An Institution's Guide to Grammarly Authorship

Grammarly is now part of Superhuman

grammarly
part of SUPERHUMAN

Welcome!

Despite widespread adoption by educators, most remain unsure how to use AI it effectively, and institutions are seeking ways to govern AI adoption at scale. Superhuman (formerly Grammarly) is committed to supporting educators and institutions as they move from reactive AI policies to intentional AI-assisted learning, fostering environments that help students get real-world experience with the tools shaping the workplace.

Since launching, Grammarly Authorship has helped bring transparency to the writing process by identifying which parts of a document were human-created, AI-generated, or AI-edited. More than 5 million Authorship reports have been generated by students to show their instructors how they're using AI responsibly. At Rowan-Cabarrus Community College, this led to a 96% reduction in academic integrity violations in one semester. This impact was recognized by the EdTech Breakthrough Awards, which named Authorship the best Overall Edtech Solution — a reflection of the trust educators and institutions have placed in transparent, process-first approaches to AI in the classroom.

To continue deepening its support for students, Superhuman launched "Docs," a workspace with agents that offer proactive, appropriate support for academic writing, giving institutions a governed space where AI use is visible, attributable, and purposeful.

We've updated this guide to reflect the latest Authorship capabilities and instructor feedback. In it, you will learn more about the challenges that Authorship solves, how it works for students, and how best to implement Authorship for your institution.

If you would like to speak with a member of the Superhuman for Education team to learn how Authorship can help your institution innovate responsibly with AI, set up a guided tour of Authorship.

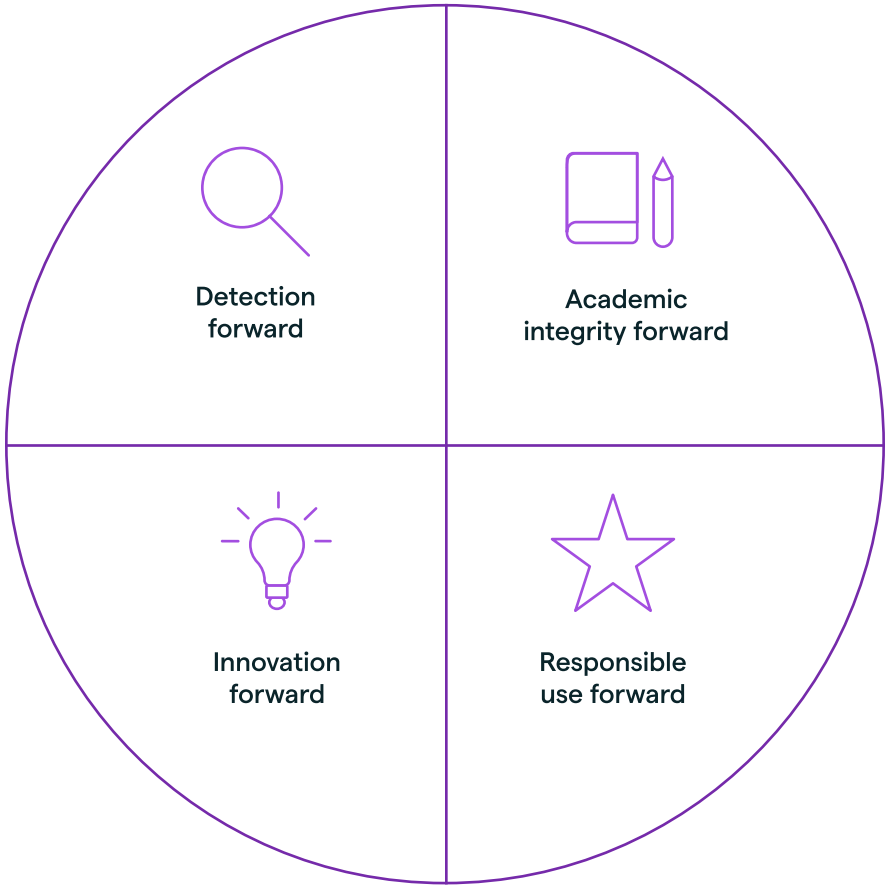
Education's AI Dilemma

There's a growing divide between those who embrace AI and those who seek to shut down its use.

AI usage rates across students, faculty, and professionals continue to grow. Students have grown accustomed to having AI embedded into the technology tools they use regularly. While equity gaps persist, having ease of access to generative AI tools is less of a barrier than understanding how to use the technology thoughtfully and effectively. And the job market is clearly signaling that it expects college graduates to possess at least baseline AI literacy, with employers across industries being "significantly more likely to offer job interviews and higher salaries to college graduates with 'AI capital.'"

The emerging AI dilemma is how to balance the imperatives of AI literacy and career preparation with the imperatives of traditional learning outcomes and academic integrity. But this balancing act is difficult to calibrate.

In our conversations with institutions, it's clear that there are four primary attitudes toward AI at the moment.



“Detection forward” instructors and administrators are dealing with economies of scale and are concerned with rampant student cheating and cognitive off-loading with AI. They tend to resort to detection as the “least worst” solution to cut down on integrity violations.

The **“academic integrity forward”** share the same concerns as the first camp about student integrity issues, but see AI as a reality that can’t be avoided.

“**Responsible use forward**” instructors encourage their students to use AI under specific circumstances, while wanting mechanisms in place to monitor inappropriate use.

And “**innovation forward**” instructors are embracing AI to reinvent their assessments and actively encouraging AI use in a wide variety of circumstances.

