

NEWS > EDUCATION • News

OpenAI launches new tool to deter cheating on its own platform – with mixed results

So far, those in the technology and education spaces are not impressed



A ChatGPT prompt is shown on a device near a public school in Brooklyn, New York, Thursday, Jan. 5, 2023. New York City school officials started blocking this week the impressive but controversial writing tool that can generate paragraphs of human-like text. (AP Photo/Peter Morgan)

By **ELISSA MIOLENE** |

emiolene@bayareanewsgroup.com | Bay Area News

Group

February 13, 2023 at 5:45 a.m.

[ChatGPT](#) — an artificial intelligence tool that can write essays, poems and emails on any subject with the click of a cursor — sent shockwaves throughout the education world when it was introduced late last year. Now, its creators have built a new program that can help catch students who use the AI bot to cheat.

But instead of quelling teachers' fears, the new detection tool has been somewhat of a letdown within the technology and education worlds. Created by San Francisco-based company OpenAI, the platform identifies AI-written text accurately only a quarter of the time — and gives a false positive for nearly 1 in 10 submissions. Even a different detection tool created by a Princeton college student works slightly better.

Some experts worry that the company's detection program could lead to wrongly accusing students of plagiarism.

"I'm surprised OpenAI would release this tool with its current performance level," said Victor Lee, the faculty lead for AI + Education at the [Stanford Accelerator for Learning](#). "If it's taken up too quickly, it could be really risky and harmful to students."

After the chatbot's release late last year, school districts across the country reacted with alarm. With the ability to churn out content in seconds, ChatGPT was seen by many as the ultimate cheat code for students — and, some believed, the nail in the coffin for original writing. Students began turning in AI-written assignments; companies began integrating ChatGPT into their copywriting protocols. And within weeks, its use was banned in schools across New York City, Los Angeles and Seattle.

Some of the world's largest tech companies scrambled to catch up. Last week, Microsoft revealed it was [integrating](#) ChatGPT's technology into its Bing search engine. Google followed with its own version, called [Bard](#).

"It's going to change everything," said Jake Carr, an English teacher in Chico, earlier this month.

By developing a tool to distinguish between human-written and AI-written text, the creators of ChatGPT, San Francisco-based [OpenAI](#), hope to deter the bot's involvement with automated misinformation campaigns, cheating in schools and other sorts of AI-infused deception.

"We're making this classifier publicly available to get feedback on whether imperfect tools like this one are useful," OpenAI said in a recent blog post. "Our work on the detection of AI-generated text will continue, and we hope to share improved methods in the future."

But the company's text classifier is late to the game. In early January, Princeton student Edward Tian built [GPTZero](#), a tool with the tagline "humans deserve the truth."

When comparing the same two ChatGPT-written essays on OpenAI's text classifier and Tian's GPTZero — three paragraphs on the dangers of plagiarism — the text classifier said the text was "possibly AI-generated," while GPTZero said it was "likely to be written entirely by AI."

On top of that, the "AI text classifier" is not particularly reliable. According to the company, the tool identified computer-written

text in 26% of their evaluations, while incorrectly labeling human-written text as AI 9% of the time. Detection is particularly poor with text shorter than 1,000 characters. Still, OpenAI said its text classifier is more accurate than other tools trying to analyze text.

Lee said he appreciates the company's openness about its classifier's limitations, but he still has concerns.

"I do worry that many people will hear that there is a tool that will detect things without looking at the fine print about its performance."

[Report an error](#)

[Policies and Standards](#)

[Contact Us](#)



Tags: [Artificial Intelligence](#), [K-12 Education](#), [Regional](#), [Technology](#)



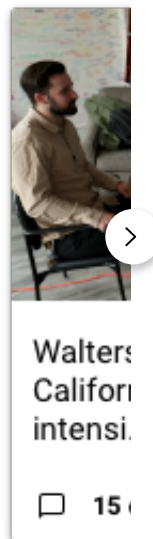
Elissa Miolene | Education Reporter

Elissa Miolene covers education for the Bay Area News Group. She previously wrote for the San Francisco Chronicle, The Wall Street Journal, and other media outlets, where she focused on migration and equity issues. Before shifting to journalism, Elissa led communications for humanitarian and United Nations agencies in the United States, East Africa, and South Asia, where she wrote about child protection, global health, and humanitarian crisis. Elissa holds master's degrees in journalism and policy from Stanford and Lehigh University respectively, along with a bachelor's degree in journalism and global studies from Lehigh.

