Organisation for Economic Co-operation and Development

EDU/EDPC(2018)45/ANN2

Unclassified **English text only**

5 July 2021

DIRECTORATE FOR EDUCATION AND SKILLS
EDUCATION POLICY COMMITTEE
EDUCATION POLICY COMMITTEE

Cancels & replaces the same document of 24 October 2018

Future of Education and Skills 2030: Conceptual Learning Framework

Education and AI: preparing for the future & AI, Attitudes and Values

8th Informal Working Group (IWG) Meeting
29-31 October 2018
OECD Conference Centre, Paris, France

This document includes two draft papers:

The first draft paper was written by Rose LUCKIN and Kim ISSROFF from the University College London. United Kingdom. This is a literature review on knowledge and skills and the relationship to Artificial Intelligence (AI). This paper describes the knowledge and skills that remain for human in a time of increasing AI. The 2nd draft paper was written by Keith MILLER, Marvin BERKOWITZ and Melinda C BIER from University of Missouri – St Louis, USA. This is a literature review on artificial intelligence, attitudes and values and ethics. This paper describes attitudes and values that have become more important in the time of increasing AI
This is still a **"working document"**.

For ACTION: participants are invited to **COMMENT** before 5 November 2018.

Miho TAGUMA, Senior Analyst, Miho.Taguma@oecd.org
Eva FERON, Analyst, Eva.Feron@oecd.org
Meow Hwee LIM, Consultant, meowrena@gmail.com

JT03479206

Abstract

In this paper, we review the existing literature to find out what knowledge and skills will remain for human in a time of increasing AI. We address some of the issues surrounding the use of AI in education, and we discuss how AI can be harnessed to improve the education and opportunities of students as they prepare to enter the workforce. We also stress the need for students, employees and society to develop the awareness and understanding that they will need in order to be effective, engaged and active citizens in a world in which AI will play an increasing role.

Introduction

The potential of AI in education has been researched, debated and discussed for nearly 30 years within the AIED academic community. Over the last few years, the debate has crept into the international public policy arena as data, sophisticated Ai algorithms that learn, computing power and access to technology has increased across the world. There are great potential benefits, but, there are risks as well as opportunities with AI in and for education. We therefore need to proceed diligently and prudently into a new educational environment where AI is used to support learners and teachers, and where we also prepare learners for a future world in which AI plays an increasing role.

"The risk is that the education system will be churning out humans who are no more than second-rate computers, so if the focus of education continues to be on transferring explicit knowledge across the generations, we will be in trouble.

The AI challenge is not just about educating more AI and computer experts, although that is important. It is also about building skills that AI cannot emulate. These are essential human skills such as teamwork, leadership, listening, staying positive, dealing with people and managing crises and conflict."

Financial Times (2017)

Some experts believe there will be a wholesale revolution in the nature of our education systems. For example, Seldon & Abidoye (2018) in a carefully researched book "The Fourth Education Revolution" analyse the role of AI and the changes that they see as inevitable in our education system. They claim that with the advent of AI, *'Barely a single facet of this education system will remain unchanged,"*

In March 2018, The Future of Education and Skills 2030 OECD project published a position paper in which it proposed an initial framework designed to help countries address two key questions: What knowledge, skills, attitudes and values will today's students need to thrive and shape their world; and How can instructional systems develop these knowledge, skills, attitudes and values effectively? The initial OECD framework is based upon a synthesis of a broad literature, and has been reviewed, tested and validated by stakeholders from around the world. The interim framework is grounded in an appreciation that societies face environmental, economic and social challenges. As a result, educational goals must be broader and they must drive individual and collective well-being. In order to ensure that all students are well prepared for the future and that they themselves can act as change agents, the interim framework identifies a broad set of knowledge, skills, attitudes and values. In particular, the interim Framework builds on previous work conducted by the OECD DeSeCo project and identifies three additional categories of competencies, also known as the "Transformative Competencies": Creating new value, reconciling tensions and dilemmas and taking responsibility.

It is interesting to consider this initial framework alongside other work that has tackled similar goals. For example, in an expert roundtable summarised by Fadel (2014), the consensus was that active citizens of the future need to be versatile as we cannot predict how and what technologies will dominate our future learning and work environments.

"We … need to replace old education standards still in general use with an educational framework that combines the acquisition of traditional knowledge with the 21st century kills of creativity, critical thinking, communication and collaboration. We will need to teach both skills and character, in addition to knowledge, with a focus on 'metacognition' which includes 'learning how to learn'. Precisely because we cannot predict what technologies will be ascendant in the future, we have to teach ourselves and our children to be versatile."

In the following sections of this report, we address some of the issues surrounding the use of AI in education, and we discuss how AI can be harnessed to improve the education and opportunities of students as they prepare to enter the workforce. We also stress the need for students, employees and society to develop the awareness and understanding that they will need in order to be effective, engaged and active citizens in a world in which AI will play an increasing role.

History and background of AI in education

Review of existing evidence

The development and application of Artificial Intelligence (AI) is bringing imminent and rapid change to almost every aspect of life, with today's children experiencing a very different life to that of their parents (Siraj, 2017). To prepare people for the anticipated changes to their lives, we must ensure that our education and training is tuned to the new demands of the workplace and society (Tucker, 2017). The landscape we must navigate is likely to be bumpy (Walsh, 2017) and there will be significant challenges, not least, those relating to ethics. Fundamental to success will be unpacking our relationship with the concept of Intelligence (Luckin, 2017). In thinking about how AI will impact on education and what sorts of knowledge and skills future citizens will need, we therefore need to look beyond the current trends towards trying to identify the jobs and skills that the world will require to the core issue of what it means to be intelligent in an AI augmented world. However, there is value in synthesizing what experts have investigated with respect to how susceptible jobs are to computerisation, because this is an important element of the context within which we need to reconceptualise human intelligence.

A seminal report from Frey & Osborne (2013) examined 702 detailed occupations, using Machine Learning AI, in the form of a Gaussian process classifier and found that 47 percent of total US employment is at risk and that wages and educational attainment exhibit a strong negative relationship with an occupation's probability of computerisation.

In a second report in 2017, in the same authors concluded that:

"In short, our findings suggest that recent development in Machine Learning will put a substantial share of employment, across a wide range of occupations, at risk in the near future. According to our estimates, however, this wave of automation will be followed by a subsequent slowdown in computers for labour substitution, due to persisting inhibiting engineering bottlenecks to computerisation." Frey and Osborne, 2017, pg 266.

The bottlenecks are discussed in detail:

1. Limitations of mobile robotics on perceptual and manipulation tasks
2. Creative intelligence tasks which AI and machine learning cannot currently achieve.
3. Social intelligence tasks (the challenge of real time recognition of human emotion and how to respond intelligently to these.)

For Frey and Osborne, the change in the nature of the job market will be rapid and then slow down. However, Fadel (2014) from a roundtable of experts made 6 predictions which provide some information about the sorts of jobs that may increase in the future:

1. Routine tasks will remain the most automatable, but some facets of innovation and creativity may be automatable.
2. Complete adoption of technologies generally takes longer than anticipated but may be deeper than first assumed

3. Robust occupations are those with challenges, new discoveries, new performances and new things to be learnt and shared
4. T shaped occupations, requiring both depth and breadth will see an increase in demand
5. A top down review will not be able to predict future job patterns. This will have to come from sector by sector analysis.
6. There are many and variable parameters which interact with one another which need to be considered in order to predict future jobs.

In the World Economic Forum report "The Future of Jobs", 2016 warns that we need to take urgent action to ensure that we are prepared for the changes to our workplaces.

"During previous industrial revolutions, it often took decades to build the training systems and labour market institutions needed to develop major new skill sets on a large scale. Given the upcoming pace and scale of disruption brought about by the Fourth Industrial Revolution, however, this simply not an option."

They discuss the different drivers of change – demographic, socio-economic and technological – and the impact of these on different job families. Their respondents predicted strong employment growth across the architecture and engineering and Computer and mathematical job families, a moderate decline in manufacturing and production roles and a significant decline in office and administrative roles.

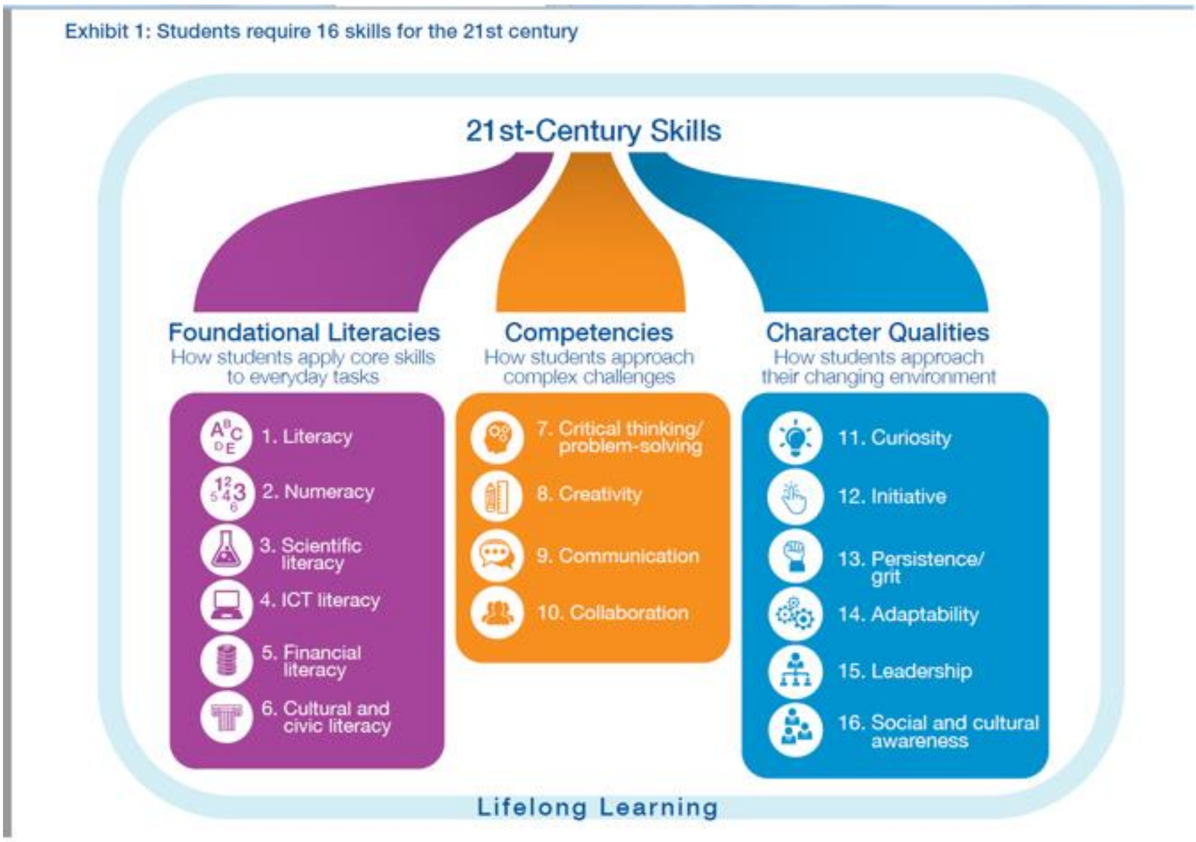
In another WEF report, "New Vision for Education: Fostering social and Emotional Learning through Technology", 2015, the authors discuss the sorts of skills that students will need in the future and argue for the importance of social and emotional learning:

"*To thrive in the 21st century, students need more than traditional academic learning. They must be adept at collaboration, communication and problem solving, which are some of the skills developed through social and emotional learning (SEL). Coupled with mastery of traditional skills, social and emotional proficiency will equip students to succeed in the swiftly evolving digital economy.*'

In a similar vein, Trilling and Fadel (2009) categorised 21st century skills into 3 groups

1. Learning and innovation skills
2. Digital literacy skills
3. Career and life skills

Figure 1. 21st century skills



These slightly different versions of the future skills, abilities, competencies and characteristics build on early ground work, such as that by Oates (2002, 2003), which identified the need to specify key competencies for the future. They also reflect an increasing interest in and appreciation of the importance of more than academic knowledge and skills. The growing range of frameworks for the future, reflect increasing concern with personal development as well as subject knowledge and skills (James et al. eds., 2011), metacognition (Tarricone, 2011), social and emotional development (see for example, Goodman et al., 2015), ad well-being (See for example, Gregory & Sadeh, 2012) and what are often referred to as soft skills (Lippman et al. 2015)

NESTA is a British charity which works worldwide as a global innovation foundation. In its 2018 on digital skills report, Djumalieva and Sleeman report propose the following 16 skills for the 21st century. They divide these into 3 classifications: Foundational Literacies, Competences and Character Qualities.

In another NESTA report, Schneider and Bakhshi (2017) discuss future skills and argue that the future of work is not only influenced by automation, but also by key trends in environmental sustainability, urbanisation, increasing inequality, political uncertainty, technological change, globalisation and demographic change. They say there will be a strong emphasis on interpersonal skills, higher order cognitive skills and systems skills.

Originality, fluency of ideas and active learning are very important. A future workforce will need broad based knowledge as well as specialist feature for specific occupations.

Their findings were similar to those of the WEF 2016 report in that the "digital skills most likely to be needed in the future are ones that are used in non-routine tasks, problem-solving and the creation of digital outputs. On the other hand, the digital skills that are linked to occupations least likely to grow tended to relate to the use of software for administrative purposes."

Reconceptualising Human Intelligence

A contrasting approach is taken by authors like Gardner (2007) and Luckin (2018) have offered alternative ways of conceptualising Intelligence to suit the modern world. Gardner suggests that we need five sorts of intelligence, or mental capacities: The Disciplinary Mind: to master academic subject knowledge, such as science, mathematics, and history, as well as at least one professional craft; The Synthesizing Mind: to enable us to integrate ideas from different disciplines into a coherent whole and to communicate this synthesized, integrated understanding to others; The Creating Mind: which imbues us with the capability to uncover and clarify new problems, questions, and phenomena; The Respectful Mind: that ensure that we are aware of and appreciate the differences among and between humans (and possibly AIs in the future); The Ethical Mind: a vital components to ensure that we fulfil our responsibilities as both a worker and a citizen. These five minds offer thought provocations for how education needs to be revised.