The truth is that there’s no single right answer with such a new and disruptive technology. What all of these attitudes require is visibility into the student writing process, whether it be to reduce instances of cheating, or whether it be to develop a better understanding of how students are using these tools creatively.

The good news for you if you’re reading this is that Grammarly Authorship provides value regardless of your current posture toward AI. The feature is meant to give insight into student writing in the AI era through a student-first tool that ensures they see what instructors see, creating two-sided transparency into text sources used and the writing process. Furthermore, as a tool that runs across web pages (and now on desktop applications), Authorship automatically categorizes text and attributes that text to the specific source without extra effort from either the student or instructor.

Ultimately, Authorship is data: data that can be used to cut down on suspected cheating, coach students to write more thoughtfully with AI, or merely understand if and how students are leveraging AI tools in their work. The feature is a trust bridge between instructors and students to navigate the challenging AI waters together.

Detection vs. Transparency

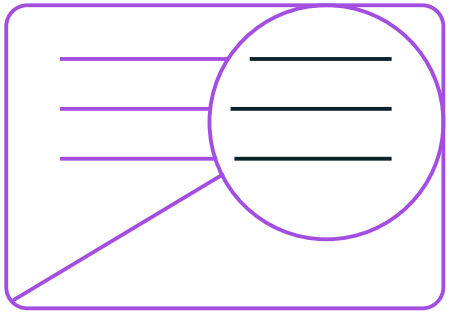
While AI detection still has its place in certain contexts, it's worth examining the issues associated with using AI detection to preserve authenticity in assignment submissions:

Detection on algorithmic logic to identify language patterns typical of AI. It offers a percentage breakdown of the entire passage that is estimated to be AI-generated and is not 100% accurate.

Process transparency provides visibility to both students and faculty into how text was created, sourced, and edited between student, AI, and the web across the entire student-AI collaboration process.

Algorithmic logic

Black-box algorithms to estimate text sources



Only professors have visibility

Transparent attribution with user content

Automatic categorization based on clipboard access

Human-typed	Sourced content
Typed	AI-generated
Typed, then edited with Grammarly	Copied from an online source
Typed, then rephrased with Grammarly's gen AI	Copied from an unknown source

Professors and students have visibility

Algorithmic-based AI detection creates a black box that provides few actionable insights for instructors and even fewer for students. Transparent authorship enables both students and faculty to have a productive dialogue on shared data.

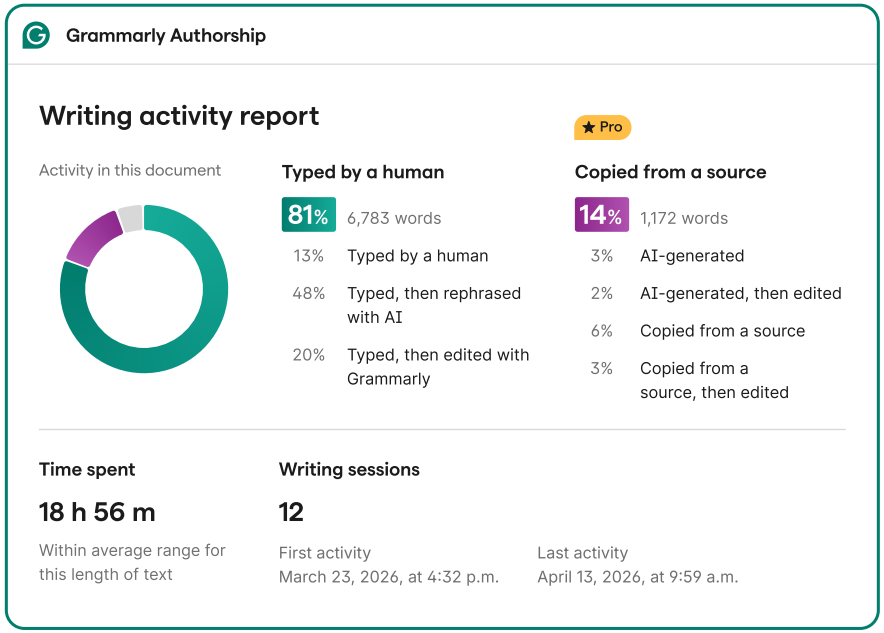
How Grammarly Authorship Works

At Grammarly, we believe there is a better way

Rather than relying on an after-the-fact estimation of which text may or may not have been created or modified with AI, Authorship relies on Grammarly's on-browser (for Google Docs) and desktop (for Microsoft Word) ubiquity to categorize text as it enters a student's doc.

Authorship does not use an inferencing model, so there is no algorithmic bias. It simply tracks how students compose their work, grouping text sources based on their origin. Similarly, none of the categorizations or data that Authorship surfaces comes with any sort of value judgment; Authorship is merely surfacing data that Grammarly is uniquely able to uncover by virtue of running across your browser or desktop, helping users as they write, and categorizing text accordingly.

When a student enables Authorship tracking in **Grammarly's docs with agents**, Microsoft Word or Google Docs, Grammarly Authorship proactively tracks the writing process as they write, categorizing text sources as they are entered into the body of a document. Because Grammarly runs on your browser, desktop, and within the body of a Google or Word Doc, Grammarly is able to distinguish between text that is typed, edited, AI-generated, and copied from AI, online, or desktop sources (note: Only tracking Copilot in Microsoft Word can attribute desktop-based sources).



The student won't see this categorization in the body of their document, but they can click the thumbprint icon in the bottom-left corner of their document anytime to see an Authorship report that will open in a new window.

