

Impact of Paraphrasing Tools on Plagiarism Detection

Assistant Professor Smita Mangesh Junnarkar¹

Department of Data Science, SIES COLLEGE OF ARTS,
SCIENCE AND COMMERCE (EMPOWERED
AUTONOMOUS), Sion(W), Mumbai.

¹smita@sies.edu.in

Surya Subramani²

Department of Data Science, SIES COLLEGE OF ARTS,
SCIENCE AND COMMERCE (EMPOWERED
AUTONOMOUS), Sion(W), Mumbai.

²suryasubramani2006@gmail.com

Abstract

Paraphrasing tools have become common in academic writing because they can quickly restructure sentences, improve fluency, and help writers express an existing idea in different language. At the same time, the same capability creates a challenge for plagiarism detection: when copied material is changed at the lexical and syntactic levels while its meaning and structure remain similar, systems based mainly on exact or near-exact word matching may produce a lower similarity signal. This paper examines the impact of automated paraphrasing on plagiarism detection from an academic-integrity perspective. Rather than treating a low similarity score as proof of originality, the study considers the difference between textual overlap, semantic similarity, and legitimate paraphrasing. A structured literature-based methodology is used to compare findings from research on plagiarism detection, automated paraphrasing, semantic similarity, and academic writing. The review indicates that paraphrase plagiarism is harder to detect than direct copying, while semantic and contextual methods provide additional evidence that lexical matching alone can miss. The paper also identifies an important distinction between using paraphrasing tools for language support and using them to disguise unattributed source material. The findings suggest that plagiarism detection should be treated as an evidence-support system rather than an automatic final judge. A combined approach involving semantic similarity, source tracing, citation checking, linguistic analysis, and human review is therefore discussed. The paper concludes with recommendations for responsible student use and directions for future research, including multilingual evaluation and the study of newer large-language-model-based paraphrasing systems.

Keywords: paraphrasing tools; plagiarism detection; academic integrity; semantic similarity; natural language processing; text matching; artificial intelligence

1. Title of the Study and Purpose of the Study

Title of the study: *Impact of Paraphrasing Tools on Plagiarism Detection*

The purpose of this study is to understand how automated paraphrasing changes the form of academic text and how those changes influence plagiarism-detection systems. The topic is important because paraphrasing has two very different academic uses. A student may use a tool to improve grammar or to understand how an idea can be expressed more clearly. However, a student may also copy an idea from a source, pass it through a paraphrasing system, remove obvious wording similarities, and submit the result without attribution. The second situation is not made academically original merely because the words are different.

This paper therefore focuses on a central question: *when paraphrasing changes the wording of a source while preserving its underlying meaning, what happens to plagiarism detection, and what forms of analysis can reduce that gap?*

Research Questions

1. How do automated paraphrasing tools change the lexical and syntactic appearance of source text?
2. Why can conventional text-matching approaches miss strongly paraphrased plagiarism?
3. How can semantic similarity and contextual NLP methods improve the identification of paraphrased source use?
4. What academic-integrity risks arise when students use paraphrasing tools without understanding citation and attribution?
5. What future research directions could make plagiarism detection more reliable against increasingly capable paraphrasing systems?

Note: This paper is written as a student research study based on a structured synthesis of published research. Numerical performance values reported from other studies are attributed to those studies and are not presented as original experiments conducted by the author.

2. Background and Context

Academic plagiarism is not limited to copying a sentence word-for-word. It can also involve taking another person's ideas, structure, argument, or wording and presenting them as one's own without appropriate acknowledgement. Computational plagiarism detection has traditionally relied on comparing a suspicious document with a collection of source documents and identifying overlapping strings or other forms of textual similarity. However, research has shown that text-matching software should be understood as a support mechanism rather than a perfect plagiarism judge. Foltýnek et al. found that detection systems can miss some plagiarism and can also flag material that is not actually plagiarized [4].

The problem becomes more complicated when a source is paraphrased. A direct copy may preserve long sequences of the same words, making lexical matching relatively straightforward. A paraphrased version can replace words, change grammatical structure, merge or split sentences, reorder information, or introduce additional words. If these operations are extensive, the surface form of the text can look substantially different even though the central proposition remains similar. Research on paraphrase plagiarism has therefore treated semantic relatedness as an important part of the detection problem [1, 2].

Automated paraphrasing tools were originally developed from broader natural-language-generation research. Paraphrasing itself is a legitimate language task: it attempts to preserve meaning while expressing the same content in a different form. Madnani and Dorr's survey describes a range of computational approaches to generating phrasal and sentential paraphrases, showing that paraphrase generation has a long history within natural language processing (NLP) [5]. Modern systems can perform much more fluent transformations than early word-substitution tools.

