



Rethinking academic integrity: Transparency and evidence in the AI era

How Mentafy closes the gap between AI detection,
fairness and students' own work



CRITICAL THINKING

WRITING SUPPORT

FAIRNESS

AGENDA

- 1 Challenges with AI
- 2 Mentafy's solution
- 3 Appendix: Software overview

1 Challenges with AI

Theses on the future of academic integrity

		YES	NO
1	Academic writing and research should still be practiced through theses. Alternative assessment formats are very valuable, of course, but they do not replace academic writing.	☐	☐
2	Students who let AI do too much of their writing risk learning too little.	☐	☐
3	Documenting how a text was written can show how much of the work is the student's own.	☐	☐
4	Nobody reviews every record. An analysis has to show whether a history is unremarkable or deserves a closer look.	☐	☐

4x yes? Then you are in exactly the right place. 3x yes? Then let's convince you!

● Challenges 2/4

Generative AI is changing education and fueling ghostwriting and plagiarism



35%

AI cheaters

35% of students surveyed have already had generative AI write term papers or theses for them

Bitkom study Digital Universities, 03.2025



94%

undetected

In one study, 94% of AI-generated submissions went undetected – and on average scored half a grade higher

A real-world test of artificial intelligence infiltration of a university examinations system: A “Turing Test” case study, Reading University 06.2024

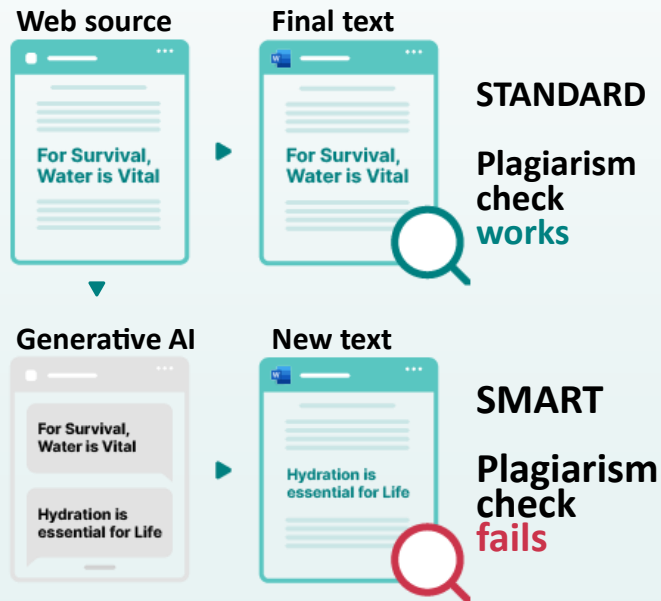
→ This is not new! And across studies (e.g. HEPI) the numbers, if changing, clearly increase since 2024.



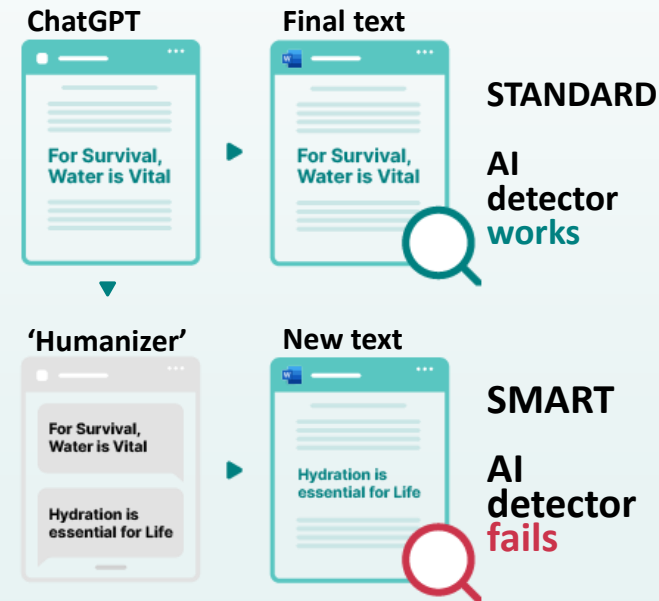
● Challenges 3/4

Generative AI pushes existing tools to their limits:

Cheaters **evade** plagiarism checks by paraphrasing with AI



Cheaters **defeat** AI detectors and **watermarks** with targeted prompts and 'humanizers' – using the same technique



When a text is AI-generated with a **targeted prompt**, the rate at which it is correctly classified as AI drops to 26%

Scarfe et al.: A real-world test of artificial intelligence infiltration of a university examinations system: A "Turing Test" case study. PLOS ONE 2024

● Challenges 4/4

We've seen this movie before

THEN: PLAGIARISM

1998

Google launches

Other people's texts are seconds away. Copying becomes trivial.

2009

PlagScan launches

The technology is there, affordable and working. Adoption: low.

2011

The plagiarism scandal

A minister resigns, then a series of cases follows. Public pressure changes the situation.

2018

Checks become standard

Seven years after the scandal, the solution is in standard use.

TODAY: AI

2022

ChatGPT

The finished text loses its value as evidence: one pass through a paraphrasing tool takes an 86% match down to zero.

2023

Mentafy Writing Process Analysis

Same founding team, same situation as in 2009: proven technology, adoption just beginning.

2026

TEQSA

Australia's regulator TEQSA names process traces as evidence of learning. Universities are in a similar position to 2011.

In 2011 it took a scandal to prompt action. This time the trigger is already on the table, as a draft law.

② Mentafy's solution

● Existing solution improved

Plagiarism checking that detects paraphrases and translations

An innovative addition to the traditional check: Our **S³ Semantic Source Search** compares meaning instead of word strings and finds the source even when the text has been reworded or translated.

Traditional plagiarism check	Mentafy S ³
Finds only verbatim matches	Detects semantically identical passages
Fails on reworded texts	Analyzes meaning and line of argument
Report without sources for paraphrases	Direct comparison: submission vs. source
A single score as the basis for judgment	Several independent indicators and evidence

SUBMISSION (SECTION 3.2)

"The way individuals **selectively absorb information** to reinforce existing beliefs **has a significant impact on decision-making** processes."

