Indian Journal of Librarianship and Administration
ISSN 2349-0411 Volume 14, Number 1 (2026), pp. 1-8

http:// www.gbspublisher.com

Promotion of Research Ethics and Evaluation of Academic Integrity in Higher Education

Dr.Ramakant Amar Navghare
Librarian (AL-13A), Changu Kana Thakur Arts, Commerce and Science College, New Panvel (Empowered Autonomous), Affiliated to University of Mumbai, Mumbai
email – ramakant.navghare@gmail.com

Dr.Mahesh M. Dalavi (Corresponding Author)
Librarian (AL-13A), Satish Pradhan Dnyansadhana College, Thane
Affiliated to University of Mumbai, Mumbai
email – dmahesh111@hotmail.com

Abstract

In the contemporary academic landscape, the dual challenges of plagiarism and the proliferation of Artificial Intelligence (AI) generated content present significant hurdles to maintaining institutional integrity. This study provides a comprehensive statistical analysis of 357 academic submissions, ranging from research papers and dissertations to project work and book chapters. Utilizing a dataset containing similarity indices and AI detection scores, we examine the prevalence of content duplication and machine-generated text. Our findings reveal a mean similarity index of 13.19% and a mean AI writing score of 20.95%. The analysis further explores the correlation between these two metrics and identifies trends across different document types and languages. The results suggest that while traditional plagiarism remains controlled, the integration of AI tools in academic writing is becoming increasingly prevalent, necessitating updated pedagogical strategies and detection frameworks.

Keywords: Academic Integrity, Plagiarism Detection, Artificial Intelligence, Research Ethics, Similarity Index, Higher Education, Machine Learning, Pedagogy

Introduction

The advent of digital information and more recently, Large Language Models (LLMs) has revolutionized the way academic content is produced. While these technologies offer unprecedented opportunities for research assistance and drafting, they also pose a fundamental threat to the authenticity of scholarly work. Academic integrity is the cornerstone of higher education, ensuring that degrees and research outputs truly reflect the knowledge and effort of the authors.

Historically, the primary concern for academic institutions was plagiarism—the act of using someone else's work or ideas without proper attribution. Tools like Turnitin, iThenticate and DrillBit became standard in identifying textual overlaps. However, the emergence of AI tools such as ChatGPT, Claude, and Gemini has introduced a "second wave" of integrity challenges. AI-generated text is often "original" in the sense that it does not directly copy existing documents, thereby bypassing traditional similarity checks, yet it lacks the human author's intellectual contribution.

This report analyzes a specific dataset of 357 submissions to understand the current state of academic writing. By examining both "Similarity" (traditional plagiarism) and "AI Scores" (machine-generated likelihood), we aim to provide a data-driven perspective on how these two phenomena coexist in the current academic ecosystem.

Methodology

The methodology for this study involved a multi-stage process of data extraction, cleaning, and quantitative analysis.

- Data Source The dataset consisted of 357 records of academic submissions. Each record included metrics such as:
- ✓ Similarity Index (%): The percentage of the document identified as matching existing sources in a global database.
- ✓ AI Score (%): The probability or percentage of the document identified as being generated by AI models.
- ✓ Metadata: Document type (Research Paper, Dissertation, Project Work, etc.), File Size, Page Count, Submission Date, and Language.

- Data Processing The raw data was processed using Python's library. Key cleaning steps included:
- ✓ Converting numeric strings (e.g., "12 %") into floating-point numbers.
- ✓ Handling missing values in the AI Score column.
- ✓ Parsing "File Size" to remove units and convert to numeric values for correlation analysis.
- ✓ Standardizing date formats to analyze submission timelines.

Analytical Framework We employed descriptive statistics to summarize the central tendencies and variability of the integrity metrics. Visualizations were generated using Matplotlib and Seaborn to identify patterns, outliers, and distributions. Correlation analysis was performed to determine if a relationship exists between high similarity and high AI usage.

Objectives

The primary objectives of this research article are:

- ✓ To quantify the average levels of plagiarism and AI involvement in a diverse set of academic submissions.
- ✓ To identify the distribution patterns of Similarity and AI scores across the dataset.

- ✓ To compare integrity metrics across different document categories (e.g., Dissertations vs. Research Papers).
- ✓ To evaluate the relationship between traditional plagiarism and modern AI-generated content.
- ✓ To provide evidence-based recommendations for academic institutions to improve their integrity monitoring.

Analysis and Results

Overview of the Dataset

The dataset comprises 357 unique submissions. The most frequent document type is the Research Paper, reflecting the academic focus of the submissions. Other significant categories include Dissertations, Project Work, and Book Chapters. The majority of documents are written in English (94.3%), with a small subset in Marathi (5.1%) and Hindi (0.6%).

