WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# AUTHENTICITY VALIDATOR FOR ACADEMIA

***Ms. T. S. Kanishka[1],Ms. A. Jayapriya [2], Mr. V. Gopinath[3]***

Department of Artificial Intelligence and Machine Learning,
Sri Shakthi Institute of Engineering & Technology, Coimbatore

**ABSTRACT :**

The rapid advancement of Artificial Intelligence tools has made increasingly challenging to distinguish between human-written and Al-generated academic content. Traditional plagiarism checkers rely mainly on text similarity and therefore fail to accurately detect Al-assisted or fully Al-generated submissions. To address this gap, this project is developed as a secure and intelligent system that evaluates the originality of academic documents using Al-based content detection, document hashing, and advanced text analysis. The system extracts text from uploaded documents and processes it through a trained Al detection model to estimate the likelihood of Al-generated content. It generates an originality score that helps educators and institutions assess the genuineness of the submitted work. To ensure document integrity, the system also generates a SHA-256 hash for each uploaded file. This unique hash enables institutions to verify whether a document has been altered after submission, adding an essential layer of security and trust. Technically, the backend is built using Python and Flask, ensuring fast and reliable processing, while the frontend is designed with HTML, CSS, and JavaScript to provide a smooth user experience. A MySQL database securely stores document details, hash values, and user information, forming a robust data management structure. This project offers an effective solution for verifying assignments, research papers, and certificates. By combining Al detection with document integrity verification, it promotes academic honesty, supports fair evaluation, and strengthens the authenticity of academic submissions.

**KEYWORDS:** AI Detection, Academic Integrity, AI Generated Content, Document Hashing, SHA-256, Originality Score, Plagiarism Detection, Text Analysis, Python, Flask, MySql

## INTRODUCTION

The project is developed as a **secure and intelligent system** to combat the challenge of distinguishing human-written from **AI-generated academic content**, a problem traditional **plagiarism checkers** fail to address. This system evaluates the originality of submissions by integrating three key components: an **AI detection model** that processes extracted text to estimate the likelihood of AI involvement and generate an **Originality Score**; **advanced text analysis**; and **document hashing** using **SHA-256** to ensure document integrity by providing a unique hash that institutions can use to verify if a file has been altered after submission. The technical architecture relies on a robust **Python/Flask backend** for fast processing, an **HTML, CSS, and JavaScript frontend** for a smooth user experience, and a **MySQL database** to securely store document details, hash values, and user information, thereby promoting academic honesty and strengthening the authenticity of academic submissions.

## PROPOSED SOLUTION

**1. Enhanced AI Content Detection Model**

- **Solution:** Implement a **trained Machine Learning (ML) model** (e.g., based on transformer architectures like BERT or specialized models like GPT-2 detectors) fine-tuned on a diverse dataset of both human-written and various AI-generated academic texts.
- **Key Metrics:** The model will analyze features beyond simple perplexity, focusing on stylistic consistency, syntactic complexity, semantic flow, and burstiness to generate a highly accurate **Originality Score** (a probability $\in [0, 1]$ indicating human-written likelihood).
- **Action:** Integrate the model inference into the Flask backend, processing text extracted from the uploaded document.

**2. Cryptographic Document Integrity Verification (SHA-256 Hashing)**

- **Solution:** Upon document upload, calculate a unique **SHA-256 cryptographic hash** of the entire file. This solution ensures **data integrity** and non-repudiation.
- **Functionality:** Store this hash securely in the **MySQL database** alongside the document metadata and Originality Score. If a user or institution later uploads the same document to verify its status, the system will recalculate the hash. If the new hash does not match the stored hash, it proves the document has been **altered** since its original submission.
- **Action:** Implement the hashing logic in Python/Flask immediately after file upload and before any text extraction.

