one (personal) human tutoring, which has long been thought to be the optimum condition for learning, although it is typically costly (at least in terms of teacher time) and so beyond the reach of most students. Famously, Benjamin Bloom (1984) calculated that students receiving personal tuition could achieve outcomes that were 2 standard deviations (2-sigma) above students taught in conventional classrooms. Although the accuracy of this has recently been challenged (VanLehn 2011), the aim of many ITS researchers has been to devise systems that answer the "*'2-sigma problem'. Can researchers and teachers devise teaching-learning conditions* *that will enable the majority of students … to attain levels of achievement that can at present be reached only under good tutoring conditions?*" (Bloom 1984). In fact, VanLehn calculates that the correct figure for human tutoring is closer to 0.8 sigma and that many ITS are already almost as effective (VanLehn 2011).

Three influential examples of personal tutors are *AutoTutor, Andes* and *CIRCSIM*, each of which has been shown to achieve at least 1.0 sigma improvement over conventional classroom teaching. ***AutoTutor*** was an online system that aimed to *"simulate the dialogue patterns of typical human tutors"* in the domain of computing (Graesser et al. 2001). The system's pedagogy model adopted the principle that it is important *"to encourage students to articulate lengthier answers that exhibit deep reasoning rather than deliver short snippets of shallow knowledge"* (ibid.), which it addressed by engaging students in a series of written exchanges and by prompting them to elaborate. Meanwhile, feedback mechanisms included providing hints, extending student responses, and correcting misunderstandings. ***Andes***, on the other hand, was an ITS focusing on the domain of physics that aimed to replace students' pencil and paper homework with an interactive and intelligent interface. The system presented students with physics problems for them to solve, each of which usually consisted of many steps (such as drawing vectors, drawing coordinate systems, defining variables and writing equations). After the student completed each step, the system gave feedback, such as hints on what was wrong with an incorrect step or what kind of step to try next. Finally, ***CIRCSIM*** was a language-based ITS for first-year medical students, which was designed to help them learn about the reflex control of blood pressure. It involved one-to-one interactions between the student and the computer, using natural language processing and generation, adopting a pedagogy model that assumes *"real understanding of something involves, at least in part, an ability to describe the basic concepts in appropriate language"* (Evens and Michael 2006). Accordingly, students were asked to solve small problems while engaging (similarly to *AutoTutor*) in a Socratic dialog (an iterative conversation of questions and responses) with the computer.

A more recent AIED example is *OpenEssayist* (Whitelock et al. 2013), which uses Natural Language Processing to provide automated meaningful feedback on draft essays. Unlike earlier AIED systems that were developed to *grade* essays and to *instruct* students how to fix problems (such as Criterion, Burstein and Marcu 2003; Summary Street, Franzke and Streeter 2006; and IntelliMetric, Rudner et al. 2006), *OpenEssayist* encourages the user to reflect on the content of their essay in order to promote self-regulated learning, self-knowledge, and metacognition. It uses linguistic technologies, graphics, animations, and interactive exercises to enable users to reflect on whether the essay adequately conveys the intended meaning, and to self-correct before submitting their essay for summative assessment. The system was based on the assumption that the quality and position of key phrases in an essay illustrate how complete and well-structured the essay is, which it determined by means of key phrase extraction, identifying which short phrases are the most suggestive of an essay's content, and extract summarisation.