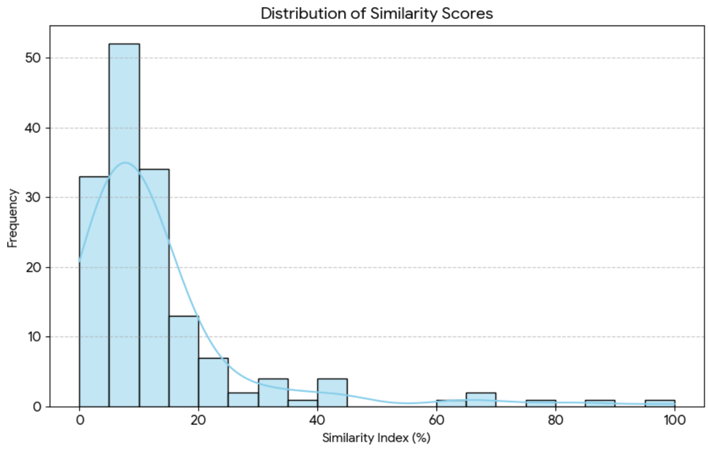
Analysis of Similarity Scores

The Similarity Index measures the overlap between the submitted work and existing literature. Mean Similarity: 13.19%

Median Similarity: 9.00%

Maximum Observed: 100.0%

The distribution of similarity scores (shown below) indicates that most submissions fall within the acceptable range of 0-20%. However, there are notable outliers where the similarity index exceeds 50%, indicating significant potential for plagiarism. The gap between the mean (13.19%) and median (9%) suggests a right-skewed distribution, driven by a few very high-similarity documents. (See Image: Distribution of Similarity Scores)



Distribution of Similarity Scores: Visualizes how plagiarism levels are spread across the dataset.

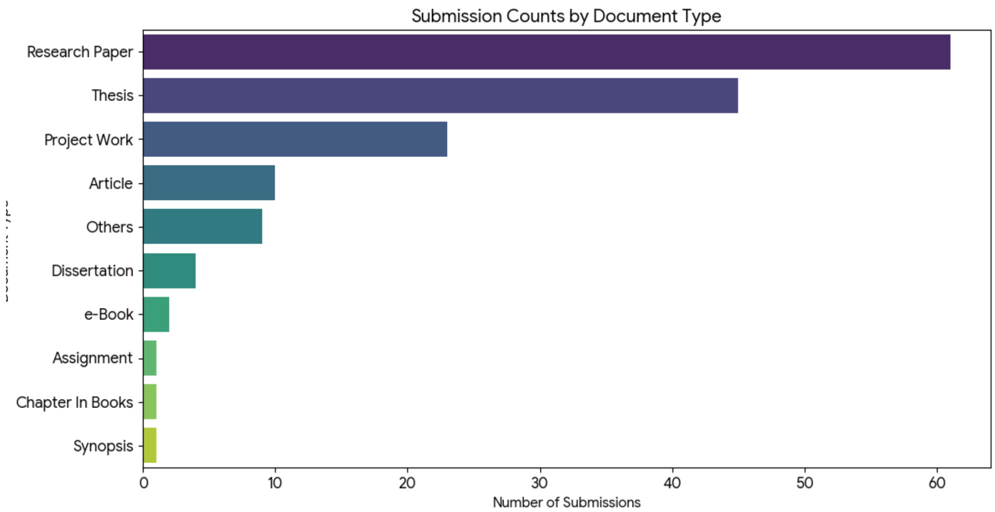
Analysis of AI Writing Scores

The AI Score represents the likelihood of machine intervention in the writing process. Mean AI Score: 20.95%

Median AI Score: 17.00%
Maximum Observed: 91.0%
The AI scores show a higher central tendency than similarity scores.

Distribution of AI Writing Scores: Shows the frequency of AI usage among the submissions.

A significant portion of the data points are clustered between 10% and 30%, suggesting that many authors are utilizing AI for drafting or refining their work. Extreme cases with AI scores above 70% likely represent documents almost entirely generated by LLMs. (See Image: Distribution of AI Writing Scores)



Submission Counts by Document Type: Identifies which types of academic works were most common.

