

# Limitations of AI Detection Algorithms

AI detection tools are increasingly used to identify potential use of large language models (LLMs) in student writing. However, these tools come with significant limitations that educators should understand before relying on them for academic integrity decisions.

---

## Key Limitations of AI Detection Tools

### False Positives

Human-written content—especially when supported by tools like Grammarly or writing centers—can be misclassified as AI-generated. This can unfairly penalize students who have not used AI tools.

### Rapid Evolution of AI

LLMs are advancing quickly, producing text that closely mimics human writing. Detection tools often lag behind these developments, reducing their reliability.

### Bias and Fairness Concerns

Detection algorithms may misinterpret writing styles of non-native English speakers or students from diverse linguistic backgrounds, leading to inaccurate assessments.

### Over-Reliance on Algorithms

Excessive dependence on detection scores can shift the focus from fostering learning to policing students, potentially eroding trust between faculty and students.

### Legal and Reputational Risks

False accusations based on algorithmic scores can harm students' academic records and reputations, and may lead to legal disputes.

---

# Understanding the Nature of AI Detection Scores

AI detection tools do not provide proof of misconduct. They offer probabilistic assessments—essentially one algorithm estimating the likelihood that another algorithm was used. These scores are not definitive and should not be used as the sole basis for disciplinary action.

## “Turnitin Statement:

“Our AI writing detection model may not always be accurate (it may misidentify both human and AI-generated text), so it should not be used as the sole basis for adverse actions against a student.”

### Sources:

- [Washington Post: ChatGPT cheating detection](#)
- [MIT Technology Review: AI detection tools are easy to fool](#)
- [Turnitin: AI Detection FAQ](#)
- [Turnitin Help Center: AI Writing Detection](#)

## Recommended Best Practices

### 1. Use Detection Tools as Indicators, Not Proof

AI scores should prompt closer review of a student’s work, not automatic penalties. Consider changes in writing style, tone, or clarity, and use your professional judgment.

### 2. Encourage Ethical Use of AI

Use flagged submissions as opportunities to discuss responsible AI use. Feedback should guide students toward improving their writing and understanding the ethical boundaries of AI assistance.

### 3. Adapt Writing Assignments

Revise prompts to require personal reflection, course-specific content, or unique perspectives. Include process-oriented tasks like outlines, drafts, or revision notes.

### Resources:

- [Critical AI: Adapting College Writing](#)
- [Mark Marino: GPT Writing Assignments](#)

## 4. Design Rubrics That Emphasize Human-Centric Writing Skills

Include criteria that LLMs typically struggle with, such as:

- Nuanced argumentation
- Integration of course-specific concepts
- Original insights or personal experiences

This allows instructors to fairly penalize weak writing while encouraging improvement, regardless of the student's writing process.

## 5. Use Document History and Quiz Logs for Contextual Evidence

Instead of relying solely on AI detection scores, consider using platform features that provide insight into the student's writing process:

- **Google Docs / Microsoft 365:** Ask students to submit document links to review version history and editing timelines.
- **Canvas Quiz Logs:** Enable quiz logs for Classic Quizzes to view time spent per question. Note that logs are not definitive proof and should be interpreted cautiously.

### Resources:

- [Canvas: Enable Feature Options](#)
- [Canvas: View Quiz Logs](#)
- [Google: Find what's changed in a file](#)
- [Microsoft: View Previous Versions](#)

## 6. Conduct Oral Examinations or Interviews

When AI use is suspected, consider asking students to:

- Explain key concepts from their work
- Describe their writing or research process
- Elaborate on specific claims or sources

This can help verify authorship and promote deeper learning.

## 7. Compare with Other In-Class and Canvas-Submitted Writing Samples

Maintaining a portfolio of student writing completed in class can help identify inconsistencies in style, tone, or complexity. If you have students submit all their writing digitally, you can also leverage GPTZero—integrated into Canvas via K16's Scaffold AI-Detection system—to compare writing samples across all submissions, including quizzes, assignments, and discussions.

This system works best when students consistently submit their work through Canvas, allowing instructors to identify shifts in writing style or voice over time. For in-class writing to be included in these comparisons, it must also be submitted digitally (e.g., uploaded as a file or entered into a Canvas assignment or discussion).

## 8. Require Student Reflection

Ask students to submit short reflections with assignments that explain:

- Their approach to the task
- How they selected sources
- What challenges they faced

These reflections promote metacognition and make AI misuse more difficult. When submitted through Canvas, reflections can also be analyzed by K16's Scaffold AI-Detection system alongside other writing samples, helping identify inconsistencies or shifts in writing style that may warrant further review.

## 9. Use Social Annotation and Concept Mapping

Encourage students to engage with texts collaboratively using tools like Hypothesis or concept maps. These activities foster critical thinking and are difficult to replicate with AI tools.

## 10. Scaffold Assignments

Break large writing tasks into smaller steps:

- Topic proposal
- Annotated bibliography
- Draft outline
- Peer review
- Final submission

This makes it easier to track student progress and harder to substitute AI-generated work.