For both students and instructors, Grammarly Authorship provides three shared vantage points—analytics, reports, and replays—for students to demonstrate how they thoughtfully composed their work and for educators to understand that process.

Authorship analytics

At-a-glance view of student text sources

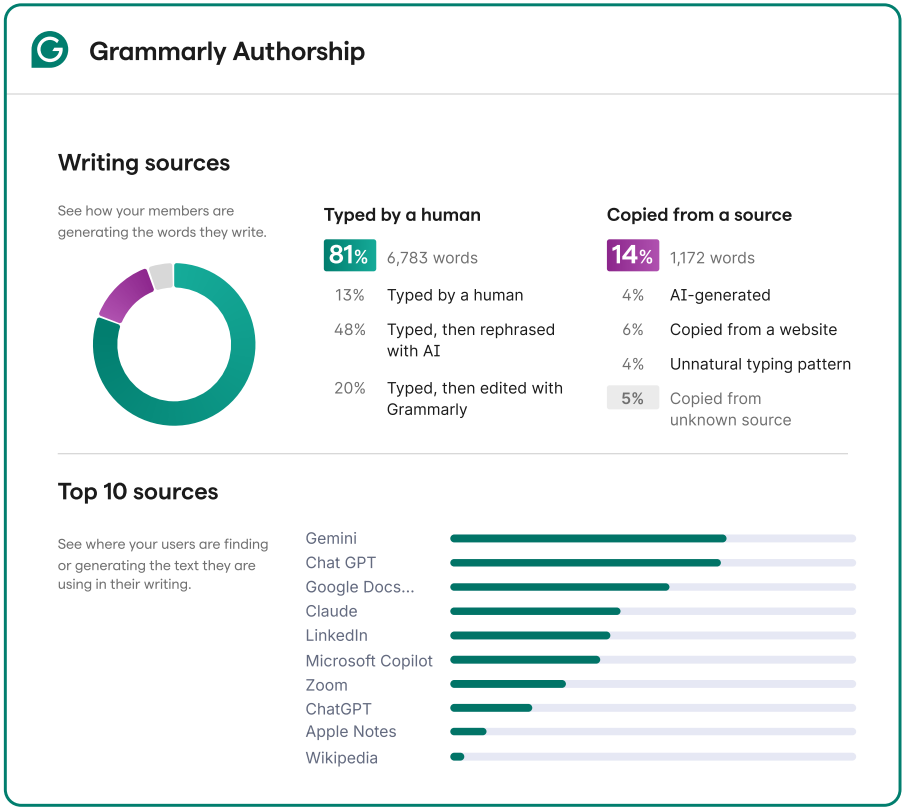
Authorship analytics aims to provide instructors with a quick snapshot of the text sources used in a given submission while encouraging them to go deeper.

At the very top of the report that a student shares, faculty will see a graph and broad data breaking down the text sources in the document. These are Authorship analytics.

Analytics includes three broad categories, with more granular categorizations broken down by percentage under each of the umbrella categories:

1. **Typed by a human**
   a. Typed by a human
   b. Typed, then rephrased with Grammarly's AI
   c. Typed, then edited with Grammarly's traditional suggestions (no large language model actions)

2. **Copied from a source**
   a. AI-generated
   b. Copied from a source (non-generative AI websites only for Google Docs and Microsoft Copilot; Word also includes paste actions from desktop applications)

