EDUsummIT 2019 Thematic Working Group 4

State of the art in thinking about machine learning – implications for education

Andrew Fluck (University of Tasmania), Mary Webb (King's College London), Irene Lee (MIT), Joyce Malyn-Smith (Education Development Center, MA), Jason Zagami (Griffith University), Raymond Trippe (Lucas Onderwijs), Johannes Magenheim (University of Paderborn), James Slotta (OISE/University of Toronto), Amma Ofori (York University), Michelle Deschênes (Laval University).

Keywords: *machine learning; education; deep learning; neural networks; social effects; policy recommendations.*

Introduction

Examples of machine learning (ML) that many people will be familiar with are 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, Busato et al. 2018), the energy industry (Cheng and Yu 2019), e-commerce (Zhang, Yang et al. 2018) , fault detection and diagnosis across most types of machinery (Zhao, Yan et al. 2019) and healthcare (Faust, Hagiwara et al. 2018). Likewise, in education, machine learning is becoming more widespread and has been used, for example, for improving curriculum design (Ball, Duhadway et al. 2019), predicting students' grades (Livieris, Drakopoulou et al. 2019), recommending higher education courses to students (Obeid, Lahoud et al. 2018); student modelling for intelligent tutoring systems (Conati, Porayska-Pomsta et al. 2018).

A survey conducted in 2016 (Grace, Salvatier et al. 2018) showed that machine learning researchers predicted that computers would outperform humans in many activities within the next 10 years. However, the survey also showed very large differences in these researchers' predictions of when "High-level machine intelligence" (HLMI) would be achieved, i.e. when unaided machines can accomplish every task better and more cheaply than human workers. Examples of activities where computers would soon outperform humans included translating languages (by 2024), writing high-school essays (by 2026), driving a truck (by 2027), working in retail (by 2031) and working as a surgeon (by 2053) but education was not discussed. A recent policy foresight report by the Joint Research Centre (JRC), the European Commission's science and knowledge service, suggests that artificial intelligence (AI) powered by machine learning will change learning, teaching and education rapidly in future creating high pressure to transform educational practices, institutions and policies (Tuomi 2018).

In this paper we examine how machine learning will have implications both for how people learn in the future and for what learners and teachers need to know about machine learning. Furthermore, we consider the implications for education policy. We start by considering definitions and characterisations of machine learning (ML) and its relationship to artificial intelligence (AI). We then focus on: mapping current applications and examining possible future developments for the next 5-10 years; emerging ethical frameworks; and finally, specific policy, practice and research recommendations for educational contexts.

Defining and characterising machine learning

There is consensus that machine learning is a subset of artificial intelligence (see Figure 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 not clearly defined and there are diverse views on its potential and risks (Kaplan and Haenlein 2019).

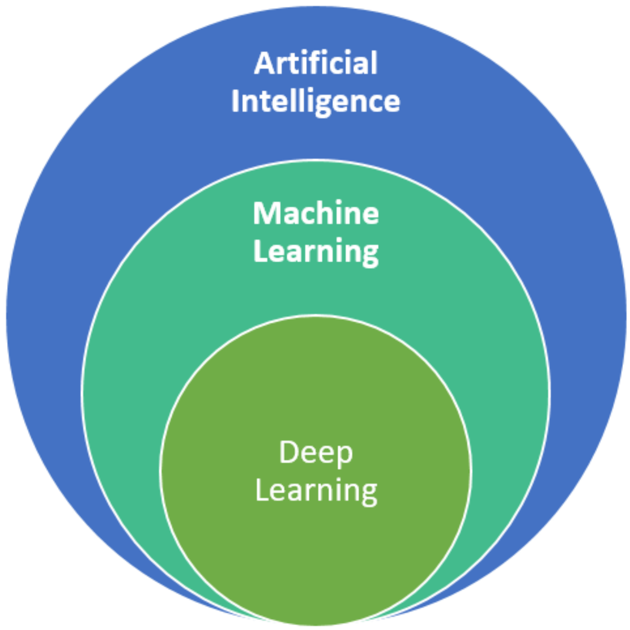

Figure 1: Relationships between terms (after Nadia Berchane 2018)

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 (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, 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, generating a representative dataset and using algorithms with statistical reliability, a cross disciplinary endeavour involving computer scientists, statisticians and data scientists:

> "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 on not only the algorithms with which it has been originally programmed, but also design decisions of the original engineers in terms of the values of learning rate parameters, the initial training regime and the choice of dataset, the context in which it is learning and subsequent upgrades to the system (Rahwan, Cebrian et al. 2019). Currently the machine learning field has no clear classification scheme for its algorithms owing to the wide range of algorithms and variations (Portugal, Alencar et al. 2018). There are many different machine learning algorithms and new algorithms are continually being introduced (Cheng and Yu 2019). Machine learning can be classified (Portugal, Alencar et al. 2018, Cheng and Yu 2019) according to the initial 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; and
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 network (DNN) (Sze, Chen et al. 2017) enabled by large multi-layered neural networks and massive amounts of data (Zhang, Yang et al. 2018). LeCun, Bengio & Hinton (2015 p.436) 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. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation 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. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech."

Deep learning is now widely used in multimedia applications such as speech recognition and computer vision and is expected to make a large contribution in many other fields in the near future (Sze, Chen et al. 2017).