↑ S³ detects semantic match

SOURCE: MILLER, J. (2023), P. 47

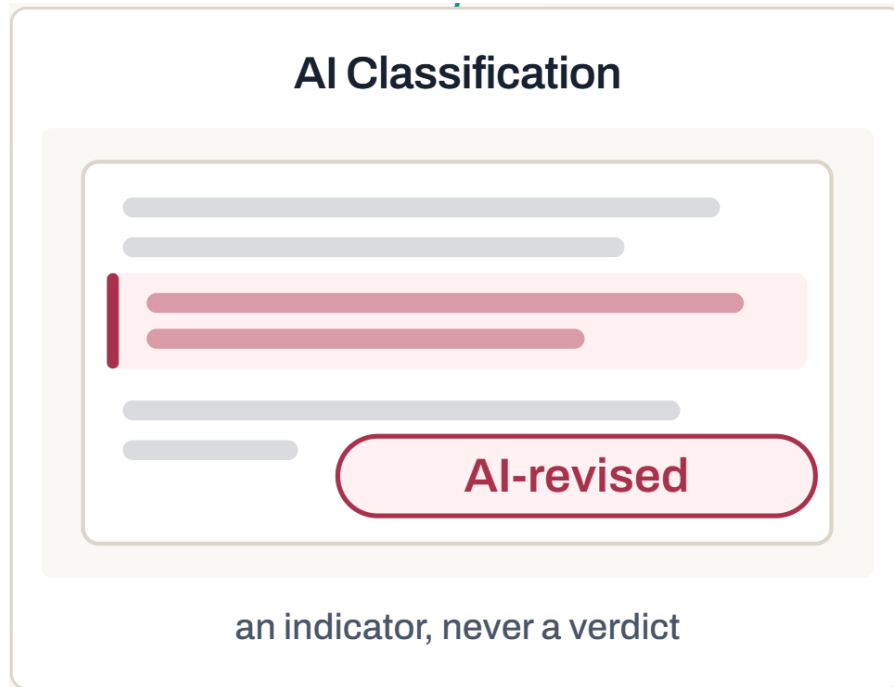
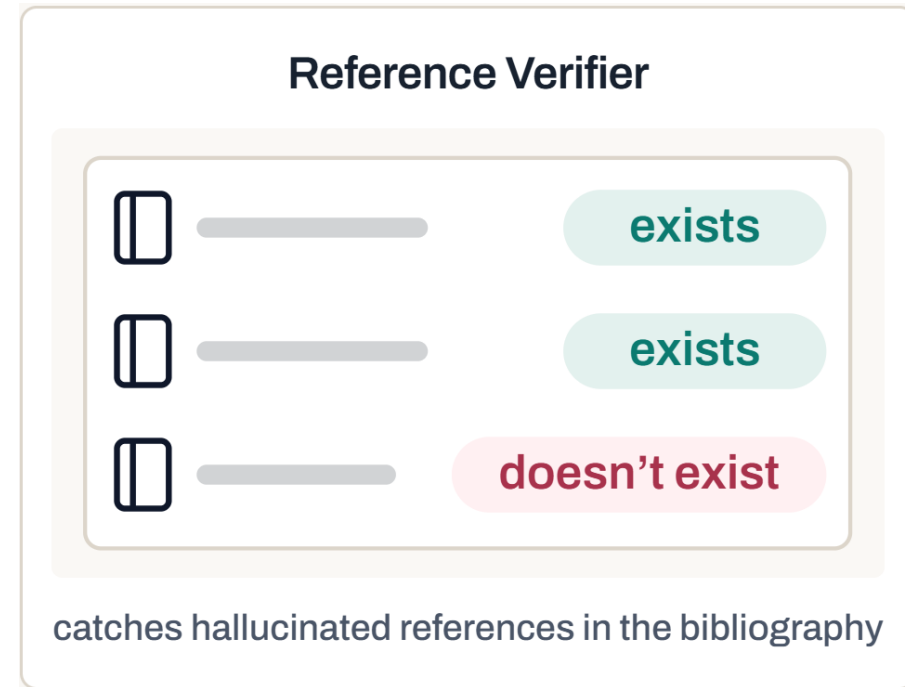
"**Confirmation bias** describes the tendency to **preferentially perceive information** that confirms existing assumptions, which systematically distorts **judgment and decision-making**."

