Copyleaks

AI DETECTION · QUICK REFERENCE

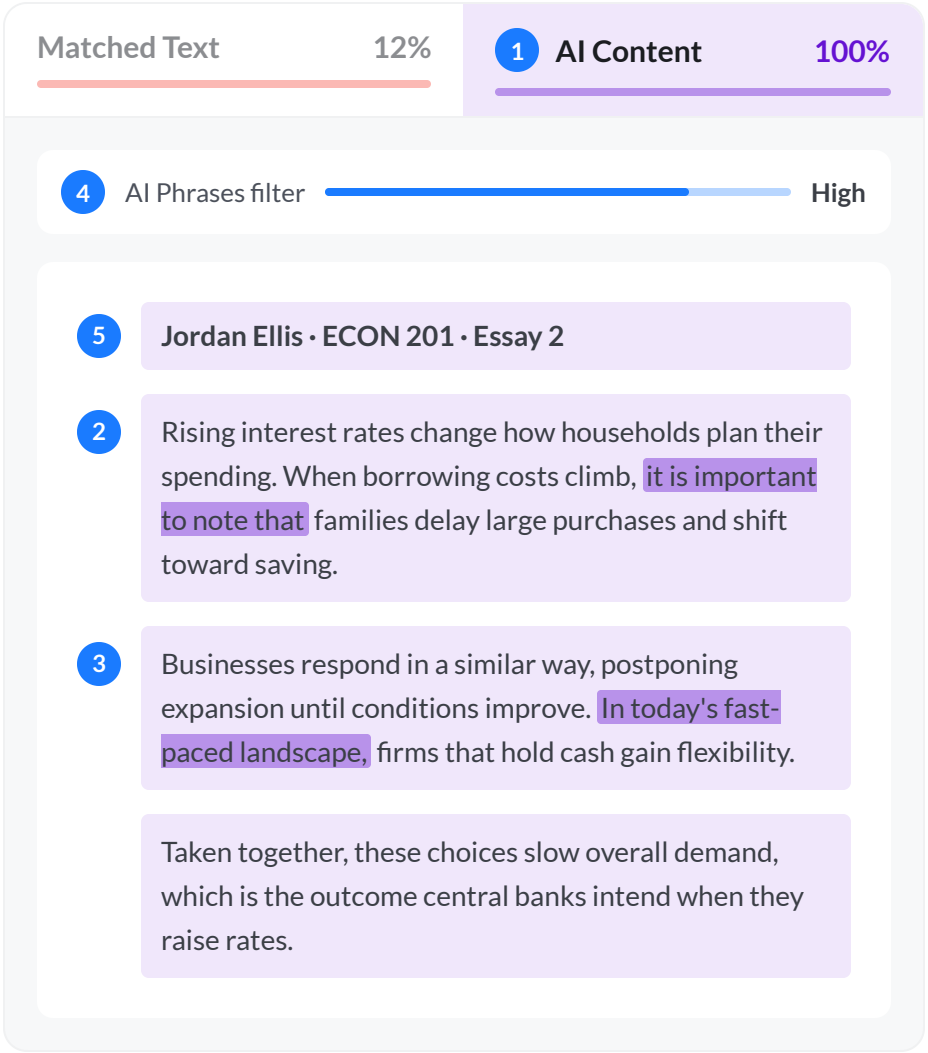
Reading an AI detection result

Plain answers to the top 10 questions users and instructors ask most about our AI scan reports.
An AI Content score should always be one data point in a broader review, never a verdict on its own.

START HERE

How to read the report: score and highlights are different things

Most disputes start here. The score counts the AI-classified text blocks. The highlighting inside those blocks is supporting evidence. A report can show 100% even when only a few phrases are dark purple.



1. **The AI Content score is the share of the text classified as AI.** It is not a confidence level. 50% means half the text was classified as AI. Whole blocks are counted, so scores often land on round numbers. Matched Text is a separate plagiarism score.
2. **Light purple is an AI-classified block.** These blocks make up the score. A block with enough AI patterns counts in full.
3. **Dark purple is AI Phrases.** These are supporting evidence and are not part of the score. They only appear in English, Spanish, Italian, German, French and Portuguese.
4. **The AI Phrases filter hides lower-confidence phrases.** It never changes the score or the light purple. That's why a filtered report can look almost empty and still show 100%.
5. **Names, headers and references are highlighted when they sit inside an AI block.** That doesn't mean those words were AI-written.