Another example is *iTalk2Learn*, an AIED system for children aged 8-12 years old who are learning fractions, which was designed to detect, analyse and respond to speech in real time in order to improve learning (Rummel et al. 2016). Specifically, the platform supported the robust learning of fractions by providing activities to help develop both *conceptual* and *procedural* knowledge of fractions. Conceptual knowledge is fostered in an exploratory learning environment called *Fractions Lab*, that facilitates students to answer given fractions tasks using virtual manipulatives (graphical representations of fractions) in any way that they choose. Procedural knowledge, on the other hand, is fostered by structured practice activities, in a commercial more linear ITS called *Maths Whizz*. A student's unique sequence of interleaved exploratory and structured practice activities is determined by an overarching intervention model (Mazziotti et al. 2015), the aim being to achieve optimum conditions for learning (avoiding students being under- or over-challenged, which may trigger either boredom or anxiety). Sequencing decisions are made according to the student's level of *challenge* and their *affective state*, both of which are inferred from the student's interaction (what they click and the actions they take on the screen) and their speech (including key words and prosodic features such as 'um's and pauses) and all of which are recorded in the student model. Throughout, the system uses a Bayesian network approach to deliver targeted formative feedback at three levels: *Socratic*, *guidance*, and *didactic* (Holmes et al. 2015). *Socratic* feedback draws on the dialogic approach to teaching (Alexander 2010), which emphasises the benefits of open questioning to encourage students to consider and verbalise possible solutions. The second level, *guidance*, reminds students of key domain-specific rules and the system's affordances. The third level, *didactic* specifies the next step that needs to be undertaken in order to move forward (this rarely-delivered final feedback also operates as a backstop, ensuring that the student is not left floundering).

Another use of AI in education is to focus on supporting teachers to support students, rather than on supporting the students directly. One example of this is the Virtual Teaching Assistant known as Jill Watson (JW), developed at Georgia Tech to address difficulties in providing automatic online assistance for large cohorts of students, particularly in online courses (Goel and Polepeddi 2017). JW was designed to monitor the online forum of a computer science course, to recognise common questions raised by the students, and to provide answers both accurately and quickly. Rather than replacing the human teaching assistants, JW aimed to relieve them of having to respond to low-level questions (such as enquiries about length of assignments, dates for submission, and required readings), which can be both time-consuming and tedious, to allow them to focus on higher-level and thus typically more interesting questions and other teaching activities. JW was originally developed using the IBM Watson AI as a service platform and broadly adopted a supervised learning approach. It was trained with two connected datasets developed over three semesters: the questions that students had asked, mapped to (labelled with) the answers that the human teaching assistants had provided. Thus trained, the system evaluates new student questions to determine if they can be mapped to question/answer dyads for which the system has confidence (because similar questions have been posed and answered many times). The appropriate answer is then selected and immediately returned to the student. On the other hand, if an appropriate answer cannot be identified with confidence, the question is referred up to a human teaching assistant without introducing any noticeable delay.

Finally, some brief examples involving AIED and two quite different learning approaches: collaborative learning and virtual reality. Research (e.g., Dillenbourg 1999) has shown that collaborative learning, which might involve two or more students undertaking a project together, can be more effective than learning alone. Collaborative learning can, for example, encourage students to articulate their thinking, to resolve differences through constructive dialogue, and to build shared knowledge. However, other research (e.g., Slavin 2010) suggests that collaboration between learners rarely happens without appropriate support. For this reason, various approaches using AI to support collaborative learning have been researched.

AI-driven adaptive group formation, for example, uses knowledge about the participants, most often in learner models, and self-learning algorithms to form a group best suited to a particular collaborative task (perhaps students are all at a similar cognitive level and have similar interests, or they bring different but complementary knowledge and skills) (Mujkanovic et al. 2012). Meanwhile, expert facilitation can involve training systems to support students collaboratively sharing knowledge. For example, Soller and colleagues (2002) developed a system using Hidden Markov Modelling (another probabilistic technique used in AI) to identify effective and ineffective knowledge sharing between students, so that intelligent guidance might be provided to foster more productive knowledge exchange. Finally, intelligent virtual agents might mediate online student interaction, or simply contribute to the dialogues by acting as a coach, a virtual peer or a teachable agent (i.e. a virtual peer that the participants might themselves teach). For example, Goodman and colleagues (2005) developed an agent that interacted with the participants when it detected something happening that was interfering with the learning (such as a student's confusion about a problem, or a participant who is dominating the discussion or not interacting with other participants).

Virtual reality (VR) and augmented reality (AR) are emerging applications of AI that are both being promoted as having potential for learning. VR can provide authentic experiences that, using VR headsets, headphones and controllers, simulate in immersive 4D (the three dimensions of space plus sound or haptics) a small part of the real world to which the user would not otherwise have access. These include places such as dangerous environments (like the interior of a volcano) or somewhere geographically or historically inaccessible (such as a black hole or the Cretaceous Period). However, while some (e.g., Hassani et al. 2013) have suggested that learning in virtual realities can enable the student to better transfer that learning to the real world (transfer of learning has long been known to be a problem), and there are examples of VR being used to support medical training (e.g., Ruthenbeck and Reynolds 2015), in a review of VR in K-12 education, Freina and Ott (2015) were unable to find any robust learning outcomes.

