



The Institute for Ethical AI in Education

Provisional guidance for educational institutions deploying artificial intelligence: a self-assessment tool

Context

The Institute for Ethical AI in Education is developing an ethical framework that will, amongst other outcomes, serve to guide educational institutions in implementing and using AI in Education (AIEd) ethically and in the best interests of learners. This framework will be published in 2021. In the meantime, it is important that educational institutions are supported to ask the right questions and put appropriate systems in place in order to be confident that they are deploying AIEd responsibly, and are strategically managing the risks that their learners may experience as a result of the use of AIEd.

This self-assessment tool draws upon the risks posed by AIEd that were identified in The Institute's Interim Report, and puts forward a set of questions related to these risks that organisations deploying AIEd should be addressing in order to protect learners from harm. It is important to emphasise that this self-assessment tool is not intended to be an exhaustive and final ethical checklist for organisations implementing AIEd. That said, the questions in this self-assessment tool do correspond to genuine risks posed to learners, and organisations will want to manage these risks appropriately in the interests of learners. As such, we hope this tool will provide valuable support.

For a number of the questions posed, organisations are asked whether they have sought sufficient assurances around the quality of AIEd tools. Currently, the most appropriate way for organisations to ensure they can answer these questions positively is to demand appropriate information from providers. In the long-term though, The Institute hopes to support both providers and deployers of AIEd by advocating for appropriate systems to be in place so that educational organisations can be confident they are using ethical AIEd resources, and so that providers of AIEd can be recognised for their ethical practice. For other questions, organisations are asked whether they have systems in place to manage/mitigate particular risks to learners. At this stage, it should be left to organisations to design and implement such systems at their discretion - with The Institute's priority being to highlight the criteria that such systems should fulfil. That said, The Institute is keen to learn how organisations have chosen to design and implement such systems and to understand the challenges in doing so. If appropriate, The Institute may decide to provide additional resources to support organisations in the design and implementation of such systems.



Questions

Autonomy

- Have you had sufficient assurance that the AI tool does not employ techniques designed to cause addiction, inappropriately manipulate learners or otherwise inappropriately affect their behaviours?
- Do you have strategies in place to monitor the impact of AI tools on learners' behaviours and general wellbeing?

Accuracy

- Have you had sufficient assurance that the AI tool has been tested and has been confirmed to perform with a sufficient level of accuracy and/or efficacy?
- In cases where the AI tool makes decisions or acts in a way that relates to a learner's safety/wellbeing or could otherwise have a significant impact on a learner, does the Institution have safeguards in place to protect against inaccurate or ineffective outcomes from an AI tool?

Privacy

- As an organisation, are you complying with all relevant data and privacy law?

Fairness

- Where appropriate, have you had sufficient assurances that the system has been trained using data that is representative of either wider society or of the specific user base?
- Have you had sufficient assurances that steps have been taken to ensure there is diversity amongst teams involved in conceptualising, designing and developing the AIED resource?
- As an organisation, do you have an ongoing process in place to review whether resources are leading to biased outcomes?

Opportunity Costs

- Do you have systems in place at an organisational level, or provide guidance at an individual level, for evaluating whether AI resources are the most appropriate resources to use in particular circumstances?



Responsibility

- In cases where AI is being used, is a human professional still ultimately responsible for outcomes? If so, are strategies in place to ensure these professionals have an appropriate understanding of the resources that they are using?

Harmful Actions

- Do you as an organisation systematically ensure that insights gained about learners are not used to deny opportunities to, stigmatise or discriminate against learners?

Transparency

- In cases where the AI resource's actions and decisions are not readily amenable to human understanding (i.e. where AI tools cannot explain their behaviours), and where the use of AI is considered to have a critical impact on learners, do you have processes in place to ensure outputs from relevant AI tools are not relied upon too heavily?
- In cases where the AI resource's actions and decisions are not readily amenable to human understanding (i.e. where AI tools cannot explain their behaviours), and where the use of AI is not considered to have a critical impact on learners, do you have processes in place to independently verify the positive impact of the tool on learners?

The Institute for Ethical AI in Education is constantly learning as it works towards developing the ethical framework for the use of AI in education, as such, it may be appropriate to release further versions of this self-assessment tool and/or supplementary material. If The Institute does decide that further resources would be appropriate, we would intend to consult with organisations using this current self-assessment tool in order to understand how it can be developed upon practically as a resource.

If you wish to get in touch with The Institute about this resource, or otherwise, we can be contacted via our website at

<https://www.buckingham.ac.uk/research/the-institute-for-ethical-ai-in-education/>.