Resist AI by rethinking assessment

While some academics embrace large language models in higher education, Angelo Pirrone suggests that take-home assessments should be given a marginal role in favour of oral exams and classroom assignments, which mitigate or avoid the risks posed by AI tools such as ChatGPT

> We're looking at how switching to oral exams could mitigate the plagiarism risks of AI tool use in take-home assessment: https://t.co/SYIgDaS3oA As an educator, what would deter you from using more oral exams to assess your students?
>
> — LSE Higher Education Blog (@LSEHEBlog) March 23, 2023

https://platform.twitter.com/widgets.js

The promise and curse of AI

For better or for worse, our world is being revolutionised by artificial intelligence (AI). For instance, in medicine, AI systems are aiding radiologists in detecting medical abnormalities. In other areas, opinions regarding the role of AI are more controversial. Take the case of the visual arts where AI has opened up new possibilities; in the news, we hear about AI winning an art contest and the polarising reactions that experts and the general public have about AI-generated art. Similarly, academic publishing is dealing with the influence of AI systems that can generate scientific articles.

Despite plenty of arguably positive or potentially positive applications in many domains, AI (mis)use is accelerating what even a few years ago would have been considered a dystopian nightmare – from self-driving cars that can re-possess themselves if the owner misses a payment, to the sophistication of AI surveillance that is invading our daily life, to AI-assisted police sketches that reinforce racial bias, and AI systems that reinforce sexism or revive dangerous pseudo-sciences such as physiognomy.

The focus of this post is the disruptive and damaging effect of large language models on academic learning and assessment. Large language models are one of the most hyped

AI applications and their risks are currently a hot topic of debate. As an example of large language models, I will refer to ChatGPT, arguably the most celebrated large language model to date and the most discussed AI system in both academic and non-academic circles.

An existential threat to the take-home assignment

In many academic systems such as the British or American ones, a commonly adopted type of academic assessment is the take-home assignment. Students complete an assignment outside the classroom, outside timetabled hours, with no invigilation, over a few hours or days, using their own equipment and software. This is a form of assessment that can accommodate most students. The rationale behind these take-home assignments is that, in engaging with the course materials and additional sources while preparing the assignment, the assignment is seen as evidence of learning and understanding and allows students to demonstrate their knowledge and abilities. Assignments generally include essays, reviews, and reports and will vary from course to course.

Academic systems that rely on these take-home assignments are greatly affected by ChatGPT and large language models in general, given that students could use AI systems such as ChatGPT to produce their assignments. A simple web search shows a multitude of posts on websites in which students discuss strategies to avoid plagiarism when using large language models and how to use additional software for masking the output of ChatGPT or extending it – for instance by adding bibliographic references, which ChatGPT (at the time of writing this post) does not generate.

> *Limiting the role of AI in academia should not be seen as punishing students or making life harder for them*

I would like to highlight that it is not the output per se that matters, but the active learning that takes place while generating the assignment that facilitates understanding and

knowledge. Therefore, the use of large language models to produce academic assignments disrupts student learning, and is also a violation of academic integrity. While there is often a grey area between plagiarism and academic support (for instance, it is fine to use a grammar or spelling check tool), generating a response to the assignment prompt using AI misses the purpose of education as students are no longer involved in the active learning process that justifies these assignments in the first place. Put in other words, students should be evaluated on their own work, not that of ChatGPT.

Limiting the role of AI in academia should not be seen as punishing students or making life harder for them; on the contrary, it is a way to protect active learning and the very purpose of higher education in the hope that scores on assignments do not become the sole focus for students. Universities could use the services of any computer programmer to develop systems that detect the use of large language models, and in some cases such tools already exist. However, in the long run, this is not a viable route. In an endless arms race between AI and detection software, new AI systems will be developed to circumvent detection software, and new detection software will be developed to detect AI-generated content.

