

# The Institute for Ethical AI in Education

## Annex: Developing The Ethical Framework for AI in Education



THE UNIVERSITY OF  
BUCKINGHAM



# Contents

2 - Section 1

4 - Section 2

5 - Section 3

6 - Section 4

8 - Section 5

9 - Section 6

11 - Section 7

12 - Section 8

13 - Section 9

14 - References

# Developing The Ethical Framework for AI in Education

## Annex to The Ethical Framework for AI in Education

The Ethical Framework for AI in Education brings together the ideas, insights and lessons learned from the Global Summit on the Ethics of AI in Education and from the series of roundtables that the Institute convened. Here follows an overview of the discussions that took place. To demonstrate how The Ethical Framework for AI in Education was derived, the discussion points have been categorised based around the objectives presented in the Framework.

While there are several definitions of AI<sup>1</sup>, the Institute has adopted a broad definition in order to engage with the ethical implications of both rules-based and machine learning-based systems. For the purposes of The Ethical Framework for AI in Education, please understand AI to mean “the use of digital technology to create systems capable of performing tasks commonly thought to require intelligence” (UK Government definition of AI<sup>2</sup>).

### **1. AI should be used to achieve well-defined educational goals based on strong societal, educational or scientific evidence that this is for the benefit of learner**

This principle is grounded in insights gained from discussions around the use of AI for teaching and learning. The Institute has extrapolated from the principles that determine how AI should be used ethically for teaching and learning, to how AI should be used ethically in broader contexts.

Throughout the roundtables and during the Global Summit a strong case was made for using AI to improve teaching and learning. In particular, AI’s potential to personalise learning and to enhance formative assessment was raised on multiple occasions.

These potential benefits of artificial intelligence for teaching and learning have been widely recognised. By analysing data on learning, AI resources are able to identify learners’ needs and therefore tailor learning experiences in order to meet those needs. The data that AI systems capture, analyse and evaluate can also be converted into granular insights into learners’ strengths and needs. These insights can then be fed back to educators or to learners themselves, allowing for further tailoring of teaching and learning.

Roundtable and Global Summit participants did raise reservations around the use of AI for certain aspects of teaching and learning, however. For instance, whilst the tailoring of learning to the needs of individual learners was generally viewed as a key benefit of AI in education, one

participant argued that there are risks to the “hyper-individualisation of learning” as this could undermine the peer-to-peer aspects of education. Another noted that approaches centred on individual learners represents an educational philosophy that is not universally agreed upon; in some parts of the world, for instance, “communitarian” learning is seen as a higher priority. Other participants noted that whilst the tailoring of learning to the needs of individuals could be an effective methodology, AI might not always be effective at achieving this, and that stakeholders should be mindful of whether sufficient data is available to enable effective tailoring in any particular case. This was not, however, the majority view.

Two repeated concerns around the use of AI for teaching and learning were that AI could be misused/overused, and that AI could marginalise/undermine teachers and, moreover, student-teacher relationships. Particularly noteworthy points were that AI should not be used where it does not support a particular aspect of learning, and that stakeholders should resist the notion that AI will inevitably be the solution to their problems. These latter points are reflected in the Framework.

During the Young People’s Roundtables, participants grappled with the question- “how should AI be used to make learning more engaging, supporting and challenging; and how should AI not be used?” Whilst participants at the roundtables for professionals highlighted the tensions between more independent learning (provided, for instance, by intelligent tutoring systems) and the more sociable, collaborative aspects of learning, the participants of the Young People’s Roundtables quickly arrived at a means of striking a balance. There was broad consensus that it would be beneficial to use AI not to change the core structure of lessons but to supplement it. There was clear demand for AI to be used to enable students who had mastered a particular aspect of learning to move onto more challenging activities. Students recognised, however, that this could lead to a situation where some learners were left behind. They suggested that learners who were not accessing the more challenging activities via AI should receive the greatest proportion of the extra time that AI had freed up for teachers.