Augmented Reality adopts a different approach. Instead of providing an alternative reality, AR overlays rich media (virtual objects such as text, still images, video clips, 3D models and animations) onto live video images of the existing reality, by means of the cameras and screens on smartphones and tablet devices, in such a way that users perceive the virtual objects as if they are coexisting with the real-world environment. There are many examples. AR techniques can be used to show textual information about a specific mountain (such as its name and maximum elevation) when a smartphone's camera is pointed at it[5]; while another AR app has been developed, for use in a university science course, that allows the user to view and interact with an anatomically correct 3D model of a human heart[6]. Nevertheless, despite the promise, again there is currently little evidence that AR leads to any notable learning gains (Bower et al. 2014; Radu 2014).

5. The future of AIED

As is clear from both the media and this brief review, AI and AIED are rapidly developing areas of research and development. In particular, AIED applications that yesterday seemed fanciful, today are being widely used by students, independently or in schools and universities. Future possibilities are limited only by the imagination and are thus difficult to predict. Here, therefore, briefly surveying four areas in which AIED has substantial potential (building on Luckin et al. 2016) will have to suffice.

21st-century skills

There is an increasing recognition that what have been called ‘21st-century skills’ (a range of skills, abilities and approaches to learning) are essential for current and future work environments (World Economic Forum and The Boston Consulting Group, 2016). Such skills include learning and innovation skills (e.g. critical thinking), digital literacy skills (e.g. information literacy) and career and life skills (e.g. initiative and self-direction). AIED can help by developing reliable and valid indicators to enable researchers to track learner progress in these 21st-century skills (including the less tangible skills like creativity and curiosity). In addition, AIED could help researchers to understand the most effective teaching approaches to support those skills.

For example, a problem-based collaborative learning experience might be monitored using AI and a combination of sources, including voice recognition (to identify who is doing and saying what in a team activity) and eye tracking (to explore which learner is focusing on which learning resources at any particular moment in time). Importantly, researchers might also consider the learning context by building context models into the AIED system. The context models might help them to identify how the combinations of technology, teachers and the environment might be adjusted to improve teaching (Luckin, 2010).

Assessment

AIED techniques have the potential to replace the ‘stop and test’ approach to assessments, the usual approach to examinations that only assesses a fraction of what has been learned by the student while at the same time causing stress to many students. Instead, AIED techniques involving learning analytics could provide continuous formative or just-in-time information about learner successes, challenges and needs that can then be used to shape the learning experience itself (see Foltz, 2014). For example, AIED will enable learning analytics to identify changes in learner confidence and motivation while learning a foreign language, say, or a tricky equation. AIED-driven assessments could also be built into meaningful learning activities, perhaps a game or a collaborative project, and will assess all of the learning that takes place, as it happens.

New insights into learning

The data gleaned from digital teaching and learning experiences will yield new insights that cannot easily be ascertained in other ways. For example, as well as identifying whether or not a learner gave the correct answer, datasets could be analysed to help teachers understand how the learner arrived at their answer. The data might also help us to better understand cognitive processes, such as remembering and forgetting, and the fundamental impact that these have on learning and student outcomes. AIED analysis might also identify if and when a student is confused, bored or frustrated, which will help teachers understand and enhance a learner's emotional readiness for learning – and possibly amend their teaching practice as a result.

Lifelong learning partners

Finally, AIED has the potential to build artificial 'learning companions' that might accompany and support individual learners throughout their studies – in and beyond school. These lifelong learning companions could be based in the cloud, accessible via mobile devices, and be designed to support teachers, removing some of the drudgery of teaching while allowing teachers to focus on the more human aspects of learning.

The learning companion might also provide students with an easy-to-access record of their learning experiences, and suggestions or ongoing guidance for future study. They might also call in specialist AIED systems, or humans with expertise in particular subject areas, as needed by the individual learner.

