

Plagiarism Detection Software: Can We Really Trust Them?



Merazga Batoul *

University Mohamed Lamine Debaghine (Algeria), Clinical psychology
laboratory

ba.merazga@univ-setif2.dz

Besbes Fatima Zohra

University Mohamed Lamine Debaghine (Algeria), Laboratory of
Psychopathology and Neuropsychology

fa.besbes@univ-setif2.dz

Date of submission: 15/09/2025 Date of acceptance: 09/12/2025 Date of publication: 16/12/2025



Abstract:

Plagiarism detection software has become a cornerstone of academic integrity enforcement across the globe. However, behind the sleek interfaces and statistical reports lies a more complex reality: one that raises profound philosophical, linguistic, and ethical questions. This article critically examines the mechanisms and logic of popular tools such as Turnitin, iThenticate, and others, questioning the assumption that textual similarity necessarily implies intellectual theft. Are ideas truly ownable? Can an algorithm distinguish between imitation and coincidence, style and structure, knowledge and noise?

As institutions increasingly rely on these systems not only for evaluation but also for control, we investigate whether such tools foster integrity—or whether they reinforce a surveillance-based model of academia, shaped by profit motives and algorithmic opacity. Furthermore, we raise concerns about the ownership of the data fed into these tools, and whether researchers' works are quietly contributing to AI training and commercial exploitation without consent. Rather than rejecting these technologies outright, the article calls for a rethinking of the values and policies behind their use. It argues for more transparent, context-sensitive, and ethically grounded approaches to academic authorship in the digital age.

* Merazga Batoul

key words: Plagiarism detection software, Academic authorship, Academic surveillance, Data ethics, Turnitin, iThenticate.

Introduction:

In contemporary academic practice, plagiarism detection software has become almost synonymous with research quality assurance. These tools are widely endorsed by universities, journals, and funding bodies, serving as automated gatekeepers of what is deemed “original.” Yet, behind their sleek dashboards and statistical similarity reports lies a deeper and more unsettling question: Can we truly trust machines to judge the authenticity of human thought?

Plagiarism is commonly defined as an unethical practice in which an individual presents the ideas or work of others as their own original production (Flint et al., 2006). It remains one of the most serious and complex challenges facing academic institutions worldwide (Park, 2004). In response, software such as Turnitin, iThenticate, PlagScan, and Urkund has emerged as the technological answer—designed to detect similarity and assign risk-based scores to academic submissions (Bretag & Mahmud, 2016; Newton, 2016). These tools are increasingly treated as the first line of defense in academic publishing and assessment.

As a result, many researchers—especially emerging scholars—invest significant effort into adjusting wording, rephrasing content, or modifying structure to avoid being flagged for similarity, even when their sources are properly cited (Flint et al., 2006). Yet few fully grasp how these systems function, the proprietary algorithms behind them, or the epistemological assumptions embedded in their design.

But the real issue may lie deeper. In a world of standardized academic writing and constrained linguistic variation, the recurrence of ideas, expressions, and reasoning patterns is not only likely but inevitable. Does such recurrence always signify plagiarism? Or are we confusing methodological convergence with ethical misconduct?

Philosopher Alain Deneault (2018) warns that academia is at risk of shifting from a space of knowledge production to one of intellectual recycling, where power dynamics allow senior figures to appropriate the work of their students without accountability. In this context, plagiarism detection software may paradoxically reinforce structural injustice—perpetuating surveillance, cultivating mistrust, and reducing intellectual labor to algorithmic compliance.

This article does not aim to discredit technology or academic institutions. Rather, it opens a deeper dialogue: What are we really detecting when we talk about plagiarism? What values underlie the tools

we use? And how might we reimagine authorship, originality, and trust in the digital academy?

1-Literature Review

The expansion of plagiarism detection software in academic settings has inspired a substantial body of research over the last two decades. Much of this literature has focused on improving the technical performance of these systems—evaluating their ability to detect textual matches, comparing detection accuracy, and analyzing their limitations in multilingual or paraphrased contexts (Foltynek et al., 2020; Chowdhury & Bhattacharyya, 2021; Pecorari, 2008). Parallel studies have explored pedagogical applications, emphasizing the need to teach students about citation practices, paraphrasing, and academic writing conventions (Flint et al., 2006; Bretag, 2016; Sivasubramaniam et al., 2020).

