**ChatGPT appears to pass medical school exams. Educators are now rethinking assessments**

ABC Science

By technology reporter James Purtill

Posted Thu 12 Jan 2023 at 3:00am, updated Mon 23 Jan 2023 at 1:42pm

The artificial intelligence (AI) tool ChatGPT is capable of passing — or at least nearly passing — the US medical licensing exams, according to researchers in the US who put it to the test.

It's the latest achievement for the publicly available and currently free technology, which was released at the end of last year and has been the subject of non-stop coverage since.

ChatGPT is one of a class of "generative AI" programs that can produce images, video and audio, make arguments, summarise books, tell jokes, write code, and generally be useful to people.

But for educators, the technology opens the door to widespread cheating on homework and take-home assignments, and many have been scrambling to rethink the nature of assessment or otherwise discourage students using the tool.

This week, the New York school system banned the use of ChatGPT, while Australian universities said they were reinstating "pen and paper" exams and beefing up cheating detection measures.

Now, in a pre-print study that has not yet been peer reviewed, researchers have explored the upper limits of ChatGPT's capabilities.

They say the AI tool achieved over 50 per cent in one of the most difficult standardised tests around: the US medical licensing exam (USMLE).

**'Comfortably within the passing range'**

Just weeks after the launch of ChatGPT in December last year, researchers at a California-based healthcare provider, Ansible Health, began experimenting with the tool in their day-to-day work.

They found it could help with tasks such as drafting payment notices, simplifying jargon-dense radiology reports, and even to brainstorm answers for "diagnostically challenging cases".

"Overall, our clinicians reported a 33 per cent decrease ... in the time required to complete documentation and indirect patient care tasks," the study authors wrote.

To test the program's ability to perform clinical reasoning, they had it sit a mock, abbreviated version of the USMLE, which is required for any doctor to obtain a license to practice medicine in the US.

The USMLE consists of three exams, with the first generally taken by second-year medical students, the second by those in their fourth year, and the last by physicians after a year of postgraduate education.

For most applicants, the tests require more than a year of dedicated preparation time. The first two tests each take a day, and the last takes two days.

The researchers fed questions from previous exams to ChatGPT and had the answers, ranging from open-ended written responses to multiple choice, independently scored by two physician adjudicators.

They also checked that the answers to those questions weren't likely to be in the dataset accessible by the AI tool when it had been trained.

In other words, ChatGPT hadn't already seen the answers.

"ChatGPT performed at or near the passing threshold for all three exams without any specialised training or reinforcement," the paper reads.

The tool received more than 50 per cent across all examinations, and approached the USMLE pass threshold of about 60 per cent.

"Therefore, ChatGPT is now comfortably within the passing range," the paper concludes.

**Just another tool, like a calculator?**

Phillip Dawson, an academic integrity researcher at Deakin University, said he wasn't able to evaluate the study itself, but that "if the authors really did what they say they've done, then that's scary stuff.

"There's a sense that this is going to be even bigger than the pandemic in terms of how it changes assessment."



*Some Australian universities say they will rely more on pen and paper examinations to discourage the use of ChatGPT.(AFP: Martin Bureau )*

Kane Murdoch, the head of academic misconduct at Macquarie University, said he was "not surprised at all" that ChatGPT could pass the USMLE.

"[And] those are pretty serious and complex exams — simpler assessments would be a piece of cake."

He and others are pushing for universities to embrace ChatGPT, rather than banning it outright.

"[ChatGPT] is like the advent of the calculator — a game changer," he said.

"Telling students that using it is forbidden won't stop usage.

"I expect it to be very heavily used until such times as universities develop new strategies for assessment."

The Tertiary Education Quality and Standards Agency (TEQSA), which regulates higher education in Australia, appears to agree that ChatGPT shouldn't be banned.

"That's not a practical or sustainable strategy," said Helen Gniel, who runs TEQSA's higher education integrity unit.

"Machine learning is only going to improve. It's going to become quite standard."