01 ACCURACY

How accurate is Copyleaks, and can you prove it?

99.7+% accurate, with a 0.026% false-positive rate at the default setting. That is about 1 in 4,000 human-written texts.

Independently tested. Independent academic studies have ranked Copyleaks the most accurate AI detector. The full method and studies are on the testing methodology page.

Covers the major AI models. The model is continuously updated and QA-validated against major LLMs, including OpenAI, Claude, Gemini and Llama models.

You choose the trade-off. Three sensitivity levels balance the risk of flagging human writing against the risk of missing AI.

AI writing assistants get flagged. If you use Grammarly or any AI writing assistant, your entry will be flagged as AI.

Sensitivity	Best for
1 · Cautious	Prioritizes human writing - flags only high-confidence AI content.
2 · Balanced (default)	Balances coverage and accuracy - flags likely AI content while protecting original writing.
3 · Maximum	Prioritizes coverage - identifies more AI content, with a higher chance of flagging original writing.

WHAT TO DO

- Treat the score as one signal or indicator in a review, not a verdict, to start a conversation with your student about academic integrity.
- Share copyleaks.com/ai-content-detector/testing-methodology with anyone who wants the evidence.

02 FALSE POSITIVE

"I wrote this myself." Why does it say 100% AI?

False positives are rare but possible, and they usually have a known cause.

Patterns, not identity. Detection looks at statistical predictability, sentence variety and structure. It can't tell who typed the text.

Generative writing tools. The rewrite, tone and sentence-completion features in AI writing assistants such as Grammarly produce AI text, and so do translators like Google Translate, humanizers, text scramblers and paraphrasers such as QuillBot. Running a human draft through these tools adds AI-like patterns. Basic spelling and grammar fixes don't.

Very short text. Under about 200 words, there isn't enough text to analyze reliably.

Structure and formatting. Numbering, bullets and outline-style headers carry more weight in short texts.

Highly formulaic writing. Heavily templated, repetitive or rigid writing can occasionally share patterns with AI output.

INSTRUCTOR

- Check that the assignment's template or prompt text is excluded. Excluded template text doesn't count toward the AI score (see card 06).
- Compare the paper with the student's earlier work.
- Ask which tools the student used.
- Have the student talk through the paper or summarize it on the spot.

STUDENT

- Check the AI Phrases heat map. Sparse highlighting supports your case.
- Gather drafts, version history, notes and outlines.
- Review the report's Resources tab with your instructor.

03 SCORE VS. HIGHLIGHTS

Why is my score 100% when only a few phrases are highlighted?

The score and the highlights measure different things. See "Reading the report" above.

The score is the share of the document classified as AI. It is not a confidence level.

Light purple marks the AI-classified blocks, and those blocks make up the score. Dark purple marks AI Phrases, which are evidence only. Each phrase shows how many times more often (×N) it appears in AI writing than in human writing.

The AI Phrases filter hides lower-confidence phrases. It never changes the score.

AI Source Match is a separate section. It shows text found in AI content that has already been published. It's only available to accounts with both AI and plagiarism detection.

The full picture. These are just a few of the many signals our system uses to detect generative AI in writing. Our deep learning models look at:

- Predictability: how closely each word matches the most statistically likely choice
- Sentence structure and syntax
- Word choice
- Phrase probability: language AI favors far more than people do
- Countless patterns too subtle to track by hand
- Consistent results across writing levels and document structures

WHAT TO DO

- Read the light-purple blocks, not just the dark phrases.
- Lower the AI Phrases filter to see all the supporting evidence.

04 CHANGING SCORES

Why does the score change when I edit, rescan, change file type, or scan part of the paper?

The model reads the whole document in context. Change the input, and the result can change.

Edits or reordering. Changing the order of paragraphs or the transitions between them changes the pattern the model reads. Big swings happen with text that was already close to the line.

Scanning only part of the paper. This removes context. One AI paragraph can be diluted in a long paper. Short sections often score 0% or 100%.