However, despite this technical and instructional focus, there remains a noticeable lack of critical inquiry into the broader ethical, political, and institutional implications of these tools. Few studies have examined how these systems operate as mechanisms of control within academic structures, or how they may affect the rights and autonomy of authors (Lipson & Reindl, 2017; Stelzer, 2021). Even fewer have explored the opacity of their algorithms, the commercial interests that shape their design, or the long-term consequences of outsourcing academic judgment to private corporations.

Moreover, the potential role of these platforms in retaining user data and contributing to commercial or AI-based models is still largely unregulated and under-researched (Kizilcec, 2020; ICAC, 2020). Questions about consent, transparency, and data exploitation remain unanswered.

This article addresses that gap. It shifts the focus from performance metrics to the ethical foundations of academic assessment, offering an interdisciplinary reflection on what plagiarism detection systems reveal about our assumptions regarding trust, authorship, and institutional power.

2-What Are the Most Widely Used Plagiarism Detection Tools—and Who Controls Them?

As plagiarism detection software becomes increasingly integrated into academic workflows, it is important to understand the origins, mechanisms, and ownership structures of these systems. Despite their

technical sophistication, they are not neutral tools. Each one reflects the priorities of the institutions, companies, and cultures that created them.

Table (1) provides a comparative overview of some of the most prominent plagiarism detection tools used in higher education and scientific publishing. It outlines the year of creation, country of origin, developer or owning entity, core detection mechanisms, and their key strengths and weaknesses.

Key Limitations	Key Strengths	Detection Method	Developer / Owner	Country	Year Established	Tool Name
High false positives; lacks algorithmic transparency	Large proprietary database; widely adopted	Matches against internal DB, web, journals	iParadigms, LLC	United States	1998	Turnitin
Not accessible to students; high subscription cost	Trusted by major publishers; detailed reports	Similarity scan of scientific content	iParadigms (for publishers)	United States	2004	iThenticate
Less effective in scientific content	Simple interface; visual similarity reports	Matches with open web + private documents	PlagScan GmbH	Germany	2009	PlagScan
Limited global presence; less familiar to individuals	Transparent process; strong in European universities	AI-based matching with educational sources	Ouriginal AB (merged entity)	Sweden	2000	Urkund (Ouriginal)
Not built for academic depth; limited scholarly databases	Freemium model; easy to use	DeepSearch™ proprietary engine	Quetext, Inc. (private)	United States	2016	Quetext
Not academic-focused;	Integrated with writing	Internet-based similarity	Grammarly Inc.	Ukraine / US	2009	Grammarly (Plagiari

weak on journal/publis her coverage	tools; intuitive interface	check				sm Tool)
---	----------------------------------	-------	--	--	--	----------

Table 1. Comparative Analysis of Major Plagiarism Detection Tools: Origins, Mechanisms, Strengths, and Limitations

Following this overview, several analytical observations emerge:

While Turnitin and iThenticate dominate academic environments due to their integration with publisher databases and institutional platforms, other tools like PlagScan or Quetext serve more general or educational uses. Most tools rely on surface-level textual matching, rather than deep semantic or stylistic analysis. Additionally, the proprietary nature of these systems means that users rarely understand how the similarity score is calculated, or how their documents are stored and used after submission.

The lack of independent international audits for these systems raises ethical questions about fairness and reliability. As these tools become increasingly influential in defining what constitutes legitimate scholarship, it becomes necessary to interrogate not only how they work, but also who benefits from their widespread adoption

3-Is Every Linguistic Similarity a Sign of Plagiarism?

One of the most problematic assumptions embedded in plagiarism detection systems is the idea that textual similarity automatically implies academic dishonesty. But is it fair—or even intellectually sound—to reduce the complex act of writing to a matter of matching phrases? In reality, the boundaries between inspiration, intertextuality, and imitation are far more fluid than these systems allow (Pecorari, 2008; Chand, 2020).

In scientific and academic writing, certain linguistic constructions are not only common but necessary. Writers frequently use shared cognitive tools such as comparison, explanation, classification, and evaluation. These mental operations often give rise to similar sentence structures and expressions across unrelated works (Flint et al., 2006). For instance, when a researcher describes a well-established phenomenon, such as "the sun rises in the east," the possibilities for paraphrasing are inherently limited. Reversing the sentence structure or changing a few synonyms does not alter the fact that the underlying truth—and its linguistic representation—are constrained by the nature of the knowledge itself (Pecorari, 2008).

