

# The Institute for Ethical AI in Education

## Interim Report

Towards a shared Vision of Ethical AI in Education



THE UNIVERSITY OF  
BUCKINGHAM



# Contents

2 - Foreword by Sir Anthony Seldon, Professor Rose Luckin, Priya Lakhani OBE, Lord Tim Clement-Jones

4 - An Introduction to AI in Education

7 - Risks and Benefits of AI in Education

18 - Developing an Ethical Framework for AI in Education

25 - Realising a Shared Vision of Ethical AI in Education

33 - References

# Foreword

The Institute for Ethical AI in Education was founded to enable all learners to benefit optimally from Artificial Intelligence (AI) in education, and to be protected against the risks this technology presents. Our definitions of ‘learners’ and ‘education’ are appropriately broad. From young children using smart toys to adults learning via adaptive tutoring software, all learners deserve to benefit from ethical innovation - from cradle to grave.

The risks posed by AI in education are profound. But so too are the benefits that could be leveraged to tackle entrenched injustices faced by learners across the globe, and here in the UK. Learners, educators and societies as a whole should without question be cautious about introducing AI into any form of learning environment. However, there could well be tremendous opportunity costs to overzealous levels of caution. The right balance needs to be struck. We want to see AI being used to increase access to education, to advance teaching and learning, and to broaden educational opportunities - thereby enabling inclusive, holistic human flourishing. But we cannot allow learning to be dehumanised or become a means of surveillance and control. And we must certainly not allow AI to widen the gaps between the privileged and the vulnerable. This is non-negotiable. It would be utterly shameful if this transformative technology becomes commonplace, but societies still routinely fail to equitably support learners who face the greatest levels of disadvantage and discrimination. That is why The Institute for Ethical AI in Education is especially committed to protecting the rights of the most vulnerable learners and enabling AI to benefit everyone, not just the privileged few.

An ethical approach is essential for achieving these goals. Ethical thinking will crystallise people's understanding of the benefits and risks of AI in education, and ensure that the concrete measures to realise responsible practice are grounded in human values. The purpose of this interim report, however, is not to put forward an ethical framework for AI in education. Our role is not to conjure up an ethical framework in isolation. It is to provide a platform for the perspectives of stakeholders and experts to be listened to and learned from. This report sets out how we can collectively develop a shared vision of ethical AI in education and together decide on the structures needed to support this vision. We encourage educators, academics, activists, technologists, parents and of course learners to contribute to this shared vision.

The interim report is also intended to inform stakeholders about AI, its applications in education, its overarching risks and benefits, and the underlying ethical implications. Harnessing the potential of AI in Education will require responsibility and application from all individuals and organisations involved in the design, development, and deployment of AI in Education. From the drawing room to the classroom, professionals involved in utilising this innovation will need to be empowered to make ethical decisions in the best interests of learners. Educating people about AI in education is hence a necessary starting point.

Sir Anthony Seldon

Professor Rose Luckin

Priya Lakhani, OBE

Lord Tim Clement-Jones

# 1. An Introduction to AI in Education

## What is artificial intelligence?

Broadly defined, AI refers to technologies that can perform tasks or functions that would usually require human intelligence or sensory abilities. Early AI systems relied on rules-based algorithms, meaning that human developers predefined explicit rules that allowed the systems to yield a specific output based on a given set of inputs. As an example, in 1966 Massachusetts Institute of Technology Professor, Joseph Weizenbaum developed a Natural Language Communication programme called Eliza. The programme responded to text inputs from a human in a way that attempted to emulate human conversation. To illustrate, in one instance a human user communicated to Eliza that “you are like my father in some ways.” to which Eliza responded “what resemblance do you see?” To select this as an appropriate response, the system had followed a set of rules including scanning the input sentence for significant key words (based on a predefined ranking), and applying transformations of sentence structure to ask a question based on the information provided. Some modern AI systems, including AIEd tutoring systems, also uses rules, but in a much more sophisticated way than these early systems.

## What is Machine Learning?

Machine Learning is a subset of AI. With Machine Learning, explicit rules for completing a task are not programmed by a human. Instead, the system is trained using large amounts of data in effect to develop its own set of rules that can be used to complete a task or reach an optimum outcome.

There are three categories of machine learning: supervised, unsupervised and reinforcement. **Supervised Machine Learning** can be used, for instance, to identify and filter out spam emails. To do this, systems will have been trained with large datasets of emails that had been pre-labelled as either spam or not spam by a human. Humans are able to use a tacit understanding of emails to identify those which are spam from those which are not. The machine learning system is then able to develop its own set of rules based on the training data to make these identifications automatically.

With **Unsupervised Machine Learning**, training data has not been previously labelled by humans, often because humans would not be able to make appropriate judgements in the

first instance. Instead, with unsupervised machine learning, systems are trained on large data-sets, and through statistical analysis can uncover and learn patterns and structures within the data. For instance, unsupervised machine learning can be used to detect fraudulent banking activity based on an evaluation of numerous statistical factors. Because humans are not capable of such multi-dimensional statistical analysis, unsupervised machine learning can often be used to provide insights humans would not have been able to achieve alone.

With **Reinforcement Machine Learning**, systems are configured to optimise a particular outcome and the machine learning algorithm uses an iterative process of trial and error in order to optimise that outcome. Reinforcement machine learning is often used in gaming, where systems learn to play a game to a standard often as good as - if not better than - humans.

AI systems can use a combination of rules-based and machine learning approaches.

## **What is AI in Education?**

Artificial Intelligence in Education (AIEd) can be defined as the application of AI for learning, for the assessment of learning or for the achievement or facilitation of other core educational goals, such as safeguarding or educational administration. AIEd can recognise and leverage the tremendous gains that have been made in our understanding of how humans learn through advances in psychology, neuroscience and education research. It is important to remember that in parallel with the advances in AI technology, there have been significant advances in our understanding of how the human brain learns. The promise of AIEd stems from the combination of these two tracks of progress.

# How is AI used in education?

## Adaptive Learning Platforms

**How are they used?** Learners access learning material that adapts based on their needs.

**Why are they used?** Learners have different strengths and needs. AI systems are able to identify, for instance, particular gaps in learners' knowledge or understanding, and tailor learning material to address such gaps.

**Where might ethical issues arise?** These systems could, for example, be used in ways that undermine teachers.

## Predictive Analytics

**How are they used?** As an example, predictive analytics can be used to identify university students at risk of not completing or passing their course. With these identifications, universities can intervene as appropriate.

**Why are they used?** Predictive analytics tools can detect patterns and spot indicators of risk based on statistical analysis; something humans would not necessarily be able to do.

**Where might ethical issues arise?** Institutions could, for example, take unfair action based on predictions that were not in the best interest of at-risk individuals. These actions could lead to people being systematically discriminated against based on factors such as gender, ethnicity, and socio-economic background.