The use of AI to help teachers understand learners’ needs also gained strong approval. Students were very clear that the student teacher relationship was highly valued, and should not be diminished by the use of AI. Teachers, they argued, were inspirations, role models, and professionals uniquely placed to develop students in a holistic and inherently human way.

Referring back to the roundtables for professionals, participants’ insights presented solutions to how AI can be used ethically within the context of teaching and learning. Numerous participants urged that learners’ needs and/or established pedagogies should guide how AI is used, and one participant stressed that “It all comes down to whether AI is actually improving the teaching and learning of young people”. With personalised learning, for instance, the use of AI has the potential to improve teaching and learning because it is grounded in effective teaching methods<sup>3</sup> that are tailored to the needs of individual learners. Likewise, due to the levels of data it can capture and process, AI can also enhance teaching and learning by building upon the successes of assessment for learning<sup>4</sup>.

The primary criterion for using AI ethically for teaching and learning must therefore be validity. AI should be used to achieve well-defined educational goals based on strong societal, educational or scientific evidence that this is for the benefit of learner. As shown by the examples of personalised learning and formative assessment, this criterion can be met by using AI to enhance established teaching and learning techniques. That said, the use of AI should not be tethered to existing educational practices. AI could for instance be used to achieve well-defined educational goals in ways that are supported by scientific research or best practice in other domains. As stated in the report *Intelligence Unleashed*, authored by authorities in AI in education at both University College London's Knowledge Lab and Pearson, AI "will continue to leverage new insights in disciplines such as psychology and educational neuroscience to better understand the learning process, and so build more accurate models that are better able to predict – and influence – a learner's progress, motivation, and perseverance."<sup>5</sup>

## **2. AI should be used to assess and recognise a broader range of learners' talents.**

The first principle "AI should be used to achieve well-defined educational goals based on strong societal, educational or scientific evidence that this is for the benefit of learner" is closely linked to the above principle .

Given that AI is adept at processing data and drawing out insight<sup>6</sup>, and that making use of insights into both learners' needs and into the efficacy of teaching practices are key ingredients of effective teaching and learning<sup>7</sup>, the Institute has highlighted this principle separately.

During the roundtables, a number of participants discussed the benefits of using AI to derive key insights into the learning process. One participant argued that "one of AI's key capacities is understanding learners' progress, how they learn, and what skills they have acquired. This data can support students and teachers alike, and enable better support and decisions, made in a more objective and data-driven way", another noted that "AI can support students to identify their own skills. This is particularly important for more independent learners".

The perspectives heard during roundtables chimed with research conducted into AI's capacity to support the fine-grained assessment of learning. Professor Luckin, for instance, has emphasised that "AI is a powerful tool to open up the 'black box of learning', by providing a deep, fine-grained understanding of when and how learning actually happens", and has demonstrated that AI can enable a broader range of learners' competencies and attributes - including collaboration, persistence, motivation and confidence - to be assessed, recognised and developed upon<sup>8</sup>.

Echoing the words of eminent educationalists Dr Peter Hill and Sir Michael Barber, the Institute is particularly keen to see AI being used to assess and recognise the "full range of student abilities", so that all learners are sufficiently prepared to thrive in a rapidly changing world<sup>9</sup>, and

so that nobody's talents are overlooked.

### **3. AI should increase the capacity of educational institutions whilst respecting human relationships**

The Institute heard that a key benefit of AI was its ability to automate many of the administrative tasks that would usually be completed by educators, thereby lightening their workloads and enabling them to reallocate time to more high-impact tasks.

A number of participants explained that the inherent features of AI make it particularly well suited to automating certain administrative tasks, noting that AI systems are better at collecting, processing and analysing data than humans.

One participant argued that a key benefit of reducing teacher workload would be the resultant increase in the attractiveness of becoming an educator. This chimed with a point raised in the Institute's interim report, in which the recruitment and retention crises were highlighted as critical problems that artificial intelligence could be used to address.

The points raised above also echo seminal research from McKinsey and Co, which estimated that up to 40% of educators' workloads could be automated by technologies that already exist. This report also demonstrated that time saved could allow teachers to attain an improved work-life balance, whilst also enabling educators' time to be "plowed back into improving education through more personalized learning and more direct coaching and mentoring"<sup>10</sup>.