For further details about these examples of AIED, you might like to read 'The next phase of AIED' in Intelligence Unleashed: An Argument for AI in Education (Luckin et al., 2016, p. 32). Please note, this is just a suggestion; you are not required to read this.

6. The Ethics of AIED

No discussion of AI in education can be complete without some consideration of the ethical implications[7]. Yet, while the range of AI technologies being introduced in schools and universities around the world are extensive and growing, the ethics are rarely investigated. There has been work around the ethics of AI in general (e.g., Bostrom and Yudkowsky 2014) and around the ethics of Learning Analytics (e.g., Slade and Prinsloo 2013). However, at the time of writing, around the world, virtually no research has been undertaken, no guidelines have been provided, no policies have been developed, and no regulations have been enacted to address the specific ethical issues raised by AIED. In short, researchers in AIED are proceeding without any fully worked out ethical groundings, and it might be argued that all AIED technologies, including those that have been introduced in this [chapter], exist in a moral vacuum.

In fact, although not as newsworthy as robots or self-driving cars, the use of artificial intelligence techniques (such as neural networks, machine learning and Bayes nets) in education has profound implications for students (their skills, knowledge and developing minds) and thus for wider society. In parallel, this also raises an indeterminate number of as yet unanswered ethical questions. To begin with, concerns exist about the large volumes of data collected to support AIED (such as the recording of student competencies, emotions, strategies, and misconceptions). Who owns and who is able to access this data, what are the privacy concerns, how should the data be analysed, interpreted and shared, and who should be considered responsible if something goes wrong? For these questions, AIED might usefully draw on the work that investigates the ethics of Learning Analytics. However, while data raises major ethical concerns for the field of AIED, AIED ethics cannot be reduced to questions about data. Other major ethical concerns include the potential for bias (conscious or unconscious) incorporated into AIED algorithms and impacting negatively on the civil rights of individual students (in terms of gender, age, race, social status, income inequality…). For these questions, AIED might usefully draw on the work that investigates the ethics of AI in general. But the AIED ethical concerns centred on data and bias are the ‘known unknowns’. What about the ‘unknown unknowns', the ethical issues raised by and specific to the field of AIED that have yet to be even identified?

One approach is to consider ethical issues in terms of the three main AIED models introduced earlier. At the pedagogical level, the impact of AIED on pedagogical relationships and how best they can be supported first needs to be addressed. For example, what kinds of pedagogical interventions are ethically warranted, what kinds of information should be used to justify an intervention, and what kinds of behavioural changes is AIED intended to bring about? At the domain level, it is important to consider how the adaptation of particular subject content amenable to AIED influences the learner experience and their understanding of that content. Finally, at the level of individual learners, issues centre on the use of personal information. In addition to the use learning analytics to profile learners, these include issues around surveillance and covert data collection (involving cutting edge technologies that are poised to collect ever more personal information), and the tension between paternalistic systems and the autonomy of the learner.

Specific AIED ethical questions include: What are the criteria for ethically acceptable AIED? How does the transient nature of student goals, interests and emotions impact on the ethics of AIED? What are the AIED ethical obligations of private organisations (developers of AIED products) and public authorities (schools and universities involved in AIED research)? How might schools, students and teachers opt out from, or challenge, how they are represented in large datasets? And, what are the ethical implications of not being able to easily interrogate how AIED deep decisions (using multi-level neural networks) are made?

Strategies are also needed for risk amelioration since AI algorithms are vulnerable to hacking and manipulation. Where AIED interventions target behavioural change (such as by 'nudging' individuals towards a particular course of action), the entire sequence of AIED enhanced pedagogical activity also needs to be ethically warranted. And finally, it is important to recognise another perspective on AIED ethical questions: in each instance, the ethical cost of inaction and failure to innovate must be balanced against the potential for AIED innovation to result in real benefits for learners, educators and educational institutions.