emiolene@bayareanewsgroup.com [Follow Elissa Miolene @elissamio](#)

[More discussion on Mercurynews.com](#)

ADVERTISEMENT



[Open ad privacy settings](#)

Conversation

FOLLOW

SIGN OUT



Start the conversation, Victor Lee

ALL COMMENTS

NEWEST



Start the conversation

ADVERTISEMENT

[Open ad privacy settings](#)

Powered by viafoura

California Do This Instead Of Buying Expensive Solar Panels (It's Genius)

California Energy Program | Sponsored

California: Don't Buy Expensive Solar Panels. Do This Instead (It's Brilliant)

SunValue | Sponsored

When You Eat Instant Oatmeal Every Day, This is What Happens

Total Restore | Sponsored

Randy Jackson: This 3 Minute Routine Transformed My Health

Unify Health Labs | Sponsored

Diabetes Is Not From Sweets! Meet The Main Enemy Of Diabetes

Health Benefits | Sponsored

The Stunning New 2023 Volkswagen Bus is Here

Volkswagen Deals | Top Searches | Sponsored

The Volkswagen Bus Is Back With A Vengeance

Take a look inside the exciting debut of the VW ID.Buzz

The New VW Bus | Sponsored

New SUVs On Sale Have Us Speechless

All Things Auto | Sponsored

The New Volvo Recharge Is A Jaw Dropper

Check out popular searches related to your favorite ...

2023 New Volvos | Sponsored

Acura Has Done It Again. This Year's Lineup Has Left Us Speechless

All Things Auto | Sponsored

Revealed: The Year's New SUVs For Seniors (See The Prices)

Senior Offers | Search Ads | Sponsored

California: Program Covers The Cost To Install Solar If You Own A Home In These Zips

Energy Bill Program | Sponsored

The New BMW Lineup Leaves Nothing To The Imagination

New 2023 BMW Lineup | Sponsored

Miss Manners: It was a simple request, and I believe my fiance's response was rude

Mercury News

Dear Abby: My disabled husband cheated, and I want to put him in a home. How do I g...

Mercury News

The Comfy Women's Sandals For Walking All Day Without Discomfort.

The promotion ends as soon as the reduced models are sold

The promotion ends as soon as the reduced models are sold out.

[Zekear](#) | Sponsored

Volkswagen has done it again. This lineup has left us speechless

[Volkswagen Deals](#) | [Top Searches](#) | Sponsored

Incredible Slimming Discovery Takes Stanford By Storm

[Slimming Discovery](#) | Sponsored

This One Commercial Made Her A Millionaire

[FamilyThis](#) | Sponsored

These 2 Vegetables Will Kill Your Belly And Arm Fat Overnight!

[Health & Welfare](#) | Sponsored

Ask Amy: My friends are making a bad parenting decision, and they won't listen to...

[Mercury News](#)

Homeowner invokes 'builders remedy' in brazen plan to build 20-unit housing compl...

[Mercury News](#)

Root Cause Of Female Hair Loss

[Hair La Vie](#) | Sponsored

Empty Alaska Cruise Cabins Cost Almost Nothing

Alaska Cruise Deals | sponsored searches | Sponsored

California: New "Grocery Allowance" Everyone on Medicare Can Apply for

Flex Allowance | Sponsored



SPONSORED CONTENT

Secure Your Business With These Proven Steps Against Cybersecurity Threats

By Comcast Business

COMCAST
BUSINESS

Public or private, large or small, no organization is immune to being targeted by hackers and cybercriminals. That's...

**SUBSCRIBE
TODAY!**
ALL ACCESS DIGITAL OFFER FOR
JUST 99 CENTS!