## Automated Essay Scoring systems

**How are they used?** These systems can grade student essays, a task that is usually performed by an educator.

**Why are they used?** Such tools could provide support with managing teachers' workloads and extending their capacities. Furthermore, these systems could provide additional flexibility and levels of support to learners, as essays could be marked quicker and at the demand of the learner.

**Where might ethical issues arise?** If ineffective, biased or inaccurate systems are used learners could be unfairly treated- particularly if these were used in high stakes situations. Also, quality of educational provision could be reduced if teachers were systematically underutilised with regards to giving feedback on essays.

## Smart Toys

**How are they used?** Children interact with toys, which are able to respond to them based on inputs such as speech.

**Why are they used?** One intended benefit is language development. Techniques such as Voice Recognition allow these toys to identify what children are saying and respond appropriately.

**Where might ethical issues arise?** Children could disclose sensitive information, for example. And this information could be used inappropriately, or hacked.

## Facial Recognition Software

**How is it used?** Learners' facial expressions are monitored whilst learning or whilst present at an educational institution.

**Why is it used?** Monitoring learners' facial expressions could provide insights into learner behaviours, which in turn could allow teaching and learning to be enhanced through relevant interventions. For instance, information gathered through this process could allow teachers to adapt their practice, or provide learners with advice on increasing their own engagement.

**Where might ethical issues arise?** This process requires highly intimate data - such as an individual's level of focus and their emotional status - to be processed and stored. This data could be used in ways that could harm individuals. For instance, oppressive states could use such insights to identify non-compliant individuals.

## 2. Risks and Benefits of AI in Education

Many bold claims have been made about the benefits that artificial intelligence (AI) could bring about for learners. UNESCO has proposed that by providing opportunities for personalised learning at scale, AI could contribute to the achievement of Sustainable Development Goal 4 - ensuring an inclusive and equitable education and promoting life-long learning opportunities for all<sup>1</sup>. NESTA has suggested that AI resources could help boost social mobility by enhancing teaching and learning and improving access to high quality learning materials<sup>2</sup>. It has also been argued that AI in Education (AIEd) could facilitate transformations in assessment systems, enable access to 'life-long learning companions' for all, help resolve the teacher recruitment and retention crisis in the UK, and address the global shortage of teachers.<sup>3</sup>

These forecasted benefits of AI in Education should not be ignored. Across the globe 617 million young people lack basic literacy and mathematics skills and 265 million children are not in school at all<sup>4</sup>. In the UK, the attainment gap at 16 between advantaged and disadvantaged children is equivalent to 18 months<sup>5</sup> - a problem that is exacerbated by the fact that 40% of teachers are considering leaving the profession within the next five years<sup>6</sup>. Globally, education systems suffer from a chronic undersupply of teachers, with UNESCO predicting that 69 million additional teachers are needed to enable all young people to receive a good standard of education<sup>7</sup>. The potential benefits of AI in Education hence need to be considered carefully.

But before leveraging AI to address these problems, we must ask two critical questions:

- To what extent can AI enable genuine solutions to these problems?
- And how can the projected benefits be achieved without causing harm to learners by undermining their development and/or violating their rights?

## Possible benefits of AI in Education

The benefits posed by AIEd stem from three fundamental factors.

- AI can **increase capacity** within education systems and increase the productivity of educators
- AIEd can **provide valuable insights** that can enhance teaching and learning, and support learners' well-rounded development
- AIEd can deliver **autonomous learning recommendations**

## Increasing Capacity and Productivity

### Why additional Capacity is needed

- Teachers face unsustainable workloads. Unmanageable workloads have led to a situation where 81% of teachers considered leaving the profession in 2018<sup>8</sup>. Workload pressures are contributing to a recruitment and retention crisis at a time when school-age population continues to rise. Resources that increase teachers' productivity could decrease workload pressures and hence contribute to solving the teacher recruitment and retention crisis.
- There is a global shortfall of 69 million teachers. It may not be possible to recruit and retain enough teachers to fill this gap, meaning tools that increase the capacity of education systems could help ensure all learners around the world have access to a high-quality education.
- As automation disrupts labour markets, there may be a need for adults to have ongoing access to educational opportunities so that they may retrain and up-skill throughout life.<sup>9</sup> This would mean significant increases on current levels of education provision.

## How AI could provide additional capacity

- Digital assistants could support educators by completing administrative and learning management tasks, such as compiling and distributing teaching materials.<sup>10</sup>
- Automated Essay Scoring systems could support educators by marking learners' work and providing feedback.<sup>11</sup>
- Life-long learning could be partially facilitated by online courses that use artificial intelligence to increase engagement and efficacy and to provide lifelong evaluation of learner progress.

## Providing Valuable Insights

### How insights can enhance education provision

- Information is key to effective teaching and learning. In the words of Paul Black and Dylan Wiliam “Teachers need to know about their pupils' progress and difficulties with learning so that they can adapt their own work to meet pupils' needs”<sup>12</sup>. Teachers, they argue in their authoritative paper *Inside the Black Box*, should always start by asking whether they have sufficient knowledge of each of their students' understanding to be able to support them, as it is this information and insight that allows educators to fully enable learning. Information on their own learning can also enable learners to become more independent, strengthening their autonomy.
- High quality feedback, which requires detailed information of learners' performances and the process of learning, is amongst the most effective interventions educators can make to improve learning.<sup>13</sup>
- According to eminent education researcher John Hattie, teachers also need information on the impact of interventions, activities and methods of instruction, so that they can constantly adapt their teaching to suit the needs of learners.<sup>14</sup>

### How AI can provide these insights

- As Professor Rose Luckin argues, “AI is a powerful tool that can open up the ‘black box’ of learning, providing a deep, fine-grained analysis of what pupils are doing as they learn, meaning their learning can be ‘unpacked’ as it happens.” This functionality can be used to provide insights to both teachers and learners.<sup>15</sup>

- Data-mining techniques informed by what we know about how human learning happens can reveal the relative impacts of different resources and interventions<sup>16</sup>
- AI Insights could provide learners and teachers with information that cannot currently be gleaned, such as insights into engagement levels, memory function, and metacognitive skills.<sup>3</sup>

As well as supporting the process of teaching and learning directly, there are clear indications that AI could support the achievement of other educational aims by providing educators with rich insights that could not have been gained without sophisticated data analysis. Examples include:

- The use of predictive analytics to identify Higher Education students at risk of dropping out of their course, which allows appropriate interventions to be designed and delivered.<sup>17</sup>
- The use of AI safeguarding systems that can support an understanding of learners' mental health.<sup>18</sup>

It is important to emphasise that these functionalities of AI could be misused, with information utilised to justify unfair or discriminatory actions. However, in principle, these functionalities could certainly achieve positive outcomes for learners if used responsibly.