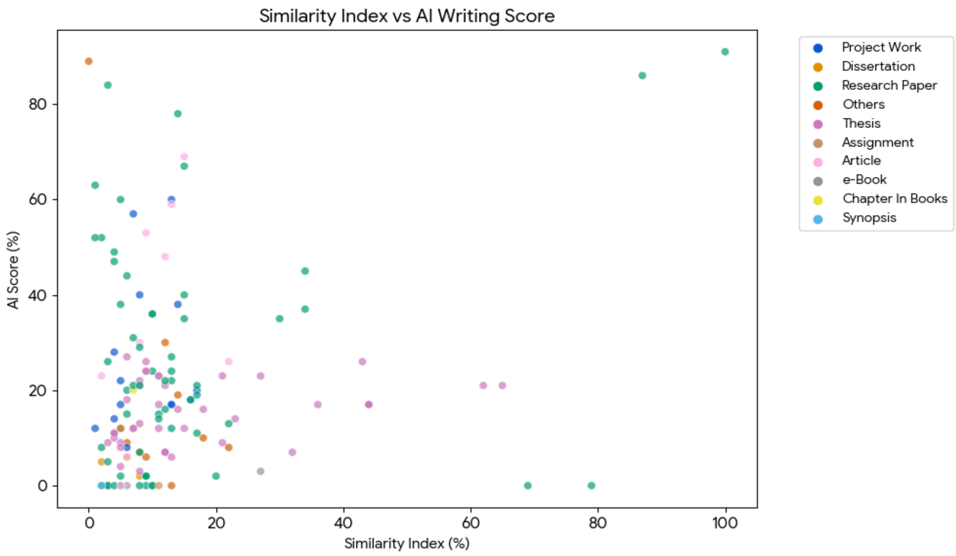
Comparative Analysis by Document Type

Different academic genres exhibit varying levels of integrity metrics.

Research papers often show higher similarity scores due to the necessity of citing existing literature and methodology sections. Conversely, project work and dissertations occasionally show higher AI scores, possibly due to students using AI to structure large volumes of text. (See Image: Submission Counts by Document Type)

The Relationship between Similarity and AI Scores

A critical question is whether authors who plagiarize are also more likely to use AI. Our scatter plot analysis shows that there is no strong linear correlation between the two. Many documents with very low similarity (near 0%) have high AI scores. This confirms that AI tools are being used to create "pseudo-original" content that evades traditional plagiarism detection. (See Image: Similarity Index vs AI Writing Score).



Similarity Index vs AI Writing Score: A scatter plot illustrating the independence of these two metrics.

Discussion

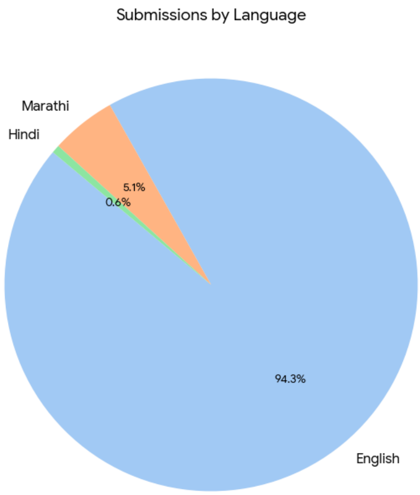
The findings of this study highlight a paradigm shift in academic dishonesty. Traditionally, the primary tool for quality control was the similarity index. If a paper showed low similarity, it was generally accepted as original work. However, our data shows a mean AI score of nearly 21%, which is significantly higher than the mean similarity score of 13%.

The "AI-Originality" Paradox

Documents can have 0% similarity but 90% AI scores. These documents appear "original" to traditional software because the text is uniquely generated by a model, but they lack the human intellectual rigor required for academic excellence. This suggests that institutions can no longer rely solely on similarity reports.

Language Variations

The inclusion of Marathi and Hindi submissions indicates a growing need for multi-lingual AI detection tools. While English detection is robust, the accuracy of detection in regional languages remains a challenge for future software development. (See Image: Submissions by Language).



Submissions by Language: A pie chart showing the linguistic diversity of the sample.

Conclusion

This study analyzed 357 academic submissions to evaluate the prevalence of plagiarism and AI-generated content. The results indicate that while blatant plagiarism (high similarity) is relatively rare, with most scores under 15%, AI-assisted writing is more widespread, with an average score of 21%. The lack of correlation between Similarity and AI scores proves that these are two distinct threats to academic integrity. As we move forward, the definition of "original work" must evolve to account for the role of generative technology in the writing process.

Recommendations

- ✓ Dual-Metric Evaluation: Academic institutions should mandate the reporting of both Similarity and AI scores for every submission.
- ✓ Policy Updates: Universities must clearly define "permissible AI use." For example, using AI for grammar correction should be distinguished from using it to generate core arguments.
- ✓ Invest in Multi-lingual Detection: As academic output in regional languages increases, detection tools must be validated for languages like Marathi and Hindi.

- ✓ Educator Training: Faculty members should be trained to recognize the "hallmarks" of AI-generated text, such as overly generic phrasing and a lack of specific, localized citations.
- ✓ Promoting Ethical AI Usage: Instead of a total ban, students should be taught how to use AI as a research assistant rather than a research replacement.

Study Limitations

- ✓ The sample size of 357, while informative, may not represent all academic disciplines equally.
- ✓ AI detection scores are probabilistic. A high score is an indicator, not definitive proof, of AI usage.
- ✓ The dataset lacks qualitative context regarding whether the AI usage was disclosed by the authors.
- ✓ Future Research: Future studies should focus on a longitudinal analysis to see how these metrics change as AI tools become more sophisticated and detection algorithms evolve to keep pace.

References

[1] Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. *Social Sciences & Humanities Open*, *11*, 101299. https://doi.org/10.1016/j.ssaho.2025.101299

[2] 2.Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society*, *75*, 102370. https://doi.org/10.1016/j.techsoc.2023.102370

[3] Braunack-Mayer, A., & Carter, S. M. (2015). Ethics and health promotion: Research, theory, policy and practice. *Health Promotion Journal of Australia*, *26*(3), 165–166. https://doi.org/10.1071/HEv26n3_ED1

[4] ChatGPT in higher education: Considerations for academic integrity and student learning. (2023). *Journal of Applied Learning & Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.17

[5] Chou, C., Lee, I. J., & Fudano, J. (2024). The present situation of and challenges in research ethics and integrity promotion: Experiences in East Asia. *Accountability in Research*, *31*(6), 576–599. https://doi.org/10.1080/08989621.2022.2155144

[6] Davis, A. (2023). Academic integrity in the time of contradictions. *Cogent Education*, *10*(2), 2289307. https://doi.org/10.1080/2331186X.2023.2289307

[7] Holden, O. L., Norris, M. E., & Kuhlmeier, V. A. (2021a). Academic Integrity in Online Assessment: A Research Review. *Frontiers in Education*, *6*, 639814. https://doi.org/10.3389/feduc.2021.639814

[8] Holden, O. L., Norris, M. E., & Kuhlmeier, V. A. (2021b). Academic Integrity in Online Assessment: A Research Review. *Frontiers in Education*, *6*, 639814. https://doi.org/10.3389/feduc.2021.639814

[9] Karkoulian, S., Sayegh, Niveen, & Sayegh, Nadeen. (2025). ChatGPT Unveiled: Understanding Perceptions of Academic Integrity in Higher Education - A Qualitative Approach. *Journal of Academic Ethics*, *23*(3), 1171–1188. https://doi.org/10.1007/s10805-024-09543-6

[10] Kolstoe, S. E., & Pugh, J. (2024). The trinity of good research: Distinguishing between research integrity, ethics, and governance. *Accountability in Research*, *31*(8), 1222–1241. https://doi.org/10.1080/08989621.2023.2239712

[11] Malik, A., Khan, M. L., Hussain, K., Qadir, J., & Tarhini, A. (2025). AI in higher education: Unveiling academicians' perspectives on teaching, research, and ethics in the age of ChatGPT. *Interactive Learning Environments*, *33*(3), 2390–2406. https://doi.org/10.1080/10494820.2024.2409407

[12] Mejía, A., & Garcés-Flórez, M. F. (2025). What do we mean by academic integrity? *International Journal for Educational Integrity*, *21*(1), 1. https://doi.org/10.1007/s40979-024-00176-1

[13] Melisa, R., Ashadi, A., Triastuti, A., Hidayati, S., Salido, A., Ero, P. E. L., Marlini, C., Zefrin, Z., & Al Fuad, Z. (2025). Critical Thinking in the Age of AI: A Systematic Review of AI's Effects on Higher Education. *Educational Process International Journal*, *14*(1). https://doi.org/10.22521/edupij.2025.14.31

[14] Murumba, J., & Alari, J. O. (2024). An Evaluation of Academic Integrity and Sustainable Quality Education in Higher Learning Institutions in Kenya: Students' Perspectives. *Kabarak Journal of Research & Innovation*, *13*(4), 81–94. https://doi.org/10.58216/kjri.v13i4.249

[15] Newson, A. J., & Lipworth, W. (2015). Why should ethics approval be required prior to publication of health promotion research? *Health Promotion Journal of Australia*, *26*(3), 170–175. https://doi.org/10.1071/HE15034

[16] Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, *20*(2). https://doi.org/10.53761/1.20.02.07

[17] Rodrigues, M., Silva, R., Borges, A. P., Franco, M., & Oliveira, C. (2025). Artificial intelligence: Threat or asset to academic integrity? A bibliometric analysis. *Kybernetes*, *54*(5), 2939–2970. https://doi.org/10.1108/K-09-2023-1666

[18] Royce Sadler, D. (2012). Assessment, evaluation and quality assurance: Implications for integrity in reporting academic achievement in higher education. *Education Inquiry*, *3*(2), 201–216. https://doi.org/10.3402/edui.v3i2.22028

[19] Saqib, M. B., & Zia, S. (2025). Evaluation of AI content generation tools for verification of academic integrity in higher education. *Journal of Applied Research in Higher Education*, *17*(4), 1430–1440. https://doi.org/10.1108/JARHE-10-2023-0470