AI Detector FAQs

Last Updated: June 24, 2025

Understanding the origin of human-written or AI-generated content is essential for maintaining enterprise-grade data security, ensuring regulatory compliance, protecting intellectual property, and upholding academic integrity.

Featured here are key questions we frequently get asked regarding our AI Detector.

How It Works

How is it possible to determine if something is AI-generated?

When a Large Language Model (LLM) generates a sentence, it draws from statistical patterns across its training data, resulting in writing that often lacks the variation and nuance of human expression. Our AI Detector identifies these statistical patterns and irregularities, especially when compared to massive volumes of verified human writing.

Visit our AI Detector Testing Methodology page to learn more.

How is your AI content detection any different from other detectors?

There are several significant differences between other detectors and our AI Detector.

For example:

- Over 99% accuracy, with third-party validation, driven by a refined model trained on trillions of human-written documents
- Enterprise-ready API and LMS integrations
- Sentence-level detection with confidence scores
- Does not flag grammar tools unless they use generative AI to rewrite passages.
- GDPR-compliant, SOC 2 and SOC 3 certified

How was the Copyleaks AI detection model trained?

We can recognize AI text patterns utilizing multiple techniques.

Since 2015, we've collected, ingested, and analyzed trillions of crawled and user-sourced content pages from thousands of universities and enterprises worldwide to train our models to understand how humans write. This allows our technology to detect irregular sentence patterns commonly used by genAI more accurately.

Also, by utilizing AI technology, our AI detector can accurately recognize the presence of other AI-generated text and the signals it leaves behind, adding an additional layer of accuracy.

How do you avoid false accusations?

The chance for content written by a human to be falsely labeled as AI-generated content is less than 0.2%. Nevertheless, we strive to inspire authenticity and digital trust by creating secure environments to share ideas and learn confidently, and that comes with the responsibility to ensure complete accuracy, particularly around false accusations. To address this, we have taken several precautions, including:

- Prioritize detection of human vs. AI, not just AI patterns
- Implement a user feedback loop to retrain our models continually.
- Only new detection updates will be released after rigorous internal testing.

What models can you detect, and what's the accuracy of each?

As of May 2025, Copyleaks can detect content from:

- GPT-4, ChatGPT
- Claude
- Gemini
- DeepSeek
- Llama (Meta)
- Bloom
- Rytr
- Jasper

...and more.

Using English text, each model's detection accuracy varies slightly from model to model, though each is above 98.0%. We don't train our AI Detector on specific language models (LLMs). Instead, we focus on the underlying text generation techniques these models use. This approach allows us to detect new AI-generated content seamlessly–if a new model utilizes an existing technique, our detector can identify it immediately. When new techniques emerge, we update our models to ensure continued accuracy and reliability.

Given the type of content being tested, you may encounter slightly different results. Accordingly, we suggest conducting several tests to determine the success rate for your specific content type.

What languages do you support, and what is the accuracy of each?

Copyleaks supports detection in over 30 languages–more than any other AI detection solution on the market. These include English, French, German, Italian, Portuguese, Spanish, Japanese, Chinese (Simplified and Traditional), Russian, Dutch, and more.

Accuracy is a top priority. Based on independent testing by our QA and Data Science teams, the average accuracy rates below reflect how well the latest version of the Copyleaks AI Detector identifies both human-written and AI-generated content across our most widely used languages.

- **English:** 99.98% (human), 98.40% (AI)
- **French:** 99.88% (human), 96.18% (AI)
- **German:** 99.94% (human), 95.63% (AI)
- **Italian:** 99.88% (human), 97.00% (AI)
- **Portuguese:** 99.95% (human), 93.08% (AI)
- **Spanish:** 99.85% (human), 98.02% (AI)

We continuously retrain and refine our models–especially in the most widely used languages–to maintain industry-leading accuracy and adapt to the latest advancements in generative AI.

Visit our AI Detector Testing Methodology page to learn more.

Is the AI Detector available for my LMS Integration? What about Microsoft Teams?

Yes. AI Detector is available for all LMS integrations and is not an extra add-on. Our LMS integration options include Canvas, Moodle, Brightspace, Blackboard, Schoology, Edsby, and Sakai.

Microsoft Teams only offers a student-view integration with no separate teacher view. Since our integrations are integral to educators and students, we currently do not offer Teams integration.