During the Young People's Roundtables, participants discussed the question "should artificial intelligence be used to mark my homework Instead of a teacher?". Participants acknowledged that using AI to mark homework (or other work) could free up time for teachers, which could make teachers' roles less stressful, and create time for teachers to provide additional support to students. But despite this benefit, students expressed numerous misgivings around using AI to mark students' work.

Whilst accepting that AI would be able to mark multiple choice, or one word answers, students questioned whether AI could accurately mark less structured work, such as essays or extended written answers. Pushed to consider whether this concern could be allayed if AI were able to demonstrate its accuracy, students raised what they considered to be further disadvantages of using AI to mark learners' work. In particular, many students argued that by removing teachers from the marking process, an opportunity is lost for teachers to gain deep insights into a student and their work - insights which teachers could have used to provide fine-tuned support to learners.

A related concern raised was that an AI might not be able to appreciate the individuality of a piece of work. A number of students countered this point, however, arguing that the current educational system encourages highly original works to be boiled down to single scores.

The key insight from these conversations was that even in cases such as the automated marking of teacher workload, where the use of AI in education is often seen as overwhelmingly positive, there are still concerns that the overuse of AI could diminish student-teacher relationships and present opportunity costs for learners. The Framework therefore stipulates the need for a thorough assessment of risks relating to the marginalisation/replacement of educators wherever AI is used to reduce workload.

#### **4. AI systems should be used in ways that promote equity between different groups of learners and not in ways that discriminate against any group of learners**

Algorithmic bias has been raised as a central ethical issue arising from the use of artificial intelligence in general contexts<sup>11</sup>. The issue of biases was raised on numerous occasions throughout the roundtable series. Where the issue was raised, participants generally saw biases as a key risk that needed to be mitigated. During one roundtable, however, a number of participants argued that biases should not be seen as inherently negative, and that some level of bias was unavoidable.

A key insight from this discussion is captured in the following comment: “With bias, it’s not about removing it. It is about identifying and removing unjustified biases”. It was therefore argued that the solution lay in identifying potential sources of unjustified bias; and then taking and documenting decisions on the best ways to address these biases. Accepting that there may be no one-size-fits-all approach to identifying and documenting biases, the Framework stipulates that organisations developing/designing AI resources should identify these biases and document how established techniques, which may include periodic monitoring and/or increased levels of transparency, are used to address these.

In their report, *Review into Bias in Algorithmic Decision-making*, The Centre for Data Ethics and Innovation (CDEI) detail a number of approaches to detecting and mitigating biases. The Institute therefore recommends that these methods be used as a starting point for addressing biases through technical means. In the report, the CDEI also emphasises that ensuring that AI systems are designed by diverse and inclusive groups of people is also a powerful means of avoiding biases<sup>12</sup> (See Criteria 9.2).

During roundtables, numerous participants also expressed concerns that an increased uptake of artificial intelligence could result in inequitable outcomes by widening, rather than narrowing, educational inequalities. It was argued that disparities in access to hardware and connectivity and in the digital skills of different groups of learners could lead to those learners who are

already advantaged enjoying greater benefits from AI to the detriment of more disadvantaged learners. One participant also pointed out that it may be that only wealthy schools have the resources to fully utilise artificial intelligence.

These notes of caution were tempered with optimism, however. It was argued that in the context of lifelong learning, AI could benefit the lowest paid workers in particular by enabling education to be delivered efficiently and flexibly. It was argued that this could be especially helpful for those who work long and irregular hours and therefore can't commit to more structured and time-consuming forms of learning. Elsewhere, participants argued that by making high quality resources more affordable and available, AI could actually narrow divides between individual learners and between different types of schools.

