AECT ASSOCIATION FOR EDUCATIONAL COMMUNICATIONS & TECHNOLOGY

Check for updates

Machine learning for human learners: opportunities, issues, tensions and threats

Mary E. Webb[1] · Andrew Fluck[2] · Johannes Magenheim[3] · Joyce Malyn-Smith[4] · Juliet Waters[5] · Michelle Deschênes[6] · Jason Zagami[7]

Abstract

Machine learning systems are infiltrating our lives and are beginning to become important in our education systems. This article, developed from a synthesis and analysis of previous research, examines the implications of recent developments in machine learning for human learners and learning. In this article we first compare deep learning in computers and humans to examine their similarities and differences. Deep learning is identified as a sub-set of machine learning, which is itself a component of artificial intelligence. Deep learning often depends on backwards propagation in weighted neural networks, so is non-deterministic—the system adapts and changes through practical experience or training. This adaptive behaviour predicates the need for explainability and accountability in such systems. Accountability is the reverse of explainability. Explainability flows through the system from inputs to output (decision) whereas accountability flows backwards, from a decision to the person taking responsibility for it. Both explainability and accountability should be incorporated in machine learning system design from the outset to meet social, ethical and legislative requirements. For students to be able to understand the nature of the systems that may be supporting their own learning as well as to act as responsible citizens in contemplating the ethical issues that machine learning raises, they need to understand key aspects of machine learning systems and have opportunities to adapt and create such systems. Therefore, some changes are needed to school curricula. The article concludes with recommendations about machine learning for teachers, students, policymakers, developers and researchers.

✉ Mary E. Webb
mary.webb@kcl.ac.uk

Extended author information available on the last page of the article

Introduction

Many people will have heard of machine learning (ML) through examples like self-driving cars, online recommendations from Amazon or Netflix, voice controlled digital assistants on mobile phones and spam filters. More broadly, applications of machine learning are widespread and increasing across most areas of human endeavour including agriculture (Liakos et al. 2018), the energy industry (Cheng and Yu 2019), e-commerce (Zhang et al. 2018), fault detection and diagnosis across most types of machinery (Zhao et al. 2019) and healthcare (Faust et al. 2018). Likewise, in education, machine learning is becoming more widespread and has been used for improving curriculum design (Ball et al. 2019), predicting students' grades (Livieris et al. 2019), recommending higher education courses to students (Obeid et al. 2018); and student modelling for intelligent tutoring systems (Conati et al. 2018).

Recent developments in machine learning can enhance or transform learning, and this possibility has implications for what learners and teachers need to understand about using machine learning systems in education. This paper was born from a meeting of experts in Quebec, Canada in October 2019. The hosting organisation, EDUsummIT, is a global community of researchers, policymakers and practitioners committed to supporting the effective integration of Information Technology (IT) in education by promoting active dissemination and use of research. The EDUsummIT thematic working group for this paper comprised experts in information technology education, computer science, assessment, STEM career development, and a Chief Knowledge Officer. The aim of the meeting was to review recent developments in machine learning, analyse opportunities and issues and to examine implications for education. Some previous analyses have looked more broadly at artificial intelligence in relation to particular aspects of education rather than focusing on the implications of recent developers in machine learning specifically. For example, Touretzky et al. (2019a, b) focused on reviewing and identifying the concepts that students should learn about artificial intelligence while Knox et al. (2019) examined how artificial intelligence could support inclusive education. A report for the European Union aimed at policymakers (Tuomi 2018) identified a number of broad issues, tensions and policy recommendations which are largely consistent with the more specific outcomes from our analysis. Focusing on recent rapid developments, a report by Hao (2020) for the MIT review examined the technologies and philosophy behind the major push by Chinese companies into artificial intelligence for learning, which is so far predominantly supporting home-based tutoring but has very broad ambitions.

The EDUsummIT group used an expert panel approach similar to that described by Galliers and Huang (2012). More specifically the process involved pre-meeting collaborative writing and intensive literature review leading to a discussion paper and structure for the meeting. During the meeting, discussions within the group and with other members of EDUsummIT enabled critical reflection to the point of theoretical saturation (Low 2019). In this way, the meeting was able to synthesize and analyse previous research to produce a report (Webb et al. 2019). Our analysis revealed that two of the major issues for machine learning in education, as well as in other fields, are explainability and accountability. As will be explored in the current article, new methods of machine learning, often called deep learning, may incorporate too much complexity for the algorithms, models and reasoning processes to be accessible and explainable to users. In order to contextualise our examination of how machine learning might contribute to human learning, we start by examining the nature of deep learning in humans that leads to understanding and the kinds

of learning experiences that foster such learning. We start by examining the term "deep learning" as it applies to how humans learn. Next, we examine how the definition of "deep learning" is quite different when applied to machine learning. Comparing and contrasting these two definitions will clarify the differences and lay the ground for discovering how the deep learning of ML might be used to reinforce deep learning in humans. We then discuss opportunities provided by current machine learning capabilities by analysing recent applications in terms of the kinds of human learning they support as well as the issues and risks associated with deploying such applications in learning contexts. Finally, we discuss policy, practice and research recommendations for deploying machine learning capability in educational contexts.

Human learning moving towards deep learning

In characterising human learning (in contrast to machine learning) the term "deep learning" has been important for some years especially in higher education (Entwistle 2005; Howie and Bagnall 2013; Marton and Säljö 1976; Webb and Ifenthaler 2018). Marton and Säljö's (1976) work focused on qualitative differences between outcomes of learning which they attributed to different levels of processing: surface processing and deep processing. Surface-level processing refers to memory recall forms of learning, while deep processing produces longer-lasting retention by, for example, focusing on the meaning of a text rather than memorisation. Since this earlier research, many researchers and practitioners in higher education have worked to promote the "deep learning" that they believe leads to a more comprehensive understanding of the subject matter (Entwistle 2005; Fullan et al. 2017; Howie and Bagnall 2013). However, Howie and Bagnall (2013) have argued that the models of deep-surface learning are underdeveloped and there is a gap in the theorisation of the underlying structure and meaning of the models. Furthermore, the terminology lacks clarity and precision. Nevertheless, many studies of students' approaches to learning, particularly in higher education, have made use of a deep learning versus surface learning model (see for example Asikainen and Gijbels 2017 for a review). Furthermore, the idea that deep learning is beneficial is prevalent in higher education where many lecturers aspire to enable deep learning in their students (Asikainen and Gijbels 2017). Fullan et al (2017) have argued for a change across compulsory education towards deep learning in order to address learning needs of all students at all levels in the twenty-first-century. Their conceptualisation of deep learning is of a process that involves higher order cognitive processing to reach a deep understanding of content and issues. Furthermore, they characterise this kind of learning as challenging, often cross-disciplinary, active, collaborative, student-centred and personally relevant and incorporating the use of digital technologies and connectivity. Fullan et al.'s (2019) characterisation of deep learning involves transforming learning so that students are engaged in creative, meaningful activities focused on real-world problems. However, research looking for examples of deep learning in American schools found very few (Mehta and Fine 2019).