This creates an important academic distinction. A paraphrase can be acceptable when a writer has understood the source, expresses the idea independently, and gives the source appropriate credit. A paraphrase can also constitute plagiarism when it is used to disguise dependence on a source or when the source's ideas are presented without attribution. Roe and Perkins describe automated paraphrasing tools as part of a developing technological arms race between techniques that can facilitate academic-integrity violations and techniques designed to identify them [8].

2.1 Why Similarity Percentage Is Not the Same as Plagiarism

A similarity score should not be interpreted as a direct percentage of plagiarism. Similarity systems identify matching or related material according to their algorithms and databases. Common phrases, titles, references, technical terminology, and properly quoted material may contribute to similarity even when no misconduct has occurred. Conversely, a document can contain unattributed ideas that are expressed with sufficiently different wording to produce a low lexical similarity score. This distinction is one of the main reasons that human interpretation remains important in academic-integrity decisions [3, 4].

2.2 The Role of Paraphrasing in Student Writing

Paraphrasing is also a genuine academic skill. Good paraphrasing requires a student to understand the original passage, identify the important idea, reconstruct it using an independent sentence structure, and retain the original meaning. Prentice and Kinden observed that automated paraphrasing could produce unusual synonym choices and other linguistic clues, while also noting that the quality of tool output varied [6]. Rogerson and McCarthy similarly showed that online paraphrasing tools can create risks when students use them as a shortcut instead of developing their own paraphrasing skills [7].

The present study treats these issues together. The question is not whether paraphrasing tools are automatically good or bad. Instead, it examines how their technical operation interacts with plagiarism detection and how that interaction affects academic integrity.

3. Review of Related Literature

3.1 Development of Plagiarism Detection

Plagiarism detection research has moved from simple string matching toward methods that can compare text at deeper semantic levels. Foltýnek, Meuschke, and Gipp reviewed 239 research papers published between 2013 and 2018 and reported substantial development in the detection of strongly obfuscated plagiarism, particularly through semantic analysis, machine learning, and analysis of additional document features [3]. This progression is relevant because paraphrasing is one of the mechanisms through which surface-level similarity can be reduced.

Traditional approaches may use character n-grams, word n-grams, term-frequency representations, or alignment methods. These methods remain useful because exact copying and light modifications can be detected efficiently. Their limitation appears when the same idea is expressed with different vocabulary and sentence structure. In such cases, the system needs to determine whether two passages are related in meaning rather than simply whether they share many words.

3.2 Paraphrase Plagiarism

Paraphrase plagiarism has been identified as a difficult form of plagiarism because the plagiarized text can be intentionally modified. Álvarez-Carmona et al. proposed semantically informed similarity and distance measures to address this problem, demonstrating why semantic information can provide features beyond ordinary lexical overlap [1]. Alvi, Stevenson, and Clough investigated paraphrase types such as synonym substitution and word reordering and proposed methods using context matching and word embeddings [2].

These approaches are important for understanding the impact of paraphrasing tools. If a paraphraser changes the words but retains the concepts and relationships between them, then a detector based only on exact word overlap loses information. An embedding-based or contextual system can represent a passage in a way that captures more of its meaning. The detector can then compare the representations of the suspicious text and candidate source rather than relying entirely on shared vocabulary.

3.3 Evidence from Educational Settings

Prentice and Kinden conducted an exploratory study after observing student essays that contained unusual language and inappropriate substitutions for discipline-specific terminology. They tested free paraphrasing tools and found that some tools replaced established medical terminology with unexpected synonyms [6]. An especially relevant observation was that some suspicious submissions received low similarity results despite showing evidence of being derived from another text. This illustrates the difference between a low matching score and genuine originality.

Rogerson and McCarthy examined internet-based paraphrasing tools and discussed the possibility that students could use them to alter existing work. Their experiment also highlighted weaknesses in originality checking when text was processed through machine paraphrasing [7]. These findings should be interpreted in their historical context because paraphrasing systems have changed significantly since those experiments. Nevertheless, the underlying technical issue remains relevant: changing surface form can reduce the effectiveness of a detector that depends heavily on surface overlap.

3.4 Modern Semantic Detection

More recent research has investigated deep-learning methods for paraphrase identification. Vrbanec and Meštrović compared numerous unsupervised approaches and reported that some deep-learning methods were more successful than statistical similarity methods for concealed paraphrase detection [9]. Other work has investigated contextual word embeddings and transformer-based models for paraphrase-plagiarism identification [2]. These studies support a broader shift from the question “*Do these texts contain the same words?*” toward “*Do these texts communicate substantially the same information?*”

At the same time, stronger detection does not mean perfect detection. Foltýnek et al. emphasized that plagiarism-support systems have limitations and that results require interpretation [4]. Therefore, semantic models should be viewed as another source of evidence rather than an automatic substitute for academic judgement.