This raises a central question: if two scholars independently arrive at similar phrasing while analyzing the same concept, is that evidence of plagiarism or simply a reflection of disciplinary thinking? When language is treated as a closed system of limited variations, the margin for what is considered “original” becomes dangerously narrow.

Thus, similarity alone cannot and should not be equated with plagiarism. Doing so risks criminalizing legitimate scholarly expression and imposing an unrealistic standard of linguistic uniqueness in fields where shared terminology and concepts are intrinsic to the discourse (Sivasubramaniam et al., 2020; Chowdhury & Bhattacharyya, 2021).

4- Can a Thought Be Stolen? Or Is Intellectual Ownership a Modern Illusion?

Much of the logic behind plagiarism detection rests on the assumption that ideas are personal, ownable, and traceable. But is this notion philosophically valid? Can a thought—an abstract cognitive act—truly be “stolen,” or is what we call “intellectual property” merely a social construct rooted in contemporary academic systems?

From a philosophical standpoint, ideas often emerge in shared contexts. Human cognition draws upon collective memory, cultural frameworks, and epistemic conventions. In many cases, researchers working independently—yet within the same disciplinary boundaries—arrive at similar observations, hypotheses, and conclusions. This phenomenon is not necessarily evidence of theft, but of parallel reasoning (Barthes, 1968; Foucault, 1969).

Moreover, the production of academic texts is shaped by standardized methods of argumentation, such as classification, synthesis, interpretation, and critique. These methods generate not only converging thoughts but converging expressions. If two authors describe a phenomenon using the same conceptual structure, is that a duplication of thought or a natural byproduct of shared scientific language?

What truly distinguishes one researcher from another is not the originality of the concept itself, but the unique manner in which it is expressed—style, tone, structure, and rhetorical approach. Yet most plagiarism detection software is blind to these dimensions. It scans the visible form, but not the intellectual fingerprint embedded in expression. Consequently, it may conflate intellectual inspiration with intentional imitation, and thus penalize honest thinkers for engaging in a communal epistemic process.

In this light, we must question whether current definitions of plagiarism are too narrow and too rigid—particularly when they are enforced by tools that cannot comprehend the nuance of human thought. Academic writing is not created in a vacuum. It is the product of countless dialogues, intertextual references, and disciplinary constraints. To assume that every resemblance is unethical is to deny the collaborative nature of knowledge itself.

5- Do These Tools Protect Us—or Enforce Hidden Surveillance?

While plagiarism detection software is often promoted as a guardian of academic honesty, its growing role within institutional ecosystems reveals another, more troubling dimension: surveillance. What began as a quality-control mechanism has gradually evolved into a disciplinary apparatus—one that may prioritize monitoring over mentorship, suspicion over trust.

These systems are rarely optional. In many universities, students and researchers are required to submit their work through a designated platform, often without full awareness of how their data is stored, analyzed, or possibly reused. This mandatory compliance, coupled with opaque algorithms, shifts the focus from developing ethical writing practices to simply avoiding being flagged (Friend & Koehler, 2009; Macdonald & Carroll, 2006).

Such dynamics raise ethical concerns about autonomy and consent. When writers must anticipate and accommodate the preferences of a machine to avoid punitive consequences, academic writing ceases to be an act of intellectual freedom. Instead, it becomes a performance calibrated to satisfy algorithmic standards—often at the expense of depth, creativity, and risk-taking in thought.

Moreover, these tools rarely distinguish between types of similarity—such as self-citation, disciplinary conventions, or common technical phrases. Their binary logic does not allow for nuance, and as a result, human judgment is replaced by numerical thresholds. Reports are interpreted as absolute verdicts, rather than prompts for contextual evaluation (Sutherland-Smith, 2010; Bretag, 2016).

This shift has profound implications. It reframes students not as developing scholars, but as potential offenders. It transforms the teacher from a mentor into a gatekeeper, and the institution into a monitoring system. Under the guise of promoting integrity, a culture of suspicion

may take hold—one that undermines trust, creativity, and the possibility of open intellectual exploration.