Luckin, in her book about machine learning and human intelligence focuses specifically on how we need to reconsider human intelligence, because of the sophisticated AI that now permeates much of society across the globe. She describes intelligence as "aligned with intellect, with complex cognitive processes, with the understanding of the knowledge, skills and abilities both of others and of ourselves. It is our intelligence that enables us to learn, to apply our knowledge, to synthesise what we know in order to solve problems, to communicate with otters, to make decisions to think, to express and t learn from experience. It is certainly about a great deal more than what we learn in school."

Luckin argues for 7 elements to human intelligence:

1. Interdisciplinary Academic intelligence
2. Social intelligence
3. Metaknowing intelligence
4. Metacognitive intelligence
5. Metasubjective intelligence
6. Meta-contextual intelligence
7. Perceived self-efficacy

She argues that these are interwoven and that AI currently only contributes to academic intelligence. However, current Ai systems excel in this area -knowledge and understanding that is multidisciplinary and interdisciplinary. However, AI has little to offer in the other areas of intelligence. For example, when a system interprets people's faces and discerns their sexuality in a way that humans don't understand e.g. Kuang (2017), Luckin suggests that this is because the machine does not have access to the contextualised and subjective knowledge that humans have and is embodied in their metasubjective and meta-contextual intelligence. AI systems do not understand themselves, cannot explain or justify their decisions and have no self-awareness.

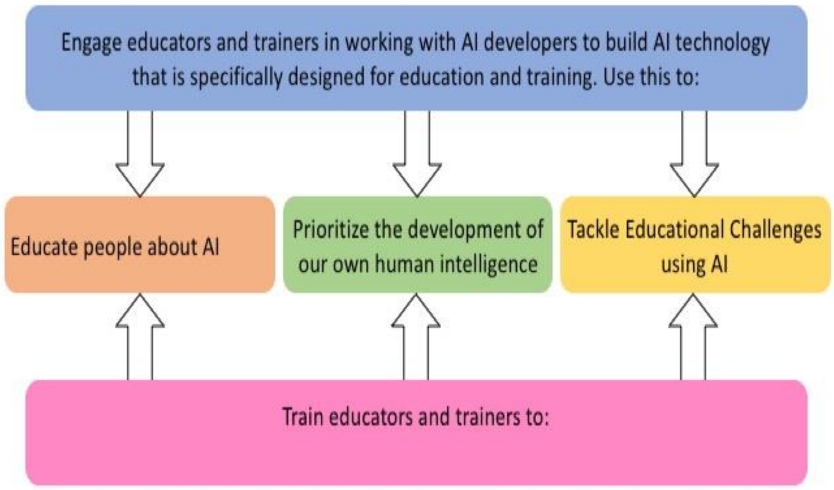
Luckin's approach to understanding human intelligence and the role and potential of AI in education contrasts with the many reports reporting on skills that will or will not be needed in the future as AI has an increasing impact on our lives. In terms of the OECD Learning Framework 2030, it bridges the left hand side and encompasses competencies as well as knowledge, skills, attitudes and values.

Knowledge, skills, attitudes and values

Human Intelligence (HI) as the Focus of attention

Now is the time when we must agree on what we want from our Human Intelligence (HI), how best this HI can work with AI, how AI and HI should complement each other and as a consequence what new knowledge and skills we must focus our attention upon. This is the key step we need to take if we wish to make informed decisions about the existing knowledge and skills that will remain important in our future AI augmented society, and to differentiate these from the knowledge and skills that we no longer need to attend to within our education and training systems. The beauty of the current situation is that because we have built highly sophisticated AI that can learn, we have also built AI that can help us develop far more sophisticated HI.

Figure 2. An intelligent approach to AI for Education and Training



An intelligent approach to AI for Education and Training

Figure 2 outlines at a high level the priorities that we need to address in order to synthesize the best of AI and HI.

Tackling Educational Challenges using Artificial Intelligence:

AI to support knowledge and skill development

The potential for AI in the teaching and development of knowledge cannot be underestimated. We distinguish between academic knowledge as distinct from social knowledge, by which we mean knowledge about the world. This is embodied in Luckin's notion of academic intelligence and is knowledge and understanding that is both multi and interdisciplinary.

It is impossible to separate out knowledge from skills and it is important to recognise their interdependence. Learners cannot use and exhibit their skills without applying them to knowledge and knowledge is only effective for learners in so far as they have the skills to use them.

However, it is useful in education terms to separate out how AI can help learners and teachers. This section is therefore divided into two sections: learners and teachers. We do not address how AI can help school leaders and administrators.

Students: Helping everyone to know

AI systems can readily acquire and process knowledge. They can rapidly build up vast representations of bodies of knowledge and these can be harnessed to help us to develop our knowledge and to learn facts. This is easiest in well-defined subject areas, such as science and maths but can also be harnessed in other disciplines which require knowledge of non-contextualised facts.

The bodies of knowledge embedded in AI systems can then be used to help students. They can be presented to students in a variety of modes for example as text augmented with pictures or audio or video. The AI system can time the presentation of the knowledge and therefore pace the learners experience as well as vary the order that the resources are presented to learners in order to maximise learning or meet some other criteria which is embedded in the design of the system.

The opportunities that this creates for learning knowledge are infinite and allows for a vast range of pedagogical actions. These can be personalised to the individual student.

Diminishing gaps between different socioeconomic groups

Ensuring equity of access to knowledge for different subgroups of our population is a major goal for education. AI systems can help to bridge this gap by providing access to knowledge as described above for those who may not have the resources either in their homes or their communities or their schools.

Providing Access to knowledge and information for disabled students and those with additional educational needs

Sophisticated AI systems have been developed to provide a range of interfaces to knowledge for students who have disabilities. For example, natural language processing software enable students with physical disabilities to use voice activation to access devices. Specialised systems have been designed to help learners with additional educational needs for example, Grammarly is an AI-powered browser plug-in designed to support people with dyslexia when writing.

Personalised learning

AI systems are designed to not only store knowledge about the domain they are teaching but also have representations of the learners who use the system. This includes the knowledge and information that has been presented to them as well as information about what the AI system thinks the learner knows (this data is gathered by the system as it records the actions of the learner in response to the system). By reasoning with the information, the system has about the learner and the information it has about the knowledge, the system can provide personalised experiences for learners which aim to help them develop the knowledge that the system knows they do not have.

Individualised feedback

AI systems store information about learners and can give feedback to learners about how they are doing both in terms of whether they have got something correct or not.

Freeing up human teachers to work with learners on other things

As AI systems are used increasingly to provide personalised learning experiences for students about academic knowledge, teachers time will be freed up as they do not need to use their time to teach this type of knowledge.

Repetition, drill and practice

Learning often requires repetition in order to consolidate knowledge or skills. AI systems can readily supply endless examples (without losing patience!) and therefore be used to consolidate knowledge and skills both for learners who are struggling and have not yet consolidated their knowledge or mastered a skill and for those who wish to practice.

Supporting Collaboration

When students work collaboratively, Ai systems can be used to monitor some indicators of collaboration and can therefore be used to both monitor and manage collaborative working and learning.

Teachers

Changing the nature of what is taught by teachers.

As AI systems are developed that store wider bodies of knowledge and enable students to practice a range of skills, teachers time can be freed up to work with learners on other aspects of their development to enable them to become active and engaged citizens. Teachers will be able to work with learners to develop the six other types of inter related intelligences, secure in the knowledge that AI will ensure that all their students have the academic knowledge and skills that they need.

Assessing and monitoring student progress

AI systems store information about what students know and can do and what information and resources have been presented to students as well as the skills they have practiced and mastered. Teachers can readily have access to this information which can be presented for individuals, particular groups and the whole cohort. It can be used for a variety of purposes which can save time if used effectively.

Teachers need to know what their students know and can do. It is important for teachers to monitor their students' progress in order to ensure that their teaching is effective. The information can be used for reporting purposes – to parents and school leaders, as the basis for conversations with learners to motivate them and engage them in their learning as well as developing their confidence in their own knowledge and skills.

Planning future teaching and interventions

On the basis of the information that teachers have about individuals and groups of learners, teachers can plan their future teaching, for example, to address gaps in knowledge or skills for an entire cohort of students or to plan interventions for individual children. They can also plan their teaching so as to move onto different areas when their students are ready.

Educating People about AI and Digital Technology

The basic AI concepts

It is crucial that citizens of the future have a basic understanding of AI concepts so that they can engage both effectively and critically with AI systems which are becoming increasingly pervasive in our daily lives.

Although there are concerns that the curriculum is overcrowded, in the future AI will become part of the existing computing/ICT/computer science curricula as they evolve.

An example of this is provided by AIinSchools http://aiinschools.com which aims to demystify AI and provides resources for schools. An example of a curriculum for year 9 children (13/14 years old is shown below.

Figure 3. Example of curriculum for year 9

Theme	**Lesson 1** **Introduction to Artificial Intelligence**	**Lesson 2** **Introduction to Neural Networks**	**Lesson 3** **The MATHS behind a neural network**	**Lesson 4** **Image Recognition**	**Lesson 5** **Different types of AI**	**Lesson 6** **Assessment**
Overview	In this lesson pupils are introduced to the "**Smart City**" and gain a basic understanding of "**Artificial Intelligence**" Pupils identify and explore the current use of AI in specific areas such as • Internet Services • Education • Medicine • Media & Entertainment • Helping the Elderly/Disabled	In this lesson pupils, will be introduced to key terminology around the field of Artificial Intelligence, Machine and Deep Learning. They will gain an overview of the **Turing Test** and **Convolutional Neural Networks**.	In this lesson pupils, focus on some of the Mathematics used in a *neural network*.	In this lesson, pupils will access remote **GPU**s via a cloud server (Amazon Web Services) to access **DIGITS** (*a front-end tool of common AI frameworks*) and use images from the web against those from a known dataset to accurately identify images.	This lesson will get pupils to investigate how **AI** will change our futures. Areas to be investigated are: •Agriculture •Shopping (automated services) •Cooking •Transport including NVIDIA driverless cars (roads) •Medical diagnosis •Assembly line	In this lesson pupils, will Review their learning through completion of a short **assessment**. Pupils will also collaborate as a class and build a "Wall of fame on women in AI"

Digital literacy

There are a multitude of definitions of digital literacy. Some authors have advocated the notion of multiple literacies including digital literacy, emotional literacy, economic literacy and spiritual literacy and there has been a long-standing research area on emergent literacy. (for a recent overview see Buckingham, D., 2018).

A commonly used definition is from Cornell University – “the ability to find, evaluate, utilize, share, and create content using information technologies and the Internet.” https://digitalliteracy.cornell.edu/

For our purposes, and especially in relation to developments in AI and robotics, we would not want to limit digital literacy to information technologies and the internet. We want our students to be engage with content in an effective, critical and creative way regardless of how and what that information is embodied in.

Data literacy

Data literacy is intimately connected with digital literacy but is particularly pertinent as AI becomes more pervasive in our society. In a detailed report on how to teach data literacy from Dalhousie University in Canada, Ridsdale et al. (2015) define it as “the ability to collect, manage, evaluate, and apply data, in a critical manner.” Students and teachers need opportunities to engage with data, manipulate and evaluate it and use it in their learning and working lives as well as being shown how to approach it critically in order to make effective use of data and to be aware of how our data is collected and used organisations.

Online safety

In a world where people increasingly live virtual lives alongside their physical selves, safety in online environments is crucial for future citizens. In order to keep our children and adults safe, they need to be aware of what is happening to them and their data, understand when they are vulnerable and have the skills to evaluate and report problems if they arise. AI has a role to play in this, both in the contribution systems may make to unsafe environments, but also in the potential there is for AI to monitor and report on unsafe behaviours.

Basic AI programming

An understanding of basic AI programming is important to help students and teacher to understand how AI systems work and to help them to reflect on their use of AI. Kahn and Winters (2018) have developed programming constructs that are suitable for use by beginners for speech synthesis, speech recognition, image recognition, and machine learning

Ethics of AI (see section xx)

All citizens will need to have a basic understanding of the ethics of AI in order to evaluate their own and others use of AI systems. This is discussed in more detail in section xx

Some people need to know how to build AI systems

Clearly as our society becomes more reliant on AI systems, we will need a workforce to develop and test these systems. This will require not just in depth understanding of programming and AI but also design professionals and in particular, people who understand how to evaluate the impact of these systems on our society, particularly in terms of learning

Work effectively with AI

In order to work effectively with AI, we need to understand not just the opportunities that AI systems create, but also the limitations of systems.

Developing the attitudes and values towards AI for citizens in 2030

Knowledgeable

As previously discussed, active citizens need to understand the basics of AI to enable them to work with AI systems and harness them for good.

Engaged

Engagement is crucial to learning and AI has huge potential to keep learners engaged with the knowledge that they are using their skills with. As AI can easily personalise the learning experience, systems can choose examples, levels of difficulty and timing that are appropriate for a particular learner to keep them motivated and engaged.

Critical

Being critical is an important skill and enables us to think about what we are doing and understand the boundaries and extent to which AI systems can be used and should (or should not) be trusted.

Ethical

Being ethical about our use of AI systems is crucial to their integration and effective use in our lives. Perhaps the greatest danger to our ability to harness AI for good use in our future society is unethical use of AI – be that to gain personal data for nefarious purposes or to enable people to access inappropriate or danger information, knowledge and skills.

While the ethical imperative is greatest for those who are designing, implementing and evaluating AI systems, an ethical attitude to AI is also essential for everyday users, who need to be able to evaluate systems, have knowledge of what is legal and illegal and have the capacity to decide when it is inappropriate to use AI systems and when to report unethical and/or dangerous systems so that other people are kept safe.

Course Example

There are very few courses on AI that are suitable for secondary school children. However, we have identified the following in addition to the AIinSchools course discussed in section xx.

1. Apps for Good (https://www.appsforgood.org/public/about-us) provide free courses for teachers to use in their classroom. The courses are designed to help 10-18 year-olds to build mobile apps and IoT (internet of things) products. Students work together in teams to find issues they care about. Students go through all key aspects of new product development, from idea generation, technical feasibility and programming to product design, deciding on business models and marketing.

2. SPARC is a Summer Program in Applied Rationality and Cognition (SPARC) set up by a group of AI researchers to help gifted teenagers understand and engage with the complexities of AI. machine learning, AI safety, and existential risks. The programme also supports teenagers in developing social skills.