87 %

Semantic similarity · Source citation generated

● Modules added

AI classification and Reference Verifier



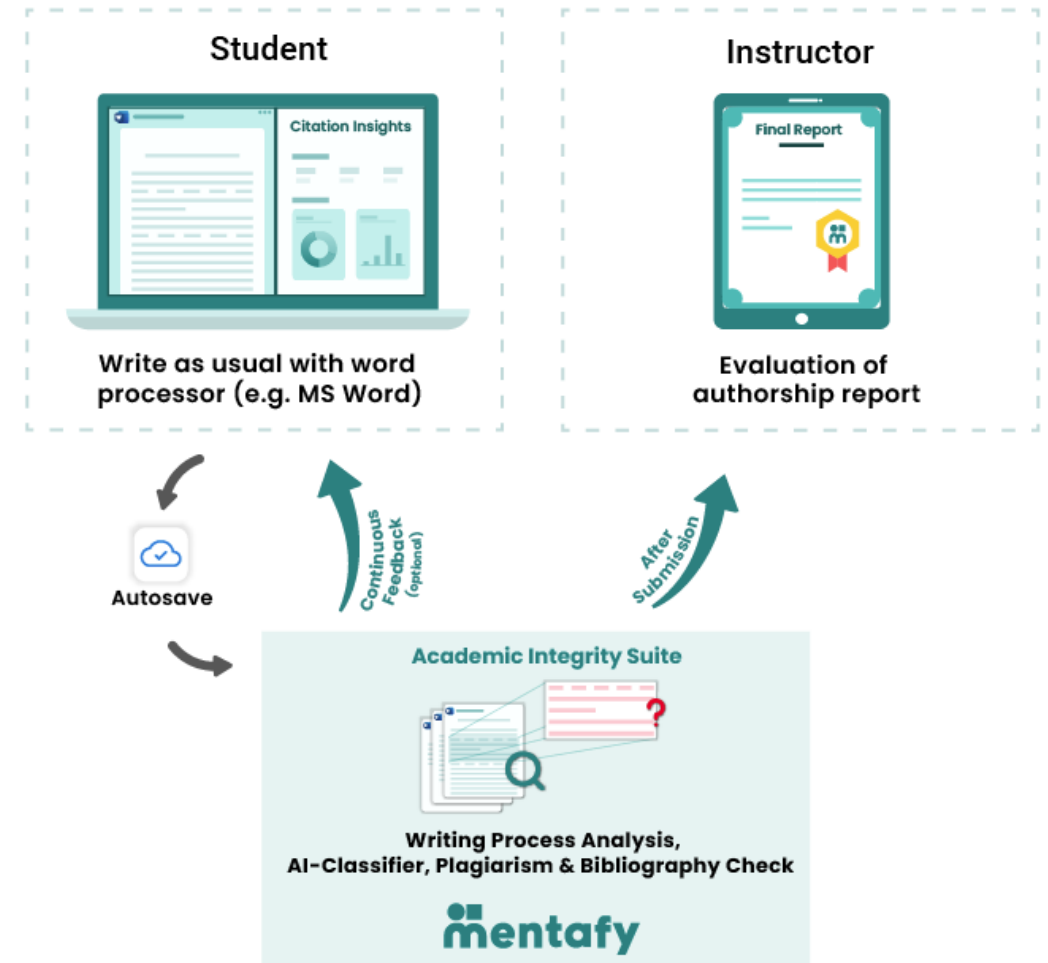
● New key technology

Our Writing Process Analysis

- ✓ Reliably distinguishes copy-and-paste and retyping from text students came up with themselves
- ✓ Can also give authors real-time feedback on working correctly (citing sources)

„The writing process itself is as important (if not more important) than the end product, because it is during the writing stage that the student will learn the most.“

Clare Johnson, Understanding academic integrity and plagiarism in the digital age:
Can **digital forensics** techniques help prevent and detect academic misconduct?
(2023), PQDT-Global



● The solution at a glance

Solutions for today and tomorrow from a single provider

Dimension	Mentafy Check (ready today)	Mentafy Trail (amend assessment policy per TEQSA)
Approach	Checks the finished document. A drop-in replacement for your current checker.	Documents how the work was written: the evidence regulation is moving toward.
Evidence	Finds plagiarism and citation errors; otherwise relies only on AI probabilities	Delivers tamper-proof evidence of students' own work: traditional check + semantics (S ³) and process
Data and documentation frequency	The finished document is checked. Everything works just like your process with plagiarism software.	You decide: automatically about once a minute, or manually. Even one version per week over a semester shows how a piece of work grew.
Analyses	Plagiarism matching, S ³ Semantic Source Search for paraphrases and translations, Reference Verifier, and AI classification optimized for German texts, among others.	Everything in Mentafy Check , plus Research Recorder and Writing Process Analysis. Students can get real-time feedback on their citations.

A combination of semantic matching (S³), automated citation validation and a transparent writing history (Version Scan and typing logs).

● The solution at a glance

Different report levels for efficient review



LEVEL 1

The traffic light

One rating. Green means no cause for concern, yellow means take a closer look, red means make sure to review it. That's all it says.

For initial screening



LEVEL 2

The one-pager

Shares of plagiarism, AI plagiarism, and quotations, the number of faulty references, and the work pattern: written over weeks, or in two days.

