Why TTU Recommends Caution with AI Detectors
Summer, 2024
Contributors: Suzanne Tapp, Alec Cattell, JaWana Green, Matt Gregory, and Brian Quinn

Our conversations about AI are often focused on academic misconduct and AI abuse. The AI Guidelines and Resources Committee suggests extreme caution when considering AI detection tools. The bottom line is that no conclusive proof can be provided by an AI detection tool at this time. Given the easy access students have to generative AI tools, we find ourselves at a crossroads in education.

Known Problems with AI Detectors
In response to concerns about academic integrity, we have seen a surge in companies who claim to be able to detect AI use such as GPTZero, ZerGPT, CrossPlag, and PassedAI. But we have also seen AI humanizer products such as Bypass GPT, HIX Bypass, Humbot.ai, Undetectable AI, and WriteHuman AI.

AI detection software in its current iteration is far from foolproof, with high false positive error rates (Edwards, 2023; Fowler, 2023). OpenAI, the company behind ChatGPT, even shut down their own AI detection software because of its poor accuracy (Nelson, 2023). AI detectors have been particularly unreliable with students who are not native English speakers (Myers, 2023), often unfairly identifying their work as AI-generated because of low perplexity scores, which measure the sophistication of the writing. However, there is some promising new evidence that tools may become more accurate in detecting AI-written content. For instance, Jiang (2024) sampled large-scale data from the Graduate Record Examinations (GRE) writing assessment with near perfect accuracy of AI-generated writing and no evidence of bias against non-native English speakers.

It is impossible to predict the future reliability of AI detectors, particularly as new versions of AI generators improve. **For now, the bottom line is that AI detectors can only *predict* whether a piece is AI generated, and these predictions are not reliable enough to support decision-making in suspected cases of academic dishonesty.**

What *can* an AI tool do?
According to the Center for Teaching Excellence at the University of Kansas ("Careful Use of AI Detectors," 2024), AI tools can indicate that instructors might need to make comparisons to a student's past work for differences in the style and quality of writing. Flagged material might also suggest that faculty need to talk with students and explain that the detector has indicated that some of the material is AI generated. In fact, we might say that the true first line of defense when unauthorized AI use is suspected is a conversation between the student and the instructor of the course. A simple non-accusatory conversation about the student's work and how the student went about completing the assignment can provide much information concerning the potential use of AI. Given how knowledgeable our faculty and instructors are in their fields, it would likely be relatively easy to discern whether a given student has the requisite

knowledge to have completed the assignment. In many instances, the student may openly admit the use of AI to complete the assignment. Instructors may want to give students a second chance to write material that is flagged by a detector. The instructor might also try to reproduce the work by attempting to use AI by inputting plausible key words or prompts. Then, the instructor can compare what AI produces with the student's submitted work. If an instructor strongly suspects that a student has misrepresented AI material as their own, a report from an AI detector can serve as one piece of evidence when an academic integrity incident report form is filed with the Office of Student Conduct.

Detecting AI Without Tools

AI detectors are easy to game and can be eluded by asking AI models to paraphrase text or to write in the style of the author. So, how can you detect text written by AI without using AI tools? You might start by checking the writing style and voice. Note any sudden improvement in writing quality, such as writing that is overly correct, elaborate, or formal. AI writing is unusually "clean" and free of mistakes, typos, and slang, making it much closer to "perfect writing" than human writing. Next, check the language used in the text. Do the words used reflect the content and terminology from your course lectures, discussions, and assigned readings? Look for telltale AI-generated words or phrases that authors have neglected to edit out, such as "regenerate" "or "regenerate response" or "As an AI language model, I am unable to…" Also, check the references. AI generates sources that may appear real but are often fabricated or contain a combination of real and fabricated information. Be aware that there is some evidence that AI generators are improving their ability to identify direct quotes and sources without hallucination (Bauschard, 2024). There are some additional indicators of AI-generated text that might stand out to instructors. One of them is implausible statements: "It takes 60 minutes to make a cup of coffee." Another is overuse of certain words or phrases, such as "the" "it" or "is" as a way to fill up space (Heikkila, 2022). The text may contain few references to current events, indicating the AI model may have been trained on earlier material. AI-generated text may seem superficial and generic, lacking any depth in analysis or interpretation. The writing may be characterized by a lack of personal experience or emotion. Machine-generated text can seem impersonal, mechanical or robotic. These are just a few tips on how to identify signs that a text may have been generated by AI. If you suspect that a student has submitted AI-generated work, the first step is to contact the student and request to meet with them. In many cases, students confess that they have submitted AI-generated work when it becomes apparent that they are unable to summarize and explain what they have submitted in conversation with their instructor. Such conversations can be uncomfortable for both the instructor and the student, but they can also be an important opportunity for the student to learn why it is important that they complete and submit their work in accordance with course policies.

The TTU Office of Student Conduct recommends the following strategies to further support instructors in discouraging the dishonest use of AI:

- Consider gathering an in-class writing sample from your students for comparison purposes, should the need arise.
- If you talk with an individual student about your concerns over inappropriate AI use, summarize and document your conversation. You might ask the student to verbally demonstrate their understanding of the topic.
- Please note that AI detector results cannot serve as the sole, definitive evidence as to whether AI was used irresponsibly by a student.