According to Cheng and Yu (2019) artificial intelligence is in its third stage which began in the 1990s when machine learning emerged. The first two stages of artificial intelligence were focused on logical reasoning and capturing human expertise. Recent advances in deep learning and big data led to a new generation of artificial intelligence (AI 2.0) (Cheng and Yu 2019). In this new era of AI, machine learning algorithms discover new knowledge directly from data rather than requiring humans to design the knowledge into knowledge-based systems (FITEE editorial staff 2018). In order to successfully discover new knowledge in most situations, machine learning needs to use data from many different sources and to make sense of many different data types i.e. Cross media knowledge (FITEE editorial staff 2018).

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 (Baltrušaitis, Ahuja et al. 2019) and low-quality datasets (Zhang, Yang et al. 2018). Furthermore, whereas humans are very well capable of coping with the typically continuous streams of data that we receive and integrating new knowledge into existing knowledge, this lifelong learning remains very challenging for the deep learning algorithms currently available (Parisi, Kemker 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, Kemker et al. 2019). It is important to recognise that for lifelong-learning, biological processes are far in advance of any machine learning yet implemented. Mechanisms for overcoming these problems are currently focused on understanding neurobiological processes (Parisi, Kemker et al. 2019). While the original model is defined by its programming, the order of training elements may affect the operating parameters in different ways, making the subsequent behaviour of the system difficult to predict. There is a need to find a way to explain these blackbox systems in both education (Rudin 2019) and medicine (Hayashi 2019).

Applications of machine learning in education: present and future

Over the last decade, many applications leveraging advances in artificial intelligence have been designed, developed, and implemented in a variety of educational contexts. In this section, we identify key areas that are being addressed by use of machine learning and present some cases that represent a good spread of current applications. The selected examples are recent (2018 and 2019) and we examine them according to the following aspects: their purpose, the context in which they are used, the machine learning methods used, and their potential benefits and limitations.

Example: automated assessment of students

Automated assessment systems, utilising machine learning techniques, are developing rapidly and are not only able to deal with multiple choice and short answer items but also long-response questions and essays (e.g, see Whitelock and Bektik 2018 for a review). Furthermore, as well as summative judgements, such tools can provide formative assessment and feedback, an aspect that is currently under-developed and should be developed further in order to support learning (Ifenthaler, Greiff et al. 2018; Whitelock and Bektik 2018). Automated assessments have focused predominantly on assessment of individuals but the Organization for Economic Cooperation and Development (OECD), has utilised artificial intelligence assisted assessments of collaborative problem solving for the Programme for International Student Assessment (PISA) exam (Webb and Gibson 2015; PISA 2017). Competency is assessed by how well the individual interacts with artificial intelligence agents during the course of problem solving. This includes achieving a shared understanding of the goals and activities, as well as efforts to pool resources and solve the problem. The USA National Center for Education Statistics (NCES) is considering adding such an assessment to the National Assessment of Educational Progress (NAEP), a federally sponsored assessment administered to a sample of students in a variety of subject areas (ref?).

Automated assessment is developing rapidly but some types of assessment remain challenging, especially where they utilise qualitative data. For example, reflective writing is recognized as an effective activity to foster students' reflection and critical thinking, but the analysis of students' writing is a major challenge. Kovanović & al. (2018) automated the analysis of students' self-reflections on the video recordings of their own musical performances using various linguistic tools and indicators. Their content analysis aimed to categorize reflections according to three types: observation, motivation, goal. Results suggested that the use of N-grams and Linguistic Inquiry and Word Count (LIWC) (Tausczik & Pennebaker, 2010) and Coh-Metrix (Graesser, McNamara, & Kulikowich, 2011) provides a good basis for the development of an automated self-reflection classification system.

There are tensions in the future of machine learning in automated assessment. On the one hand, it offers consistency, speed and cost reductions (Lazendic, Justus & Rabinowitz 2018, p.13). On the other hand, there could be considerable legal obstacles to its adoption if the operational characteristics 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:

> 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 Article 13, 2018).

This requirement may be in tension with the commercial confidentiality of the automated assessment system's internal workings. Overall, the future of automated assessment appears to rest upon the 'explainability' of machine learning systems.

Example: personalized learning and intelligent tutoring
Personalized learning falls into three categories: Adaptive learning apps, Recommender systems and assistive tutors. Adaptive learning apps are used to adjust learning experiences based on a student's needs and allow for differentiation within a classroom. Content Technologies and Carnegie Learning are two companies currently developing these types of instructional systems. Their instructional systems include providing learning experiences, testing, and feedback to students. The learning experiences (such as in-class assignments) and tests can also be auto-generated for each student's particular needs. Recommender systems, much like a Netflix for education, operate on a learner's past interests, learning and performance to suggest additional content, tools and resources for students to try next. Assistive tutors support individual learners by providing assistance with homework or test preparation. Through individualized support and instantaneous feedback, this method of tutoring aspires to increase student progress in learning.

Several co-tutoring and adaptive learning systems are already in operation. Amira is a reading assistant for K-3 students that listens, assesses and coaches to accelerate reading mastery. (https://www.amiralearning.com/). Duolingo is an application for learning a foreign language. The lessons adapt to the user's learning style. Data from 300 million users enables the system to discover new insights about the nature of language and learning (https://www.duolingo.com/).

