ChatGPT: Student builds app to sniff out AI-written essays

- By Nadine Yousif
- BBC News

13 January 2023
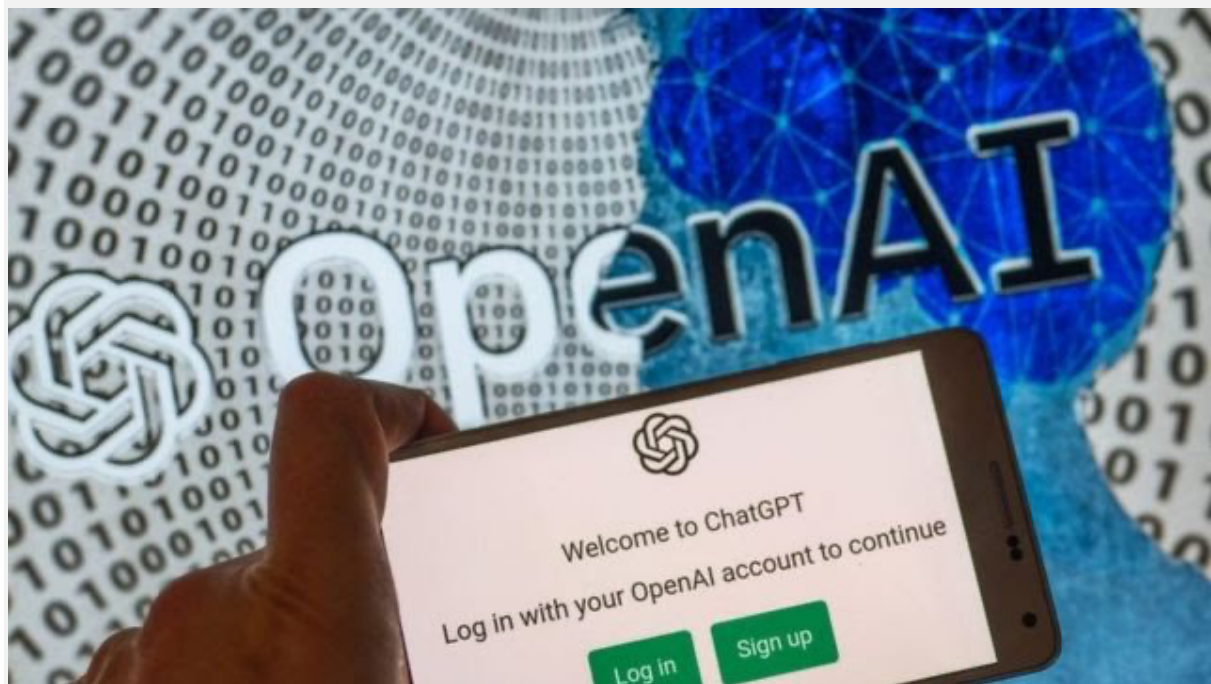
Edward Tian's tool has gone viral on social media, and has garnered global attention from school teachers and university admissions officers

Edward Tian has been thinking about the power of artificial intelligence for a number of years.

But it was in a packed lecture last year that the computer science student at Princeton University saw how advanced this technology had become. His thesis adviser displayed a set of text in front of the class and asked the students to differentiate between what had been written by a human and what had been AI generated.

Many students guessed wrong. He realised then that there was a problem that needed a solution.

"This technology is only going to get better and better, AI is here to stay. This is the future," Mr Tian told the BBC.

"But at the same time, I believe we need to enter this future responsibly."

This is why the 22-year-old spent his winter holiday break at a coffee shop in his hometown of Toronto working on an application that can determine, with high accuracy, if a text was written by a human or a bot.

He created it in response to the emergence of ChatGPT late last year - a free online chatbot that can expertly write almost anything, from English essays and news articles, to meal plans and computer code, all from a simple prompt.

- ChatGPT: New AI chatbot has everyone talking to it
- How human-like are the most sophisticated chatbots?

The popularity of ChatGPT since its launch has been met with alarm, including from some US schools who have blocked it on their servers in a bid to prevent students from cheating. Others worry the bot will take jobs away from writers and creatives, or will be used for more sinister purposes by hackers to write harmful malware.

Mr Tian, who is in his senior year at Princeton, said the app he developed, GPTZero, was the first step to address a host of concerns that could arise as artificial intelligence becomes smarter and more easily accessible.

IMAGE SOURCE, EDWARD TIAN

Edward Tian is a 22-year-old senior studying computer science at Princeton University

The app works by looking at two variables in a text - perplexity and burstiness - and it assigns each of those variables a score.

First, the app measures how familiar it is with the text presented given what it has seen during training. The less familiar it is then the higher the text's perplexity is, meaning "it's more likely to be human-written", Mr Tian said.

It then measures burstiness by scanning the text to see how variable it is. For example, does the text have a mix of short versus long sentences? Or does the writing appear to be more levelled and uniform?

"If you plot precisely over time, a human-written article will vary a lot," Mr Tian said. "It would go up and down, it would have sudden spikes."

He is still working on improving GPTZero, but he has released a beta version for public use. In a tweet, he demonstrated how the app can successfully sniff out the difference between an essay published in the New Yorker magazine versus a letter written by ChatGPT.

He said he has also since tested it out by feeding the app BBC articles written by journalists, versus articles written by ChatGPT using the same headline as a prompt. (Mr Tian formerly worked with the BBC's investigations unit). He said the app successfully guessed the difference between the texts with a less than 2% false positive rate.

Since its launch, Mr Tian's app has been used by thousands of people. He said he has since been contacted by teachers and university admissions officers from around the world who are interested in how it works.

While GPTZero was created to combat academic plagiarism, Mr Tian said he sees apps like his being used to address other issues that will come with the rising popularity of artificial intelligence such as online disinformation campaigns.
He is, however, not opposed to artificial intelligence - in fact, he said he was very excited about its emergence, and has found it useful in helping him to write computer code and solve other problems. But he said it was important to develop safeguards for any new technology as it gives its use a sense of credibility.

But he said that, above all, the popularity of his app speaks to "a human urge to know the truth".