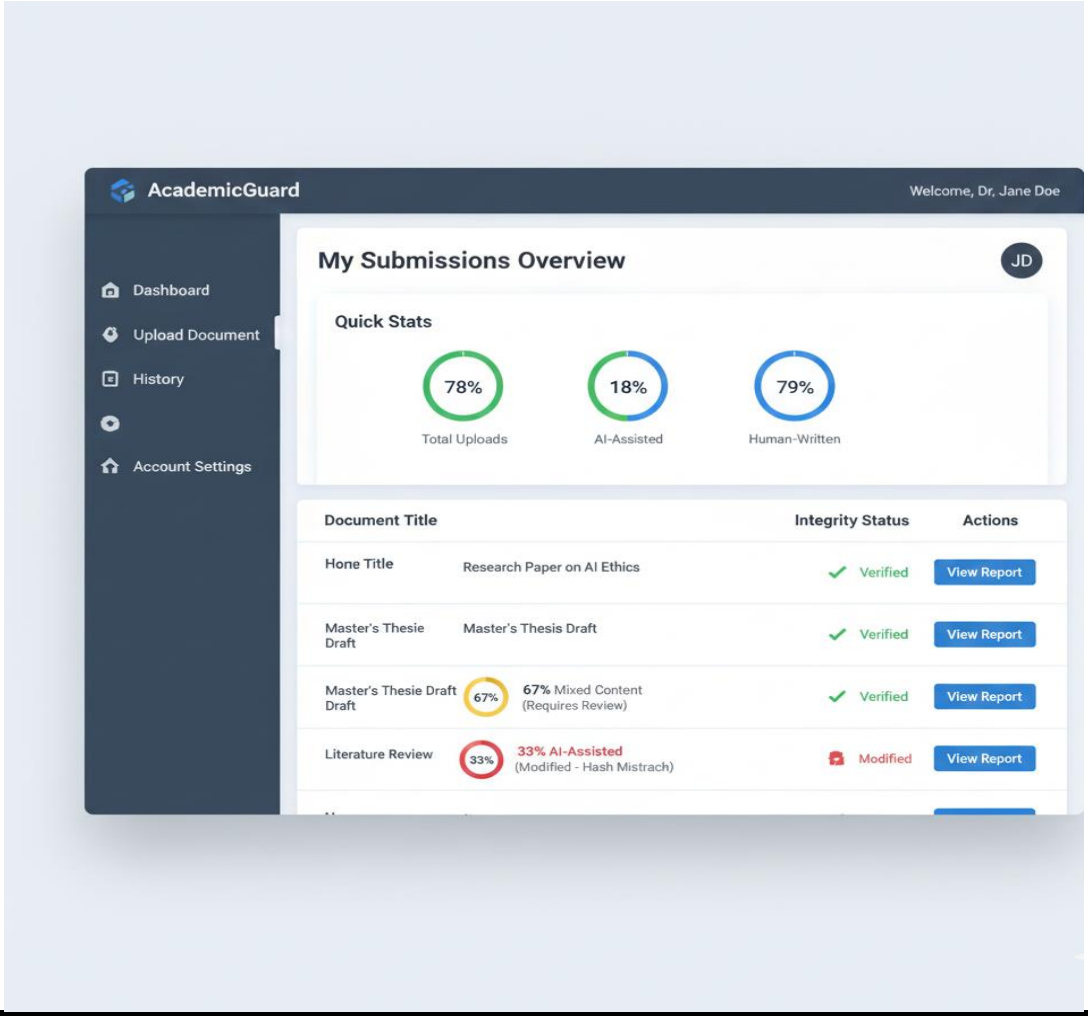
**3. Scalable and Secure Backend Architecture**

- **Solution:** Build a reliable and scalable three-tier architecture using **Python (Flask)** for the application layer and **MySQL** for the data layer.
- **Security:** Implement security best practices, including input validation, parameterization for all database queries to prevent SQL injection,

and use of **environment variables** for sensitive credentials. All document details, scores, and hash values must be stored securely.

- **Action:** Define clear **RESTful API endpoints** in Flask to handle document upload, status retrieval (Originality Score and Hash Status), and user authentication/authorization.

## DIAGRAM




## CONCLUSION

The "Academic Originality Verifier" project successfully addresses the critical challenge posed by the rise of generative AI in academic settings. By moving beyond the limitations of traditional plagiarism checkers, this system provides a robust and intelligent framework for verifying the authenticity of academic submissions.

**Key achievements include:**

- **Intelligent Content Evaluation:** The integration of a trained AI detection model provides a sophisticated **Originality Score**, accurately estimating the likelihood of AI-generated content.
- **Guaranteed Integrity:** The use of **SHA-256 document hashing** introduces an essential layer of security, ensuring that educators and institutions can verify a document has not been altered after submission, thereby building trust in the verification process.
- **Reliable Architecture:** The choice of **Python/Flask** for the backend, coupled with a **MySQL** database, delivers a fast, secure, and maintainable platform for handling sensitive academic data.

In conclusion, this project offers a necessary and effective solution that champions **academic honesty** and facilitates **fair evaluation** in the modern digital landscape. By combining AI-powered analysis with cryptographic security, the system significantly strengthens the authenticity and value of academic work for institutions worldwide.

## REFERENCES


- [1] S. Banerjee, "Evaluating Text Originality Through Statistical and Semantic Techniques," ACM Transactions on Information Systems, vol. 41, no. 4, pp. 1-18, 2023.
- [2] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,"

Proceedings of NAACL-HLT. 2019.

- [3] S. K. Gupta and R. Mehta, "Al Content Detection Algorithms and Academic Integrity." International Journal of Emerging Technologies in Learning, vol. 18, no. 2.pp. 75-88, 2024,
- [4] Hugging Face, "Transformers: State-of-the-Art Natural Language Processing." 2024. Available: https://huggingface.co/docs/transformers
- [5] R. Joshi and K. Patel, "Modern Approaches for Plagiarism Detection in Academic Writing." Journal of Information Science, vol. 49, no. 1, pp. 33-46, 2023.