3. CS4FN is an online magazine for teenagers which includes a range of articles and examples about AI http://www.cs4fn.org/ai/

4. Sapere is a British organisation which aims to teach Philosophy 4 Children (P4C). In P4C the teacher facilitates a student-led discussion on a philosophical question. P4C builds higher order thinking, questioning, speaking and listening skills. They train teachers and some of the examples they use are relevant to AI and can be used as an introduction to the ethical issues surrounding AI.

Fostering and expanding our Human Intelligence

Learning is the holy grail of success and intelligence. If we are good at learning, the world is our oyster and we can continually progress. Learning is also what sets modern AI aside from the earlier Good Old-Fashioned AI (GOFAI). The reason that AlphaGo beat Master Go player Lee Sedol in March 2016 (Wikipedia, n.d.), is because AlphaGo was phenomenally good at learning. If we are to foster and further expand our Human Intelligence, then our ability to learn is the key to our success.

Machines can learn thanks to AI, and they can learn faster and they can recall what they have learnt more accurately, than humans. However, this learning is currently focussed within the sphere of academic intelligence: knowledge about the world. Machines can mimic some of the features of other elements from the human intelligence, such as emotions, but they feel no emotions, and have no awareness of the subjective experience of any emotions. Societies must fulfil their responsibility to their members by designing and implementing education systems that effectively develop people's human intelligence. To achieve this, education systems need progression models that constantly promote growth across and between all seven intelligence elements. Embracing the AI augmented world is not simple however, and whilst educators are unlikely to be amongst the early white-collar victims of AI replacement, but their lives must and will change forever. They will need to teach different material, as well as some of the material they already teach, and they will need to teach differently.

A sophisticated personal epistemology helps people develop sophisticated knowledgeable understanding and skills from their academic Studies. And it is beyond AI. To extend the initially simple personal epistemologies of our students, we need to explicitly teach them about the potential sources of knowledge and the ways in which they can justify that knowledge is justified. We need to help people design and ask good questions that probe the information they are presented with in an appropriate and useful manner. We and they must recognise the contextual nature of their knowledge and its inconsistency.

The final and most important most important element of human intelligence is perceived self-efficacy. It pulls together all the other intelligence elements and is way beyond the powers of AI. Self-efficacy is important for teachers as well as learners. We can help learners to develop a greater understanding of their own self-efficacy through developing the other six intelligence elements. It also needs to be the focus of specific and explicit teaching. It should be the intelligence that we strive for throughout our lives, within and beyond our formal education and training.

Moving to an intelligence-based curriculum of the sort outlined in Luckin (2018 will require a transformational change, for which we must plan now. And, as if this were not challenging enough, we also need to teach people about AI, including and as the highest priority, we need to teach the teachers and trainers about AI. Education about AI must include: teaching people how to work effectively with AI systems; giving people a voice in what AI should and should not be designed to do; and helping some people to build the next generation of AI systems.

AI can help us build our future education systems based on the progression models that include all seven intelligence elements. It is technically straightforward to develop AI to teach academic, interdisciplinary knowledgeable understanding and skills including the provision of detailed continuous assessment about each individual's progress towards each goal. The use of such systems would free our human educators to focus on the holistic development of their students' intelligence.

Imagination and Creativity

Creativity and imagination are essential human capacities, Einstein is believed to have equated intelligence with imagination. Creativity can be assisted through large bodies of knowledge that have been securely committed to memory. Creativity and imagination enable us to express our thoughts, feelings and desires and they underpin scientific and technological development too. They are not however a separate sort of intelligence, but rather the result of the development of all of our holistic human intelligence. Creativity and imagination can be nurtured by education, although systems that focus primarily on knowledge acquisition where there is an emphasis on testing and examinations can hamper learners' capacity to be imaginative and creative. There are some excellent books about how imagination and creativity can be nurtured, (see for example Cochrane & Cockett, 2007 Cochrane, 2012; Hannon, et al. 2013; Lucas, et al., 2013; QCA, 2004; Sefton-Green, et al, 2011; Sorrel, et al., 2014; Sternberg, 1999). Some of the key aspects of behaviour that have been identified as being associated with creativity include being curious and questioning, being willing to explore and challenge one's assumptions. Persistence is also important, as is being confident enough to be different and capable of coping with a degree of uncertainty as well as having the ability to focus and direct one's attention.

Ethics

We have made the case above that an understanding of the ethics of AI is crucial to our future use of AI, both in terms of how systems are developed, but also to ensure that users can make good and effective use of the AI systems.

There are a variety of ways of teaching ethics but these need to be age and stage appropriate. There are many examples of using science fiction from films and books to get students to reason with the consequences of the rapid development of AI.

One of the most powerful ways to learn is learning by doing and by using AI systems or creating them, learners understand how they work and begin to appreciate the boundaries and potential of AI systems.

Another approach is to support students in thinking through the consequences of relevant thought experiments - both positive and negative – and modelling reasoning about potential scenarios. Students can then be supported in practicing their own ethical reasoning. This will almost certainly involve having to confront and discuss difficult scenarios with uncomfortable consequences and needs skilled and expert teachers to manage the conversations.

Finally, older children can learn about the different ethical frameworks (utilitarianism, deontology and virtue ethics) within which we can evaluate the ethics of AI which will enable them to evaluate their use of AI by a variety of measures.

As Luckin (2018) points out: “Any failure to recognize and address the urgent and critical teaching and training requirements precipitated by the advancement and growth of AI is likely to result in a failure to galvanize the prosperity that should accompany the AI revolution.” P 124

So, what would a curriculum for secondary education look like?

As we have discussed, the curriculum would have to include resources on:

- Basic concepts of AI
- Basic AI programming
- How to work with AI systems effectively
- Digital and data literacy
- Ethical reasoning

However, the most crucial part of this curriculum would be to enable students to explore AI systems, to give them the opportunities to think about their use of AI and how it can be harnessed for good, as well as the understanding and wherewithal to know what to do when AI is not being used for legal and ethical purposes.

Conclusion

AI will make a significant contribution to how students engage with knowledge, develop academic knowledge related skills and experience their academic education in terms of personalised learning.

This will free human teachers to support children to develop and monitor the other aspects of intelligence described by Luckin (2018).

It is impossible to predict what AI will bring to our futures given the rapid pace of change and development but it is clear that we need to support active citizens of the future in harnessing and engaging with AI for good ethical purposes.

References

Buckingham, D. (2018) Defining digital literacy: What do young people need to know about digital media? Nordic Journal of Digital Literacy, 2015(4), 21-34.

Djumalieva, J. and Sleeman, C. (2018) “Which digital skills do you really need? Exploring employer demand for digital skills and occupation growth prospects.” NESTA https://www.nesta.org.uk/report/which-digital-skills-do-you-really-need/

Financial Times, 2017 Education must transform to make people ready for AI. https://www.ft.com/content/ab5daa64-d100-11e7-947e-f1ea5435bcc7

Frey, C. & Osborne, M. (2013) The Future of Employment, University of Oxford.

Frey, C., and Osborne. M. (2017. “The Future of Employment: How Susceptible are Jobs to Computerisation?” Technological Forecasting & Social Change 114: 254–280.

Gardner, H. (2007). Five minds for the future.

Goodman, A. et al. (2015), *Social and emotional skills in childhood and their long-term effects on adult life*, Institute of Education, UCL.

Gregory, A. and A. Sadeh (2012), “Sleep, emotional and behavioural difficulties in children and adolescents”, Sleep Medicine Reviews, Vol. 16/2, pp. 129-136

James, M. et al. (eds.) (2011), *The Framework for the National Curriculum: A Report by the Expert Panel for the National Curriculum Review*, Department for Education, UK.

Kahn, K. and Winters, N. (2018) AI Programming by Children. Proceedings of the Constructionism 2018 Conference, https://ecraft2learn.github.io/ai/) accessed Sept 2018.

Kuang, C. (2017) Can AI be taught to explain itself?” New York Times Magazine, 21 November. Online https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html (accessed 1 September 2018)

Lippman, L. et al. (2015), *Key Soft Skills that Foster Youth Workforce Success; Toward a Consensus Across Fields*, Child Trends Publishing, Washington, DC, https://www.childtrends.org/wp-content/uploads/2015/06/2015-24WFCSoftSkills1.pdf (accessed on 11 October 2018).

Medicine Reviews, Vol. 16/2, pp. 129-136, http://dx.doi.org/10.1016/j.smrv.2011.03.007.

Oates, T. (2002), “Contributions to the Second DeSeCo Symposium Definition and Selection of Key Competencies”, *Contributions to the Second DeSeCo Symposium*, https://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf (accessed on 10 October 2018).

Oates, T. (2003), "Key Skills/Key Competencies: Avoiding the Pitfalls of Current Initiatives", in Swiss Federal Statistical Office (SFSO) and A. Education Statistics Services Institute (ESSI) (eds.), *Contributions to the Second DeSeCo Symposium Definition and Selection of Key Competencies*, Swiss Federal Statistical Office (SFSO), Neuchâte, http://www.oecd.org/education/skills-beyond-school/41529505.pdf (accessed on 10 October 2018).

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., and Wuetherick, B., (2015) Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report, Dalhousie University. https://dalspace.library.dal.ca/bitstream/handle/10222/64578/Strategies%20and%20Best%20Practices%20for%20Data%20Literacy%20Education.pdf?sequence=1&isAllowed=y Accessed September 2018.

Schneider, P. & Bakhshi, H. (2017) The Future of Skills: Employment in 2030 NESTA https://www.nesta.org.uk/report/the-future-of-skills-employment-in-2030/

Seldon, A. & Abidoye, O. (2018) The Fourth Education Revolution: Will Artificial Intelligence liberate or infantilise humanity? The University of Buckingham Press.

Tarricone, P. (2011) The Taxonomy of Metacognition. New York: Psychology Press.

World Economic Forum (2015) New Vision for Education: Fostering social and Emotional Learning through Technology https://www.weforum.org/reports/new-vision-for-education-fostering-social-and-emotional-learning-through-technology

World Economic Forum (2016) The Future of Jobs Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution, https://www.weforum.org/reports/the-future-of-jobs

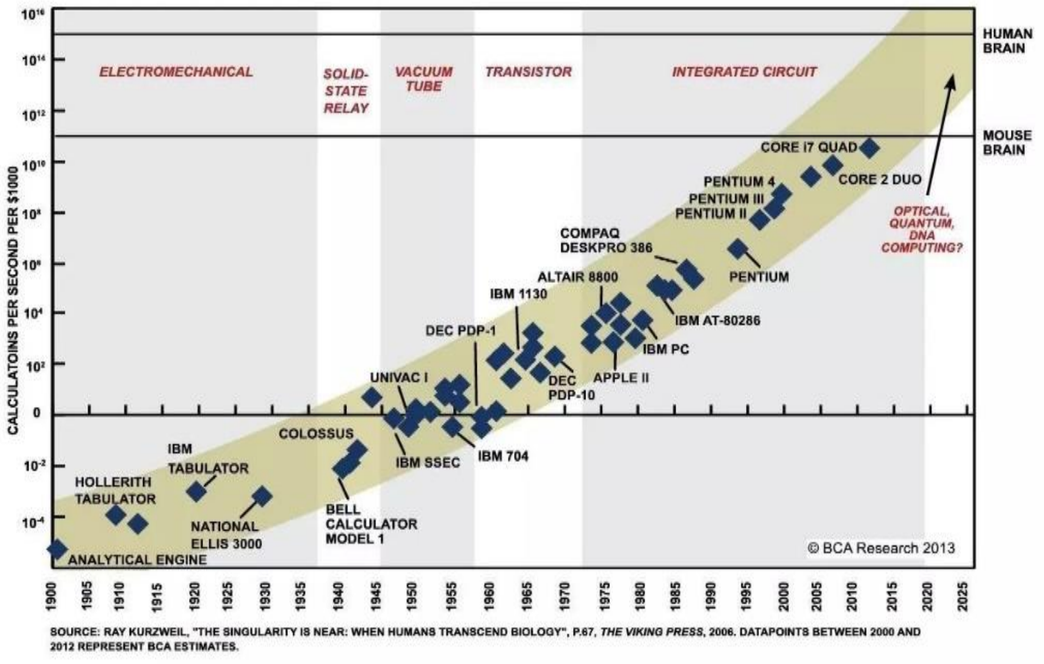
Abstract

AI is becoming a far more concrete reality today as machines that “think” are becoming increasingly sophisticated, and widely available. In this paper, we explore how that sophistication and availability has already made significant changes to human societies, what those changes might look like by 2030, and particularly the relevance of AI and character development. We discuss attitudes and values that will become more important in the time of increasing AI and look at the risk or concerns regarding ethics. In the light of agency, how we can ensure ethical behaviour of AI when the AI itself has agency.

Introduction

Technology is progressing exponentially fast. Gordon Moore (1975) observed that computing power (defined by the number of transistors per integrated circuit) was doubling approximately every two years, a pattern that remains consistent to this day (known as “Moore’s Law”). If the trend continues, the processing capacity of machines *will* surpass the capacity of the human brain, and continue well beyond; each year making more progress than it did the year before (this intimidating process is illustrated in Figure 1). Exponential progress is notoriously difficult for humans—who are adapted to thinking linearly-to appreciate (Wagenaar & Sagaria, 1975). Technology and society are heavily interconnected, so faster changes in technology often translate to faster changes in society. Moreover, change, more often than not, translates to greater unpredictability in the form of VUCA. To the human mind that is seeking to make sense of the world, this is a potentially troubling state of affairs (as we will soon see). With growing automation, machines also have the potential to take away the pursuits that bring meaning and purpose to the lives of millions of people—their careers, their passions, and their way of life. While it is clear that there are potential costs to technological progress, there are also many benefits. Technology can transform society for the better by improving decision-making, healthcare, transport, security, and education. Artificial intelligence can liberate humans from many tasks that would ordinarily be demanded of them, potentially returning us to a society that is “time rich” instead of “time poor”. The key then, is in making the most of the benefits while being well prepared for the hazards.

Figure 1. The exponential progress of computing power from 1900 to 2013, with projections into 2025, where artificial intelligence is projected to parallel the processing capacity of the human brain.