REFERENCES

Alexander R (2010) Dialogic teaching essentials. Singapore: National Institute of Education

Bereiter C, Scardamalia M (1989) Intentional learning as a goal of instruction. Knowing Learn Instr Essays Honor Robert Glas 361–392

Black H (1986) Assessment for learning. Assess Educ Achiev 7–18

Bloom BS (1984) The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. Educ Res 13:4–16

Bostrom N, Yudkowsky E (2014) The ethics of artificial intelligence. Camb Handb Artif Intell 316–334

Bower M, Howe C, McCredie N, et al (2014) Augmented Reality in education–cases, places and potentials. Educ Media Int 51:1–15

Bruner JS (1961) The Act of Discovery. Harv Educ Rev 31:21–32

Burstein J, Marcu D (2003) A machine learning approach for identification thesis and conclusion statements in student essays. Comput Humanit 37:455–467

Cumming G, McDougall A (2000) Mainstreaming AIED into education? Int J Artif Intell Educ IJAIED 11:197–207

Dillenbourg P (1999) What do you mean by collaborative learning? Collab-Learn Cogn Comput Approaches 1–19

Dimitrova V, Mccalla G, Bull S (2007) Preface:" Open learner models: Future research directions" Special Issue of the IJAIED (Part 2). Int J Artif Intell Educ

Domingos P (2017) The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World, 01 edition. Penguin, London

du Boulay B (2016) Artificial Intelligence as an Effective Classroom Assistant. IEEE Intell Syst 31:76–81. doi: 10.1109/MIS.2016.93

Ericsson KA, Krampe RT, Tesch-Römer C (1993) The role of deliberate practice in the acquisition of expert performance. Psychol Rev 100:363

Evens M, Michael J (2006) One-on-One Tutoring by Humans and Computers. Psychology Press

Fiorillo CD (2003) Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. Science 299:1898–1902. doi: 10.1126/science.1077349

Franzke M, Streeter LA (2006) Building student summarization, writing and reading comprehension skills with guided practice and automated feedback. Highlights Res Univ Colo White Pap Pearson Knowl Technol

Freina L, Ott M (2015) A literature review on immersive virtual reality in education: state of the art and perspectives. In: The International Scientific Conference eLearning and Software for Education. " Carol I" National Defence University, p 133

Gaved M, Luley P, Efremidis S, et al (2014) Challenges in context-aware mobile language learning: The MASELTOV approach. In: International Conference on Mobile and Contextual Learning. Springer, pp 351–364

Goel AK, Polepeddi L (2017) Jill Watson: A Virtual Teaching Assistant for Online Education. 21

Goodman BA, Linton FN, Gaimari RD, et al (2005) Using Dialogue Features to Predict Trouble During Collaborative Learning. User Model User-Adapt Interact 15:85–134. doi: 10.1007/s11257-004-5269-x

Graesser AC, VanLehn K, Rosé CP, et al (2001) Intelligent tutoring systems with conversational dialogue. AI Mag 22:39

Hassani K, Nahvi A, Ahmadi A (2013) Design and implementation of an intelligent virtual environment for improving speaking and listening skills. Interact Learn Environ 1–20. doi: 10.1080/10494820.2013.846265

Holmes W, Anastopoulou S, Mavrikis M, Heike (2018) Technology-enhanced Personalised Learning. Untangling the Evidence. Robert Bosch Stftung, Stuttgart

Holmes W, Mavrikis M, Hansen A, Grawemeyer B (2015) Purpose and Level of Feedback in an Exploratory Learning Environment for Fractions. In: Conati C, Heffernan N, Mitrovic A, Verdejo MF (eds) Artificial Intelligence in Education. Springer International Publishing, Cham, pp 620–623

Kapur M (2008) Productive Failure. Cogn Instr 26:379–424. doi: 10.1080/07370000802212669

Landauer TK, Lochbaum KE, Dooley S (2009) A New Formative Assessment Technology for Reading and Writing. Theory Pract 48:44–52. doi: 10.1080/00405840802577593

Luckin R, Holmes W, Griffiths M, Forcier LB (2016) Intelligence Unleashed. An argument for AI in Education. Pearson, London

Mayer RE, Moreno R (2003) Nine ways to reduce cognitive load in multimedia learning. Educ Psychol 38:43–52

