HAKEMLİ Denetişim
Ortak Aklın Harmanı
Kamu İç Denetçileri Derneği Meşrutiyet Caddesi Konur Sokak No: 36/6 Kızılay - ANKARA
www.kidder.org.tr/denetisim/ • denetisim@kidder.org.tr
ISSN 1308-8335
Yıl: 17, Sayı: 35, 438-458, 2026
Gönderilme Tarihi: 1 Haziran 2026 Kabul Tarihi: 10 Temmuz 2026

Research Article

AI-ASSISTED REWRITING AND IDEA LAUNDERING IN ACADEMIA: A PROOF-OF-CONCEPT STUDY FROM THE PERSPECTIVE OF PUBLICATION CONTROLS

(YAPAY ZEKÂ DESTEKLİ YENİDEN YAZIM VE AKADEMİDE FİKİR AKLAMA: YAYIN KONTROLLERİ AÇISINDAN BİR KAVRAM KANITI ÇALIŞMASI)

Yusuf Mert VELİOĞLU[1], Hakan VELİOĞLU[2], Selçuk OLUM[3], Ali Kasım ARKIN[4]

ABSTRACT

This study examines the extent to which the similarity reports and AI writing reports used in academic publishing can provide assurance against the risk of "idea laundering" that may be carried out through AI-assisted rewriting. Its central claim is that AI-assisted writing does not, in itself, constitute plagiarism; the ethical problem lies in rewriting another author's academic idea, argument, or contribution without attribution and presenting it as an original contribution. To this end, the discussion/conclusion sections of twelve English-language social science articles in the fields of internal audit, corporate governance, and risk management were used. Three separate files were prepared for each source text: the original text taken from the source article, a naive LLM rewriting produced without any specific instruction, and a structural rewriting created to observe the risk of idea laundering under controlled conditions. This yielded thirty-six core evaluation files for the twelve source articles, together with five AI-assisted control texts grounded in the authors' own original ideas. All files were uploaded to Turnitin in "No Repository" mode. The findings show that the original texts were recognized by Turnitin with an average similarity of 99.75%, that the average similarity score fell to 44.50% under the naive rewriting condition, and that it reached 19.08% under the structural rewriting condition. Two independent expert raters found that the main idea, argument flow, and concluding claim were largely preserved across all twelve cases. While the control texts returned a similarity score of 0%, their high AI writing scores support the position that an AI score cannot be interpreted as an indicator of plagiarism. The study argues that countering the risk of idea laundering requires a multi-layered publication governance approach grounded in semantic evaluation, source–argument traceability, editorial judgment, and an internal control perspective.

Keywords: Idea laundering, Semantic plagiarism, AI-assisted rewriting, Academic publication governance.
JEL Codes: M42, M40, O33, I23.

ÖZ

Çalışma, akademik yayın süreçlerinde kullanılan benzerlik raporları ve AI (Artificial Intelligence, Yapay Zekâ) yazım raporlarının, yapay zekâ destekli yeniden yazım yoluyla gerçekleştirilebilecek "fikir aklama" riskine karşı ne ölçüde güvence sağlayabildiğini incelemektedir. Çalışmanın temel iddiası, yapay zekâ destekli yazımın tek başına intihal olmadığı; etik sorunun, başkasına ait akademik fikir, argüman veya katkının kaynak gösterilmeksizin yeniden yazılarak özgün katkı gibi sunulması olduğudur. Bu amaçla çalışmada; iç denetim, kurumsal yönetişim, risk yönetimi alanlarına ilişkin yazılmış olan 12 adet İngilizce sosyal bilimler makalesinin tartışma/sonuç bölümleri kullanılmıştır. Her kaynak metin için üç ayrı dosya hazırlanmıştır: kaynak makaleden alınan orijinal metin, özel bir yönlendirme içermeyen naif LLM (Large Language Model-Büyük Dil Modeli) yeniden yazımı ve fikir aklama riskini kontrollü biçimde gözlemlemek amacıyla oluşturulan yapısal yeniden yazım metnidir. Bu şekilde 12 kaynak makale için toplam 36 ana değerlendirme dosyası elde edilmiş, ayrıca yazarların özgün fikirlerine dayanan AI destekli 5 kontrol metni hazırlanmıştır. Tüm dosyalar Turnitin'e depolama olmaksızın yüklenmiştir. Bulgular, orijinal metinlerin Turnitin tarafından ortalama %99,75 benzerlikle tanındığını; naif yeniden yazım koşulunda ortalama benzerlik skoru %44,50'ye düştüğünü ve yapısal yeniden yazım koşulunda ise ortalama benzerlik skorunun %19,08 olarak gerçekleştiğini göstermektedir. İki bağımsız uzman değerlendirici, 12 vakanın tamamında ana fikir, argüman akışı ve sonuç iddiasının büyük ölçüde korunduğunu belirlemiştir. Kontrol metinlerinde benzerlik skoru %0 iken AI yazım skorlarının yüksek çıkması, AI skorunun intihal göstergesi olarak yorumlanamayacağını desteklemektedir. Çalışma, fikir aklama riskine karşı semantik değerlendirme, kaynak-argüman izlenebilirliği, editöryal muhakeme ve iç kontrol perspektifine dayalı çok katmanlı yayın yönetişimi yaklaşımının gerekli olduğunu savunmaktadır.

Anahtar Kelimeler: Fikir Aklama, Semantik İntihal, Yapay Zekâ Destekli Yeniden Yazım, Akademik Yayın Yönetişimi.
JEL Kodları: M42, M40, O33, I23.

1 Msc., ORCID ID: 0009-0008-9687-7101, mertvelioglu20@gmail.com
2 Dr., Internal Auditor, Ministry of Agriculture and Forestry, Ankara, ORCID ID: 0000-0003-2220-3602, hakan.velioglu@tarimorman.gov.tr
3 Dr., Internal Auditor, Ministry of Agriculture and Forestry, Ankara, ORCID ID: 0000-0002-1442-3982, selcuk.olum@tarimorman.gov.tr
4 Internal Auditor, Düzce University, Düzce, ORCID ID: 0000-0002-6826-0998, aliarkin@duzce.edu.tr

1. INTRODUCTION

The growing presence of AI-assisted text generation tools in academic writing has moved the debate over academic integrity into a new phase. This debate is often conducted around the question of whether text written with AI constitutes plagiarism. Yet the question itself risks narrowing the ethical core of academic plagiarism. Plagiarism is not merely a matter of which tool a text was written with; it concerns claiming another's idea, argument, finding, interpretation, or academic contribution as one's own without attribution. Equating AI-assisted text generation directly with plagiarism can therefore create two basic problems. The first is the undue criminalization of legitimate academic writing practices. The second, and more dangerous, is that it obscures the risk that ideas belonging to others may be transformed through AI and presented as an original contribution.

The central claim of this study is that AI-assisted writing cannot, on its own, be judged to be plagiarism. What is decisive for academic ethics is not the tool with which a text was written, but to whom the idea, the argument, and the academic contribution presented in that text belong. When the idea, the analysis, and the contribution are the author's own, drawing on AI for language, structuring, or expression does not amount to plagiarism. By contrast, when an academic idea belonging to someone else is rewritten through an LLM or a paraphrasing tool so that its textual similarity is reduced and it is then presented as an original contribution, this is a serious ethical violation despite the low similarity score. This study refers to that second situation as idea laundering.

This normative position is not unique to the present study. Major publishers and publication ethics bodies have adopted a broadly similar stance: AI assistance in language, readability, and structuring is permissible subject to disclosure, AI tools cannot be credited as authors, and the technology must not replace the authors' own critical thinking, expertise, and evaluation (Committee on Publication Ethics [COPE], 2023; Elsevier, n.d.). The point of departure of this study is therefore not the normative distinction itself but its control-side counterpart. Policies of this kind presuppose that the ownership of the contribution presented in a manuscript can be meaningfully assessed; yet the automated indicators available in publication processes measure textual overlap and the likely mode of production, not contribution ownership. How this verification gap can be exploited, and what it implies for publication governance, is the question this study takes up.

Idea laundering should be situated within, rather than outside, the established understanding of plagiarism. Plagiarism has never been limited to the copying of words; it also encompasses the unattributed appropriation of another's ideas, arguments, structures, and contributions. What this study terms classical textual plagiarism is the most readily detectable form of the violation: another author's text is used directly, or in very close form, without attribution. Text matching systems can be particularly useful in detecting this kind of surface overlap. In a case of idea laundering, however, the traces of the source text at the level of words and sentences are reduced, while the main idea, the argument flow, the academic contribution, and the concluding claim can be preserved. Here a low similarity score does not show that the text is original in its ideas; it shows only that overlap at the textual surface has decreased. The distinctive feature of idea laundering is therefore not that it constitutes a new category of ethical violation; it is that AI-assisted rewriting allows this long-recognized form of misappropriation to be carried out quickly, at scale, and in a way that systematically escapes surface-oriented controls designed to detect textual overlap.