For grading



LEVEL 3

The full text

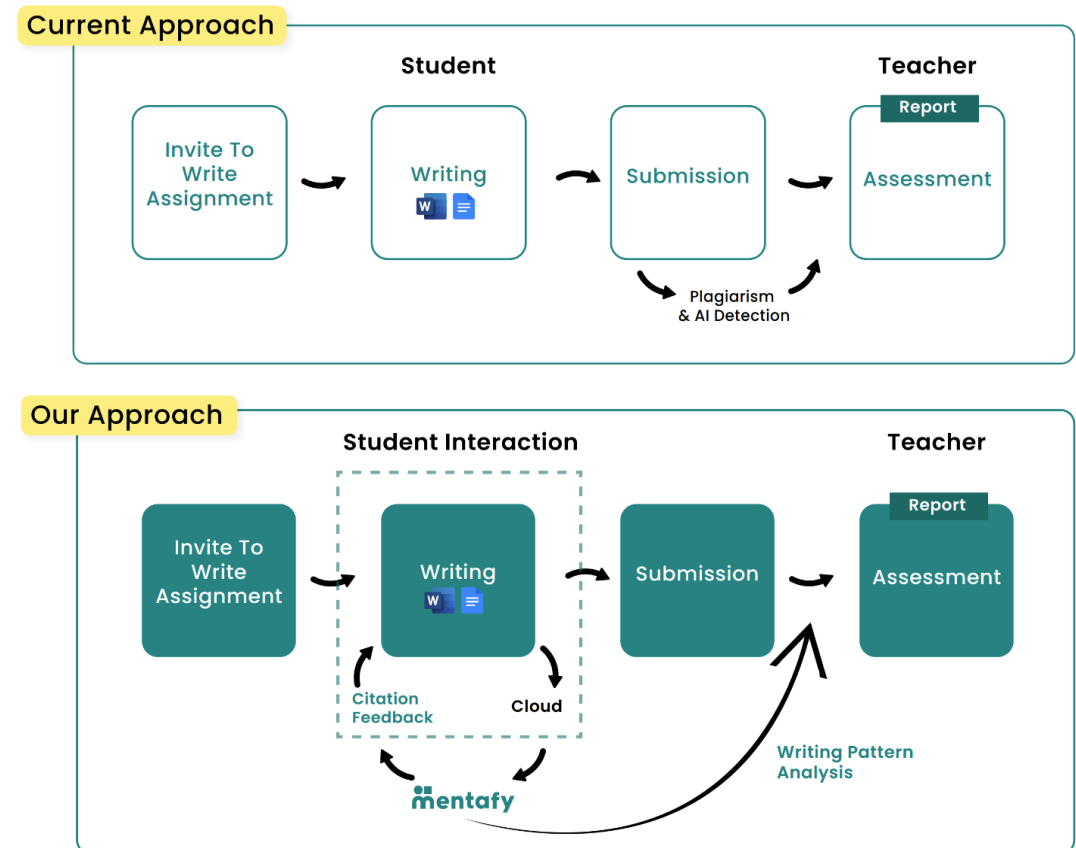
The text with all process data, the sources in full text with links, the AI classification results, and every faulty reference.

For disputes

● Assessment process

Only minimal changes to the workflow needed: students write in their preferred word processor

- Mentafy still offers this way of checking, i.e. plagiarism checking with AI detection on the finished document (e.g. PDF).
- We recommend working with a version history in this process, or requiring it to be submitted, to substantially enrich the data available.
- If we adapt the assessment process so we can take part in real time while the text is being written, we gain these benefits:
 - Optional: students can get real-time feedback to correct mistakes as they write
 - Transparency on source use, AI revision and fully generated text.