### **Transforming assessments**

AI's ability to extract meaningful insights from data has led to forecasts of AI giving rise to a transformation in how we formally assess learning and how we understand the learning process.

- As noted by Professor Luckin, "the prevailing exam paradigm is stressful, unpleasant, can turn students away from education, and requires that both students and teachers take time away from learning." With AI supported assessment systems, AI could continuously monitor and assess learning over a long period of time to build up a representative picture of a student's knowledge, skills and abilities- and of how they learn.<sup>15</sup>
- AI could also transform how education institutions are assessed. Supervised Machine Learning systems are already being used within the school inspection process to identify schools at risk of underperformance.<sup>19</sup>

## Autonomous Learning Recommendations

The Education Endowment Foundation has shown that one to one tuition delivered by experienced professionals in coordination with classroom teaching, can result in the equivalent of five additional months of progress for learners per year. With such interventions, learners benefit from support that is tailored to their individual needs. The problem, however, is that these interventions are highly expensive, with costs estimated to be in the range of £700 per pupil.<sup>20</sup> The Education Endowment Foundation has also identified individualised instruction as an effective intervention for pupils. With individualised instruction, individual learners complete tasks, and receive support and feedback based on their own needs. Individualised instruction can be achieved in a number of ways and is often associated with 'very low' costs. That said, although the average impact is equivalent to three additional months of progress for learners, there is significant variation in the impact of this intervention. Interventions that use digital technology to individualise instruction have tended to show higher impacts.<sup>21</sup>

Through autonomous learning recommendations, AIEd enables individualised instruction and personalised learning at scale. These systems use a broad range of data inputs from learners to adapt learning experiences to individual users. Individualised outputs could include changing the difficulty of questions; adapting the order in which learners move through content; offering the most appropriate learning resources; and providing all-important feedback to questions, which can be personalised, scaffolded and presented at the optimum time.

Platforms that achieve personalised learning utilise both AIEd's ability to increase capacity in education and to harness rich insights on learning. Effective platforms are able to achieve a high quality learning experience because they develop insights into learners' needs through the intelligent analysis of data about individuals' interactions with the platform based on a sound understanding of how people learn. They take appropriate action to personalise learning based on this intelligent analysis. These resources also increase capacity because it would not be possible for educators to analyse information on learners and make decisions on how best to personalise learning at the rate that AI platforms can. For instance, a teacher could never provide every student in a class of 30 with tailored feedback simultaneously. Importantly, Intelligent Tutoring Systems and Adaptive Learning Platforms provide opportunities for learners to access personalised learning on-demand, with flexibility, and in unlimited quantities; this is a key reason why Intelligent Tutoring Systems and Adaptive Learning Platforms present credible opportunities for life-long learning.

It should also be recognised that personalised learning does not dilute teachers' importance in the classroom. Empowering teachers with AI may be the key to optimally personalised education. AIED's propensity to increase teacher productivity by performing routine and administrative tasks, could lead to a situation where teachers have sufficient capacity to devote time to providing learners with one to one support. In providing this highly personalised support, teachers would be able to make use of rich insights into their students' needs. These insights, which could be gained by utilising AI resources, could be represented in the form of an 'open-learner model' that demonstrates students' performance relative to a wide range of attributes and competencies<sup>15</sup>.

**In summary**, there is good reason to believe that AI can systematically have a positive impact on learners, and play a role in solving major educational challenges. But this does not justify harnessing AI without understanding, mitigating and preventing the risks it presents to learners.

## Possible risks of AI in Education

The broader conversation around ethical AI provides a strong foundation from which to understand the harms that AI could cause to learners. Numerous authoritative ethical guidelines and frameworks have been produced including the Asimolar AI Principles<sup>22</sup>, The Responsible Machine Learning Principles<sup>23</sup>, The Montreal Declaration for Responsible Development of Artificial Intelligence<sup>24</sup>, and the "five overarching principles for an AI Code" put forward by The House of Lords Select Committee on Artificial Intelligence<sup>25</sup>. Of the available frameworks for ethical AI, The Ethics Guidelines for Trustworthy AI<sup>26</sup> (developed by The Higher Level Expert Group on AI, a group which was established by, reports to and advises on policy for The European Commission) provides a strong model for understanding the risks AI poses as it 'represents a convergence of ethical thinking'<sup>27</sup>.

The Ethics Guidelines for Trustworthy AI puts forward seven Requirements of Trustworthy AI. These requirements allow us to begin to understand the risks that AI could pose to learners. It should be noted that the risks outlined below are not intended to be an exhaustive list of ethical risks posed to learners by Artificial Intelligence that is used unethically, and that further steps will be taken to understand how likely it is that these risks will lead to learners experiencing harm.

## Requirements of ethical AI and possible risks posed to learners

### **Requirement: Human Agency and Oversight**

**Description:** “ AI systems should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy... and allow for human oversight.”

#### **Potential harms to learners**

- Learners’ agency could be decreased if AI systems reduce independent and introspective thought, and lead to the underdevelopment of higher-order thinking skills and self-regulation.
- Learners could become over reliant on AI systems, hence diminishing their overall autonomy.
- Education could become depersonalised and less effective if human interaction were undermined.

### **Requirement: Accountability**

**Description:** This requirement “necessitates that mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use.”

#### **Potential harms to learners**

- AIEd could erode accountability structures in schools, colleges, universities and other educational institutions; and cause difficulty with deciding who is accountable for poor performance, or inappropriate or harmful outcomes.

### **Requirement: Technical Robustness and Safety**

**Description:** Technical robustness and safety consists in AI systems being secure and resilient to attack, having a fallback plan in case of problems, being sufficiently accurate, and able to yield reliable and reproducible results.

#### **Potential harms to learners**

- Highly sensitive data relating to individuals’ learning processes (including data on emotions) could be hacked by malicious parties.
- AI systems could make invalid recommendations - e.g. on which university course to study, which career to pursue or on how to address a mental health condition - that could harm learners.
- Pastoral AI systems could fail to identify urgent safeguarding needs- such as risks of self harm- thereby missing opportunities to protect vulnerable people.

**Requirement: Diversity, non-discrimination and fairness**

**Description:** AI systems should treat all individuals fairly, and in particular should not be used in ways that discriminate against different groups of people.

**Potential harms to learners**

- Due to algorithmic bias, some groups of learners could be unfairly discriminated against. For example, a predictive analytics tool that predicts university students who are most at risk of dropping out of university may overstate the likelihood of pupils from lower socio-economic backgrounds dropping out and therefore lead to discriminatory actions.
- An AI system could be more effective for male students due to being trained on datasets that included more males than females.