Links have been made between pedagogy embedded with formative assessment techniques and deep learning because such pedagogy can achieve the engagement, student autonomous learning and self-regulated learning that enables the development of understanding (Shepard 2019). Over recent years the importance and nature of formative assessment has been examined and has become an accepted element of classroom practice in many countries. More recently the importance of integrating formative assessment into

pedagogy has been emphasised (Black and Wiliam 2018). Black and Wiliam (ibid.), in examining pedagogy emphasising formative assessment, also identified some of the challenges for teachers in designing activities, enabling peer collaboration, making use of dialogue and providing appropriate feedback that leads to learning. Thus, they have provided some explanation of why deep learning may be scarce in schools. For formative assessment to be effective in supporting learning, whether mediated by humans or computers, the feedback needs to be part of a process of communication and interaction in which cognitive and affective factors are important and the students come to understand what they have achieved and what are the next steps for learning (Webb and Ifenthaler 2018).

Deep learning then, in educational terms, is arguably based on models that are as yet ill-defined. While earlier explanations of deep learning were limited predominantly to achieving good understanding, more recent explanations focus on defining the learning processes that can achieve engagement and lead to understanding in learners. Such learning is being characterised as a transformational process that is rarely achieved. While this transformational approach to deep learning will be recognised by many teachers as a pedagogical approach that they would support and aspire to, we have identified reasons why teachers may find this approach difficult to implement. A vision for the future of assessment to support learning outlined by Webb and Ifenthaler (2018) is for technology to support teachers and students working together to understand their learning needs, move their learning forward and develop evidence of their achievements. The deep understanding that we argue is possible through such a pedagogical approach is dependent on meaningful feedback that addresses both cognitive and affective needs of students. We now turn to deep learning in machines which, as we will discuss, is very different from human learning and is based on statistical reasoning processes that identify patterns and draw conclusions from massive datasets.

Understanding machine learning

Definitions

There is consensus that machine learning is a subset of artificial intelligence, and that deep learning is a subset of machine learning (see Fig. 1).

The origin of the term 'artificial intelligence' is attributed to a conference at Dartmouth College (USA) in 1956 and refers to studies where computers behave like humans. Following more than 50 years of research, and abundant articles, artificial intelligence is still not clearly defined and there are diverse views on its potential and risks (Kaplan and Haenlein 2019). An early ubiquitous definition of machine learning which is often quoted and emphasises outcomes is:

> Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively the next time. (Simon 1983, p. 28).

But as Wang and Tao (2008) explain, such a definition is inadequate for computer scientists who are focusing on designing algorithms and analysing problems that can be solved by machine learning. In education also, we need more functional definitions that characterise the machine learning processes as well as the outcomes. While, just like in human cognition, perceptual capabilities and access to data are also necessary for artificial intelligence,

Fig. 1 Relationships between key terms
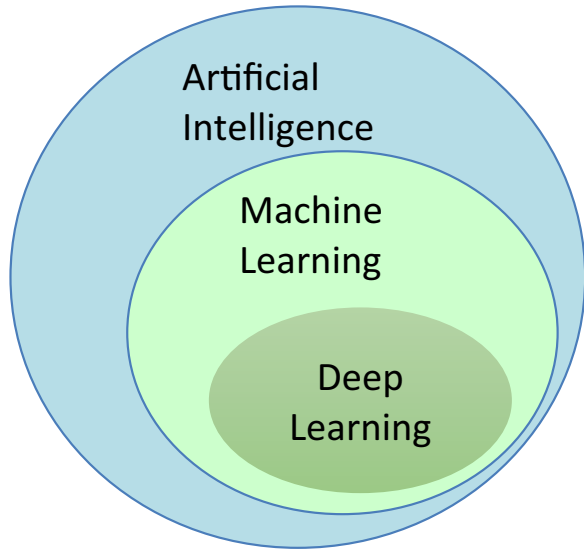
it is the machine learning processes that determine not only the nature of the outcomes and judgements that are made by the system but also our access to how such judgements were made. Wang and Tao's definition emphasises the practicalities of implementing machine learning: developing a model that is true to the real-world problem being solved and generating a representative dataset that can be used for training the model and using algorithms with statistical reliability:

> the process (algorithm) of estimating a model that's true to the real-world problem with a certain probability from a data set (or sample) generated by finite observations in a noisy environment. (Wang and Tao 2008, p. 49).

Thus, the nature of the machine learning in any particular system depends not only on the algorithms with which it has been originally programmed and the architecture specified, but also on the design decisions of the original engineers in terms of the values of learning rate parameters, the initial training regime, the choice of dataset, the context in which it is learning and subsequent upgrades to the system (Rahwan et al. 2019). The training regime refers to the way in which the machine is trained using a dataset selected to be representative of the overall dataset. Typically, data will be supplied to a machine learning system in batches and as the machine evaluates each batch of data it generates an error value for the difference between the existing model and the model generated by the new data. The learning rate parameter controls the proportion of the error value with which the model is updated at each iteration. Thus, the learning rate parameter controls the rate of machine learning: a faster rate would be expected to be less accurate.

Types of machine learning

Machine learning can be classified according to the inputs that it learns from as:

1. supervised learning where both training data and correct answers are supplied;
2. unsupervised learning where machines learn from a dataset on their own;

3. semi-supervised learning where the training set has some missing data and the algorithms are still able to learn from the incomplete data; and
4. reinforcement learning based on feedback from the environment.

A major focus for current research and development in machine learning is 'deep learning' (DL) or 'deep neural networks' (DNN). LeCun et al. (2015) provide a useful characterisation of deep learning that primarily uses neural networks:

> Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction….. Deep learning discovers intricate structure in large data sets by using the back-propagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer (LeCun et al. 2015, p. 436).

Deep learning is now widely used in speech recognition and computer vision and is expected to make a large contribution in many other fields in the near future (Sze et al. 2017). Back-propagation (see Fig. 2) is a process that takes place during the training phase of a neural network. As each element in a data set is processed through the neural network, the resulting prediction is compared with the actual known target value. Once the difference between the prediction and actual value is determined, the weights (strength) of links in the neural network are adjusted in a "backwards" direction, between adjacent layers, to minimize error between the prediction and target value. The complexity of multi-layered neural networks as well as the probabilistic nature of the models means that typically deep learning operates as a "black box" such that the basis of outcome decisions is not accessible, so decisions may have limited or no explainability. Using deep learning, significant progress has been made for handling bimodal data and dealing with very large datasets, but major challenges remain for dealing with high velocity data, multimodal data sources
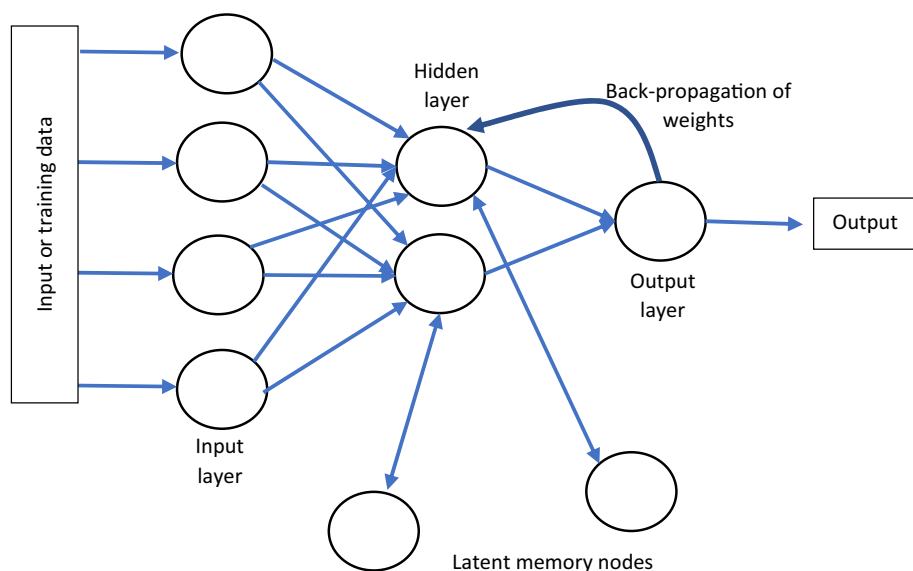