3.5 Recent AI-Paraphrasing Concerns

The arrival of generative AI has expanded the paraphrasing problem. Modern systems can rewrite passages with more natural syntax and can perform substantial structural changes while maintaining the main message. Research on AI-generated-text detection has also shown that automated detectors can be affected by paraphrasing and other forms of text transformation [10]. Turnitin introduced an AI-paraphrasing detection feature in 2024 specifically to help identify situations in which AI-generated text may have been modified through paraphrasing tools [11]. The existence of such a feature does not establish that every paraphrased passage is misconduct; it demonstrates that paraphrasing has become an explicit concern in integrity technology.

4. Conceptual Framework

The study uses a simple conceptual framework linking four stages: source material, paraphrasing transformation, detection response, and human academic judgement. The framework is useful because it separates the technical effect of paraphrasing from the ethical question of whether the underlying source was properly acknowledged.

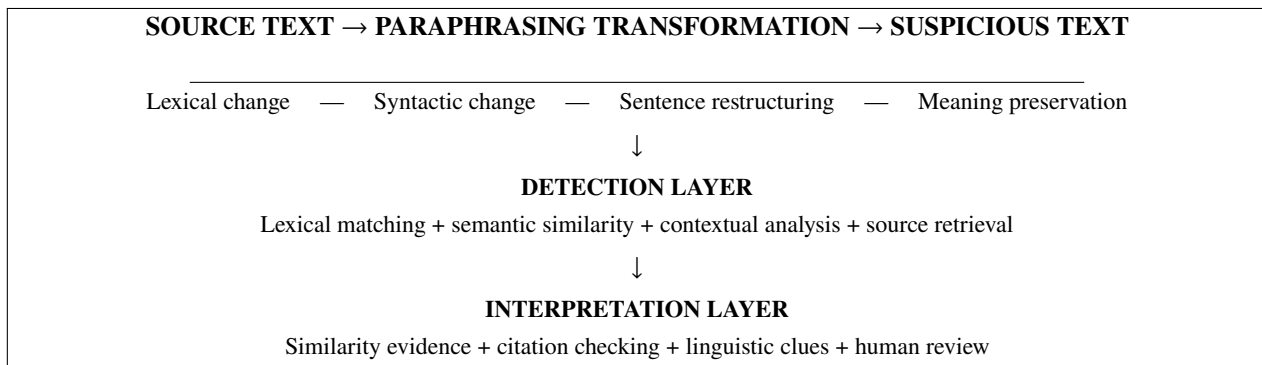


Figure 1: Conceptual framework for analysing paraphrasing and plagiarism detection.

4.1 Layer 1: Lexical Similarity

Lexical similarity examines visible word overlap. A simplified representation is the cosine similarity between two document vectors:

$$\text{Cosine}(A, B) = \frac{A \cdot B}{\|A\| \|B\|}. \quad (1)$$

When two passages use many of the same terms, their vector representations can be close. However, synonym substitution and sentence restructuring can lower this score even when the underlying idea is substantially preserved. Lexical similarity is therefore useful but incomplete.

4.2 Layer 2: Semantic Similarity

Semantic similarity attempts to represent meaning rather than only word identity. Modern NLP models can convert sentences or passages into numerical embeddings. Two texts can then be compared in an embedding space. If the representations remain close after paraphrasing, the detector has a stronger opportunity to identify a relationship between the suspicious passage and its source.

This does not mean that a high semantic similarity score automatically proves plagiarism. Two independent authors may discuss the same well-known concept using similar meaning. Technical definitions may also have limited acceptable variation. The

score should therefore be combined with source evidence and contextual interpretation.

4.3 Layer 3: Context and Structure

Paraphrasing may preserve more than meaning. A transformed passage may retain the order of ideas, unusual examples, citation sequence, argument structure, or discipline-specific concepts. These features can be useful clues. Alvi et al. specifically studied paraphrase types involving synonym substitution and word reordering, illustrating how identifying transformation patterns can complement ordinary similarity measures [2].

4.4 Layer 4: Human Review

The final layer is human interpretation. A reviewer can examine the source, the suspicious passage, the citations, the similarity evidence, and the context in which the work was submitted. This is particularly important when automated systems produce false positives or false negatives. The literature consistently supports treating detection software as evidence-support technology rather than as an infallible decision-maker [4].

Table 1: Comparison of major detection perspectives

Approach	Main signal	Strength	Limitation
Exact / lexical matching	Shared words or phrases	Effective for direct copying	Sensitive to wording changes
N-gram / alignment	Local sequence overlap	Detects light modification	Weaker after strong rewording
Semantic embeddings	Meaning-related representation	Better for paraphrased text	Can confuse related but independently written ideas
Contextual models	Meaning plus context	Captures complex relationships	Computational and model-dependent
Human review	Source, context and evidence	Can integrate multiple clues	Time-consuming and subjective

5. Methodology

This study follows a structured literature-review and comparative analysis methodology. The objective is not to claim an original benchmark score for a particular commercial detector. Instead, the method combines established research findings to explain how paraphrasing affects detection and which technical approaches are relevant to the problem.