Different file type. PDF, Word and plain-text files extract different spacing, headers, line breaks and tracked changes. The text isn't identical, so the scores won't be either.

WHAT TO DO

- Compare like with like: the same file, in the same format, scanned whole.
- To check one paragraph, scan it on its own. Don't expect it to match the full-paper score.

05 MISSED AI

Why didn't Copyleaks catch AI that another tool flagged? Can I trust a 0%?

Detectors use different models, training data and thresholds, so they won't always agree. Copyleaks is tuned to avoid incorrectly flagging human writing.

Missed AI is rare. At the default Balanced setting, the false-negative rate is 0.64%. That means about 99 in 100 AI-written texts are caught, according to our published testing methodology.

Always improving. Our AI text detection model is continuously trained on the latest LLM releases, improving with each one. It is also retrained against humanizers and tested against every major new AI model.

Other tools differ. Other detectors use different models and training data, so their results won't always match ours.

What a 0% means. The text shows the natural variety typical of human writing. Heavily mixed writing, such as human-edited AI drafts, or text below the minimum length can also score low.

IF YOU SUSPECT MISSED AI

- Rescan at Sensitivity 3. It's built to catch humanized or heavily edited AI text.
- Scan the suspect section on its own, so it isn't diluted by the rest of the document.

06 NAMES, CITATIONS, TEMPLATES

Why are my name, citations or the assignment template flagged?

Text inside an AI-classified block is highlighted along with the rest of the block. That doesn't mean each word was AI-written.

Names, titles, headers and references are highlighted because they sit inside an AI-classified block.

How AI detection works. The model splits a document into chunks and classifies each chunk as a whole, not word by word. Chunks can cross sentence and paragraph boundaries. A human-written sentence right after an AI-written one can land in the same chunk and be flagged along with it.

Citations and quotes can be left out of plagiarism matching with the Omit filters in scan settings (references, quotes, citations, titles and table of contents). Omitted sections appear in gray and are excluded from the plagiarism score.

Prompts and boilerplate such as assignment prompts, cover sheets, standard headers and bibliographies can match web sources or other students' submissions in the plagiarism report.

Templates can be excluded so the scan doesn't flag the assignment's own template text.

WHAT TO DO

- Instructors: exclude up to 3 templates per assignment.
- Admins: exclude up to 5 templates across the institution at lti.copyleaks.com under Default Scan Settings.

07 SHARED DATA HUB

Why does a paper match my own earlier upload or the Shared Data Hub?