● Authorship Report

Compact report formats for instructors and students

The report answers the following questions:

- How much time was invested, and how is it distributed?
- How much research was done, and is the bibliography authentic?
- How much text was actually composed at the keyboard?
 - Quotations vs. plagiarism vs. ghostwriting vs. AI-generated/AI-revised

→ Evidence that holds up before examination boards instead of vague “black-box” percentages

→ Results can always be shown and checked directly in the text

General

2385

Word count

2024-03-04 - 2024-04-05

Duration of Project

9.17

Hours spent writing

Research

34

Notes

10

Total number of files

4

Webpages

21

References

Writing Statistics

72%

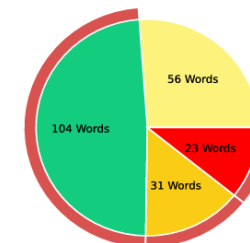
Flagged Text

Text flagged due to unusual input patterns (red border)

The input patterns captured during the preparation of this document are unusual and require further investigation.

Further 26% unusual patterns were auto-resolved.

Sources classification



- Categorized as cited (auto-detected)
- Categorized as AI generated
- Unresolved Web Sources
- Unresolved

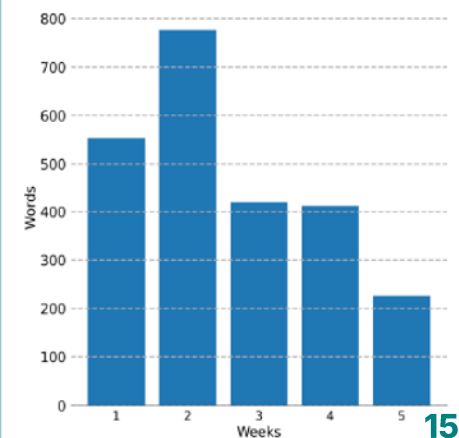
15

Words per sequence

on average in the Document

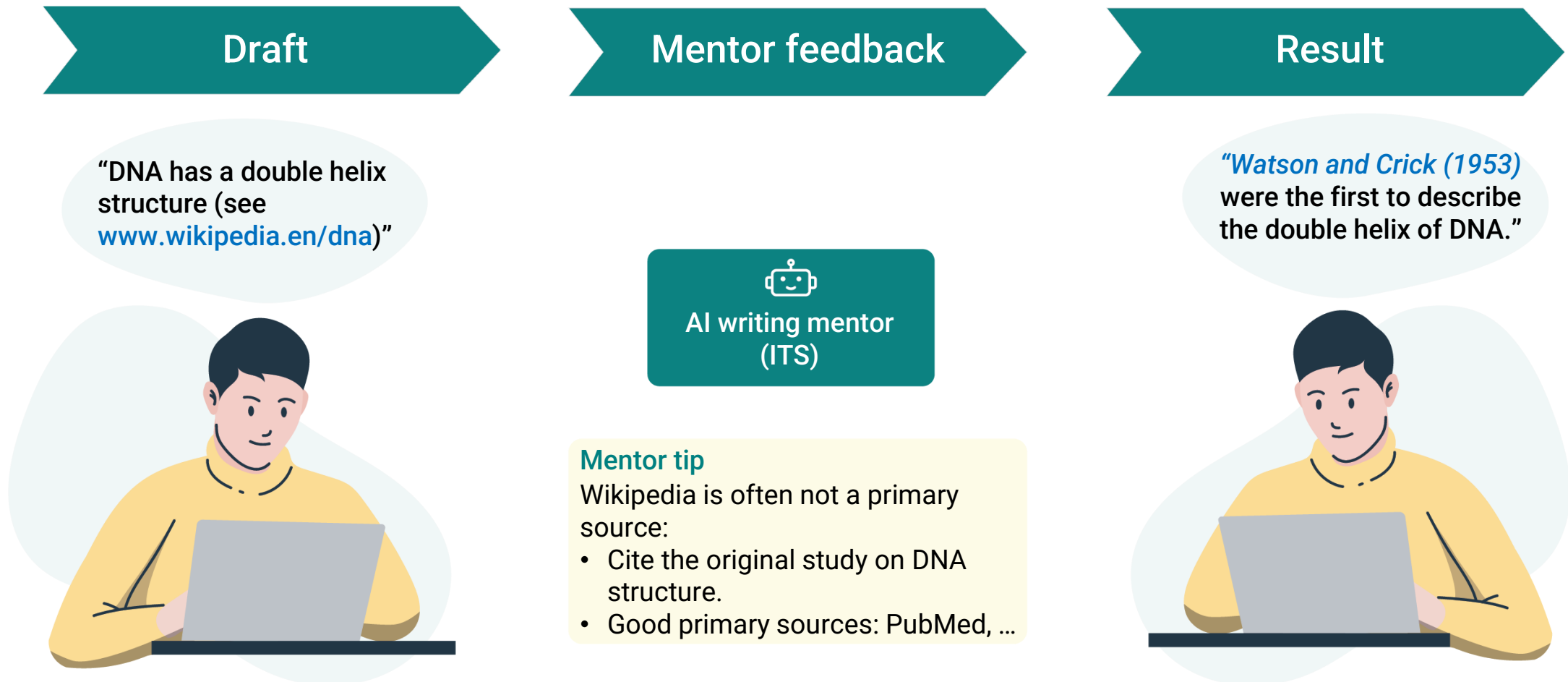
The average is slightly higher than expected, but there may be good reasons for this - further investigation is recommended.