Source: Figure adapted from (Kurzweil, 2005)

The acronym VUCA has been used often in the context of leadership theories (Pasmore & O'Shea, 2010), and the US Army originally used the term to refer to the increasingly VUCA-esque world following the Cold War (Bennis & Nanus, 1985). What are the different aspects of VUCA? Volatility is the liability of something to change rapidly and unpredictably. Stock markets, for example, are considered volatile because of how quickly they change and therefore how notoriously challenging they are to predict. Uncertainty relates to the quality of information one has—or the degree to which the outcome of an event is knowable in advance. Complexity increases when there is a greater number of relevant variables or interrelationships; the more variables, the more complex the situation. For instance, managing a classroom with five children is much less complicated than managing 100, and doing so is simpler if each student speaks the same language. Ambiguity occurs when an event, situation, or context is unclear, either because information is missing, inconsistent, contradictory, or obscured in some way. For humans—and indeed machines—each of these components of VUCA ultimately make for a less predictable world.

The following paper is divided into two parts (1: VUCA and 2: Agency), followed by a general discussion. Part 1 begins with a succinct introduction to a theory of the brain known as predictive processing. The aim here is to provide scaffolding for understanding how humans respond in VUCA situations. We then briefly review the psychology, neuro-chemistry and neuro-physiology literature to evaluate the negative and positive consequences of VUCA for human well-being. We then consider how people can learn to respond to VUCA effectively. In the final section of Part 1, we provide a broad overview regarding the current state of artificial intelligence (AI)—including specific and general AI—and then discuss how autonomous machines perform in VUCA situations. In Part 2, we turn our attention to Agency, and consider the possibility that a sense of agency might serve a valuable function for humans in uncertain situations. We also consider whether the processes that underlie autonomy in machines resemble primitive forms of agency, and whether AI is prepared for morally ambiguous situations. In the general discussion, we synthesise the review and propose that a malleable attribute of humans—their ability to learn new skills and apply meta-learning principles—will be particularly valuable for ensuring that humans can adapt to change in a VUCA world. We suggest that one of the key goals of education ought to be to prepare students to be adaptable learners equipped with meta-learning skills (Maudsley, 1980), so that when inevitable changes occur, people and communities are robust enough to adopt new skills and practices, and effectively transfer learning across situations. Meta-learning capacities may also help provide a sense of agency and self-efficacy in an otherwise chaotic world—it may not be possible to know *what* to expect, but it is perhaps possible to learn *how* to respond when the unexpected occurs.

Introduction

The idea of artificial intelligence (AI) clearly is nothing new to humans. Hephaestus, the Greek god of metalworking, fire, and sculptors, is said to have produced mechanical dogs, horses, and “golden maids to do his bidding” (Mayor, 2016). In Jewish folklore, a golem is a humanoid made of clay and animated by mystical means (Collins & Pinch, 2012). Despite its ancient origins, AI is becoming a far more concrete reality today as machines that “think” are becoming increasingly sophisticated, and widely available. This report explores how that sophistication and availability has already made significant changes to human societies, what those changes might look like by 2030, and particularly the relevance of AI and character development for each other.

Three sets of questions frame our thinking about this report:

1. What attitudes and values become more important in the time of (increasing) AI (technology, robotics, big data etc.)?

2. What are the risks or concerns regarding ethics in the time of (increasing) AI?

3. How can we ensure the ethical behaviour of AI? Can that be done? Especially in the light of agency: can we ensure ethical behaviour of AI when the AI itself has agency?

These questions will be explored and revised in the rest of this report. The report continues with definitions, and then moves through three sections, one dedicated to each of the three questions above. The report concludes by discussing how what we know and predict about AI, robots, and big data fits into some of the OECD frameworks.

We will look to many different sources in order to better understand the past, present, and future of AI. Computer scientists, philosophers, psychiatrists, psychologists, and sociologists will all be cited. An indication of the power of AI is how many scholarly disciplines are eager to explore its impacts.

Definitions

AI is developing rapidly, and there are no universally accepted definitions for terms that will be important to this report (Parnas, 2017). We will assume the definitions below for convenience in this document. We have annotated the definitions and provide clarifications when necessary.

Artificial Intelligence: The ability of a computer or other machine to perform those activities that are normally thought to require [human] intelligence. Adapted from (American Heritage, 2018).

This augmented dictionary definition of artificial intelligence contains ambiguities. For example, "normally thought" can be interpreted in many different ways. However, because the popular use of the term covers a broad range of machines, we will adopt this admittedly loose language because tight restrictions on the machines considered will not be helpful for this report. When we need to be more specific, we will narrow the scope of this definition.

The history of artificial intelligence is filled with ambiguous terminology and claims that were later shown to be at the very least premature. Already in the 1980's, some computer scientists noticed a trend. Discussing the use of "artificial intelligence," the software engineering pioneer David Parnas (1985, p. 1332) wrote, "…once we see how the program works and understand the problem, we will not think of it as AI anymore." Despite these cautions, it is helpful to begin with some basic definitions.

Intelligence: The ability to apply knowledge to manipulate one's environment or to think abstractly as measured by objective criteria (such as tests) (Merriam Webster, 2018).

This is the second meaning from the Merriam Webster dictionary. We prefer it because the idea of tests and criteria determining the presence or absence of AI has a long history, a history we will review in this report.

Robot: A physical machine that has sensors and can move with at least one degree of freedom. It is often programmed, but might also be controlled remotely.

There is even less agreement about the definition of "robot" than about "artificial intelligence."

Bot: A computer program that is capable of interacting with servers and sometimes with human users, on the World Wide Web.

Autonomous: An entity (in our case, a robot or bot) that is designed to execute its programs for relatively long periods of time without direct human intervention.

Please note that our definition of "autonomous" intentionally evades any discussion of whether or not a machine or program requires characteristics know as "free will" or "intention" when observed in humans. Discussions about this issue have a long scholarly history; for example, see Sullins (2006). We will avoid those philosophical debates in this report.

Big Data: "… data sets that are so big and complex that traditional data-processing application software are inadequate to deal with them " (Wikipedia, 2018).

Sophisticated machines: Machines controlled by computing, where the software is routinely classified as artificially intelligent. Examples include robots, web bots, networked devices with voice recognition, games with AI players, and systems that generate virtual reality environments.

A careful reader may object to these definitions as vague and imprecise. Since considerable time and treasure are being invested in AI, robotics, and big data, we could reasonably expect that these kinds of projects should have some widely agreed upon definitions. Unfortunately, they do not. We will make do with the somewhat loose definitions above, including when appropriate examples that should help the reader understand what the terms mean in specific situations.

What attitudes and values become more important in a time of increasing use of AI, robotics, and big data?

This question could be approached as a descriptive question, an invitation to predict how people are likely to react to the increased use of sophisticated machines. In this report, we take a different approach: we interpret the question as an invitation to prescribe. We will describe how we think attitudes and values should change as AI becomes more sophisticated.

The technologies of AI, robots, and big data interact synergistically. Artificial intelligence is enhanced, and in some cases enabled, by the use of big data. AI that is embedded in an autonomous robot can increase the capabilities of the robot, and the mobility and sensors of a physical robot can enhance the amount and quality of data collected. In answering this question, we will be commenting on the increasing sophistication of machines, with the underlying assumption that AI, robotics, and big data each contribute to that sophistication.

Before we discuss attitudes and values specific to AI, we need to assert that in the past, technologies have often challenged existing attitudes and values. Viewed from a wide historical angle, the societal challenges of AI are similar to (though, we will argue, also distinct from) the challenges society faced in the Industrial Revolution, or at the dawn of agriculture. We need ethical people to design, develop, and deploy AI just as we needed ethical people to design, develop, and deploy looms and telephone systems. We have already lived through the advent of television, automobiles, the Internet, and nuclear weapons. Yes, AI has unique challenges, and we discuss them below. But we should keep in mind the importance of approaching these challenges in a way that does not undermine valuable attitudes and principles that have weathered previous technological innovations.

Another way of addressing this historical continuity of ethical concerns (even panic) over emerging technologies is to consider the role of humans in engaging these technologies, including AI. In all cases, whether horseless carriages, telephones, cloning or AI, a central aspect of the ethics of the technologies falls on the ethics of the human users of the technologies. It is certainly not exclusively a question of moral psychology, but it is centrally so. We can likely never build a moral world without moral people, and that starts with the moral socialization of children. "There is no future without children, and there is no moral future without children of character" (Berkowitz, 2012, p. 12). Whether it is cloning, nuclear energy, or AI, a check on unethical application in the moral character of the people who will apply them is critical. As Franklin Roosevelt (2018) said, "We cannot always build the future for our youth, but we can build our youth for the future. Such character is holistic more than it is modular. Isolating particular virtues, values, attitudes, or character strengths is conceptually interesting and even enlightening, but we need to be cautious in throwing out the holistic ethical baby with the particular virtue/value/attitude bathwater. Nonetheless, we can (and do below) identify specific ethical characteristics of relevance to this debate.

ANSWER 1A: As machines become more sophisticated, creativity and originality will become more important for people.

Moor (2001) identifies logical malleability as a central characteristic of any computational device. When we create a digital world, we can defy the physical laws that constrain us in the real world (at least within the confines of a simulation). Furthermore, AI has an unprecedented range of application compared to other technologies, which also can only be maximized through the creativity and imagination of the users and designers. This malleability is a major advantage for AI, robotics, and big data, because they are all powered by a digital model of the world. But we can only harness that power if our imaginations are up to the tasks of creating and implementing original, visionary ideas. It makes sense for us to look, then, to the intellectual virtues (Baehr, 2011; Zagzebski, 2003) of creativity and imagination.

The malleability of sophisticated machines means, among other things, that the machines can be used for good or ill. Some believe that AI has great potential for good (Taddeo & Floridi, 2018). To guide the development of AI and other technologies in appropriate ways, and to accurately predict the consequences of a particular technology, creative thinking is required. And adapting a technology someone else developed to a new purpose also requires creativity and originality.

Unfortunately, the use of computing in education so far has not always risen to the challenge for creativity and originality. Schneiderman (2003) started his book Leondardo's Laptop: Human Needs and the New Computing Technologies with this observation: "Computing today is about what computer can do; the new computing will be about what people can do." Although Schneiderman wrote that sentence 15 years ago, it can be argued that most computer use in education has not yet reached his description of "the new computing," but there is research that identifies some progress (Masoumi, 2015; McVeigh & Walsh, 2000; O'Hara, 2008).

Here are three examples of when education and technology have been used creatively in education:

1. A number of experiments have been attempted using robots to engage children diagnosed with autism spectrum disorder (Moorthy & Pugazhenthi, 2017). The predictability of robot behaviors seems to help some children become comfortable, and some of the children eventually improve their interactions with people after interacting with a robot. Results of this use of robotics are still preliminary (Talaei-Khoei, 2017), but this is one effort to use sophisticated machines in a non-obvious way for the benefit of a particular group of people. This is an example of using sophisticated machines creatively in a way that benefits a particularly vulnerable group of people.

2. Computer programs that use AI to help tutor students are an explicitly educational innovation that leverages advances in cognitive science and AI for learning (e.g., Crow et al., 2018). AI adds value to these systems by customizing the content and assessments for individual learners without the direct intervention of a human teacher. There is empirical evidence that demonstrates the effectiveness of several modern intelligent tutoring systems (Craig et al., 2018). Many of these systems are designed for a particular, some would say narrow, range of content. Research continues on making these systems more general, and more effective. One recent effort is to use intelligent tutoring with collaborative learning (Olsen et al., 2015); most intelligent tutoring focuses on individual learning. One of the possible advantages of automated tutoring is that it allows people freedom to fail at tasks without embarrassment. A particular kind of automated tutoring that encourages failure as learning is the use of online educational games, which has proven useful in some settings (González, 2014; Majdoub, 2016). Clearly, this is an imaginative way to use AI. However, removing a human teacher from this process of student learning has risks. For many years, research has indicated problems for students, especially children, who spend more time with machines and less time with people (Oppenheimer, 2003).

3. Big data holds promise for planning and delivering education at all levels, and there are examples already of useful results, especially in higher education. For example, Siemans and Long (2011) explain how institutions can identify students at risk, and can tailor interventions using big data analytics. However, enthusiasm for further work in this area is tempered by concerns about security, privacy (Wang, 2016), and inappropriate educational assessments (Gipps, 2002).

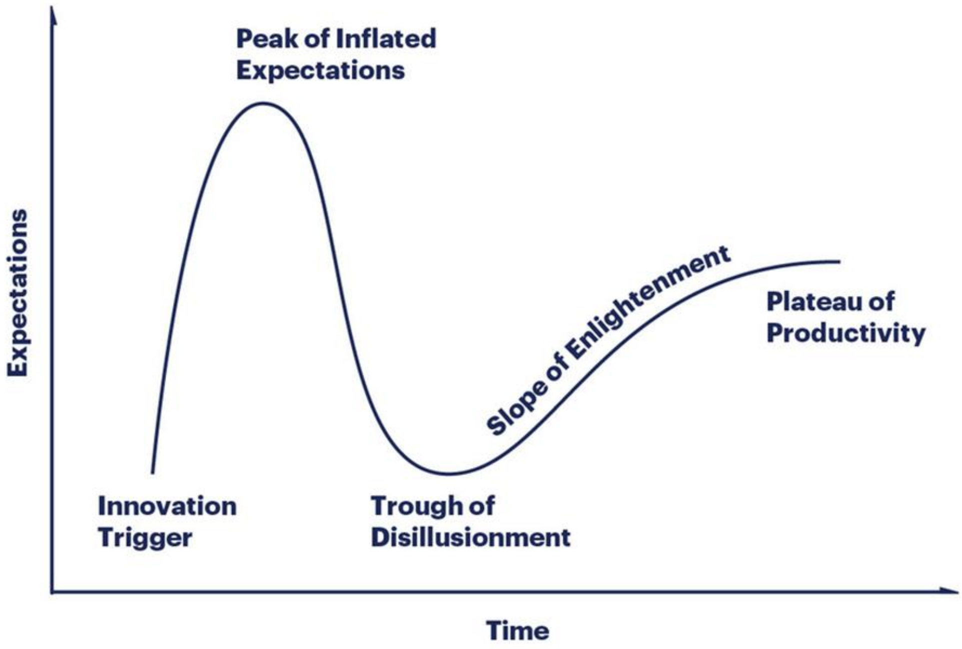
There are at least two aspects of education using creativity and originality with sophisticated machines that we wish to stress: encouraging educational professionals to use creativity and originality when employing sophisticated machines in education; and encouraging students to use creativity and originality as they use these machines during formal learning, information learning, and in their subsequent professional lives.