5.1 Research Design

The research design contains four stages:

1. **Problem definition:** Define paraphrase plagiarism as the use of another source's content or ideas in altered language without adequate attribution.
2. **Literature identification:** Select peer-reviewed research dealing with plagiarism detection, paraphrase detection, automated paraphrasing, semantic similarity, and academic integrity.
3. **Thematic comparison:** Group the evidence into detection methods, paraphrasing effects, educational risks, and possible improvements.
4. **Synthesis:** Connect the technical evidence to a practical academic-integrity framework that combines automated signals with human review.

5.2 Evidence Categories

The selected literature was organized into five evidence categories. First, foundational work was used to define computational plagiarism detection and paraphrase generation. Second, studies of paraphrase plagiarism were used to understand why lexical similarity can fail. Third, educational studies were used to connect the technical problem to student writing. Fourth, recent AI-related work was considered because modern generative systems can produce more fluent transformations. Finally, institutional technology documentation was used only to describe documented product developments, such as the introduction of AI-paraphrasing detection features.

Table 2: Representative studies used in the literature synthesis

Study	Year	Focus	Relevant contribution	Use in this paper
Madnani & Dorr	2010	Paraphrase generation	Reviews computational paraphrase-generation methods	Defines paraphrasing as an NLP task
Rogerson & McCarthy	2017	Online paraphrasing	Examines integrity risks and detection limitations	Establishes educational context
Prentice & Kinden	2018	Student writing	Examines linguistic clues from paraphrasing tools	Supports discussion of tool-generated language
Foltýnek et al.	2019/2020	Plagiarism detection	Reviews and tests text-matching systems	Establishes limits of automated detection
Alvi et al.	2021	Paraphrase types	Uses context and embeddings to identify transformations	Supports semantic/contextual detection
Roe & Perkins	2022	Automated paraphrasing	Reviews APTs and academic integrity	Frames the technological arms race
Vrbanec & Meštrović	2023	Paraphrase detection	Compares similarity and deep-learning methods	Supports semantic approaches
Weber-Wulff et al.	2023	AI-text detection	Tests detection under obfuscation/paraphrasing	Shows reliability concerns
Turnitin	2024	AI paraphrasing feature	Documents a product-level detection development	Shows practical industry response

5.3 Analytical Criteria

The literature was compared using five criteria: (1) ability to preserve or identify meaning, (2) sensitivity to lexical changes, (3) ability to identify source relationships, (4) interpretability of the evidence, and (5) suitability for educational use. These criteria were chosen because a detection system should not only produce a score; it should also help an instructor understand why a passage has been flagged.

5.4 Ethical and Academic Integrity Considerations

The study deliberately avoids treating paraphrasing as synonymous with plagiarism. The ethical concern arises from the relationship between the source and the writer’s submitted work. A student who uses a tool to improve grammar in a paragraph they genuinely wrote is in a different situation from a student who copies a published paragraph, runs it through a paraphraser, and submits it without citation. Institutional rules may differ regarding acceptable AI assistance, so students should follow the policy of their institution and course.

5.5 Limitations of the Method

Because this is a literature-based study, it does not independently test a fixed sample of documents across a fixed set of detectors. Results from different studies cannot be directly combined into one accuracy number because their datasets, languages, tools, thresholds, and evaluation criteria differ. The paper therefore reports patterns across the literature rather than creating a new performance benchmark.

6. Key Findings

The literature synthesis produces six main findings. These findings are presented as a reasoned interpretation of published evidence rather than as results from an original experiment.

6.1 Finding 1: Paraphrasing Can Reduce Surface-Level Detectability

The first finding is that paraphrasing can reduce the visible overlap between a source and a submitted passage. When words are replaced and sentence structures are changed, systems that depend heavily on matching word sequences have fewer identical strings to compare. Rogerson and McCarthy’s experimental discussion and Prentice and Kinden’s exploratory work both illustrate why machine paraphrasing can create a gap between actual source dependence and the similarity signal returned by a text-matching system [7, 6].

This does not mean that paraphrasing always defeats detection. Light paraphrasing may leave enough wording overlap for conventional systems to identify the source. The impact depends on the degree and type of transformation, the detector’s database, and the detector’s underlying algorithm.