Fig. 2 Back-propagation & memory in machine learning

(Baltrušaitis et al. 2019) and low-quality datasets. Furthermore, whereas humans are very capable of coping with the typically continuous streams of data that we receive and can integrate new knowledge into existing knowledge, this lifelong learning remains very challenging for the deep learning algorithms currently available (Parisi et al. 2019). Typically, models are "trained" with static datasets and incorporating new data requires retraining which often results in catastrophic forgetting or catastrophic interference with existing knowledge (Parisi et al. 2019).

Issues in machine learning

As explained earlier, a machine making decisions and/or predictions may operate as a "black box" because deep learning algorithms and models can be very complex and inscrutable thus inhibiting traceability of reasoning processes. In many situations there is a need for transparency of the reasoning processes as well as the data used so that decisions and conclusions made by machines can be explained. This transparency is essential to minimize bias and ensure that decision making based on machine learning is fair, interpretable and accessible for all. There could be considerable legal obstacles to its adoption if the operational characteristics of a machine learning system cannot be explained. The European General Data Protection Regulation applies nearly worldwide, since it has implications for all European trading partners through its extraterritorial applicability. There is debate about the implied 'right to explanation' of algorithmic decisions in the EU General Data protection Regulation (GDPR):

> The controller shall … provide the data subject with the following further information necessary to ensure fair and transparent processing: … f) the existence of automated decision-making … and meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject. (EU GDPR 2018, Article 13).

Harm caused by algorithmic activity can be hard to detect and find its cause. Furthermore, it is rarely straightforward to trace who should be held responsible for any such harm caused owing to the multiple actors involved in the design and development of the system. Creators of machine learning systems/models should be held accountable for any issues of bias and transparency. Key issues for machine learning in education are therefore explainability and accountability.

Explainability in machine learning

Explainability is basically the ability to understand and explain 'in human terms' what is happening within the machine learning model; how exactly it works under the hood (Itransition 2019). The problem is not unique to education: predictions in medicine also present a dilemma in which complex deep learning predictions are becoming very accurate, for example, for determining cancer risk but prioritising accuracy over interpretability and transparency has been severely criticised (Hayashi 2019). Interpretability of machine learning systems is currently a strong focus for research (Carvalho et al. 2019; Doshi-Velez and Kim 2017). Doshi-Velez and Kim (2017) argue that not all systems need to be interpretable, for example aircraft collision avoidance systems function without human intervention.

More specifically, explanations are not necessary where: (1) there are no severe consequences for incorrect results or (2) when the problem is very well studied and validated in real applications so that the system's decisions can be trusted (ibid.). However, systems where fundamentally there is incompleteness in problem formalisation which therefore results in unquantified bias, require interpretability so that the gaps in problem formalisation are accessible to humans (ibid.).

Fundamentally, learning and assessment of learning are problems that can never be completely formalised, therefore arguably some level of interpretability and explainability is necessary for machine learning systems used in education unless there are no significant consequences for incorrect results. As we discuss later machine learning is used for a range of different types of human learning situations. Arguably, the use of machine learning in any system that supports student learning will have important consequences for those students but obviously some systems such as those used for high stakes assessment may have life changing consequences.

There is much research on ways of developing rule-based systems to explain black box models but Rudin (2019) argues that, for high stakes decision-making, such systems are high-risk as they are very prone to inaccuracy. Therefore, Rudin (2019) argues for developing machine learning that is inherently interpretable. Initial guidance for industry also emphasises the importance of providing explainability in their artificial intelligence systems for engendering trust and empowering people to understand and evaluate such systems and contribute to the debate surrounding their use, and also for legal compliance. This guidance identifies six main types of explanation (Itransition 2019; ICO and Turing 2019, p. 19) as follows.

- Rationale explanation: the reasons that led to a decision, delivered in an accessible and non-technical way.
- Responsibility explanation: who is involved in the development, management and implementation of an artificial intelligence system, and who to contact for a human review of a decision.
- Data explanation: what data has been used in a particular decision and how; what data has been used to train and test the artificial intelligence model and how.
- Fairness explanation: steps taken across the design and implementation of an artificial intelligence system to ensure that the decisions it supports are generally unbiased and fair, and whether or not an individual has been treated equitably.
- Safety and performance explanation: steps taken across the design and implementation of an artificial intelligence system to maximise the accuracy, reliability, security and robustness of its decisions and behaviours.
- Impact explanation: the impact that the use of an artificial intelligence system and its decisions has or may have on an individual, and on wider society.

What is apparent however, is that there are no agreed metrics for the quality of explanation methods (Carvalho et al. 2019) and comparisons are difficult because explanations can vary in format so much across different contexts. For example, Holzinger's (2018, p. 61) explanation is presented as a heatmap of molecule properties with a dendrogram for groups and colour coding for molecular values. It is hard to conceive a metric which would compare this visualisation with the percentage likelihood of identification output from Machine Learning for Kids. Conati et al. (2018) describe how example tracing evaluates students' problem-solving steps against typical examples of correct problem-solving steps, represented by behaviour graphs. This Bayesian Knowledge Tracing is reflected back to

AECT

Fig. 3 TARDIS video recordings with synchronised user-inspected data (Porayska-Pomsta and Rajendran 2019, p. 53)

the student as a series of 'skill bars' to help them identify areas where more learning is required. As is clear from these diverse examples, explainability is context specific so what is needed is a set of approaches to explanations together with a general framework that can identify an appropriate approach for any particular system taking into account the type of domain, use case and type of user (Carvalho et al. 2019).

Accountability in machine learning

Accountability is required when the outcome of a machine learning system is challenged. As the system is designed, trained and applied, different potential fault sources can be identified. Much current debate focusses on the reasonable representativeness of training sets. Unfair outcomes can emerge if the population sample used for training is insufficient or biased for wider application. Faults can also arise at other stages, particularly if the source data fed into the system is erroneous (in the 'apply') stage. Two ontological perspectives on accountability are prevalent. The first relates to the post-factum accountability, involving a blameable agent. The second is the normative one, with society assessing machine learning systems performance against norms of fairness, justice etc. (Porayska-Pomsta and Rajendran 2019).