The use of sophisticated machines in education are not a panacea, and they should be used carefully. Issues of privacy of students is a major concern. Also, the use of technology should never become an excuse for removing the human touch from education. Otherwise, educational technology could increase the risk of an issue discussed above: the devaluation of humanness.

ANSWER 1B: As machines become more sophisticated, attitudes of prudent caution and intelligent scepticism will become increasingly important.

For many years, AI, robots, and big data have all suffered from inflated claims and repeated cycles of disappointment followed by new waves of enthusiasm (Hopgood, 2003). Figure 1 is a graph depicting "Gartner's Hype Cycle" for technology (Gartner, 2018). Figure 2 shows many AI technologies mapped onto particular parts of the hype cycle by Alex Woodie (2017). For our purposes, the exact location of each technology on that curve is not crucial; but it is relevant that so many of the technologies lumped under AI are new enough that they are considered in the early stages of technological hype. When hype is intense, caution is called for.

Figure 2. The Gartner Hype Cycle, from Gartner's website. (2018)



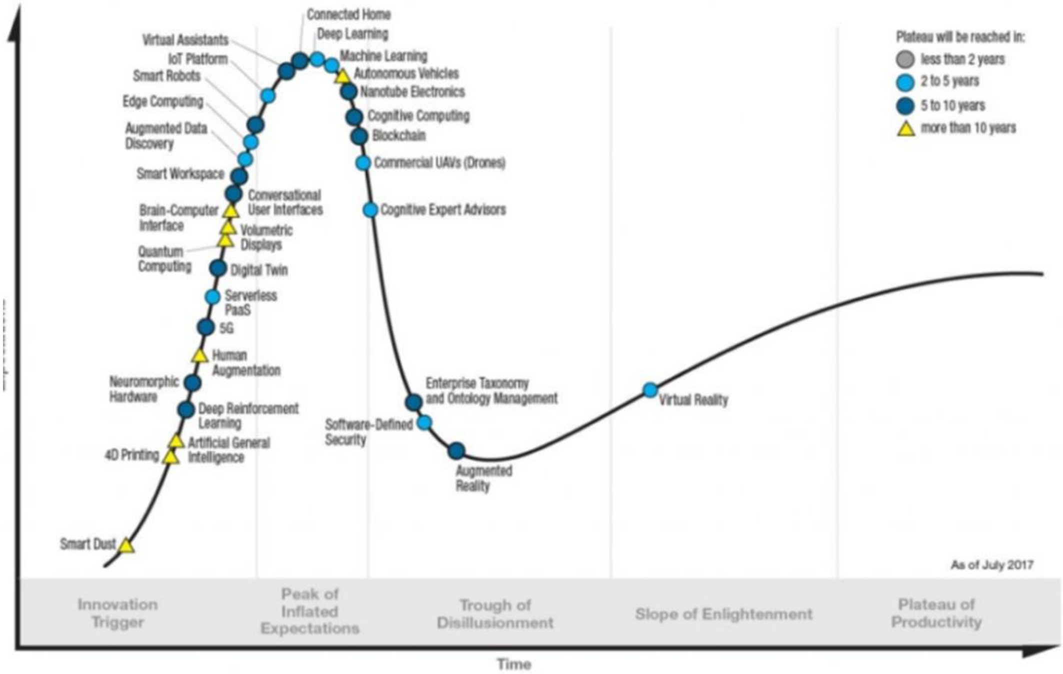
Veteran observers of technology know that early in their development, AI artefacts can become famous long before they are well understood (Business Insider, 2018). As we shall see later in this report, as sophisticated machines more successfully imitate human behaviours and habits, mysteries and misunderstandings about their true capabilities will multiply. A recent paper in a leading computer science journal warned about "overtrust in the robotic age" (Wagner et al., 2018).

A prerequisite for humans to make wise decisions about technology in general, and sophisticated machines in particular, is that humans have a realistic, fairly accurate picture of what these machines can and cannot do. Realism can be challenging to accomplish in the face of relentless hype and influential fictions about sophisticated machines.

Informed citizens who think critically will be crucial if we are to have any hope of wisely employing sophisticated machines. Intellectual virtues, taught in schools and practiced by all, will help us to select wisely among the many possibilities, and to apply them judiciously.

For all of these reasons, caution and scepticism about sophisticated machines will be increasingly important for the public in general, and for professional educators in particular. Buying into the hype of a particular technology before it has become mature and reliable is wasteful and distracting. Furthermore, relying on these technologies can be dangerous when they behave erratically and unexpectedly. (We discuss the reliability of sophisticated machines later in this report.)

Figure 3. AI technologies mapped onto the hype cycle. (Woodie 2017)



ANSWER 1C. As machines become increasingly sophisticated, it will become increasingly important for humans to recognize the value of human contact, both physical contact and social contact.

In a recent book called Alone Together, Sherry Turkle (2017) warns about the dangers of what she calls social robots. The subtitle to the book is “Why we expect more from technology and less from each other.” There is a danger that the increasing importance of sophisticated machines will encourage humans to devalue other humans. This devaluation can happen in several different ways; some scholars (including Turkle) are convinced this devaluation is already occurring. If these scholars are right, then it will be increasingly important for people to recognize the value of their own humanness and the value of others’ humanness (Putman, 2000). This is important for individual well-being, and for the well-being of human societies and organizations.

One aspect of human devaluation is manifest when machines take the place of people in the workplace. The most obvious impact of such replacements is that people lose their jobs. Automated grocery check-outs (Heppermann, 2012) and self-driving mining trucks (Morell, 2017) have already replaced people with machines. By 2030, this trend is likely to increase (Semuels, 2017), especially in transportation, logistics, office workers, and factory employees (Frey & Osborne, 2017).

But this economic displacement is only one symptom of a deeper cause: humans who own and manage businesses and organizations no longer value their human employees as much as they value the machines that replace those workers. This is essentially a modern variation on the Marxist concern about the devaluing those who produce commodities. And customers (for example, in retail) who accept (or even demand) this automation participate in the devaluing. The example of grocery checkout is illustrative of this point: when customers scan their own groceries without the help of a store employee (except one employee who is in charge of multiple automated stations), then those customers have less human interaction when they purchase groceries. (A similar situation occurs with unattended ATM machines replacing bank clerks.) If the human touch is valued, this kind of automation constitutes a loss for customers. If that loss is felt with sufficient clarity, customers could avoid stores that do not have human clerks, and stores might change their procedures. But if customers do not feel this loss acutely, or do not communicate their displeasure to retailers, then increasingly clerks will be replaced and retail purchases will become less likely to include human interaction. In a classic study, Fisher, Rytting, and Heslin (1976) reported the positive impact of slight touches by librarians at book checkout (most notably for females).

Another manifestation of the de-humanization of retail is the growth of online selling at the expense of brick-and-mortar businesses (Hannam, 2017). It may be that some online retailers treat their employees well, but there are reports that some large online retailers do not treat employees humanely (Head, 2014). If these reports are accurate, that reinforces the idea that automation is being used to replace people, not to enhance their working experience.

There is a temptation to explain the replacement of human workers by machines as being made solely on the basis of efficiency (Byrne, 2018). Another argument is that automation creates more jobs than it destroys, pointing to decades of historical data (Allen, 2015). Both these arguments are, we contend, less compelling as machines become more sophisticated. When machines replaced humans in jobs that were largely boring and dangerous, and when the resulting technological improvements quickly opened up new jobs for humans, both arguments seemed fairly strong. But now and in the future, sophisticated machines are increasingly taking jobs that many people found enjoyable and stimulating, including jobs that include significant human interaction. Also, it is less clear today than twenty years ago that when these kinds of jobs are taken by machines, that there will be different jobs created for people (Acemoglu & Restrepo, 2017).

If sophisticated machines do indeed reduce the opportunities for people to have fulfilling work, that is a significant loss for individuals and societies. Surely a healthy work-life balance (as described by the OECD (ND)) is difficult to establish when work, particularly valued work, is less available to people.

If we recognize the value of human interaction during business transactions, a robot cannot deliver that value, despite any efficiencies. If we do not value human interaction, then the calculation of cost and benefit changes, and automation will continue to replace humans in jobs we previously assumed were for people. It is our contention that an enhanced appreciation for human-to-human interactions is becoming increasingly important to make wise decisions about sophisticated machines. This concern dovetails with concerns about the deterioration of human interactions that predate the rise of AI and big data (Putnam, 2000).

Despite Turkle's objections, there are people who may benefit from an increased use of AI in social interactions. For example, people who find it difficult, either through illness or temperament, to relate to other people may find it helpful to either relate to AI, or to relate to other people with computer mediation. Thus AI can complement, rather than merely replace, human contact. One example of the potential of this use of AI is the use of robotics with children diagnosed with autism (Tucker, 2015). Another is the use of AI as triage in emotionally charged behaviour incidents in schools (Hylen, 2008).

What are the risks or concerns regarding ethics on the time of (increasing) AI?

While answering Question 1, we have already introduced some of the ethical issues that arise with AI, robotics, and big data. We will expand on some of those issues as we deal with Question 2.

Privacy: Privacy concerns about technology have a long history in the study of computer ethics (Nissenbaum, 2009). But as machines become more sophisticated, privacy risks are likely to increase (Skirpan, 2018). Simply put, machines will be capable of tying together large collections of data in ways that will reveal personal data that would be hidden from people and from less sophisticated machines. Privacy will also be threatened when people become less cautious about guarding their private information from machines that will appear "friendly," and which will be around us in many different forms, some of them hidden from view. The technologies sometimes referred to as "the Internet of things" increase the amount of data that can be collected about our daily lives, and increase these privacy concerns (Federal Trade Commission, 2015). Furthermore, as machines become more intelligent, and people are more "connected", machines will find more ways to invade privacy and gather information.

Human dignity: When people prefer electronic entertainment and electronic "companions" to human contact, people are choosing to value machines over other people; this diminishes the dignity of humans. When employers choose machines to replace people in jobs the people want, again some human's dignity is harmed. Ethics, particularly virtue ethics, has long focused on human dignity as a fundamental concept (Düwell, 2017). The increased use of sophisticated machines does not have to harm human dignity; however, that is clearly a risk. For example, Sharkey (2014) explored ways in which using robots in elder care could result in harm to elders' dignity, but also could improve elders' dignity. In human affairs, different control and leadership frames offer different threats (or lack thereof) to human dignity of those being led or controlled (e.g.,, in sports or business organizations).

Concentration of economic power: The concentration of economic power is a political and economic development that has important ethical implications. When economic power and wealth are concentrated in relatively few individuals and corporations, the resulting inequities tend to reduce fairness and the public good (Richard, 2017). Although technology is not the only driver of this concentration, it is an important factor (Katz et al., 2017). And as machines become increasingly sophisticated, the people and corporations that develop and own the most sophisticated machines become all the more powerful. This cycle can increase concentrated power, which is an ethical concern. Several of the ethical risks mentioned in this report become more ominous when sophisticated machines are developed and controlled by organizations that focus on profit to the exclusion of human thriving (Rowe, 1995). It can be argued that not all corporations ignore the public good all of the time, and that corporations could use their power for socially laudable aims (Banerjee, 2008). However, the increasing concentration of power that sophisticated machines enable is clearly a risk.

There is a possibility that some sophisticated machines could be used to empower people who would otherwise be relatively powerless. If sophisticated machines could be made widely available (either by subsidies or because the technology is relatively inexpensive), then their economic and cultural power could be more widely shared. One way in which sophisticated machines could be relatively inexpensive and widely available is if they used open source or free software (Grodzinsky et al., 2003). The choice of which kind of software is used to power AI, robots, and big data – proprietary software on the one hand, and free and open source software on the other hand – is a choice with ethical significance.

Algorithmic bias: Recent events (Knight, 2017) have raised concerns about "biased algorithms." The concerns can be characterized with two questions: "Are algorithms treating us fairly?" and "Can we tell if algorithms are treating us fairly?" (Kirkpatrick, 2016). There is evidence that software developers routinely embed their own biases in the systems they create, and that data generated and stored digitally can embody unfair biases (Baeza-Yates, 2016).

System reliability: Software engineers constantly face the question "how good is good enough?" Computer ethics scholars have long recognized that this seemingly technical question has fundamentally ethical implications (Collins et al., 1994). It is unlikely that software of any complexity will be error free (Miller et al., 1992), and as sophisticated machines add capabilities, their underlying software takes on increasing complexity. Engineers often talk about the tradeoffs between making their products quickly, making them better, and making them more inexpensively. The common wisdom is that at best you can get two of the three in a product; it is just too difficult to get a product "faster, better, and cheaper" simultaneously (Swink et al., 2006). Complexity pushes developers' speed, cost, and quality, so system reliability is likely to be a challenging and ethically significant issue as machine complexity increases in AI, robotics, and big data applications.

Another factor in system reliability is the flexibility that is being programmed into machines designed to "learn." "Machine learning" is a term of art in AI; Faggella (2018) explains it as follows: "Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions." In order to improve itself without human intervention, a machine must be able to change its programming "on the fly," based on information gathered by its sensors over time. Trying to decide reasons for behaviours of a machine driven by machine learning algorithms becomes increasingly impractical as the cycle of behaviour, change, and self-programming continues (often multiple times in a single second). Self-modifying code is notoriously difficult to debug, or even to understand why the machine has reached its current state. This type of reliability issue is clearly an ethical problem because software safety and quality control is so difficult (if not completely impractical) with machines that learn without direct human intervention (Wolf et al., 2017).

.

How to ensure the ethical behaviour of AI? Can that be done? Especially in the light of agency: can we ensure ethical behaviour of AI when the AI itself has agency?

What does it mean to ensure the ethical behaviour of AI?

In their book Moral Machines: Teaching Robots Right from Wrong, Wallach and Allen (2008) tackled the philosophically difficult issue of whether or not a robot can be developed so that they can recognize the difference between right and wrong. Further, if robots can recognize such differences, how can we best program them (or "teach" them) how to act right?

As Wallach and Allen ably described, we are not yet capable of answering either of those important questions with authority. However, we can discuss with more precision a more limited set of questions around the issue of whether or not we can make sophisticated machines act as if they had internalized the difference between right and wrong. Now we have recast the questions to be about behaviours and not about the internal motivations and moral compass inside the machine's programming; that is a step forward for our task, since demonstrating that a machine has motivations and a moral compass is a daunting task. We will avoid that by adopting the less challenging (but still difficult) questions around machine behaviours.