To see the future of such systems, we note that to date research in Human Language Technology (HLT) has provided improvements to such things as machine translation, name/entity and event detection, and knowledge base creation; and have resulted in various applications in personal assistants (Siri and Alexa) call and reservation centres. However, HTL has no regard for meaning as it is reliant on statistical aspects of language. GAILA is a research initiative that will overcome these weaknesses by grounding events in visual cues thereby enabling computers to acquire and make meaning of language the way children acquire language. GAILA does not start with vocabulary, a dictionary or taxonomy. GAILA Language will be acquired by aligning visuals with words (sounds/auditory input) and by labelling and interpreting distinctions among visual representations through sequencing information, varying word forms, using verbs, nouns, adjectives and prepositional phrases. Machine learning will associate text and spoken input with specific elements of images and will use logic/heuristics/inference to describe previously unseen entities, relations and events. For example after seeing a black table, a white table, and a black chair, machine learning should be able to learn enough about the meanings of the words "black," "white," "table," and "chair" that it is able to take the colour name from "white table" and the object name from "black chair" so that it can recognize a newly-observed white chair and describe it as such. GAILA will make information more useable by automated analytics (DARPA-PA-18-02-06).

Example: optimizing resource deployment
Machine learning can assist with various IT and human services offered in schools by finding patterns in complex data in order to create data-driven forecasts. These forecasts can support setting up of class schedules, seating assignments, and streamlining automated ordering of supplies and cafeteria food. It has also been proposed that artificial intelligence can assist with identifying student mood and affect based on facial recognition techniques (Bosch et al. 2016; Al-Alwani 2016). Counselling offices can be alerted to students found to be in distress, and teachers can be alerted to students struggling to grasp a concept.

Gray and Perkins (2019) investigated the potential of applying modern machine learning tools and techniques to student attendance data. They aimed to identify students at risk of dropping out as early as possible with maximum precision. The authors used WEKA workbench, classifiers, and tools (Frank et al., 2016, p. 553) to evaluate several candidate machine learning methods: Sequential Forward Selection (Kittler, 1986; Whitney, 1971), the Nearest Neighbour classifier (Kudo and Sklansky, 2000), and the C4.5 Pruned Tree classifier (Quinlan, 1993). The F-Measure statistic (Joshi, 2002) was used to measure performance. Gray and Perkins used data from a full academic year (4877 instances / students), school and the Bangor Engagement Metric for (non)-attendance over the first three weeks. The authors demonstrated that it was possible to identify students at risk with about 97% accuracy, thus showing the potential of machine learning in a context of student retention.

In the future, these systems are likely to become real-time enabling students at risk to be identified earlier and support services more quickly deployed to assist them. The USA National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA) have developed the Real-Time Machine Learning Program (RTML) to develop foundational breakthroughs in hardware and machine learning needed to build systems that respond and adapt in real time. Through this effort they will create a processor that can interpret and learn from data in real-time, solve unfamiliar problems using what it has learned, and operate with the energy efficiency of the human brain (DARPA - HR001119S0037).

Issues with developing future applications of machine learning in education

The developments outlined here have the potential to free up teachers' time and capitalize on efficiencies in school settings. However, they do not attend to aspects of social learning, collective sense-making and managing full-class discussions.

Just as these multiple computer systems interact, it is not too difficult to perceive a time when multiple machine learning systems will need to interact as well. How would that work? Would multiple artificial intelligence agents be able to collaborate with their deductions and the supporting rationales with estimated probabilities of accuracy? Would one method be the creation of an over-arching software neural network that takes the output of assessment and other machine learning system as input?

It has been argued that machine learning systems need to report on explanations for their findings. By extension, any over-arching software neural network (OASNN) would need to weave explanations from its sub-systems into a rational explanation for its own outputs. To this extent, such an OASNN would be aware of its own internal workings.

Emerging ethical frameworks for machine Learning

Ethical questions of machine learning/deep learning are only understandable if they are considered in a broader context, concerning their internal software, training data sets and data collection technologies. When we talk about ethical issues of machine learning/deep learning, we should therefore be aware that it is a mixed-method technology used in specific application contexts.

To consider machine learning/artificial intelligence systems, one must go beyond the applied algorithms and computer models. It is important to also consider the data underlying these calculations, the question of their origin and reliability, as well as the transparency of the results and the credibility of the data interpretation.

In this context, Katal, Wazid, Goudar (2013) describe essential technical characteristics in the use of Big Data as variety, volume, velocity, flexibility and complexity. Variety encompasses the disparate nature of sensor readings and different degrees of structure in data records. Volume refers to the remorseless quantity of data produced, particularly in social networking. Velocity indicates the speed at which systems are generating data, causing technical challenges. Variability in data flows over time causes similar challenges – planning for the maximum flow may be expensive! Furthermore, the complexity of data relationships is poorly understood, yet vital to their interpretation.

Basic ethical issues that arise when dealing with Big Data and machine learning/deep learning are well presented in the work of Mittelstadt et al. (2016) who summarise as follows:

> "In information societies, operations, decisions and choices previously left to humans are increasingly delegated to algorithms, which may advise, if not decide, about how data should be interpreted and what actions should be taken as a result. More and more often, algorithms mediate social processes, business transactions, governmental decisions, and how we perceive, understand, and interact among ourselves and with the environment. Gaps between the design and operation of algorithms and our understanding of their ethical implications can have severe consequences affecting individuals as well as groups and whole societies." (p 1).

Mittelstadt et al. address six types of ethical concerns raised by the use of algorithms (see Figure 2).

