The Institute for Ethical AI in Education

Interim Report: Summary

Towards a Shared Vision of Ethical AI in Education

THE UNIVERSITY OF BUCKINGHAM

Contents

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

Risks and Benefits of AI in Education

Possible 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[1]. NESTA has suggested that AI resources could help boost social mobility by enhancing teaching and learning and improving access to high quality learning materials[2]. 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.[3]

The benefits posed by AIEd stem from three fundamental factors.

AI can increase capacity within education systems and increase the productivity of educators. By reducing teacher workload and extending affordable and high quality learning opportunities, AI tools could address teacher recruitment and retention problems, and enable life-long learning provision globally.

AIEd can provide valuable insights that can enhance teaching and learning, and support learners' well-rounded development. 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."[4] This functionality can be used to provide insights to both teachers and learners.

AIEd can deliver 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. Through autonomous learning recommendations, AIEd enables individualised instruction and personalised learning at scale.

Possible risks of AI in Education

The Ethics Guidelines for Trustworthy AI puts forward seven Requirements of Trustworthy AI.[5] These requirements allow us to begin to understand the risks that AI could pose to learners.

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 oversight were undermined

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 inaccurate 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: "In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle."
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.

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 school, and cause difficulty with deciding who is accountable for poor performance, or inappropriate or harmful outcomes.

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.[6] 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.

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

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 upon these recommendations 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, and it also stipulates that there must be active monitoring of whether the criteria laid out in the blueprint have been met.

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 could outweigh risks at an individual level, and vice-versa. At an operational level, 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.

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: I 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

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- Luckin. R (2017), Towards Artificial Intelligence-based assessment systems, Nature Human Behaviour

5- Ethics Guidelines for Trustworthy AI (European Commission, 2019); https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai)

6- 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.