The key mitigating differences between artificial intelligence and human decision-making are that human decisions involve individual flexibility, context-relevant judgements, empathy, as well as complex moral judgements, missing from artificial intelligence. In some cases, the machine learning system must operate very quickly (autonomous driving, aircraft control systems) and immediate accountability is impractical. However, post-facto accountability is crucial, since it can be used to diagnose a crash and improve future systems. In other cases, for example in the TARDIS project (Porayska-Pomsta and Rajendran 2019; Porayska-Pomsta and Chryssafidou 2018) (Fig. 3), the designers successfully used the Open Learner Model approach. This was facilitated through an off-the-shelf Microsoft Kinect and a high-quality microphone, to provide young people at risk of exclusion from education, employment or training with insight into their social interaction skills in job

interview settings. TARDIS used machine learning based agents (shown on the computer screen) acting as job recruiters. Data was gathered on the quality of young people's specific verbal and non-verbal behaviours (e.g. length of answer to specific interview questions, facial expressions, quality of gestures, posture, and voice respectively) as they interacted with the machine learning agent. For accountability, users had access to interactive time-lines of their interview simulations, including precise information on all the actions that they and the machine learning agent performed, moment by moment. A comparison study showed improvement in the quality of participant interview answers, verbal and nonverbal behaviours.

Joshua Kroll provides a three-layer model of accountability for machine learning. The bottom recording layer keeps track of what the system does and how it has been changed over time. The middle analytical layer analyses these records and matches them with output performance. The top responsibility layer determines who made which change and the consequence enhancement or degradation of performance. This has to be done whilst protecting commercially confidential algorithms or trained model weights. By applying these principles to the USA entry visa lottery, he was able to show the accountability circuit would require a computational overhead of only .12% (Kroll 2012, p. 186). As we have seen, although we don't yet have all the answers, research is beginning to address these issues of explainability and accountability in some contexts. Next, we will examine learning opportunities enabled by machine learning focusing on both the potential for supporting deep human learning and how important explainability and accountability are for each type of application. More specifically, key questions are: how important is explainability for a particular system? For example, if using a system produces better learning, does it matter if the system acts like a black box so that the student simply follows the system's instructions/advice and cannot question the approach? Similarly, if the teacher has no way of examining or questioning the system's approach and is thus disempowered and unable to manage the students' learning, does that matter if the students' learning improves? With regard to accountability in education, currently teachers are held accountable for the learning achievements of their students. If the students are learning from machine learning systems who is accountable? With respect to assessment, currently it is usual for the assessment criteria, marking schemes, marked scripts etc. to be accessible to scrutiny, so they can be checked by examiners. For high stakes assessments such arrangements are essential for accountability. However, high stakes assessments are notoriously unreliable, so if machine learning systems could produce more accurate and reliable assessments, would the importance of accuracy be preferred over explainability and accountability?

Learning opportunities enabled by machine learning

First, we will identify the range of types of applications of machine learning systems that may be used in education and then we will examine specific examples in more depth in order to analyse the learning opportunities provided and whether and how the issues of explainability and accountability are addressed. Machine learning systems can: adapt the learning process and resources to provide personalised learning perhaps with assistive tutors; recommend courses or resources to students based on students' characteristics; recommend groupings to teachers or students (Rajagopal et al. 2017); predict students' grades or behaviours (Han et al. 2011; Livieris et al. 2019) and interact with students to provide feedback. Furthermore, some more standard online learning

AECT

environments may have machine learning elements for specific purposes, for example, to measure learners' engagement (Dewan et al. 2019). Machine learning systems don't necessarily need to act alone: one promising approach is to combine machine learning with crowdsourcing by utilising the different strengths of each approach to create a hybrid tool, for example for analysing student interactions in online systems especially Massive Open Online Courses (MOOCs) or for detecting cheating in online exams (Alenezi and Faisal 2020). Kloos et al. (2019) discussed three main educational purposes of machine learning for supporting human learning in more complex contexts: (1) mixed realities; (2) multimodal interaction and (3) mixed social networks. Mixed realities enable student interaction involving immersion in a virtual reality as well as the contextualisation of the real world while machine learning supports personalisation based on students' data. Multimodal interaction and in particular the use of voice assistants that use machine learning for natural language processing enable a more natural way of interacting with computer systems. In mixed social networks especially, for example, in MOOCs, machine learning systems can analyse interactions between large numbers of students and identify patterns in social interaction and students' behaviours. Most of these examples are still being developed and researched with a view to identifying the most appropriate machine learning methods and their degree of accuracy (Kloos et al. 2019; Nájera and de la Calleja Mora 2017) so currently the specific types of machine learning algorithms in use are still under consideration.

Examples of how deep learning systems are already becoming important in educational contexts include assistive tutors, such as Amira and Duolingo that aspire to increase student progress through individualized support and instantaneous feedback. Amira is a reading assistant (chat bot), for K-3 students that listens, assesses and coaches to accelerate reading mastery using deep learning to determine its interventions. Duolingo is an application for learning a foreign language that adapts to the user's capabilities, using users' data and deep learning, to predict, for example, whether the user will remember a word. Data from 300 million users also enables the Duolingo system to use deep learning to discover new insights about the nature of language and learning. Both Amira and Duolingo have been developed over some years with the addition of increasingly sophisticated machine learning capabilities that are not detailed in the literature. These assistive tutors are used predominantly for informal learning and they have had some success for aspects of language learning that require frequent practice (García Botero et al. 2019) i.e. relatively shallow human learning rather than deep human learning. Nevertheless, for some aspects of language learning e.g. developing vocabulary, these relatively shallow skill-based approaches are considered useful (García Botero et al. 2019). Enabling students to be self-motivated to use such systems remains a challenge (ibid.). Duolingo has implemented a basic tracking system for teachers to review students' progress through the tasks, but the machine learning that it employs acts as a black box. Arguably, while such systems are used only in informal contexts or as extensions to formal classroom work, incorrect results have no severe consequences so their lack of explainability it is not a significant concern (Doshi-Velez and Kim 2017).

Another system has been described by Bosch et al. (2016) which used computer vision, learning analytics, and machine learning to detect students' affect (emotional state) in the real-world environment of a school computer lab that contained as many as thirty students at a time. The system could identify student boredom, confusion, delight, engagement, and frustration in natural environments up to 98% of the time. This example of measuring and characterising affective factors shows that machine learning components can go further

than the delivery of content and assist with a range of educational aspects hitherto presumed the sole domain of teachers.

Learning about machine learning

Many countries have recently redeveloped their computer science curricula to respond to the need for a better understanding of computer science among citizens as well as the need for more computer scientists (Webb et al. 2018), but curricula will need to adapt further in order to address the changing emphasis in computer science brought about by machine learning and artificial intelligence. There are signs that some countries are responding to developments in artificial intelligence with curriculum initiatives. China has stated its intention of becoming a world leader in artificial intelligence by 2030 and is introducing artificial intelligence into curricula in primary and secondary schools (Jing 2018; Yang 2019). In its articulation of a new national curriculum, China identified the need to respond to the fact that artificial intelligence is replacing humans in many areas and this affects the kind of competences that humans should develop (Wang 2019). China's approach to its curriculum for artificial intelligence is within an "environment of reflection on the relationship between artificial intelligence and human intelligence, the collaboration between man and machine, and shared development of the future" (Yang 2019). Within this environment, instead of focusing on computers and the Internet, the curriculum for artificial intelligence focuses on data, algorithms, information systems and the information society (ibid.). Through a module focused on AI, within STEM education, students learn about the concepts and historical development of artificial intelligence and gain practical experience of developing simple artificial intelligence applications through a problem-solving approach with some elements of computational thinking (Yu and Chen 2018).