Text Creation Timeline



● Writing mentor: an outlook

An intelligent tutoring system (ITS) that supports learning instead of policing it – beta MVP in Q1 2027



● Flexible implementation

Options for process documentation – you choose the depth

COMPLETE

Full cloud integration

Automated analysis of writing and research processes, integrated directly into Google Drive or OneDrive.

For institutions that need a comprehensive analysis of writing and research integrity.

DOCUMENTATION FREQUENCY

Your choice

- uploaded by students (all formats supported, as with Check)
 - Weekly
 - Daily
 - Student chooses
- via add-in

For maximum local data protection and minimal impact on IT infrastructure.

FINISHED DOCUMENT

Retrospective analysis

After submission: analyzing the version history gives retrospective insight into the writing process.

Note: if no version history is available, traditional analyses apply:

- Plagiarism check
- AI classification
- Reference Verifier

For all text formats; analysis of the finished document and versions, if available

How does Mentafy compare with the competition?

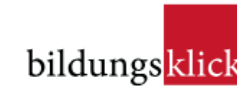
Reliably assessing students' own work

	MENTAFY	TURNITIN	GRAMMARLY	GPTZERO	PANGRAM	PLAGAWARE
AI transparency	✓	✓ (low evidence)	✓ (low evidence)	✓ (low evidence)	✓ (low evidence)	✓ (low evidence)
Separating AI revision from AI generation	✓	-	-	-	-	-
Evidence from version history	✓	✓ (announced for Q4 2025)	-	✓ (in beta)	-	-
Servers in Germany	✓	-	-	-	-	✓
Research Recorder	✓	-	-	-	-	-

● Pilot successfully completed in 2025



Hochschule Konstanz
Technik, Wirtschaft und Gestaltung



Teaching

- Superior, innovative technology (S³ + writing process)
- Minimal effort (just write in Word, LibreOffice, etc.)
- **No false accusations of AI misuse**



Procurement

- Full price transparency: **from €1 per student per year**
- No unpredictable module costs, and easy to purchase at department level too.
- All-in-one: a complete academic integrity suite.



Administration

- **GDPR-compliant:** designed for data minimization, servers in Germany
- Assessment policy changes just like for plagiarism checks
- Easy integration (e.g. single sign-on, LMS)



● Founding team

Experienced founders in academic integrity

We spotted the copy-and-paste plagiarism problem early, built PlagScan and helped build Ouriginal



Markus Goldbach

Supplied plagiarism software to 2,000+ institutions worldwide.

PlagScan Ouriginal



LinkedIn



Johannes Knabe, PhD

Built the award-winning plagiarism software PlagScan and was CTO at Ouriginal.

