



Mahavir Education Trust's  
**Shah & Anchor Kutchhi Engineering College**  
An Autonomous Institute Affiliated to University of Mumbai

**UG Program in Computer Engineering**  
(Accredited by NBA for 3 years from AY 2022-23)

**Roll no:** 25

**Name 1:** KATKAM SUYASH SRINIVAS

**Roll no:** 28

**Name 2:** KULKARNI ANUGRAH ARUNKUMAR

**Roll no :** 29

**Name 3:** LIMBACHIYA KARTIK MALKESH

**Roll no :** 44

**Name 4:** PATIL SUJAL KISHOR

**Class Batch:** TY9 - B

**Paper Id:** TY9-B1

**Title:**

**Optimization of LLM & Increasing Its Efficiency Using ZSL Over Traditional  
Keyword-Based & TF-IDF Grading Models**

**Problem Statement:**

Manual grading of exam scripts is a time-consuming, subjective, and inconsistent process, consuming approximately 20% of an educator's time and reducing productivity in teaching and research. Traditional grading methods are prone to human bias and errors, leading to potential unfairness in assessments. With increasing student numbers and complex assessment structures, maintaining fair and standardized grading has become a significant challenge. To address these issues, AI-driven automated exam evaluation offers a promising alternative by enhancing accuracy, efficiency, and fairness in grading. This research investigates whether an LLM-based automated grading system using Zero-Shot Learning (ZSL) and Generative AI can outperform traditional manual grading methods in terms of accuracy, consistency, and scalability.

**Intervention:**

The study proposes an AI-powered automated grading system that leverages GPT-4 with Zero-Shot Learning (ZSL) and Generative AI to evaluate student responses without prior training. The system applies Algorithm 1: Semantic Similarity Measurement, which calculates cosine similarity between student responses and

model-generated answers. Additionally, it implements Algorithm 2: Synonym Similarity Measurement, which assigns scores based on exact matches, synonym matches, and contextual similarity. These techniques ensure that non-standard but contextually correct answers are fairly evaluated. The system also incorporates iterative refinement techniques, improving answer evaluation over time to enhance grading accuracy and fairness.

### **Comparison:**

The performance of the AI-powered grading system is compared against traditional human grading. The study evaluates 20 descriptive questions, divided into 10 annotated and 10 non-annotated questions. The system's effectiveness is measured using absolute and relative error calculations, comparing AI-generated scores with those of experienced educators.

### **Outcome:**

The proposed automated grading system demonstrates a high accuracy of 94.32%, with a low relative error of 1.29% for annotated and 1.67% for non-annotated questions. The system provides fair, unbiased, and consistent grading, significantly reducing grading time and allowing educators to focus on teaching and research. With Zero-Shot Learning, the model adapts to new questions without additional training, making it a scalable solution for educational assessments. The study concludes that AI-driven grading can effectively replace traditional grading methods, ensuring efficiency, fairness, and accuracy in academic evaluations.

### **Abstract:**

Manual grading of exam scripts is a time-consuming and labor-intensive process that is prone to subjectivity and inconsistencies, often leading to unfair assessments. This research presents an automated grading system using GPT-4, integrated with Zero-Shot Learning (ZSL) and Generative AI, to enhance the accuracy, efficiency, and fairness of exam evaluation. The system employs semantic similarity measures, cosine similarity calculations, and synonym recognition techniques to assess student responses objectively. Performance evaluations indicate a low relative error of 1.29% for annotated and 1.67% for non-annotated questions compared to human graders, achieving an overall accuracy of 94.32%. These results demonstrate the transformative potential of AI-driven assessment systems in education by reducing grading time, eliminating human bias, and ensuring a consistent evaluation process. The study highlights how adaptive AI and iterative refinement techniques can improve grading accuracy over time, making it a scalable and reliable solution for modern educational institutions.

### **Introduction:**

Manual grading of exam scripts is time-intensive and prone to human bias, consuming nearly 20% of educators' time and leading to inconsistent assessments. As student enrollment grows, the need for an efficient, standardized, and unbiased grading system becomes more critical. This study presents an LLM-based automated grading system using GPT-4, integrating Zero-Shot Learning (ZSL) and Generative AI to evaluate exam responses without additional

training. The system applies semantic similarity and synonym recognition to ensure accurate assessment of non-standard answers. Tested on 20 descriptive questions, the system achieved a 94.32% accuracy rate, with a low relative error of 1.29% for annotated and 1.67% for non-annotated questions, closely matching human grading. These results highlight AI-driven grading's potential to enhance efficiency, fairness, and consistency in educational assessments, offering a scalable and reliable alternative to traditional methods.

**Key words:**

Automated Exam Grading, Large Language Models (LLMs), Zero-Shot Learning (ZSL), Generative AI, Semantic Similarity, Educational Assessment, Fairness in Grading, AI-Powered Evaluation

**References**

- [1] H. Hardison, "How teachers spend their time: A breakdown," Education Week, 2022.
- [2] R. Shrestha, R. Gupta, and P. Kumari, "Automatic AnswerSheet Checker," EasyChair, 2022.
- [3] S. Chamarathi, R. Balaji, and S. Kumar, "Automatic answer checker for descriptive answers," Natural Language Processing Journal, 2021.
- [4] M. M. Rahman and F. H. Siddiqui, "NLP-based automatic answer script evaluation," DUET Journal, 2020.
- [5] S. Pulman and J. Sukkarieh, "Automatic short answer marking," in Proceedings of the Second Workshop on Building Educational Applications Using NLP, 2005, pp. 9–16.
- [6] S. N. Nobel, S. Sultana, M. A. Tasir, and M. S. Rahman, "Next Word Prediction in Bangla Using Hybrid Approach," in 2023 26th International Conference on Computer and Information Technology (ICCIT), pp. 1–6, IEEE, 2023.
- [7] System Analysis — System Design - GeeksforGeeks, "System Analysis — System Design," accessed Aug. 19, 2024. [Online]. Available: <https://www.geeksforgeeks.org/system-analysis-system-design/>.
- [8] S. Sultana, J. Hossain, M. Billah, H. Hossain Shajeeb, S. Rahman, K. Ansari, and K. F. Hasan, "A Decentralized Blockchain-Enabled Federated Learning Approach for Vehicular Networks," in 2023