**Requirement: Privacy and data governance**

**Description:** Privacy is “a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy.”

**Potential harms to learners**

- Highly intimate data - including a learners’ strengths, vulnerabilities, behavioural habits, and biometric information - could be inappropriately used to exploit, manipulate or oppress learners.
- Poor governance of learner data could lead to ineffective outcomes.
- Learners could lose control over who had access to their academic performance data, which could lead to employers or educational institutions having access to a disproportionately high level of information on applicants against their will.

**Requirement: Transparency**

**Description:** Transparency requires that “the data sets and the processes that yield the AI system’s decision, including those of data gathering and data labelling as well as the algorithms used, should be documented to the best possible standard to allow for traceability and an increase in transparency” (traceability); “that the decisions made by an AI system can be understood and traced by human beings” (explainability); and that “AI systems should not represent themselves as humans to users”.

**Potential harms to learners**

- A learner’s essay could have been marked by an AI system and the learner and relevant educators may not be able to understand the verdict. This could lead to unfair outcomes or missed opportunities for further development based on feedback.
- A learner could be denied a place on a particular university course, and have no means of seeking redress or an explanation due to the inherent opacity of the system.

**Requirement: Societal and environmental wellbeing**

**Description:** “ The broader society, other sentient beings and the environment should be also considered as stakeholders throughout the AI system’s life cycle. Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, such as for instance the Sustainable Development Goals.”

**Potential harms to learners**

- AIEd could weaken humans’ abilities to interact with each other, due to a reduction - or lack of emphasis on - interpersonal skills.
- AI could lead to greater educational inequalities in society, thereby creating further divisions, if highly effective AIEd were more accessible to some groups than others.

**Amplifying the risks of less advanced technologies**

Evidence of the impacts of technologies that are currently in widespread use by children and other learners also allows us to anticipate possible risks of AI in Education. Persuasive design techniques - including infinite scrolls, aggressive summons, social obligation mechanisms, and having no means of saving one’s progress - employed in technologies such as social media, have been found to have a negative impact on users’ mental health and ability to form strong relationships<sup>30</sup>. There is a risk that persuasive design techniques could be strengthened by AI systems that can predict, with high levels of accuracy, which persuasion techniques will be most effective.

# Perspectives from Educators

“AI Summative Judgements. At present educational platforms are using AI to make formative judgements on individual students (based on what is likely to be a globally unrepresentative dataset) in order to tailor learning to the needs of the individual student. There is a risk down the line that we may see AI educational systems replace the present public examination system, making summative judgments that have a significant impact on life choices such as access to higher education or certain levels of employment. Furthermore, judgments might be made on wider criteria than presently deployed. For example, rather than just assessing whether or not a student has mastered a particular skill or unit of knowledge, it would be possible to assess how fast a student takes to so.”

- Mark Steed, Principal and CEO at Kellett School, Hong Kong

“AI will extend educational opportunities beyond anything we have yet imagined.

However, an inspirational learning relationship between teacher and student always has - and always will- characterise the very best education. AI can assist a gifted teacher; it will never replace one. Humans are an imitative species, never more so than in their most formative student years. The world over, people always have - and always will - remember a great teacher.”

- Vivienne Durham, Chief Executive of the Girls' Schools Association

“Providers of informal learning are likely to adopt and market education “powered by AI”. Unencumbered by formal regulation, informal providers will be less averse to the ethical risks inherent in cutting edge AI. In response to such competition, providers in regulated education should urgently define and promote to learners the benefits of ethical AI approaches”

- Peter Horrocks CBE, Broadcast Executive and Former Vice Chancellor of The Open University

## Grounds for optimism

If artificial intelligence is used effectively, responsibly and with learners' best interests at heart then real progress could be made towards achieving a more effective, well-rounded and equitable education for all learners. But if we get this wrong, societies could become more divided, education could be used as a weapon of mass manipulation, and human flourishing could suffer profoundly.

At The Institute for Ethical AI in Education we conclude that there are grounds for optimism. Why? Because societies still have the power to shape and steer the development of ethical AIEd. To achieve this positive outcome, those designing, developing and deploying AIEd need to be equipped with the tools to allow them to act responsibly and make ethical decisions; learners, parents and those who represent their interests must be empowered to understand the ethical issues associated with AI in Education and know when ethical practice is being achieved; societies as a whole need to appreciate the importance of ethical AIEd and demand nothing less than the best deal possible for learners and for all members of society; and industry could benefit immensely from continuing to develop an understanding of the ethical risks presented by AI in Education so that they are supported to innovate safely - thus realising the benefits of AI for learners without exposing learners to potential harms.

In the next section we put forward our approach to developing an ethical framework for AI in Education, the ultimate purpose of which will be to allow learners to experience optimal benefits from AI in Education, and to protect learners and society as a whole from the credible risks that AI poses.

### 3. Developing an Ethical Framework for AI in Education

“That AI will have a major impact on society is no longer in question. Current debate turns instead on how far this impact will be positive or negative, for whom, in which ways, in which places, and on what timescale.”

AI4People—An Ethical Framework for a Good AI Society

AI4People’s Ethical Framework for a Good AI Society<sup>31</sup> provides a model for developing an ethical framework that aims to simultaneously facilitate the benefits of AIEd and safeguard learners against its risks. It argues that AI can be used to support human flourishing, but that AI can also be underused - thus creating opportunity costs- or cause harm via misuse or overuse. Ethical approaches to AI can do more than just safeguard users against the harms that unethical AI could cause. They confer the ‘dual advantage’ of allowing societies to understand and realise the beneficial opportunities posed by artificial intelligence, whilst also guarding against applications of artificial intelligence which could cause harm at an individual or societal level.

This approach is in line with the point made in *The Ethics Guidelines for Trustworthy AI*, that AI ethics can ‘stimulate reflection’ on how to protect individuals and groups from the potential threats of artificial intelligence, and that it can stimulate innovative uses of AI, which may foster and promote societal values.

#### Towards an effective ethical framework

Given the gulf between the best case scenario for learners from AIEd and the worse, there is a moral responsibility to develop frameworks that genuinely have a positive impact on how AI is developed, deployed and used. For learners’ sakes, one cannot afford to establish frameworks that seem credible in theory but fall apart in practice. In this section, some common pitfalls of ethical frameworks are explored along with measures to counteract these problems.

#### Proactively addressing ‘ethical risks’

Floridi et al outline a set of intentional behaviours- “unethical risks”<sup>32</sup> - that could undermine ethical practice. Two key examples of such unethical risks are those of ethics shopping and ethics bluewashing.

The problem of ethics shopping is in part due to the proliferation of ethical frameworks, and consists in organisations cherry-picking aspects of different frameworks where principles or ethical requirements are least demanding for the organisation to fulfil. An emergent consequence is that collectively, ethical frameworks merely have the effect of legitimising current practice, rather than steering improved, ethical decision-making.