PlagScan Ouriginal



LinkedIn

Start a no-obligation pilot in your faculty or department

Markus Goldbach

CEO

Markus.Goldbach@mentafy.com

+49 175 5888661

Gottfried-Hagen-Str. 60-62
51105 Cologne, Germany



[Book a webinar now](#)



QUESTIONS & ANSWERS

Limited access

- Complete solution: access to one dedicated folder in the cloud
- Add-in: real-time access to exactly one document
- Version history: access to exactly one document
- Access is used solely to create the Authorship Report.

Folder management

- Mentafy creates a separate folder for writing projects.
- Read access: limited entirely to this folder
- Uploaded materials are stored only in this folder.

Data protection and rights:

- Mentafy keeps access to a minimum, as set out in the data processing agreement (DPA), and needs no further administrative rights.
- Students' privacy is fully protected.

**What access rights does
Mentafy receive at
registration?**

Handwritten preparation

- Mentafy is designed for digital processes. Notes and handwritten drafts are of course a proven method; we recommend uploading them alongside (a mobile app is available) or recording them in the Writing Journal.

Auto-save turned off

- If auto-save is turned off, the detailed version history is missing. Instructors can make this clear: auto-save is the default and should stay on.

Add-in not activated

- With this option, no documentation is created without the add-in.

→ Signals showing whether documentation is set up correctly can be viewed in the LMS or in Mentafy's assignment management.

**What happens if
students make mistakes
using Mentafy and the
documentation is
incomplete as a result?**

GDPR compliance

- Full compliance with the General Data Protection Regulation (GDPR).
 - We store only the data that is strictly necessary
 - We are 100% transparent about what we store
 - All data is always deleted as agreed
- Our servers are hosted in Germany by Hetzner Online AG.
- We sign a data processing agreement (DPA).
- We regularly run comprehensive risk analyses to identify and minimize potential vulnerabilities.

**How do you ensure
compliance with data
protection rules?**

Evidence instead of suspicion

- Transparency tools secure traceable evidence (sources, history). They don't assume anything; they document.

Fairness for honest students

- Without checks, cheating pays off. Consistent procedures protect students' own work and ensure fair grades.

Legal and procedural certainty

- Verifiable criteria allow for a hearing, counter-evidence and appeal – more sound than gut feeling or black-box verdicts.
- Legal strength doesn't come from "the one score", but from corroborated evidence (process data, text matches and AI indicators support each other) with a verifiable derivation. This turns suspicion into solid evidence – or, conversely, clears the student.

**By using tools like
Mentafy, am I not putting
every student under
suspicion?**

Transparent assessments

- Turnitin and others rely on unreliable AI pattern recognition.
- Mentafy delivers verifiable, data-based evidence through comprehensive Writing Process Analysis
- Only our approach can tell AI revision apart from AI-ghostwriting

Tamper-proof Writing Journal

- Students create it in a standard format with Mentafy
- Instructors get clear insight into students' own work (including their research).

Support instead of control

- Focus on active support through source tips and project management.
- The goal is to foster independence rather than primarily to detect misconduct.
- No locked-down environment: students write in Word and similar programs.

Focus on data protection

- No keyloggers or invasive methods.
- Deleted text stays deleted with us (unlike, for example, the version history that Google and Microsoft keep)

**How is Mentafy
different from the
competition?**

Strong interest

- Demand for a solution like Mentafy is huge, and feedback confirms its relevance in education. 100+ organizations have checked 10,000+ documents.

Integration challenges

- Educational institutions find it hard to switch to approaches that accompany the writing.
- Institutions with Microsoft or Google licenses can start right away.
- For other providers, we develop custom solutions.

Easy-to-read reports

- The Mentafy report is intuitive and takes minimal effort.
- In-depth analysis is rarely needed, as the report usually provides enough information.
- If needed, we help with detailed questions.

**How have existing
users received the
product?**