The Institute is optimistic that AI can be used to narrow socioeconomic educational inequalities. That said, the Institute firmly acknowledges that divides in access to hardware, infrastructure and connectivity create major barriers to this. As such, the priority must be to help enable equality of uptake by working towards a situation where the hardware, infrastructure and connectivity needed to derive the benefits of AI are available to all learners. The Framework therefore calls upon those applying AI in education and those designing AI for education to take decisive steps to close this divide. As emphasised in the Institute's introduction to the Framework, governments have a central role to play in addressing disparities in access.

During the Global Summit and the roundtables, participants also expressed both optimism and concern around the use of AI to support learners with additional needs. As well as stressing how important it is that AI systems do not exhibit biases against learners with additional needs, participants were also keen to ensure that AI systems were applied in ways that were well suited to these groups of learners.

The Institute urges decision-makers to follow objective 1 - "AI should be used to achieve well-defined educational goals based on strong societal, educational or scientific evidence that this is for the benefit of learner" - whenever making decisions around how to support students with additional needs via the use of AI. By following this objective, practitioners and leaders can have confidence that AI is being used to address the complex needs of learners in valid ways that are supported by evidence. To enable equitable uses of AI in education, the Institute also stipulates that those procuring and applying AI in educational settings should ensure that those with additional needs are not excluded from using and benefiting from AI in education.

## 5. AI should be used to increase the level of control that learners have over their learning and development

*The Ethics Guidelines for Trustworthy AI* - an authoritative framework developed by The High Level Expert Group on Artificial Intelligence, which was set up by the European Commission - puts forward “respect for human autonomy” as a key principle for ethical artificial intelligence<sup>13</sup>. During the roundtables, a number of issues that were raised related to the interplay between autonomy and the use of artificial intelligence in education.

Participants were concerned about the risk of AI being used intentionally, or inadvertently, to manipulate and coerce learners, to “exploit their vulnerabilities”, or to “brainwash” them. One participant did note, however, that AI systems could influence learners’ behaviours for positive purposes, for instance, reducing users’ levels of stress. The Institute therefore stipulates that whilst it is crucial that all forms of coercion be outlawed, it may be appropriate to use AI to influence learners’ behaviours for beneficial purposes, providing that there is a clear scientific basis for any methods employed.

Participants also discussed the fact that AI systems can make predictions and/or recommendations based on extrapolations from past experiences, and that this has significant implications for the autonomy of learners. For instance, AI can be used to predict which students are most likely to fail or drop out of a university course. These predictions could enable pre-emptive action to be taken, therefore empowering learners with greater control over their developments. That said, such predictions could also be used to penalise learners, through expulsions, for instance<sup>14</sup>.

Influenced by this example, at the Young People’s Roundtables, the Institute asked students “would it matter if AI predicted that I was not going to do well in my exams?” This provoked many enlightening responses and spirited discussions. There was a strong concern that such predictions could affect students’ motivations, mindsets and/or self-perceptions. For those who were knowingly predicted to do well, there might be a risk of complacency. For those at the other end of the spectrum, unfavourable predictions could sap their motivation, leading to a self-fulfilling prophecy.

Students, however, respected that predictions also provided opportunities for support, and that, in cases where predictions were already routinely made, there was an argument for utilising AI, because human predictions alone can be biased (perhaps due to favouritism or unconscious assumptions about students). That aside, students did note that there might be a problem of invalid predictions being made about their future successes, and therefore questioned whether predictions could be given too much weight. There also expressed concerns that the antidote to invalid predictions - more intimate data - posed its own set of risks.

The following comments were also made:

“...scaling this up, if an AI system predicted you were going to commit a crime, I think that this would be unjust to convict based on this prediction. Scaling it back down, I think it could be dangerous to judge people based on data-driven predictions of how they might perform in their exams. What if there are other factors that haven’t been considered, what if the prediction is not accurate?”

“Let’s say if you have a student who struggled at first and then who studied really hard closer to the exam, then they could be penalised by a prediction.”

A key insight to draw from these discussions is that learners should not be held to account for predictions. These predictions do not represent actual achievements, disappointments or transgressions, and so should not be treated as such. Resonating with this sentiment, during the Global Summit on the Ethics of AI in Education, former Education Secretary Damian Hinds emphasised that “No child should be held back from their potential by some predictive algorithm”. The Institute firmly agrees with this pronouncement.

