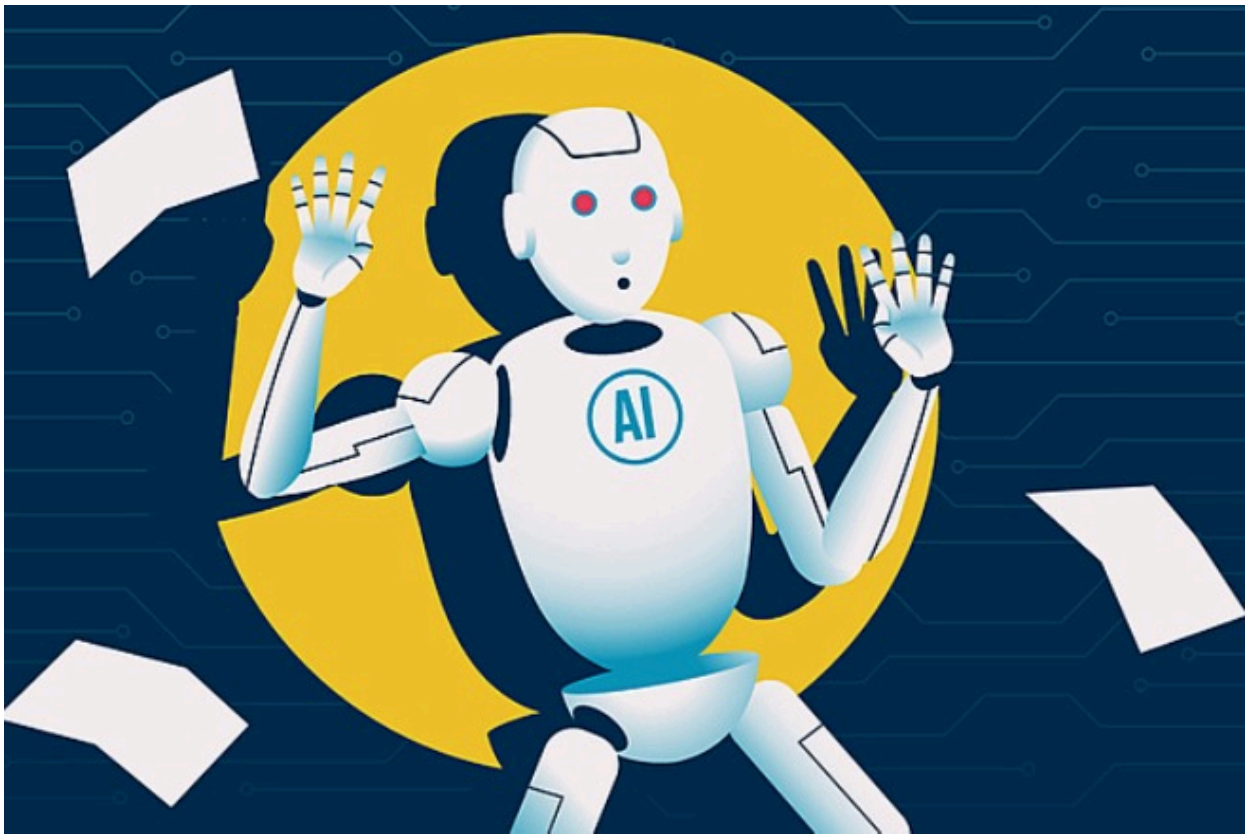
Universities must help shut down the illicit AI detection economy

timeshighereducation.com/opinion/universities-must-help-shut-down-illicit-ai-detection-economy

June 24, 2026



Source: Moor Studio/Getty Images

It only took a few months after the release of ChatGPT in late 2022 for vendors of academic integrity tools to embed artificial intelligence (AI) detection features into their systems.

Many institutions now treat the AI-likelihood scores they generate as evidence in academic misconduct procedures, and some universities are setting explicit thresholds above which disciplinary action may follow. However, the role of these tools is not always articulated clearly to students (or sometimes even to instructors). And this generates anxiety – and a black market in managing it.

Students are naturally worried about the degree to which their use of AI will be deemed appropriate by markers and whether declaring certain AI uses may lead to academic integrity investigations and, potentially, disciplinary action. Moreover, while “academic

integrity tools" advertise their systems as highly accurate, cases of false positives are frequently shared on social media. Many students will be aware that even the US Declaration of Independence has been flagged as AI-generated. And the tools have been found to be particularly prone to error when assessing texts written by non-native speakers of English – a particular issue in Hong Kong.

As instructors ourselves, we have noticed that students are increasingly asking us questions about their AI-likelihood score. And the combination of high uncertainty and high stakes is precisely the condition needed for a workaround market to emerge.

Sure enough, we have recently become aware that, for a few dollars, a student can buy their score from the same proprietary integrity tool as their university uses. Numerous sellers offer this service on popular e-commerce platforms – either as one-off reports, bulk purchases or even monthly subscriptions. How these sellers are obtaining access to the proprietary academic integrity tools that are only available to official institutions is unclear. But it is abundantly clear that they are doing so.

Students purchase a report, review it, edit their essay accordingly, and only then submit through the institution's integrity system. Some submit the same assignment to these paid services multiple times, tweaking it until the score drops to a level they consider "safe".

Comments on the e-commerce sites suggest the agents take no accountability for their services. Vendors often promise that submitted texts will not be archived in the integrity systems – which would instantly flag them as having been plagiarised when they are resubmitted through the university's own system. But they cannot credibly guarantee this, and there is no recourse if drafts are retained or leaked. Comments under some vendors' advertisements left by disgruntled clients suggest that texts submitted by them have indeed been archived, leading to plagiarism accusations.

Although it is difficult to know how big this market is, the number of vendors on e-commerce sites and our conversations with students and colleagues suggest that using these services is becoming a common practice. And we have noticed that many agents follow dynamic pricing strategies, raising their rates during assessment periods, when everyone is submitting final essays.

All this is yet another instance of the hugely frustrating cat-and-mouse game around academic integrity that AI has led us to. The situation is further muddied by the emergence of commercial AI writing tools that include "AI humaniser" or "AI rewriter" features, which identify text that reads as if it is AI-generated and provide suggestions on

how it can be rewritten. The outcome is a system in which one set of AI tools is being used to counter another, with the student left to make difficult choices without all the required information needed to make good ones.

The most troubling consequence of what we call the AI detection economy is pedagogical. Anyone involved in academic writing will know that revision – reconsidering arguments, tightening evidence, sharpening voice – is one of the most valuable cognitive activities students engage in. However, conversations with students reveal that they now treat revision primarily as an exercise in reducing their AI score. On social media, students share AI prompts engineered not to improve an argument but to make a paragraph less “AI-like”, such as deliberately adding grammatical mistakes. We have observed students celebrating on social media their ability to achieve low AI scores (despite having used AI) and passing tips on how to do so through word of mouth, thereby normalising the practice.

Although universities have not deliberately created the AI detection economy, they must address the fact that they have inadvertently done so by explaining clearly what they use academic integrity tools for, how AI-likelihood scores are interpreted and what they cannot reveal. At least this way students will be clearer about what they should be worried about (AI text generation) and what they should not (AI language editing).

In addition, the pedagogical value of detection tools needs an honest re-examination. Trust between students and instructors has already been reduced by inconsistent AI declarations by the former and ambiguous AI use policies from the latter. If the use of AI detection tools is making that situation worse – and, at the same time, damaging learning – that cost has to be weighed against whatever deterrent effect the tools offer around cheating. Recognising that dilemma, some universities have already banned, disabled or discouraged the use of their academic integrity systems’ AI detection features. This might be a good way to go.

Finally, vendors of integrity tools must take responsibility for who can access their products. Allowing proprietary detection scores to be sold on open marketplaces directly undermines the systems they are paid by universities to protect. They must investigate and cut off this practice.

As for students, we have a lot of sympathy for them. They have been faced with a very complex and often contradictory reality since the release of AI tools. That’s why, in the coming year, we plan to offer workshops directly addressing academic integrity in the age of AI.

We hope that by engaging in dialogue, clarifying the role of AI detection and answering questions, we can reduce students' anxiety and reassure them that we are on their side.

That way, we also hope to push the AI detection economy towards a natural death.

Benjamin Luke Moorhouse is an associate professor and James Mian Jia is an assistant professor in the department of English at the City University of Hong Kong.