

Position Statement on AI Detection

AI in Teaching & Learning Committee, Fall 2024

Introduction

Generative AI tools have become widely accessible. These tools have a variety of applications, and within the classroom, many instructors have found ways to incorporate these new technologies into their pedagogy in innovative ways. While, for a number of reasons (based on discipline, content area, and/or the knowledge, skills, and habits the student is learning in a particular class), many faculty encourage the use of AI, other faculty would prefer not to allow student use of generative-AI tools in some or all of their courses. In addition, all faculty, regardless of approach to Generative AI, desire that students accurately represent their work as their own and appropriately cite work they have sourced elsewhere. At many institutions of higher education, and at Boise State, many are interested in the potential for tools that purport to detect the use of Generative AI in the creation of written text.

The AI in Teaching & Learning Committee, after consulting with the Boise State Faculty Senate Steering Committee, currently recommends against the institutional adoption of AI detection tools, and against the use of AI detection tools by individual educators as a method for policing student use of AI.

AI Detection Software is unreliable

A growing body of literature points to the unreliability of AI detection software, the preponderance of both false negatives and false positives, and the difficulty of utilizing results from AI detection tools (Abdelaal, Gamage, & Mills, 2022, Akben, 2024, Elkhatat, Elsaid, & Saeed, 2023, Sadasivan et al., 2023, Thompson & Hsu, 2023, Walters, 2023). As an example, currently Turnitin's AI detector has a 98% confidence rate that text may be AI, with the resulting percentage likely it is AI having a margin of error of 15 (Turnitin, 2023). If Turnitin's AI detector detects 23% AI-generated content in a writing sample, it could contain anywhere from 8% - 38% AI-generated content. A 2024 position statement by the MLA-CCCC Joint Task Force on Writing and AI states that:

Because detection software is inherently flawed and can be easily fooled (Thompson and Hsu), colleges and universities must be mindful of the power that instructors and administrators relinquish when detection software is used uncritically to evaluate GAI within student writing. The fallibility of detection software, like *Turnitin*, creates a number of opportunities for vulnerable student identities to be profiled and accused without serious consideration of process, source evaluation and synthesis, or development of student reading and attribution practices (Coley; "How"). Even systems that are tested and trusted for accuracy can be evaded through software designed to mediate the detection of AI language and phrasing. Many students are already aware of how to run AI content through paraphrase or translation software.

As AI language models improve, they produce more human-like outputs, which makes AI-generated text harder to detect. This means that AI content detectors may be unable to identify text generated by the latest large language model versions. Additionally, electronic tools

such as [Quillbot](#) can rewrite AI-generated text, which makes detecting AI-generated text more difficult.

AI Detection Software may be biased against non-native speakers of English

Second-language writers are more likely than their peers to be accused of academic misconduct (Foltýnek et al., 2019, Tzanni, 2022). This is a pattern likely to be exacerbated by the finding that AI detectors are likelier to identify prose written by non-native speakers of English as AI generated (Liang et al. 2023, Weber-Wulff et al., 2023).

Additionally, use of AI detection tools may undermine pedagogically helpful uses of translation tools. The previously cited MLA/CCCC position statement asserts that:

Additionally, many second language and multilingual students already use such translation software to aid their adaptation to academic forms of writing. They may be flagged for using online translation tools that are essential to their acquisition of academic writing literacies.

There is a potential for AI Detection Software to undermine educator-student relationships

It's well-established in the K-12 and higher education literature that student-teacher relationships play an incredibly important role in student success (Mazer et al., 2014, Quinlan, 2016, Pekrun et al., 2006, Rocca, 2004, Snyder et al., 2002, Titsworth et al., 2010, Titsworth et al., 2013, Witt et al., 2004). Composition scholars have pointed out that AI detection tools can create a culture of distrust (Hubbard, 2023), something particularly important to avoid as generative AI is a new arena in which norms have not been set. If students believe their writing might be falsely identified as AI generated, this may cause them to distrust instructors. Quoting again the MLA-CCCC report: "All forms of surveillance place stress and pressure on individuals. AI technologies and language surveillance present a particularly complex situation. Placing students under a high-pressure surveilled writing situation can create negative feelings toward the writing process and undermine investment in transparency." [As recently reported in Bloomberg](#), students are feeling significant stress related to being reported for using generative AI in a punitive fashion.

