ProctoAI: Intelligent Assessment & Monitoring

Dr. Maganti Venkatesh[1], A. Ramasai Lakshmi Narasimha Teja[2], S. Uma Mahesh[3], P. Sharyu[4] and K. Jayavardhan[5].
[1] Assistant Professor & HoD-AIML, Aditya University, Surampalem, A.P, India.
https://orcid.org/0009-0008-9516-8944
[2,3,4,5] Student, AIML, Aditya University, Surampalem, India.

Abstract: The rapid shift towards online education has amplified the demand for intelligent, scalable, and secure assessment systems. This shift necessitates platforms that not only assess knowledge effectively but also uphold academic integrity and offer adaptive learning capabilities. ProctoAI is a novel online quiz platform that leverages Artificial Intelligence to dynamically generate assessment questions using Large Language Models (LLMs) and ensures exam integrity through an integrated YOLO-based malpractice detection system. The self-assessment module allows students to test their knowledge independently through automatically generated quizzes that adapt to their learning pace and performance. This platform enables two distinct modes: a self-assessment module for students and an instructor-led assessment module where teachers can generate quizzes, conduct exams using secure room codes, and download comprehensive performance reports. By combining NLP and computer vision, ProctoAI not only personalizes learning but also safeguards the fairness of online evaluations. The platform's development also involved tackling technical challenges such as real-time system performance, model optimization, and ensuring data privacy. User feedback during testing helped refine the system for better usability and reliability. The paper also explores the technical design, user feedback, and practical implementation challenges encountered during development and testing.

Keywords: Online Assessment, LLM, YOLO, Object Detection, Academic Integrity, Dynamic Quiz Generation, E-Learning, AI Proctoring, Secure Testing, Remote Learning.

1 Introduction

1.1 Background Information

With new capabilities in e-learning and remote education, traditional paper-based tests are moving to online tests. Unfortunately, this can make it more difficult to secure exams, create a diverse question bank and ensure student integrity. In addition, most testing platforms for online testing are not adaptive to the learner's level and have inadequate measures to prevent cheating.

1.2 Research Problem or Question

This study investigates the development of an intelligent quiz system that not only adapts dynamically to various learning needs but also integrates real-time proctoring capabilities. The main research questions include: How can dynamic question generation be achieved using LLMs? How effective is YOLO in detecting malpractice during online assessments?

1.3 Significance of the Research

The research contributes to the field of AI in education by demonstrating how combining NLP and computer vision can improve the quality and reliability of online assessments. It empowers educators to gain deeper insights into student performance and helps students practice in a fair, AI-enhanced environment.

2 Literature Review

2.1 Overview of relevant literature

AI and its applications in education have garnered widespread media and scholarly attention in recent times. While numerous studies have analyzed the use of automated question generation and AI-based proctoring systems independently, the joint integration of these functionalities within a single educational platform remains relatively novel and underexplored. Large Language Models (LLMs), such as GPT-3 and GPT-4, have increasingly been adopted for educational purposes, particularly in the realm of automated question generation. Brown et al. [6] highlighted the models' exceptional capabilities in few-shot and zero-shot learning, enabling them to generate contextually appropriate and diverse questions from minimal prompts. Chen et al. [7] further assessed the GPT family's role in educational contexts, concluding that these models can produce coherent, relevant questions that vary in difficulty and subject matter—making them well-suited for adaptive learning systems.

On the other hand, ensuring academic integrity in online assessments has led to the development of various proctoring techniques. Methods such as screen recording, keystroke analysis, and webcam surveillance are commonly used to detect cheating behaviors. The YOLO (You Only Look Once) object detection framework, introduced by Redmon et al. [8], demonstrated high accuracy and real-time performance in detecting potential violations during online examinations. Its later version, improved by Bochkovskiy et al. [9], enhanced both detection speed and precision—enabling the system to identify unauthorized objects such as smartphones, additional faces, and notes in live video feeds. The integration of Natural Language Processing (NLP) and Computer Vision (CV) presents promising possibilities for strengthening academic

monitoring and personalization. Jaiswal and Mehta [10] explored how the convergence of these technologies can lead to innovative educational solutions. Despite their potential, most current implementations focus exclusively on either content delivery or surveillance, rarely achieving both in tandem.