What data protection does Copyleaks have?

At Copyleaks, our products routinely undergo independent verification of privacy, security, and compliance control to achieve certifications against global standards and earn and retain the trust of the millions of Copyleaks customers worldwide.

Our current Copyleaks certifications and compliance standards include:

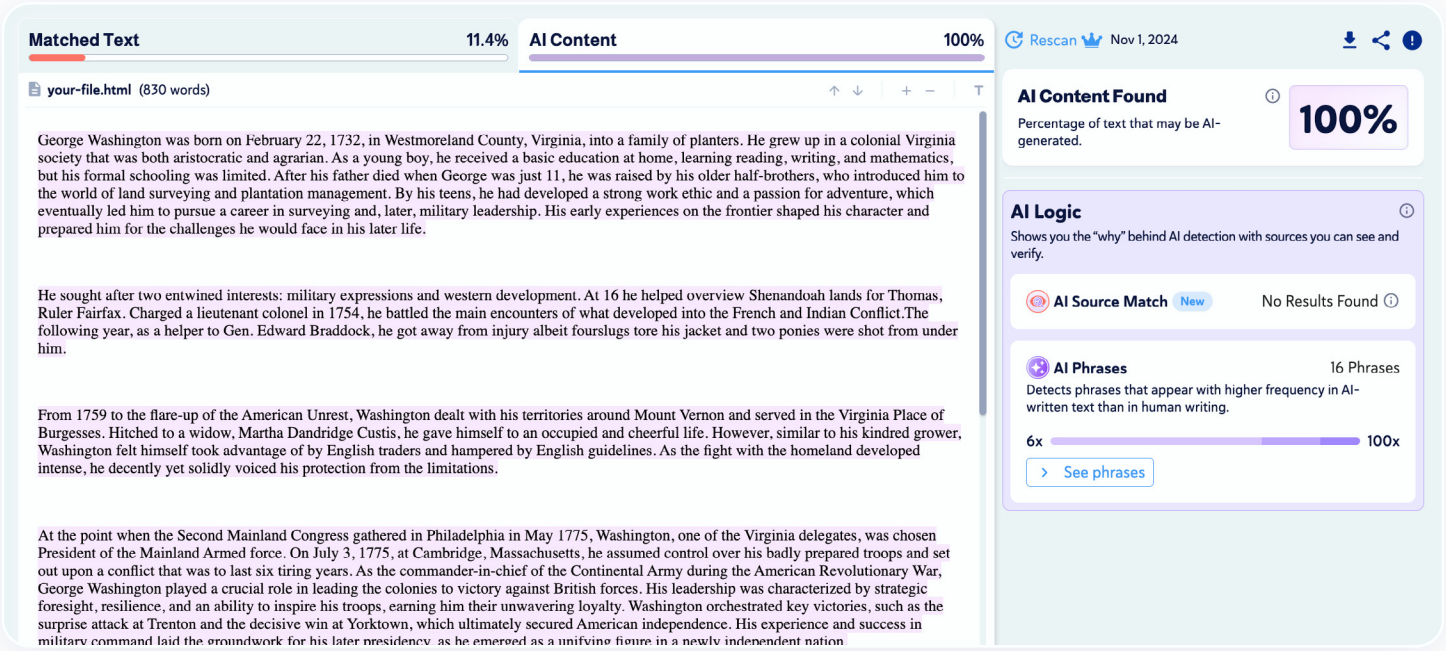
- SOC 2
- SOC 3
- GDPR
- PCI Payment Card Industry Data Security Standard
- NIST Risk Management Framework (RMF)

Please visit our Compliance and Certifications and Security Practices pages to learn more.

Understanding the Results

How will I know if AI content has been detected?

You will be notified on the Similarity Report if AI content is detected. See below.



AI Detectors can offer a lot of insight and data to encourage essential conversations in classrooms and boardrooms alike to determine the rules and regulations around AI. When your report states that AI content was found, take the time to investigate further. Again, the data provided by AI detectors should be used to inform the situation and offer the option for a learning opportunity and alignment on expectations.

Can you detect mixed text where human-created text has been amended with AI-generated text?

Yes. Our detection can discern specific elements of text written by humans and AI, even if the text has been interspersed.

