The Center for Learning & Teaching Excellence
at ANDERSON UNIVERSITY

Guidance on Generative AI Content Detection and Alternative Approaches

This guide covers some basic info., strengths, and weakness for several commonly used AI detection software solutions: Turnitin, Quillbot, GPT Zero, Copyleaks, and, Winston AI, as well as, and perhaps more important, alternatives to relying on detection software.

A common theme in the weaknesses and accuracy metrics of tools is the reality that no detection software is 100% effective, and each *will* invariably produce both false positive and false negative results. Therefore, if you suspect unauthorized and unacknowledged/uncited use of AI in course work that students submitted as original work, AI content reports generated by detection software cannot be your *only* evidence for making an accusation of a violation of academic integrity. Use of such tools can be a *part* of your verification process, but they should not be the end result.

Other considerations and alternatives to using detection software tools can be found at the end of this document. (Arguable, this may be the most important information in this document.)

Basic Considerations:

1. Reliability is questionably for both free and paid versions any detection tool. No detector is 100% reliable. Even the best detectors have limitations and can be circumvented.

2. Turn-it-in is not as effective for shorter writing samples. Quillbot it not as effective for longer, academic writing samples.

3. False positives are a serious concern and potentially damaging to reputations, trust, and relationships.

4. Detection reports are biased against non-native English speakers, which is especially problematic for AU's international students and students in the U-Hub Degree programs.

5. Free versions are limited in how many words or characters you can submit monthly. The full version of Turnitin is available to all AU faculty (but again, reliability is inconsistent).

6. Use these tools as part of a broader, complementary strategy that includes revised assignment designs, scaffolded assessments, and discussions about AI and integrity. See page 5.

7. Be transparency with students about detection methods and policies that can help maintain trust and academic integrity.

8. Both AI writing and detection tools are rapidly evolving, requiring ongoing evaluation of effectiveness. Note the revision date of this document in the footnote.

Turnitin

1. Tools & Information Provided:
    a. Confidence scores for AI content: A percentage of AI-generated only by an LLM, and a percentage of AI-generated by an LLM and then likely revised by an additional LLM.
    a. To reduce the occurrence of false-positives, it will not give a percentage score if it falls below 20% of content is AI generated.
    b. Integration with existing plagiarism reports - Produces a separate similarity report to check for plagiarism.
2. Best Suited For:
    a. Formal academic papers and essays
    b. Institutions already using Turnitin for plagiarism
3. Strengths:
    a. Combines AI detection with plagiarism checking
    b. Familiar to many faculty and students
    c. Integrated into many existing LMS platforms
4. Weaknesses:
    a. Subscription costs can be significant
    b. Only works with well with long-form prose. Must be at least 30 words. Code, poetry, tables, bulleted lists, etc. will not be assessed.
    c. May have higher false positive rates than some competitors
    d. Less transparent about detection methodology
5. Accuracy Metrics:
    a. Limited information is available at this time
    b. Generally considered reliable but with higher false positive rates than some competitors
6. Pricing:
    a. Institutional subscription model

Quillbot

1. Tools & Information Provided:
    a. Basic AI detection scoring
    b. Limited detailed analysis
    c. AI detection is one of ten tools for students:
        1. Paraphraser/Paraphrasing Tool (uses AI to rewrite and rephrase sentences while maintaining meaning)
        2. Grammar Checker
        3. AI Detector
        4. Plagiarism Checker
        5. AI Humanizer (designed to make AI-generated text appear more human-written)
        6. AI Chat (an interactive AI assistant for writing help)
        7. Translator (for converting text between languages)
        8. Summarizer
        9. Citation Generator
        10. "QuillBot Flow" (an "AI writing assistant")
2. Best Suited For:
    a. Easy user interface.
    b. 4 languages (English, Spanish, German and French)
    c. Shorter, less technical content
    d. Quick preliminary checks
3. Strengths:
    a. Integrated with other writing tools

    b. User-friendly interface
4. Weaknesses:
    a. Basic AI detection only – might only handle short essays well
    b. Less accurate with technical/specialized content, lengthier academic paper with citations
    c. Works best with English text - May flag non-English text as "cannot determine"
5. Accuracy Metrics:
    a. Limited information is available at this time
    b. May struggle with non-English content
6. Pricing:
    a. Limited free version (AI Detector for 1,200 words per month)
    b. Premium $8.33/month
    c. Team plans available

