

HOW RELIABLE ARE AI DETECTORS?

written by Dr. Bruce Ellis

AI (artificial intelligence) is quickly making an impact on our lives. Though AI has been around for many years, easy access and potential for mass use have caused ripples throughout many industries – with education being no exception. While AI is getting better at producing human-like written language, it's not yet fully able to do so without being detected. Plus, as AI models refine their algorithms, AI detectors will need to adjust as well, to accurately predict whether a piece of text is likely written by AI or a person.

A recent example of AI detection gone wrong? As **announced last week**, a professor failed half his class based on the AI detector analysis of his students' work, which turned out to be inaccurate. That brings me to two questions. How does AI generate information on a topic or prompt that you give it? And how reliable are AI detectors? Let's take a look at some of the current tools.

“

As AI models refine their algorithms, AI detectors will need to adjust as well to accurately predict whether a piece of text is likely written by AI or a person.

”

Testing AI Detectors with AI-Generated Text

To test out some AI detectors, I used the same prompt for **ChatGPT-3.5**, **ChatGPT-4**, **Google Bard**, and **Microsoft Bing**. As a prompt, I chose to ask it for information that would address an 8th grade history standard:

Write a 100 word essay on the 8th grade Texas history standard: The student understands the foundations of representative government in the United States. The student is expected to explain the reasons for the growth of representative government and institutions during the colonial period.

I did not ask the models to regenerate the response; I took the initial response that was produced by each. You can **view the actual responses** that I used to test the AI detectors and compare the similarities and differences between them.

The Results

I looked at free AI detectors, since these are what most teachers may rely on unless their district or school provides them with a subscription to a different detector model. For some of these tools, I did have to create an account, while others were publicly accessible. Here are the results for each tool when fed the four different AI-generated texts. It is important to note that for some (many?) detectors, just a sampling of the text is reviewed to give an overall score.

	ChatGPT-3.5 Text	ChatGPT-4 Text	Google Bard Text	Microsoft Bing Text
<u>Copyleaks AI Content Detector</u>	Human text	Human text	96.4% Probability for AI	Human text
<u>Hugging Face</u>	81.9% Probability for Human	99.5% Probability for Human	99.9% Probability for AI	58.6% Probability for AI
<u>Content at Scale AI Detector</u>	82% Probability for Human	82% Probability for Human	68% Probability for AI	78% Probability for AI
<u>OpenAI Text Classifier</u>	Likely AI- generated	Likely AI- generated	Likely AI- generated	Unclear it if is AI-generated
<u>GPTZero</u>	Likely to be written entirely by a human	Likely to be written entirely by a human	Likely to be written entirely by AI	Likely to be written entirely by a human

<u>Writer</u>	100% Human- generated content	0% Human- generated content	100% Human- generated content	100% Human generated content
<u>Sapling</u>	Fake: 100%	Fake: 0%	Fake: 99.8%	Fake: 99.6%

Results when AI detectors were fed AI-generated text related to education.

Can we rely on AI detectors?

AI detection seems to be a cat-and-mouse game where the models are striving to get increasingly human-like while AI detectors are struggling to maintain any level of accuracy and reliability. As of yet, you would be better off flipping a coin to determine if a text is AI-generated. The best strategy is to know your students. Does what they turned in fit with the whole of their work? Does it align with the skills they demonstrate in class?

Just like teaching students how to effectively use search engines, now students need to know how to ethically use AI tools as part of their productivity process. Consider how you and your students can use AI tools in the classroom to enhance the learning experience. Have conversations as to how and when using AI tools is appropriate for your classroom. Discuss what “cheating” looks like in your classroom and encourage them to have integrity. Whether we like it or not, AI is here to stay. It is now part

of the landscape – which includes the classroom but also includes their future careers.