Further concerns raised included the risk of learners becoming addicted to using AI platforms or over reliant on these systems to learn. Based upon previous research into persuasive design techniques employed by software designed for young people<sup>15</sup>, there are concerns that advancements in AI could lead to heightened levels of addiction if left unchecked. The Institute therefore stipulates measures that should be taken to protect learners from addiction and practices designed to compel learners to extend their use of a resource beyond a point that is beneficial for their learning.

## **6. A balance should be struck between privacy and the legitimate use of data for achieving well-defined and desirable educational goals**

As previously noted, and as was emphasised at numerous points during roundtables and the Global Summit, a key strength of artificial intelligence is its ability to capture, analyse and draw insights from data.

Adding to the points heard about the benefits of AI in the context of formative assessment, the Institute also heard arguments that efforts should be made to derive greater levels of shared benefit from data, and that there is a “moral obligation” on educational institutions not to hoard data. That said, risks to learners’ privacy were emphasised strongly.

During the Young People’s Roundtables, students tended to agree that a balance needed to be struck between using data to enhance their education, and not using data in a way that encroached upon their privacy. It was generally accepted that some amount of learners’ data

was needed in order for AI systems to function effectively, and that data gathered through AI systems could also enable teachers to support learners more effectively.

Concerns were voiced too, however. Students expressed misgivings about high levels of monitoring and surveillance. One student noted that if AI were monitoring all aspects of their learning (referring to the context of continuous assessment), then they would feel as though big brother was watching them.

Students observed that protecting privacy often has more to do with upholding rights than mitigating risks. Students broadly agreed that they should have oversight over what data is collected on them, who it is shared with and how it is used. As part of this right, it was acknowledged that learners would need a strong baseline of knowledge about AI and how it uses data in order to exercise this right fully (see the Informed Participation section in the Framework).

Throughout the roundtable series, many participants cautioned against the use of AI to surveil learners. That said, it was clear from discussions that there were tensions between the benefits of monitoring of learning and the risks associated with surveillance.

A number of participants explained that artificial intelligence could facilitate a shift away from high-stakes assessments and towards continuous assessment of learning. This, it was argued, would enable a broader range of learners' competencies to be assessed and valued, and could make assessments a less stressful process for learners.

The benefits of AI-enabled continuous assessment aside, one participant argued that if students were constantly being assessed, they may feel as though they cannot "bring their full selves to the table" during lessons. Notwithstanding the significant benefits posed by continuous assessments, based on this concern, the Institute has decided to endorse a point made by one participant who argued that even if a shift does take place to continuous assessments, there should be safe spaces where learners are not being monitored.

Concerns were also raised that highly sensitive data could be gathered through the use of AI in education, and that this data could be used against learners. A related point that was made provides a potential response to this concern. It was argued that the risks related to users' data rights and privacy apply to almost all domains of life, not just education, and that the Institute should not be trying to tailor approaches for learners' data rights in the context of education.

The Institute broadly agreed with this point, however, as noted above, we do think there are specific issues relating to users' privacy that arise in the context of AI in education: namely, the need to be mindful of the line between beneficial ongoing monitoring and surveillance, and the need to minimise the depth of sensitive insights into learners that can be garnered through the use of AI in education.

## 7. Humans are ultimately responsible for educational outcomes and should therefore have an appropriate level of oversight of how AI systems operate

Throughout the Institute's consultations, there was a general consensus that AI should be used primarily to augment and empower educators, and not to marginalise or replace them. In the context of AI in education, however, human responsibility can sometimes be difficult to maintain due to the limited levels of oversight humans have over how AI systems operate, and because of the possibility that humans will not fully exercise their own judgement in response to AI-based predictions or recommendations.

Whilst some participants of roundtables argued that AI should always be explainable (i.e. able to explain why it produced a particular output), others argued that this was not always possible or preferable. Some expressed that less transparent AI systems can often be more accurate, and others argued that transparent AI systems would be more open to gaming, which could significantly diminish the learning experience and create unfair advantages for those who were better at cheating the system.