This, of course, assumes programmers and others involved in AI design, programming and usage want the intelligent machines in question to be ethical. That in turn brings us back to the fundamental question of educating, parenting and otherwise socializing people, both in general and in the AI realm in particular, to develop moral character; i.e., to have the knowledge, skills, values and attitudes necessary for moral agency. As Berkowitz and Bier (2014) have defined it, moral character is "the set of psychological characteristics that motivate and enable an individual to function as a competent moral agent" (p.250). Nevertheless, we will address this question of insuring ethical AI by assuming those responsible are indeed motivated and able to do so.

Early in the history of computing science, Alan Turing (1950) published what became know as the "Turing test." Based on a parlour game popular at the time called the "imitation game," Turing's test had a judge receive typed messages from two sources (both hidden from the judge): a human in one room and a machine in another room. If the judge could not tell the difference (guessed wrong as often as right), then the machine was said to be acting intelligently. Turing explicitly denied that this proved that the machine was truly intelligent, in the same way that humans are intelligent. He merely thought this was a practical way to decide if the machine appeared to be acting intelligent in this specific game, which is quite a different matter (Oppy & Dowe, 2018). Despite Turing's objections, the test (or at least one subsequent version of the procedure) is often referred to as "Turing's test for artificial intelligence."

Allen, Varner and Zinser (2000) suggested that we adapt Turing's deflection of the "are machines intelligent?" question in a similar way, but with respect to machines' ethical behaviour. They proposed a "moral Turing test" or MTT, in which ethically significant questions are put to two entities, one human and one a machine. If a human judge cannot reliably determine which of the entities is a human, than the machine "passes" the MTT (Allen et al., 2000, p. 254). Allen et al. admit that there are philosophical problems with the MTT, since (as with the original Turing test), it is only outward behaviour that is being tested, not any inner states or motivations. But the practical advantages of having such a test make it potentially useful, if not ethically definitive about the possible interior significance of the outward behaviours.

Another objection to the MTT is that it seeks to force human ethics onto a machine, when a different sort of ethic might be more appropriate to a sophisticated machine. We assert that human ethics is precisely what we should enforce on machines. If we allow machines that follow a different ethical path from humans, it seems unlikely that the machines will enhance human flourishing (Barrat, 2013).

One possible set of criteria for the experts to use in judging the MTT would be Kohlberg's stages of moral development (Kohlberg et al., 1983). Using Kohlberg's stages, the ethics experts would not base their judgment on what action the machine suggested, but instead would examine the nature of the justifications the machine gave for preferring that action.

Because of its obvious advantages in answering the question of this subsection, the MTT is adopted for this report. We can then answer the question of whether or not a machine is likely to pass the MTT (or one of several possible refinements of the MTT mentioned by Allen et al.) in the foreseeable future.

Allen and et al. (2000) seem cautiously optimistic about the possibility that machines could be developed to pass the MTT. They hope that the resulting machine will be designed to be more than merely convincing to a human judge and will be built to act like a "praiseworthy moral agent" (p. 261), clearly a higher bar.

In order to continue our discussion about this and subsequent questions in this report, we will assume that if a machine of the future can appear to a panel of human experts in ethics to be answering questions about ethical questions as if the answers were from what they consider a praiseworthy moral agent as described by Allen et al. (2000), then that machine is said to be behaving ethically (in the form of its answers).

It would require extensive additional justifications to demonstrate that a machine capable of passing the MTT by giving answers, could also be capable of transforming these answers into actions that would be judged to be consistent with the answers. We will not attempt those justifications here, but we will merely state our opinion that such a transformation would be non-trivial and would require significantly more work; but also that this additional work would be feasible given considerable time, effort and resources.

At the writing of this report, we are not aware of any machines that could pass a rigorous MTT. However, we also do not see any theoretical reason why some future sophisticated machine could not be developed that could successfully mimic human reactions (presumably acting as a moral agent) to questions related to ethical decisions.

One way to convince ourselves about a future machine that might pass the MTT is to examine the history of machines challenging the original Turing test. The Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB) runs an annual contest to award its Loebner Prize for the computer program that most successfully mimics a human being typing via chat to another human (AISB, 2018). The first Loebner competition was run by Hugh Loebner in 1991. The contest has its detractors; Marvin Minsky called it a "publicity stunt," in an article entitled "Artificial Stupidity" (Sundman, 2003). Despite these objections, the existence of the annual event, and the AI programs submitted to the contest, do show how these programs trying to fool humans are becoming more sophisticated and more successful. This suggests (though does not prove) that machines are likely to improve their performances as mimics of humans. That at least gives some indication that the Turing Test and the MTT might reveal machines that at least sometimes pass these (somewhat limited and arbitrary) tests.

Let us make the assumption that some machines in the future can with some regularity pass the MTT. What kind of answers might they be producing that experts would find convincing? Notice that philosophers and other experts do not always agree on what action would be most ethical in particular situations, nor do they always agree on the proper justification for such an action. Indeed, the philosophical literature on ethics is filled with arguments about such matters. Ethicists who emphasize the consequences of an act (for example, utilitarians) often clash with ethicists who follow Kantian ethics (for example, deontologists who focus on the inherent nature of an act). Such disputes date back at least to ancient Greece (Parry, 2014). Therefore, we should not expect that all machines that pass an MTT will give the same answer for particular questions about ethics. We merely suppose that a machine or machines will give answers that experts find convincing as ethically justifiable answers. Just as experts can disagree with colleagues about their analysis of an ethical problem without calling their opponents unethical, so might they judge a machine as "mistaken, but making a good faith effort" to solve a problem.

Given our recasting of the question, we now have a target for sophisticated machines that developers are trying to make into "moral machines:" passing the MTT as judged by experts in human ethics.

Is it possible to ensure the ethical behaviour of sophisticated machines?

We are now ready to approach the question of whether or not it is possible to ensure the ethical behaviour of sophisticated machines. We answer with a highly qualified "yes:" if a machine was built that routinely passed the MTT; and if that machine was stable (that is, it did not change its programming over time); and if the machine was capable of matching its behaviour to the ethical justifications it generated in passing the MTT; then that machine is likely to have behaviours that experts would generally label as "ethical." As with people's behaviours, not all ethics experts would agree on all of this machine's actions. But we expect a machine that matches each of these "ifs" would be judged as generally acting ethically.

There are other cautions that refer back to previous issues in this report. A machine that behaved ethically at one moment might act unethically in the next if the underlying program was unreliable due to faulty programming. A machine that learned autonomously (not under direct human control) could learn unethical behaviours. A machine could be hacked, and then instructed to behave unethically. Finally, a machine could be programmed to behave ethically for a period of time to establish trust, and then behave unethically to gain an advantage for itself or its owners.

How do we establish accountability for the behaviour of sophisticated machines? Is there a point after which a machine with AI becomes ethically "untethered" from its human creators, thereby attaining some kind of "agency?"

The wording of the original Question 3 makes certain assumptions that it is important to unpack as we explore our answers to that group of queries. First, there is the implicit claim that sophisticated machines will eventually have (if they do not already have) some form of agency. In the literature about computer ethics, the issue of moral agency of machines is by no means a settled question. (For example, see Miller & Larson 2005; Johnson & Miller, 2008; and Grodzinsky et al., 2012.) We do not seek to recreate those disputes here. Instead we will assume a limited view of agency in order to make progress on these questions. For our purposes, we will assume “agency” means that a sophisticated machine can act without direct human intervention over significant periods of time. For such machines, how do we envision issues of accountability and responsibility when a machine acts?

Michael Davis lists nine separate but related forms of the concept of “responsibility” (Davis, 2010). We will focus on two of Davis’s concepts of responsibility when discussing sophisticated machines in this report:

Accountability: “you should explain because you are responsible for what happened”

Liability: “you should pay because you are responsible for what happened”

Both of those quotes are from Davis (2010, p. 15); the emphases shown here are in the original.

Sophisticated machines that are designed for moral accountability could be designed to record both their inner states of computation and the input from all their sensors in real time. This would be similar, but far more elaborate than, body cameras worn by police. The machines’ recordings could be made available to appropriate people and organisations in cases where people wanted to know how and why a particular event happened when the machine was present and turned on. There would be no guarantee that a machine (or possibly machines) present at an event would have recordings that would unequivocally determine all that someone wanted to know about an event. But presumably evidence available from machines might be helpful in establishing some aspects of past events. When this is so, sophisticated machines will add to the account of the events, perhaps including the effects (if any) that the machine itself had on the unfolding of the event.

Merely recording sensor inputs may not, however, be the kind of accountability some people would desire. Such recordings might help reconstruct what happened without revealing much if anything about why something happened. Some relevant “why” questions might require information about why the machine was present, who owns the machine, who programmed the machine, what the machine was designed to do, when the machine was given explicit instructions from a human or another machine, and so on. A particular owner or manufacturer of a machine might be more or less willing to answer those questions; the machine itself might facilitate or impede the collection of such information from its memory. The more helpful the humans involved and the machine itself are to get these questions answered, the more “accountable” they are.

This sharing of accountability by humans and machines emphasizes the importance of a broader focus than merely focusing on the machine. The machine is part of a sociotechnical system, and that system, not the machine, bears responsibility for the machine’s actions. In this, we agree with Johnson (2006) and Johnson (2011): a sophisticated machine and the people who design, develop, and deploy that machine are not untethered by time and space after humans stop giving direct orders to the machine. No matter how long the machine runs without interventions, and no matter what changes occur inside the machine and its programming during execution (whether those changes were designed or were mistakes), the entire sociotechnical system is morally responsible for the actions of that machine (Ad Hoc Committee, 2010; Grodzinsky et al., 2012). And this again raises the point that ultimately, machine ethics will depend at least in part on human moral character.

Not all scholars agree with this eternal moral connection between developers and their machines. For example, Floridi and Sanders (2004) use levels of abstractions to model (from at least some perspectives) a moral distance between seemingly autonomous machines and their past creators and controllers. But Johnson and Miller (2008) do not accept that an abstract view of the situation should determine a machine's agency. Indeed, Johnson insists in this and other works that assigning moral agency is not a discovery of a fact of nature; the assigning of moral agency is a human choice, a decision that can be made and unmade. And Johnson maintains (and we agree) that the assigning of agency to a machine should not be allowed to untether relevant humans for their responsibilities (in all of Davis's nine variations) for the actions of their machines. Human law is rife with criteria for assigning culpability and responsibility that can be applied here, but which in any depth are well beyond the scope of this analysis.

However, our discussion here has been about moral responsibility, in the sense of accountability. There is also a moral sense of responsibility related to liability (where the duty to pay is a moral duty), but the legal liability of laws, fines, and suing for damages, is closely related to our moral sense of who (or what) is to blame for bad actions and bad consequences. Ideally, the law and ethics are closely related. Although this is not always the case, we can hope that future policies and laws will establish and enforce the kinds of responsibility we have outlined here, a responsibility that developers and owners of sophisticated machines will not be allowed to disavow.

How issues of AI, robots, and big data interact with some important frameworks?

When considering gamification and virtue development, Shields (2011) describes five character dimensions, four at the level of individuals, and one at the group level. We revisit those dimensions here and suggest how sophisticated machines matter in each.

Performance character influences how people work on tasks. Sophisticated machines change how people work with some tasks, and the increasing presence of these machines, and their increased functionality, will result in more people doing more tasks using AI, robots, and big data. If tasks are taken over from humans by the machines, then the people formerly involved in those tasks may become less motivated, less skilled, and ultimately less fulfilled with respect to those tasks. If instead, the people and the machines form a well-integrated team (de Laat, 2015), it could lead to people gaining new confidence and motivation, as well as capabilities, to accomplish this sort of task. This illustrates the dual potential of sophisticated machines: they may degrade performance character; they may enhance performance character. It is crucial to understand how the machines are employed before deciding if their use will be beneficial or detrimental to human flourishing.

Intellectual character influences how people interact with information. Greenman (2010) worries that the Internet kills curiosity, especially in youngsters; it is just too easy to look up easy answers. One way to avoid this problem is to ask more complex questions. But as sophisticated machines are used to focus Internet visitors into smaller and smaller subsets based on perceived interests, we fear that serendipity and intellectual courage could be discouraged, and easy questions and answers are encouraged. However, this need not be the case. If instead of narrowing our focus, intelligent machines could be programmed make suggestions that aim to broaden our view, to encourage us to explore alternative ideas, and to challenge our assumptions. As for performance character, machines offer positive and negative possibilities that depend on choices we make with our technology. If we tune the machines properly, it can encourage the flames of curiosity instead of quickly extinguishing them.

Civic character influences the degree to which people commit to a community. Some people interact with sophisticated machines (for example, with AI-driven online games) and largely withdraw from face-to-face communities. This discourages civic engagement. But some people use sophisticated machines to help them support relationships with people. Thus people and machines can, once again improve or harm civic character. In large representative democracies, AI can offer very sophisticated ways to participate in the common space searching for the common good and hence fertile grounds for the development and enactment of civic character. Or it can be used to prey on civic and intellectual character weaknesses to dupe people into counter-productive civic engagement or to stifle engagement altogether.

Moral character is that aspect of the self that motivates and enables one to be effectively ethical. It includes kindness and empathy. In so much as intelligent machines "nudge" their users towards kindness (Borenstein & Arkin, 2016), the machines encourage moral character. If instead, machines encourage meanness and alienation, moral character is diminished. (Research, such as (Dill et al., 2008), suggests that some virtual realities have detrimental effects on some gamers.) Much as is true of many informational media, such as television, the content and message can determine whether that technology enhances or thwarts the development of moral agency and character.

Collective character influences how groups of people interact, or "patterns of group life" (Bier & Coulter, 2014, p. 4). Examples include families and schools. The collective character of a group could be affected, positively or negatively, by sophisticated machines, either shared by individuals in the group, or developed and deployed by the group itself.

AI, robotics, and big data have become mainstream technologies. As they increase in sophistication, and as they become involved in ever more aspects of our lives, it becomes crucial that the people and organizations developing, buying, and deploying these machines do so prudently, giving careful consideration to the consequences of how they use them. The character and integrity of the people who make and use these machines should be a central concern of education today and in the foreseeable future.

References

AAAI. A brief history of AI. https://aitopics.org/misc/brief-history, accessed 10 August 2018.

