



Addendum to Evaluating the Effectiveness of Turnitin's AI Writing Indicator Model

In addition to our assessment of Turnitin's AI detector capability, we also investigated additional AI detector tools: ZeroGPT, Copyleaks AI Content Detector, Crossplag AI Content Detector, and Content at Scale's AI Detector. Using the same fully human, fully AI, and hybrid artifacts that we processed with Turnitin's engine, we tested these additional tools. These detectors were seemingly less reliable in their ability to properly discern the AI, human, and hybrid writing samples. For example, not one of these four detectors correctly identified all the human artifacts as, in fact, being human written; one detector did not identify any of the human generated writing samples as being fully written by humans. Similar issues occurred with the fully AI generated texts, with not one of the detectors identifying all the artifacts correctly, and some identifying none accurately. The hybrid artifacts (text written/generated by both human and AI) exhibited similar problems as those observed with Turnitin. For a discussion on the issues concerning hybrid artifacts, please see our Turnitin report: [REPORT: Evaluating the Effectiveness of Turnitin's AI Writing Indicator Model | Center for the Advancement of Teaching](#). In addition, the narrative outputs of these four detectors were more confusing than Turnitin's. In summary, in our test of these AI detectors, they lack the ability to reliably discriminate between what is human written and what is not. We recommend proactive measures to reduce academic dishonesty rather than reactive measures that rely on AI detectors. To note, we ran the above narrative (fully human written) through an additional AI detector (GPTZero) just to see its results; it reported a 48% probability the text was entirely written by AI.