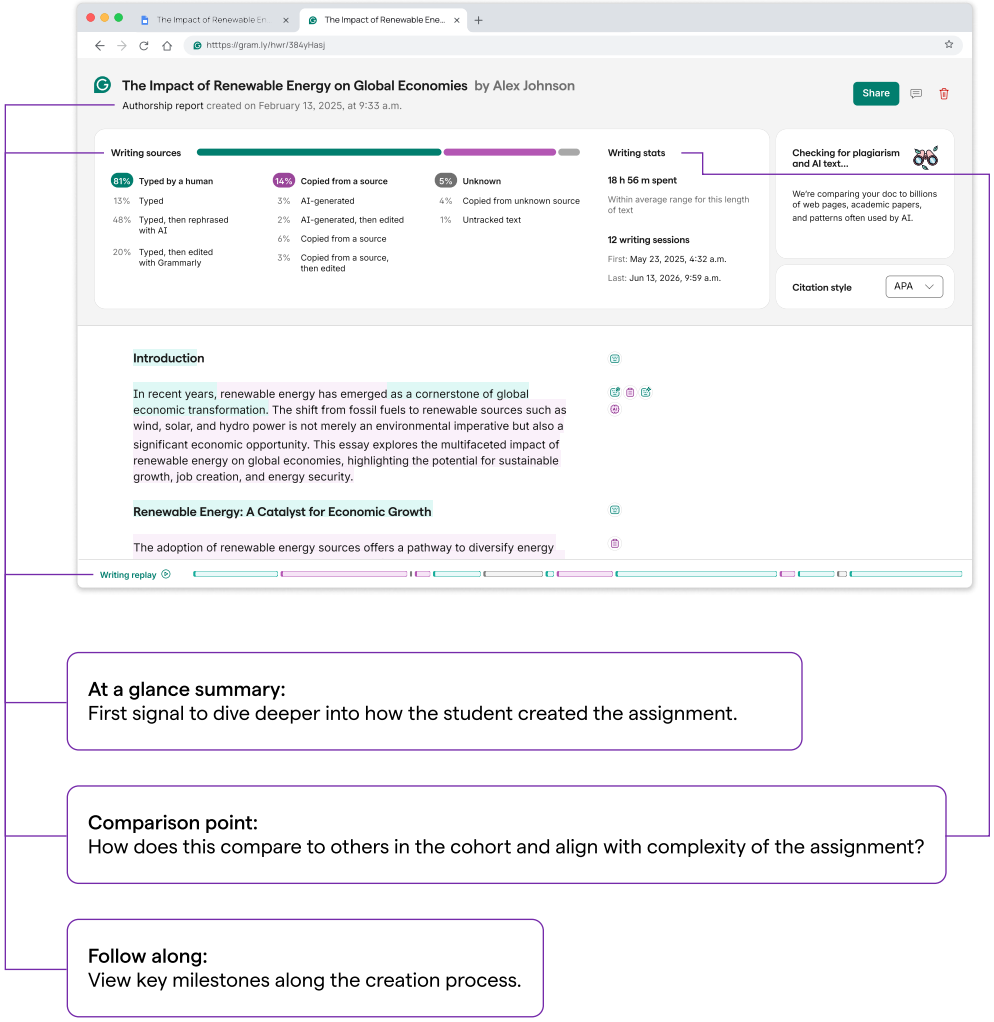


How to use analytics

Authorship report of student writing

We strongly encourage faculty to use the high-level data within Authorship analytics as a signal for a more thorough review if text categories appear misaligned with the assignment. For example, if a professor explicitly stated that they expect to see a diverse set of text sources in a document but then encounter a paper that is 95% typed, that is a signal for them to look at the Authorship report and understand more deeply what the student did in their assignment.

It's important to note that the high-level data is grouped intentionally into broad categories and that the "Copied from a source" category combines gen AI and traditional sources into one high-level number. The reason for this is that Grammarly does not want faculty to make snap judgments based on a single number, even if we believe that number to be more objective than AI detectors. Rather, Authorship analytics are intended to give time-strapped faculty the ability to gloss over reports that are in line with expectations and go deeper on the ones that appear to be misaligned at first glance.



At a glance summary:
First signal to dive deeper into how the student created the assignment.

Comparison point:
How does this compare to others in the cohort and align with complexity of the assignment?

Follow along:
View key milestones along the creation process.

Authorship reports

Deeper examination of student text and editing history

Once you scroll down from analytics, you will see the entirety of a student's submission color-coded based on how it is sourced or typed, mirroring the high-level categories in the analytics.

Each paragraph has at least one "insight card" noting the specific categorization, along with additional data about the number of words that may have been typed or copied, the number of words that were then edited, and total editing time for that particular section of text. This is intended to help reviewers see where text came from and better understand how specific text may have been modified once typed or pasted into the document.

Types of Insight Cards

Insight cards help writers and reviewers understand how typed text was adjusted with AI

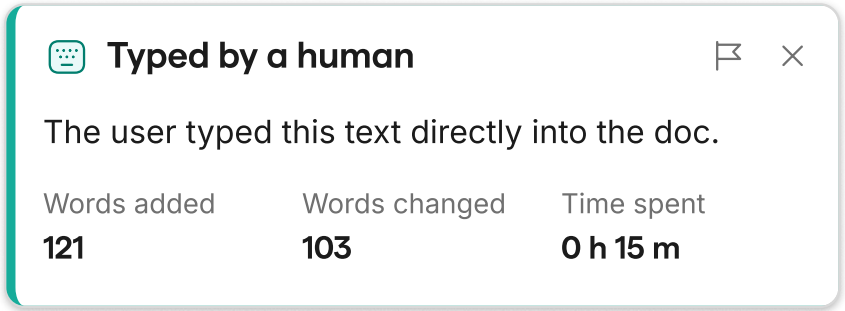
There are a variety of detailed insight cards that may show up in Authorship reports. Some cards are not available to see unless you and the end user who submitted the report have Grammarly for Education or Grammarly Pro plans. You will also notice slight variances in the insights depending on which writing platform a student used (Grammarly's docs with agents, Google Docs or Microsoft Word).

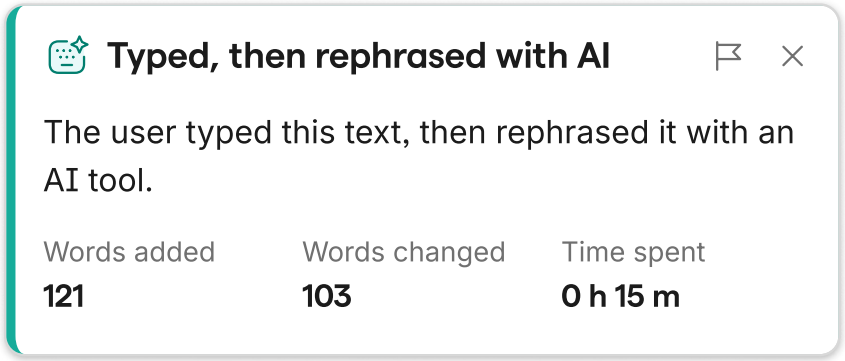
Typed by a human cards: These insights are available with all Grammarly plans and in both Google Docs and Microsoft Word.

1. **Typed by a human:** The user typed the text directly into their document.

2. **Typed by a human, then edited with Grammarly:** The user typed the text directly into their document, then made changes with Grammarly's traditional in-line suggestions.

