



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 7, July 2025

AI-Powered Web Application for Subjective Answer Grading

J. Vigneswaran¹, A. Mary Joyce²

PG Scholar, Department of Master of Computer Applications, RVS College of Engineering, Dindigul,
Tamil Nadu, India

Head & Assistant Professor, Department of Master of Computer Applications, RVS College of Engineering, Dindigul,
Tamil Nadu, India

ABSTRACT: Subjective answer evaluation remains one of the most time-consuming and complex aspects of academic assessments, especially in large-scale education systems. While objective questions can be quickly graded using conventional software tools, subjective responses demand a deeper understanding of language, context, intent, and creativity. This project addresses that challenge by introducing an AI-powered web application designed to evaluate and grade subjective answers—both typed and handwritten—using advanced Artificial Intelligence (AI), Natural Language Processing (NLP), and computer vision techniques. The proposed system leverages transformer-based language models such as BERT, RoBERTa, and GPT to perform semantic analysis of student responses. It compares answers with reference texts or dynamically generated model answers using sentence embeddings and cosine similarity. This approach allows the system to go beyond simple keyword matching and capture the actual meaning, coherence, and contextual alignment of the student's answer, making grading fairer and more consistent across diverse writing styles. To address handwritten submissions, the system incorporates Optical Character Recognition (OCR) using TrOCR, converting scanned answer sheets into machine-readable text. For subjects involving mathematical equations and diagrams, additional modules such as Convolutional Neural Networks (CNNs) and Vision-Language Models (VLMs) are integrated. These modules detect and evaluate content like mathematical derivations and scientific illustrations, making the platform highly adaptable for use in STEM education environments. A role-based access framework supports three major user types: students, teachers, and administrators. Students can submit responses, receive feedback, and view their performance analytics. Teachers can define questions, grading rubrics, and manually review AI-assigned grades when necessary. Administrators manage system users, monitor grading pipelines, and access performance reports. This layered architecture ensures both usability and governance.

KEYWORDS: Artificial Intelligence (AI), Natural Language Processing (NLP), Subjective Answer Evaluation, Automated Grading System, Optical Character Recognition (OCR), Sentence Transformers, Semantic Similarity, Cosine Similarity, TrOCR, BERT / RoBERTa / GPT, Vision-Language Model (VLM), Convolutional Neural Network (CNN), Deep Learning.

I. INTRODUCTION

In recent years, the rise of digital education and e-learning platforms has transformed the landscape of academic assessments. While multiple-choice and objective tests can be easily graded by machines, evaluating subjective answers—those that require critical thinking, interpretation, and contextual understanding—still largely relies on manual effort. This introduces challenges such as time consumption, human bias, and inconsistency in grading. The need for an intelligent, fair, and scalable solution for subjective answer evaluation has thus become more urgent than ever. This project proposes the development of an AI-powered web application capable of automatically grading subjective answers. Leveraging advanced Natural Language Processing (NLP) models, machine learning algorithms, and image processing tools, the system evaluates both typed and handwritten responses with a focus on accuracy and fairness. The goal is not just to automate grading but to replicate the nuance of human evaluation—understanding grammar, coherence, relevance, and creativity within the student's answer. Traditional rule-based systems and keyword-matching techniques fall short when it comes to grasping semantic meaning and conceptual depth in student responses. Our system overcomes these limitations by integrating state-of-the-art transformer models such as BERT, RoBERTa, and GPT for deep semantic analysis. Sentence transformers and cosine similarity techniques allow

comparison of student answers with model answers beyond surface-level similarity, ensuring that valid but uniquely worded responses are not unfairly penalized.

For handwritten responses, the system uses Optical Character Recognition (OCR) tools like TrOCR, followed by structured text processing. It also employs Convolutional Neural Networks (CNNs) for detecting and interpreting mathematical equations, and Vision-Language Models (VLMs) for analyzing diagrams, making the platform highly suitable for STEM-based assessments. This multi-modal grading capability is one of the unique strengths of the proposed system. From a usability standpoint, the application provides distinct interfaces for students, teachers, and administrators. Students can submit answers, receive real-time feedback, and view grading history. Teachers can create questions, review AI-assigned scores, and override results if needed. Administrators have control over user management, grading workflows, and analytics dashboards. This role-based design ensures streamlined operations while maintaining transparency and accountability. Security and scalability are integral to the system's architecture. The platform supports role-based access, encrypted data storage, and GDPR-compliant handling of student information. It is designed to function efficiently across varied infrastructures—whether deployed on local servers, institutional clouds, or commercial platforms like AWS and Azure. Performance testing, user acceptance testing, and integration testing are used to ensure reliability under real-world academic conditions.