How do the Similarity Score and AI content detection percentage differ? Are these completely distinct metrics, or is there an overlap in calculating them?

The Similarity Score shows the percentage of text in a document that matches other online sources or sources stored in our Copyleaks Shared Hub. It factors in identical text, minor changes, and paraphrasing.

The AI detection percentage is different. It estimates the total content in a document that generative AI may have created. The AI percentage does not influence the Similarity Score calculation, nor does the Similarity Score change the AI percentage. They are independent metrics produced by separate analyses.

What is AI Logic?

AI Logic is Copyleaks' advanced detection ecosystem designed to meet the evolving needs of organizations in education, publishing, business, and beyond. It moves beyond basic AI detection by providing transparent insights into why content is flagged as AI-generated, helping users make informed, confident decisions.

AI Logic is powered by two core technologies: **AI Phrases** and **AI Source Match**.

- **AI Phrases** highlights specific phrases that are most likely generated by AI, offering sentence-level insights to help contextualize detection results. Previously known as AI Insights, this feature gives users deeper visibility into AI involvement and supports better decision-making through explainability.

- **AI Source Match** enhances detection by checking whether AI-generated content resembles known large language model (LLM) output that has been publicly shared or widely circulated. While it does not confirm the exact origin or authorship of the content, it provides valuable signals for verifying originality, supporting compliance, and informing copyright and licensing review.

Will AI detection be a different workflow than the one we currently use with the Copyleaks report?

The workflow will remain the same. The only change will be within the report, where you will see a section for AI content detection alerts. Additionally, you can choose how the AI content alert is shown if you are working with the API.

Does adding AI content detection alter how I utilize the Similarity Report?

No, your workflow and use of the Similarity Report will not change with the addition of AI detection. It does not impact how you interpret or act on the Similarity Report. The core functionality and value of the report remain the same. The AI percentage notification provides supplemental information but does not modify the Similarity Score or how you leverage the report.

Detection Capabilities & Limitations

I've heard that AI Detection is just vaporware/snake oil. Is that true?

No. While AI detection continues to evolve, recent peer-reviewed research confirms that it can be highly effective when used responsibly and in the appropriate context.

For example, a 2024 study published on arXiv introduced the RAID benchmark, one of the most comprehensive evaluations of AI-generated text detectors to date.

Here's what the research shows:

- **Detection works, but context matters.** Detectors trained for specific domains, such as academic or scientific writing, have achieved accuracy rates above 98-99% in controlled environments.
- **False positives are a known risk.** Studies show that human-written content–especially from non-native English speakers–can be mistakenly flagged as AI-generated.
- **Simple edits can impact results.** Even minor paraphrasing or reformatting can reduce detection accuracy for some AI detectors, particularly for general-purpose models.
- **Ongoing validation is essential** and that's where Copyleaks stands out. Our AI Detector has consistently ranked among the most accurate in independent evaluations, with a proven track record across education, publishing, and enterprise. While no tool should be used in isolation for high-stakes decisions, Copyleaks is trusted by leading organizations to deliver reliable results backed by continuous improvement and transparency.

Bottom line, AI detection is a scientifically grounded, useful technology when applied with transparency and human oversight. It's not perfect, but it's far from snake oil.

Has a third party tested the accuracy of the AI Detector?

Yes. Copyleaks has been independently validated in multiple third-party studies conducted by researchers across the globe. To date, multiple academic and industry-led evaluations have confirmed the accuracy and effectiveness of the Copyleaks AI Detector in identifying AI-generated text–often ranking it as the most reliable among leading tools. These studies highlight Copyleaks' consistent performance across various content types and use cases, including education, publishing, and enterprise.

To explore the research and results, read the full breakdown here.

Can the AI Detector read source code and detect AI-generated code?

Yes. Copyleaks can detect AI-generated and reused code, even when it's been changed. Our system analyzes source code at the function level to identify copied, modified, or rewritten code–including content created by large language models (LLMs) and used without proper licensing or attribution.

By surfacing where code comes from and how it's been altered, Copyleaks helps organizations maintain IP integrity, enforce licensing compliance, and reduce the risk of hidden vulnerabilities in AI-assisted development.

Can the AI Detector detect AI within other content formats, such as video?

At this time, the AI Detector can only detect text.

However, we are always developing new features for future product updates.