We should rethink academic assessment so that reliance on intrusive and damaging technology is minimised

Resisting AI

We need simple but effective solutions, rather than having to outsmart each AI system on a case-by-case basis. A key aspect of operational security is to avoid reliance on unnecessary technology whenever possible. For instance, concern about corporations hoarding personal data can be mitigated by using self-hosted, libre software rather than proprietary software. Similarly, we should rethink academic assessment so that reliance on intrusive and damaging technology is minimised. Some countries are already discussing and implementing similar proposals. Some Australian universities are dealing with the risks to academic integrity brought by large language models by re-adopting

paper and pen assignments in certain cases.

I see two ways in which this principle could be translated into the classroom – oral exams and invigilated exams completed in the classroom. In both cases, there is still room for take-home assignments, but their overall role and impact are significantly decreased. In invigilated assignments or oral exams, students could expand on and build on the take-home assignment or engage with different aspects of the course material.

Like every other form of assessment, oral exams have their pros and cons which have been described and discussed in detail in the literature. Oral exams are an opportunity for students to learn to communicate verbally and to do so extemporaneously, often without rehearsal or preparation – a skill that is going to be appreciated in the workplace. On the negative side, oral exams are more time-consuming and may not be feasible for introductory courses and courses with a high number of students. For those courses, classroom exams coupled with take-home assignments may be more appropriate. That said, I see no a priori reasons to oppose oral exams. Elsewhere, for example in Germany or Italy, oral exams have long been an important type of assessment in higher education and surveys show that, on average, students are supportive of oral exams.

> *In Germany or Italy, oral exams have long been an important type of assessment in higher education*

One of the common critiques of oral exams is that students may be evaluated according to different questions and this may lead to unfair outcomes for students. From this perspective, standardised assignments are a fairer form of assessment as students are presented with the same set of questions. However, I personally believe that precisely due to their unstructured nature, oral exams allow a more genuine appreciation of students' knowledge and abilities. If anything, the fact that oral exams allow teachers to adjust questions on the spot to assess students' knowledge is what makes this form of exam AI-proof and an attractive substitute to take-home assignments. While at first the reply to a topic from large language models (aka stochastic parrots) may seem impressive, subsequent systematically ambiguous replies indicate a system that probabilistically combines information without reference to meaning.

Precisely due to their unstructured nature, oral exams allow a more genuine appreciation of students' knowledge and abilities

Both oral and in-classroom exams are naturally AI resistant; for oral exams, it is straightforward to understand why. In the future, AI systems could be devised that disrupt even this form of assessment (say, auditory devices) but practically we are far from such applications. For classroom assignments that require the use of computers, university system administrators could block specific websites (eg known AI chatbots) or limit internet access to all programs other than those needed to complete and submit the assignment. These proposals limit the potential impact of AI systems during classroom exams. That said, we should not be overzealous in imposing limits or we risk giving exams too much importance.

Regarding inclusivity, another often mentioned limitation of oral exams, no single exam type can accommodate all students. While an oral exam may be a better prospect for dysgraphic students or students with non-verbal learning disorders, it may disproportionately disadvantage students with specific types of social disabilities. However, it should be noted that, contrary to popular opinion, introverted students do not always have a problem with public speaking or oral exams. Special care should be exercised to recognise cases in which oral exams may not be suitable and make sure that options are put in place; this could take the form of specially designed oral exams (eg so that a series of questions are known in advance), or the possibility to replace oral exams with other forms of exams such as in-classroom assignments. Similar adjustments should be made for in-classroom exams when necessary; computer equipment with support software should be offered to students that require it.

Special care should be exercised to recognise cases in which oral exams may not be suitable and make sure that options are put in place

As a closing note, if the form of assessment is the oral or in-classroom exam, then large language models could even be used by students to foster active learning, together with activities such as classroom discussions that are known to promote active learning. Again, the risk that needs to be avoided is to equate learning exclusively with performance on written assignments given that nowadays written assignments can be entirely produced by AI systems. We often hear or read that it is inevitable that most aspects of our life are going to be affected or even dominated by AI. While in some cases that may be acceptable, in others, such as the case of academic assessment, our responsibility as teachers, lecturers, and students may be to resist AI rather than embrace it.

Image source: main photo by Clem Onojeghuo on Unsplash

_____

_____