Mayer-Schonberger V, Cukier K (2013) Big Data: A Revolution That Will Transform How We Live, Work and Think. John Murray, London

Mazziotti C, Holmes W, Wiedmann M, et al (2015) Robust Student Knowledge: Adapting to Individual Student Needs as They Explore the Concepts and Practice the Procedures of Fractions (workshop paper). Madrid

Mujkanovic A, Lowe D, Willey K, Guetl C (2012) Unsupervised learning algorithm for adaptive group formation: Collaborative learning support in remotely accessible laboratories. In: International Conference on Information Society (i-Society 2012). pp 50–57

Müller VC, Bostrom N (2016) Future progress in artificial intelligence: A survey of expert opinion. In: Fundamental issues of artificial intelligence. Springer, pp 553–570

O'Neil C (2017) Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, 01 edition. Penguin, London

Pane JF, Griffin BA, McCaffrey DF, Karam R (2014) Effectiveness of Cognitive Tutor Algebra I at Scale. Educ Eval Policy Anal 36:127–144. doi: 10.3102/0162373713507480

Radu I (2014) Augmented reality in education: a meta-review and cross-media analysis. Pers Ubiquitous Comput 18:1533–1543

Rohrer D, Taylor K (2007) The shuffling of mathematics problems improves learning. Instr Sci 35:481–498

Rudner LM, Garcia V, Welch C (2006) An Evaluation of the IntelliMetricSM Essay Scoring System. 22

Rummel N, Mavrikis M, Wiedmann M, et al (2016) Combining Exploratory Learning with Structured Practice to Foster Conceptual and Procedural Fractions Knowledge. In: ICLS. Singapore

Russell S, Norvig P (2016) Artificial Intelligence: A Modern Approach, 3 edition. Pearson, Boston etc

Ruthenbeck GS, Reynolds KJ (2015) Virtual reality for medical training: the state-of-the-art. J Simul 9:16–26. doi: 10.1057/jos.2014.14

Shute VJ (2008) Focus on Formative Feedback. Rev Educ Res 78:153–189. doi: 10.3102/0034654307313795

Siemens G (2011) 1st International Conference on Learning Analytics and Knowledge 2011 | Connecting the technical, pedagogical, and social dimensions of learning analytics. https://tekri.athabascau.ca/analytics/about. Accessed 29 Oct 2017

Slade S, Prinsloo P (2013) Learning analytics ethical issues and dilemmas. Am Behav Sci 57:1510–1529

Slavin RE (2010) Co-operative Learning. What Makes Group-Work Work? In: Dumont H, Istance D, Benavides F (eds) The Nature of Learning. Using Research to Inspire Practice. OECD, pp 161–178

Soller A, Wiebe J, Lesgold A (2002) A machine learning approach to assessing knowledge sharing during collaborative learning activities. Association for Computational Linguistics, p 128

VanLehn K (2011) The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educ Psychol 46:197–221. doi: 10.1080/00461520.2011.611369

Vygotsky LS (1978) Mind in Society: Development of Higher Psychological Processes. Harvard University Press, Cambridge, MA

Whitelock D, Field D, Pulman S, et al (2013) OpenEssayist: An automated feedback system that supports university students as they write summative essays

Whitelock D, Twiner A, Richardson JT, et al (2015) OpenEssayist: a supply and demand learning analytics tool for drafting academic essays. In: Proceedings of the Fifth International Conference on Learning Analytics And Knowledge. ACM, pp 208–212

Woolf B (1988) Intelligent tutoring systems: A survey. In: Exploring artificial intelligence. Elsevier, pp 1–43

Woolf BP (2008) Building Intelligent Interactive Tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann, Amsterdam ; Boston

[1] http://www.carnegielearning.com

[2] http://www.open.ac.uk/iet/main/research-innovation/research-projects/maseltov

[3] http://slp.bnu.edu.cn (NB Only available to students with accounts)

[4] http://edition.cnn.com/2006/TECH/science/07/24/ai.bostrom/index.html

[5] https://www.peakfinder.org/mobile

[6] https://appstore.open.ac.uk/humanheart

[7] The author is grateful to Dr Rob Farrow for his significant contribution to an earlier document that informed this section.