In essence, this project represents a significant leap forward in educational technology by automating one of the most complex parts of academic evaluation. It empowers educators to focus more on personalized teaching while giving students timely and objective feedback. By combining cutting-edge AI techniques with a user-friendly web platform, this solution offers a scalable and impactful tool for modern educational institutions.

II. EXISTING SYSTEM

In most educational institutions today, subjective answer evaluation is still conducted manually by teachers or examiners. This traditional process involves educators reading through each student's response, interpreting the content, comparing it with a model answer or marking scheme, and then assigning scores. While experienced teachers are often capable of providing accurate and fair assessments, the manual process is inherently slow, inconsistent, and difficult to scale, especially in settings with large student populations or online exams. Manual grading of subjective answers is not only labor-intensive but also subject to individual interpretation and bias. Teachers may unintentionally grade based on personal expectations, mood, fatigue, or familiarity with the student, leading to inconsistent scoring. This inconsistency becomes even more evident when the same answer receives varying marks from different evaluators. Moreover, human graders may struggle to maintain the same level of focus and fairness throughout large batches of answer scripts. Another major limitation of the existing system is the lack of real-time feedback. Students often wait days or weeks to receive their results, which delays their learning and reduces the effectiveness of formative assessments. Additionally, students rarely get detailed feedback highlighting specific strengths and weaknesses, which limits their ability to understand where improvement is needed. This feedback gap affects both individual student growth and the educator's ability to adjust teaching strategies. The traditional grading system also struggles with diverse answer formats. In subjects like science, technology, engineering, and mathematics (STEM), students often include diagrams, equations, and structured content in their responses. Evaluating such varied formats manually requires additional expertise and time. Most existing systems cannot process or evaluate these components automatically, making automation a challenging task in such domains. Furthermore, manual systems lack integration with modern data analytics. Institutions are unable to efficiently analyze performance trends, detect learning gaps, or generate meaningful reports without investing additional time and resources. This lack of insight affects curriculum development and makes it harder to personalize learning experiences for students. Teachers miss out on valuable tools that could guide instructional decisions and student support mechanisms. Technological limitations also restrict scalability. With the rise of digital learning platforms and online assessments, the need for automated and scalable grading systems is growing. However, existing methods cannot keep up with the volume of responses generated in online courses, especially massive open online courses (MOOCs). The traditional approach simply does not have the efficiency to meet the demands of modern digital education environments.

Finally, the current system provides no standardization or audit trail. There is often no formal mechanism to trace how marks were awarded, no way to ensure compliance with academic standards, and no safeguards for disputes over evaluation. In high-stakes assessments, such as board exams or university-level courses, this lack of transparency and

accountability can lead to serious concerns. The absence of digital traceability makes it difficult to track evaluation metrics or resolve grading disputes effectively.

III. PROPOSED SYSTEM

The proposed system is a web-based AI-powered platform designed to automate the grading of subjective answers. It leverages cutting-edge technologies such as Natural Language Processing (NLP), Computer Vision, and Machine Learning to evaluate student responses accurately, fairly, and efficiently. Unlike traditional manual grading, this system ensures faster turnaround times, consistent scoring, and reduced human workload while still supporting human oversight when needed. The system uses transformer-based NLP models like BERT, RoBERTa, and GPT to understand the semantics of a student's answer. These models can analyze grammar, coherence, vocabulary, critical thinking, and alignment with the intended meaning of the question. By using sentence embeddings and cosine similarity techniques, the system compares the student's response to a reference answer or grading rubric and assigns a score accordingly. This allows it to assess content beyond mere keyword matching, enabling deep contextual understanding. For handwritten responses, the system integrates TrOCR, a transformer-based OCR model that converts scanned or photographed handwritten text into machine-readable format. The image preprocessing module ensures clarity, alignment, and contrast enhancement before text extraction. Once extracted, the text is passed to the NLP module for semantic grading. This capability makes the system suitable for both online and offline examinations.