This distinction matters for the assurance function of the automated control tools used in academic publishing. Systems such as Turnitin can produce strong technical indicators of textual similarity, but these indicators are not themselves a verdict of plagiarism. In the same way, AI writing reports can provide limited signals that a text may have been AI generated or AI modified/rewritten, yet they do not directly measure to whom the idea in the text belongs. How both a low similarity score and a high AI writing score are to be interpreted in the publishing process is therefore not only a technical question but also an audit and governance question.

The study approaches the subject not only in the context of educational technology or academic ethics, but as a problem of internal control, information systems audit, and assurance within the publishing process. When the academic publishing process is conceived as a control environment, the Turnitin similarity report and the AI writing report are automated detective controls, while editorial and peer review constitute the manual, judgmental control layer. Within this control architecture, the risk of idea laundering emerges as a control gap that can exploit precisely the limits of the automated detective controls.

The literature on enterprise risk management and information systems increasingly emphasizes that emerging technologies open up new risk surfaces for organizations, and that these risks should be anticipated and assessed proactively before they materialize (Karahan et al., 2025). This study carries a similar logic of proactive risk anticipation into academic publishing. AI-assisted rewriting and paraphrasing can no longer be treated as a merely prospective scenario: large-scale analyses estimate that a measurable and growing share of scholarly texts is already being modified

with LLMs (Liang et al., 2024), and hybrid human–AI writing is increasingly described as an emerging norm of academic text production (Eaton, 2023; Lund et al., 2023). What the study seeks to anticipate is therefore not the emergence of the practice itself, but the consolidation of an assurance gap around an already visible challenge: the risk that publication controls continue to rely on similarity and AI writing indicators while idea laundering falls outside what those indicators measure.

Within this framework, the main research question of the study is the following: when the Turnitin similarity report and the AI writing report are evaluated together, can they provide adequate assurance against the risk of idea laundering, understood as rewriting another author's academic idea through LLM or paraphrasing tools and presenting it as original? In connection with this main question, the study examines the extent to which the original source texts are recognized by Turnitin, the extent to which similarity scores fall under the rewriting conditions, the extent to which the source idea is preserved despite low similarity scores, and how far AI writing scores serve to make this risk visible.

The study is designed as a small-sample proof-of-concept and vulnerability assessment. Its aim is not to show that Turnitin is wholly ineffective, or that all LLMs can circumvent similarity controls in every case. Nor is the study an AI model comparison. The aim is to observe whether low textual similarity and high semantic preservation can arise together under realistic publication-control conditions, and to discuss what this means for publication governance.

The remainder of the article is structured as follows. The second section presents the literature review and conceptual framework, discussing the distinction among AI-assisted writing, classical textual plagiarism, and idea laundering. The third section sets out the research design, the sample selection, the LLMs used, the prompts, the Turnitin application protocol, and the semantic preservation coding. The fourth section presents the findings. The fifth section discusses the findings in the context of publication governance, internal control, and source–argument traceability. The final section summarizes the study's conclusions, limitations, and suggestions for future research.

2. LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

2.1. Textual Similarity, Plagiarism, and Academic Judgment

Academic plagiarism most often becomes visible through textual similarity; yet textual similarity and plagiarism are not the same thing. A text that closely resembles another source may give rise to a suspicion of plagiarism, but the final ethical judgment must take into account citation form, attribution, context, shared terminology, methodological necessities, and the author's academic contribution. Similarity reports should therefore be seen not as automated decision tools that take the place of academic judgment, but as technical indicators that support evaluation.

Bretag and Mahmud (2009) emphasize that electronic detection tools should be evaluated together with academic judgment. In a similar vein, Meo and Talha (2019) note that Turnitin is essentially a text matching tool and does not, in itself, deliver a verdict of plagiarism. Turnitin's own guides likewise state that the similarity score shows the extent to which a text matches existing sources, but that plagiarism cannot be determined from the score alone (Turnitin, n.d.-a; Turnitin, n.d.-b). This approach is consistent with the central assumption of the present study: the similarity score is an important control indicator, but it is not the whole of an assessment of originality or plagiarism.

This distinction has become more critical in the current context, in which LLMs and paraphrasing tools have proliferated. A low textual similarity score may not mean that the source idea has been lost. The textual surface may have changed, sentence structures may have been transformed, and word overlap may have decreased; yet the idea the text carries, its argument flow, and its contribution claim may remain largely the same. In such a case, control based on surface similarity alone may fail to make the problem of academic contribution ownership visible.

2.2. Patchwriting, Paraphrasing, and Dependence on the Source Text

Patchwriting refers to a form of source-dependent writing that arises when another author's text is partially rearranged, when some words are replaced, or when sentence structures are transformed. Howard (1995) addressed the concept of patchwriting specifically within the contested space between academic writing development, source use, and plagiarism debates. Pecorari (2003) and Shi (2012) show that, in the context of second-language academic writing, source-dependent rewriting practices complicate the boundaries among plagiarism, the learning process, and text production. Jamieson (2015) likewise stresses that the concept of patchwriting needs to be treated more meaningfully in order to understand the gray area between plagiarism and academic writing development. According to work along these lines, paraphrasing can be part of legitimate academic writing; but where the source idea is reappropriated without adequate attribution, the ethical problem remains. Academic integrity therefore depends not only on whether words have been changed, but on whether the source of the intellectual contribution has been clearly stated.

In this connection, Walker (1998) shows that approaches relying solely on punitive or technical tools may fall short in combating student plagiarism, and that institutional policy, academic guidance, and assessment processes need to be considered together. This approach holds in the current context of AI-assisted writing as well. Important as technical detection tools are, the need for institutional and editorial judgment regarding academic contribution ownership and source use does not disappear.

2.3. The Limits of Paraphrasing Tools and Text Matching Systems

Internet-based paraphrasing tools and language translation tools have formed an important field of research that makes the limits of textual similarity controls visible. Rogerson and McCarthy (2017) show that internet-based paraphrasing tools can transform a source text at the surface level and thereby make it difficult for classical text matching systems to establish the link to the source. Prentice and Kinden (2018), in turn, argue that even though paraphrasing and translation tools can produce low similarity scores, the resulting texts may remain dependent on the source text at a semantic or structural level.

These studies establish that a low similarity score does not mean independence from the source text. Paraphrasing tools in particular can reduce the traces of a text at the level of words and sentences; but whether the source idea or argument structure has been preserved must be assessed separately. Although text matching systems thus provide an important initial screening control against the risk of classical textual plagiarism, they may not, on their own, be adequate against the risk of semantic plagiarism or idea laundering.

2.4. AI Text Detection, Paraphrasing, and the Risk of Idea Laundering

With the spread of LLMs, the dimension of AI text detection has been added to the debate over academic integrity. Work in this area discusses the extent to which AI-generated texts can be distinguished by existing detection tools, how paraphrasing operations affect those tools, and how reliable AI writing indicators are. Khalil and Er (2023) argue that texts produced by ChatGPT can yield low similarity scores in classical plagiarism detection tools, and that this calls for rethinking the existing detection paradigm.

Krishna et al. (2023) show that when AI-generated texts are paraphrased, the performance of existing AI text detection tools can decline. Their study demonstrates that paraphrasing operations can largely preserve the semantic content of a text while weakening the surface or generative traces on which detection tools rely. The authors further argue that retrieval-based defense approaches may be more robust against the effects of paraphrasing. This literature documents, in a broader context, the fragility of detection logic based on the textual surface or on generative traces in the face of paraphrasing-type transformations.

The literature that has accumulated since the public release of ChatGPT at the end of 2022 reinforces these concerns on a broader empirical basis. In one of the most comprehensive tests to date, covering fourteen detection tools including Turnitin, Weber-Wulff et al. (2023) conclude that the available detectors are neither sufficiently accurate nor sufficiently reliable to ground high-stakes integrity decisions, and that detection performance deteriorates markedly when texts are machine-paraphrased or manually edited. Gao et al. (2023) show that research abstracts generated by ChatGPT received very high originality scores from a plagiarism detector and that even blinded human reviewers misclassified a non-trivial share of them. Working with an explicit adversary model, Sadasivan et al. (2023) demonstrate that recursive paraphrasing can break a wide range of detectors, including watermarking-based and retrieval-based schemes, and they offer a theoretical argument that, as LLMs improve, even the best possible detector may perform only marginally better than a random classifier. In an applied higher education setting, Perkins et al. (2024) find that when submissions are produced with prompting techniques intended to reduce detectability, the Turnitin AI indicator identifies only part of the AI-generated content, and academic staff report only about half of the experimental submissions to misconduct processes.