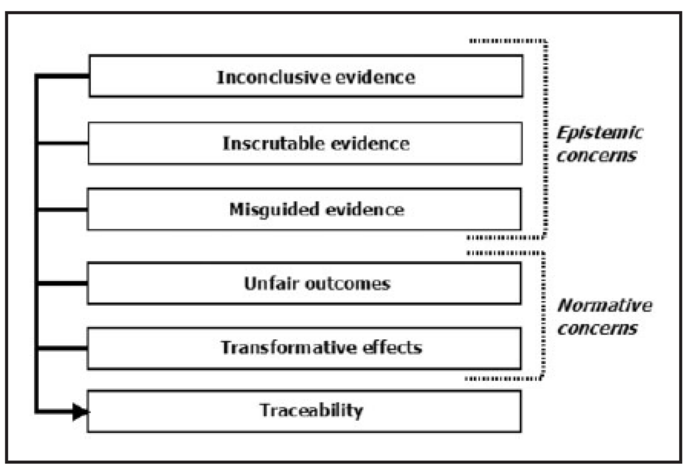

Figure 2: Six types of ethical concerns raised by algorithms (Mittelstadt et al. 2016 p. 4)

Algorithms can produce inconclusive evidence when they use inferential statistics or machine learning techniques because these methods can identify correlation but not causation. Furthermore, the problem of explainability, discussed earlier, gives rise to inscrutable evidence. Conclusions can only be as reliable (but also as neutral) as the data they are based on, thus providing possibilities for misguided evidence. In order to examine the evidence, users need to know the provenance and quality of training data, and how each of the many data-points used by a machine-learning algorithm contribute to the conclusion it generates.

In addition to the three epistemic concerns outlined above, actions driven by algorithms are assessed according to ethical criteria and principles of society which evaluate fairness and discrimination. For example, even if based on conclusive, scrutable and well-founded evidence, an action driven by an algorithm may be regarded as discriminatory, for example, solely from its effect on a class of people that is protected in that particular societal context. Harm caused by algorithmic activity can be hard to detect and find its cause. It is rarely straightforward to trace who should be held responsible for any such harm caused owing to the multiple actors involved. Mittelstadt, Allo et al. (2016), in a review of the literature, further discussed the ethical issues surrounding the most important consequences of the application of algorithms, which they summarized as follows:

- Inconclusive evidence leading to unjustified actions
- Inscrutable evidence leading to opacity
- Misguided evidence leading to bias
- Unfair outcomes leading to discrimination
- Transformative effects leading to challenges for autonomy
- Transformative effects leading to challenges for informational privacy.

These abstract categories form a reasonable basis for limiting and typifying ethical questions in connection with machine learning/deep learning. The use of these categories for the needs of Computer Science Education requires a further concretization and contextualization of the problems. A grid developed at the International Federation of Information Processing (IFIP) conference SUZA 2019 in Zanzibar by members of the IFIP education committee (TC3) (Figure 3) may help with this necessary process. The grid serves as a basis for identifying problems that arise in connection with the future development of ICT and the resulting social challenges. The two dimensions T (technology) and S (society) describe cells T_iS_i into which real social impacts can be entered as a consequence of technological developments. Each of these problems will then be examined with regard to its significance for informatics education and the development of future Computer Science curricula.

	S1:	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
	(Cyber)Security	Privacy	Social / Surveillance	Decent work	Health Care	e-government/ e-administration e-law	environment/ ecology/ sustainability	ethics / deontology/ governance/ code of conduct	mobility	Digital equity	Gender Equity	perenity and quality of information	place of automatisation in the society	Future Society component(s)	Energy	LifeLong Learning and Education	Professional Education	Economy and e-Society
T1 IoT - Internet of Things																		
T2 Quantum Computing																		
T3 Communication (eg 5G)																		
T4 Implication of Big Data																		
T5 Blockchain																		
T6 3D / 4D Printing																		
T7 VR virtual reality - AR Artificial Reality																		
T8 Autonomous Systems																		
T9 Cloud Computing																		
T10 Humanoids																		
T11 Transhumanism																		
T12 Nano Technologies																		
T13 Technologies of Recognition (tracking)																		
T14 Future Technologies																		

Figure 3: Technology and Society dimensions for identifying problems

Ethical questions arising from the application of the mixed-technology machine learning/deep learning can also be located in a number of these cells of the grid. This concerns cells of line T_4 (Implications of Big Data) in particular, but also other cells such as T_1 (IoT, Internet of Things) and T_{13} (Technologies of Recognition / tracking).

Learning about machine learning systems

in addition to learning about the ethical considerations outlined above, the rising importance of machine learning, has major more general consequences for computer science education as Tom Mitchell, Professor of Machine Learning at Carnegie Mellon explained:

Beyond its obvious role as a method for software development, machine learning is also likely to help reshape our view of Computer Science more generally. By shifting the question from "how to program computers" to "how to allow them to program themselves," machine learning emphasizes the design of self-monitoring systems that self-diagnose and self-repair, and on approaches that model their users, and take advantage of the steady stream of data flowing through the program rather than simply processing it (Mitchell 2006 p. 3).

Many countries have recently redeveloped their computer science curricula to respond to the need for better understanding of computer science among citizens as well as the need for more computer scientists (Webb, Davis et al. 2017; Webb 2019) but curricula will need to adapt further in order to address the changing emphasis in computer science by machine learning. There is already an initiative (ai4k12.org), and work in progress, sponsored by AAAI and CSTA, to develop a framework for AI for K-12, focusing on what students should know. A position paper from this initiative has identified big ideas of artificial intelligence: perception, representation and reasoning, learning, natural interaction, societal impact, which they consider cover the richness of the field, while being small enough to be manageable by teachers (Touretzky, Martin et al. 2019).