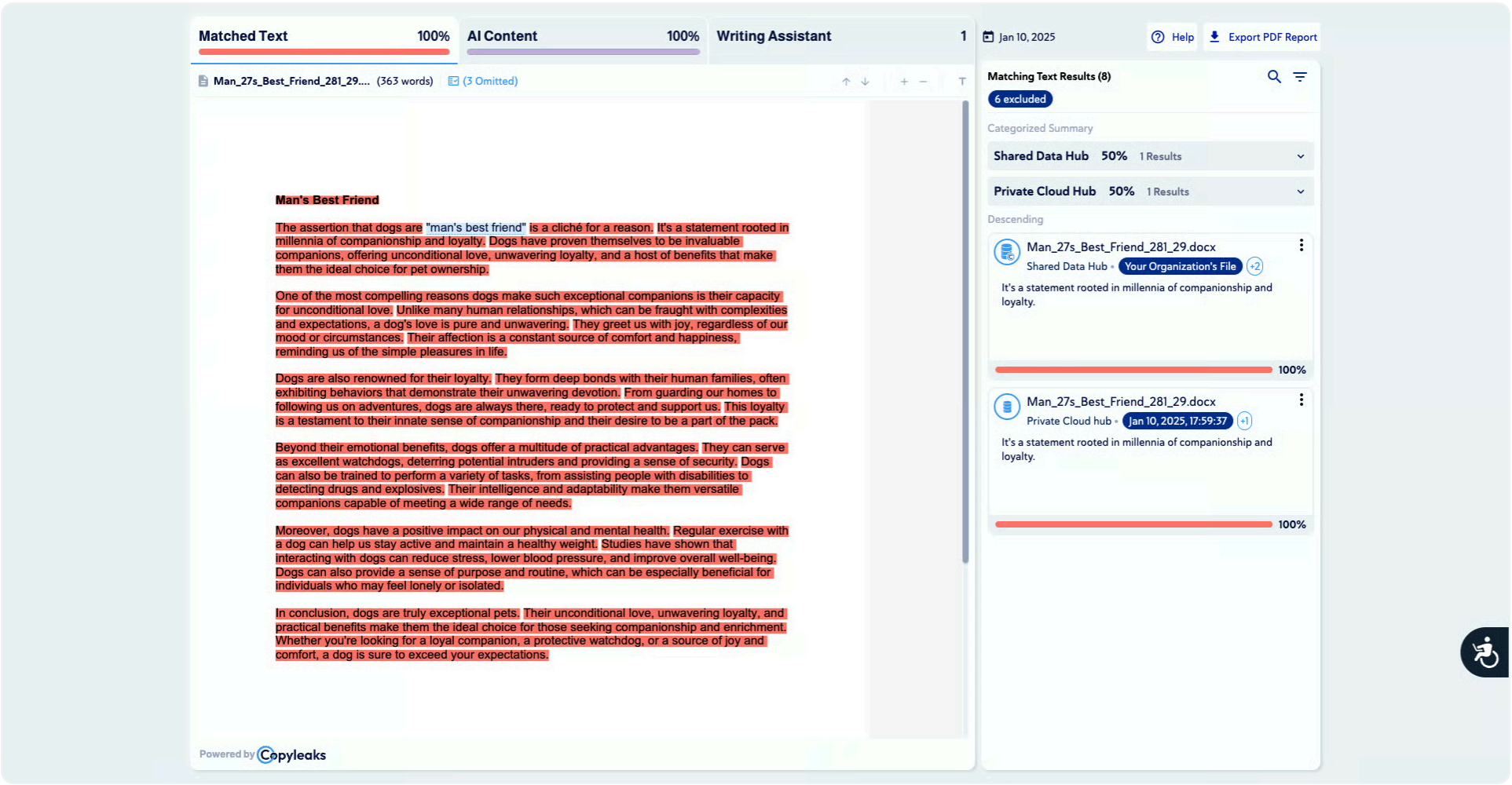
Each match label in the Matched Text tab tells you where the matching text came from.

"Your file [name] submitted" means the scan matched a document you saved to the Hub earlier. Scanning against the Hub also saves your document into it.

"Your Organization's File" means someone at your organization uploaded matching content. Admins can turn on student identification to see who.

"Other Files" means the match is another organization's upload. That document is masked, and only the identical text is shown.

Matching your own earlier work. A submission can still match your own earlier one if it was submitted in a different course or a different year. This mainly applies to LMS integration customers.



The source label appears on each result card in the Matched Text tab.

WHAT TO DO

- Delete your own submissions from the Hub. API customers can use purge: true.
- Matches against your own or your institution's earlier scans can be set aside from the final score in user or organization settings, or in the API with includeMySubmissions.
- Since September 22, 2026, only admins control the Hub setting, and a student's earlier submissions in the same course no longer match against themselves.

08 N/A OR DISABLED

What does "AI Content: N/A / Disabled" mean?

AI detection didn't run on this scan, for one of three reasons.

Too short. AI detection needs at least 200 words.

Unsupported file type. The report shows "AI Detection Not Executed: File Type not Supported."

Not turned on. API customers turn it on with aiGeneratedText.detect: true. LMS admins turn it on in scan settings. It can also be turned off in organization settings, the scan profile or the LMS assignment settings, or the plan may not include AI detection.

WHAT TO DO

- Check the length and file type, and confirm AI detection is on. Then rescan.

09 LANGUAGES

Does AI detection work in my language?

AI detection works in 30+ languages, but some features, such as AI Phrases, vary by language. See the full list of supported languages.

AI Phrases runs only in these six languages. Other languages show "Unsupported Language" for AI Phrases, but the AI score is still produced.

Sensitivity levels are English only. In a mixed-language document, the language that makes up most of the text is the one scanned.

Language	Human text	AI text
English	99.98%	98.40%
Spanish	99.85%	98.02%
Italian	99.88%	97.00%
French	99.88%	96.18%
German	99.94%	95.63%
Portuguese	99.95%	93.08%