An initiative based in the USA (ai4k12.org), sponsored by the Association for the Advancement of Artificial Intelligence (AAAI) and the Computer Science Teachers Association (CSTA), to develop a framework for artificial intelligence for K-12, has identified big ideas of artificial intelligence which they claim cover the richness of the field while being small enough to be manageable by teachers (Touretzky et al. 2019b) as part of computer or data science education (Magenheim and Schulte 2020). As with the Chinese curriculum, this approach also emphasises strongly the need for students to experience artificial intelligence, not only through interacting with artificial intelligence, but also through adapting and creating artificial intelligence systems. Table 1 summarises these big ideas, which have been suggested to frame the development of the curriculum for artificial intelligence, and what students should understand (Touretzky et al. 2019a).

These developments and especially the broad range of topics shown in Table 1 illustrate the complexity of developing a curriculum that provides students with appropriate knowledge and skills. The broad range of topics also suggests the need for an interdisciplinary approach as well as for significant adaptations to computing curricula. This curriculum challenge is also illustrated at university level where Langley (2019), in an analysis of existing introductory artificial intelligence courses in universities, has identified a number of problems in the content and structure of these courses that may detract from students developing appropriate background understanding and capabilities. More specifically, Langley argues that current introductory courses focus on students being consumers of artificial intelligence rather than producers. This approach fails to include many of the

Table 1 Big ideas in AI and what students should understand (based on Touretzky et al. 2019a)

Key topics	Big idea	What students should understand
Perception and enactment	Computers perceive the world using sensors, and act upon it using various equipment	Machine perception requires extensive domain knowledge e.g. for language there are sounds, vocabulary, grammar and usage patterns. Machine outputs can include control signals to devices like robotic actuators
Representation and reasoning	Agents maintain model/representations of the world and use them for reasoning	The concept of representation e.g. maps, diagrams. Computers construct representations using data from which algorithms can derive new information. This reasoning can be very powerful but is very different from human thinking
Learning	Computers can learn from data	Machine learning is a "kind of statistical inference" that finds patterns in data and allows computers to create their own representations
Natural interaction	Making agents interact comfortably with humans is a substantial challenge for developers	While computers can understand natural language to a limited extent, at present they lack the general reasoning and conversational capabilities of even a child
Societal impact	AI applications can impact society in both positive and negative ways	The ethical construction of AI systems that make decisions affecting people's lives requires attention to the issues of transparency and fairness

discipline's basic and important ideas. It tends to cover isolated elements that are easy to teach, rather than looking at the integrated nature of the subject.

Research on how children around the world interact with artificial intelligence driven smart toys and home systems like Alexa, indicate that children who have experience constructing algorithms with block coding are more likely to believe that artificial intelligence machines are capable of understanding them, and have an easier time understanding artificial intelligence concepts (Druga et al. 2019). A large range of resources have now been developed to enable school students to learn about machine learning through tinkering with applications (Touretzky et al. 2019a). Such hands-on experiences are very important but Jatzlau et al's (2019) analysis shows that the machine learning elements remain as a black box because typically these tools use, for example, an API call. Therefore, how the machine learning works and its potential explainability is not accessible. The online tool, Machine Learning For Kids, for example, enables uses to train a system to recognise images. After training with data, the system provides a percentage likelihood for the identification of a new image. However, the user is not able to access the models and algorithms that have led to this outcome. Other kinds of explanation method are used, such as feature summary (statistical indicators), model internals (weights between layers), example data-points or analogies with surrogate interpretable models. Model-agnostic explanations are generally post-hoc, being decoupled from the 'black box'. Thus, students can explore the output capabilities of machine learning models through training them but cannot scrutinise how they work. Jatzlau et al (2019) have developed an approach for students aged 14 + to use a block-based programming to explore the reinforcement learning paradigm of machine learning. In this system it is possible for students to examine and edit the machine learning algorithms.

The rapid developments in machine learning and the future expectations of widespread use of machine learning-based applications throughout society, mean there is a need for all students to develop some basic literacies relevant to machine learning by the end of their compulsory education. These literacies are needed to enable everyone to: (1) understand the nature of the machine learning processes that may be supporting their own learning and (2) act as responsible citizens in contemplating the ethical issues that machine learning raises. There have been calls for "algorithmic literacy", for example, from specialists in media literacy concerned with how artificial intelligence algorithms are now constructing media environments and enabling misinformation campaigns (Cohen 2018; Wilson 2019).

Questioning algorithms in the context of media consumption can provide a starting point for some educators who might feel overwhelmed by the challenges of understanding and teaching a complex subject. The Algorithm Literacy Project, in Canada, encourages students and teachers to think about how online choices become data that algorithms then use to predict preferences (see https://algorithmliteracy.org/).

Issues, tensions and threats

As discussed above, where a machine learning application is used to develop some specific capabilities as an adjunct to formal education or is used entirely in an informal setting, requirements for explainability might be lower. Users may accept its value based on either their own experience of using the system or on studies that have compared its use with other learning approaches. The use of machine learning systems for high stakes assessment

represents the other extreme in education where such systems are expected to be able to explain and justify their decisions and be held accountable.

An example of conflict created by this expectation comes from Australia where Lazendic et al (2018) investigated the Constructed-Response Automated Scoring Engine (CRASE® by Pacific Metrics), an artificial intelligence system for assessing school student writing in a national test. They found it provided scoring outcomes that had consistency and reliability equivalent to those produced by independent groups of very experienced markers. Furthermore, CRASE® was resilient to attempts to manipulate marking and the latent structure of criterion-based automated scores was the same as that of the human markers. Despite this, the system was not used because of a conflict identified by the Australian Computer Society between commercial confidentiality and the need for explainability. This tension between two regulatory principles illustrates the difficulties faced by new disruptive technologies. Resolution will depend upon societal understanding, political will and the lobbying of commercial interests.

With respect to accountability, the TARDIS project described above (Porayska-Pomsta and Chryssafidou 2018) showed how user referencing can provide checks and balances to an educational machine learning system. This is very much in the spirit of the GDPR cited earlier, by giving the recipient of a machine learning decision some elements of control over its enactment. Once again, there are likely to be tensions between parties concerning the resolution of this power dispute.

Deep learning in both machine and human learning can be of great benefit to society. The balancing act we are attempting will be to enhance and support these benefits, while doing everything in our power to avoid deleterious consequences.

