Hype, or the future of teaching and learning? 3 Limits to AI's ability to write students essays

The growth of original, AI-generated content demands that we reconsider methods of student assessment. **Clare Williams** *explores the impact of ChatGPT on pedagogy, identifying three current weaknesses that instructors might bear in mind when setting student assessment questions.*

Recently, an instructor from a higher education institution in the United States posted this on Reddit:
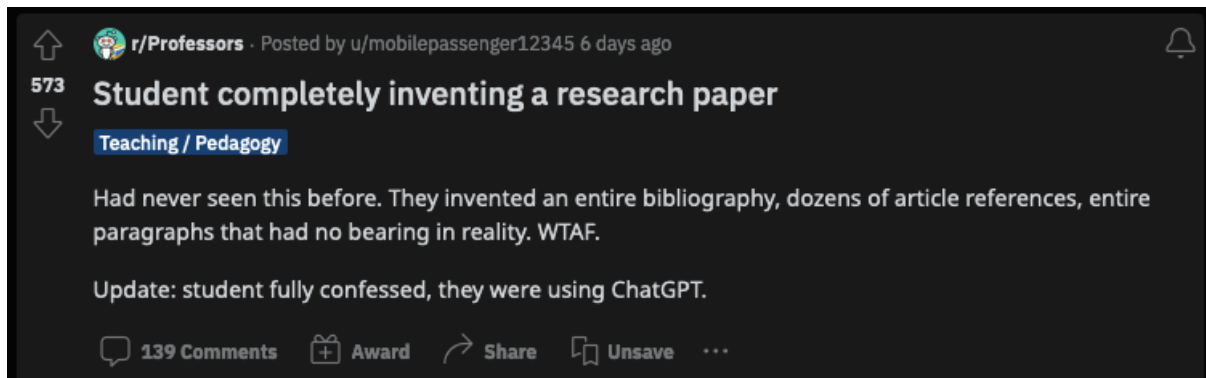

Figure 1: Screengrab from Reddit, post by u/mobilepassenger12345 posted to the sub r/Professors

Even before u/mobilepassenger12345 updated their post to confirm that the student's essay had indeed been written using artificial intelligence (AI), replies were quick to ask whether an algorithm, specifically OpenAI's latest chat tool, ChatGPT, had been the culprit.

Posts on this blog have explored how AI is reshaping many areas of everyday life, including education, notably by challenging established pedagogies through the creation of original student essays. ChatGPT is the latest iteration of natural language processing (NLP) artificial intelligence (AI) capable of not only holding a realistic conversation with a human, but of generating credible creative content. It has received significant coverage for its ability to pen original stories and poetry, write and debug original computer code, even feign sentience.

And yet, for each claim that ChatGPT can produce profoundly original and human-like content, there is a similar number of posts and articles claiming that the content is flawed, untrustworthy, or simply plain wrong. One conclusion we can take from the current debates is that, despite not being perfect, tools like ChatGPT are improving and are here to stay. Moreover, this is only a small step further than the AI we already expect students to use in their essays, such as spelling and grammar checkers in Microsoft Word, or apps like Grammarly.

What does this mean for pedagogy?

The implications for how we assess student knowledge and understanding are profound though: if we ask students to submit a piece of original work, how can we be sure that it is their own work, and not a (more or less) doctored response scripted by AI?

There are three potential areas that an instructor might keep in mind when designing assessment questions to ensure that the student must answer the question without relying too heavily on AI: current affairs, referencing, and critical analysis. To illustrate, I used a typical short essay prompt, as can be seen in Figure 2 below. The essay prompt is displayed next to the yellow box, with ChatGPT's response displayed next to the green box, reproduced in full:

CL Discuss the main trends and changes that have occured in regulatory governance in the last 60 years

There have been many significant trends and changes in regulatory governance over the last 60 years. Here are a few of the main ones:

1. Increased globalization: In the last 60 years, there has been a significant increase in the interconnectedness of economies and the movement of goods, services, and people across national borders. This has led to the development of international regulatory frameworks and the establishment of supranational regulatory bodies, such as the World Trade Organization (WTO) and the European Union (EU).

2. Growing complexity of regulatory issues: The last 60 years have also seen an increase in the complexity of regulatory issues, with more and more areas of society and the economy coming under the purview of regulatory authorities. This has led to the development of specialized regulatory agencies and the growth of regulatory expertise.

3. Greater emphasis on risk management: In recent decades, there has been a shift towards a more proactive, risk-based approach to regulatory governance. This involves identifying and addressing potential risks before they become problems, rather than simply reacting to problems as they occur.

4. Increased focus on transparency and accountability: In recent years, there has been a growing emphasis on transparency and accountability in regulatory governance. This includes the use of open data and open government initiatives to increase transparency and the adoption of measures to ensure that regulatory authorities are accountable to the public.

5. Emerging technologies: The rapid pace of technological change in recent years has brought new challenges and opportunities for regulatory governance. Regulators have had to grapple with the regulation of emerging technologies, such as artificial intelligence and the Internet of Things, and how to balance the benefits of these technologies with the potential risks they pose.

Figure 2: Essay prompt and ChatGPT response from December 2022

Critiquing ChatGPT: what to look out for

As we can see from Figure 2, AI tools like ChatGPT still have limitations that instructors can be aware of when setting student essay questions. As ChatGPT is trained on data up to mid- to late-2021 and is not connected to the internet, it currently has no knowledge of the world after that date. In other words, ChatGPT is not performing a quick Google search in the background as it formulates a response, and this gives instructors the possibility of assessing the application of knowledge and understanding to more recent events that lie beyond the capabilities of ChatGPT. The technology also seemingly struggles with in-depth critical analysis demanded by prompts. While it can retrieve, compare, and contrast

viewpoints of different authors, it cannot engage deeply with the arguments, again offering instructors a point of entry for assessing their students' understanding. In the example given above, we might amend the assessment question to ask students to demonstrate their knowledge and understanding of changes in regulatory governance by applying it to ongoing events or current affairs.

Another sure sign that all is not well in a student's essay is their referencing. While some instructors have found that ChatGPT can offer reliable references for some material, the majority, like u/mobilepassenger12345 posting on Reddit above, seem to find that the algorithm simply invents sources that look convincing to the untrained, non-expert eye, but which look suspicious to anyone with in-depth knowledge of the subject. Additionally, any requirement that references are presented in a different format according to disciplinary norms, for example OSCOLA in law, fall apart as all ChatGPT referencing collapses into a haphazard Harvard style. At the present time, this can be one of the first places to look to check whether AI has been a core element in the production of an essay.

Do we need to worry?

Current affairs, referencing, and critical reasoning seem to offer three key weaknesses of current AI models like ChatGPT that instructors can take advantage of to ensure that the student essay remain a robust and reliable means of assessment. However, this approach is unlikely to be fool proof, and the emergent qualities of foundation models and their downstream applications mean that such weaknesses might no longer be as reliable next year. GPT4 is just around the corner, and likely to be trained on data from interactions with the current iteration. While AI-generated detectors are emerging in response, they lag behind NLP AI and have, so far, not achieved a sufficiently high degree of accuracy to be reliable. Does this mean that the demise of the student essay as a method of assessment is nigh?

The growth of AI-generated original content could mean that the student essay will not be a reliable way of assessing learning for much longer. The impact of this will vary from field to field, but might have particular relevance for degrees that include professional qualifications, like the qualifying law degree (QLD). Professional and regulatory bodies might do well to work with higher education institutions to (re)consider what it means to learn and understand in a world where AI can generate the content for us, and for our students. HEIs might also want to urgently consider the boundaries and borderlands of academic misconduct in an age of AI. If a student uses AI to generate an (original) essay outline and then fleshes out their answer, does this constitute academic misconduct? How much AI is too much?

How can we respond? Instead of banning the use of AI-generated content in education assessment, we might do well to consider harnessing tools like ChatGPT as a starting point for student essays. Students are likely to be graduating into a world that is increasingly reliant on AI, for example, law graduates, will be competing with tools like DoNotPay, "The World's First Robot Lawyer". As tools like ChatGPT become embedded in the professional spheres of competencies in which students will work, it seems pointless and even counterintuitive to expect students to avoid AI altogether in their studies. Why not offer the

opportunity to work *with* the technology, learning how to use it safely and constructively? So, for example, a research essay might expect students to use ChatGPT to create an initial essay outline that they then critique and correct, update and extend. One thing is certain: ChatGPT should prompt instructors to revisit how students are assessed if rigorous, robust assessment of knowledge and understanding is to be continued in the era of AI.