Do I have to go back to handwritten, in-class assessments?

Oftentimes we see suggestions in the literature about returning to "old school" assessments to remove opportunities for inappropriate AI use. Faculty who teach large sections or have a heavy workload simply may not be able to facilitate this type of change. Further, reliance on handwritten exams, in-class writing, or oral presentations can raise equity concerns (Ceres, 2023). Timed, handwritten exams may be challenging for students who are unaccustomed to composing and writing by hand or unable to write quickly (Tai et al., 2022). Oral presentations may cause anxiety, particularly for non-native speakers of English. (Grieve et al, 2021). Perhaps a better solution is provide a variety of assessment strategies to level the field for all students. The AI Resources and Guidelines Committee has curated a list of sample assignments, rubrics, and assessment tools from faculty across campus who have developed these to account for the presence of AI. These may serve as templates as you consider enacting shifts in your course assignments and assessments.

Does Our New Learning Management System Have an AI Detector?

Most learning management systems have a built-in plagiarism detector. For example, SimCheck (which is a TurnItIn product) is the integrated plagiarism tool in Canvas. These detectors typically provide an AI Score, which is the percentage of work that is predicted to be written by AI. However, a quick search on Google or even TikTok will identify ways to bypass the detector by humanizing AI-generated content.

Is It Time to Reconsider Plagiarism?

What is the future of academic integrity? Although outside the scope of this white paper, we acknowledge that there are broader conversations about the ethical implications of AI on academia for students *and* for faculty. Some suggest that we are entering a new era of postplagiarism in which technologies such as AI are viewed as normal and even inseparable from daily life. Eaton (2023) offers an accessible introduction to concepts of postplagiarism, including a framework that suggests that hybrid human-AI writing will be expected. Soon enough, we will have students who have never known school without AI, and we must continue to grapple with our conceptualization of ethics and AI as this reality takes shape.

Now what?

We understand the frustration that faculty likely feel as they encounter AI misrepresentation. AI is changing rapidly, and we are all adjusting. We share your hope that we can help students

understand the importance of mastering foundational information firsthand and learning how to think and communicate independently. We encourage you to talk about academic integrity with your students and be transparent about your concerns. A few ideas include:

- Explain your rationale about your AI policies and the "why" behind your learning goals.
- Provide examples of citation styles for your discipline.
- Brainstorm with students about when/how to use AI for particular assignments in your course.
- Engage your students in a conversation about the benefits and dangers of new technologies and encourage them to consider previous examples (such as the printing press or the Internet) for comparison. Help them to understand that new technologies inevitably shift the way that we do things and encourage them to reflect on their role in shaping the use of AI in ways that are ethical and equitable.

References

Baushard, S. (2024, May). AIs can now pull quotes from articles without hallucination.
    Education Disrupted: Teaching and Learning in An AI World. 19
    Stefanbauschard@substack.com

Careful use of AI detectors. (2024, May 7). Retrieved from https://cte.ku.edu/careful-use-ai-
    detectors.

Ceres, P. (2023, January 26). *ChatGPT is Coming for Classrooms. Don't
    Panic.* Wired. https://www.wired.com/story/chatgpt-is-coming-for-classrooms-dont-
    panic

Eaton, S. (2023). Postplagiarism: Transdisciplinary ethics and integrity in the age of artificial
    intelligence and neurotechnology. International Journal for Educational Integrity. 19, 23.
    https://doi.org/10.1007/s40979-023-00144-1

Edwards, B. (2023, July 14). *Why AI detectors think the US Constitution was written by AI.* Ars
    Technica. https://arstechnica.com/information-technology/2023/07/why-ai-detectors-
    think-the-us-constitution-was-written-by-ai

Fowler, G. A. (2023, April 14). We tested a new ChatGPT-detector for teachers. It flagged an
    innocent student. *The Washington
    Post.* https://www.washingtonpost.com/technology/2023/04/01/chatgpt-cheating-
    detection-turnitin

Grieve, R., Woodley, J., Hunt, S. E., & McKay, A. (2021). Student fears of oral presentations and
    public speaking in higher education: A qualitative survey. *Journal of Further and Higher
    Education, 45*(9), 1281-1293. https://doi.org/10.1080/0309877x.2021.1948509

Heikkila, Melissa. (2022). How to spot an AI-generated text. *MIT Technology Review.*
    https://www.technologyreview.com/2022/12/19/1065596/how-to-spot-ai-generated-
    text/

Jiang, Y, Hao, J., Fauss, M. and Li, C. (2024). Detecting ChatGPT-generated essays in a large-scale
    writing assessment: is there a bias against non-native speaker? Computers & Education.
    217. https://doi.org/10.1016/j.compedu.2024.105070.

Myers, A. (May 15, 2023). AI-Detectors biased against non-native English writers. Stanford
    University Human-Centered Artificial Intelligence. https://hai.stanford.edu/news/ai-
    detectors-biased-against-non-native-english-writers

Tai, J., Mahoney, P., Ajjawi, R., Bearman, M., Dargusch, J., Dracup, M., & Harris, L. (2022). How are examinations inclusive for students with disabilities in higher education? A sociomaterial analysis. *Assessment & Evaluation in Higher Education, 48*(3), 390–402. https://doi.org/10.1080/02602938.2022.2077910