Discussion and conclusion

Naturally, the expert panel format has limitations. Principally these relate to the composition of the group and its time-bounded activities. Although the group comprised experts from Europe, Oceania and North America, the voices of Asia and other lands were missing. Also, machine learning is making huge strides as new applications emerge almost daily. However, within these confines, we considered what teachers and students need to understand in order to make appropriate use of machine learning for their own learning and to understand the broader uses in society. It is clear from the foregoing that societal knowledge and understanding of machine learning will be crucial in resolving the tensions we have described in relation to the human learning that may be supported by machine learning systems and the importance of explainability and accountability of such systems for different educational purposes. There are both parallels and major differences between human deep learning and deep machine learning. Furthermore, computer science is informing neuroscience and vice versa. As we educate our students about machine learning, they can be encouraged to find out more about their own mental processes. An increased coverage of basic elements of neuroscience starting in primary schools could support students' developing understanding of both human learning and machine learning. We need to characterise and define emerging literacies relating to machine learning, algorithms, data/big data, and modelling.

For the wider societal knowledge and understanding of machine learning to be achieved, we will need to reform curricula to ensure all students develop a strong background in machine learning and the range of literacies that support this understanding. Furthermore,

in order to develop their conceptual understanding of algorithms, models and how machine learning works, students must have opportunities not only to use and apply machine learning but also to create their own examples. Recent research, discussed in this paper, suggests that children aged 11 upwards can undertake such activities but developing associated basic literacies including algorithmic literacy can start much earlier. The specific content and sequencing of such curricula are topics for future research and development as discussed later in this paper. As a powerful tool that may not be used to its full potential, there is a need for students to understand how machine learning can be used to identify and solve real-world problems.

In practical terms, our analysis suggests that the capability for explaining its decisions should be programmed into a machine learning system during its design. Several explainability methods from 'rationale' to 'impact' have been described above, each of which can be made comprehensible to the recipient of the machine's decision without infringing on commercially protected algorithms. The ability of a system to provide explanations will be important in the next phase of machine learning development, which is likely to encompass the inter-connection of several machine learning systems. Each system would need to 'explain' to the next the basis for its output, in order that reliabilities can be estimated and compared. For instance, an intelligent tutoring system might take as input a student's emotional state (Bosch et al. 2016, as described above), and combine it with attendance and achievement data. The weight ascribed to each of these three inputs can be adjusted according to the explanation supplied.

Likewise, accountability mechanisms need to be built in during system design and Kroll's (2012) three-layer system of accountability could provide a suitable architecture for this purpose. Accountability is the reverse of explainability. Explainability flows through the system from inputs to output (decision) whereas accountability flows backwards, from decision to the person taking responsibility for it. In this sense there is a human dimension to any machine learning system, so control and legislation are important for managing its use in education. Therefore, in order to keep pace with developments, we will need to update policies and practices. The development of a Code of Conduct for machine learning in education for users and developers is likely to be an important element of this process. As argued above, essential components of educational reform in relation to machine learning are professional development and resources for teachers, educational leaders and other key stakeholders. It will also be necessary to support educators and learners in conducting risk analysis in the use of machine learning in education. Finally, research into the nature of relevant policy and practice across different countries will be important to identify best practices and to enable opportunities and mitigate risks globally for machine learning in education.

As educators introduce more and more learning tools into their teaching repertoire, machine learning will play an increasingly important role in future learning of individuals both in and out of school. What does this mean for teachers, students, and other stakeholders in our education system? Our recommendations are:

For teachers and students: To identify and use applications for learning that incorporate machine learning, all teachers will need an introduction to machine learning for education as part of pre-service teacher education and in-service teacher professional development. This introduction should provide a basic understanding of machine learning and discussion/orientation to commonly used machine learning applications for learning. Equally important will be preparing students for success in both learning and future work for which purpose teachers will need to explore and understand how machine learning is used to conduct routine tasks and solve problems in the workplace. Teachers will need to integrate

examples into their curricula so that students develop basic skills in using the affordances of machine learning in classrooms, developing at the same time an understanding of how machine learning works and its possible uses in the world of work. This basic understanding along with opportunities to explore and practice machine learning in workplace applications and problem solving will prepare them for success in work. Teacher externships, providing teachers with opportunities to spend school vacations and/or summers working in high technology industries will help educators develop the skills, knowledge and examples they can bring into their classrooms.

For policymakers: Transforming learning through the integration of applications that incorporate machine learning will require careful planning and allocation of resources. Policymakers need to understand and make connections between machine learning used for learning in schools and the machine learning needed to drive local economies and help local business and industry thrive. Aligning education and industry interests may leverage investments in both. Developing partnerships between education and business or industry can yield productive results with broad social impact.

For machine learning developers: We urge that they consider the moral, ethical and likely legal requirements for explainability and accountability to be designed into new systems from the outset.

For researchers: The developments that we envisage in this article create a need for research into a range of educational and social consequences of new developments in machine learning including new learning opportunities, effects on practice, cognitive and affective aspects of learning with applications that incorporate machine learning as well as the pedagogy of how to develop learners' understanding of machine learning and at what age particular conceptual areas can best be incorporated. Collaboration between researchers and other stakeholders will be important for focusing and prioritising research agendas and providing evidence to support rapid change in education in response to technological development.

Final comments

This article has considered both specialised machine learning systems which are already fairly extensively used in education as well as generalised machine learning systems which can be turned to any task. These generalised systems may comprise multiple inter-connected specialised machine learning systems. There has been a growth in autonomous artificial intelligence systems in military and space applications especially when communication with a guiding human is impossible or impractical. As we move into a future increasingly pervaded by systems incorporating machine learning, we understand that machines will be better than people at doing many things. Machines will "understand" complex things that we cannot reason through as quickly; detect processes and make inferences about issues that we won't know about; informing the decisions we make. They will direct what we do based on insights into data they analyse. But how will machine learning systems be integrated into the workplace in ways that allow us to do what we do best? How will they help us thrive in workplaces that allow machines to do what they do best? Although these issues seem to be far in the future, implications of what is needed for workplace success need to be driving both how machine learning is used for teaching and articulation of the understanding and skill set required by students preparing for life and work.

The twin issues of explainability and accountability for machine learning in education are not going to be resolved quickly. However, technical developments are fast out-running legislation, so there is urgency in the situation. UNESCO's Executive Board (2019) has studied the issues and invited a new standard-setting instrument on the ethics of artificial intelligence to be considered by the General Conference at its 41st session. We trust this instrument will address explainability and accountability issues and promote a global discussion and consensus.

Compliance with ethical standards

Conflict of interest None.

Research involving human participants and/or animals None.

References

Alenezi, H. S., & Faisal, M. H. (2020). Utilizing crowdsourcing and machine learning in education: Literature review. *Education and Information Technologies, 25,* 2971–2986. https://doi.org/10.1007/s10639-020-10102-w

Asikainen, H., & Gijbels, D. (2017). Do students develop towards more deep approaches to learning during studies? A systematic review on the development of students' deep and surface approaches to learning in Higher Education. *Educational Psychology Review, 29*(2), 205–234. https://doi.org/10.1007/s10648-017-9406-6

Ball, R., Duhadway, L., Feuz, K., Jensen, J., Rague, B., & Weidman, D. (2019). *Applying machine learning to improve curriculum design.* Paper presented at the proceedings of the 50th ACM technical symposium on computer science education, Minneapolis, MN, USA.