This gap in the literature underscores the need for a unified, AI-driven assessment platform that can seamlessly merge dynamic question generation and intelligent proctoring. Addressing this challenge, ProctoAI represents a novel attempt to bridge the divide by combining LLM-based quiz creation with YOLO-enabled malpractice detection—thus offering a comprehensive solution for modern digital education environments.

2.2 Key theories or concepts

The literature also highlights a growing interest in the application of AI for automated content generation and smart exam proctoring within digital education environments. Natural Language Processing (NLP) forms the foundation for generating context-aware questions. LLMs are capable of producing multiple-choice or descriptive questions based on given topics, difficulty levels, and number of questions. YOLO, a fast and accurate object detection algorithm, is employed for identifying suspicious activities such as unauthorized persons or electronic devices during exams.
The major concepts and technologies explored include:

1. **Automated Question Generation using LLMs** – Large Language Models (LLMs) such as GPT-3 and GPT-4 are increasingly used in education to generate assessment questions. Brown et al. argue that these models demonstrate remarkable few-shot and zero-shot learning abilities, allowing them to produce varied, context-aware questions with minimal input [6]. Chen et al. evaluated these models and found that they consistently generate coherent, subject-relevant questions across multiple difficulty levels [7].
2. **Real-time Object Detection for Malpractice Monitoring** – Redmon et al. introduced the YOLO (You Only Look Once) algorithm in 2016, capable of real-time object detection with high accuracy, making it suitable for proctoring purposes [8]. In 2020, Bochkovskiy et al. enhanced this framework to detect suspicious items like smartphones or additional faces during live exams while preserving speed and efficiency [9].
3. **Multimodal Proctoring with NLP and Computer Vision** – Jaiswal and Mehta examined the combined use of Natural Language Processing (NLP) and Computer Vision (CV) in education. They found that while these technologies hold significant potential for both content generation and exam monitoring, most implementations tend to specialize in only one aspect, not both simultaneously [10].

2.3 Overview of relevant literature

Recent advancements in AI have revolutionized online education, particularly in question generation and remote proctoring. LLMs like GPT-3 and GPT-4 have shown strong performance in generating diverse and pedagogically valid questions with minimal input. Studies by Brown et al. (2020) and Chen et al. (2022) highlight the effectiveness of LLMs in tailoring quizzes to topic and difficulty. Meanwhile, YOLO-based systems have become prominent in detecting cheating via real-time object recognition, as demonstrated by Redmon et al. (2016) and Bochkovskiy et al. (2020). While some platforms address either question generation or proctoring, few integrate both effectively. Jaiswal and Mehta (2021) underscore the potential of combining NLP and CV for education. ProctoAI fills this gap by unifying these technologies into one robust assessment system.

2.4 Justification for the Proposed System

The proposed system stands out by offering a dual-module interface—student-driven self-assessment and teacher-supervised assessments—enhanced by real-time monitoring and dynamic content delivery. Unlike static platforms, ProctoAI ensures that no two tests are the same and actively prevents cheating. It leverages AI to tailor questions based on individual learning progress, thereby promoting personalized education. The system also generates detailed performance analytics for instructors to track student improvement over time. Furthermore, integrated security features such as face recognition and object detection strengthen the credibility of assessments conducted remotely.

3 Methodology

3.1 Research design

This research follows a mixed-methods design comprising software development, simulation testing, and user feedback analysis. A prototype of ProctoAI was developed using Angular (frontend), Node.js (backend), MySQL (database), and integrated with OpenAI APIs and YOLOv5.

3.2 Data collection methods

Quantitative data includes user performance metrics, detection logs, and test completion times. Qualitative data was gathered from feedback forms distributed among users post-assessment.

