

CAN AI ESSAY SCORING BE BOTH EFFECTIVE AND TRANSPARENT?

ABBREVIATIONS

AI - Artificial Intelligence
AES - Automated Essay Scoring
EU - European Union

THE EU GUIDELINES FOR TRUSTWORTHY AI AND HIGHER EDUCATION ASSESSMENT

A group project applying the *EU Guidelines for Trustworthy AI* (2019) to assessment in higher education. The guidelines consist of seven interrelated areas of concern in all artificial intelligence systems: 1 Human agency and oversight • 2 Technical robustness and safety • 3 Privacy and data governance • **4 Transparency** • 5 Diversity, non-discrimination and fairness • 6 Societal and environmental wellbeing • 7 Accountability
This poster looks particularly at area 4. Transparency. This is applied to higher education assessment, specifically the marking of essays via artificial intelligence, through reviewing the EU guidelines as well as relevant literature on essay marking AI.

INTRODUCTION

Artificial Intelligence can perform tasks to standards beyond human capability. In higher education assessment, AI systems, replace the academics in marking student essays, in a manner that is theoretically more efficient, cheaper and less prone to bias. Such systems not only reduce the toll on academics, who may spend the time and effort elsewhere, but also are fairer to students, in terms of the speed of feedback, consistency of marking and lack of human subjectivity (Zhang, 2013).

However appealing the case for AES may be, is it ethically sound? Is it possible, for AES to perform adequately as well as adhere to transparency principles?

A TRANSPARENT ESSAY MARKING AI

Transparency in AI is not a single concept, but encompasses a range of principles, including, but not limited to algorithmic transparency, interpretability, trust, explainability, and accountability. In other words, not only is it important that **developers create trustworthy algorithms, but also that the whole system and its life cycle is ethically transparent and accountable** (Larsson & Heintz, 2020).

According to the *EU Guidelines for Trustworthy AI* (2019), transparency includes:

- **traceability** - documentation of the algorithms and the data flow through the system
- **explainability** - humans to be able to trace and understand the outputs of the AI system
- **communication** - stakeholders to understand that they are interacting with AI

In AES, communication is perhaps the least concerning area. The interaction between the students and the AI system is minimal, at least in the traditional assessment sense, and they only need to be informed that their essay is to be marked by a AI system and an explanation of its limitations.

AN EFFECTIVE ESSAY MARKING AI

AI is difficult to define, and has been done so in many different ways. A possible definition is that AI concerns theoretical computer processes that perform traditionally human tasks, to a better standard and with more efficiency.

Then in terms of AES, a fully effective system would be one which marks essays accurately, removing the subjectivity and errors associated with human involvement, and in a much more thorough and time efficient manner. The system should be able to learn from the input data and develop accordingly. Further, if automation is prioritised, then fully effective AI system should do all of the above with minimal or no human intervention. In other words, **the ideal AES would replicate the marking of an academic who is not limited by traits such as stereotyping, bias, fatigue and inconsistency.**

IMPLICATIONS FOR TRANSPARENCY

The main concern in transparency of AES systems is concerned with principles of traceability and explainability. Even if the algorithms are documented and theoretically understood by specialists, **the highly complex and mathematical processes** (Hussein et al. 2019) **are likely not to be understood by the main stakeholders.** These are often called black-box algorithms: where the input and the output is visible, but not the actual process. It is easy to see how receiving a single mark when submitting an essay to an AI system may cause scepticism at the very least. The more effective AI algorithms are, the more complex and untraceable they are likely to be.

This issue cannot be easily solved, if all, however, it may be bypassed by **focussing on alternate areas of transparency.** As described, transparency is not solely concerned with the traceability of the algorithmic processes, but also the whole life cycle of the system (EU Commission, 2019; Larsson & Heintz, 2019). For example, AES requires a corpus of pre-scored essays which essentially calibrate the system (Hussein et al., 2019). These essays are chosen and marked by humans, and so human errors and bias may be carried through to the AES. They also, currently, have limited accuracy, and search for specific features of the essay rather than viewing it as a whole. **Making information such as this clearly and readily available to the stakeholders increases the transparency even for black box algorithms.**

Another approach is to refer to transparency during the development of the system. For example, AES systems have been developed to not only output a score for the essay, but also specific feedback (Doewes & Pechenizkiy, 2020; Östling & Grigonyté, 2017). Indeed, some papers suggest that **the act of the AES to provide feedback of any type to users improves the transparency of the system.** A certain system colour codes individual words to show how they contribute positively and negatively towards the score. This is greater transparency than even human markers can offer.

