CENTER FOR FACULTY DEVELOPMENT | UNIVERSITY OF SOUTHERN MISSISSIPPI

What Turnitin Can and Cannot Do

Key Takeaways from the CFD Turnitin Workshop

Our Turnitin workshop gave faculty a chance to look carefully at something that rarely comes through in policy conversations, which is what AI detection tools actually do at a technical level and, just as importantly, where the boundaries of that work are. Turnitin's AI detection model analyzes patterns in submitted text to generate a probability estimate, meaning a statistical likelihood that portions of a document were produced by a generative AI system, and that output is not a determination of authorship, not a finding of intent, and not evidence that can stand alone in an academic integrity decision. One of the most practically significant things Turnitin shared directly with us is a boundary built into the model itself: the AI detection feature is not designed to assess submissions shorter than 300 words or longer than 30,000 words, and applying the tool outside of that range produces output that is not reliable and should not be used to inform an integrity review. Discussion posts, short reflections, and papers under 300 words are not within the model's operating window, and neither are large research papers and dissertation-length documents above 30,000 words. Knowing where the tool is designed to work is the most basic condition for using it well.

There is a second finding from the workshop that deserves its own moment, because the procedural risk that detection creates does not land equally across the student population. Research on widely used detection systems has found that these tools misclassify a disproportionate share of essays written by non-native English speakers as AI-generated, and the reason that happens is worth understanding, because it is not a flaw in Turnitin specifically so much as a consequence of what the tool is actually doing, which is looking for the same kind of grammatical regularity, syntactic restraint, and vocabulary patterns that show up both in AI-generated text and in writing produced by someone still developing fluency in a second or third language. What that produces is a situation where a student can turn in work that is entirely their own and still end up in a conversation they were not expecting, and the students most likely to find themselves there, international students, multilingual writers, anyone who uses grammar or editing tools to strengthen their prose, are also the students who typically have the least institutional support for navigating what comes next.

What all of this points toward is rethinking how we think about this process, which is that a detection score is a starting point and not an answer, and faculty who approach it that way are going to be in a stronger position every time, both because the integrity process holds up better when it rests on actual evidence and because students get a fairer shot at a review that examines the situation rather than just the flag. That means looking at draft history, prior submissions, the nature of the assignment, and the context that a number on a screen cannot carry on its own. Making sure your syllabus clearly names which of USM's three AI use options applies to your course is the foundation that makes all of that possible, because documented expectations are what give a review its institutional legs, and that structure is already in place and already designed to support you through exactly this kind of concern.

BEFORE YOU SUBMIT ANYTHING TO ACADEMIC INTEGRITY

A checklist for faculty

☐ **Policy on file:** Is your AI use option (Encouraged, Permissible, or Human-Generated Only) documented in your syllabus for this course?

☐ **Word count:** Is the submission between 300 and 30,000 words? Turnitin's detection model is not designed to assess submissions outside that range and the output will not be reliable.

☐ **Equity check:** Is this student population at elevated risk for a false positive? International students, multilingual writers, and students who use grammar tools are disproportionately flagged.

☐ **Score as flag:** Are you treating the detection score as the start of an inquiry, not a conclusion? A score is a reason to look closer, not a finding of fact.

☐ **Corroborating evidence:** Do you have supporting context beyond the score, such as draft history, prior submissions, or assignment-specific factors?

Read the full paper: *Beyond Detection: AI Literacy, Academic Integrity, and the Scholars We Actually Want to Graduate*