Does the Copyleaks AI Detector flag writing assistant tools like Grammarly as AI content?

It depends on how the writing assistant is used. Some tools–like Grammarly–offer generative AI features that can rewrite or rephrase text. When these features are used, the revised content may be flagged by the Copyleaks AI Detector as AI-generated, since it was partially written by a large language model.

However, simple grammar or spelling corrections made by tools like Grammarly typically won't trigger detection. Similarly, the Copyleaks Writing Assistant is designed to support learning without generating content on behalf of the user–so it won't get flagged as AI.

To learn more, read our analysis about writing assistant tools getting flagged as AI.

No AI was used, but my text is getting flagged for AI. Why?

Even if you didn't use an AI tool like ChatGPT or Gemini directly, your content may still be flagged if other tools involved in the writing process use large language models (LLMs) behind the scenes.

Here are some common examples:

- **Language Translators:** Tools that translate entire passages often rely on LLMs to preserve meaning and tone, which can result in text being flagged as AI-generated.
- **Advanced Writing Assistants:** While basic grammar and spelling corrections are typically not flagged, tools like GrammarlyGO that offer rewriting, auto-complete, or paraphrasing features use genAI and may trigger detection.

We recommend reviewing your process and the tools used to identify where AI assistance might have occurred. The Copyleaks report provides phrase-level insights to help you investigate further.

Can the AI Detector catch text that's been spun or includes intentional typos?

Yes. The Copyleaks AI Detector is designed to recognize paraphrased content, including text that has been processed through spinners or modified with deliberate typos. These tactics are often used to disguise AI-generated or copied content, but our system can still detect underlying patterns that suggest AI involvement.

That said, detection models are continuously evolving. We regularly update our system to improve accuracy in identifying manipulated content, including increasingly sophisticated paraphrasing techniques.

What are the limitations of the AI Detector?

While the Copyleaks AI Detector is highly accurate–exceeding 99% in many scenarios–there are still a few limitations to keep in mind:

- **Text length matters.** Detection accuracy improves with longer content. For best results, we recommend scanning text that's at least 350 words.
- **Creative writing is more complex.** Poems, song lyrics, and highly stylized writing can be harder to classify due to their abstract structure. We're actively training our models to better handle these formats.
- **Language accuracy varies.** English currently yields the highest accuracy. As we expand our language training, performance across additional languages will continue to improve.
- **AI-written text may sometimes appear human.** While false positives are rare (0.2%), highly sophisticated AI content can occasionally be misclassified as human-written. We continuously update our models to reduce these "false negatives."

As with any detection technology, Copyleaks is most effective when paired with human judgment and used as part of a broader content evaluation process.

Why is there a minimum and maximum text requirement for some AI content scans?

To accurately detect AI-generated content, our models need enough text to analyze language patterns and structure. The more text available, the more confidently our system can determine whether it was written by a human or AI. That's why minimum and maximum character limits vary by platform.

Here are the current requirements:

- **AI Detector Browser Extension**
  Minimum: 350 characters
  Maximum: 25,000 characters

- **AI Detector Web-Based Platform:**
  Minimum: 255 characters
  Maximum: Up to 2,000 pages (no character limit within this range)

Providing adequate text ensures more reliable detection and higher confidence in the results.

What will we have to do to support new product updates?

Nothing on your end. Copyleaks handles all product updates automatically. Updates are deployed seamlessly in the background, ensuring your experience remains uninterrupted. When new features or improvements are released, we'll share detailed release notes so you're always informed about what's new and how it enhances your workflow.

Will Copyleaks be able to detect newer models that will come out?

Yes. Thanks to machine learning, we train the system to detect new genAI models accurately once released.

What other AI content detection capabilities are you working on?

We're continually enhancing our detection technology to stay ahead of evolving AI tools and writing tactics. Current areas of development include:

- **Improved detection of manipulated content,** including paraphrased text, spun content, and deliberate typos

- **Expanded language support** to ensure high accuracy across more global markets

- **Broader model compatibility** to detect content generated by emerging and lesser-known LLMs

- **Ongoing accuracy optimization** across all content types and formats

We actively monitor trends and listen to customer feedback to ensure our detection tools remain at the forefront of reliability and innovation.

sales@copyleaks.com • copyleaks.com