A common idea in AES literature is that despite many benefits, these systems are only replicating traditional forms of assessment. AI has the potential to change these arguably outdated forms of assessment into processes more suited to higher education and industry today. These potential new AI assisted methods of assessment could prove to be more transparent than what is obtainable today (Bearman & Luckin, 2020).

CONCLUSION

The fully effective and transparent AES cannot be realised currently. As suggested in the EU guidelines, trade-offs will likely be required between cutting edge AI and traceability. While algorithmic transparency is difficult to attain, good communication, explainability, and consideration for wider ethical principles in the development and deployment of these systems is likely to prove satisfactory for stakeholders. As the fields of AI and machine behaviour progress, and trust in AES systems grow, **the human need for full transparency may also subside, as transparency is replaced with trust in the accountability of AI.**

LITERATURE REVIEW

AUTHOR	TITLE	KEY QUOTE	KEY POINTS
Bearman & Luckin (2020)	Preparing University Assessment for a World with AI	"current machine learning approaches align with the assessment we already do, rather than promoting the assessment we will need into the future"	AI has its own advantages, however is limited in terms of assessing creative and innovative work, and in defining quality. At the current moment, AI only replicates human forms of assessment, but has the potential to introduce new forms of more relevant and effective assessment.
Doewes & Pechenizkiy (2020)	Structural Explanation of Automated Essay Scoring	"Feedback in AES system provides transparency about the grading process."	AES often lacks transparency in the reasoning behind the essay score. Systems which are designed to provide feedback improves transparency and therefore accountability and trustworthiness, for both students and teachers, as it allows understanding of where there may be faults in the essay as well as the scoring system.
Hussein et al. (2019)	Automated language essay scoring systems: A literature review.	"AES systems do not assess the intrinsic qualities of an essay directly as human-raters do, but they utilize the correlation coefficients of the intrinsic qualities to predict the score"	There are many different types of techniques in AES and several systems have been developed. What these have in common is the complex mathematical nature of the algorithms, and the need for the system to be trained using a set of pre-scored essays. They also have in common the current inability to score creative propositions and evaluate their practicality.
Larsson & Heintz (2020)	Transparency in Artificial Intelligence	"Transparency in AI... takes a system's perspective rather than focusing on the individual algorithms or components used."	Algorithmic transparency is only one part of AI transparency, which includes, but is not limited to transparency of: data flow and storage, goals and outcomes, and compliance with regulation. It is important to take a systems perspective rather than focussing on individual components.
Östling & Grigonyté (2017)	Transparent text quality assessment with convolutional neural networks.	"Since one of our primary concerns is transparency, we choose a fixed-width convolutional neural network so that it is easy to infer how each part of the text contributes to the model's estimate."	Transparency can be achieved even with highly effective AI techniques, if the system itself is designed to provide feedback and show what parts of the essay contributes positively or negatively to the given score. In this way, developing AI systems with transparency and wider ethics in mind is likely to yield systems that are trusted by stakeholders.
Zhang (2013)	Contrasting Automated and Human Scoring of Essays	"The primary strength of automated scoring... lies in its efficiency, absolute consistency in applying the same evaluation criteria across essay submissions and over time, as well as its ability to provide fine-grained, instantaneous feedback."	Both human scoring and AES have their own respective strengths and weaknesses. AES is a viable possibility for the future so long as the fundamental difference between it and human scoring is understood. Advances into this field may show that AI essay scoring can be more transparent than human scoring: the concept of black box algorithms can easily be applied to the human mind as well.

Key References: Bearman, M. & Luckin, R. (2020) Preparing university assessment for a world with AI: tasks for human intelligence. In: Re-imagining University Assessment in a Digital World. Cham: Springer: 49-63. • Doewes, A. & Pechenizkiy, M. (2020) Structural Explanation of Automated Essay Scoring. In: Proceedings of The 13th International Conference on Educational Data Mining. • European Commission (2019) Ethics Guidelines for Trustworthy AI. • Hussein, M.A., Hassan, H. & Nassef, M. (2019) Automated language essay scoring systems: A literature review. PeerJ Computer Science, 5, p.e208. <http://doi.org/10.7717/peerj-cs.208> • Larsson, S. & Heintz, F. (2020) Transparency in artificial intelligence. Internet Policy Review, 9(2): 1-16. • Östling, R. & Grigonyté, G. (2017) Transparent text quality assessment with convolutional neural networks. In: Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications. • Zhang, M. (2013) Contrasting Automated and Human Scoring of Essays. R & D Connections, 2(12).