In combination with the above points, we suggest that at the current time the potential harm of AI detections outweigh the potential benefit.

This doesn't mean that instructors should not pay attention to the issue of students using Generative AI when it is against course policies to do so. We encourage instructors to take advantage of the many resources available for assignment design that disincentivize shortcuts (whether those be the unethical use of AI, or traditional plagiarism), and encourage development of the whole student.

Recommendations

The cost to the institution of any of the available AI detections tools is significant. At this time, we suggest that there are more pressing uses for such funds (new AI-oriented positions, stipending faculty development, bringing in experts to campus, etc.). Given this point, and the issues addressed above, we do not find it to be the most appropriate use of resources. Instead, it

is recommended that instructors rely on developmental conversations and connections with their students to assess learning and potential academic misconduct (either from their own expertise, or guided by an institutionally available [Suggested Conversation Guide](#) using person-first language). This interpersonal skillset is free/institutionally available, and the most useful and developmental way for instructors to determine responsibility for student academic misconduct. Instructors have opportunities to set clear expectations and refine those through conversations with students, which will help students develop understandings of what plagiarism is, why it matters, and how generative AI can be acceptably used for learning. Many Boise State resources are available to help faculty adjust curriculum and assessment methodology, thus mitigating the need for AI-generated text detection:

- [Center for Teaching and Learning](#)
- [eCampus Center](#)
- [Academic Integrity Program](#)
- [AI in Education](#)

References and Further Reading

- Abdelaal, E., Gamage, S. & Mills, J. (2022). Assisting academics to identify computer generated writing. *European Journal of Engineering Education*. 47. 1-21.
10.1080/03043797.2022.2046709.
- Akben, M. (2024). Recent Advancements, AI Tools, and AI Detectors. AI Assessment Workshop.
- Coley, M. (2023, August). Guidance on AI Detection and Why We're Disabling Turnitin's AI Detector. Brightspace, Vanderbilt University.
www.vanderbilt.edu/brightspace/2023/08/16/guidance-on-ai-detection-and-why-were-disabling-turnitins-ai-detector/.
- Foltýnek, T., Meuschke, N., & Gipp, B. (2019). Academic Plagiarism Detection: A Systematic Literature Review. *ACM Computing Surveys* 52(6), pp. 1–42.
dl.acm.org/doi/pdf/10.1145/3345317.
- Henshall, W. (2023, August 2). 4 Charts That Show Why AI Progress Is Unlikely to Slow Down. *Time*. <https://time.com/6300942/ai-progress-charts/>
- Elkhatat, A. M., Elsaid, K., & Saeed Al-Meer. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, 19(1). <https://doi.org/10.1007/s40979-023-00140-5>
- Han, B., Nawaz, S., Buchanan, G., & McKay, D. (2023). Ethical and Pedagogical Impacts of AI in Education. *Lecture Notes in Computer Science*, 667–673.
https://doi.org/10.1007/978-3-031-36272-9_54
- Hubbard, J. (2023, June). The Pedagogical Dangers of AI Detectors for the Teaching of Writing. *Composition Studies Online*.
compstudiesjournal.com/2023/06/30/the-pedagogical-dangers-of-ai-detectors-for-the-teaching-of-writing/.