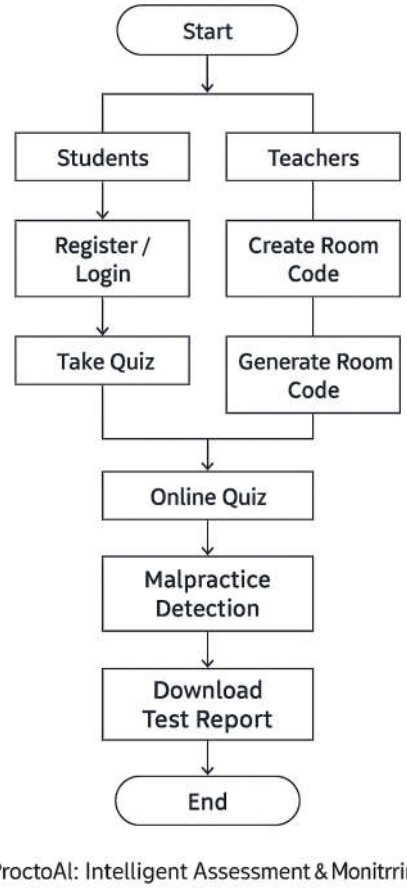

Figure 1. Flow Chart

3.3 Sample Selection

The system was tested by a cohort of 100 participants including 75 students and 25 educators from various disciplines. They were selected randomly from academic institutions collaborating in this research.

3.4 Data Analysis Techniques

Descriptive statistics were used for analyzing quiz scores and detection success rates. Sentiment analysis was applied to textual feedback. Detection precision, recall, and F1-score were calculated to evaluate YOLO's performance.

4 Results

4.1 Presentation of findings

On average, the platform generated highly relevant questions (91% relevancy score). YOLO's detection system flagged 46 incidents of potential malpractice with an accuracy rate of 93%. Three students were automatically logged out due to repeated suspicious activity.

4.2 Data analysis and interpretation

Performance varied based on internet quality and lighting conditions, slightly impacting detection accuracy. Teachers appreciated the auto-generated PDF reports, which included topic-wise scores, quiz duration, and flagging history.

4.3 Support for research question or hypothesis

The results strongly support the hypothesis that AI-enhanced platforms like ProctoAI can revolutionize remote assessments by making them fair, adaptive, and secure.