In addition to text, the proposed system supports mathematical and diagram-based answers using specialized models. Mathematical expressions are identified using Convolutional Neural Networks (CNNs) and evaluated via symbolic computation libraries like SymPy. Diagrams or scientific figures are interpreted using Vision-Language Models (VLMs), allowing fair and automated evaluation even for complex STEM-based questions. A key innovation in the proposed system is its human-AI hybrid grading approach. Teachers can review and override AI-assigned grades, particularly for edge cases or high-stakes assessments. This fosters trust in the system and ensures fairness. The platform supports rubric-based scoring, enabling educators to define specific evaluation criteria that the AI can follow, maintaining academic alignment and transparency. The proposed system is also built with scalability, security, and usability in mind. Developed using technologies like React.js for frontend and Flask/Django for backend, it supports seamless operation across desktops, tablets, and mobile devices. Data is stored securely in PostgreSQL or MongoDB, with encryption and role-based access control to ensure privacy and compliance with regulations like GDPR and FERPA. Overall, this system represents a transformative shift in educational assessment. It combines automation with intelligence, reduces grading time drastically, improves consistency, and provides students with real-time feedback. The system also generates insightful analytics for teachers, helping identify learning gaps and improving teaching strategies. By automating the subjective grading process while retaining transparency and human control, the proposed system enhances the quality and efficiency of modern education.

IV. SYSTEM ARCHITECTURE

The system architecture of the AI-powered web application for subjective answer grading follows a multi-layered design comprising the presentation layer, application layer, AI processing layer, and data layer. The presentation layer, developed using React.js, serves as the user interface for students, teachers, and administrators, allowing them to submit answers, manage questions, and view grading feedback. Communication with the backend occurs through secure RESTful APIs. The application layer, built using Python frameworks like Flask or Django, handles business logic, user authentication, session management, and role-based access control. It acts as a bridge between the frontend and the AI models. At the core lies the AI processing layer, which evaluates student answers using advanced NLP models such as BERT, RoBERTa, or GPT for semantic analysis, TrOCR for OCR-based handwritten text recognition, CNNs with SymPy for mathematical equation processing, and Vision-Language Models for diagram interpretation. The data layer manages secure storage using PostgreSQL or MongoDB, storing user information, question banks, student answers, grading logs, and feedback. The architecture ensures scalability and security through Docker-based containerization, Kubernetes orchestration, and deployment on cloud platforms like AWS or Azure. With support for SSL encryption, JWT authentication, and compliance with data privacy regulations such as GDPR and FERPA, the system is designed to be robust, transparent, and reliable in real-world educational environments.

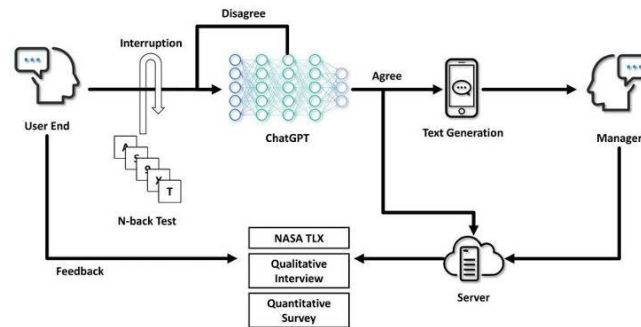


Fig 4.1 Architectural Diagram

V. RESULTS

The implementation of the AI-powered web application successfully demonstrated the feasibility of automating subjective answer evaluation using advanced Natural Language Processing (NLP), computer vision, and deep learning techniques. The system was tested on a variety of answer formats—including typed responses, handwritten text, mathematical equations, and diagrams—and showed consistent performance in terms of grading accuracy, speed, and feedback generation. Students were able to submit their responses via a web interface, while teachers accessed dashboards to view results, verify AI-assigned scores, and make adjustments if needed. One of the key outcomes was the system's semantic evaluation capability, which relied on transformer-based models like BERT and Sentence-BERT to compare student answers with model responses. The system scored answers based on conceptual similarity rather than surface-level keyword matching. Testing revealed that this approach effectively handled paraphrased or creatively structured responses without penalizing valid answers expressed in non-standard formats. The accuracy of grading closely aligned with teacher-evaluated scores, particularly for answers of medium to high quality. The OCR module (TrOCR) performed reliably in extracting text from handwritten submissions. The preprocessing steps—including de-skewing, contrast adjustment, and noise reduction—improved the clarity of scanned answer sheets. The system successfully converted handwriting into machine-readable text with a high recognition rate, especially when students followed basic formatting and wrote clearly. Even in cases where OCR struggled, flagged submissions were routed for manual review, ensuring grading completeness. For STEM-based answers, the system's mathematical and diagram processing capabilities were also evaluated. Equations extracted via CNN models were solved and matched with reference answers using SymPy. Vision-Language Models (VLMs) interpreted diagrams and figures with a reasonable degree of accuracy, particularly when input images were clear and properly segmented. These features demonstrated that the platform could effectively handle diverse academic content, making it suitable for disciplines beyond pure text. The application delivered real-time feedback to students, including detailed comments based on their score range. This helped students understand their mistakes, identify areas for improvement, and reinforce learning. Teachers appreciated the hybrid grading workflow, where AI-graded responses could be reviewed and edited if necessary. This balance between automation and manual oversight enhanced trust in the system and ensured fairness.