Baltrušaitis, T., Ahuja, C., & Morency, L. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Bosch, N., D'Mello, S. K., Baker, R. S., Ocumpaugh, J., Shute, V., Ventura, M., et al. (2016). Detecting student emotions in computer-enabled classrooms. International Joint Conference on Artificial Intelligence, July 9–15, New York (pp. 4125–4129). https://www.ijcai.org/Proceedings/16/Papers/615.pdf.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics, 8*(8), 832.

Cheng, L., & Yu, T. (2019). A new generation of AI: A review and perspective on machine learning technologies applied to smart energy and electric power systems. *International Journal of Energy Research, 43,* 1928–1973. https://doi.org/10.1002/er.4333

Cohen, J. N. (2018). Exploring echo-systems: How algorithms shape immersive media environments. *Journal of Media Literacy Education, 10*(2), 139–151.

Conati, C., Porayska-Pomsta, K., & Mavrikis, M. (2018). AI in Education needs interpretable machine learning: Lessons from open learner modelling. arXiv preprint https://arxi.org/1807.00154.

Dewan, M. A. A., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: A review. *Smart Learning Environments, 6*(1), 1. https://doi.org/10.1186/s40561-018-0080-z

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint https://arxiv.org/1702.08608.

Druga, S., Vu, S. T., Likhith, E., & Qiu, T. (2019). Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019* (pp. 104–111).

Entwistle, N. (2005). Learning outcomes and ways of thinking across contrasting disciplines and settings in higher education. *Curriculum Journal, 16*(1), 67–82. https://doi.org/10.1080/0958517042000336818

EU GDPR. (2018). Regulation (EU) 2016/679 of the European Parliament and of the Council: General Data Protection Regulation. *Official Journal of the European Union*, L 119/1. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine, 161,* 1–13. https://doi.org/10.1016/j.cmpb.2018.04.005

Fullan, M., Gardner, M., & Drummy, M. (2019). Going deeper. *Education Leadership, 76*(8), 64–68.

Fullan, M., Hill, P., & Rincón-Gallardo, S. (2017). *Deep learning: Shaking the foundations*. Ontario, Canada. www.ndpl.global.

Galliers, R. D., & Huang, J. C. (2012). The teaching of qualitative research methods in information systems: An explorative study utilizing learning theory. *European Journal of Information Systems, 21,* 119–134. https://doi.org/10.1057/ejis.2011.44

GarcíaBotero, G., Questier, F., & Zhu, C. (2019). Self-directed language learning in a mobile-assisted, out-of-class context: Do students walk the talk? *Computer Assisted Language Learning, 32*(1–2), 71–97. https://doi.org/10.1080/09588221.2018.1485707

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Amsterdam: Elsevier.

Hao, K. (2020). Born in China, taught by AI. *MIT Technology Review, 123*(1), 24–29.

Hayashi, Y. (2019). The right direction needed to develop white-box deep learning in radiology, pathology, and ophthalmology: a short review [mini review]. *Frontiers in Robotics and AI*. https://doi.org/10.3389/frobt.2019.00024

Holzinger, A. (2018). From machine learning to explainable AI. In *2018 world symposium on digital intelligence for systems and machines (DISA),* (pp. 55–66). IEEE.

Howie, P., & Bagnall, R. (2013). A critique of the deep and surface approaches to learning model. *Teaching in Higher Education, 18*(4), 389–400. https://doi.org/10.1080/13562517.2012.733689

ICO & Turing. (2019). Explaining decisions made with AI: Draft guidance for consultation - Part 1: The basics of explaining AI. Wilmslow, Cheshire, UK: (Information Commissioner's Office and The Alan Turing Institute).

Itransition. (2019). Explainable AI and the future of machine learning. *Ai Weekly Newsletter*.

Jatzlau, S., Michaeli, T., Seegerer, S., & Romeike, R. (2019). It's not magic after all-machine learning in snap! Using reinforcement learning. In *2019 IEEE blocks and beyond workshop (B&B),* (pp. 37–41). IEEE.

Jing, M. (2018). China looks to school kids to win the global AI race. *South China Morning Post*.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kloos, C. D., Alario-Hoyos, C., Muñoz-Merino, P. J., Ibáñez, M., Estévez-Ayres, I., & Crespo-García, R. M. (2019). What can you do with educational technology that is getting more human? In *2019 IEEE global engineering education conference (EDUCON), 8–11 April* (pp. 1480–1487). https://doi.org/https://doi.org/10.1109/EDUCON.2019.8725188

Knox, J., Wang, Y., & Gallagher, M. (Eds.). (2019). *Artificial intelligence and inclusive education: Speculative futures and emerging practices*. Berlin: Springer.

Kroll, J. (2012). *Accountable algorithms*. Princeton: Princeton University.

Langley, P. (2019). An integrative framework for artificial intelligence education. In *Proceedings of the AAAI Conference on Artificial Intelligence, 2019* (Vol. 33, pp. 9670–9677).

Lazendic, G., Justus, J.-A., & Rabinowitz, S. (2018). NAPLAN online Automated scoring research program: research report. https://www.nap.edu.au/docs/default-source/default-document-library/naplan-online-aes-research-report-final.pdf?sfvrsn=0.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors, 18*(8), 2674.

Livieris, I. E., Drakopoulou, K., Tampakas, V. T., Mikropoulos, T. A., & Pintelas, P. (2019). Predicting secondary school students' performance utilizing a semi-supervised learning approach. *Journal of Educational Computing Research, 57*(2), 448–470. https://doi.org/10.1177/0735633117752614

Low, J. (2019). A pragmatic definition of the concept of theoretical saturation. *Sociological Focus, 52*(2), 131–139. https://doi.org/10.1080/00380237.2018.1544514

Magenheim, J., & Schulte, C. (2020). Data science education. In A. Tattnall (Ed.), *Encyclopedia of Education and Information Technologies*. Cham: Springer.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—outcome and process. *British Journal of Educational Psychology, 46*(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Mehta, J., & Fine, S. (2019). *In search of deeper learning: The quest to remake the American high school.* Cambridge: Harvard University Press.

Nájera, A. B. U., & de la Calleja Mora, J. (2017). Brief review of educational applications using data mining and machine learning. *RevistaElectrónica de InvestigaciónEducativa (REDIE), 19*(4), 84–96.

Obeid, C., Lahoud, I., Khoury, H. E., & Champin, P.-A. (2018). Ontology-based recommender system in Higher Education. Paper presented at the *Companion Proceedings of the The Web Conference* 2018, Lyon, France. https://doi.org/https://doi.org/10.1145/3184558.3191533

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks, 113,* 54–71. https://doi.org/10.1016/j.neunet.2019.01.012

Porayska-Pomsta, K., & Chryssafidou, E. (2018). Adolescents' self-regulation during job interviews through an AI coaching environment. *International conference on artificial intelligence in education, 201* (pp. 281–285). Cham: Springer.

