

# Legal Technology

edited by  
VICTOR LI  
victor.li@americanbar.org



16

## ARTIFICIAL INTELLIGENCE

# False Positive

AI detection tools might not be the answer just yet

BY DANIELLE BRAFF

**A**rtificial intelligence detection tools purport to be able to act as digital gatekeepers in law classrooms, courtrooms and beyond. But there's a glitch—liter-

ally. Though vendors say their detectors accurately determine when someone has used AI, real-world use tells a different story.

Antony Haynes, partner and head of the cybersecurity, data privacy & AI practice group at Dorf Nelson & Zauderer, recently asked ChatGPT which of his 500-word excerpts from his pre-pandemic writing was written by AI. “For each of the seven excerpts, ChatGPT said there was at least a 30% chance that they were written by AI,” Haynes says. “And for one of these excerpts, ChatGPT said there was a 60% to 80% likelihood that this text of mine, written before the release of ChatGPT, was written by AI.”

AI-generated text detectors are used to detect suspected academic dishonesty in schools across the country. What happens when the software gets it wrong?

Legal experts say we're entering murky constitutional territory: Students may be disciplined without meaningful evidence, and proof of misconduct comes from tools that can't even explain their own decisions.

A study published this year by University of Maryland researchers found that text written by humans was flagged incorrectly as AI-generated by the AI detector Giant Language Model Test Room about 7% of the time. And in 2023, AI detection tool Turnitin

acknowledged its software had a sentence-level false-positive rate of 4%.

As a result, Turnitin advises against using its detector to accuse students of AI plagiarism.

“Our AI writing detection model may not always be accurate (it may misidentify human-written, AI-generated and AI-paraphrased text) so it should not be used as the sole basis for adverse actions against a student,” the company says on its website. “It takes further scrutiny and human judgment in conjunction with an organization’s application of its specific academic policies to determine whether any academic misconduct has occurred.”

Students are fighting back. In April, a University at Buffalo senior was accused of using AI to cheat on three class assignments based on findings by the Turnitin AI detection tool. She was able to prove to the school that she did not cheat; then she launched an online petition calling for it to end its use of Turnitin. At press time, the petition had received more than 1,000 signatures.

“To ensure fairness, the university does not rely solely on AI-detection software when adjudicating cases of alleged academic dishonesty,” the University at Buffalo said in a statement provided to the *ABA Journal* in August. “To reach our standard of preponderance of evidence, instructors must have additional evidence that a student used AI in an unauthorized way.”

### Flawed framework

AI detectors are largely unreliable because they rely on probabilistic language models that attempt to identify patterns consistent with machine-generated writing, says Keith Altman, founder and lead attorney at K Altman Law in Michigan, a law firm that has represented students accused of improperly using generative AI tools like ChatGPT.

The tools are far from reliable, Altman says, explaining that they often flag legitimate student work as AI-assisted simply because it exhibits formal grammar, structured sentences or predictive phrasing. This especially impacts students whose first language

isn’t English and those who write with the help of tutors or learning centers.

Students are being penalized based solely on detector outputs, with some schools requiring them to submit drafts or version histories to avoid failing.

“We’ve also seen cases where professors rely too heavily on these flawed tools, leading to accusations that collapse under scrutiny,” Altman says. “Unfortunately, these tools lack transparency and consistency, making them inappropriate as stand-alone evidence.”

Often, says Jason Ostendorf, a consumer appellate attorney in Hunt Valley, Maryland, the more likely these flawed tools are to flag it as AI-generated.” Ostendorf compares the enthusiasm surrounding AI detectors to “an academic witch hunt justified by algorithms.”

Many autistic students are being targeted, as AI uses patterns—and these students have been taught to write in a straightforward manner, triggering the detectors, Altman says.

“This is a systemic issue that is escalating quickly,” he says.

Meanwhile, 90% of college students report using ChatGPT for assignments, according to a 2023 survey from Study.com. A separate study from the American Association of Colleges and Universities and Elon University found that 56% of university leaders say their institutions are not prepared to use generative AI tools to prepare students for the future. The result: a disconnect, leading to widespread confusion, unfair penalties and a lack of consistent enforcement, Altman says.

Even educators are using AI to create tests and classroom materials (one student at Northeastern University demanded a tuition refund for a class after learning its professor used ChatGPT to create course materials; she did not receive a refund, the *New York Times* reported), blurring the line of acceptable use and underscoring the need for institutions to set clear policies.

Luci Vágnerová, a Brooklyn, New York-based education adviser and academic difficulty consultant with BKT Education, worked with a student

at a highly selective law school in the United States who was accused of using AI when writing an appellate brief. Fortunately, Vágnerová says, the student had a large amount of documentation along with instructor feedback for her brief, which took her eight weeks to write. Vágnerová says the case against the student was dismissed in its entirety.

For reasons like this, Fort Lauderdale, Florida-based personal injury and product liability lawyer Jason Turchin recommends students keep clear records of their search history along with any electronic or hard copy paper trail of their work and edits. By demonstrating the work was not AI-created, students can increase their chances of vindication.

If they’re accused of using AI for their work, students should immediately request the evidence and methodology used to make the accusation, Ostendorf says. Those who are challenged based on the word of an AI detector should confront the school, as these tools are fallible and lack transparency, he says. Students should also consider seeking legal counsel to evaluate whether their 14th Amendment due process rights were violated, especially if they attend a public university or law school.

Ostendorf says attorneys can help by attacking the reliability and admissibility of the AI detection tools, raising procedural defenses, due process violations and in some cases, defamation or contract-based claims against schools. But, he says, there’s an even bigger issue.

“The real question is, ‘Why are we punishing students for using AI in the first place?’” he asks, arguing that instead of waging a losing war against progress, law schools and universities should embrace AI use as long as students grasp the topic. The ABA adopted a resolution during its annual meeting in August urging law schools to adopt clear generative AI policies while understanding that students will have to use these tools when they enter the legal profession. “AI is now woven into everyday life. Any attorney not using AI is already behind in terms of efficiency and productivity.” ■