GPTZero

1. Tools & Information Provided:
    a. Sentence-level analysis
    b. Scans of submissions produce a confidence score, a probability score, a categorization of whether content was likely to be AI-generated, human-generated, or both.
    c. Detailed perplexity and burstiness scores
    (*Burstiness* refers to the variation in sentence structure, length, and complexity, indicating how varied the writing pattern is. Human writers tend to be "bursty," varing between simple and complex sentences. In contrast AI LLMs (especially earlier versions) tend to lack variety in sentence structure throughout the text produced.
    Perplexity measures how predictable or unpredictable the text is. Humans tend to write in less predictable ways. Our writing is more often more surprising or confusing, often using unique phrases, making unexpected, awkward, or not quite correct word choices, and it may incorporate personal insights. AI-generated text tends to follow more predictable patterns.)
2. Best Suited For:
    a. Mixed content where students might blend AI and human writing
    b. Longer academic papers and essays
    c. Multi-language detection needs
3. Strengths:
    a. AI detection trained on academic writing
    b. Uses advanced metrics of perplexity and burstiness for detection
    c. Particularly strong at detecting mixed content (96.5% accuracy with mixed AI/human documents)
    d. Recent benchmarks show 98.6% accuracy with 0% false positive rate against ChatGPT o1
    e. Batch processing capabilities for educators
    f. Sentence-by-sentence analysis for granular feedback
4. Weaknesses:
    a. Requires minimum input of 60 words
    b. Limited functionality in free version
    c. Maximum scan of 5,000 words may limit use for longer academic papers
5. Accuracy Metrics:
    a. 99% accuracy rate with only 1% false positive rate (company claim)
    b. Independent benchmarks show 98.6% accuracy with 0% false positives against ChatGPT o1
    c. Promotes itself as a tool to highlight the *possible* use of AI in writing. It cannot be considered 100% reliable. It *will* produce both false positive and false negative results.
6. Pricing:
    a. Basic AI detection is free (10K words/month)
    b. Educational subscriptions available for advanced features

Copyleaks

1. Tools & Information Provided:
    a. AI-generated probability scores
    b. Highlights portions of text likely written by AI
    c. Combines transformer-based classifiers and statistical analysis (Transformer-based classifier models have superior pattern recognition, “understand” how words relate to each other in context, can be fine-tuned to detect content from specific AI models, and can work across different languages. However, as AI writing improves, detection becomes more difficult, and they do also generate false positives and false negative results.)
2. Best Suited For:
    a. Technical writing samples
    b. Multi-language detection needs
3. Strengths:
    a. Multi-language support
    b. Chrome extension available
4. Weaknesses:
    a. Recent benchmarks show 89.1% accuracy with 5% false positive rate
    b. Some inconsistency in detection results across different tests
5. Accuracy Metrics:
    a. 89.1% accuracy with 83.3% recall and 5% false positive rate against ChatGPT o1
    b. Some independent tests show inconsistent results
6. Pricing:
    a. Subscription-based model
    b. App version and Chrome extension available

Winston AI

1. Tools & Information Provided:
    a. AI detection likelihood scores
    b. Character-limited analysis in free version
2. Best Suited For:
    a. Individual faculty needs
    b. Quick checks of shorter submissions
3. Strengths:
    a. User-friendly interface
    b. Chrome extension available
    c. Strong accuracy in some independent tests
4. Weaknesses:
    a. Less established than some competitors
    b. Limited information on methodology
5. Accuracy Metrics:
    a. *Some* independent tests show 100% accuracy for both AI-generated and human-written text
    b. Limited actual information on false positive/negative rates
6. Pricing:
    a. Free monthly plan with 1,800-character limit
    b. Paid plans starting at $18/month (80k credits) with annual subscription
    c. Up to $32/month for higher tiers

AI detectors are not fully reliable, so what are the recommended best practices?