Regarding the question of how to resolve trade-offs associated with the transparency and explainability of AI systems, the Institute's judgement has been informed by insights from the Information Commissioner's Office. In their Project Explain Interim Report<sup>16</sup>, they found that in contexts such as recruitment and criminal justice, stakeholders expected a balance between explainability and accuracy due to the expectation that people should be able to learn from feedback, and improve their behaviours/qualities - processes that are enabled by explanations. The Institute therefore considers that where AI is used for educational purposes, accuracy should not be prioritised over explainability outright. Instead, levels of both accuracy and explainability should be reasonably high.

As "reasonably high" will be interpreted differently in varying contexts, it is appropriate for decision makers in educational settings to exercise their own judgement in regards to the most appropriate trade-off between accuracy and explainability. To aid them in doing so, the Institute stipulates that those procuring AI systems should insist that suppliers make explicit whether there were any trade-offs between accuracy and explainability in the design of the AI resource, specifying where any compromises have been made and providing a justification.

Overlaying the issue of opacity is the issue of deference. As Thea Snow has argued, they are: "Not good at predicting rare events[,] often trained on incomplete data[,] and] often trained on biased data, resulting in discriminatory tools." In her report, *Decision Making in the Age of the Algorithm*<sup>17</sup>, she makes the case that "human-machine interaction" is the "optimal approach", whereby professional judgment should be married with AI-based insights and recommendations, synergistically utilising AI's ability to analyse and evaluate large amounts of information with the ability of human experts to contextualise information and identify nuances. Snow argues that, as part of this process, "humans must challenge, scrutinise and question the guidance that the tool is offering", but humans must also be open to reflecting on the potential

flaws in their own decision making (due to unconscious biases for instance) which could be unearthed through the use of AI<sup>18</sup>. The Institute therefore stipulates that all practitioners using AI in education should be trained to scrutinise AI's behaviours and decisions (see criteria 8.2).

## **8. Learners and educators should have a reasonable understanding of artificial intelligence and its implications**

During the Global Summit on the Ethics of AI in Education, three working parties were convened with the specific remit of recommending mechanisms that would drive the ethical design and application of AI in practice. One of the recommendations was that strategic efforts should be made to raise awareness of AI's capabilities and to provide training for stakeholders. This included teaching students themselves about artificial intelligence and its societal and ethical implications.

The need to inform stakeholders about artificial intelligence and its implications also shone through during roundtable discussions. Participants argued that training and education was needed to enable educators to use AI effectively, to allay nerves, to increase confidence and trust in using AI, and to ensure practitioners would use AI critically. A further point was made that if stakeholders were better educated about AI, they would be in a position to hold providers of AI products/services to account and steer ethical design through the procurement process.

The insights heard during the roundtables and the Global Summit provide clear grounds for stipulating that educators need to be trained and upskilled on how to use AI discerningly and effectively. That said, the Institute is firmly of the position that learners too deserve to be educated about artificial intelligence and its implications. Indeed, it is their data that will often be processed, and it is they who will ultimately be impacted by the use of AI in education. As such, the Institute urges that learners have a right to understand what artificial intelligence is, how it uses data, and how it can impact them throughout their lives, not just as learners. Whilst educating learners in this way places demands on the curriculum, the Institute emphasises that these demands are justified.

Digital skills will be in high demand throughout society, as such empowering learners with such skills needs to be a priority. Digital skills go beyond coding. Learners need to be literate in the principles of how technologies operate and the societal implications of their use. The Institute therefore echoes the words of 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."<sup>19</sup>

## **9. AI resources should be designed by people who understand the impacts these resources will have**

Numerous roundtable participants urged that stakeholders, including learners and educators, should be involved in the design of AI products. It was argued that this would confer a number of benefits including increased alignment between the needs of learners and the capacities of AI resources, resources being designed with fairness in mind, and increased trust in AI systems amongst users.