6.2 Finding 2: Meaning Is More Stable Than Wording

The second finding is that semantic relationships can remain detectable even when surface wording changes. Research on paraphrase plagiarism has therefore moved toward semantic similarity, word embeddings, contextual representations, and neural models [1, 2, 9]. The basic idea is straightforward: two passages may use different words but still describe the same event, process, relationship, or conclusion.

For example, consider the following simplified transformation:

Source: “Regular physical activity can reduce the risk of several chronic diseases.”

Paraphrase: “Maintaining an active lifestyle is associated with a lower likelihood of developing a number of long-term illnesses.”

The words are different, but the central proposition remains closely related. A semantic system is designed to notice that relationship. A lexical system may observe less direct overlap.

6.3 Finding 3: Strong Detection Requires More Than One Signal

The third finding is that no single signal is sufficient for every case. Lexical matching is useful for direct copying; semantic similarity is useful for paraphrased relationships; citation analysis can identify missing attribution; and human review can examine

context. Foltýnek et al.’s evaluation of support tools demonstrates that automated systems can miss plagiarism and can also raise problematic matches [4]. This supports a layered detection model.

Table 3: Expected effect of paraphrasing intensity on detection signals

Text condition	Lexical overlap	Semantic relationship	Human source tracing
Direct copy	Very high	Very high	Straightforward
Light rewording	High to medium	High	Usually straightforward
Strong paraphrase	Low to medium	Potentially high	Requires closer review
Structural rewrite	Low	Medium to high	Requires contextual evidence
Independent writing on same topic	Variable	Variable	Needs judgement

6.4 Finding 4: A Low Similarity Score Does Not Prove Originality

A low similarity result can be reassuring only when interpreted in context. If a detector has a limited source database, it may not contain the source material at all. A heavily paraphrased passage can also produce low lexical overlap. Consequently, similarity scores should be treated as indicators of detected overlap rather than as complete measurements of originality. This distinction is particularly important when decisions about academic misconduct are based on software output [3, 4].

6.5 Finding 5: Paraphrasing Tools Can Affect Writing Quality as Well as Detection

The impact is not purely technical. Prentice and Kinden found that some automated paraphrasing outputs contained unusual synonyms and discipline-inappropriate terminology [6]. Such linguistic changes can themselves become clues. A paper may have a low similarity score while still displaying unnatural vocabulary, inconsistent terminology, or sudden shifts in writing style.

6.6 Finding 6: Generative AI Makes the Problem More Complex

Modern AI systems can generate more fluent paraphrases than older word-substitution tools. Research into AI-generated-text detection has shown that paraphrasing and other obfuscation methods can reduce detector reliability [10]. The practical implication is that future plagiarism detection cannot depend exclusively on obvious word substitution patterns. It needs to consider meaning, source retrieval, writing context, and multiple forms of evidence.

7. Discussion

The central idea emerging from this study is that plagiarism detection and paraphrase detection are related but not identical problems. Plagiarism is an academic and ethical concept, while paraphrase detection is a computational task. A system can identify that two passages are semantically related, but it cannot determine by itself whether the relationship is acceptable. Context, attribution, authorship, and institutional rules still matter.

7.1 Why Traditional Matching Can Struggle

Traditional matching methods are particularly strong when the suspicious document retains phrases from the source. Their logic is similar to finding matching pieces in two strings. When a paraphraser changes the words and grammar, fewer pieces remain identical. This creates a technical gap: the source relationship is still present at the level of meaning, but the evidence has moved away from exact wording.

The gap can be reduced by representing text semantically. Embedding-based systems map text into numerical vectors. Passages with similar meanings can be close in the resulting representation space even when they use different vocabulary. Research on paraphrase plagiarism has repeatedly explored this principle [1, 2].

However, semantic similarity introduces its own challenge. Similar meaning is not necessarily evidence of copying. Two students answering the same question may naturally produce semantically similar explanations. A detector therefore needs source evidence and context. In other words, a semantic score can identify a relationship worth investigating, but it should not be interpreted as a final accusation.

7.2 The Importance of Source Retrieval

One useful direction is to combine semantic similarity with source retrieval. Instead of asking only whether a student’s document is similar to documents in a database, a system can retrieve candidate sources and show the specific passages that are related. This makes the result more interpretable. An instructor can then inspect whether the student’s passage represents legitimate paraphrasing with a citation or unattributed reuse.

This idea is consistent with research showing that plagiarism detection is increasingly concerned with strongly obfuscated text and semantic analysis [3]. It also fits the broader principle that a detection system should provide evidence rather than merely a numerical score.

7.3 Paraphrasing as a Learning Tool

The findings should not lead to the conclusion that students should never use paraphrasing technology. There are legitimate educational uses. A student may use a tool to see alternative sentence structures, improve readability, identify repetitive wording, or compare their own draft with a clearer formulation. The key issue is authorship and understanding. If the student cannot explain the source idea in their own words, the tool may be replacing rather than supporting the learning process.

Recent research on student perceptions also reports potential benefits such as time savings and vocabulary support while raising concerns about over-reliance and originality [12]. This suggests that institutional guidance should distinguish between language assistance and source obfuscation rather than treating every use of AI paraphrasing as equivalent.

7.4 A Practical Detection Model

Based on the literature, a practical academic workflow can contain five steps:

1. Run conventional text matching to identify direct or lightly modified overlap.
2. Use semantic retrieval or embedding-based comparison to locate passages with related meaning.
3. Inspect citations and references to determine whether the source relationship is acknowledged.
4. Examine linguistic and structural clues, including unusual terminology, preserved argument order, or abrupt changes in style.
5. Conduct human review before making an academic-integrity decision.

Table 4: Proposed layered workflow for academic review

Stage	Technique	Question answered	Evidence produced
1	Lexical matching	Is wording directly or closely repeated?	Matching phrases
2	Semantic retrieval	Is a source passage similar in meaning?	Candidate source passages
3	Citation analysis	Has the source been acknowledged?	Citation/source relationship
4	Linguistic review	Are there transformation clues?	Style and terminology evidence
5	Human review	What does the complete context show?	Final academic interpretation

8. Implications for Students, Teachers and Institutions

8.1 For Students

Students need to understand that changing words does not automatically make an idea their own. A responsible paraphrase begins with comprehension. One practical approach is to read the source, identify the main idea, close the source, explain the idea in one's own words, and then compare the draft with the source to check accuracy. The citation should still be retained because the underlying idea came from another author.

Paraphrasing software can be used as a language-support tool, but the student should review every generated sentence. Automatic changes can introduce incorrect meanings, inappropriate terminology, or claims that were not present in the source. Prentice and Kinden's observations about unusual synonym substitutions demonstrate this risk in an educational setting [6].

8.2 For Teachers

Teachers should avoid interpreting a similarity percentage as a complete academic-integrity diagnosis. A high score may contain legitimate quotations or standard terminology, while a low score does not guarantee that source ideas have not been reused. Assignments can be designed to require source evaluation, explanation of methodology, reflection on the student's reasoning, and appropriate citation. These tasks make it more difficult to substitute automated rewriting for actual understanding.

When a suspicious passage is identified, the source and the student's work should be examined together. The question should be whether the student has appropriately represented and acknowledged the source, not simply whether a software percentage crossed a particular threshold.

8.3 For Institutions

Institutions should provide clear policies about acceptable use of paraphrasing and generative-AI tools. Policies can distinguish between proofreading, language assistance, idea generation, source paraphrasing, and automated production of assessed work. They should also explain how students should acknowledge tool use when institutional rules require disclosure.

Institutions should also be cautious about privacy. Students may upload unpublished assignments, research ideas, personal data, or confidential material to third-party websites. Tool selection should therefore consider data handling and institutional requirements as well as detection performance.

8.4 Why Human Review Remains Necessary

The literature provides a consistent reason for retaining human judgement: detection systems operate on signals, not on complete knowledge of authorship and intention. Even advanced models can produce false positives and false negatives. Foltýnek et al. explicitly describe text-matching software as support for identifying potential plagiarism rather than as a system that independently determines plagiarism [4].

The same principle becomes more important as AI-generated text becomes more fluent. A detector may identify a suspicious pattern, but an instructor can examine the assignment context, student's previous writing, cited sources, draft history, and explanation of the work. None of these should be reduced to one automated score.

8.5 Academic Integrity as a Process

The most useful way to think about plagiarism detection is as a process of evidence collection. The software provides leads. Semantic analysis can broaden the search for related source passages. Citation analysis can establish whether the source was acknowledged. Human review can connect these signals to the academic context. This process is more defensible than treating a similarity percentage as a final verdict.

Table 5: Responsible use of paraphrasing tools

Practice	Potentially appropriate use	Academic risk
Grammar improvement	Improve clarity of student’s own draft	Low, subject to policy
Vocabulary assistance	Learn alternative wording while checking meaning	Moderate if used without understanding
Paraphrase of cited source	Restate an idea while retaining attribution	Appropriate when policy permits and meaning is accurate
Paraphrase of copied text without citation	Change wording to hide source dependence	High integrity risk
Submitting generated text as independent work	Replacing student’s own assessed writing	High integrity risk

9. Future Research

The impact of paraphrasing tools on plagiarism detection is likely to remain an active research problem because both sides of the technology are changing. Future research should not only build stronger detectors but should also evaluate whether those detectors provide meaningful, interpretable, and fair evidence for academic decisions.