6- Has Plagiarism Detection Become a Profit-Driven Industry?

What began as a response to academic dishonesty has gradually transformed into a lucrative commercial model. The proliferation of plagiarism detection tools, particularly in English-speaking academia, has generated a competitive marketplace in which private companies—often operating with little external regulation—profit from licensing their algorithms, subscriptions, and institutional contracts.

Turnitin, for example, is owned by Advance Publications, a major media conglomerate. Its services are used by over 15,000 institutions in more than 140 countries, and it maintains a vast database of student papers, journal articles, and internet sources. iThenticate, a tool under the same corporate umbrella, is used by most of the world's leading academic publishers. This consolidation of market power raises serious concerns about monopolization, data ownership, and the privatization of academic oversight (Lipson & Reindl, 2017; Stelzer, 2021).

Furthermore, users—particularly students—often submit their work to these systems without knowing that their texts may be retained indefinitely and used to expand the companies' internal databases. This raises difficult questions: Who owns the content once it is uploaded? What rights do authors retain over their intellectual labor? And can private companies ethically profit from unpaid academic contributions?

Some critics argue that plagiarism detection software represents a form of extractive capitalism in academia. While marketed as tools for preserving integrity, these platforms also commodify trust itself—packaging it as a purchasable service, rather than a pedagogical value cultivated within the academic community (Eaton, 2020; Lipson & Reindl, 2017).

In this context, plagiarism detection may no longer be merely about ethics. It is also about economics—about who controls the tools of evaluation, who benefits from their widespread adoption, and who pays the price when integrity becomes a product.

7- Are Researchers' Texts Being Used to Train AI Without Their Knowledge?

In the age of artificial intelligence, data has become one of the most valuable resources in the world. Within academia, this raises a new and urgent concern: Are the texts submitted to plagiarism detection systems being silently used to train AI models, without the knowledge or consent of their authors?

Most commercial plagiarism detection tools operate with closed and proprietary databases. When researchers or students submit their work, it is often stored indefinitely—ostensibly to improve future comparisons. But where does this data actually go? And what guarantees exist to prevent its reuse for purposes beyond plagiarism detection?

Some scholars have begun to warn that the massive corpora collected by systems like Turnitin and iThenticate may one day be used to train machine learning models in academic writing, text classification, or even automated authorship scoring. These possibilities, though still speculative, are plausible given the economic value of such data and the current opacity surrounding its use (Kizilcec, 2020; ICAC, 2020).

This raises profound ethical questions about data ownership, academic autonomy, and digital consent. If a student's thesis is uploaded to a detection system and then later absorbed into an AI model without their knowledge, does this constitute exploitation? Have we, under the guise of ensuring academic integrity, created a mechanism for quietly extracting intellectual labor?

The paradox becomes clear: researchers are required to submit their work for plagiarism screening in the name of transparency and integrity—but they are rarely granted transparency in return. There is little clarity about how long their work is retained, who can access it, or whether it becomes part of commercial datasets used for AI development.

These concerns do not imply conspiracy, but they do demand oversight. As the boundaries between educational tools and commercial data mining blur, academic institutions must demand greater transparency, legal safeguards, and ethical standards. Trust in plagiarism detection cannot be blind—it must be reciprocal.

8- What Do We Need to Build a More Ethical Academic System?

This article does not seek to demonize plagiarism detection software, nor to undermine the institutions that have adopted such tools in good faith. Instead, it raises a set of critical, yet unresolved, questions that demand attention in a rapidly evolving academic landscape.

The concerns discussed here—about linguistic standardization, epistemic ownership, institutional surveillance, data exploitation, and algorithmic opacity—do not stem from hostility to technology. They emerge from a deep commitment to academic freedom, transparency, and the cultivation of genuine scholarly integrity.

Plagiarism detection tools can be helpful when used as diagnostic supports, guided by human judgment and ethical nuance. But they must not be allowed to operate as unquestionable authorities. A similarity score is not a moral verdict, and an algorithm is not a conscience. As such, we must resist the temptation to outsource academic evaluation to machines whose logic remains inaccessible and whose motivations may not align with the values of education.

This paper calls not for rejection, but for reform. For universities and publishers to develop clearer policies around consent, data retention, and usage rights. For researchers to be treated not as suspects, but as intellectual partners. And for institutions to invest in tools that recognize not just textual patterns, but the distinctiveness of human voice, reasoning, and intention.

Ultimately, we must ask: What kind of academic culture are we building when we measure integrity through algorithms? The answer lies not in abandoning technology, but in reclaiming control over how it is used—and for whose benefit.

Recommendations:

The issues raised throughout this paper highlight the need for a more balanced and ethically grounded approach to evaluating academic integrity. Rather than relying solely on automated systems, institutions must foster human-centered policies that respect authorship, preserve intellectual autonomy, and ensure technological accountability. The following recommendations aim to support such a transformation:

- Re-evaluate institutional policies governing the use of plagiarism detection software to ensure transparency, informed consent, and respect for author rights.

- Invest in the development of open-source, ethically designed tools under academic governance, with clear oversight and accountability mechanisms.

-Incorporate education on authorship ethics and critical digital literacy into curricula, helping students understand both the risks and responsibilities of scholarly writing.

-Acknowledge the limitations of similarity-based detection, and promote the use of human review—especially in borderline cases where interpretation is required.

-Encourage international dialogue between universities, publishers, and researchers to standardize ethical frameworks for evaluating originality without compromising trust or autonomy.

Conclusion

This paper has sought not to deliver judgments, but to raise questions—questions that are increasingly urgent in an academic culture that leans heavily on algorithmic tools to assess integrity. The widespread use of plagiarism detection software, while initially intended to preserve the values of originality and authorship, has brought with it a new set of ethical, epistemological, and political concerns.

When similarity becomes synonymous with guilt, and when authors are evaluated not by human discernment but by opaque algorithms, we must pause and ask: What are we really protecting? And at what cost?

The true value of academic work lies not only in its novelty, but in its engagement with knowledge, its reasoning, its style, and its contribution to collective understanding. Reducing this richness to numerical scores risks flattening scholarly expression and suppressing intellectual diversity.

This is not a call to reject technology. It is a call to treat technology with the same critical rigor that academia demands of everything else. We must not mistake automation for objectivity, nor convenience for justice. If we are to foster a truly ethical academic system, we must ensure that our tools serve scholarship—not the other way around.

Ultimately, this is not a technical issue, but a philosophical one. What kind of academic world do we wish to build? One governed by trust, dialogue, and nuance—or one shaped by metrics, suspicion, and control?

Bibliography List :

Books

- Barthes, R. *Writing Degree Zero*. Hill and Wang, 1968.
- Deneault, A. *Mediocracy: The Politics of the Extreme Centre* (C. Browne, Trans.). Between the Lines, 2018. (Original work published 2015).
- Eaton, S. E. *Academic Integrity in Canada: An Enduring and Essential Challenge*. Springer, 2020. <https://doi.org/10.1007/978-3-030-58790-6>
- McCabe, D. L. *Cheating in College: Why Students Do It and What Educators Can Do About It*. Johns Hopkins University Press, 2022.
- Pecorari, D. *Academic Writing and Plagiarism: A Linguistic Analysis*. Continuum, 2008.

Journal Articles

- Ali, P. A. *Academic integrity and cultural perspectives: A view from the Middle East*. *International Journal for Educational Research*, 14(1), 2018, pp. 1–7. <https://doi.org/10.1007/s40979-018-0032-4>
- Bretag, T. *Challenges in addressing plagiarism in education*. *PLOS ONE*, 11(6), 2016. <https://doi.org/10.1371/journal.pone.0157540>
- Bretag, T., & Mahmud, S. *A conceptual framework for implementing exemplary academic integrity policy in higher education*. *Journal of Higher Education Policy and Management*, 38(4), 2016, pp. 406–419. <https://doi.org/10.1080/1360080X.2016.1181880>
- Chand, A. *Language variation and challenges in detecting academic plagiarism in multilingual contexts*. *Language and Education*, 34(3), 2020, pp. 217–234. <https://doi.org/10.1080/09500782.2020.1744520>
- Chowdhury, A., & Bhattacharyya, D. *A survey on plagiarism detection: A linguistic perspective*. *Journal of Information Science*, 47(6), 2021, pp. 824–840. <https://doi.org/10.1177/0165551519897915>
- Flint, A., Clegg, S., & Macdonald, R. *Exploring student engagement: A sociology of education perspective*. *Teaching in Higher Education*, 11(4), 2006, pp. 395–408. <https://doi.org/10.1080/13562510600874202>
- Foltynnek, T., Meuschke, N., & Gipp, B. *Academic plagiarism detection: A systematic literature review*. *ACM Computing Surveys*, 52(6), 2020, pp. 1–42. <https://doi.org/10.1145/3345317>
- Friend, C. L., & Koehler, D. *The ethics of Turnitin in the classroom: A student perspective*. *Teaching English in the Two-Year College*, 37(1), 2009, pp. 58–72.
- Haissam, N. *Ethics of scientific research in the age of artificial intelligence: ChatGPT application and the issue of plagiarism*. *Qabas Journal for Human and Social Studies*, 8(1), 2024, pp. 1164–1184. <https://asjp.cerist.dz/en/article/250780>
- Helgesson, G., & Eriksson, S. *Plagiarism in research: A survey of recent literature*. *Science and Engineering Ethics*, 28(3), 2022, pp. 1–17. <https://doi.org/10.1007/s11948-022-00353-2>
- Kizilcec, R. F. *How much information? Effects of transparency on trust in automated plagiarism detection*. *Computers & Education*, 148, 2020, Article 103791. <https://doi.org/10.1016/j.compedu.2019.103791>
- Laadjal, I. E., & Elhamza, M. *Informational pollution as a new challenge to scientific publishing: The case of predatory journals*. *Qabas Journal for Human and Social Studies*, 5(1), 2021, pp. 518–540. <https://asjp.cerist.dz/en/article/159703>

- Lipson, C., & Reindl, T. *Ownership, consent, and control: Plagiarism software and the rights of student authors. Ethics and Information Technology*, 19(3), 2017, pp. 217–226. <https://doi.org/10.1007/s10676-017-9432-y>
- Macdonald, R., & Carroll, J. *Plagiarism—a complex issue requiring a holistic institutional approach. Assessment & Evaluation in Higher Education*, 31(2), 2006, pp. 233–245. <https://doi.org/10.1080/02602930500262536>
- Meuschke, N., & Gipp, B. *State-of-the-art in detecting academic plagiarism. International Journal for Educational Integrity*, 9(1), 2013, pp. 50–71.
- Newton, P. M. *Academic integrity: A quantitative study of confidence and understanding in students. Studies in Higher Education*, 41(4), 2016, pp. 657–670. <https://doi.org/10.1080/03075079.2014.927850>
- Pupovac, V., & Fanelli, D. *Scientists admitting to plagiarism: A meta-analysis of surveys. Science and Engineering Ethics*, 21(6), 2015, pp. 1331–1352. <https://doi.org/10.1007/s11948-014-9600-6>
- Sedika, A. *The role of ChatGPT in enriching Islamic scientific research. Qabas Journal for Human and Social Studies*, 9(1), 2025, pp. 786–804. <https://asjp.cerist.dz/en/article/268989>
- Sivasubramaniam, S., Kostka, I., & Marzano, D. *Academic integrity in a digital age: Student voices. International Journal for Educational Integrity*, 16(1), 2020, pp. 1–15. <https://doi.org/10.1007/s40979-020-00059-7>
- Stelzer, R. *Plagiarism detection and the neoliberal university: A critical perspective. Journal of Academic Ethics*, 19(2), 2021, pp. 145–163. <https://doi.org/10.1007/s10805-021-09377-1>
- Sutherland-Smith, W. *Retribution, deterrence and reform: The dilemmas of plagiarism management in universities. Journal of Higher Education Policy and - Management*, 32(1), 2010, pp. 5–16. <https://doi.org/10.1080/13600800903440519>
- Walker, J. *Measuring plagiarism: Researching what students do, not what they say they do. Studies in Higher Education*, 35(1), 2010, pp. 41–59. <https://doi.org/10.1080/03075070902906717>

Internet Websites

- Trip Gabreil, *Plagiarism Lines Blur for Students in Digital Age*. 1 Aug, 2010 (consulted on 03/06/2025/10:00), <https://www.nytimes.com/2010/08/02/education/02cheat.html>
- ICAC – International Center for Academic Conduct. *Guidelines for Ethical Use of Plagiarism Detection Software* (consulted on 03/06/2025/10:00), <https://academicintegrity.org>