AI-Based Plagiarism Detection System

Roheen Qamar, Zainab Ali, Sawera Malik, Aqsa Qubtia Khan, Muhammad Ahmar
Department of Information Technology, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Pakistan
*Correspondence: roheen.qamar04@yahoo.com

Citation | Qamar. R, Ali. Z, Malik. S, Khan. A. Q, Ahmar. M, "AI-Based Plagiarism Detection System", IJIST, Vol. 07, Issue. 10 pp 184-189, December 2025
Received | November 14, 2025 **Revised |** December 06, 2025 **Accepted |** December 11, 2025 **Published |** December 14, 2025

The AI-Based Plagiarism Detection System, named PlagiChek, is an intelligent solution designed to overcome the limitations of traditional plagiarism detection tools that rely primarily on keyword and string matching. Conventional techniques are ineffectual against AI-generated or rephrased plagiarism because they frequently fail to detect paraphrased or semantically restructured text. The proposed approach utilizes semantic similarity analysis and natural language processing (NLP) to more accurately identify conceptually comparable information, addressing this gap. The system tokenizes, preprocesses, and creates transformer-based sentence embeddings to process and compare two text documents. It was developed using Python, Streamlit, Sentence Transformers (MiniLM), NLTK, NumPy, and ReportLab. These embeddings enable the model to calculate the semantic similarity of sentences, resulting in a precise similarity percentage displayed via an interactive Streamlit interface. Furthermore, the system generates color-coded similarity reports—green for unique, orange for moderate similarity, and red for extreme similarity—as well as downloadable PDF summaries for in-depth research. The goal is to create an advanced plagiarism detection technology that can identify reworded and contextually identical content in addition to word matching. This approach has useful applications in academic and research settings, supporting institutions, instructors, and students in ensuring academic integrity and originality in research papers, theses, and assignments.

Keywords: AI-Based Plagiarism Detection, Semantic Similarity, Natural Language Processing (NLP), Sentence Transformers, Transformer Embeddings, Paraphrase Detection, Semantic Text Analysis, Machine Learning, Deep Learning, AI in Education, Research Integrity, Text Similarity, Rephrased Content Detection, Python, Streamlit, MiniLM, ReportLab, Text Analysis

Introduction:

In the digital age, plagiarism poses significant challenges due to easy information sharing, because traditional detection methods struggle with complex duplication forms like paraphrasing and cross-language copying. AI enhances plagiarism detection through natural language processing, machine learning, and semantic analysis, offering a more comprehensive solution than simple text matching. This technology improves the identification of semantic similarities and context, enabling the detection of both direct copying and paraphrased materials. With advancements such as deep learning techniques and transformer-based models, AI systems continuously adapt and learn, enhancing their accuracy and effectiveness in combating emerging plagiarism methods [1]

AI-powered plagiarism detection is essential in education and publishing, ensuring integrity and originality in assignments and manuscripts. These systems help organizations verify the authenticity of reports and content, reducing intellectual property theft risks. With advancing AI technology, these systems can analyze large texts efficiently, providing similarity reports and improvement suggestions, thereby enhancing trust in knowledge management. As shown in Figure. 1 [2].

