BY JOEL WALMSLEY

"ChatGPT and the Philosophers"

Artificial Intelligence (AI) is rapidly becoming an integral part of many industries, and the field of actuarial science is no exception. One of the most exciting developments in AI is the advent of language models, such as Open AI's ChatGPT. These models have the ability to understand and respond to natural language inputs, making them ideal for use in a variety of applications, including chatbots, virtual assistants, and language translation. In addition to its potential use in the field of actuarial science, ChatGPT is also being used in education to help students improve their language skills and writing abilities.

J. Walmsley is a Philosopher at University College Cork, in Ireland.

... Or so the enthusiasts say. The doomsayers, by contrast, are sceptical. In my own field of university education, for example, I've heard many a colleague echoing Socrates, who, in Plato's dialogue Phaedrus (370 BC), expresses similar worries about the invention of ... wait for it ... *writing*. He says:

"This invention will produce forgetfulness in the minds of those who learn to use it ... you offer your pupils the appearance of wisdom, not true wisdom, for they will read many things without instruction and will therefore seem to know many things, when they are for the most part ignorant and hard to get along with, since they are not wise, but only appear wise."

LINGUISTIC ABILITY IS WHAT DISTINGUISHES HUMANS FROM ANIMALS

Fast forward a couple of millennia and the worry is the same; students will use this new technology to write their essays for them, in ways that escape the notice of standard plagiarism-detection tools, giving rise to a new form of turbo-charged AI-powered cheating.

What's particularly unsettling is that language use has long been regarded as the hallmark of intelligence. The philosopher René Descartes (who you may remember from such lines as "I think, therefore I am") held that linguistic ability is what distinguishes humans from animals; he thought it was evidence that we have immortal souls whilst they are mere automata. In addition, in his 1637 book *Discourse on the Method*, he famously claimed that "it is inconceivable that a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as even the dullest of men can do." Uncannily enough, nearly 400 years later, such a thing is not only conceivable: it's *actual*.

Despite his general scepticism about AI, Descartes thereby anticipated another famous twentieth century development in the field: the Turing Test. In his landmark 1950 essay entitled "Computing Machinery and Intelligence" Alan Turing suggested that instead of asking the somewhat nebulous question "Can a machine think?" we'd be much better off addressing the more concrete issue: "Could a machine deceive a human interlocutor into thinking that they were conversing with another human?" Turing's idea was that we can never really be certain what, or that, another person (or machine) is actually thinking, but that conversational ability is like the tip of an iceberg: a pretty good and generally reliable way of inferring that there's a whole lot more going on beneath the evidence at the surface.

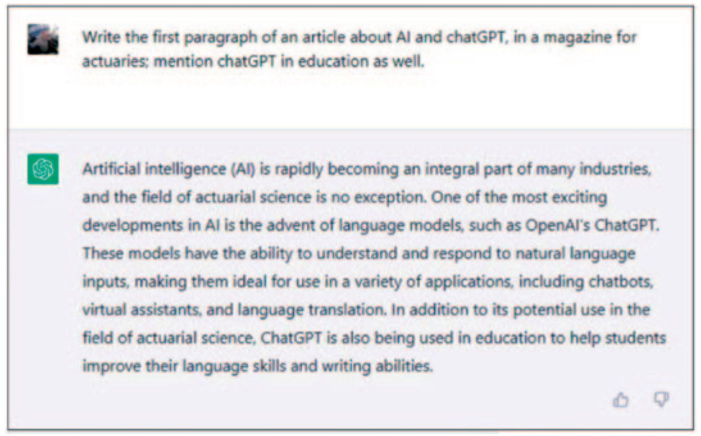
I sometimes joke with my students that the essays and exams that I ask them to write serve much the same purpose as the Turing Test. I can't really be certain that they've learnt and understood everything that we've covered, but the 2500 words that they hand in at the end of the semester is like the tip of an iceberg; if I construct the assignment carefully, it gives me a pretty good and generally reliable way of inferring that they've learnt a whole lot more, as a backdrop to the evidence presented in the essay. So the worry I mentioned earlier is that since ChatGPT can clearly pass the Turing Test, then it could also pass university courses -- on behalf of cheating students -- in the same way.[1]

WE CAN REDUCE THE LIKELIHOOD OF CHEATING, MAKE THE ASSIGNMENTS MORE INTERESTING

... Or could it? Notice that I said "...*if I construct the assignment carefully*." Many "standard" paper topics (e.g., "Describe and explain Philosopher X's idea about Y") are the sorts of prompts for which ChatGPT can produce a perfectly serviceable passing essay in response. But these were *never* good essay questions, because they encourage exactly what Socrates was worried about: rote learning and boring, formulaic answers. Instead, by finding ways for students to work with ChatGPT, we can reduce the likelihood of cheating, make the assignments more interesting, *and* really develop the critical analytic abilities that we were seeking all along.

In one of my classes this term, students must first ask ChatGPT to answer as if it were René Descartes (it refuses to impersonate someone, but it responds well if you start with "Let's play a game..."). They must then conduct an interview about its (his?) views on the nature of the mind, using ChatGPT to role-play a dialogue, and then critique the answers based on what they know from the philosophical texts. In another class, I've asked students to conduct an actual Turing Test with ChatGPT, and then to evaluate its performance based both on what they've read in Turing's paper, and on what they know about how ChatGPT works.

So rather than trying to ban the use of ChatGPT, or developing yet more sophisticated plagiarism detection software (and risking precipitating some kind of technological arms race), there is a better way to work with the technology, just as we did with writing (contrary to Socrates). For that reason, on balance and contrary to the doomsayers, I am inclined to side with the slightly more optimistic opening paragraph above, even though it's a sentiment for which, I'm willing to admit, I had a "co-author":



Write the first paragraph of an article about AI and chatGPT, in a magazine for actuaries; mention chatGPT in education as well.

Artificial intelligence (AI) is rapidly becoming an integral part of many industries, and the field of actuarial science is no exception. One of the most exciting developments in AI is the advent of language models, such as OpenAI's ChatGPT. These models have the ability to understand and respond to natural language inputs, making them ideal for use in a variety of applications, including chatbots, virtual assistants, and language translation. In addition to its potential use in the field of actuarial science, ChatGPT is also being used in education to help students improve their language skills and writing abilities.

■

[1] – See https://www.nbcnews.com/tech/tech-news/chatgpt-passes-mba-exam-wharton-professor-rcna67036 and https://edition.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html for example.