Acemoglu, D., & Restrepo, P. (2017). Robots and jobs: Evidence from US labor markets. http://pascual.scripts.mit.edu/research/robots_jobs/robots_and_jobs_2017.pdf, accessed 24 September 2018.

The Ad Hoc Committee for Responsible Computing (2010), Moral responsibility for computing artifacts: The rules, https://edocs.uis.edu/kmill2/www/TheRules/, accessed 4 September 2018.

AISB (2018) Loebner prize. http://aisb.org.uk/events/loebner-prize, accessed 20 August 2018.

Allen, C., Varner G., and Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental and Theoretical Artificial Intelligence, 12*: 251-261.

Allen, Katie. (2015) Technology has created more jobs than it has destroyed, says 140 years of data. *The Guardian.* https://www.theguardian.com/business/2015/aug/17/technology-created-more-jobs-than-destroyed-140-years-data-census, accessed 14 August 2018.

American Heritage. Artificial Intelligence. *American Heritage Dictionary of the English Language, 5th Edition.* Houghton Miffline Harcourt Publishing Co. (2018), https://www.ahdictionary.com/word/search.html?q=artificial+intelligence, accessed 10 August 2018.

Baehr, J. (2011). *The inquiring mind: On intellectral virtues & virtue epistemology.* New York: Oxford University Press.

Baeza-Yates, R. (2016). Data and algorithmic bias in the web. In *Proceedings of the 8th ACM Conference on Web Science*, 1.

Banerjee, S. B. (2008). Corporate social responsibility: The good, the bad and the ugly. *Critical sociology*, *34*(1), 51-79.

Barrat, James. (2013) *Our Final Invention: Artificial Intelligence And The End Of The Human Era*. Macmillan, 2013.

Beir, M., & Coulter, R. (2014) The Gamification of Virtue Development. *Proceedings of Jubilee Center's 'Can Virtue be Measured?'*

Berkowitz, M.W., & Bier, M.C. (2014). Research-based fundamentals of the effective promotion of character development in schools. In L. Nucci, D. Narvaez, & T. Krettenauer (Eds.), *Handbook of moral and character education (Second Edition), (pp. 248-260).* New York: Routledge.

Borenstein, J., & Arkin, R. (2016). Robotic nudges: the ethics of engineering a more socially just human being. *Science and engineering ethics*, *22*(1), 31-46.

Business Insider Nordic. (2018) A Swedish bank has fired its world-famous AI assistant, Amelia. https://nordic.businessinsider.com/a-swedish-bank-just-fired-its-top-ranked-ai-colleague--heres-why--/ , accessed 3 September 2018.

Byrne, Patrick. (2018) 12 ways to automate your business and boost efficiency. *Entrepreneur*. https://www.entrepreneur.com/article/307286, accessed 15 August 2018.

Collins, Harry M., and Trevor Pinch. (2012) *The golem: What you should know about science*. Cambridge University Press.

Collins, W.R., K. W. Miller, B. Spielman, and P. Wherry. (1994) How good is good enough? An ethical analysis of software construction and use. *Communications of the ACM*, Vol. 37, No. 1, 81-91.

Craig, S. D., Graesser, A. C., & Perez, R. S. (2018). Advances from the Office of Naval Research STEM Grand Challenge: expanding the boundaries of intelligent tutoring systems. *International Journal of STEM Education*, *5*(1), 11.

Crow, T., Luxton-Reilly, A., & Wuensche, B. (2018, January). Intelligent tutoring systems for programming education: a systematic review. In *Proceedings of the 20th Australasian Computing Education Conference* (pp. 53-62).

Davis, M. (2012) "Ain't no one here but us social forces": Constructing the professional responsibility of engineers", *Science and Engineering Ethics* 18: 13-34.

de Laat, P. B. (2016). Trusting the (ro) botic other: by assumption? *ACM SIGCAS Computers and Society*, *45*(3), 255-260.

Dill, K. E., Brown, B. P., & Collins, M. A. (2008). Effects of exposure to sex-stereotyped video game characters on tolerance of sexual harassment. *Journal of Experimental Social Psychology*, *44*(5), 1402-1408.

Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2017). Concentrating on the Fall of the Labor Share. *American Economic Review*, *107*(5), 180-85.

Düwell, M. (2017). Human Dignity and the Ethics and Regulation of Technology. *The Oxford Handbook of Law, Regulation and Technology*, 177.

Etzioni, Amitai, and Oren Etzioni. "Incorporating ethics into artificial intelligence." *The Journal of Ethics* 21.4 (2017): 403-418.

Faggella, D. (2018) What is machine learning? *TechEmergence*, https://www.techemergence.com/what-is-machine-learning/, accessed 4 September 2018.

Federal Trade Commission (2015) *Internet of Things: Privacy and security in a connected world*. Technical report, https://www.ftc.gov/news-events/events-calendar/2013/11/internet-things-privacy-security-connected-world, accessed 25 September 2018.

Fisher, J.D., Rytting, M., & Heslin, R. (1976). Hands touching hands: Affective and evaluative effects of an interpersonal touch. *Sociometry, 39,* 416-421.

Frey, Carl Benedikt amd Michael A. Osborne. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*: 254-280.

Gartner. (2018) Gartner hype cycle. https://www.gartner.com/en/research/methodologies/gartner-hype-cycle, accessed 12 August 2018.

Gipps, C. (2002). *Beyond testing: Towards a theory of educational assessment*. Routledge.

González, C., Mora, A., & Toledo, P. (2014). Gamification in intelligent tutoring systems. In *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality*, 221-225.

Greenman, Ben (2010) Online curiosity killer. *New York Times Magazine*, https://www.nytimes.com/2010/09/19/magazine/19lives-t.html, accessed 26 September 2018.

Grodzinsky, F. S., Miller, K., & Wolf, M. J. (2003). Ethical issues in open source software. *Journal of Information, Communication and Ethics in Society*, *1*(4), 193-205.

Grodzinsky, F. S., K. Miller, and M. J. Wolf (2012). Moral responsibility for computing artifacts: “the rules” and issues of trust. *SIGCAS Comput. Soc*., *42*, 2 , 15-25.

Hannam, Keshia (2017) A record amount of brick and mortar stores will close in 2017. *Fortune*, http://fortune.com/2017/10/26/a-record-amount-of-brick-and-mortar-stores-will-close-in-2017/, accessed 25 September 2018.

Head, Simon (2014) Worse than Wal-Mart: Amazon’s sick brutality and secret history of ruthlessly intimidating workers. *Salon*, https://www.salon.com/2014/02/23/worse_than_wal_mart_amazons_sick_brutality_and_secret_history_of_ruthlessly_intimidating_workers/, accessed 25 September 2018.

Heppermann, Ann. (2012) Robots ate my job: Taking humans out of the supermarket checkout. *Marketplace.* https://www.marketplace.org/2012/03/27/tech/robots-ate-my-job/taking-humans-out-supermarket-checkout, accessed 12 August 2018.

Hopgood, Adrian A. "Artificial intelligence: hype or reality?" *Computer* 36.5 (2003): 24-28.

Hylen, M.G. (2008). *The impact of a character education based interactive discipline program on at-risk student behavior in an alternative school.* Dissertations. 531. https//irl.umsl.edu/dissertations/531.

Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and information technology*, *8*(4), 195-204.

Johnson, D. G. (2011). Computer systems: moral entities, but not moral agents. *Machine ethics*, 168-183.

Johnson, D. G., & Miller, K. W. (2008). Un-making artificial moral agents. *Ethics and Information Technology*, *10*(2-3), 123-133.

Kirkpatrick, K. (2016). Battling algorithmic bias: How do we ensure algorithms treat us fairly?. *Communications of the ACM*, *59*(10), 16-17

Knight, W. (2017). Biased algorithms are everywhere, and no one seems to care. *MIT Technology Review*. https://www. technologyreview. com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/, accessed 20 September 2018.

Kohlberg, L., Levine, C., & Hewer, A. (1983). Moral stages: A current formulation and a response to critics.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, *86*(1), 42-78.

Majdoub, M. (2016). *Promoting High School ESL Learners' Motivation and Engagement Through the Use of Gamified Instructional Design* (Doctoral dissertation, Concordia University).

Maly, Tim (2010). The Internet as curiosity machine. *The Atlantic*, https://www.theatlantic.com/technology/archive/2010/09/the-internet-as-curiosity-machine/63421/, accessed 26 September 2018.

Masoumi, Davoud (2015). Preschool teachers' use of ICTs: Towards a typology of practice. *Contemporary Issues in Early Childhood 16*, 1: 5-17.

Mayor, Adrienne. Bio-techne. *Aeon* (16 May 2016), https://aeon.co/essays/replicants-and-robots-what-can-the-ancient-greeks-teach-us, accessed 10 August 2018.

McVeigh, Tracy and Nick Paton Walsh (2000). Computers kill pupils' creativity. *The Guardian*. https://www.theguardian.com/uk/2000/sep/24/schools.news, accessed 15 August 2018.

Merriam Webster. Intelligence. *https://www.merriam-webster.com/dictionary/intelligence*, accessed 10 Aug. 2018.

Miller, K. and D. Larson. Angels and artifacts: Moral agents in the age of computers and networks. Journal of Information, Communication & Ethics in Society, Vol. 3, No. 3 (July, 2005), 151-157.

Miller, K.W., L. Morell, R. Noonan, S. Park, D. Nicol, B. Murrill, and J. Voas. (1992), Estimating the probability of failure when testing reveals no errors. *IEEE Trans. on Software Engineering*, Vol. 18, No. 1, 33-43.

Minsky, Marvin. *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon and Schuster, 2007.

Moor, J. H. (2001). The future of computer ethics: You ain't seen nothin'yet!. *Ethics and Information Technology*, *3*(2), 89-91.

Moorthy, R., & Pugazhenthi, S. (2017). Teaching psychomotor skills to autistic children by employing a robotic training kit: A pilot study. *International Journal of Social Robotics*, *9*(1), 97-108.

Morell, John. (2017) Self-driving mining trucks. *ASME*. https://www.asme.org/engineering-topics/articles/robotics/selfdriving-mining-trucks, accessed 14 August 2018.

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

OECD (ND) Work-life balance. http://www.oecdbetterlifeindex.org/topics/work-life-balance/, accessed 22 September 2018.

O'Hara, Mark (2008) Young children, learning and ICT: a case study in the UK maintained sector, Technology, Pedagogy and Education, 17:1, 29-40.

Oppenheimer, Todd. (2003) *The Flickering Mind: The False Promise of Technology in the Classroom and How Learning Can Be Saved.* Random House.

Oppy, G. and D. Dowe (2018) The Turing test. *The Stanford Encyclopedia of Philosophy,* Ed. E. Zalta, https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=turing-test, accessed 29 August 2018.

Parnas, David Lorge. (1985): "Software aspects of strategic defense systems." *Communications of the ACM* 28.12: 1326-1335.

Parnas, David L. (2017) The real risks of artificial intelligence. *Communications of the ACM, 10,* 10: 27-31.

Parry, Richard (2014) Ancient ethical theory", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2014/entries/ethics-ancient/, accessed 2 September 2018.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon and Schuster.

Richard, J. A. (2017). Equality and equal opportunity for welfare. In *Theories of Justice* (pp. 75-91). Routledge.

Roosevelt, F.D. (2018). Franklin D. Roosevelt quotes. https://www.quotes.net/quote/9563, accessed 27 September 2018.

Rowe, Jonathan (1995). Corporations and the public interest: a look at how the original purpose behind corporate charters has been lost. https://www.context.org/iclib/ic41/rowe/, accessed 23 September 2018.

Russell, S., Hauert, S., Altman, R., & Veloso, M. (2015). Ethics of artificial intelligence. *Nature, 521*(7553), 415-416.

Shneiderman, B. (2003). *Leonardo's laptop: human needs and the new computing technologies*. Mit Press.

Semuels, Alana (2017) The parts of America most susceptible to automation. *The Atlantic.* https://www.theatlantic.com/business/archive/2017/05/the-parts-of-america-most-susceptible-to-automation/525168/, accessed 11 August 2018.

Sharkey, A. (2014). Robots and human dignity: a consideration of the effects of robot care on the dignity of older people. *Ethics and Information Technology, 16*(1), 63-75.

Shields, D. (2011). Character as the aim of education. *Kappan, 92*(8), 48-53.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, *46*(5), 30.

Skirpan, M. W., Yeh, T., & Fiesler, C. (2018) What's at Stake: Characterizing Risk Perceptions of Emerging Technologies. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 70). ACM.

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, *106*(2), 331.

Sullins, John P. (2006) When is a robot a moral agent? *International Review of Information Ethics 8*: 24-30.

Sundman J. Artificial stupidity (2003), *Salon.com*, https://www.salon.com/2003/02/26/loebner_part_one/, accessed 4 September 2018.

Swink, M., Talluri, S., & Pandejpong, T. (2006). Faster, better, cheaper: A study of NPD project efficiency and performance tradeoffs. *Journal of Operations Management*, *24*(5), 542-562.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, *361*(6404), 751-752.

Talaei-Khoei, A., Lewis, L., Kaul, M., Daniel, J., & Sharma, R. (2017). Use of Lean Robotic Communication to Improve Social Response of Children with Autism.

Torresen, Jim. "A Review of Future and Ethical Perspectives of Robotics and AI." *Frontiers in Robotics and AI* 4 (2018): 75. https://www.frontiersin.org/articles/10.3389/frobt.2017.00075/full

Tucker, Eleanor. (2015). How robots are helping children with autism. *The Guardian.* https://www.theguardian.com/lifeandstyle/2015/feb/01/how-robots-helping-children-with-autism, accessed 22 September 2018.

Turing, A. (1950). Computing machinery and intelligence. *Mind, LIX,* (236), 433-460.

Turkle, Sherry. "Artificial intelligence and psychoanalysis: A new alliance." *Daedalus* (1988): 241-268.

Turkle, Sherry. *The second self: Computers and the human spirit*. Mit Press, 2005.

Turkle, Sherry. "Always-on/always-on-you: The tethered self." *Handbook of mobile communication studies* 121 (2008).

Turkle, Sherry. *Life on the Screen*. Simon and Schuster, 2011.
(Google Scholar lists nearly 13000 citations to this book.)

Turkle, Sherry. *Alone together: Why we expect more from technology and less from each other*. Hachette UK, 2017.