9.1 1. Larger Benchmark Datasets

Future studies should create benchmark datasets containing direct copies, lightly paraphrased text, strongly paraphrased text, human-written independent text, and AI-generated paraphrases. Each document should be linked to its true source and labelled using a clear annotation procedure. Such a dataset would allow different detection systems to be compared under the same conditions.

9.2 2. Multilingual and Cross-Language Paraphrasing

Much research has focused on English, while students and researchers often work across languages. Cross-language plagiarism can involve translation followed by paraphrasing, making lexical matching even more difficult. Multilingual embeddings and cross-lingual retrieval should therefore be evaluated systematically. This is especially relevant in diverse higher-education environments.

9.3 3. Transformer and Embedding-Based Detection

Future work can compare TF-IDF and n-gram baselines with sentence embeddings, transformer encoders, cross-encoders, and newer long-context models. The goal should not be to replace traditional methods but to measure whether combining lexical and semantic evidence improves recall without creating excessive false positives.

9.4 4. Explainable Detection Reports

A useful detector should show why a passage was flagged. Instead of reporting only a number, future systems could display the candidate source, the related passage, semantic similarity, shared concepts, citation status, and the type of transformation detected. This would make the output easier for instructors and students to interpret.

9.5 5. Robustness Against Iterative Paraphrasing

A single paraphrase pass is only one possible transformation. A future benchmark should test repeated paraphrasing, combinations of paraphrasing and translation, and manual editing after machine paraphrasing. The purpose should be to understand detector robustness rather than to develop methods for evading detection.

9.6 6. Human–AI Evaluation

Research should compare automated systems with trained human reviewers. The goal is not necessarily to identify one winner but to understand where each approach succeeds and fails. Human reviewers can provide contextual judgement, while automated systems can process large document collections quickly. A combined workflow may therefore be more practical than either approach alone.

9.7 7. Student Learning Outcomes

Future educational studies should measure whether responsible use of paraphrasing tools actually improves students’ paraphrasing ability. It is possible for a tool to improve the final sentence while reducing the student’s need to practise the underlying skill. Longitudinal studies could compare students who use tools as language feedback with students who rely on automatic rewriting.

9.8 8. Privacy and Governance

As academic documents increasingly pass through cloud-based AI services, privacy should become part of evaluation. Future research should examine data retention, training use, document ownership, and institutional controls. Detection effectiveness alone is not sufficient for an academic tool if sensitive research material is handled inappropriately.

9.9 9. Evaluation of AI-Paraphrase Detection

Commercial systems are beginning to introduce dedicated features for identifying AI-assisted paraphrasing. Turnitin, for example, announced an AI-paraphrasing detection feature in 2024 [11]. Future independent research should evaluate such features using transparent datasets and reproducible methods. Vendor claims and independent academic measurements should be kept separate when interpreting results.

9.10 10. Towards a Unified Academic Integrity Framework

The long-term research direction should move from a single “plagiarism score” toward a multi-evidence framework. Such a framework could combine lexical overlap, semantic similarity, source retrieval, citation analysis, writing-pattern evidence, document

history, and human review. The final objective should be better evidence and better academic practice, not simply higher detection percentages.

10. Conclusion

Paraphrasing tools have changed the relationship between writing assistance and plagiarism detection. Their central function is to transform language while attempting to preserve meaning. That function can support legitimate academic writing, but it can also be misused to disguise unattributed source material. The technical consequence is that the visible wording of a source may become less similar even when the underlying ideas remain closely related.

The literature reviewed in this paper shows why plagiarism detection cannot depend entirely on exact word matching. Research on paraphrase plagiarism has increasingly explored semantic similarity, contextual embeddings, word representations, and machine-learning approaches because these methods can examine relationships that remain after wording has changed [1, 2, 9]. At the same time, studies evaluating plagiarism-detection and AI-detection systems show that automated tools have limitations and can generate both missed cases and questionable flags [4, 10].

A major conclusion of this study is therefore that a similarity percentage should be treated as evidence, not as a definition of plagiarism. A low score does not prove that a document is original, and a high score does not automatically prove misconduct. The source, the student's citations, the transformation of the text, and the academic context all need to be considered.

From a student's perspective, the most responsible use of paraphrasing technology is to treat it as a learning and language-support tool rather than as a method for hiding source use. Genuine paraphrasing requires understanding. If the original idea came from another author, that contribution should remain acknowledged even when the wording has been substantially changed.

For plagiarism-detection systems, the future is likely to involve hybrid methods. Lexical matching can efficiently identify direct copying; semantic retrieval can identify related passages after paraphrasing; contextual models can capture more complex relationships; and human review can determine whether the relationship represents acceptable scholarship or an academic-integrity concern. This layered approach is more informative than relying on a single automated score.