Two implications of this literature matter for the present study. The first is that the reliability problems of AI text detection are structural rather than incidental: the indicators rest on surface or generative traces that rewriting is specifically able to weaken. The second is that this body of work is concerned overwhelmingly with whether a text was produced by AI. The question examined here is different and, from a publication governance standpoint, prior: to whom do the idea, the argument, and the contribution claim carried by the text belong? Even a perfectly reliable AI detector would not answer that question, because AI-assisted writing grounded in the author's own ideas is not plagiarism. The post-2022 detection literature therefore sharpens, rather than resolves, the assurance problem addressed in this study.

2.5. A Threefold Conceptual Distinction: AI-Assisted Writing, Classical Plagiarism, and Idea Laundering

This study conceptually distinguishes three different situations in academic text production. The first is AI-assisted writing. Here the idea, the analysis, and the academic contribution belong to the author; AI is used to develop language,

to structure the text, to strengthen the exposition, or to improve academic style. Such use may be debated separately in terms of transparency and institutional policy, but it cannot, on its own, be characterized as plagiarism.

The second is classical textual plagiarism. The term denotes here not the whole of plagiarism, which extends to the unattributed appropriation of ideas, arguments, and structures as well as words, but its most surface-visible form: another author's text is used directly, or in very close form, without attribution. Text matching systems such as Turnitin are oriented mainly toward detecting this kind of surface textual overlap. Textual similarity reports can therefore serve an important initial screening control function against the risk of classical plagiarism; but these reports are not themselves a verdict of plagiarism.

The third is the risk of idea laundering, or semantic plagiarism, which has become more visible with the rapid development of AI tools. Here another's idea, argument structure, theoretical contribution, or concluding claim is syntactically transformed through LLM or paraphrasing tools; textual similarity is reduced; but the intellectual core of the source contribution is preserved. The ethical violation here lies not in the verbatim repetition of words, but in concealing the source of the academic contribution. This threefold distinction makes it necessary to assess the risk of idea laundering separately from classical textual plagiarism and from legitimate AI-assisted writing practices.

2.6. The Publishing Process as a Control Environment

The academic publishing process is not only a process of scientific evaluation but also a control environment. Within this environment, editors, reviewers, similarity control systems, AI writing reports, ethics statements, contribution statements, and reference checks each take on different control functions. Turnitin and similar systems can be thought of as automated detective controls. Editorial and peer review, in turn, constitute the judgmental, manual control layer.

Publisher policies on AI-assisted writing form the policy layer of this control environment. COPE (2023) states that AI tools cannot be listed as authors and that authors remain fully responsible for the content of their manuscripts, including the parts produced with AI support. Elsevier's policy similarly permits generative AI in the writing process only to improve the readability and language of the work, requires disclosure, and stresses that the technology must be applied with human oversight and must not replace key authoring tasks (Elsevier, n.d.). Such policies are indispensable in defining expected behavior and allocating accountability. From an internal control perspective, however, they are preventive and declarative in character: their assurance value depends on honest disclosure and on the capacity of the surrounding controls to verify what the policies presuppose, namely that the contribution presented is the authors' own. Against an author who strips the textual traces of a source and does not disclose AI use, the policy layer has no self-executing force; the burden falls back on the detective controls, whose limits this study examines.

From an internal control perspective, no single control can be expected to address all risks. The Committee of Sponsoring Organizations of the Treadway Commission (COSO) framework stresses that the control environment, risk assessment, control activities, information and communication, and monitoring must work together. The Institute of Internal Auditors (IIA) Three Lines Model, in turn, sets out that assurance and responsibilities should be distributed across different lines. Applied to academic publishing processes, this approach shows that similarity and AI writing reports should be evaluated not as standalone decision tools but as components of a broader publication governance architecture.

The risk of idea laundering can exploit, within this control environment, precisely the limits of the automated detective controls. Accordingly, publishing processes require not only textual similarity or AI writing indicators, but also source–argument traceability, semantic evaluation, and editorial judgment grounded in domain expertise.

2.7. From the Literature to the Research Gap: The Position of the Study

The literature reviewed above shows that text matching systems and AI writing indicators have two distinct but related limits. The first limit is that the traces of textual similarity can weaken following paraphrasing and structural transformation. The second is that, although AI writing indicators can provide a signal about how a text was produced, they cannot directly determine the true source of the idea or academic contribution appropriated in the text. A low similarity score therefore does not prove intellectual independence from the source text; nor does a high or low AI score, on its own, offer an adequate indicator of plagiarism or contribution ownership.

This gap shows the need, in academic publication controls, for a broader evaluation framework that moves from textual similarity toward contribution traceability. The problem is not limited to the question of whether a paraphrasing tool reduces similarity; the truly critical question is whether the academic idea, the argument flow, and the contribution claim in the source text are preserved despite a low similarity score. This framework makes it necessary to address the distinction among AI-assisted writing, classical textual plagiarism, and idea laundering together with a publication governance and internal control perspective.

In this regard, the principal contribution of the study is that it conceptualizes the risk of idea laundering as a problem of academic publication control that extends from textual similarity to contribution traceability. Within this framework, similarity and AI writing reports should be evaluated not as final decision tools in their own right, but as limited detective controls within a broader assurance architecture.

This framing also makes explicit how the study relates to existing publisher policies and to the recent detection literature. The study does not claim novelty for the normative position that AI-assisted writing is not in itself plagiarism; that position is now embedded in publisher policy (COPE, 2023; Elsevier, n.d.). Its contribution is to move from the policy statement to the control question. It demonstrates empirically, under realistic publication-control conditions, that the automated indicators available to editors cannot verify contribution ownership; it conceptualizes the resulting exposure, idea laundering, as a detective-control gap within publication governance; and it translates the policy expectation into operational evaluation questions for editors and reviewers. In doing so, the study extends the post-2022 detection debate (Gao et al., 2023; Perkins et al., 2024; Sadasivan et al., 2023; Weber-Wulff et al., 2023) from the question of production mode to the question of contribution ownership.

3. METHOD

3.1. Research Design

In line with this conceptual framework, the research is designed as a small-scale proof-of-concept and vulnerability assessment intended to make visible the risk of idea laundering that may arise in academic publication controls. Calling the design a "vulnerability assessment" does not mean that the study aims to produce a guide to misuse. On the contrary, the methodological conditions were deliberately kept limited and were framed within the means available to an average computer user without advanced technical knowledge. For this reason, no API was used, no model parameters were adjusted, a single run was carried out under each condition, the first output was taken as the basis, and no output optimization was pursued through regeneration or repeated attempts. Accordingly, the study focused not on producing a systematic evasion strategy, but on observing the assurance gap that may arise under realistic and limited publication-control conditions.

The research question was formulated as follows: when the Turnitin similarity report and the AI writing report are evaluated together, can they provide adequate assurance against the risk of idea laundering, understood as rewriting another author's academic idea, argument, and contribution claim through LLM or paraphrasing tools and presenting it as original? This main question was examined at three sub-levels. First, the extent to which the original source texts were recognized by Turnitin was measured. Second, the extent to which the similarity scores of those same texts changed under the naive rewriting and structural rewriting conditions was assessed. Third, the extent to which the source idea, the argument flow, and the concluding claim were preserved in texts that produced low similarity scores was coded by two independent expert raters.

The findings obtained were therefore interpreted not in order to measure the general accuracy of AI detection tools, but in order to assess the limit of the assurance that similarity and AI writing reports provide in the face of the risk of idea laundering.

3.2. Sample Selection

The study's sample was selected from a preliminary pool of forty candidate articles formed in the first stage. This pool was composed of English-language social science articles related to the fields of internal audit, corporate governance, risk management, ethics, financial reporting, and auditing. The choice of fields rests on the aim of forming a sample consistent with the study's publication governance, internal control, and academic audit perspective.

At the preliminary evaluation stage, the discussion and conclusion sections of each article in particular were examined. Articles whose discussion/conclusion sections consisted predominantly of tables, technical result lists, or numerical output, and which did not textually represent the study's main idea, argument flow, and contribution claim well enough, were eliminated. In the same way, it was judged that where the relevant section was too short, sufficient argument density for the semantic preservation assessment could not be provided, while where it was too long, problems could arise in terms of the comparability of the experimental files, the Turnitin upload process, and methodological consistency.

Following this preliminary screening, twelve articles whose discussion/conclusion sections were judged more suitable in terms of textual argument density, suitability for the rewriting conditions, field diversity, and balance of citation level were included in the final sample. The article codes were used by preserving the candidate codes assigned in the preliminary pool. As a result, the final sample contains code skips, such as the jump from A10 to A14 or from A18 to

A40. These skips do not indicate missing data or the loss of an experimental file; they result only from not renumbering the codes of candidate articles that appeared in the preliminary pool but were not included in the final sample.

The source articles were identified using Harzing's Publish or Perish software, Windows GUI Edition version 8.19.5300.9483. In the search process, key terms related to themes such as internal audit, auditing, corporate governance, risk management, ethics, and financial reporting were used. The year range was set as 2015–2020. This range was chosen because it provides sufficient time for the articles to enter academic circulation and accumulate citations, while not letting the literature lose its currency entirely.