Vanderelst, Dieter, and Alan Winfield. "The dark side of ethical robots." *arXiv preprint arXiv:1606.02583* (2016).

Wagner, A.R., J. Borenstein, and A. Howard. Overtrust in the robotic age. *Communications of the ACM, 61*, 9, 22-24.

Wallach, W., & Allen, C. (2008). *Moral machines: Teaching robots right from wrong*. Oxford University Press.

Wang, Y. (2016). Big opportunities and big concerns of big data in education. *TechTrends*, *60*(4), 381-384.

Wigan, M. R. and R. Clarke, "Big Data's Big Unintended Consequences," in *Computer*, vol. 46, no. 6, pp. 46-53, June 2013.

Wikipedia contributors. (2018, August 8). Big data. In *Wikipedia, The Free Encyclopedia*. Retrieved 23:17, August 12, 2018, from https://en.wikipedia.org/w/index.php?title=Big_data&oldid=854006873.

M. J. Wolf, K. Miller, and F. S. Grodzinsky (2017), Why we should have seen that coming: comments on Microsoft's Tay "experiment," and wider implications. *SIGCAS Comput. Soc. 47*, 3, 54-64.

Wolfe, Alan. "Mind, self, society, and computer: Artificial intelligence and the sociology of mind." *American Journal of Sociology* 96.5 (1991): 1073-1096.

Woodie, Alex. (2017) How AI fares in Gartner's lastest hype cycle. *Datanami.* https://www.datanami.com/2017/08/29/ai-fares-gartners-latest-hype-cycle/, accessed 12 August 2018.

Zagzebski, M. R. D. L. T. (2003). *Intellectual virtue: Perspectives from ethics and epistemology*. Clarendon Press.

Ananthaswamy, A. (2017). That's a termite colony between your ears. *New Scientist, 233*(3112), 42-43.

Aston-Jones, G., & Cohen, J. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403-450.

Augustine, D., Chrona, J., C, H., & Williams, L. (2018). *Meaningful Reconciliation: Indigenous knowledges flourishing in B.C.'s K-12 education system for the betterment of all students.*

Avvisati, F., Jacotin, G., & Vincent-Lancrin, S. (2013). Educating Higher Education Students for Innovative Economies: What International Data Tell Us. *Tuning Journal for Higher Education*(1), 223-240.

Baumeister, R., & Brewer, L. (2012). Believing versus disbelieving in free will: Correlates and consequences. *Social and Personality Psychology, 6*(10), 736-745.

Baumeister, R., Crescioni, A., & Alquist, J. (2011). Free will as advanced action control for human social life and culture. *Neuroethics, 4*(1), 1-11.

Behrens, T., Woolrich, M., Walton, M., & Rushworth, M. (2007). Learning the value of information in an uncertain world. *Nature neuroscience, 10*(9), 1214.

Bennis, W. G., & Nanus, B. (1985). *Leaders: The Strategies for Taking Charge.* New York: Harper & Row.

Bestmann, S., Ruge, D., Rothwell, J., & Galea, J. (2014). The role of dopamine in motor flexibility. *Journal of cognitive neuroscience, 27*(2), 365-376.

Biggs, J. (1985). The role of metalearning in study processes. *British Journal of Educational Psychology, 55*, 185-212.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2), 199-231.

Brooks, A. (2014). Get excited: Reappraising pre-performance anxiety as excitement. *Journal of Experimental Psychology: General, 143*(3), 1144.

Brooks, A., Schroeder, J., Risen, J., Gino, F., Galinsky, A., Norton, M., & Schweitzer, M. (2016). Don't stop believing: Rituals improve performance by decreasing anxiety. *Organizational Behavior and Human Decison Processes, 137*, 71-85.

Brown, A., & Kane, M. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20*(4), 493-523.

Buhn, K., & Dugas, M. (2002). The intolerance of uncertainty scale: Psychometric properties of the English version. *Behaviour research and therapy, 40*(8), 931-945.

Campbell, M., Hoane, J., & Hsu, F. (2002). Deep Blue. *Artificial intelligence, 134*(1-2), 57-83.

Clark, A. (2013). Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences, 36*(3), 181-204.

Collins, A., & Frank, M. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological review, 120*(1), 190.

De Berker, A. O. (2016). Acute stress selectively impairs learning to act. *Scientific Reports, 6*(29816).

de Berker, A., Rutledge, R., Mathys, C., Marshall, L., Cross, G., Dolan, R., & Bestmann, S. (2016). Computations of uncertainty mediate acute stress responses in humans. *Nature communications, 7*, 10996.

de Freytas-Tamura, K. (2018, March 19). What’s Next for Humanity: Automation, New Morality and a ‘Global Useless Class’. *New York Times*.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and learning, 3*(3), 231-264.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: changes in grey matter induced by training. *Nature, 427*(6972), 311.

Dugas, M., Gagnon, F., Ladouceur, R., & Freeston, M. (1998). Generalized anxiety disorder: A preliminary test of a conceptual model. *Behaviour research and therapy, 36*(2), 215-226.

Dweck, C., Davidson, W., Nelson, S., & Enna, B. (1978). Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. *Developmental psychology, 14*(3), 268.

Edcuation2030_2. (2018). Progress report 7th IWG.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science, 338*(6111), 1202-1205.

Eysenck, M., Derakshan, N., Santos, R., & Calvo, M. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion, 7*(2), 336.

Freeston, M., Rheaume, J., Letarte, H., Dugas, M., & Ladouceur, R. (1994). Why do people worry? *Personality and individual differences, 17*(6), 791-802.

Frey, C., & Osborne, M. (2013). The Future of Employment.

Frey, C., & Osborne, M. (2013, September 2013). *The Future of Employment: How Susceptible are jobs to Computerisation.* Retrieved from Oxford Martin Programme on Technology and Employment: http://sep4u.gr/wp-content/uploads/The_Future_of_Employment_ox_2013.pdf

Frey, C., & Osbourne, M. (2017). The Future of Employment: How susceptible are Jobs to Computerization? *Technological Forcastong & Social Change, 114*, 254-280.

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325-1352.

Friston, K. (2005). A Theory of Cortical Responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1456), 815-836.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences, 13*(7), 293-301.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127.

Friston, K. (2013). Learning and inference in the brain. *16*(9), 1325-1352.

Frith, C. (2012). Explaining delusions of control: The comparator model 20 years on. *Consciousness and cognition*, *21*(1), 52-54.

Greene, B. (2006, October 20). *The Universe on a String.* Retrieved from The New York Times: https://www.nytimes.com/2006/10/20/opinion/20greenehed.html

Halgreen, T. (2018). *Mathematics in PISA, presentation at the Mathematics Curriculum Document Analysis workshop* .

Haste, H. (2018). *Attitudes and Values and the OECD Learning Framework 2030: A critical review of definitions, concepts and data.*

Heine, S. J. (2006). The meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review, 10*(2), 88-110.

Hirsch, E. (1988). *Cultural literacy: What every American Needs to Know.* Random House.

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological review, 119*(2), 304-320.

Hirsh, J., Mar, R., & Peterson, J. (2013). Personal narratives as the highest level of cognitive integration. *Behavioral and Brain Sciences, 36*(3), 216-217.

Hohwy, J. (2013). *The predictive mind.* New York, NY, US: Oxford University Press.

Hsu, F., Campbell, M., & Hoane, A. J. (1995). Deep Blue system overview. *Proceedings of the 9th international conference on Supercomputing* (pp. 240-244). ACM.

Inzlicht, M., & Tullett, A. (2010). Reflecting on God: Religious primes can reduce neurophysiological response to errors. *Psychological Science, 21*(8), 1184-1190.

Inzlicht, M., McGregor, I., Hirsh, J., & Nash, K. (2009). Neural markers of religious conviction. *Psychological Science, 20*(3), 385-392.

Johansen, B. &. (2013). Navigating the VUCA world. *Research-Technology Management, 56*(1), 10-15.

Karpicke, J., & Blunt, J. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 1199327.

Kasdan, T., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical psychology review, 30*(7), 865-878.

Kossowska, M., Bukowski, M., Guinote, A., Dragon, P., & Kruglanski, A. (2016). Self-image threat decreases stereotyping: The role of motivation toward closure. *Motivation and emotion, 40*(6), 830-841.

Kossowska, M., Szwed, P., Wronka, E., Czarnek, G., & Wyczesany, M. (2016). Anxiolytic function of fundamentalist beliefs: neurocognitive evidence. *Personality and Individual Differences, 101*, 390-395.

Kurzweil, R. (2005). *The Singularity is Near.* New York: New York: Viking Books.

Lake, B., Ullman, T., Tenenbaum, J. D., & Gershman, S. J. (2017). Building Machines That Learn and Think Like People. *Behavioral and Brain Sciences, 40*.

Lang, M., Kratky, J., Shaver, J., Jerotijevic, D., & Xygalatas, D. (2015). Effects of anxiety on spontaneous ritualized behavior. *Current Biology, 25*(14), 1892-1897.

Laukkonen, R., Biddell, H., & Gallagher, R. (2018). *Preparing humanity for change and artificial intelligence: Learning to learn as a safeguard against volatility, uncertainty, complexity and ambiguity.*

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature, 521*, 436-444.

Lynn, M., Muhle-Karbe, P., Aarts, H., & Brass, M. (2014). Priming determinist beliefs diminishes implicit (but not explicit) components of self-agency. *Frontiers in psychology, 5*, 1483.

Marshall, J., & Oberwinkler, J. (1999). Ultraviolet vision: The colourful world of the mantis shrimp. *Nature, 401*(6756), 873.

Marshall, L., Mathys, C., Ruge, D., de Berker, A., Dayan, P., Stephan, K., & Bestmann, S. (2016). Pharmacological fingerprints of contextual uncertainty. *PLoS Biology, 14*(11), e1002575.

Maudsley, D. (1979). A theory of meta-learning and principles of facilitation: An organismic perspective (Unpublished Master's thesis). University of Toronto, Canada.

Maudsley, D. (1980). A theory of meta-learning and principles of facilitation: An organismic perspective. *Unpublished Masters Thesis*.

McEvoy, P., & Mahoney, A. (2011). Achieving certainty about the structure of intolerance of uncertainty in a treatment-seeking sample with anxiety and depression. *Journal of Anxiety Disorders, 25*(1), 112-122.

McEwen, B. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York academy of sciences, 840*(1), 33-44.

McEwen, B., & Stellar, E. (1993). Stress and the individual: mechanisms leading to disease. *Archives of Internal Medicine, 153*(18), 2093-2101.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., . . . Peterson, S. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529.

Moore, G. (1975). Progress In Digital Integrated Electronics. *Technical Digest*, 11-13.

Moore, J., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and cognition, 17*(1), 136-144.

Nasuno, K. (2017). *Analysing students' knowledge acquisition in online Learning platforms using deep learning. Paper commissioned by OECD* .

Newell, A., & Simon, H. A. (1961). GPS, a program that simulates human thought.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior, 32*(2), 303-330.

Palmer, J., & Chakravarty, A. (2014). Supervised machine learning. In A. Haney, D. Bowman, & A. Chakravarty, *An Introduction To High Content Screening: Imaging Technology, Assay Development, and Data Analysis in Biology and Drug Discovery* (p. 231). Hoboken, NJ: John Wiley & Sons Inc.

Pasmore, B., & O'Shea, T. (2010). Leadership agility: A business imperative for a VUCA world. *People and Strategy, 33*(4), 32.

Peters, A., & McEwen, B. (2015). Stress habituation, body shape and cardiovascular mortality. *Neuroscience & Biobehavioral Reviews, 56*, 139-150.

Peters, A., McEwen, B., & Friston, K. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in neurobiology, 156*, 164-188.

Peters, A., Schweiger, U., Pellerin, L., Hubold, C., Oltmanns, K., Conrad, M., . . . Fehm, H. (2004). The selfish brain: competition for energy resources. *Neuroscience & Biobehavioral Reviews, 28*(2), 143-180.

Proulx, T., & Heine, S. (2009). Connections from Kafka: Exposure to meaning threats improves implicit learning of an artificial grammar. *Psychological Science, 20*(9), 1125-1131.

Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in cognitive sciences, 16*(5), 285-291.

Roediger III, H., & Butler, A. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences, 15*(1), 20-27.

Rougier, N., Noelle, D., Braver, T., Cohen, J., & O'Reilly, R. (2005). refrontal cortex and flexible cognitive control: Rules without symbols. *roceedings of the National Academy of Sciences, 102*(20), 7338-7343.

Schneider, P., & Bakhshi, H. (2017). *The Future of Skills: Employment in 2030, NESTA.*

Servan-Schreiber, D., Printz, H., & Cohen, J. (1990). A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science, 249*(4971), 892-895.

Shariff, A., Greene, J., Karremans, J., Luguri, J., Clark, C., Schooler, J., . . . Vohs, K. (2014). Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological Science, 25*(8), 1564-1570.

Swanson, L. (2016). The predictive processing paradigm has roots in Kant. *Frontiers in systems neuroscience, 10*, 79.

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: an interplay between prediction and postdiction. *Frontiers in psychology, 4*, 127.

Tolin, D., Abramowitz, J., Brigidi, B., & Foa, E. (2003). Intolerance of uncertainty in obsessive-compulsive disorder. *Journal of anxiety disorders, 17*(2), 233-242.

Turing, A. M. (1950). COMPUTING MACHINERY AND INTELLIGENCE. *Mind*.

Vohs, K., & Schooler, J. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science, 19*(1), 49-54.

Voogt, J., Nieveen, N., & Klopping, S. (2016). *Curriculum overload: a literature study.*

Wagenaar, W., & Sagaria, S. (1975). Misperception of exponential growth. *Perception & Psychophysics, 18*(6), 416-422.

Wang, X. Z., Ashfaq, R. A., & Fu, A. M. (n.d.). Fuzziness based sample categorization for classifier performance improvement. *ournal of Intelligent & Fuzzy Systems, 29*(3), 1185-1196.

Wegner, D., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist, 54*(7), 480.

World Economic Forum. (2016). *The Future of Jobs, Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution.*

Young, M. (2013). Overcoming the crisis in curriculum theory: a knowledge-based approach. *Journal of Curriculum Studies, 45*(2).

Young, M., & Muller, J. (2016). *Curriculum and the specialization of knowledge.* Routledge.

Yu, A., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681-692.

Zimmerman, B. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology, 25*(1), 82-91.