Porayska-Pomsta, K., & Rajendran, G. (2019). Accountability in human and Artificial Intelligence decision-making as the basis for diversity and educational inclusion. In J. Knox, Y. Wang, & M. Gallagher (Eds.), *Artificial Intelligence and inclusive education: Perspectives on rethinking and reforming education*. Singapore: Springer.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., et al. (2019). Machine behaviour. *Nature, 568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

Rajagopal, K., van Bruggen, J. M., & Sloep, P. B. (2017). Recommending peers for learning: Matching on dissimilarity in interpretations to provoke breakdown. *British Journal of Educational Technology, 48*(2), 385–406.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Shepard, L. A. (2019). Classroom assessment to Support Teaching and Learning. *The ANNALS of the American Academy of Political and Social Science, 683*(1), 183–200. https://doi.org/10.1177/0002716219843818

Simon, H. A. (1983). Why should machines learn? *Machine learning* (pp. 25–37). Amsterdam: Elsevier.

Sze, V., Chen, Y., Yang, T., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE, 105*(12), 2295–2329. https://doi.org/10.1109/JPROC.2017.2761740

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence, 33,* 9795–9799.

Touretzky, D., Martin, F., Seehorn, D., Breazeal, C., & Posner, T. (2019b). Special session: AI for K-12 guidelines initiative. In *Proceedings of the 50th ACM technical symposium on computer science education,* pp. 492–493: ACM.

Tuomi, I. (2018). *The impact of artificial intelligence on learning, teaching, and education: Policies for the future*. Luxembourg: Publications Office of the European Union.

UNESCO Executive Board. (2019). *Preliminary study on the technical and legal aspects relating to the desirability of a standard-setting instrument on the ethics of artificial intelligence*: 206 EX/42. Paris. https://unesdoc.unesco.org/ark:/48223/pf0000367422.

Wang, J., & Tao, Q. (2008). Machine learning: The state of the art. *IEEE Intelligent Systems, 23*(6), 49–55. https://doi.org/10.1109/MIS.2008.107

Wang, T. (2019). Competence for students' future: Curriculum change and policy redesign in China. *ECNU Review of Education, 2*(2), 234–245. https://doi.org/10.1177/2096531119850905

Webb, M. E., Bell, T., Davis, N., Katz Yaacov, J., Fluck, A., SysłoMaciej, M., et al. (2018). Tensions in specifying computing curricula for K-12: Towards a principled approach for objectives. *Information Technology, 60*(2), 59–68.

Webb, M. E., Fluck, A., Deschenes, M., Kheirallah, S., Lee, I., Magenheim, J., et al. (2019). Thematic Working Group 4-State of the Art in Thinking About Machine Learning: Implications for Education. In P. Fisser, & M. Phillips (Eds.), *Learners and learning contexts: New alignments for the digital age Report of EDUsummIT 2019, Québec City, Canada, September 29th–October 2nd, 2019* (pp. 127–136)

Webb, M. E., & Ifenthaler, D. (2018). Assessment as, for, and of Twenty-First-Century learning using information technology: An overview. In J. Voogt, G. Knezek, R. Christensen, & K.-W. Lai (Eds.), *Second Handbook of Information Technology in Primary and Secondary Education* (pp. 1–20). Cham: Springer.

Wilson, C. (2019). Media and Information Literacy: Challenges and opportunities for the World of Education. https://mediaethics.ca/media-and-information-literacy.

Yang, X. (2019). Accelerated move for AI education in China. *ECNU Review of Education, 2*(3), 347–352. https://doi.org/10.1177/2096531119878590

Yu, Y., & Chen, Y. (2018). Design and development of high school artificial intelligence textbook based on computational thinking. *Open Access Library Journal, 5*(9), 1–15.

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion, 42,* 146–157. https://doi.org/10.1016/j.inffus.2017.10.006

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing, 115,* 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mary E. Webb is Reader in Information Technology in Education at King's College London and on the International Federation of Information processing (IFIP) Education Executive Committee leading the IFIP Task Force on Computer Science in the Curriculum. Mary is internationally recognised for her research on pedagogy and formative assessment across both Computer Science as a subject and the uses of new technologies for learning. Mary's research has resulted in over 150 publications. Recent research has focused on analysing computing curricula internationally as well as the use of haptics in virtual reality simulations in schools. See more details at https://www.kcl.ac.uk/people/mary-webb.

Andrew Fluck has taught science, mathematics and computing in Nigeria, England and Australia. He teaches and researches information technology education in the Faculty of Education at the University of Tasmania. Dr. Fluck's funded research investigates the transformative use of computers to teach integral calculus and quantum mechanics in primary schools; and eExams, where students take their own computers into the exam hall. He is the previous chair of Working Group 3.3 (research into educational applications of information technologies) for IFIP/UNESCO. Andrew is an avid longbow archer, continental archery judge and Director for Archery Australia. See more details at http://Andrew.Fluck.id.au.

Johannes Magenheim is a Senior-Professor for Computer Science Education at the Institute of Computer Science at the University of Paderborn. His primary areas of research and teaching are Computer Science Education, E-Learning, and organisational knowledge management. He conducted numerous research projects in these areas which have been funded e.g. by the German Research Foundation, the German Ministry for Research and Education, the European Union, and the industry. He is also member of different working groups of the German Informatics Society (GI) and German National Representative to IFIP TC3. In these functions he had responsibilities in various commissions for curriculum development of computing education in primary, secondary and higher education in Germany.

Joyce Malyn-Smith is a national expert on STEM career development and workforce education. She leads a body of work that explores how to enhance STEM learning and support people in using STEM skills, knowledge, and dispositions to pursue productive and rewarding careers. Her projects develop industry/education connections in computational thinking, big data, artificial intelligence (AI), and other advanced technologies.

Juliet Waters brings over twenty years of journalism experience to her passion for creating sustainable communities of practice in digital learning. She has written about her own adventure learning to code for The New York Times. As Chief Knowledge Officer for the Canadian not-for-profit Kids Code Jeunesse, she

created professional development materials and led workshops across Canada as part of the CanCode program launched by Innovation, Science and Economic Development Canada.

Michelle Deschênes is a doctoral student in educational technology at Laval University. Her research focuses on teachers' agency and professional development. She is interested in learning analytics, communities of practice, and artificial intelligence. She is a lecturer in pre-service and in-service teacher education at the college and secondary level and teaches web development at the college level.

Jason Zagami is a senior lecturer in Educational Technologies and Computer Science Education with research interests in Educational Technology Trends, Artificial Intelligence, Robotics, and the use of games in education. Jason is editor of Australian Educational Computing. See more details at jason.zagami.info.

Affiliations

Mary E. Webb[1] · **Andrew Fluck[2]** · **Johannes Magenheim[3]** · **Joyce Malyn-Smith[4]** · **Juliet Waters[5]** · **Michelle Deschênes[6]** · **Jason Zagami[7]**

Andrew Fluck
andrew.fluck@utas.edu.au

Johannes Magenheim
jsm@uni-paderborn.de

Jason Zagami
j.zagami@griffith.edu.au

[1]   Reader IT and Education, School of Education, Communication and Society, King's College London, Franklin-Wilkins Building - Waterloo Bridge Wing, Waterloo Rd, London SE1 9NN, UK

[2]   University of Tasmania, Launceston, Australia

[3]   University of Paderborn, Paderborn, Germany

[4]   Education Development Center (EDC), Waltham, MA, USA

[5]   Kids Code Jeunesse, Montreal, Canada

[6]   Laval University, Quebec City, Canada

[7]   Griffith University, Gold Coast, Australia