The sample includes both highly cited and less cited articles (Table 1). Six articles fall in the category of 300 or more citations, two articles in the 200–300 citation range, and four articles in the 20–60 citation range. This distribution ensures that the source texts consist of examples that are both highly visible academically and of comparatively more limited circulation. The study makes no claim to a statistical comparison between these citation categories; citation levels were used only to show sample diversity and balance in terms of source visibility.

Table 1. Bibliometric and Thematic Characteristics of the Source Articles Included in the Sample

Code	Short topic/theme	Year	Citations	Citations per year
A01	Whistle-blowing and corporate governance	2019	26	3.71
A02	Whistle-blowing regulation	2020	51	8.50
A03	Internal audit and risk culture	2020	55	9.17
A04	Auditing and corporate governance literature	2017	56	6.22
A08	Enterprise risk management	2018	225	28.13
A10	International corporate governance	2019	296	42.29
A14	Internal audit quality and audit delay	2015	387	35.18
A15	Effectiveness of ethics programs	2015	407	37.00
A16	Enterprise risk management investment	2015	432	39.27
A17	Enterprise risk management maturity	2015	496	45.09
A18	Synthesis of the internal audit effectiveness literature	2015	500	45.45
A40	Enterprise risk management and financial reporting	2017	335	37.22

3.3. Source Text Selection

For each article, the discussion, conclusion, or contribution-oriented closing sections were selected as the source text. Selecting these sections rather than the full text is a deliberate methodological choice. In academic articles, original interpretation, theoretical contribution, the concluding claim, the contribution to the literature, and practical implications are mostly concentrated in the discussion and conclusion sections. The risk of idea laundering, too, most often arises not merely at the level of words or sentences, but at the level of the argument and the contribution claim in these sections.

The length of the source texts was generally kept within the 700–1,100 word range. This length was judged suitable both for the Turnitin and AI writing reports to be generated and for the texts to carry the integrity of an academic argument. In selecting the texts, passages were preferred that contained not the article's main methodological details, but rather its more interpretive and contribution-focused sections. The experiment thus focused on the preservation of the idea, the argument, and the concluding claim rather than on the repetition of technical terminology.

3.4. Experimental Conditions and Models Used

Three main conditions were created for each source text. The "O condition" denotes the original discussion/conclusion text selected from the source article. The "N condition" is the naive LLM rewriting; in this condition the LLM was told to rewrite the text in academic English while preserving the meaning, the argument flow, and the conclusion. The "A condition" is the structural rewriting condition. In this condition the text was to be reconstructed not only at the level of words but also in terms of sentence structure, paragraph transitions, and order of exposition, so that it would read like an independently written academic discussion.

The design of the structural rewriting condition is a deliberate choice in terms of the study's ethical positioning. In the real world, a malicious user might construct the prompt with adversarial wording—that is, with explicit instructions intended to mislead a particular detection mechanism, lower the similarity score, or conceal the trace of the source. Such use would in all likelihood expose the risk of idea laundering more sharply. In this study, however, that path was not chosen, for reasons of academic ethics. Publishing examples of prompts that contain operational evasion carries the risk of turning the article into a guide to misuse. The study therefore adopted a "structural rewriting" condition that preserves the meaning of the source text while reconstructing only the structure and the exposition, and that carries no direct intent to circumvent. That even this condition can jointly produce low similarity and high semantic preservation suggests that the risk associated with explicitly adversarial uses constructed in the real world may be sharper than the pattern observed in this study.

Three popular and current LLMs were used in the study: ChatGPT 5.5 Thinking, Claude Opus 4.7, and Gemini 3.1 Pro. All of the models were used through their web interfaces; no API was used, and no model settings that might affect the diversity or randomness of the output were adjusted. A single run was carried out under each condition, the first output was taken as the basis, and the option to regenerate a response was not used (Table 2). This design rests on the aim of not optimizing model performance and of preserving realistic user conditions.

Table 2. Distribution of Models in the Main Experimental Set

Model	Article codes used	Number of cases
ChatGPT 5.5 Thinking	A01, A04, A14, A18	4
Claude Opus 4.7	A02, A08, A15, A40	4
Gemini 3.1 Pro	A03, A10, A16, A17	4

As Table 2 shows, although the distribution of models was kept balanced, the study was not designed as a contest among models. Because each model was used in only four cases, it is not methodologically sound to make claims of superiority or weakness at the model level. Model information was reported in order to provide context for the interpretation of the findings.

3.5. Prompts and Control Texts

The prompts used in the naive and structural rewriting conditions were held constant. The prompts explicitly stated that no new literature, new citations, new data, new examples, or external information was to be added (Table 3). This constraint is necessary so that the rewriting outputs preserve their relationship with the idea and argument structure of the source text, and so that the experiment measures only the effect of rewriting.

In addition to the main experimental set, five control texts were created. The control texts were not produced in order to rewrite the idea or argument structure of any source article. Instead, original seed ideas derived from the study's own conceptual arguments were given to the AI tools, and these ideas were to be written in the form of an academic discussion. The purpose of the control texts is to show that the AI writing score is not an indicator of plagiarism, and to make visible the conceptual distinction between AI-assisted writing and idea laundering.

Table 3. Prompts Used in the Naive, Structural Rewriting, and Control Conditions

Condition	Prompt used
Naive LLM rewriting	You are given an excerpt from an academic article. Rewrite the text in formal academic English while preserving the original meaning, argument flow, conceptual claims, and conclusion. Do not summarize the text. Do not add new claims, examples, data, citations, references, or external information. Do not remove substantive arguments. Do not change the overall position of the author. Keep the rewritten version approximately similar in length to the original text. Output only the rewritten academic text.
Structural rewriting	You are given an excerpt from an academic article. Produce a substantially restructured academic version of the text. Preserve the underlying idea, conceptual contribution, argument sequence, reasoning logic, and conclusion of the original passage. However, reformulate the wording, sentence structures, paragraph transitions, and exposition style so that the resulting passage reads as an independently written academic discussion. Do not add new literature, new citations, new empirical claims, new examples, or external information. Do not distort the original argument. Do not summarize; produce a full rewritten version with approximately similar length. Avoid direct reuse of distinctive phrases unless they are unavoidable technical terms. Output only the rewritten academic text.
Control text prompt template	You are given an original conceptual argument developed by the author. Write it as a formal academic discussion. Do not add new literature, citations, empirical claims, examples, or external information. Preserve the author's original idea and argument. Output only the academic text.

3.6. Turnitin Application Protocol and Measured Variables

All files were uploaded to Turnitin in "No Repository" mode. This choice was made so that the experimental texts would not be added to the Turnitin database and so that no artificial matches would arise in subsequent uploads. For each file, the Turnitin similarity score, whether there was a direct match with the source article, the highest source match, the decrease in similarity relative to the original text, and the Turnitin AI writing score were recorded. In cases where the AI writing report disclosed them, the "AI generated" and "AI modified/rewritten" sub-indicators were also noted separately.

The AI writing reports were evaluated as a limited production-mode indicator that Turnitin generates under its relevant file and text eligibility conditions (Turnitin, n.d.-c; Turnitin, n.d.-d). Because these scores do not directly show to whom the idea presented in the text belongs or whether plagiarism is present, they were not interpreted as an indicator of plagiarism. This distinction is central to the study's conceptual framework: the AI score signals at the level of the mode of production, the similarity score at the level of textual overlap, and semantic preservation at the level of whether the source idea has been preserved—each a different level of signal.

A total of forty-one files were uploaded to Turnitin in the study. Thirty-six of these belong to the main experimental set: 12 original source texts, 12 naive rewriting outputs, and 12 structural rewriting outputs. The remaining 5 files consist of the original idea + AI-assisted writing control texts. This number is consistent with the proof-of-concept nature of the study and kept the Turnitin uploads at a manageable level.

3.7. Semantic Preservation Coding

The study's most critical measurement is not the Turnitin similarity score alone. Whether the source idea was preserved in a text that produced a low similarity score was also assessed. For this purpose, semantic preservation coding was carried out for the rewritten texts in the A condition. The coding aims to measure not textual similarity, but the extent to which the main idea, the argument flow, the academic contribution or conceptual emphasis, and the concluding claim of the source text were preserved in the candidate text.

The semantic preservation coding was conducted by two subject-matter experts independent of the authors. The raters were presented with pairs consisting of the source text and the rewritten candidate text; however, they were not given information about the Turnitin similarity scores, the AI writing scores, the model used, or the experimental condition to which the texts belonged. The evaluation was thus conducted on a blind coding procedure, stripped of condition labels and automated report results.

The coding rubric used in the study has four levels: 0 = the source idea has largely been lost; 1 = some concepts have been preserved, but the main argument has been disrupted; 2 = the main idea has been preserved, but the reasoning or emphasis has partly changed; 3 = the main idea, the argument flow, and the concluding claim have largely been preserved.