3. **Typed by a human, then rephrased with Grammarly's generative AI:** The user typed the text into their document, then used one of Grammarly's generative AI prompts to rephrase the text.

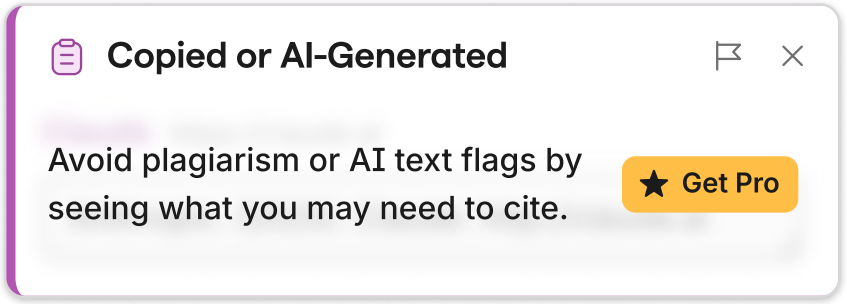


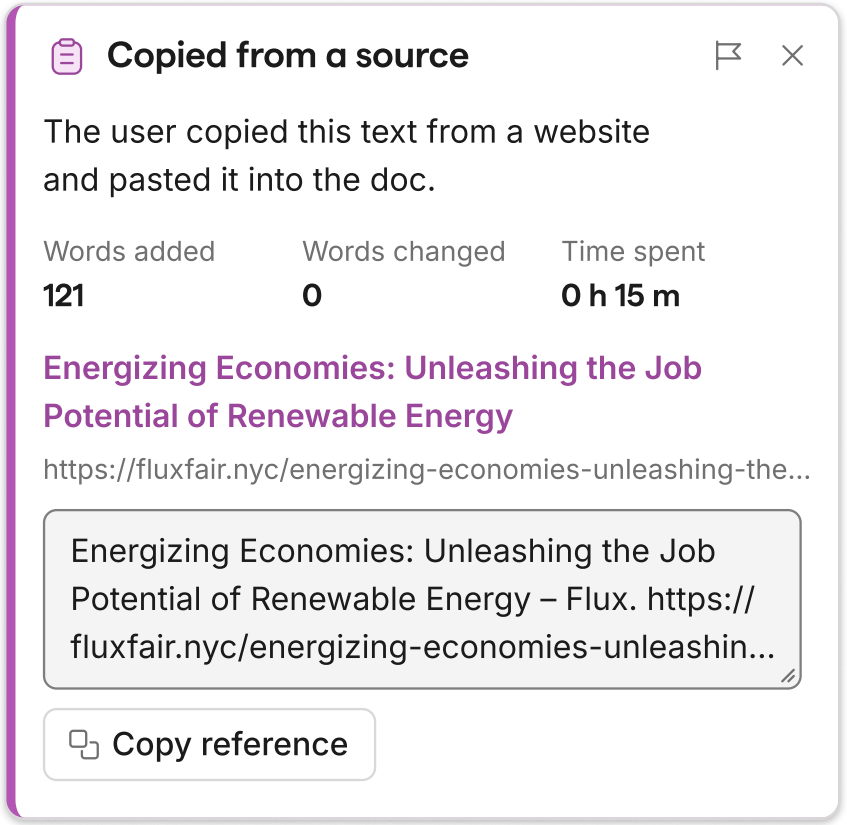


Copied from a source cards: Instructors who have Grammarly Pro or Grammarly for Education plans will see all of these insights, when appropriate, regardless of what plan was used to create the report.

1. **Copied or AI-generated**: Users and reviewers on free plans will see cards labeled “Copied or AI-generated,” but will not see the specific source of the text. Reviewers can assume that this text was copied from a known online or desktop source (Word only), including from potentially AI sources.

2. **Copied from a source** (Grammarly Pro and Grammarly for Education; Docs with agents; Google Docs and Word): These cards will attribute text to non-generative sources and include the source of the text on the card. All cards of this type from Google Docs will be from websites, while cards for documents that were written in Word can include website attribution as well as desktop application.