The solution put forward to this problem lies in establishing “clear, shared, and publicly accepted ethical standards.” If stakeholder groups and the public at large have a clear and robust expectation of the ethical principles that should be abided by, it will be significantly more difficult for these principles to be sidestepped.

Ethics bluewashing is another unethical risk and refers to the practice of systematically making claims that overstate an organisation’s ethical credentials, and purposefully making an organisation appear more ethical than it actually is. The principles embedded in a poorly designed ethical framework could therefore remain uninstantiated if organisations were free to - and actively chose to - misguide audiences on how these principles were being put into action.

Transparency and education are put forward as a two-pronged defence against ethics bluewashing. It is essential that claims about ethical practice are open to independent verification (and where possible public scrutiny) to ensure organisations can be held to account for unethical conduct. This, however, cannot be achieved in full unless stakeholders and the wider public are educated so that they can understand what ethical practice looks like, and have the awareness, information and tools to identify where actions and decisions are falling foul of what is expected.

The following conclusions can be drawn from examining these two unethical risks. Firstly, to ensure that there is a strong expectation that the spirit of an ethical framework should be carried out in full, stakeholders affected by AI in Education must accept and give assent to the principles underlying these frameworks. Next, it should be possible to verify whether or not the framework’s principles are genuinely being put into practice. This will require mechanisms for evidencing and interrogating whether the systems’ behaviours, and the decisions of those responsible for how AI is developed or deployed, are in line with the framework’s principles. This would allow for transparency around organisations’ or individuals’ adherence to the framework, and enable decision makers to be held to account.

The third conclusion is that educating key stakeholders - learners, parents, educators, policymakers, software engineers, and more - about the ethical implications of AI in

Education will be essential for ensuring any frameworks can be properly implemented.

Stakeholder education will be needed not only to implement such frameworks once they have been created, but also to develop them in the first instance. In order for stakeholders to develop a shared expectation of how AI should be ethically used in education they will need to understand both the technology and the domain in which it is applied. Not only will learners, parents, teachers, lecturers and education leaders need to sufficiently understand artificial intelligence; experts in AI will need to understand the process of learning and the goals of education in order to develop technologies that will truly benefit learners.

### **Addressing inherent tensions between principles**

In addition to problems caused by parties willingly frustrating the process of putting ethical principles into practice, a further set of problems arise from the inherent properties of ethical frameworks themselves.

In the report, *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*<sup>33</sup>, published by the Nuffield Foundation and The Leverhulme Centre for The Future of Intelligence, three key problems inherent in many frameworks for ethical AI are identified.

- ‘A lack of clarity or consensus around the meaning of central ethical concepts and how they apply in specific situations.’
- ‘Insufficient attention given to tensions between ideals and values.’
- ‘Insufficient evidence on both (a) key technological capabilities and impacts, and (b) the perspectives of different publics.’

Regarding the first point, an ethical concept for AI in Education might be that AI should enhance learners’ knowledge, understanding and skills. However, it is difficult to see how such a principle on its own could guide decision-making in concrete situations. What knowledge, understanding and skills should learners be developing? What applications of AI are best placed to develop these? What applications of AI are likely to interfere with their development?

An example of a tension for point 2 is that of equity and support for the vulnerable vs non-discrimination and fairness. For instance, an AI system used to support pastoral care could systematically identify people of a particular demographic as being more

vulnerable than those from other demographics. However, measures to reduce this bias through reconfigurations of the system could lead to people from these demographics being disproportionately overlooked.

As well as there being tensions between achieving benefits and avoiding risks, there may also be cases where it is unclear whether benefits have been conferred at all. It is possible that AI will lead to a shift away from grades to more holistic representations of learners' skills, competencies and attributes. Before any such shift is made, however, we should try to thoroughly understand who will benefit and what risks are posed.

A key part of the solution put forward for these problems is, again, to involve those whose lives will be affected by AI in Education in the process of defining central ethical concepts, and determining how best to resolve tensions between principles. It will not be sufficient to impose a framework on stakeholders and expect it to gain their ascent and structure their expectations of ethical AI. Stakeholders' views need to be taken into account to determine what net benefits consist in; where the lines are between underuse, effective use and overuse; and what misuse entails and how it can be mitigated. As stated by The Centre for Data Ethics and Innovation, "the public's views [should] form the governance of data driven technology". What's more, the public's views need to be informed by evidence of the technological capabilities and impacts of artificial intelligence used in education.

## Criteria for the framework for ethical AI in Education

To effectively enable learners to experience optimal benefits from AI in Education, the framework for ethical AI in Education will need to:

- Recognise that there are both benefits and risks of AI in Education, and provide clear guidance for ethical decision-making that will result in learners experiencing optimal benefits from AI in Education. This includes providing a means by which tensions and trade-offs can be resolved.
- Reflect the views and values of stakeholders and hence represent a shared vision of ethical AI in Education. These views and values should be informed, as far as possible, by evidence of the 'technological capabilities and impacts' of AI in Education.
- Be designed such that it can be verified whether decisions and actions are compliant with the framework.

In order to develop the framework and ensure its effective application, stakeholders will need to be educated so that they can make their own informed judgements of what ethical AI in education consists in, and can determine whether it is being achieved in practice.

Additionally, to ensure learners get the best possible deal from AI in education, ethical frameworks themselves may not be sufficient. Frameworks may guide ethical decision-making, but in some cases it may be necessary to compel ethical action, and stimulate ethical innovation. Further mechanisms may hence be needed to guarantee ethical practice. Appropriate mechanisms may include regulation, standards, laws, and awareness campaigns. To be effective, such mechanisms must embody the principles of the framework. In turn, the frameworks themselves should be developed so that formal mechanisms can be derived from them.

## A blueprint

The following blueprint is intended to form the basis of an ethical framework that fulfils the criteria laid out above.

- AI should **only** be used for educational purposes where there are clear indications that it will genuinely benefit learners either at an individual or collective level.
- AI should **not** be used for educational purposes in cases where the risks posed to learners are at all significant.
- To support the achievement of an optimal balance between underuse, and overuse/misuse of AI in Education, AI may only be used in cases where there is any risk of harm to learners - at an individual or collective level - if the following conditions are met:
  - A) Decisive steps have been taken - and openly communicated in an understandable way - to mitigate the risks involved;
  - B) The benefits to learners outweigh these risks;
  - C) The risks to learners are not significant;
  - D) Monitoring of conditions (A)-(C) is in place;
  - E) If monitoring in (D) shows that conditions (A)-(C) are not fulfilled, the use of AI is to be improved or terminated.