The two expert raters assigned a score of 3 to all twelve cases in the A condition. Inter-rater percent agreement was therefore 100%. Because the score distribution was concentrated in a single category, the interpretive value of coefficients such as Cohen's kappa was judged to be limited, and the level of agreement was reported as percent agreement.

This coding was used so that the risk of idea laundering could be assessed not only through textual similarity scores but also through the preservation of the source idea and argument structure. In this way, the extent to which the intellectual core of the source contribution persisted in texts that produced low similarity scores was examined separately.

3.8. Methodological Limits

The study makes no claim to firm generalization. The sample is limited to twelve articles; only discussion/conclusion sections were used as source texts; and the experiments were conducted on a single-run, first-output basis. These limitations clearly establish the proof-of-concept nature of the study. Different results may be obtained in different disciplines, in different languages, with different model versions, with different prompts, or in scenarios involving manual editing.

Because of the rapid development of LLMs, this study should be regarded as a snapshot obtained with the model versions accessible within a particular time frame. Moreover, because the study is not a model comparison, no model-level performance claim was made. The interpretation of the findings is directed toward showing the limits of a control logic based on Turnitin similarity and AI writing reports in the face of the risk of idea laundering.

4. FINDINGS

4.1. Turnitin Similarity Scores in the Original and Rewritten Texts

The research first examined the extent to which the original source texts were recognized by Turnitin. This check is important; for in a case where the original text is not recognized by Turnitin with high similarity, the interpretive value of a low similarity score after rewriting would remain limited. The findings show that the original texts were recognized by Turnitin at almost an exact-match level. The average similarity score of the twelve original texts is 99.75% (Table 4).

As Table 4 shows, the average similarity score fell to 44.50% under the naive rewriting condition. Under the structural rewriting condition, the average similarity score came to 19.08%. In other words, the average decrease relative to the original texts under the structural rewriting condition is 80.67 percentage points. Despite this, all cases in the A condition were coded with a semantic preservation score of 3. This finding shows that the marked decrease in textual similarity does not mean that the main idea, the argument flow, and the concluding claim of the source text have been lost.

Table 4. Turnitin Similarity Scores and Semantic Preservation Under the O/N/A Conditions, with Model Information

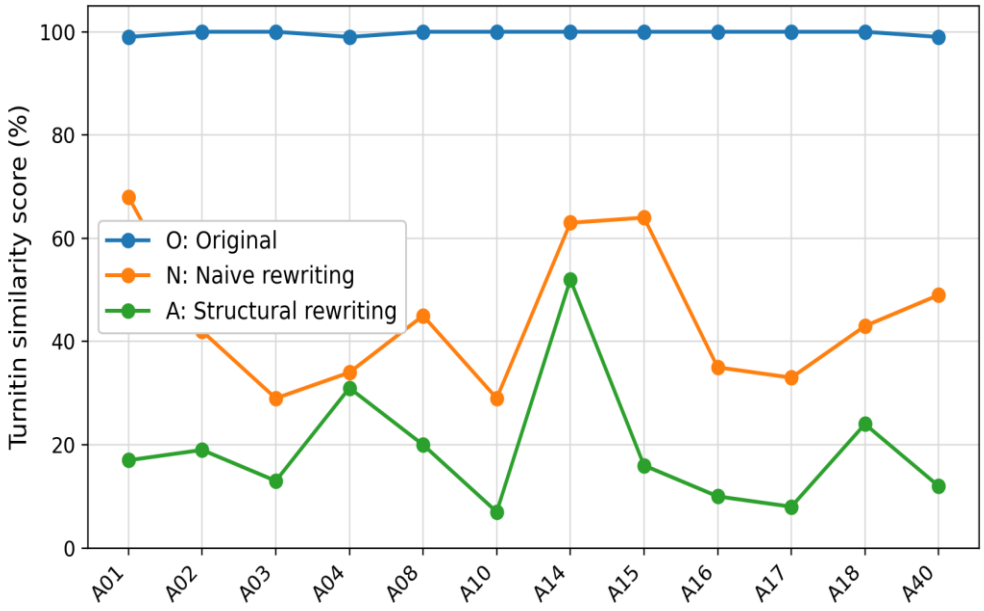
Article	Model	Original text similarity score (O condition) (%)	Similarity score under the N condition (%)	Similarity score under the A condition (%)	Semantic preservation score under the A condition
A01	ChatGPT 5.5 Thinking	99	68	17	3
A02	Claude Opus 4.7	100	42	19	3

[2] A coding rubric is a guiding table created to organize, make sense of, and categorize raw data (interviews, survey responses, focus group discussions, and the like), particularly in qualitative research and data analysis processes.

Article	Model	Original text similarity score (O condition) (%)	Similarity score under the N condition (%)	Similarity score under the A condition (%)	Semantic preservation score under the A condition
A03	Gemini 3.1 Pro	100	29	13	3
A04	ChatGPT 5.5 Thinking	99	34	31	3
A08	Claude Opus 4.7	100	45	20	3
A10	Gemini 3.1 Pro	100	29	7	3
A14	ChatGPT 5.5 Thinking	100	63	52	3
A15	Claude Opus 4.7	100	64	16	3
A16	Gemini 3.1 Pro	100	35	10	3
A17	Gemini 3.1 Pro	100	33	8	3
A18	ChatGPT 5.5 Thinking	100	43	24	3
A40	Claude Opus 4.7	99	49	12	3
Average		99.75	44.50	19.08	3

As can be seen in Figure 1, while similarity scores in the original texts are at almost an exact-match level, the scores fall markedly under the N and especially the A conditions. The A14 case, in turn, stands apart within this general pattern as a limiting/contrasting case.

Figure 1. Case-by-Case Variation in Turnitin Similarity Scores Under the O, N, and A Conditions



(Figure created by the authors)

4.2. Case-by-Case Ranking Under the Structural Rewriting Condition

The case-by-case results under the structural rewriting condition show that the similarity scores vary across a fairly wide range. Under the A condition, the lowest similarity score was 7% in A10, and the highest similarity score was 52% in A14. This table should not be interpreted as any institutional plagiarism threshold or Turnitin decision criterion. The aim is to make visible, descriptively, the case-by-case scores obtained under the structural rewriting condition.

Table 5. List of Cases Ranked by Similarity Score Under the A Condition

Rank	Article	Model	Turnitin similarity rate under the A condition (%)	Semantic preservation score under the A condition
1	A10	Gemini 3.1 Pro	7	3
2	A17	Gemini 3.1 Pro	8	3
3	A16	Gemini 3.1 Pro	10	3
4	A40	Claude Opus 4.7	12	3
5	A03	Gemini 3.1 Pro	13	3
6	A15	Claude Opus 4.7	16	3
7	A01	ChatGPT 5.5 Thinking	17	3
8	A02	Claude Opus 4.7	19	3
9	A08	Claude Opus 4.7	20	3
10	A18	ChatGPT 5.5 Thinking	24	3
11	A04	ChatGPT 5.5 Thinking	31	3
12	A14	ChatGPT 5.5 Thinking	52	3

Table 5 shows that, under the structural rewriting condition, low similarity and high semantic preservation can be observed together. The study's main finding is therefore not merely that the similarity score falls. The critical finding is that this decrease occurs together with the preservation of the source idea, the argument flow, and the concluding claim. What makes the risk of idea laundering visible is precisely this simultaneity: the reduction of textual traces and the persistence of intellectual dependence.

4.3. Turnitin AI Writing Scores Under the N and A Conditions

In the main experimental set, the Turnitin AI writing scores were also recorded. These scores were evaluated as a limited production-mode indicator that a text may have been AI generated or AI modified/rewritten. However, these scores were not interpreted as an indicator of plagiarism, because the AI writing score does not directly measure to whom the idea or academic contribution in the text belongs.

Table 6. Turnitin AI Writing Scores Under the N and A Conditions

Article	Model	N condition Turnitin similarity (%)	Rate of content detected as AI under the N condition (%)	A condition Turnitin similarity (%)	Rate of content detected as AI under the A condition (%)	Rate of content detected as modified under the A condition (%)
A01	ChatGPT 5.5 Thinking	68	0	17	30	0
A02	Claude Opus 4.7	42	0	19	43	0
A03	Gemini 3.1 Pro	29	88	13	96	0
A04	ChatGPT 5.5 Thinking	34	72	31	70	0
A08	Claude Opus 4.7	45	26	20	Not disclosed (<20)	Not disclosed (<20)
A10	Gemini 3.1 Pro	29	85	7	96	0
A14	ChatGPT 5.5 Thinking	63	40	52	66	0
A15	Claude Opus 4.7	64	Not disclosed (<20)	16	43	0
A16	Gemini 3.1 Pro	35	100	10	100	0
A17	Gemini 3.1 Pro	33	100	8	100	0
A18	ChatGPT 5.5 Thinking	43	24	24	Not disclosed (<20)	Not disclosed (<20)
A40	Claude Opus 4.7	49	38	12	73	0

In Table 6, the Turnitin AI writing report did not produce a detailed rate for some files because the total AI score remained below 20%. For this reason, the expression “Not disclosed (<20)” is used in the relevant cells. The AI modified/rewritten sub-indicator did not produce a positive value in any of the A-condition files for which it was disclosed. This should be interpreted as a secondary and limited finding, indicating that the relevant sub-indicator does not systematically flag the idea laundering or structural rewriting conditions.

In the N-condition files for which the total AI score was disclosed, the average AI score was approximately 52%, and in the A-condition files it was approximately 72%. Even so, the case-level distribution ranges widely, from 0% or the undisclosed <20% level up to 100%. This distribution supports the conclusion that the AI writing score does not, on its own, reliably indicate the risk of idea laundering.

In all of the original source texts, the Turnitin AI writing score was reported as 0%. This finding shows that the O condition serves as the baseline point of comparison; however, because the study was not designed as an AI detection validation test, this result should not be interpreted as independent evidence of the AI detection tool’s general accuracy.

4.4. Control Texts: Original Idea and AI-Assisted Writing

As explained in the method section, the control texts were used in order to make visible the conceptual distinction between AI-assisted writing and idea laundering. This section reports the findings obtained for these texts through their Turnitin similarity and AI writing scores.

As Table 7 shows, the Turnitin similarity score came to 0% in all of the control texts. By contrast, the AI writing scores are quite high: an AI score of 100% was observed in four control texts and 64% in one. This result clearly shows that having been written with AI is not an indicator of plagiarism. A text may have been written with the help of AI; but if its intellectual core belongs to the author, this situation cannot in itself be interpreted as plagiarism.

Table 7. Turnitin Similarity and AI Writing Scores in the Control Texts

File	Tool	Turnitin similarity rate (%)	AI score (%)	Comment
CTRL-01	ChatGPT	0	100	Original idea + AI writing
CTRL-02	ChatGPT	0	100	Original idea + AI writing
CTRL-03	Claude	0	100	Original idea + AI writing
CTRL-04	Claude	0	64	Original idea + AI writing
CTRL-05	Gemini	0	100	Original idea + AI writing
Average		0	92.8	

The control texts support the study's central conceptual distinction. A high AI score shows that AI assistance may have been present in producing the text; but it does not show that the idea in the text belongs to someone else. By contrast, some of the low similarity scores observed under the A condition do not mean that the source idea has been lost. Accordingly, neither a low similarity score nor a high AI score is, on its own, sufficient for an assessment of academic ethics.

4.5. The A14 Case: A Limiting/Contrasting Case Interpretation

The A14 case is an important boundary condition that balances the study's results. In this case, the original text produced 100% similarity; under naive rewriting, similarity came to 63%, and under structural rewriting, 52%. In the A14 A condition, semantic preservation was again coded as 3; however, the similarity score did not fall to a low level. The case was therefore evaluated as a limiting/contrasting case showing that structural rewriting does not produce a low similarity score in every instance.

In interpreting the A14 case, it is not possible to separate the model effect entirely from the source-text effect. Nonetheless, it should be borne in mind that A14 was processed with the ChatGPT 5.5 Thinking model and that, in the other cases processed with the same model, much lower residual similarity scores were produced. In the other three cases processed with the same model, the A-condition similarity scores are 17% in A01, 31% in A04, and 24% in A18. The average A similarity score of these three cases is 24%, and the average similarity decrease is approximately 75.3 percentage points. By contrast, in A14 the A-condition similarity score remained at 52% and the decrease was 48 percentage points. The high residual similarity in A14 should therefore be evaluated as a boundary case that can be explained by the structural features of the source text or by a particular model–text interaction, rather than by a general inadequacy of the model.

It is not sound to offer a firm causal explanation for this result; but several possible interpretations can be put forward. First, the A14 source text may have contained more fixed conceptual or terminological expressions than some of the other cases. Such expressions can lead to the preservation of textual links with the source text even after rewriting. Second, the structural rewriting output may have preserved certain distinctive expression patterns or the order of the argument to a greater degree. Third, Turnitin may have produced matches not only with the source article directly, but also with secondary sources carrying the same textual or conceptual structure. None of these possibilities has been confirmed on its own; but the A14 case is important in showing that the risk of idea laundering does not become invisible to the same degree in every text.

Retaining A14 in the study is a deliberate methodological choice. This case shows that the research does not construct an extreme claim such as "LLM rewriting always circumvents Turnitin." The strength of the findings derives not from all cases producing the same result, but from the explicit reporting of the fact that, while the similarity score fell sharply in most cases, this decrease remained more limited in one boundary case.

4.6. Overall Assessment of the Findings

When the findings are evaluated together, three conclusions stand out. First, Turnitin was able to recognize the original source texts with high similarity rates. This result shows that the study cannot be interpreted as meaning that Turnitin

could not recognize the source texts at all. Second, when the same source texts were rewritten through an LLM, the similarity scores fell sharply, especially under the structural rewriting condition. Third, despite the low similarity scores, the main idea, the argument flow, the academic contribution, and the concluding claim of the source text were largely preserved.

These three conclusions together show why the risk of idea laundering cannot be assessed through textual similarity alone. When the Turnitin similarity score is low, this does not prove that the text rests on an original idea. In the same way, when the AI writing score is high, this does not show that an idea belonging to someone else has been appropriated. In academic publishing processes, the similarity score and the AI writing score should therefore be interpreted together with an assessment of semantic preservation and contribution ownership.

5. DISCUSSION

The findings of this study show that the assurance function attributed to similarity and AI writing reports in academic publishing needs to be rethought with care. The main conclusion to be discussed is not that the use of AI renders academic writing unethical in itself. On the contrary, the findings support the view that the use of AI and the assessment of plagiarism should be conceptually separated. That a text has been written with the help of AI does not show that the idea in it belongs to someone else. In the same way, that a text produces a low similarity score does not prove that the idea in it is original. What is decisive for academic ethics is not the tool of production or the surface similarity rate, but the true source of the intellectual contribution.

Publication controls therefore cannot be reduced to two indicators alone: the similarity score and the AI writing score. The similarity score produces a limited signal regarding textual overlap. The AI score produces a limited signal regarding the likely mode of production of a text. But idea laundering occurs in an area that neither of these indicators measures directly: the ownership of the source idea, the argument architecture, and the claim to academic contribution. The study's proposal is therefore not to reject automated controls, but to develop a stronger publication governance approach that complements these controls with semantic evaluation and source–argument traceability.

5.1. The Assurance Limit of the Similarity Score

The findings of this study should not be interpreted as meaning that Turnitin or similar text matching tools are wholly ineffective. On the contrary, that the original texts produced similarity scores in the 99%–100% range shows that Turnitin can recognize accessible source texts with high accuracy. The problem that emerges in this study is therefore not that the tool cannot recognize original sources. The real problem is the limit that arises when a control tool based on textual similarity is assigned the function of guaranteeing the intellectual originality of an academic contribution.

This distinction is especially important, because similarity reports are not a verdict of plagiarism but technical indicators that must be interpreted in context. Turnitin's own guides likewise state that the similarity score does not amount to a plagiarism decision, that the score should be evaluated within its academic context, and that a low similarity score does not guarantee the absence of plagiarism (Turnitin, n.d.-b). Just as high similarity may arise from legitimate citation, shared terminology, or the necessity of methodological expression, low similarity does not prove that the academic contribution is original. The critique offered in this study is therefore directed less at Turnitin's own stated purpose of use than at the expanded assurance role that similarity scores can effectively take on in publishing processes.

In publishing processes, similarity scores can at times be used as a de facto acceptance, rejection, or revision threshold in editorial pre-screening. In such cases, a technically limited indicator becomes, in practice, a broader decision tool regarding academic originality and contribution ownership. This score-centered control practice may be functional against the risk of classical textual plagiarism; but it cannot directly assess whether the idea, the argument flow, or the concluding claim has been taken from another source. The assurance gap therefore arises less from the tool's capacity to generate a report than from the expansion of the meaning that this report carries in academic publishing decisions.

The findings show this limit clearly. Turnitin recognized the original texts at almost an exact match; but in the structurally rewritten versions of those same texts produced through an LLM, the average similarity score fell to 19.08%. Despite this, the semantic preservation coding shows that the main idea, the argument flow, and the concluding claim of the source texts were largely preserved. The study's critical finding is not merely that the similarity score falls; it is that this decrease occurs together with the preservation of the source idea. A low similarity score therefore shows that textual traces have decreased; but it does not prove intellectual originality or contribution ownership.

5.2. The Conceptual Limit of the AI Score

The second important conclusion of the findings is that the AI writing score cannot be interpreted as an indicator of plagiarism. AI writing reports can provide a signal about the likely mode of production of a text. But this signal does

not directly determine to whom the idea presented in the text belongs, from which source the argument derives, or whether the academic contribution is original. The AI score and the assessment of plagiarism should therefore be conceptually separated.

The control texts made this distinction especially visible. In all five control texts, which rest on the study's own original seed ideas, the Turnitin similarity score came to 0%. By contrast, the AI writing scores of these texts are quite high; an AI score of 100% was observed in four texts and 64% in one. This result establishes that a high AI score does not show that an idea belonging to someone else has been appropriated. The control texts may have been written with AI assistance; but their intellectual core consists of the study's own arguments.

The main experimental set supports the same distinction. In some rewritten texts the AI score was high, while in some files the AI rate remained too low to be disclosed. Even so, what is decisive for idea laundering is not whether the AI score is high or low, but whether the source idea has been preserved. A high AI score is therefore not evidence of plagiarism; nor does a low AI score show that the risk of idea laundering is absent. AI writing reports may offer auxiliary signals in publishing processes; but they cannot take the place of an assessment of plagiarism, contribution ownership, or idea laundering.

5.3. The Shift Toward a Paradigm of Idea Laundering and Contribution Ownership

The conceptual contribution of this study is that it extends the debate over academic plagiarism from a paradigm of textual similarity toward a paradigm of contribution ownership. Classical textual plagiarism concerns the direct, or very close, transfer of expressions belonging to someone else. In such violations, text similarity reports can produce strong warning signals. Idea laundering, however, concerns the appropriation not of a text's words but of its academic contribution. The same control logic may therefore fall short in capturing this risk.

Idea laundering is the stripping of the textual traces of an academic idea, argument structure, or contribution claim belonging to another researcher, through rewriting, and its presentation as an original contribution. Here the ethical violation lies not merely in the similarity of sentences, but in concealing the source of the intellectual labor. Academic integrity requires attribution and traceability not only at the level of words, but also at the level of contribution.

This approach requires the concept of "originality" to be rethought in publishing processes. Originality is not a text's appearing different from earlier texts, but the author's offering their own intellectual contribution to the literature. This contribution may draw on earlier work; the production of academic knowledge most often develops in relation to the prior literature. But this relationship must be open, attributed, and traceable. Idea laundering poses a serious risk to the credibility of academic publishing precisely because it breaks this traceability.

5.4. Publication Governance, Internal Control, and Source–Argument Traceability

When the publishing process is treated as a control environment, these findings make the design limits of the existing automated controls more visible. The problem is not that similarity reports or AI writing reports are wholly ineffective; the problem is that these reports cannot, on their own, take on the function of guaranteeing the source of the intellectual contribution. The risk of idea laundering should therefore be evaluated, within the existing control architecture, as a contribution ownership risk that cannot be monitored solely through indicators of direct textual overlap or mode of production.

The proposed approach is therefore not to abandon automated tools, but to support these tools with complementary controls within a stronger publication governance architecture. Similarity and AI writing reports can serve an early-warning function; but against the risk of idea laundering, they need to be supported by source–argument traceability, semantic evaluation, editorial judgment grounded in domain expertise, and, where necessary, stylometric or retrieval-based techniques.

This proposal carries a methodological parallel with the retrieval-based approaches developed against paraphrasing in the detection of AI-generated text (Krishna et al., 2023), yet it is directed at a different assurance problem: not the diagnosis of AI production, but the traceability, at the level of source and argument, of an academic contribution produced by a human. The approach proposed here therefore aims less at designing an AI text detector than at strengthening the traceability of intellectual contribution ownership in academic publishing processes.

Looked at prospectively, this suggests a direction in which text matching systems themselves may evolve. Retrieval-based tracing would mean that a control system no longer asks only which sentences of a manuscript match existing sources, but also which published works are the closest semantic and argumentative neighbors of the manuscript, returning candidate sources for editorial comparison even when word-level overlap has been stripped away. Semantic embedding indexes of the published literature, argument-level comparison, and stylometric signals could support such

tracing. Krishna et al. (2023) provide evidence that retrieval over a reference corpus is considerably more robust to paraphrasing than detectors that rely on surface or generative traces, although even retrieval-based defenses can be degraded by recursive paraphrasing (Sadasivan et al., 2023). The realistic function of such systems in publishing is therefore not automated adjudication but candidate generation: bringing the most plausible source texts in front of the editor or reviewer, which is precisely the point at which the risk of idea laundering currently escapes the control architecture.

This framework is also consistent with the multi-layered understanding of assurance in the internal control literature. When the publishing process is evaluated as a control environment, the similarity and AI reports are detective controls that are not sufficient on their own. Editorial review, peer review, the contribution statement, source–argument comparison, and domain expertise, in turn, are assurance layers that complement these controls. This framework is consistent with the COSO (2013) internal control approach and the IIA (2020) Three Lines Model. An effective approach against the risk of idea laundering is not a decision based on a single score, but a publication governance model in which these layers work together.

5.5. A Draft Semantic Evaluation Checklist for Editors and Reviewers

The multi-layered governance approach defended above remains abstract unless it can be translated into questions that editors and reviewers can actually ask. As a first step in that direction, and as a draft rather than a validated instrument, Table 8 gathers the study's conceptual framework into a semantic evaluation and source–argument traceability checklist. The checklist is a judgment-support tool: it does not produce a score, it does not replace domain expertise, and it is not a decision algorithm. Its purpose is to structure the editorial reading of a manuscript around contribution ownership rather than around the two automated indicators alone.

Table 8. Draft Semantic Evaluation and Source–Argument Traceability Checklist for Editors and Reviewers

Control Area	Guiding Questions for Editors and Reviewers
Contribution Claim	What is the manuscript's explicit contribution claim? Is it articulated in the authors' own analytical voice and connected to their own data, reasoning, or conceptual work?
Nearest Neighbors in The Literature	Which published works are conceptually closest to the manuscript's main argument? Are they cited, and is the manuscript's difference from them stated?
Argument-Flow Comparison	Does the sequence of premises, reasoning steps, and concluding claim parallel a specific prior work, even where the wording does not?
Source–Argument Traceability	Can each substantive claim be traced either to a cited source or to the authors' own analysis? Are there load-bearing arguments with no visible origin?
Citation Texture	Does the manuscript substantively engage the sources it cites, or do citations remain peripheral while the core argument resembles an uncited work?
AI-Use Disclosure	Is the declared use of AI tools consistent with journal and publisher policy, and plausible given the characteristics of the text?
Similarity Report İnterpretation	Is the similarity score treated as an initial screening indicator rather than an originality verdict? Is a low score explicitly treated as inconclusive with respect to contribution ownership?
AI Writing Report İnterpretation	Is the AI writing score treated strictly as a production-mode signal, and explicitly not as an indicator of plagiarism or contribution ownership?
Escalation Path	Where doubts about contribution ownership persist, is there a defined escalation step (author query, an additional domain reviewer, referral to an ethics body) rather than reliance on automated indicators alone?

(The table was created by the authors.)

The checklist deliberately places the two automated reports late in the sequence. In a control environment where idea laundering is a recognized risk, the similarity report and the AI writing report function as initial screening controls whose output requires interpretation; the substantive assurance work lies in the traceability of the contribution claim. Because the checklist is derived from a small proof-of-concept study, it should be treated as a starting point: its applicability, item wording, and inter-rater consistency remain to be tested in real editorial settings.

To operationalize this checklist without increasing the administrative burden on reviewers, journals could consider a tiered evaluation: initial screening dimensions (such as AI disclosures and similarity reports) can be processed during the editorial triage, while deep semantic dimensions (such as argument-flow and citation texture) are routed to domain experts. Future validation studies should investigate the inter-rater reliability of this instrument across different disciplinary traditions, particularly comparing high-context humanities versus more structured, data-driven or quantitative social sciences.

5.6. Limitations and Future Research

This study is built on a limited sample and a controlled experimental protocol. There are essentially four limitations within the scope of the study. First, the sample is limited to twelve social science articles, and only discussion/conclusion sections were used as source texts. Results may differ for source texts in different disciplines, of different text types, in different languages, or of different lengths. The findings should therefore not be generalized to the entire universe of academic publishing.

The second limitation is that the rewriting conditions rest on a single-run, first-output basis. This choice was adopted in order to preserve realistic conditions of use and to avoid optimizing the results. However, different prompts, different models, different user interventions, or manual editing may produce different results. The study was carried out with the current, publicly available LLM versions accessible during the period in which it was conducted. Because of the rapid evolution of LLMs, the findings should not be generalized directly to future model versions or to different modes of operation; in this respect, the study is in the nature of a snapshot showing the publication-control risk at a particular point in time. In addition, the structural prompt used in the study does not, for ethical reasons, contain operational evasion instructions; it is likely that the risk of idea laundering that explicitly adversarial prompts would expose in the real world is sharper than the pattern observed in this study. The emerging evidence on adversarial techniques supports this expectation. Deliberate manipulations that require no advanced technical skill, such as prompting strategies chosen to reduce detectability, chaining rewriting tools in sequence, regenerating output until an automated indicator falls, or lightly hand-editing machine output, have been shown to degrade the accuracy of detection systems substantially (Perkins et al., 2024; Sadasivan et al., 2023). A determined actor can, in other words, treat the detection layer itself as an optimization target. The single-run, first-output protocol used here should therefore be read as a conservative lower bound on the exposure rather than as its ceiling.

The third limitation is that the semantic preservation coding rests on the structured evaluation of two expert raters. Achieving 100% agreement between the raters strengthens the consistency of the findings. At the same time, the concentration of all the scores in a single category limits the interpretive value of more complex reliability coefficients. Future studies are advised to use a larger sample, more coders, and the measurement of semantic preservation by breaking it down into sub-dimensions such as the main idea, the reasoning, the contribution claim, and the concluding claim.

The fourth limitation is that the study was conducted on the basis of Turnitin outputs. Turnitin is a representative tool widely used in academic publishing processes; but the findings should not be generalized directly to all text matching or AI writing detection systems. The databases, algorithms, language support, and reporting formats of different systems may produce different results. It should also be borne in mind that Turnitin AI writing reports can be generated only under certain file and text eligibility conditions (Turnitin, n.d.-d). For this reason, rather than delivering a final verdict on a particular piece of software, the study shows the limits of a control logic based on similarity and AI writing reports in the face of the risk of idea laundering.

In future research, the sample may be enlarged, different disciplines and languages may be compared, multiple text matching systems may be evaluated together, and the question of how retrieval-based source–argument tracking tools might be used in publishing processes may be investigated. In addition, the draft semantic evaluation checklist proposed in Section 5.5 may be validated in real editorial settings, its items refined, and its inter-rater consistency examined.

6. CONCLUSION

This study examined, through a limited proof-of-concept design, the extent to which the combined use of the similarity report and the AI writing report in academic publishing processes can provide assurance against the risk of idea laundering. The findings showed that Turnitin can recognize the original source texts with high similarity rates; but that when the same texts are structurally rewritten through an LLM, the similarity scores fall sharply in most cases. Despite this, the semantic preservation assessment carried out by two independent experts revealed that the main idea, the argument flow, and the concluding claim of the source texts were largely preserved.

This result shows that a low similarity score should not be interpreted as a certificate of academic originality. Low similarity shows that textual overlap is limited; but it does not prove that the idea, the argument, or the academic contribution in the text is original. In the same way, the AI writing score is not an indicator of plagiarism. As seen in the control texts, AI-assisted texts that rest on an original idea can produce a high AI writing score. By contrast, a text that carries an idea belonging to someone else can produce low similarity or a low AI score. An assessment of academic ethics therefore cannot be confined to score-based technical indicators alone.

The findings of this study support, at the level of a limited proof of concept, the view that the debate over academic integrity cannot be confined to the question of surface textual similarity or of whether a text was produced by AI; it must be broadened to include the dimension of the source of the intellectual contribution and of contribution ownership. The problem is not writing with AI-assisted tools, but rewriting an academic idea belonging to someone else through AI and presenting it as an original contribution. The fundamental question in publishing processes should therefore not be only "how similar is the text?" or "was the text written with AI?" The more critical question is: "to whom do the idea, the argument, and the contribution claim appropriated in this text belong?"

From an audit and internal control perspective, the Turnitin similarity report and the AI writing report are automated detective controls. These controls are valuable; but they are not sufficient on their own. Publication governance must support these controls with complementary control layers such as editorial judgment, domain expertise, semantic evaluation, the contribution statement, and source–argument traceability. The draft checklist presented in Section 5.5 of this study is intended as a first operational step in that direction. Such an approach both avoids the unnecessary criminalization of legitimate AI-assisted writing practices and provides stronger assurance against the risk of an idea belonging to someone else being appropriated by reducing its textual traces.

In conclusion, the control approach that academic publishing processes require in the age of AI cannot rest on tools that measure only textual similarity or the likelihood of AI production. These tools acquire their meaning within a broader audit of publication governance and academic contribution ownership. The risk of idea laundering calls for extending the debate over academic ethics from the textual surface toward contribution traceability.

References

Bretag, T., & Mahmud, S. (2009). A model for determining student plagiarism: Electronic detection and academic judgement. Journal of University Teaching & Learning Practice, 6(1), 57–69. https://doi.org/10.53761/1.6.1.6

Committee of Sponsoring Organizations of the Treadway Commission. (2013). Internal control—Integrated framework: Executive summary.

Committee on Publication Ethics. (2023, February 13). Authorship and AI tools: COPE position statement. https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools

Eaton, S. E. (2023). Postplagiarism: Transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. International Journal for Educational Integrity, 19, Article 23. https://doi.org/10.1007/s40979-023-00144-1

Elsevier. (n.d.). The use of generative AI and AI-assisted technologies in writing for Elsevier. Retrieved July 4, 2026, from https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. npj Digital Medicine, 6, Article 75. https://doi.org/10.1038/s41746-023-00819-6

Howard, R. M. (1995). Plagiarisms, authorships, and the academic death penalty. College English, 57(7), 788–806. https://doi.org/10.2307/378403

Institute of Internal Auditors. (2020). The IIA's three lines model: An update of the three lines of defense.

Jamieson, S. (2015). Is it plagiarism or patchwriting? Toward a nuanced definition. In T. Bretag (Ed.), Handbook of academic integrity. Springer, Singapore. https://doi.org/10.1007/978-981-287-079-7_68-1

Karahan, Ç., Velioğlu, H., & Arslan, Y. (2025). Unveiling the shadows: Anticipating future cyber risks and their impacts on businesses. In H. Kıral & G. Yılmaz (Eds.), *Futurisks: Risk management in the digital age* (pp. 17–55). Springer. https://doi.org/10.1007/978-981-96-6448-1_2

Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking of plagiarism detection. In P. Zaphiris & A. Ioannou (Eds.), Learning and collaboration technologies (pp. 475–487). Springer, Cham. https://doi.org/10.1007/978-3-031-34411-4_32

Krishna, K., Song, Y., Karpinska, M., Wieting, J., & Iyyer, M. (2023). Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense. arXiv preprint arXiv:2303.13408. https://doi.org/10.48550/arXiv.2303.13408

Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., Cao, H., Liu, S., He, S., Huang, Z., Yang, D., Potts, C., Manning, C. D., & Zou, J. Y. (2024). Mapping the increasing use of LLMs in scientific papers. arXiv preprint arXiv:2404.01268. https://doi.org/10.48550/arXiv.2404.01268

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. Journal of the Association for Information Science and Technology, 74(5), 570–581. https://doi.org/10.1002/asi.24750

Meo, S. A., & Talha, M. (2019). Turnitin: Is it a text matching or plagiarism detection tool? Saudi Journal of Anaesthesia, 13(Suppl 1), S48–S51. https://doi.org/10.4103/sja.SJA_772_18

Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. Journal of Second Language Writing, 12(4), 317–345. https://doi.org/10.1016/j.jslw.2003.08.004

Perkins, M., Roe, J., Postma, D., McGaughran, J., & Hickerson, D. (2024). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. Journal of Academic Ethics, 22(1), 89–113. https://doi.org/10.1007/s10805-023-09492-6

Prentice, F. M., & Kinden, C. E. (2018). Paraphrasing tools, language translation tools and plagiarism: An exploratory study. International Journal for Educational Integrity, 14, Article 11. https://doi.org/10.1007/s40979-018-0036-7

Rogerson, A. M., & McCarthy, G. (2017). Using internet based paraphrasing tools: Original work, patchwriting or facilitated plagiarism? International Journal for Educational Integrity, 13, Article 2. https://doi.org/10.1007/s40979-016-0013-y

Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-generated text be reliably detected? arXiv preprint arXiv:2303.11156. https://doi.org/10.48550/arXiv.2303.11156

Shi, L. (2012). Rewriting and paraphrasing source texts in second language writing. Journal of Second Language Writing, 21(2), 134–148. https://doi.org/10.1016/j.jslw.2012.03.003

Turnitin. (n.d.-a). Understanding the similarity score. Retrieved May 25, 2026, from https://guides.turnitin.com/hc/en-us/articles/23435833938701-Understanding-the-similarity-score

Turnitin. (n.d.-b). Turnitin and plagiarism. Retrieved May 25, 2026, from https://guides.turnitin.com/hc/en-us/articles/34400565079053-Turnitin-and-plagiarism

Turnitin. (n.d.-c). Using the AI writing report. Retrieved May 25, 2026, from https://guides.turnitin.com/hc/en-us/articles/22774058814093-Using-the-AI-Writing-Report

Turnitin. (n.d.-d). File requirements for an AI writing report. Retrieved May 25, 2026, from https://guides.turnitin.com/hc/en-us/articles/28234943089933-File-requirements-for-an-AI-writing-report

Walker, J. (1998). Student plagiarism in universities: What are we doing about it? Higher Education Research & Development, 17(1), 89–106. https://doi.org/10.1080/0729436980170105

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19, Article 26. https://doi.org/10.1007/s40979-023-00146-z