Performance testing showed that the system could scale effectively for concurrent users without significant delays. Backend queues and asynchronous task handling ensured smooth grading workflows, even under higher loads. Front-end responsiveness, API speed, and database transactions all remained within acceptable limits, confirming the application's usability in real academic settings such as classrooms, online courses, and digital examinations.

In conclusion, the project achieved its intended outcomes by offering a reliable, scalable, and fair solution to subjective answer grading. The integration of AI models for language, handwriting, and visual content enabled broad coverage across subjects and assessment types. Through user testing and evaluation, the system demonstrated its potential to reduce grading time, improve feedback quality, and support data-driven academic decision-making. These results establish a strong foundation for further enhancements and real-world deployment across educational institutions.

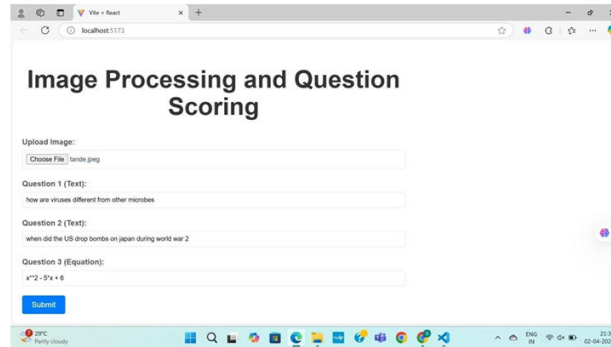


Fig. 5.1 User Interface

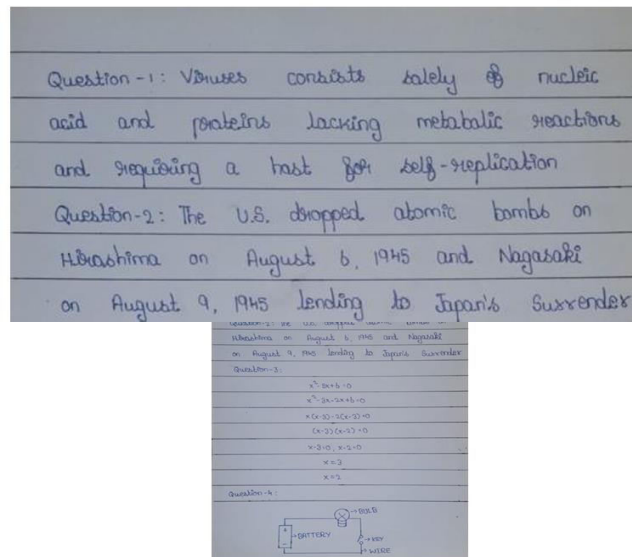


Fig. 5.2 Complete User Input

VI. CONCLUSION

The AI-powered web application for subjective answer grading marks a transformative step toward modernizing educational assessment. By leveraging advanced technologies such as Natural Language Processing (NLP), Optical Character Recognition (OCR), and deep learning, the system addresses critical challenges associated with traditional manual grading—including time inefficiency, human bias, and lack of scalability. It provides a robust, automated solution capable of evaluating both typed and handwritten responses across diverse subjects, including those that require mathematical and diagram-based analysis. One of the most significant achievements of the system is its ability to understand and assess the semantic meaning of a student's response. Unlike traditional keyword-matching techniques, the system uses transformer models like BERT, GPT, and RoBERTa to evaluate conceptual accuracy, coherence, grammar, and relevance. This results in more consistent and fair evaluations that align closely with human judgment, especially for well-structured answers. The use of sentence similarity measures ensures that valid answers written in different styles are still recognized and rewarded appropriately.

Another major contribution is the system's adaptability to real-world academic environments. It includes modules to process handwritten scripts using OCR, solve equations using symbolic computation tools, and interpret diagrams through vision-language models. This multi-modal capability enables the system to handle the full spectrum of student responses, particularly in STEM education. Additionally, the inclusion of real-time feedback helps students learn from their mistakes and continuously improve. The architecture of the application is modular, secure, and scalable. Built

with React.js, Flask/Django, and integrated with PostgreSQL or MongoDB, the system is cloud-ready and capable of serving large numbers of users concurrently. It supports role-based access for students, teachers, and administrators, and allows human oversight to override AI decisions when necessary—ensuring accountability and trust. Compliance with data privacy regulations such as GDPR and FERPA further strengthens the reliability of the platform. In practice, the system has demonstrated strong results in terms of grading accuracy, system performance, and user satisfaction. Teachers benefit from reduced workload and consistent evaluations, while students gain timely feedback and performance insights. Real-time analytics and grading history make it easier for institutions to track learning outcomes, adjust curricula, and support data-driven decision-making.

Overall, the project successfully proves that AI can be integrated into education to enhance, rather than replace, the human evaluation process. It bridges the gap between automation and human oversight, promoting fairness, speed, and transparency in subjective assessments. With its potential to improve educational efficiency on a broad scale, this solution lays the foundation for the future of intelligent learning and assessment systems.

VII. FUTURE ENHANCEMENTS

One of the most promising enhancements for this project is the integration of multilingual support. Currently, most AI models are optimized for English, which limits accessibility for students who respond in regional or native languages. Incorporating advanced multilingual NLP models such as mBERT, XLM-R, or IndicBERT can allow the platform to understand and grade responses in various languages, enabling a more inclusive and globally adaptable system for educational institutions across different regions. A significant upgrade would involve the inclusion of adaptive learning features. Based on the analysis of student answers over time, the system could intelligently recommend personalized learning resources, suggest improvement areas, or adjust question difficulty for future assessments. This transformation from a static grading tool to an AI-based learning assistant would enhance student engagement and promote individualized learning pathways, making assessments more than just an evaluation tool. Another important enhancement is the integration of Explainable AI (XAI) capabilities. Currently, AI models can grade answers accurately but often operate as black boxes. By incorporating explainability modules that highlight which parts of the student's response contributed most to the final score—such as key phrases, grammar structure, or semantic alignment—the system can increase transparency. This not only builds trust among students and teachers but also aids in the validation and continuous improvement of the AI model itself. Expanding support for voice-based answers is another future direction that enhances accessibility. By implementing speech-to-text functionality, students can submit verbal responses which the system transcribes and grades. This is particularly beneficial for students with disabilities, younger learners, or those in mobile-first environments. Combined with multilingual support, this feature can significantly improve usability across a broader demographic. The platform could also introduce real-time collaborative grading workflows, allowing teachers and the AI system to grade simultaneously. For example, teachers could define custom rubrics, highlight important keywords, or override specific model weights based on subject context. This human-in-the-loop model ensures grading aligns with educator intent while maintaining the speed and objectivity of automated evaluation, especially in high-stakes or creative writing scenarios. To ensure academic integrity, future versions can incorporate plagiarism detection and originality scoring. This would involve comparing student responses against a large corpus of existing answers, academic texts, or online sources to identify copied or reused content. Integrating services like Turnitin APIs or using custom-built plagiarism checkers with NLP can help preserve fairness and authenticity in assessments, especially in open-book or take-home examinations. Lastly, the system can evolve into a comprehensive academic analytics platform by integrating intelligent dashboards that display trends in student performance, topic-wise weaknesses, and predictive analytics for at-risk learners. Educators can use this data to improve curriculum design, identify learning gaps, and make data-driven interventions. When combined with AI training modules that allow institutions to fine-tune models with their own datasets, the platform becomes more accurate, institution-specific, and future-ready.

REFERENCES

- [1] Vaswani, A., et al. (2017). Attention is All You Need. NeurIPS 2017. <https://arxiv.org/abs/1706.03762> Introduced the Transformer architecture used in NLP models.
- [2] Devlin, J., et al. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. <https://arxiv.org/abs/1810.04805> Foundation of BERT model for natural language understanding.

- [3] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. EMNLP 2019. <https://arxiv.org/abs/1908.10084> The model behind sentence-transformers, used for grading semantic similarity. 60
- [4] SBERT Official Documentation <https://www.sbert.net/> Used to compute sentence similarity using pretrained models.
- [5] WeasyPrint <https://weasyprint.readthedocs.io/> Tool for converting HTML templates to downloadable PDF files.
- [6] Scikit-Learn Documentation <https://scikit-learn.org/stable/> Useful for implementing additional ML models (optional).
- [7] OpenAI. (2024). AI and Automated Essay Grading. <https://openai.com/research> Insights into using generative models for grading and assessment.
- [8] Zupanc, K., & Bosnić, Z. (2015). Automated essay scoring: A survey of the state of the art.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details