- How can educators respond to students presenting AI-generated content as their own? OpenAI Help Center. (n.d.). Help.openai.com.
<https://help.openai.com/en/articles/8313351-how-can-educators-respond-to-students-presenting-ai-generated-content-as-their-own>
- Knox, L. (2023, April 3). *Can Turnitin Cure Higher Ed's AI Fever?* Inside Higher Ed.
<https://www.insidehighered.com/news/2023/04/03/turnitins-solution-ai-cheating-raises-faculty-concerns>
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *ArXiv* (Cornell University).
<https://doi.org/10.48550/arxiv.2304.02819>
- Mazer, J. P., McKenna-Buchanan, T. P., Quinlan, M. M., & Titsworth, S. (2014). The dark side of emotion in the classroom: Emotional processes as mediators of teacher communication behaviors and student negative emotions. *Communication Education*, 63(3), 149–168.
- MLA-CCCC Joint Task Force on Writing and AI, “Generative AI and Policy Development: Guidance from the MLA-CCCC Task Force,” April 2024,
<https://aiandwriting.hcommons.org/working-paper-2/>
- Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology*, 101(1), 115–135. <https://doi.org/10.1037/a0013383>.
- Quinlan, K. (2016). How emotion matters in four key relationships in teaching and learning in higher education. *College Teaching*, 64(3), 101–111.
- Rocca, K. A. (2004). College student attendance: Impact of instructor immediacy and verbal aggression. *Communication Education*, 53(2), 185–195.
- Sadasivan, Vinu Sankar, Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-Generated Text be Reliably Detected? *ArXiv*, 2303.11156v3.
<https://doi.org/10.48550/arxiv.2303.11156>
- Snyder, C. R., Shorey, H. S., Cheavens, J., Pulvers, K. M., Adams, V. H., & Wiklund, C. (2002). Hope and academic success in college. *Journal of Educational Psychology*, 94(4), 820–826.
<https://doi.org/10.1037/0022-0663.94.4.820>.
- Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, 18(1). <https://doi.org/10.1186/s41239-021-00292-9>
- Steere, E. (2023, October 18). The Trouble With AI Writing Detection. Inside HigheEd.
<https://www.insidehighered.com/opinion/career-advice/teaching/2023/10/18/faculty-should-know-tools-students-use-beat-ai-detection>
- Thompson, S. A. & Hsu, T. (2023, June 28). How Easy Is It to Fool A.I.-Detection Tools? The New York Times, 28 June 2023.
<https://www.nytimes.com/interactive/2023/06/28/technology/ai-detection-midjourney-stable-diffusion-dalle.html>

- Titworth, S., McKenna, T. P., Mazer, J. P., & Quinlan, M. M. (2013). The bright side of emotion in the classroom: Do teachers' behaviors predict students' enjoyment, hope, and pride? *Communication Education*, 62(2), 191–209.
- Titworth, S., Quinlan, M. M., & Mazer, J. P. (2010). Emotion in teaching and learning: development and validation of the classroom emotions scale. *Communication Education*, 59(4), 431–452.
- Tzanni, P. (2022). Discrimination and Native-Speakerism in English for Academic Purposes (EAP). *Studies in Technology Enhanced Learning* 2(3).
<https://doi.org/10.21428/8c225f6e.755eca93>.
- Turnitin. (2023). FAQs for Turnitin's AI writing detection capabilities. University of Kansas Center for Teaching Excellence.
[https://cte.ku.edu/sites/cte/files/images/2023/AI%20Writing%20FAQs%20March%202023%20\(1\).pdf](https://cte.ku.edu/sites/cte/files/images/2023/AI%20Writing%20FAQs%20March%202023%20(1).pdf)
- Walters, W. H. (2023). The Effectiveness of Software Designed to Detect AI-Generated Writing: A Comparison of 16 AI Text Detectors. *Open Information Science*, 7(1).
<https://doi.org/10.1515/opis-2022-0158>
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023) Testing of Detection Tools for AI-Generated Text. *International Journal for Education Integrity* 19(26).
<https://doi.org/10.1007/s40979-023-00146-z>.
- Witt, P. L., Wheelless, L. R., & Allen, M. (2004). A meta-analytical review of the relationship between teacher immediacy and student learning. *Communication Monographs*, 71(2), 184–207.
- Wood, C. (2024, September 9). AI Detectors are Easily Fooled, Researchers Find. Edscoop.
https://edscoop.com/ai-detectors-are-easily-fooled-researchers-find/?mkt_tok=MTgwLUxTVi02NzIAAAGVn-IWNrJ-lq8yT5YAZKa1zgKzhDtx5l3lZfpLDHjKL__fNkothmFCbhRcI82SO_1yAnCg8ttSGb595uiN9IhAMPnHJ9Sd9QlpzqP4jvZV_8L9nrH8gQ