Learning to live with AI in high-stakes assessment

Artificial intelligence (AI) technology is set to revolutionise how we work and live. This includes its potential to drive creative change in the educational sector. However, AI is accelerating much faster than our capacity to understand and regulate it.

Our research focus

At AQA, we've been researching the use of AI in assessment for some time. We know that AI has the potential to mark some types of exam questions and to support the re-mark process, where a student has appealed the grade awarded to them. We've also considered how AI might assist with item writing.

Our work has highlighted a number of challenges to overcome if we want to steer AI for positive change.

Challenges to using AI systems

They can't explain their decisions

An AI system can't explain or justify the basis for its judgements in a way that we would find acceptable. They have no real-world context to draw upon, beyond the data they have been trained on. For example, an AI system can highlight the key words in a student's response that it has used to determine the mark, but it can't evaluate the quality of an argument as a human marker could.
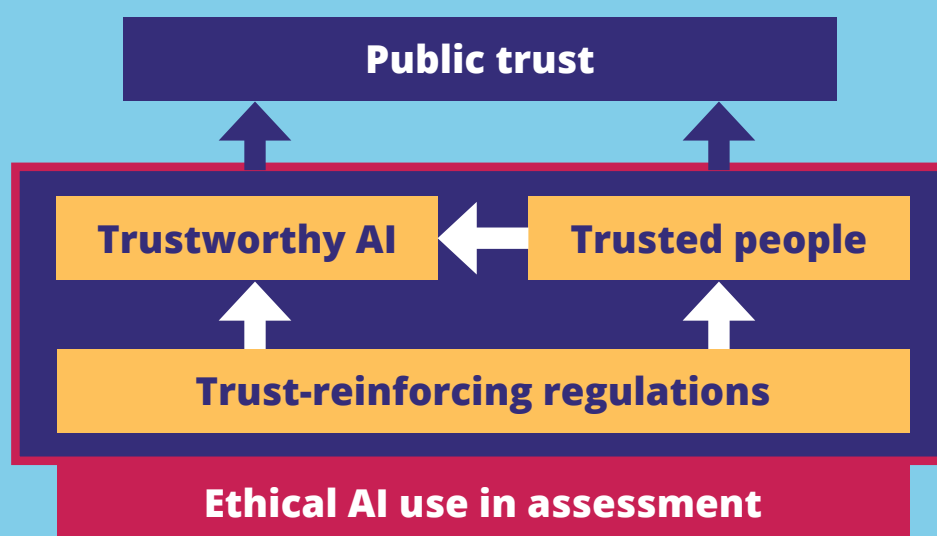
They are unreliable

Standard AI technologies are unable to tell the difference between important or unimportant content in a text. Newer generative AI technologies can't tell fact from fiction, and often repeat falsehoods.

Training them to behave ethically is not easy

AI systems 'learn' and reflect back wider societal attitudes that already exist in the data sets they're trained on. This means an AI system could treat some groups of people more favourably or discriminate against them, based on characteristics such as sex, ethnicity or religious beliefs. Continuous monitoring and correcting would be needed to prevent this.

AQA
Questions matter

What will it take to integrate AI into high-stakes assessment?

Public trust

Trustworthy AI ← **Trusted people**

Trust-reinforcing regulations

Ethical AI use in assessment

Questions around trust and accountability

- Who is responsible for the AI's 'behaviour'?

- Who should be expected to quality-assure the AI system?

- Who should have the final say in the case of a marking review – the human marker or the AI?

- Who holds the intellectual property of AI-generated content?

Embedding assessment expertise

The AI developers who train, assess and evaluate their systems are not experts in assessment. This skills gap can lead to unrealistic claims of AI competencies. Given that these systems will be used to perform tasks alongside or in place of people, it is vital to embed the right expertise at an early stage.

Looking to the future

AI holds great promise for an exciting future, with the potential to develop more personalised educational tools as well as opportunities to enhance writing and marking processes. However, research will need to focus on the issues of unreliability, bias and low explainability to maintain transparency and accountability in the exam system.