As well as stipulating that stakeholders should be involved in the design of AI resources, the Institute also stipulates that the designers themselves (i.e. software developers) should represent the diversity of potential users. This draws upon a key stipulation in *The Ethics Guidelines for Trustworthy AI*, which states that “it is critical that, as AI systems perform more tasks on their own, the teams that design, develop, test and maintain, deploy and procure these systems reflect the diversity of users and of society in general.” This is also supported by guidance from McKinsey and Co, who state that “A more diverse AI community will be better equipped to anticipate, spot, and review issues of unfair bias and better able to engage communities likely affected by bias”<sup>20</sup>.

The Institute further stipulates that those designing AI resources should have an understanding of the ethical issues arising from the use of AI in education.

## References

- 1- The UK Government gives a broad definition of AI (Government Digital Service, Office for Artificial Intelligence, *A guide to using artificial intelligence in the public sector, 2019*); as explained in House of Lords Artificial Intelligence Committee, *AI in the UK: ready, willing and able?, 2018*, AI can also be understood as “the capacity to learn or adapt to new experiences or stimuli”
- 2- Government Digital Service, Office for Artificial Intelligence, *A guide to using artificial intelligence in the public sector, 2019*
- 3- The Education Endowment Foundation, Individualised Instruction, viewed at <https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/individualised-instruction/>
- 4- Luckin. R (2017), *Towards Artificial Intelligence-based assessment systems*, Nature Human Behaviour
- 5- Luckin, R., Holmes, W., Griffiths, M. & Forcier, L. B. (2016). *Intelligence Unleashed. An argument for AI in Education*. London: Pearson
- 6- Luckin. R (2017), *Towards Artificial Intelligence-based assessment systems*, Nature Human Behaviour
- 7 - See Black. P, Wiliam. D, *Inside the Black Box: Raising Standards Through Classroom Assessment*, Kings College London School of Education; Hattie. J, *Teachers Make a Difference, What is the Research Evidence*, Australian Council for Educational Research Research (ACER), 2003
- 8- Luckin. R (2017), *Towards Artificial Intelligence-based assessment systems*, Nature Human Behaviour
- 9 - Hill. P, Barber. M, *Preparing for a Renaissance in Assessment*, 2014, Pearson
- 10- McKinsey and Company, *How Artificial Intelligence Will Impact K-12 learners*, 2020, viewed at <https://www.mckinsey.com/industries/public-and-social-sector/our-insights/how-artificial-intelligence-will-impact-k-12-teachers>
- 11- See The Higher Level Expert Group on AI Ethics, *Ethics Guidelines for Trustworthy AI*, European Commission, 2019; and The Aletheia Framework, Rolls-Royce, viewed at <https://www.rolls-royce.com/~media/Files/R/Rolls-Royce/documents/stand-alone-pages/aletheia-framework-worksheet-v1.pdf>
- 12- Centre for Data Ethics and Innovation, *Review into Bias in Algorithmic Decision-making*, 2020
- 13- The Higher Level Expert Group on AI Ethics, *Ethics Guidelines for Trustworthy AI*, European Commission, 2019
- 14- Ekowo. M, Palmer. I, (2016), *The Promise and Peril of Predictive Analytics in Higher Education*, New America
- 15- Kidron. B, Evans. A, Afia. J, *Disrupted Childhood: The cost of Persuasive design*, 5Rights Foundation, 2018
- 16- The Information Commissioner's Office (ICO), *Project Explain: Interim Report*, 2019
- 17- Snow. T, *Decision-making in the Age of the Algorithm*, Nesta, 2019
- 18- Ibid
- 19- House of Lords Artificial Intelligence Committee, *AI in the UK: ready, willing and able?*, 2018
- 20- McKinsey and Company, *Tasking Bias in Artificial Intelligence and in Humans*, 2019 <https://>

[www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans](https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans)

### **Further Credit**

- The structure of The Ethical Framework for AI in Education was informed and influenced by The Aletheia Framework by Rolls-Royce (<https://www.rolls-royce.com/sustainability/ethics-and-compliance/the-aletheia-framework.aspx>)