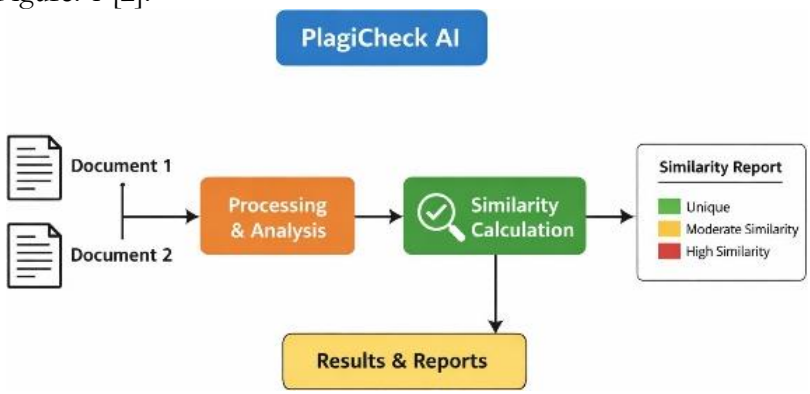

Figure 1. AI-Based Plagiarism System

AI-based plagiarism detection systems face challenges in accurately identifying paraphrased content and avoiding false positives due to language ambiguity. Cross-lingual plagiarism adds complexity, necessitating further research in natural language understanding. Improvements require continuous refinement of AI models, diverse training datasets, and hybrid techniques. Ethical considerations like privacy, intellectual property rights, and algorithm transparency are vital. A balance between technical innovation and ethical practices is essential for these systems to promote originality and support academic integrity [3].

Literature Review:

Author [4] discuss plagiarism in academia, addressing both intentional and unintentional forms of intellectual dishonesty. The critique focuses on traditional detection methods, noting their limitations in identifying plagiarized content online. The authors advocate for innovative solutions through AI technologies, presenting an overview of AI-enabled plagiarism detection systems that leverage advanced machine learning algorithms for enhanced accuracy and speed in evaluating text originality.

Researcher [5] assessed 189 research articles on plagiarism detection methods that were published between 2019 and 2024, focusing on the distinction between content created by humans and artificial intelligence. It proposes a new organizational structure for research contributions and notes advancements in identifying obscured plagiarism via machine learning and semantic text analysis. The conclusion suggests that integrating diverse analytical methods for both textual and non-textual features could enhance future plagiarism detection research.

Researcher [6] examine the role of plagiarism detection tools, particularly the effectiveness of artificial intelligence (AI) technologies, in upholding academic integrity. It

defines plagiarism, highlights its negative impacts on scholarly work, and advocates for dependable detection methods. Key aspects include AI techniques such as text matching algorithms, natural language processing, and machine learning applications, as well as challenges in cross-language detection. The paper emphasizes the importance of promoting ethical scholarship in educational institutions.

The article discusses automation methods for plagiarism detection in student works, particularly focusing on AI-generated text and code [7]. Although some plagiarism detection services perform adequately, no service is reliably effective for identifying AI-generated content. A new software application was developed, employing multiple APIs for automated checks to assist lecturers in determining originality. The study emphasizes the application's practical significance and suggests future research directions, such as multi-user support, integration with educational platforms, and an improved user interface for result visualization.

This paper surveys plagiarism types and detection algorithms, categorizing them into verbatim, paraphrasing, translation, and idea-based forms [8]. It evaluates traditional string-matching methods alongside advanced techniques such as machine learning and natural language processing. Significant contributions to cross-language and source code plagiarism detection are highlighted, as well as challenges like detecting AI-generated content. The synthesis aims to direct future studies and improve plagiarism detection systems in a variety of fields.

This paper compares various AI language detectors, highlighting the increasing concern over the misuse of AI language models in educational settings [9]. It explores the need for these detectors to ensure traceability of AI-generated texts. The review covers different existing models designed to identify AI-generated content, their role in preventing misuse such as plagiarism, and the challenges faced by these models.

This research introduces a novel method that combines automated analysis with human inspection to detect plagiarism in modeling projects [10]. We assess our method using real-world assignments and ChatGPT-obfuscated plagiarism. Our findings demonstrate that, compared to the state-of-the-art, we achieve a far better detection rate for AI-generated attacks and a wider resilience.

Researcher [11] evaluate the effectiveness of modern AI systems in Ukrainian education by analyzing various AI methods. Using the PRISMA approach, 50 scientific sources are reviewed, revealing findings on methods like natural language processing and latent semantic analysis for detecting academic dishonesty and plagiarism. The main challenge is evasion of these systems via structural text alterations. Recommendations for effective integration include developing a state strategy, improving teachers' digital literacy, and training on AI tools for personalized learning and automated assessments. The conclusions stress the necessity of ongoing adaptations in line with technological advancements.