Building the framework around this blueprint is intended to make it applicable-by-design, with the key objective of bringing about net benefits from AI for learners.

The first recommendation intends to orientate the use of AI in Education towards benefits for learners. Here we intend to set the bar high. Without this point, a framework may still be able to safeguard learners from significant harm, but risks missing the opportunity to enable and stimulate beneficial use.

The second and third recommendations acknowledge that achieving zero risk to learners may not be possible. However, these points are intended to collectively safeguard learners from risks that are at all significant, whilst also providing a set of criteria for addressing how benefits can still be achieved in cases where zero risk cannot be guaranteed. The third recommendation specifically provides an overarching mechanism by which trade-offs and tensions can be resolved within specific contexts. It also stipulates that there must be active monitoring of whether the criteria laid out in the blueprint have been met, which lays the foundation for processes that enable verifiability and accountability.

Together, these recommendations are intended to drive ethical decision-making in a way that is practical and will result in optimal advantage being taken of the opportunities presented by AI in education whilst minimising and mitigating the risks. They also aim to provide clear red-lines so that learners can be protected against levels of risk that would be considered intolerable regardless of what benefits might be achieved as a result.

These recommendations merely provide a foundation for an ethical framework. We have yet to define what is meant by benefit to learners or provide a means of judging whether such benefits have been realised. We are not yet in a position to decide whether a particular benefit outweighs a particular risk, and we cannot currently draw lines in the sand between tolerable and intolerable levels of risk. And although the blueprint acknowledges that both risks and benefits can be posed at an individual or collective level, deliberation will be needed to decide, for instance, where benefits at a collective level might outweigh risks at an individual level, and vice-versa. In operational terms, it will need to be established how monitoring can take place and enable systematic processes for verification and accountability. To address these points, the conversation has to be opened up to learners, educators, policymakers, software developers, academics, professional bodies, advocacy groups, and industry. The Institute for Ethical AI in Education will not impose a singular vision of ethical AI upon societies. Only a shared vision of ethical AI in Education will have sufficient gravitas to enable learners to experience optimal benefits from AI in Education whilst being safeguarded against the risks posed.

## 4. Realising a Shared Vision of Ethical AI in Education

### Mechanisms for realising ethical AI in education

Developing a shared vision for ethical AI in education entails both establishing consensus on what it means to design and deploy AI ethically, and converging upon an agreement on how best to realise ethical practice. It means answering both “what does ethical AIEd look like?” and “how can it be achieved?” As such, a critical step is to allow stakeholders and experts to review and evaluate possible mechanisms for realising ethical AI in education in order to establish which mechanisms are likely to be effective and how existing instruments may need to be amended for the context of AI in education, and to ascertain where tailor-made processes will need to be developed.

The options available range from doing nothing and hoping for the best, to producing global regulatory frameworks that tightly control how AI can be designed, developed and deployed for educational purposes. Below is a list of possible options.

The following mechanisms were suggested in The Ethics Guidelines for Trustworthy AI and could be adapted for the context of ethical AI in Education.

#### **Mechanism: Regulation**

---

**In the context of Education:** Regulation could be “revised, adapted or introduced” so that organisations designing, developing or deploying AI in Education would be compelled to act and make decisions ethically.

---

#### **Mechanism: Codes of Conduct**

---

**In the context of Education:** Organisations designing and developing AIEd systems - including academic research groups, commercial companies and education institutes - could voluntarily, or be required to, sign up to the ethical framework for AI in education, and adapt their governance practices to ensure ethical practice.

---

#### **Mechanisms: Certification**

---

**In the context of Education:** Organisations, possibly acting on behalf of central education authorities, could be tasked with certifying that AIEd systems conformed to certain standards, such as those relating to transparency and fairness.

---

### **Mechanism: Standardisation**

---

**In the context of Education:** Standards (in line with those developed by The Institute of Electrical and Electronics Engineers (IEEE) and The British Standards Institute (BSI)) could be developed for AIEd resources. These standards could focus on factors such as the transparency of AIEd systems; levels of algorithmic bias; or the means by which data is collected, processed and used. Standards may be required specifically for the context of AIEd, or broader standards may be deemed to be readily applicable in the context of education. Such standards might include the IEEE's P7000 series, which includes standards relating to 'Transparency for Autonomous Systems', 'Data privacy process', 'Algorithmic Bias Considerations', 'Child and Student Data Governance', and 'Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems'.

---

### **Mechanism: Education and Awareness for Stakeholders**

---

**In the context of Education:** Organisations could provide learners, parents, educators, administrators, etc. with educational support to understand AI and its ethical implications in education. This could organically foster expectations of ethical AIEd and ensure that key stakeholders are in a position to distinguish between ethical and unethical products and practices.

---

### **Mechanism: Diversity and Inclusive Teams**

---

**In the context of Education:** Organisations developing and deploying AI for education could commit to charters which facilitate the ongoing development of diverse and inclusive teams. The Tech Talent Charter may provide an effective platform for this to be achieved. Teams developing AIEd products and services could be encouraged to sign up to the Charter.

---

### **Mechanism: Architectures for Trustworthy AI**

---

**In the context of Education:** At the design and development stages, certain states and behaviours could be either prohibited (via blacklists) or compelled (via whitelists), with the intention of limiting a system to procedures that are considered ethical. Such measures could directly affect what types of data on learners is collected and how it is used. These measures might also be effective at guarding against persuasive design techniques and other methods of manipulation.

---

---

**Mechanism: Quality of Service Indicators**

---

**In the context of Education:** These could include indicators that provide a metric of how secure learner data is, how transparent a system is, or how accurate a system is at achieving a particular educational outcome. These instruments could be particularly effective at supporting organisations to make informed decisions about how to balance trade-offs between competing factors, given that decision-making would be supported by quantifiable metrics.

---

---

**Mechanism: Explanation Methods**

---

**In the context of Education:** AI systems could be compelled to be designed in such a way that significant decisions could always be investigated so that people understand why the decisions were made. In the context of education, it may be expected that significant decisions - such as the grading of formal assessments, selection or rejection of students by an institution, or processes that highlighted students for strategic interventions - are sufficiently explainable. This in turn may provide a means by which learners can query decisions and seek redress for unfair actions taken.

---

## Further mechanisms

---

**Mechanism: Procurement procedures**

---

**Explanation:** Educational Institutions and government bodies could be provided with procurement guidelines to assist them in assessing whether particular AI tools were ethical. Similarly, tools could only be allowed to be part of procurement platforms if they had been expertly verified as ethical.

---

---

**Mechanism: Training about AI and about the use of AI for educators**

---

**Explanation:** The way in which AI is deployed in education has significant ethical implications. Educators could hence be trained to use AI so that it can have a beneficial impact, and risks can be mitigated. This training could be part of Initial Teacher Training (ITT) programmes and/or as part of continuous professional development.