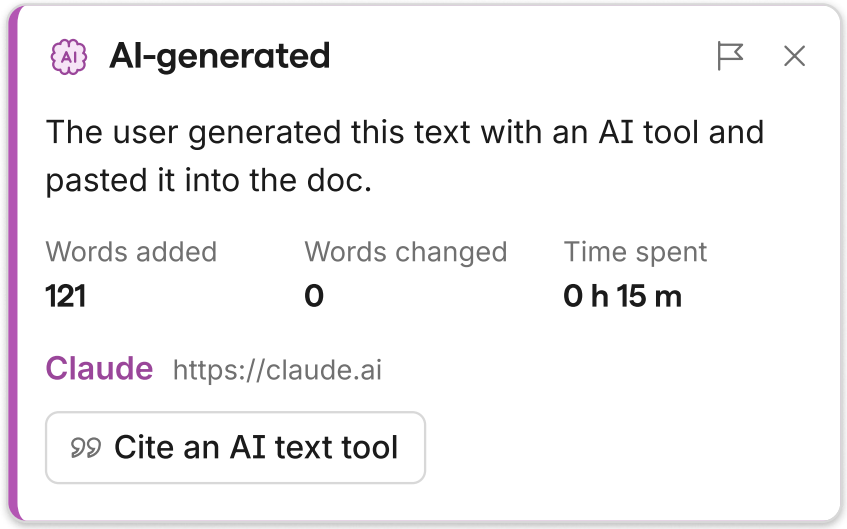


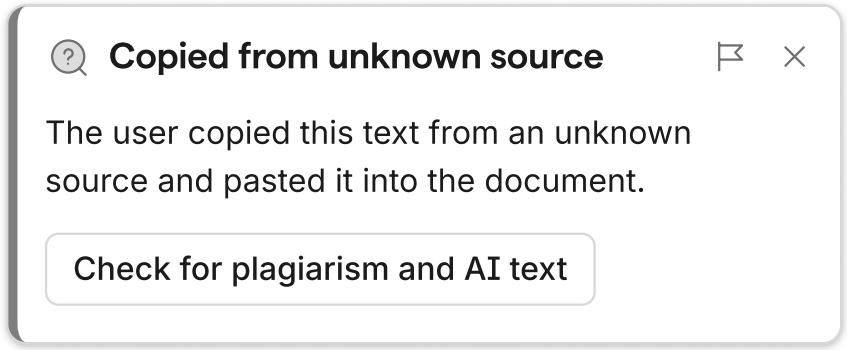
3. **AI-generated** (Grammarly Pro and Grammarly for Education; Google Docs and Word): These cards will show any text that was copied directly from an AI-generated chatbot (ChatGPT, CoPilot, Claude, Gemini, Perplexity, Grammarly). Note: Only reports generated from Microsoft Word will be able to attribute AI-generated text from AI Chatbot desktop apps. Reports generated from Google Docs will be able to show sources only from web-versions of AI chatbots.

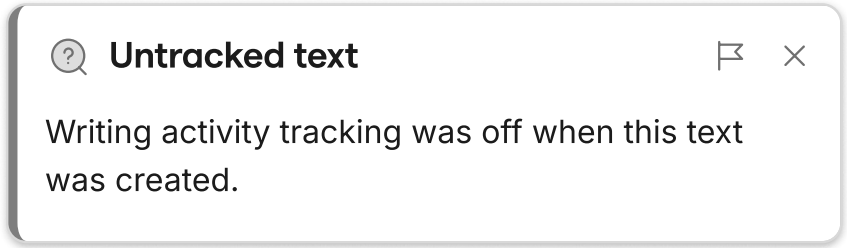
Unknown cards: These are cards that show text that was either pasted from a source that Grammarly doesn't have access to or that appeared when tracking was not turned on.

1. **Copied from an unknown source:** The text was pasted from a source either online or on a desktop where Grammarly doesn't run.
    a. Example: A student pasted text from a desktop application (e.g., a Microsoft Word doc) into their Google Doc.
        i. Because the Grammarly Extension runs Authorship within Google Docs and does not run on the user's desktop, Grammarly wouldn't be able to attribute this text but would know it was pasted, not typed.

2. **Untracked text:** This text entered the document when tracking was not turned on. Most of the time, this is because the student started Authorship tracking after they had started typing or enabled tracking on a document that already had considerable text in it.







How to use reports

Quick walk-through of text color in the report

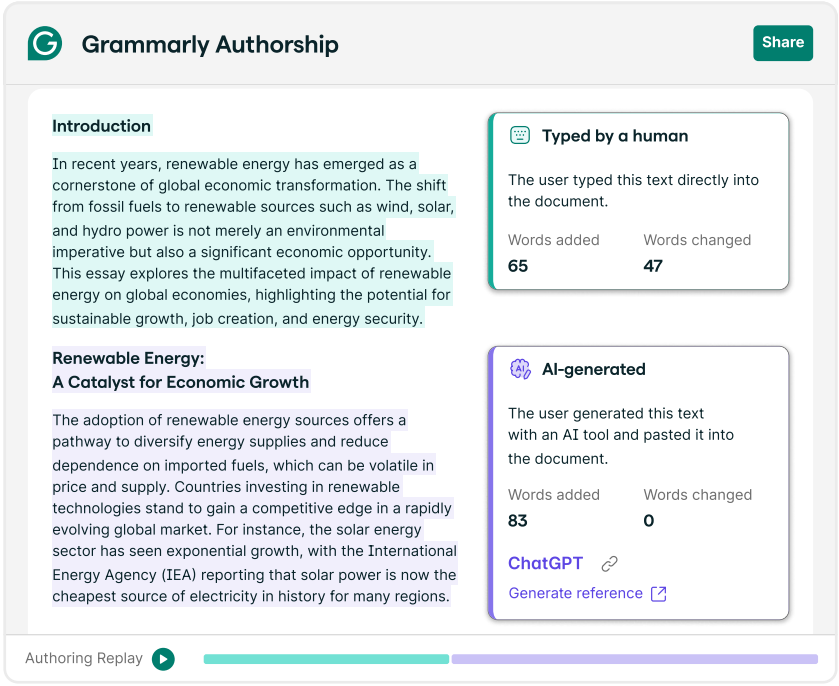
Lilac text
Lilac text indicates it came from an external source, and the corresponding card will show the source and, if applicable, URL for that specific text. If there is AI-generated text, students will be able to see their exact prompt history by clicking the link, while reviewers will see only the source (e.g., ChatGPT). If text is from a generic website like Wikipedia, both students and professors will see the link to the article present.

Green text
Green text indicates text that was typed into the document by a human.

Gray text
Any text that was pasted into the document from a source that Grammarly cannot track will show up as gray and will be labeled “Copied from an unknown source.” This doesn’t mean that anything ill-intentioned was done by a student; it merely means that they pasted text from a window that Grammarly cannot access.