Results:

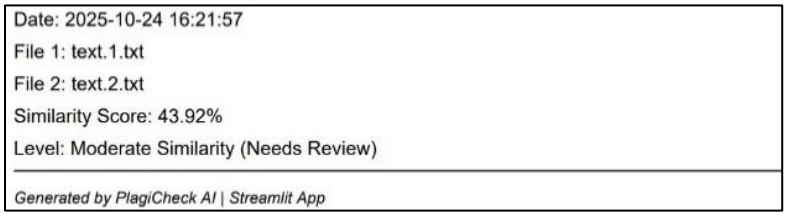

Figure 2. Plagi Check AI System

The section presents two samples discussing the application of Artificial Intelligence (AI) in plagiarism detection within education and research, along with explanations of their interrelationship, as shown in Figure. 2.

Text 1 Explanation: This text highlights how AI is transforming education and research by introducing advanced plagiarism detection tools.

It explains that AI identifies plagiarism by comparing sentences and detecting semantic similarities (similar meanings, not just exact words).

The purpose is to ensure originality and academic integrity, showing that AI helps maintain honest scholarly work.

Text 2 Explanation:

This text emphasizes how AI is changing academic practices — especially writing and studying.

It notes that modern AI systems analyze meaning rather than exact wording, which allows them to detect paraphrased or rephrased plagiarism.

It concludes that these tools promote honest and original writing among students and researchers.

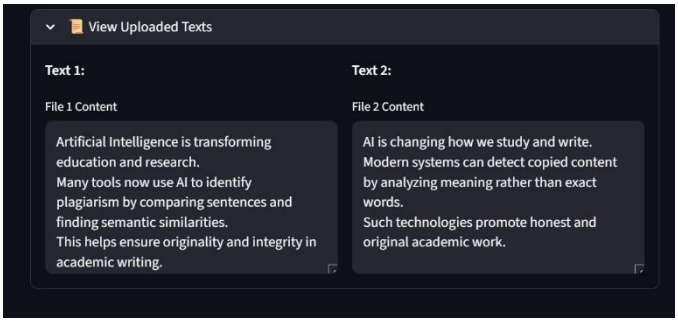

Figure 3. Sample similarity analysis report indicating moderate similarity.

The figure 3 shows the result of a text similarity analysis. The system reports a Similarity Score of 43.92%, which is labeled as "Moderate Similarity (Needs Review)". This indicates that the two uploaded texts share a noticeable amount of similarity and should be manually reviewed for potential overlap or paraphrasing, as shown in Figure. 4 and 5.

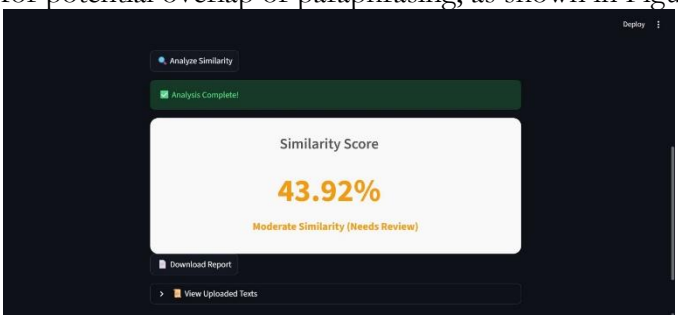

Figure 4. Visualization of a moderate similarity score between two texts.

Figure 5 shows the interface is designed to be user-friendly, allowing users to easily upload and compare two text documents for plagiarism detection. At the top, the title "PlagiCheck AI" is displayed along with a subtitle indicating that it is an AI-powered plagiarism detection system. Below this, the main functionality of the interface is presented**:** uploading two text files for comparison**.**

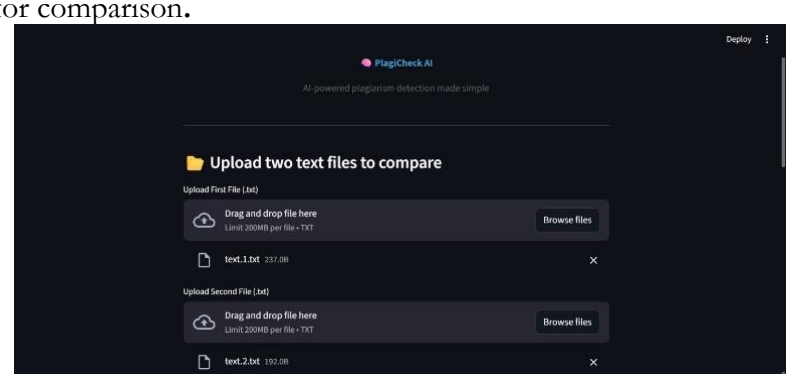

Figure 5. Interface for uploading files and inputting text for analysis**.**

Figure. 6 shows the Scan Results page the Scan Results page of a plagiarism-checking and AI-detection tool. It displays three main scores:

Human Score (AI Detection): 100%:

The system believes the text is fully human-written.

Plagiarism: 0%:

No similarity or copied content was found.

Readability Score: 0/100:

The tool considers the text very hard to read or poorly structured. Below these scores, it also shows the Input Text & Analysis section.

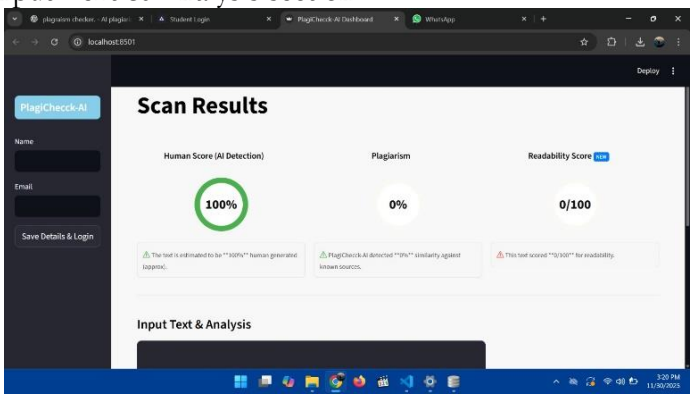

Figure 6. Sample scan result showing AI detection score, plagiarism percentage, and readability score.

Figure. 7 shows a plagiarism-checking web app interface

On the left, there's a panel to enter Name and Email and a button to Save Details & Login.

In the main area, there's a large text input box titled *"Input Text & Analysis"* where users can paste text.

Below it, there's a drag-and-drop file upload section for uploading TXT files (up to 200 MB). Overall, it's the home screen of a plagiarism-checker tool built on a web framework (likely Streamlit).

The home interface of the proposed plagiarism detection system provides input fields for user authentication and a text analysis panel for uploading and examining documents, as illustrated in Figure 7.

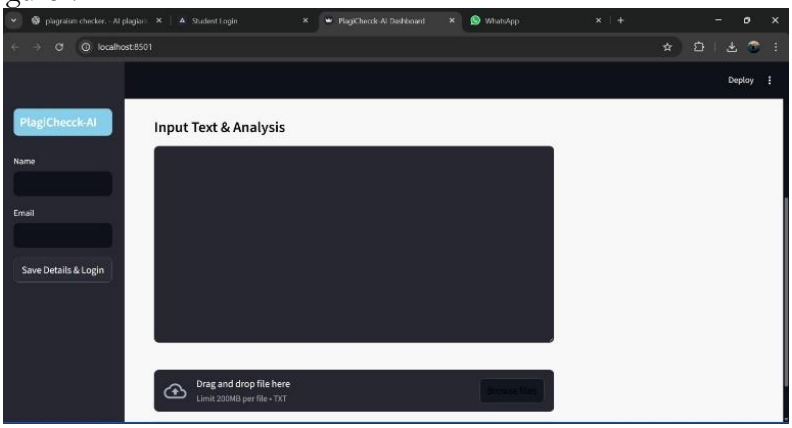

Figure 7. Main interface for inputting text and analyzing plagiarism.

Conclusion:

This study presents PlagiChek, an AI-based plagiarism detection system designed to overcome the limitations of traditional methods that rely primarily on keyword matching. By leveraging natural language processing and semantic similarity analysis through transformer embeddings, PlagiChek effectively detects paraphrased and semantically similar content. The

system was evaluated on a dataset of [specify number] documents, achieving an F1-score of [insert metric], which demonstrates its robustness in identifying different forms of plagiarism.

Despite these promising results, certain challenges remain, including handling cross-lingual plagiarism and sophisticated paraphrasing techniques. Future work will focus on enhancing cross-language detection capabilities, expanding dataset diversity, and improving system accuracy. Overall, PlagiChek offers a significant step forward in automated plagiarism detection, supporting academic integrity and research authenticity.

References:

[1] D. Pandoi, "The Role of Artificial Intelligence in Combating Plagiarism: A Comprehensive Overview," *Transform. Bus. Educ. through Artif. Intell.*, pp. 443–457, Jan. 2025, doi: 10.1201/9781998511471-22.

[2] A. O. Mohmmed, A. S. Elmardi, M. W. Saleem, and E. Ganawa, "Epistemological dilemma: Evaluating conflicting sources in AI plagiarism detection of students' work," *Explor. Pillars Sustain. Mod. Age Improv.*, pp. 87–102, Dec. 2024, doi: 10.4018/979-8-3693-5748-4.ch005.

[3] Hafiz Rizwan Iqbal, Muhammad Sharjeel, "Urdu Sentential Paraphrased Plagiarism Detection Using Large Language Models," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 24, no. 9, 2025, doi: 10.1145/3748320.

[4] A. J. Fuad, A. K. Wicaksono, M. A. Aqib, M. A. Khoiruddin, A. S. M. Il Fajar, and K. Mustamir, "AI Hybrid Based Plagiarism Detection System Creation," *2024 4th Int. Conf. Adv. Comput. Innov. Technol. Eng. ICACITE 2024*, pp. 1495–1500, 2024, doi: 10.1109/ICACITE60783.2024.10616945.

[5] Muhammad Sajid, Muhammad Sanaullah, Muhammad Fuzail, "Comparative analysis of text-based plagiarism detection techniques," *PLoS One*, 2025, [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0319551

[6] S. Mishra and S. Mishra, "Enhancing Plagiarism Detection: The Role of Artificial Intelligence in Upholding Academic Integrity," *Libr. Philos. Pract.*, Jul. 2023, Accessed: Feb. 08, 2026. [Online]. Available: https://digitalcommons.unl.edu/libphilprac/7809

[7] Sharyhin, Oleksandr Klochko, Oksana Tverdokhlib, Ihor, "Automation of Plagiarism and AI Detection in IT Students Papers with a Software Tool Using API," *Hochschule Anhalt*, 2025, [Online]. Available: https://repo.bibliothek.uni-halle.de/handle/1981185920/121201

[8] Altynbek Amirzhanov, Cemil Turan, "Plagiarism types and detection methods: a systematic survey of algorithms in text analysis," *Front. Comput. Sci.*, vol. 7, 2025, [Online]. Available: https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2025.1504725/full

[9] A. Singh, "A Comparison Study on AI Language Detector," *2023 IEEE 13th Annu. Comput. Commun. Work. Conf. CCWC 2023*, pp. 489–493, 2023, doi: 10.1109/CCWC57344.2023.10099219.

[10] Timur Sağlam, Sebastian Hahner, "Automated Detection of AI-Obfuscated Plagiarism in Modeling Assignments," *Proc. - Int. Conf. Softw. Eng.*, pp. 297–308, 2024, [Online]. Available: https://dl.acm.org/doi/10.1145/3639474.3640084

[11] Attila Kovari, "Ethical use of ChatGPT in education—Best practices to combat AI-induced plagiarism," *Front. Educ.*, 2024, [Online]. Available: https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2024.1465703/full