---

---

**Mechanism: Safeguarding approach**

---

**Explanation:** This would involve including protocols for Ethical AIEd within statutory safeguarding duties. This would mean that schools, colleges, universities and other educational providers have a responsibility for ensuring learners are safe when using AIEd. For instance, statutory requirements could include systems to prevent adverse mental health impacts of AIEd.

---

---

**Mechanism: Workers' rights approach**

---

**Explanation:** Particularly relevant in the context of corporate learning and development, stipulations around how professional learners' data can be used to monitor and manage learners could be included in regulations around workers' rights.

---

---

**Mechanism: Kite marks**

---

**Explanation:** Particularly relevant in the context of informal learning and business-to-consumer education markets, kite marks could be used to inform learners about which providers have developed ethical AIEd products and services. This would then help inform purchasing decisions, which could lead to unethical products and untrustworthy organisations failing to gain a sufficient share of the market.

---

---

**Mechanism: Monitoring and reporting on macroscopic impacts of AIEd**

---

**Explanation:** Organisations could be tasked with monitoring the macroscopic impact of AIEd and making policy recommendations to government. For instance, The Social Mobility Commission could monitor the impacts of AIEd on life chances and inclusivity, and make recommendations on how AIEd could further support social mobility.

---

---

**Mechanism: Ensure the ethics of AI in Education is part of the curriculum**

---

**Explanation:** As stated in the report *Ready, Willing and Able?* from The House of Lords' Select Committee on Artificial Intelligence, "All citizens have the right to be educated to enable them to flourish mentally, emotionally and economically alongside artificial intelligence." A logical starting point to achieve this would be to ensure that throughout formal education, learners are taught about the ways in which they may be using AI and the ethical implications of doing so. This would include learning about the ethics of AIEd. By empowering learners directly with this knowledge and understanding, learners may be in a better position to judge the appropriateness of using AI in Education, and could make decisions on when not to use AI.

---

## Data Governance

Robust frameworks for the governance of learners' data will be essential for realising ethical practice wherever AI is used in education. Existing legal frameworks already serve to regulate how learners' data can be collected, stored and processed; although it should be considered that additional ethical frameworks/guidelines may be needed to ensure that use of learners' data is not only legal, but ethical too.

## Towards a tentative solution for ethical AI in education

The tentative solution below will provide a launch pad for a programme of research, consultation and deliberation. During this process, The Institute expects these positions to be questioned, interrogated, amended, extended, clarified and perhaps changed beyond recognition.

### Achieving benefits

Providing risks have been prevented and/or sufficiently mitigated, the use of AI in education should be considered beneficial in cases where:

- AI facilitates shifts towards a more holistic development of learners without compromising the academic skills that are the current focus of education systems; and where it allows for a broader range of socially, economically and personally useful skills to be developed and recognised.
- The progress of learners can be enhanced through improved insights into learners' needs, and through increased levels of individualised support.
- Access to educational opportunities is limited, and where such limitations could not be easily remedied by modest interventions or reforms. For example, providing access to personal tuition for all could meet this criteria.

### Drawing red Lines

- Data that is collected to enhance learning and provide educational support may be used to support related social services where vulnerable individuals are the primary concern, and which could therefore be seen as an extension of safeguarding duties in education. For instance, information on learners could be used by social and health services. **However**, data collected to enhance learning and provide educational support should not be used for commercial purposes (including to recommend upgrades of already used platforms), or for government monitoring purposes (including policing or welfare purposes).
- In cases where AIEd tools exhibit biases towards different groups of individuals and create/widen gaps in progress and performance, the tool should cease to be used or a clear strategy should be put in place to reverse the emergent effects of this bias. To achieve this, processes for auditing the outcomes of AIEd tools will need to be in place- which will require consideration of how to isolate the impacts of the tools themselves - and there will need to be effective oversight of

how monitoring and intervention strategies are designed and implemented.

- Whenever an individual interacts directly with an AIEd resource, they ought to be clearly informed that the resource is using AI. As part of this principle, where AIEd resources present themselves as a human character, there should be clear measures in place to remind users that they are not interacting with a natural person, and AIEd systems should never take the form of actual living people.

### **Securing ethical AIEd**

- All learners in formal education should, from an appropriate age, learn about AI and its implications, and there should be a particular focus on the areas in which they use AI - which may include AIEd.
- Wherever AI systems are used to make significant decisions about a learner's development or educational opportunities, the decision should be explainable and systems should exist to allow learners (and where appropriate, their parents) to have access to all available information related to the decision.
- A process should be established whereby public institutions are advised to only procure AIEd solutions that have been certified as ethical and beneficial.

## Developing a shared vision of ethical AI in education

During the programme of research, consultation and deliberation, the following overarching questions will be asked - with the tentative solution providing a reference point.

- What genuine benefits can be achieved through AI in Education and how can stakeholders be confident that those benefits will be achieved through any particular use of AI?
- What are different stakeholders' perspectives on the risks posed by AIEd, on how these risks should be mitigated, and on how individuals can be confident said risks have been mitigated?
- Which risks to learners are considered to be intolerable, and how can learners be safeguarded against these risks and be confident that appropriate safeguards are in place?
- How can it be decided whether the benefits of AIEd outweigh its risks in any specific case and what processes are needed to allow such trade-offs to be decided in practice in a variety of contexts?
- By what means can it be verified whether decisions relevant to the development or deployment of AI in Education are compliant with the ethical framework, and what evidence is required to verify whether a particular AIEd resource is compliant?
- What practical methods and mechanisms can be employed to support learners and promote ethical practice?

These questions stem from The Institute's Blueprint for the Ethical Framework for AI in Education, and aim towards developing a fully-fledged ethical framework, along with recommendations on practical mechanisms to ensure that learners are supported as directed by the framework.

To answer these questions, the Institute will listen to and take on board the perspectives, views and values of learners (of all ages), educators (from all sectors), advocacy groups, education leaders, government and policymakers, policy and academic researchers, academics, software developers, ethics specialists and other relevant stakeholders and experts. We will also listen carefully to the considerations of industry and technology experts to ensure that the frameworks for ethical AI in education are both ambitious and realistic.

The Institute will work towards ensuring all stakeholders' views are informed by relevant evidence and information, which will include insights into ethical practice and governance from other sectors, such as healthcare.

To ensure the voices of these groups of stakeholders are heard and listened to, the Institute will be delivering a programme of engagement, consultation and deliberation. This programme will include:

- A call for evidence that will give a wide range of stakeholders the opportunity to express their understanding and views on the benefits and risks of AIEd, and how best to realise ethical practice.
- A series of interviews with domain experts.
- A series of focus groups with stakeholders from different domains.
- Roundtable events, which will bring together a cross-section of stakeholders and provide opportunities for deliberation on key issues.
- Meetings of an International Advisory Group that will ensure global perspectives are drawn upon in developing the framework for ethical AI in education.
- An International Conference that will bring stakeholders and experts from across the globe together to converge upon a global vision for ethical AI in education.