The Authorship report provides teachable moments on how to attribute text properly. Authorship makes it easy for students to see where they’ve used external sources and easily grab a preformatted citation to include in their references section—but they still have to take that action on their own before they submit.

The report is also a good opportunity for writing instructors to see where students may have brought in external text and then modified it with their own words and knowledge, something that is shown in greater depth in the Authorship replay.



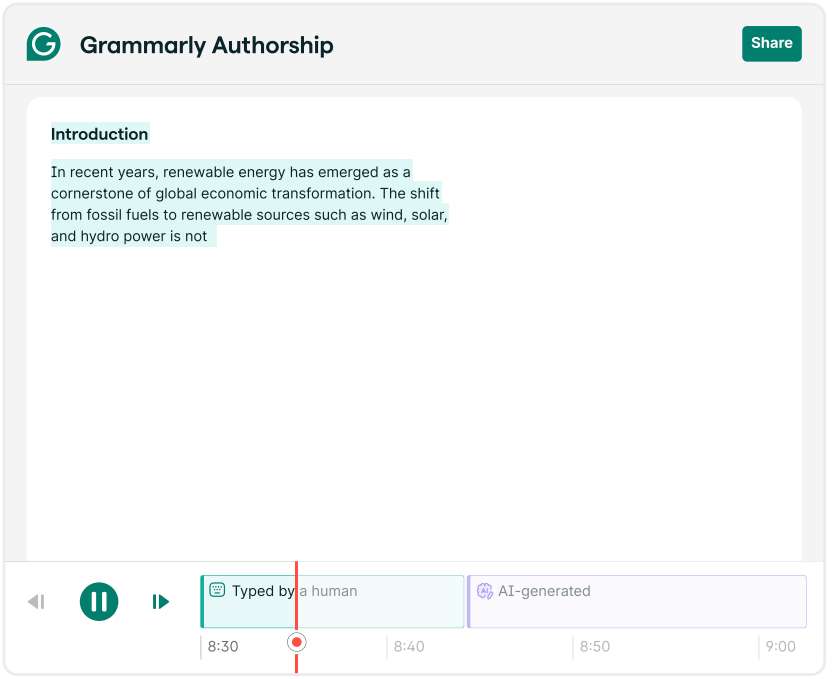
Authorship Replays

In-depth review of the writing process

For faculty who want insight into the full writing process, there is a full Authorship replay available. Upon hitting play, reviewers will see all text actions taken in the document from the moment that Authorship tracking was enabled by the student.

If the student enables Authorship after they've added text (for example, after they've created an outline), then any text they already have in the body of their document will be displayed there without attribution.

Reviewers can watch the replay from start to finish or jump to specific text actions of interest. For example, if the report showed an instance where the student used AI-generated text, the reviewer can jump ahead to that moment in the replay and see how the student inserted the text and then modified it.

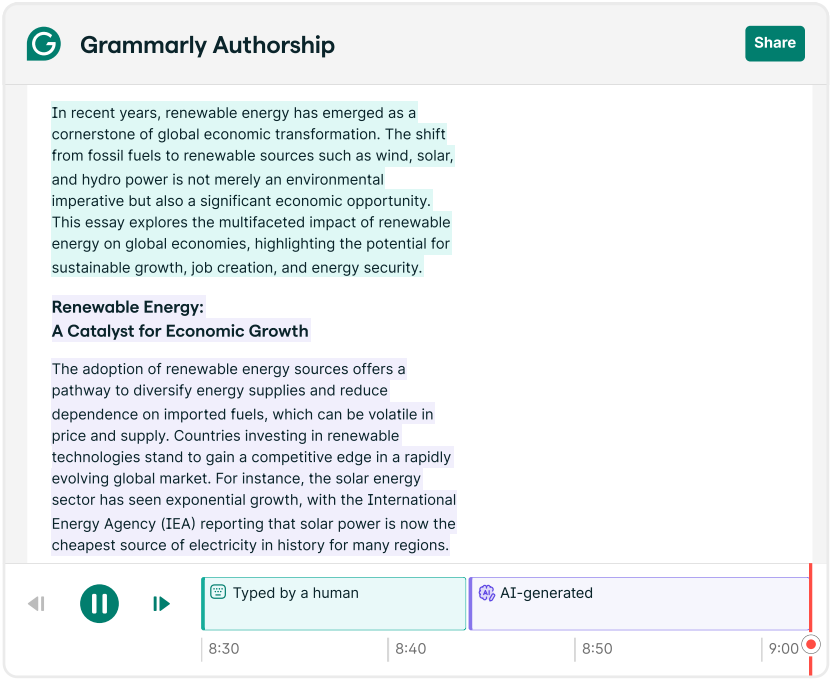


How to use replays

Playing back the writing process

The replay also offers a great opportunity for instructors to see how students are leaning on AI to revise/rewrite their own words. In addition, instructors will be able to identify patterns where individual students may get stuck in the writing process. The replay will show instances where students continuously rewrite sentences and paragraphs, struggle to cite efficiently, or where they use a scratch pad to type out ideas that are eventually deleted from their papers. All of these activities provide a window into the student writing process for professors whose goal is to teach more effective writing practices.

It's important to note that, due to early-stage feedback from both students and faculty, **students have the opportunity to exclude the replay when sharing their Authorship report with their professors.** Numerous students and faculty expressed reservations about reviewers being able to see their full editing history, which can admittedly be messy for many writers. Grammarly's recommendation is to avoid penalizing students who do not share their replay. By encouraging students to share it, it will provide faculty with more insight into where students may need more support in becoming better writers and editors.



Guidance for instructors

How to support student learning and AI literacy

Authorship is a student-first tool for the AI era that is meant to foster greater transparency between students and instructors in the writing process. By design, students have agency over tracking. They can view their reports before sharing them with their instructors. Importantly, however, they cannot change or manipulate any of the data in the report. The replay will capture all activity completed while tracking was turned on. So, students will know before they submit how closely they adhered to the guidelines of the particular writing assignment, creating a shared understanding with instructors if any questions arise.

So, if you're an instructor reading this, you may be wondering: When should I use this with my students? Well, after pressure testing this with millions of students and getting a lot of feedback from instructors we have some suggestions!

Three actionable strategies to support AI-assisted learning

1. **Cut down on inappropriate AI use.** If you teach a course in which AI use is not appropriate, you can require that students use Authorship as a way of cutting down on student AI use. By telling your students that they must submit their written assignments alongside a shared Authorship link, you will be dramatically reducing the incentive to "cognitively off-load" the writing to an AI chatbot because students will know that copy and paste actions will be captured by the tool.

2. **Encourage thoughtful AI use with full transparency.** If you want to meaningfully integrate AI into your assignments as a genuine learning tool — Grammarly's Docs with Agents gives you a governed space to do it. Students write with access to institutionally approved AI writing agents, while Authorship automatically captures every interaction from the very first word. No setup required on the student's side: tracking is on by default, and agent-specific attribution makes it easy to see exactly when and how AI was used at each stage of the writing process. Rather than guessing whether a student engaged critically with AI, you'll have a clear, shared record of it — creating the foundation for richer, more honest conversations about what good AI-assisted writing actually looks like.

3. **Identify areas of improvement in the writing process.** Particularly for writing faculty, Authorship is a great way to help students improve their writing and ensure the vision of "writing as thinking" comes to life. With Authorship replays, writing faculty can zoom in on areas of the student writing process that may be particularly challenging for them and see how they modify, rewrite, and continue editing, offering detailed opportunities to provide more specific support to each student.

> "From introducing Grammarly with its own AI detection and Authorship, we went from 27 academic integrity violations to one. It opened up more transparent conversations with students. That's one of the biggest wins we've had here recently. It wasn't a tool used to catch students—it was a tool to make students responsible."

– Jenny B,
Program Chair of Associate in Arts & Division Chair of English and Study Skills
Rowan-Cabarrus Community College

We recommend that faculty be transparent with their students about why they should use Authorship, focusing on the potential benefits to students and their own learning:

1. Provides students protection against any false positives from AI detectors that may be used by faculty because students can present a record of their editing history and text sources with minimal added effort.
2. Allows faculty to get a more objective, clearer view of text sources used in an assignment without having to rely on an AI detector.
3. Gives faculty more in-depth insight into how students wrote a particular assignment, enabling them to identify course-level trends and key instructional points more quickly.
4. Ensures that faculty and students can have more substantive conversations about the writing choices made on a particular assignment because both parties have the same information ahead of time.
5. Has the potential to illuminate individual areas of improvement for specific students via the Authorship replay, which can inform more personalized instruction to help students improve their drafting and editing skills.
6. Helps faculty to quickly identify cases when students may have employed sources that were out of step with the particular assignment and quickly address this with students before it becomes a formal academic integrity violation, creating a learning opportunity instead.

How to go beyond detection to support trust

The primary goal of Authorship is to facilitate more authentic, responsible writing assignments by surfacing a more transparent method of text attribution.

Trust through transparency can be achieved only if:

1. Students enable Authorship, write according to their instructors' guidelines, and trust that their instructors will accept their submissions accordingly.
2. Instructors use the insights provided by Authorship to help students improve their writing and have meaningful conversations about what good writing and responsible technology use look like in the AI era.

Grammarly does not recommend that faculty use Authorship as a way to police student work in any way that disrupts substantive conversations rooted in student learning.

In other words, Authorship should not be used as a way of screening student work for too much or too little use of AI or as a single data point that is used to penalize students for inappropriate use of AI. Furthermore, faculty should not use the data that Authorship provides as singular data points to make decisions about grading or referring for academic integrity violations. While this advice is consistent with advice on how to deploy effective AI detection, the biggest difference is that students will have visibility into what their Authorship report contains. As a result, there should not be any surprises when instructors come to them to raise questions about certain actions taken or point out where their use of a source may be out of step with their guidelines.

Where authorship works

Authorship meets students and instructors where they already are — across writing surfaces, browsers, and learning management systems.

Works where students write
Authorship works natively in Google Docs (via the browser extension), Microsoft Word on Mac and Windows (via the desktop app), and Grammarly's own AI-powered writing surface, Docs with Agents. In Docs with Agents, tracking is on by default — no student setup required.

Works in the browsers they use
Authorship runs in Chrome and Edge today. Firefox support is arriving with the next browser release.

Integrates with your LMS
Authorship connects directly with Canvas LMS - instructors enable it per assignment with a single toggle, and Authorship reports are automatically surfaced in a faculty dashboard at submission. Blackboard integration is expected in 2026.

We welcome any and all feedback. Please let us know what you think and share your feedback via this survey!

Get a guided tour of Grammarly authorship

If you would like to speak with a member of the Superhuman for Education team to learn how Authorship can help you innovate responsibly with AI at your institution, set up a demo using the link below.

Get a Guided Tour of Grammarly Authorship →