Figure 4: *User's feedback on interaction with Korbit.*

Approaches to teaching about machine learning are being developed at all levels of education. For example, Martínez-Tenor, Cruz-Martín and Fernández-Madrigal (2019) describe a scenario for interactive learning in engineering at graduate level using Lego® robots to complement the theoretical teaching of Cognitive Robotics. Outside of school, an initiative for the teaching of deep learning was launched by Korbit, a Montreal company. The course, free and open to all, is 4 weeks long and contains lectures and over 100 interactive exercises https://www.korbit.ai/machinelearning . The deep learning tutor guides students through the course, so this is an online course on machine learning taught by a machine learning tutor. It is an initiative that helps to democratize the understanding of artificial intelligence and deep learning. The text analysis remains imperfect so far: adding "not" in the expected answer, the system does not notice that we have just given an incorrect answer. Korbit is, however, in continuous improvement, so all interactions can be evaluated by the learner, as shown in Figure 4.

A range of other materials are available online to support the learning and teaching of machine learning. *Machine learning for kids* is a free online application based on Scratch that introduces machine learning by providing hands-on experiences for training machine learning systems and building things with them (https://machinelearningforkids.co.uk/). *Calypso for Cozmo* is a simple tile-based user interface for the Cozmo robot that incorporates multiple artificial intelligence technologies to learn more about robot logic and behavior (https://Calypso.software). *Cognimates* is platform based on Scratch for building games, programming robots & training artificial intelligence models (http://cognimates.me/home/). *Teachable Machine* is a free online application to start exploring how machine learning works by using a webcam as an input to train a machine learning model and running neural nets (https://teachablemachine.withgoogle.com/). *TensorFlow Playground* is an open-source interactive visualization of neural networks. It contains a tiny neural network library that meets the demands of this educational visualization (https://playground.tensorflow.org).

Emerging areas and questions for discussion at EDUsummIT

While machine learning has many potential benefits for education, there are also issues and problems that educators need to be aware of and developers and researchers need to address. From the examples and future directions presented in this paper, we propose four main interrelated areas for consideration during EduSummit: explainability; trust; the future of educators and their work; machines language and privacy.

Explainability: As machine learning algorithms have become more complex, their reasoning and basis of their judgements has become less accessible and difficult to scrutinise. This so-called "black box" problem is exacerbated by the recent developments in deep learning that make use of very complex multi-layered neural networks. The problem is not unique to education: predictions in medicine also present a dilemma in which prioritising accuracy enabled by complex deep learning over interpretability and transparency has been severely criticised (Hayashi 2019). There is currently 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.

Fundamentally, in the current situation, as machines learn they are not yet able to communicate in human-understandable terms, nor can they self-assess task competency and strategy. Being able to do so is essential if we want to engage with machines in a more accessible and intuitive way and if we are to work together seamlessly with machines as trusted, collaborative partners. To address these types of challenges and to build competency-based trusted machine learning systems DARPA (the U.S. Defense Advanced Research Projects Agency) is developing Competency-Aware Machine Learning (CAML) systems. Such an autonomous system would be able to "self-assess its task competency and strategy, and express both in a human-understandable form, for a given task under given conditions" (Special Notice DARPA (SN) DARPA-SN-19-26 p.1).

Developing machine capability is only part of the equation: we need to understand human requirements for communication from machine learning systems. Research into student modelling with the focus on "Open Learner Modelling" in which student models are accessible with varying levels of interactivity, has started to determine key considerations for designing what information to reveal to the user, how and why and how closely the presentation should represent the machine learning and data models (Conati, Porayska-Pomsta et al. 2018). These key considerations are: 1) why the Open Learner Model is being built; 2) which aspects of the model are made available to the user; 3) how is the model accessed; 4) who has access to the model.

Trust: If humans and machines are to work as true partners in solving problems, a greater degree of trust needs to be developed in the ways machines are solving problems. This is difficult to do when machines are not yet able to describe how they go about coming to a conclusion or predicting an outcome. For example, through machine learning we know that telemetry devices attached to newborns in neonatal units can predict which infants will develop an infection 48 hours before symptoms present themselves. Medical professionals are put in a position to have to trust these predictions without fully understanding the causal relationship. Without this deeper understanding of how machine learning addresses and solves a problem, humans are not likely to be able to fully trust machines as partners. It is difficult to establish trust with machines if they cannot explain and therefore if we cannot understand how they come to conclusions.

The future of educators and their work: machine learning specialists have no doubt that developments in machine learning will impact work:

Automated stores and restaurants and driverless cars will become more commonplace. Although some jobs may disappear, the majority will be disrupted wherein some work tasks will be taken over by machines and the “job" will be shared with humans, where humans do what they do best, and machines do what they do best. For example, a physician’s communication skills in extracting symptoms from a patient and offering that to machine learning for diagnosis, a warehouse picker riding pre-programmed automated vehicles to locate more efficiently merchandize in warehouses for shipping (interview with Kai Fu Li, 2019).

Machines, language and privacy: Machines will be able to derive meaning from words and phrases, not simply translate words. This means that colloquialisms, phrases that combines words in ways to change meaning and words in context will be better understood which has significant implications for machine learning applications. This means personal privacy and equity will become increasingly important as people struggle with decisions between the convenience that machine learning provides and personal privacy that is given up when taking advantage of these conveniences; and as we learn more about the biases that are programmed into machine learning systems.

While examining the four areas identified above and the capabilities and future development of machine learning, the following key questions have emerged for consideration:

1. How can deep learning systems based on complex neural networks explain their workings to each other and humans? Are such explanations indeed possible?
2. What tensions arise from the current activities and future potential?
3. How do we evaluate the balance of benefits and harms arising from applying machine learning in education?
4. What strategies and policies with respect to machine learning in education can direct its development in overall beneficial ways for humankind?
5. Who will be teaching whom, who will be leading whom and what roles will be attributable to learning companions (bots)?
6. How will new human computer interfaces change the learning interactions and scope and range of learning opportunities?
7. How can we prevent bias and unfair and discriminatory outcomes from machine learning systems?
8. What do teachers and students need to know about machine learning?
9. What do all citizens need to understand about machine learning?
10. What are the implications of machine learning for the computer science curriculum?[1]
11. What are the likely relationships between quantum computing and machine learning?
12. Do we have a moral obligation to nurture consciousness in machine learning systems?

Recommendations for Policy, Practice and Research

We endorse Robert Murphy’s three policy recommendations in relation to artificial intelligence in education identified in a RAND perspective paper (2019) that provide general principles that we endorse. The first recommendation urges developers to focus artificial intelligence on solving important teaching problems, for which they are most suited, such as adaptive online instruction, automated text scoring and early identification of students at risk. His second recommendation promotes the importance of ‘explainability’ of machine learning systems to stakeholders. The third suggestion is for further research into the effects of artificial intelligence in learning and teaching. It is our intention at EDUsummIT 2019 to further clarify these three policy recommendations as we examine the issues and questions outlined above.

[1] See for instance: aiforgood.com.au

References

Al-Alwani, A. (2016). Mood Extraction Using Facial Features to Improve Learning Curves of Students in E-Learning Systems. International Journal of Advanced Computer Science and Applications 7(11)444-453. https://thesai.org/Publications/ViewPaper?Volume=7&Issue=11&Code=ijacsa&SerialNo=57.

Ball, R., L. Duhadway, K. Feuz, J. Jensen, B. Rague and D. Weidman (2019). Applying Machine Learning to Improve Curriculum Design. Proceedings of the 50th ACM Technical Symposium on Computer Science Education. Minneapolis, MN, USA, ACM**:** 787-793.

Baltrušaitis, T., C. Ahuja and L. Morency (2019). "Multimodal Machine Learning: A Survey and Taxonomy." IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(2): 423-443.

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671), 195.

Biever, C. (15May 2013) Consciousness: Why we need to build sentient machines. *New Scientist*.

Bosch, N., D'Mello, SK, Baker, RS, Ocumpaugh, J., Shute, V., Ventura, M., Wang, L. & Zhao, W. (2016). Detecting Student Emotions in Computer-Enabled Classrooms. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16). https://www.ijcai.org/Proceedings/16/Papers/615.pdf

Cheng, L. and T. Yu (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.

Conati, C., K. Porayska-Pomsta and M. Mavrikis (2018). "AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling." arXiv preprint arXiv:1807.00154.

Cruz-Martín, A., Fernández-Madrigal, J. A., Galindo, C., González-Jiménez, J., Stockmans-Daou, C., & Blanco-Claraco, J. L. (2012). A Lego Mindstorms NXT approach for teaching at data acquisition, control systems engineering and real-time systems undergraduate courses. *Computers & Education, 59*(3), 974–988. https://doi.org/10.1016/j.compedu.2012.03.026

Demertzi *et al*. Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*. Published online February 6, 2019. doi:10.1126/sciadv.aat7603.

EU GDPR (2018) Article 13: Information to be provided where personal data are collected from the data subject. https://web.archive.org/web/20170801183401/http://www.privacy-regulation.eu/en/13.htm

Faust, O., Y. Hagiwara, T. J. Hong, O. S. Lih and U. R. Acharya (2018). "Deep learning for healthcare applications based on physiological signals: A review." Computer Methods and Programs in Biomedicine **161**: 1-13.

FITEE editorial staff (2018). "An interview with Dr. Raj Reddy on artificial intelligence." Frontiers of Information Technology & Electronic Engineering **19**(1): 3-5.

Floridi, L (2014) The Fourth Revolution: How the Infosphere is Reshaping Human Reality, Oxford: OUP.

Frank, E., Hall, M. A., & Witten, I. H. (2016). *The WEKA Workbench. Online Appendix, data mining: Practical machine learning tools and Techniques.* Fourth Edition.

Grace, K., J. Salvatier, A. Dafoe, B. Zhang and O. Evans (2018). "When will AI exceed human performance? Evidence from AI experts." Journal of Artificial Intelligence Research **62**: 729-754.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational researcher, 40*(5), 223-234. https://doi.org/10.3102/0013189X11413260.

Gray, C. C., & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education*, *131*, 22-32. https://doi.org/10.1016/j.compedu.2018.12.006

Hayashi, Y. (2019). "The Right Direction Needed to Develop White-Box Deep Learning in Radiology, Pathology, and Ophthalmology: A Short Review." Frontiers in Robotics and AI **6**(24).

Ifenthaler, D., S. Greiff and D. Gibson (2018). "Making use of data for assessments: Harnessing analytics and data science." Second Handbook of Information Technology in Primary and Secondary Education: 1-16.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R.* Corrected edition. New York: Springer.

Joshi, M. V. (2002). On evaluating performance of classifiers for rare classes. In *2002 IEEE International Conference on Data Mining, 2002. Proceedings.* (pp. 641-644). IEEE. https://doi.org/10.1109/ICDM.2002.1184018

Kaltenegger, L. and O'Malley-James, J. (2019). Lessons from early Earth: UV surface radiation should not limit the habitability of active M star systems. *Monthly Notices of the Royal Astronomical Society* Volume 485, Issue 4, June 2019, Pages 5598–5603, https://doi.org/10.1093/mnras/stz724.

Kaplan, A. and M. Haenlein (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.

Karl, L. (1997). DNA computing: arrival of biological mathematics. The mathematical intelligencer, 19(2), 9-22.

Katal, A., Wazid, M. and Goudar, R.H. (2013) Big Data: Issues, Challenges, Tools and Good Practices. 2013 Sixth International Conference on Contemporary Computing (IC3), Noida, August 2013, 404-409. http://dx.doi.org/10.1109/ic3.2013.6612229

Kittler, J. (1986). Feature selection and extraction. In Young, T.Y., and Fu, K.S. (Eds.), *Handbook of Pattern Recognition and Image Processing*, Academic Press, 59-83.

Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., & Dawson, S. (2018, March). Understand students' self-reflections through learning analytics. In *Proceedings of the*

8th international conference on learning analytics and knowledge (pp. 389-398). https://doi.org/10.1145/3170358.3170374

Kudo, M., & Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition, 33*(1), 25-41. https://doi.org/10.1016/S0031-3203(99)00041-2.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Penguin.

Lazendic, G., Justus, J-A., & Rabinowitz, S. (2018). NAPLAN online automated scoring research program: research report. Australian Curriculum Assessment and Reporting Authority. 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 volume 521, pages 436–444 https://doi.org/10.1038/nature14539

Liakos, K. G., P. Busato, D. Moshou, S. Pearson and D. Bochtis (2018). "Machine Learning in Agriculture: A Review." Sensors **18**(8): 2674.

Livieris, I. E., K. Drakopoulou, V. T. Tampakas, T. A. Mikropoulos and P. Pintelas (2019). "Predicting Secondary School Students' Performance Utilizing a Semi-supervised Learning Approach." Journal of Educational Computing Research **57**(2): 448-470.

Mansell, J., Spencer, D., Plante, B., Diaz, A., Bellardo, J. and Betts, B. (2019): Orbit Raising and Attitude Performance of the LightSail 2 Solar Sail Spacecraft. Paper submitted to the 70th International Astronautical Congress, Washington, D.C. https://planetary.s3.amazonaws.com/projects/lightsail/papers/Mansell-et-al-2019-orbit-raising.pdf

Martínez-Tenor, Á., Cruz-Martín, A., & Fernández-Madrigal, J. A. (2019). Teaching machine learning in robotics interactively: the case of reinforcement learning with Lego® Mindstorms. *Interactive Learning Environments*, *27*(3), 293-306. https://doi.org/10.1080/10494820.2018.1525411

Mitchell, T. M. (2006). The discipline of machine learning, Carnegie Mellon University, School of Computer Science, Machine Learning ….

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), p 2053951716679679. doi:10.1177/2053951716679679

Monroe, D. (2014). Neuromorphic computing gets ready for the (really) big time. *Communications of the ACM*, 57(6), 13-15.

Muggleton, S. H. (2006). 2020 Computing: Exceeding human limits. *Nature*, 440(7083), 409.

Murphy, R. (2019). Artificial Intelligence Applications to Support K–12 Teachers and Teaching: A Review of Promising Applications, Opportunities, and Challenges. Perspective. RAND corporation. https://doi.org/10.7249/PE315

Obeid, C., I. Lahoud, H. E. Khoury and P.-A. Champin (2018). Ontology-based Recommender System in Higher Education. Companion Proceedings of the The Web Conference 2018. Lyon, France, International World Wide Web Conferences Steering Committee**:** 1031-1034.

Parisi, G. I., R. Kemker, J. L. Part, C. Kanan and S. Wermter (2019). "Continual lifelong learning with neural networks: A review." Neural Networks **113**: 54-71.

PISA (2017). PISA 2015 Collaborative Problem-Solving Framework, Organisation for Economic Co-operation and Development (OECD).

Portugal, I., P. Alencar and D. Cowan (2018). "The use of machine learning algorithms in recommender systems: A systematic review." Expert Systems with Applications **97**: 205-227.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Rahwan, I., M. Cebrian, N. Obradovich, J. Bongard, J.-F. Bonnefon, C. Breazeal, J. W. Crandall, N. A. Christakis, I. D. Couzin, M. O. Jackson, N. R. Jennings, E. Kamar, I. M. Kloumann, H. Larochelle, D. Lazer, R. McElreath, A. Mislove, D. C. Parkes, A. S. Pentland, M. E. Roberts, A. Shariff, J. B. Tenenbaum and M. Wellman (2019). "Machine behaviour." Nature **568**(7753): 477-486.

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.

Salas, E. E., & Fiore, S. M. (2004). Team cognition: Understanding the factors that drive process and performance. *American Psychological Association*.

Shannon, C.E. and Weaver, W. (1964). The mathematical theory of communication. The university of Illinois press: Ubana.

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

Sze, V., Y. Chen, T. Yang and J. S. Emer (2017). "Efficient Processing of Deep Neural Networks: A Tutorial and Survey." Proceedings of the IEEE **105**(12): 2295-2329.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24-54. https://doi.org/10.1177/0261927X09351676

Touretzky, D., F. Martin, D. Seehorn, C. Breazeal and T. Posner (2019). Special Session: AI for K-12 Guidelines Initiative. Proceedings of the 50th ACM Technical Symposium on Computer Science Education, ACM.

Tuomi, I. (2018). The Impact of Artificial Intelligence on Learning, Teaching, and Education: Policies for the Future.

Valiant L.G. (1993) A View of Computational Learning Theory. In: Meyrowitz A.L., Chipman S. (eds) Foundations of Knowledge Acquisition. The Springer International Series in Engineering and Computer Science, vol 195. Springer, Boston, MA. https://doi.org/10.1007/978-0-585-27366-2_8.

Wang, J. and Q. Tao (2008). "Machine Learning: The State of the Art." IEEE Intelligent Systems **23**(6): 49-55.

Webb, M. E. (2019). Curricula in Computer Science. Encyclopedia of Education and Information Technologies. A. Tatnall. Cham, Springer International Publishing**:** 1-7.

Webb, M. E. and D. C. Gibson (2015). "Technology enhanced assessment in complex collaborative settings." Education and Information Technologies **20**(4): 675-695.

Webb, M., N. Davis, T. Bell, Y. J. Katz, N. Reynolds, D. P. Chambers and M. M. Sysło (2017). "Computer science in K-12 school curricula of the 2lst century: Why, what and when?" Education and Information Technologies **22**(2): 445-468.

Whitelock, D. and D. Bektik (2018). "Progress and challenges for automated scoring and feedback systems for large-scale assessments." Second Handbook of Information Technology in Primary and Secondary Education: 617.

Whitney, A. W. (1971). A direct method of nonparametric measurement selection. *IEEE Transactions on Computers C, 20*(9), 1100–1103. https://doi.org/10.1109/T-C.1971.223410.

Wiggers, K. (2019). Google begins selling the $150 Coral Dev Board, a hardware kit for accelerated AI edge computing. https://venturebeat.com/2019/03/06/google-begins-selling-the-150-coral-dev-board-a-hardware-kit-for-accelerated-ai-edge-computing.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4), 625-636.

Zhang, Q., L. T. Yang, Z. Chen and P. Li (2018). "A survey on deep learning for big data." Information Fusion **42**: 146-157.

Zhao, R., R. Yan, Z. Chen, K. Mao, P. Wang and R. X. Gao (2019). "Deep learning and its applications to machine health monitoring." Mechanical Systems and Signal Processing **115**: 213-237.

Additional reading – Key papers

These papers are particularly useful to read or skim prior to discussions at EDUsummIT.

Cave, S. & ÓhÉigeartaigh, S.S. (2019) Bridging near- and long-term concerns about AI. *Nature machine intelligence* 1, pp.5-6.

Ifenthaler, D., S. Greiff and D. Gibson (2018). "Making use of data for assessments: Harnessing analytics and data science." Second Handbook of Information Technology in Primary and Secondary Education: 1-16.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. Nature volume 521, pages 436–444 https://doi.org/10.1038/nature14539

Whitelock, D. and D. Bektik (2018). "Progress and challenges for automated scoring and feedback systems for large-scale assessments." Second Handbook of Information Technology in Primary and Secondary Education: 617.

Further reading

Ada Lovelace Institute & Leverhulme centre for the future of intelligence (2019) *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research Summary report*. Nuffield Foundation.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Damacharla, P. et al. (2019) Effects of voice-based synthetic assistant on performance of emergency care provider in training. *Int J Artif Intell Educ* 29:122–143. https://doi.org/10.1007/s40593-018-0166-3

DeFalco, J.A. et al. (2018) Detecting and Addressing Frustration in a Serious Game for Military Training. *Int J Artif Intell Educ* 28:152–193. DOI 10.1007/s40593-017-0152-1

Dermeval, D. et al. (2018) Authoring Tools for Designing Intelligent Tutoring Systems: a Systematic Review of the Literature. *Int J Artif Intell Educ* 28:336–384. https://doi.org/10.1007/s40593-017-0157-9

Filius, R.M. et al. (2018) Promoting deep learning through online feedback in SPOCs. *Frontline learning research* 6(2)92-113. https://doi.org/10.14786/flr.v6i2.350

Fletcher, J.D. & Sottilare, R.A. (2018) Shared mental models in support of adaptive instruction for teams using the GIFT tutoring architecture. *Int J Artif Intell Educ* 28:265–285. DOI 10.1007/s40593-017-0147-y

Gonzalez, A.J. et al. (2017). AI in informal science education: Bringing Turing back to life to perform the Turing test. *Int J Artif Intell Educ* 27:353–384. DOI 10.1007/s40593-017-0144-1

Gunning, D. (2017). *Explainable artificial intelligence (XAI): DARPA/I2O - Program update November 2017*. DARPA.

Pinkwart, N. (2016). Another 25 years of AIED? Challenges and opportunities for intelligent educational technologies of the future. *Int J Artif Intell Educ* 26:771–783. DOI 10.1007/s40593-016-0099-7.

Roll, I. & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *Int J Artif Intell Educ* 26:582–599. DOI 10.1007/s40593-016-0110-3.

Snow, E. et al. (2016). Taking control: Stealth assessment of deterministic behaviors within a game-based system. *Int J Artif Intell Educ* 26:1011–1032. DOI 10.1007/s40593-015-0085-5.

Spector, JM and Ifenthaler, D and Sampson, C and Yang, LJ and Mukama, E and Warusavitarana, A and Dona, KL and Eichhorn, K and Fluck, A and Huang, R and Bridges, S and Lu, J and Ren, Y and Gui, X and Deneen, CC and San Diego, J and Gibson, DC (2016). Technology enhanced formative assessment for 21st Century learning. *Educational Technology and Society*, **19** (3) 57-71.