Ultimately, the impact of paraphrasing tools is not simply that they make plagiarism "harder to detect." Their deeper impact is that they change what evidence a detection system needs to examine. As wording becomes easier to transform, meaning, context, source relationships, and transparent academic practice become more important. Future research should therefore focus on reliable, explainable, multilingual, and privacy-conscious detection systems while continuing to teach students the underlying skill of writing and citing ideas responsibly.

Student Researcher Reflection

From the perspective of a Data Science student, this topic shows how a real academic problem can be approached as a data and NLP problem. The main understanding developed through this study is that text similarity is not one-dimensional. Two documents can have low word overlap but still be strongly related in meaning. This makes the problem suitable for combining traditional similarity measures with embeddings, machine learning, and human interpretation. At the same time, the ethical side of the problem cannot be solved by an algorithm alone. A technically strong model still needs a clear definition of what counts as evidence and how that evidence should be used.

Final Statement

Paraphrasing technology should not be viewed only as a threat to academic integrity or only as a writing convenience. Its effect depends on how it is used and how detection systems interpret transformed text. The most reliable academic environment is one in which students are taught to understand sources, paraphrase independently, cite appropriately, and use technology transparently, while institutions use detection tools as part of a broader evidence-based review process.

References

- [1] M. A. Álvarez-Carmona, M. Franco-Salvador, E. Villatoro-Tello, M. Montes-y-Gómez, P. Rosso, and L. Villaseñor-Pineda, “Semantically-informed distance and similarity measures for paraphrase plagiarism identification,” *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 5, 2018, doi: 10.3233/JIFS-169483.
- [2] F. Alvi, M. Stevenson, and P. Clough, “Paraphrase type identification for plagiarism detection using contexts and word embeddings,” *International Journal of Educational Technology in Higher Education*, vol. 18, art. 42, 2021, doi: 10.1186/s41239-021-00277-8.
- [3] T. Foltýnek, N. Meuschke, and B. Gipp, “Academic plagiarism detection: A systematic literature review,” *ACM Computing Surveys*, vol. 52, no. 6, art. 112, 2019, doi: 10.1145/3345317.
- [4] T. Foltýnek, D. Dlabolová, A. Anohina-Naumeca, S. Razí, J. Kravjar, L. Kamzola, J. Guerrero-Dib, Ø. Çelik, and D. Weber-Wulff, “Testing of support tools for plagiarism detection,” *International Journal of Educational Technology in Higher Education*, vol. 17, art. 46, 2020, doi: 10.1186/s41239-020-00192-4.
- [5] N. Madnani and B. J. Dorr, “Generating phrasal and sentential paraphrases: A survey of data-driven methods,” *Computational Linguistics*, vol. 36, no. 3, pp. 341–387, 2010, doi: 10.1162/coli_a.00002.
- [6] F. M. Prentice and C. E. Kinden, “Paraphrasing tools, language translation tools and plagiarism: An exploratory study,” *International Journal for Educational Integrity*, vol. 14, art. 11, 2018, doi: 10.1007/s40979-018-0036-7.
- [7] A. M. Rogerson and G. McCarthy, “Using Internet based paraphrasing tools: Original work, patchwriting or facilitated plagiarism?” *International Journal for Educational Integrity*, vol. 13, art. 2, 2017, doi: 10.1007/s40979-016-0013-y.
- [8] J. Roe and M. Perkins, “What are Automated Paraphrasing Tools and how do we address them? A review of a growing threat to academic integrity,” *International Journal for Educational Integrity*, vol. 18, art. 15, 2022, doi: 10.1007/s40979-022-00109-w.
- [9] T. Vrbanec and A. Meštrović, “Comparison study of unsupervised paraphrase detection: Deep learning—The key for semantic similarity detection,” *Expert Systems*, 2023, doi: 10.1111/exsy.13386.
- [10] D. Weber-Wulff et al., “Testing of detection tools for AI-generated text,” *International Journal for Educational Integrity*, 2023, doi: 10.1007/s40979-023-00146-z.
- [11] Turnitin, “Turnitin helps educators and publishers advance critical thinking with new AI paraphrasing detection feature,” Press Release, July 16, 2024. Available: <https://in.turnitin.com/press/turnitin-new-ai-paraphrasing-detection-feature>
- [12] I. C. Tresna, I. Usman, N. Rachmah, and T. Siswono, “Quill & Bytes: A qualitative analysis on the perceived impacts of AI-based paraphrasing tools in academic writing and performance toward higher education students,” *Procedia Computer Science*, vol. 263, pp. 664–671, 2025, doi: 10.1016/j.procs.2025.07.079.
- [13] B. N. Wijaya and R. Hidayat, “Paraphrase plagiarism detection in student assignments using IndoBERT and two-level cosine similarity,” *bit-Tech*, vol. 9, no. 1, pp. 1478–1489, 2026, doi: 10.32877/bt.v9i1.4130.