1. Establish Clear Expectations.
Unclear expectations about assignment products are among the most common student complaints. As use of AI is still relatively new, with a *wide* variety of expectations for us among faculty, you must set clear expectations for when, how much, if, when AI can be used. It is highly recommended to not limit your thinking about student use of AI only to how it can be used to cheat and circumvent work. However, when it comes to students violating academic integrity expectations, one of the surest ways to encourage cheating it to never talk about cheating, integrity, or your expectations. When discussing expectations for AI use, it is highly recommended to refer to the Generative AI Use Scale found in Thrift Library's AI@AU Resource Hub.

2. Be Transparent.
Similarly, if you will use AI detection, let students know. But make it clear that it will not be the only part of your process, if you have concerns about work submitted. Normalize having conversations with students about how the created their work, what they learned from it, and whether their use of AI helped them achieve the learning objectives, or just bypass them, allowing them to create a product without understanding the material they submitted.

3. Set a Baseline.
Compare work to an early writing sample, submitted in class, written by hand or without access to internet. Perhaps consider such activities as being similar to a pre-test.

4. Vary Your Assessments.
Include several types of assignments. How else can you assess learning? It is already a best practice to measure learning in several ways. The old-school mid-term exam/term paper/final exam model has long been somewhat obsolete, as it usually doesn't focused on active learning, does not provide enough opportunities to demonstrate learning, and does not offer feedback on ways for students to improve. Some faculty are again relying on using more bluebooks or oral presentations. But be mindful about what you are actually assessing. An in-class assignment or quiz written by hand is not a good measure of student writing ability, but it may be a good way to measure student's understanding of terminology or basic content, for example. An oral presentation might be a good way to assess communication skills, but may not be a good way to assess the ability to generate original content or a students' grasp on material.

5. Do the Work of Re-working Assignments.
Create more localized assignments – use very specific and unique examples in case studies or assignment prompts.
Consider alternatives to essay assignments. Several alterative are discussed in the post below, including performance-based assessments, a portfolio or writing journal, project-based assessments, observation-based assessments, and visual essays:
https://leonfurze.com/2023/10/04/rethinking-assessment-for-generative-ai-beyond-the-essay/
or various approaches to discussion-based and oral assessment:
https://leonfurze.com/2023/09/27/rethinking-assessment-for-generative-ai-orals-and-discussions/

6. Focus on Process First, not Product.
Focus your teaching and your students' work on the process first, rather than the product first. Scaffold assignments in stages and provide iterative feedback along the way. Overly emphasizing a to be product being turned in, without any feedback along the way, runs the risk of turning learning into a transactional endeavor, rather than a relational, developing process. If the emphasis is only on the product, they are more likely to rely on whatever quick and convenient tools are available – whether they learn from the experience or not. But if

students can learn to trust the process, and you can provide the guidance needed, the results they want will follow.

6. Incorporate Reflections on Learning .

Incorporate assignments or activities that ask students to reflect on both the draft iterations or their work and on their final results. Like many items on this list, reflecting on learning is a best practice that reinforces learning, regardless of discussions about AI or academic integrity. When allowing or asking student to use AI, ask them to reflect on how they used it, whether it was useful/accurate/helpful, and what they learned from using it – about both the content and about the advantages and limitations of AI. Such reflections could be class discussions, think-pair-share activities, structured online discussion boards, one-minute papers in class, or a stand-alone short writing assignment.

7. Consider Version Tracking.

Use Google Docs and review the version history. Again, be transparent: Let students know in advance that you will be looking at version history. Your goal is to guide them through the learning/writing/creative process from the beginning, not to just set up a situation where you can easily "bust" them after the fact.

8. Become Familiar with the "tells."

Check for common flags – hallucinations (invented) information, common phrases, unique synonyms, etc. Check the validity of references, as they may often be fabricated. Currently, frequent occurrences of em-dashes (–) are also common indicator of AI content. Check citations for accuracy. AI will sometimes hallucinate fictional sources as it attempts to fulfill a prompt.

9. *Talk* to your Students.

Always discuss any suspicions with students, before submitting a report. In fact, students are required to complete section 3 (or at least have the option to complete it) before an Infraction Report can be submitted. Confrontations are never pleasant for faculty or students. But often students will share once you begin talking with them. Discussions about AI can still come down to "your word against mine," of course. Some things have not changed.

For more info on concerns with over-relying on AI detection, see also

AI Detectors - Part 1: Accuracy, Deterrence, and Fairness in Practice
AI Detection in Education is a Dead End