**Can it be detected?**

ChatGPT can generate plausible academic writing that's generally very hard for educators or existing academic integrity software to detect, although the style is bland and formulaic, and it has a habit of making up facts and references.

Kane Murdoch, whose job at Macquarie University includes detecting the use of AI text-generators, said most academics "don't know what they're looking for" and would fail to notice when a student has used ChatGPT.

"What I'm looking for is really gross errors of fact," he said.

Turnitin is one of the world's largest plagiarism detection services, widely used by Australian schools and universities.

James Thorley, Turnitin's Asia-Pacific regional vice president, said generative AI marks the start of a third era for the company, which was founded more than 20 years ago in the early days of the internet.

In the first era, plagiarism amounted to copying and pasting digitised information.

In the second, which began around 2010, students began paying others to complete their assignments, which is known as contract cheating.

Generative AI is much harder to detect, he said.

"These AI tools are much more impressive — they're much less obvious."



*An example of what ChatGPT will generate from a text prompt.(Supplied: OpenAI)*

Turnitin will launch a new service this year to detect the use of AI paraphrasing tools, sometimes called "text spinners". This particular service won't help with detecting the use of AI-generated text.

AI-generated text could be detected though, Mr Thorley said.

"AI-generated text is much more predictable than human-generated text at the moment."

But eventually it'll be "entirely indistinguishable" from human writing, he added.

"I think the bigger question here is [around] the incorporation of these kinds of tools.

"What is the expectation of the student in terms of being able to write originally and being able to communicate original thought?"

**Does using it count as cheating?**

That depends on the rules of the education institution, but undeclared use of ChatGPT or similar tools would, in general, be considered academic misconduct, Mr Murdoch said.

"I think universities need to move beyond this way of thinking," he added.

Cath Ellis, a professor with UNSW's school of the arts and media, agreed.

"These tools exist — they're out there in the world. They're being used in professions or they're going to be used."

High schools are also grappling with the question of whether using generative AI tools counts as cheating.

Rob Barugh, director of learning technology at Hale School in Perth, said ChatGPT may count as a "support tool", similar to being helped by a parent or older sibling.

"I'm not going to call it outright cheating if they've used a support tool," he said.

But he added that using generative AI to complete assignments could undermine "the neurological aspect of deep learning", and this would be a problem.

"That's a tricky one for us to get our heads around."

Asked how schools would respond to students using ChatGPT, WA Department of Education Chief Information Officer David Dans said schools can choose to block specific sites and applications.

Each school is required to develop a policy for senior secondary assessments, which includes how it will address any cheating and plagiarism matters.

In NSW, the Department of Education said it was reviewing student access to AI software including ChatGPT.

“The Department of Education takes cheating and malpractice in academic work and exams very seriously, with robust measures in place to deal with this," a spokesperson said.

**The death of the essay?**

A syllabus that embraced the use of AI tools would probably have more oral assessments and fewer standard take-home assignments, said Professor Ellis from UNSW, who's an expert on academic integrity.

That could mean fewer essays.

"You're given a question by a teacher, put it into chatbot and less than five minutes later you have a 1,500-word essay with apparently logical looking references in it," she said.

"It doesn't actually produce learning.

"We're going to have to be more prepared to have conversations with students where we ask them to explain verbally to us what their work says."

She and others said student cheating was reflective of a broader issue of universities and educational institutions emphasising assessment over "genuine learning".

"What if we change how we assess students, and instead of reading what they've claimed to have written, we spend time talking to them about what they've learned?" Professor Ellis said.

Macquarie University's Kane Murdoch agreed.

"We're currently focusing students' minds on assessment rather than learning," he said.

"We need to better understand the pros and cons, consider how such tools fit into professional contexts, and engage in a very serious change process with regard to how and what we assess."

But how quickly educators can change their assessment methods remains unclear.

Meanwhile, the academic year starts within weeks.

And when students return, ChatGPT will be coming with them.