If you and/or your organisation would like to participate in this programme of engagement, consultation and deliberation - or would otherwise like to get in touch - please do contact The Institute via our website at <https://www.buckingham.ac.uk/research/the-institute-for-ethical-ai-in-education/>

## References.

- 1- Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development (2019). UNESCO. [https://backend.educ.ar/refactor\\_resource/getBook/1097](https://backend.educ.ar/refactor_resource/getBook/1097)
- 2- Anissa, N. ,Baker, T. , Smith, L. , (2019). Educ-AI-tion Rebooted: Exploring the future of artificial intelligence in schools and colleges. NESTA.
- 3- Luckin, R., Holmes, W., Griffiths, M. & Forcier, L. B. (2016). Intelligence Unleashed. An argument for AI in Education. London: Pearson.
- 4- United Nations 2020, Quality Education, Viewed January 2020 <<https://www.un.org/sustainabledevelopment/education/>>
- 5- Avinash, A., Bonetti, S., Crenna-Jennings, W., Hutchinson, J. (2019). Education in England: 2019 Annual Report. Education Policy Institute
- 6- Richardson S 2019, Four out of 10 teachers plan to quit, survey suggests, BBC News, Viewed January 2020 <<https://www.bbc.co.uk/news/education-47936211>>
- 7 - UNESCO Institute for Statistics 2016, The World Needs Almost 69 Million New Teachers To Reach The 2030 Education Goals, Viewed January 2020 <<https://unesdoc.unesco.org/ark:/48223/pf0000246124>>
- 8- National Education Union 2018, Teacher Recruitment and Retention, Viewed January 2020 <<https://neu.org.uk/policy/teacher-recruitment-and-retention>>
- 9- The Future of Work: A Vision for The National Retraining Scheme- Part III: Building a Workforce for The Future (2019), The Centre for Social Justice, <https://www.centreforsocialjustice.org.uk/core/wp-content/uploads/2019/02/CSJJ6899-Future-Of-Work-WEB-190215.pdf>
- 10 - Aftab Hussain 2018, Digital Assistants for Teachers, [aftabhussain.com](http://www.aftabhussain.com/digital_assistant_for_teachers.html), Viewed January 2020, <[http://www.aftabhussain.com/digital\\_assistant\\_for\\_teachers.html](http://www.aftabhussain.com/digital_assistant_for_teachers.html)>
- 11- Dickli. S (2016), Automated Essay Scoring, Turkish Online Journal of Distance Education, Viewed January 2020 <<https://pdfs.semanticscholar.org/cc05/3772ceac5d3803917a10ec48cdd6eca193a5.pdf>>
- 12- Black. P, Wiliam. D (1998), Inside The Black Box: Raising Standards Through Classroom Assessment, Phi Delta Kappa International
- 13 - Teaching and Learning Toolkit, Feedback, Education Endowment Foundation, Viewed January 2020 <<https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/>>
- 14 - <https://visible-learning.org/>
- 15- Luckin. R (2017), Towards Artificial Intelligence-based assessment systems, Nature Human Behaviour
- 16 - System Upgrade: Realising the Vision for UK Education (2012), Technology Enhanced Learning, Viewed January 2020 <https://discovery.ucl.ac.uk/id/eprint/1475950/1/System%20Upgrade%20Final.pdf>
- 17- Ekowo. M, Palmer. I, (2016), The Promise and Peril of Predictive Analytics in Higher Education, New America
- 18- Rowland Manthorpe 2019, Artificial intelligence being used in schools to detect self-harm and bullying, Sky News, Viewed January 2020 <<https://news.sky.com/story/artificial-intelligence-being-used-in-schools-to-detect-self-harm-and-bullying-11815865>>
- 19- Ofsted. Methodology note: the risk assessment process for good and outstanding maintained schools and academies. 2018. <[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/737583/Methodology\\_note\\_risk\\_assessment\\_of\\_good\\_and\\_outstanding\\_maintained\\_schools\\_and\\_academies\\_030918.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/737583/Methodology_note_risk_assessment_of_good_and_outstanding_maintained_schools_and_academies_030918.pdf)> Accessed January, 2020

- 20- Teaching and Learning Toolkit, One to One Tuition, Education Endowment Foundation, Viewed January 2020 <<https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/one-to-one-tuition/>>
- 21- Teaching and Learning Toolkit, Individualised Instruction, Education Endowment Foundation, Viewed January 2020 <<https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/>>
- 22- Asilomar AI Principles, Future of Life Institute (2017), <https://futureoflife.org/aiprinciples/>.
- 23- The Responsible Machine Learning Principles, The Institute for Ethical AI and Machine Learning <https://ethical.institute/principles.html>
- 24- Montreal Declaration for a Responsible Development of Artificial Intelligence. (2017). Announced at the conclusion of the Forum on the Socially Responsible Development of AI. (<https://nouvelles.umontreal.ca/en/article/2018/12/04/developing-ai-in-a-responsible-way/>)
- 25- House of Lords Artificial Intelligence Committee. (2018). AI in the UK: ready, willing and able? (<https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>)
- 26- Ethics Guidelines for Trustworthy AI (European Commission, 2019); <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- 27 - Clement-Jones, T., Floridi, F. (2019) "The five principles key to any ethical framework for AI", New Statesman <https://tech.newstatesman.com/policy/ai-ethics-framework>
- 28 - UN General Assembly, Convention on the Rights of the Child, 20 November 1989, United Nations, Treaty Series, vol. 1577, p. 3, available at: <https://www.refworld.org/docid/3ae6b38f0.html>
- 29- Charlotte Santry 2018, Artificial intelligence 'risks school apartheid', TES, viewed January 2020, <<https://www.tes.com/news/artificial-intelligence-risks-school-apartheid>>
- 30- Baroness, Beeban & Evans, Alexandra & Afia, Jenny & Bowden-Jones, Henrietta & Hackett, Liam & Juj, Anisha & Przybylski, Andrew & Rudkin, Angharad & Group, Young. (2018). Disrupted Childhood: The cost of Persuasive design.
- 31 - Floridi, L., Cowls, J., Beltrametti, M. et al. Minds & Machines (2018) 28: 689. <https://doi.org/10.1007/s11023-018-9482-5>
- 32 - Floridi, L. Philos. Technol. (2019) 32: 185. <https://doi.org/10.1007/s13347-019-00354-x>
- 33 - Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K. and Cave, S. (2019) Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research. <http://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf>