-----

GPTZero is an artificial intelligence detection software developed to identify artificially generated text, such as those produced by large language models . [ 3 ] [ 4 ] [ 5 ] [ 6 ]

While GPTZero was praised for its efforts to prevent academic dishonesty, many news outlets criticized the tool's false positive rate, which can be especially harmful in academic settings. [ 7 ]

History

GPTZero was developed by Edward Tian, a Princeton University undergraduate student, and launched online in January 2023 in response to concerns about AI-generated usage in academic plagiarism . [ 8 ] [ 2 ] GPTZero said in May 2023 it raised over 3.5 million dollars in seed funding . [ 9 ] [ 10 ]

In the first week of its release, the GPTZero experienced 30,000 uses, which led to a crash. It was supported by the web app company Streamlit, who allocated more server resources in response. [ 11 ] In July 2024, it had 4 million users, compared to 1 million one year earlier. [ 12 ]

In summer 2024, GPTZero raised $10 million in Series A round funding. [ 13 ]

In September 2024, GPTZero announced an authorship tracking software that enables "to compile and share data about their writing process such as their copy/paste history, the number of editors they had, and how long editing took", in an effort "to move away from an all-or-nothing paradigm around AI writing towards a more nuanced one." [ 13 ]

Mechanism

GPTZero uses qualities it terms perplexity and burstiness to attempt determining if a passage was written by a AI. [ 14 ] According to the company, perplexity is how random the text in the sentence is, and whether the way the sentence is constructed is unusual or "surprising" for the application. Texts with language that is more chaotic or unfamiliar to language models -- i.e., that are likely to "perplex" the model -- are deemed more likely to be written by humans. [ 15 ] In contrast, burstiness compares sentences with each other, determining their similarity. Human text is more discontinuous, meaning humans tend to write with more sentence variation than AI . [ 2 ]

Use cases

The academic community has attempted using GPTZero to tackle concerns about AI-generated content for plagiarism . [ 16 ] [ 15 ] [ 14 ] Educational institutions, including Princeton University , have discussed the use of GPTZero to combat AI-generated content in academic settings, with mixed opinions. [ 2 ] [ 17 ] In October 2023, GPTZero had partnered with the American Federation of Teachers . [ 18 ]

By 2024, Tian reported that GPTZero also "received a lot of adoption with hiring managers, with recruiting [and] cover letter analysis." [ 13 ]

Efficacy

In a March 2023 paper named "Can AI-Generated Text be Reliably Detected?", [ 19 ] computer scientists Vinu Sankar Sadasivan, Aounon Kumar, Sriram Balasubramanian, Wenxiao Wang , and Soheil Feizi from the University of Maryland demonstrate empirically and theoretically that several AI-text detectors are not reliable in practical scenarios. [ 20 ] [ 21 ]

Tech website Futurism tested the tool, and said that while the "results are impressive", based on its error rate, teachers relying on the tool would end up "falsely accusing nearly 20 percent of [innocent students] of academic misconduct". [ 22 ]

The Washington Post noted in August 2023 that GPTZero suffers from false positives, emphasizing that "even a small 'false positive' error rate means that some students could be wrongly accused [of academic misconduct]". [ 7 ]

News website Ars Technica commented that humans can still write sentences in a highly regular way, leading to false positives. The writer, Benj Edwards, went on to state that the perplexity score only concerns itself with what is "surprising" for the AI, leading to instances where highly common texts, such as the US Constitution , are labeled as likely AI-generated. [ 23 ]

See also

Artificial intelligence content detection

Ethics of artificial intelligence

Natural language processing

ChatGPT

Text generation

Turing test

References

External links

Cassidy, Caitlin (January 11, 2023). "College student claims app can detect essays written by chatbot ChatGPT" . The Guardian .