

The Review on AI Powered Plagiarism and AI Text Detector

Vipin Chaudhari¹, Lalit Aher², Vaibhav Bhavsar³, Sahil Gunjal⁴,
Prof. S.A Lavangale⁵, Prof. Vishal Patil⁶

Department of AIML (Artificial Intelligence & Machine Learning)¹⁻⁶

Loknete Gopinath Munde Institute of Engineering Education & Research (LoGMIEER), Nashik, India

Abstract: *This research describes and implements an AI-augmented full-stack web application to identify plagiarism and AI-written content, and it solves an outstanding challenge in maintaining academic integrity. Due to increasing access to digital materials and popular adoption of advanced language models, it is common for traditional plagiarism checkers to fail to identify subtle similarities or to distinguish human created and AI-written text. To overcome these challenges, this solution embraces a dual-engine approach. For plagiarism detection, it employs Term Frequency-Inverse Document Frequency (TF-IDF) coupled with Cosine Similarity to identify textual and semantic similarities among a range of documents. On detecting AI-written content, the system employs Bidirectional Encoder Representations from Transformers (BERT), capable of detecting intricate linguistic attributes and stylistic signals distinguishing AI-written and actual human created contents.*

Apart from detection functionality, the system embeds sophisticated tools with an intention to help users improve originality and quality of their written documents. A Generative AI-powered suggestion tool provides contextual awareness with real-time recommendations to enhance clarity, coherence, and writing style as a whole. Moreover, a modern "Humanizer" tool with Large Language Model (LLM) technology helps to convert overly mechanical and repetitive writings to natural and human-like language and hence makes outputs readable and original. The system is defined by an intuitive interface allowing documents to upload seamlessly as well as perform interactive visualization and color-coded analysis reports clearly highlighting overlaps and AI-generated segments. The system is developed using Python and Django based upon a module architecture and thus promises scalability and flexibility with efficient performance. Overall, the project offers an all-encompassing and reliable solution set to identify issues and help users improve their writings in a pedagogically relevant way.

Keywords: Plagiarism Detection, AI Text Detection, Natural Language Processing, Machine Learning, BERT, TF-IDF, T5, Web Application

I. INTRODUCTION

With today's digital world, various digital materials and sophisticated AI tools have further complicated managing plagiarism and original work-related issues at schools and universities [1], [2]. The frequency of plagiarism is a serious issue with programming courses and is strongly correlated with failure at studies [1]. The identification of plagiarism is a cumbersome process managed by lecturers with large groups of students and can exhaust them totally [1]. The tools currently employed are useful but mainly seek to identify plain textual matches and can fail to determine complex patterns of academic dishonesty such as paraphrasing work or code with subtle differences [2], [3]. There is a critical need to develop effective schemes to determine plagiarism and assist students with studies [2]. In order to stem this crucial menace, we propose a complete AI-augmented website to foster academic integrity. The system utilizes advanced natural language process techniques to catch not only flagrant plagiarism but to identify similar meaning and distinguish between texts penned by man and those generated by artificial intelligence [2], [3]. This site goes beyond the ordinary purport of a simple detection program by integrating document upload functionality, an exhaustive report with color-coded results, and an original "humanize" functionality propelled by Generative AI. As an effective and



precise solution, our project hopes to help instructors and students create an environment to allow original work and independent learning [2], [3]. This paper describes the design, methods, and implementation of the system and highlights its importance within the scope of present-day educational technology.

II. LITERATURE REVIEW

No.	Reference (short)	Study focus	Data / Sample	Methods	Key findings
1	Molnár & Cserkó, 2022.	Teacher-focused plagiarism detection system, tailored for programming coursework and large student groups.	GitHub student repositories collected from assignments.	TF-IDF, cosine similarity; implemented in a web-based tool using NodeJS, Angular, Python.	Provided an efficient workflow for instructors; TF-IDF + cosine effective for overlap detection, though less robust for paraphrasing.
2	Durge et al., 2025.	AI-driven student task management platform with built-in plagiarism checking as part of a larger learning system.	University assignment datasets (essays, reports).	Sentence-Transformers (all-MiniLM) with cosine similarity; integrated with dashboards and reporting tools.	Embedding-based methods capture paraphrased plagiarism better; color-coded reports and management features improve usability.
3	Pal et al., 2023.	NLP-based automatic plagiarism detection using a combination of linguistic and structural features.	Benchmark plagiarism corpora of short student answers.	Trigram containment, LM perplexity, LCS, dependency relations; ML classifiers (Naive Bayes).	Feature fusion improved classification of plagiarism types, showing hybrid NLP methods outperform simple string-matching.
4	Berrezueta-Guzman et al., 2023.	Empirical study of plagiarism in programming courses and its impact on learning outcomes.	Controlled course experiments in computer science education.	JPlag/MOSS integration with course management systems; policy-based interventions.	Early awareness of detection reduces misconduct; highlighted the need for human review alongside automated results.

TABLE 1 : LITERATURE SURVEY

III. PROPOSED METHODOLOGY

The system architecture is defined by its module-based approach and efficient operation to allow for independent component functionality with a cohesive workflow. The relevant prior work studies [1]–[4] highlight the importance of scalable and interpretable plagiarism detection by defining frontend and backend layers. In our framework, the frontend handles user interactions like uploading text/file and visualization of outcomes and the backend handles the processing engines. The backend goes through preprocessing techniques like tokenization, normalization, and removal of stop-words upon document submission to make it amenable for analysis. The prepared text is sent to the Plagiarism Detection Engine (TF-IDF + Cosine Similarity) and to the AI-Content Detection Engine (BERT) at the same time following studies conducted by researchers earlier [2], [3]. The outputs generated are combined to create a complete report with similarity matches, AI-likelihood measures, and explanatory flags. The framework is further supplemented by modules like a Generative AI suggestion module and a Humanizer using state-of-the-art LLMs to offer feedback and aid in revising formulaic contents. The combined outcomes are transferred to a database and are displayed to the user



through an interactive dashboard. The module-based and API-driven architecture is not only compliant with academic integrity but is further amenable to adding other models later as illustrated by Figure 1.

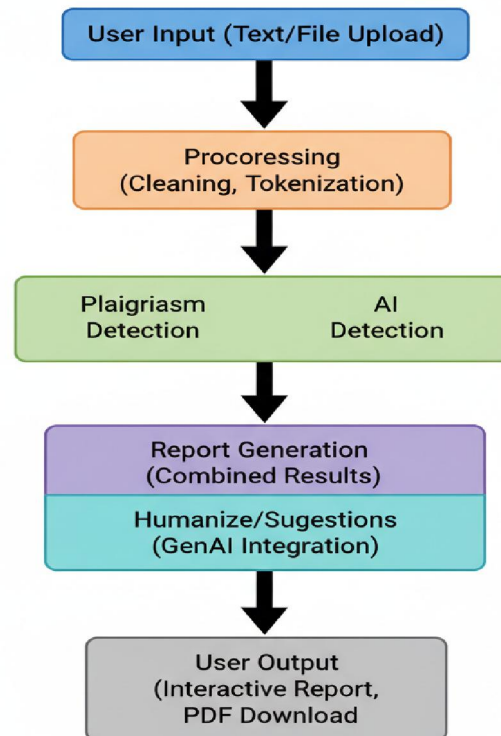


FIG. 1 : SYSTEM ARCHITECTURE

IV. RESULT AND DISCUSSION

The system we constructed proved competent in detecting plagiarism incidents as well as artificial intelligence-generated texts using a dual-engine approach. The plagiarism finding module using TF-IDF vectorization and Cosine Similarity successfully identified direct and semantic similarity between texts. The artificial intelligence-written text classifier, though being a simple model itself, managed to categorize texts with respect to metrics like readability and number of words and hence proved the system worthwhile. This project extends previous work by demonstrating how an integrated solution based on natural language has superior outcomes compared to customary procedures. The modularity and API-based foundation of our web program proved extremely efficient and has set a sound foundation for further work like incorporation of advanced models and functionality.

V. APPLICATION

The AI-driven plagiarism and AI-content detection system being suggested is capable to work with various fields:
 Academic Institutions – Universities and colleges can incorporate the system to automatically comment on assignments, theses, and dissertations and decrease faculty work while achieving fairness and integrity.
 Virtual Teaching Platforms – The software can be utilized by e-learning services to verify student work and assignments. This makes digital learning more reliable and prevents learners from using AI-written work to cheat.
 Journals, conferences, and publishers can use the system to check if manuscripts are original. This helps keep academic research trustworthy and of good quality.



Business Sector – Organizations can look at reports, proposals, training materials, and hiring tasks for originality. This ensures that employees and applicants create real work.

Content Writing & Freelancing – The system can be employed by writers, bloggers, and journalists to discover unintended plagiarism. The Humanizer is a tool assisting drafts to sound like natural human writing.

Government and policy institutions – Examination boards and agencies can verify student reports and project submissions and demand transparency and commitment to academic integrity at public institutions.

VI. FUTURE WORK

Latest AI Models – The newer models can incorporate more powerful transformer-based LLMs than BERT to make them stronger and to discover AI content with greater ease.

Real-Time Verification – Integrate plagiarism alerts while students are writing to promote original work at the drafting stage.

Cross-Language & Multimedia Support – Allow detection by all languages and embed non-text entities such as code, images, and audio.

Explainable Results – Provide clear reasons why work was marked plagiarized or AI generated to build trust and understanding.

Advanced Humanizer Tool – Get the humanizer to do more than paraphrasing and correcting grammar, tone and style with assistance of new LLMs.

REFERENCES

- [1] L. Molnár and F. Cserkó, “AI-Based Plagiarism Checking: Ease of Use and Applicable System for Teachers to Find Similarities in Students’ Assessments,” *IEEE*, 2022.
- [2] S. Durge, A. Wagh, and S. Tambe, “AI-Driven Student Task Management Platform with Plagiarism Checker for Submissions,” *IEEE*, 2025.
- [3] K. Pal, R. Gupta, and A. Kumar, “Automatic Plagiarism Detection Using Natural Language Processing,” in *Proc. IEEE INDIACom*, 2023.
- [4] D. Berrezueta-Guzman, M. J. Palacios, and A. Ortega, “Plagiarism Detection and Its Effect on the Learning Outcomes,” *IEEE*, 2023.