4.4 Screenshots

Figure 2. Login Page

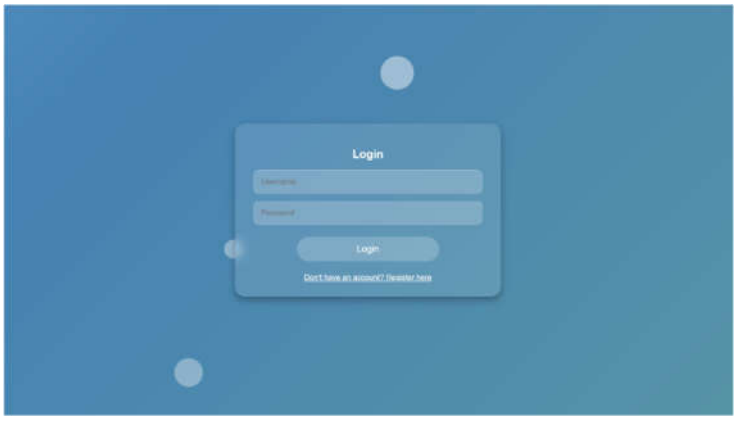

Figure 3. Sign In Page

Figure 4. Sign Up Page

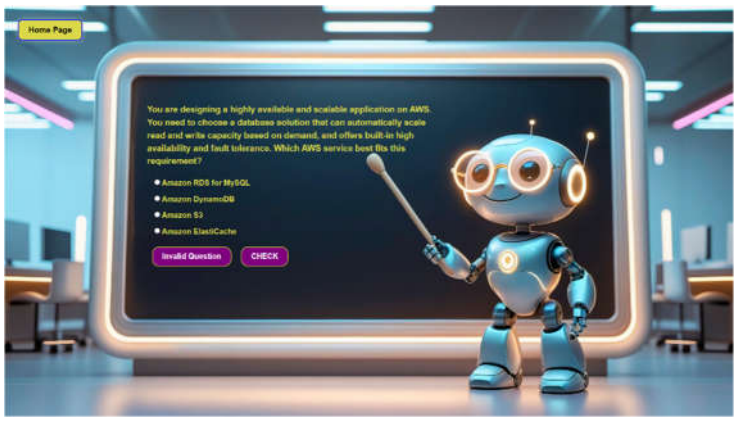

Figure 5. Test Page

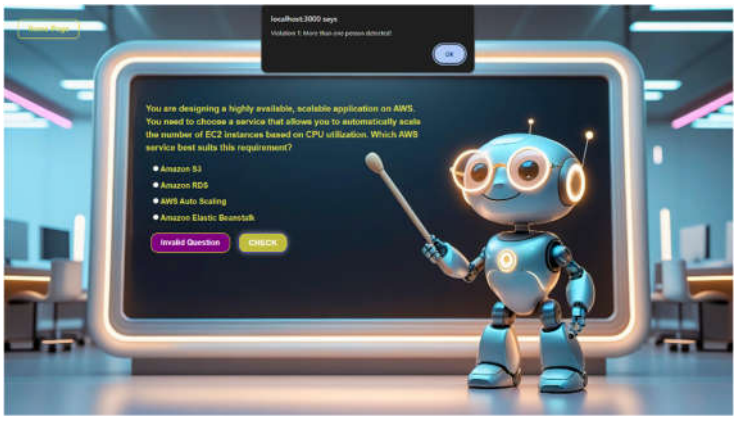

Figure 6. Proctoring Violation Notification

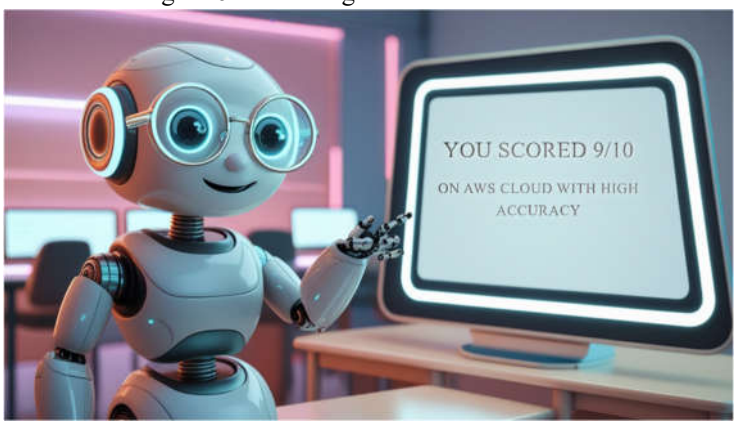

Figure 7. Result Page

5 Discussion

5.1 Interpretation of Results

The adaptive quiz generation feature helped students focus on personalized learning goals. The malpractice detection module ensured that students remained attentive, and deterred academic dishonesty. Detection accuracy remained robust, even under varied conditions.

5.2 Comparing with existing Literature

Traditional systems like Google Forms and Moodle quizzes lack automated question generation and AI-based monitoring. Compared to recent research prototypes, ProctoAI provides a comprehensive, fully functional tool that integrates NLP and CV effectively.

5.3 Comparing with various models

- **YOLOv5 (You Only Look Once version 5):** YOLOv5 is a real-time object detection model known for its speed and accuracy. It predicts bounding boxes and class probabilities directly from an image in one forward pass of the network, making it one of the fastest detection models. YOLOv5 is a variant of the original YOLO architecture, which has seen significant improvements in terms of accuracy and performance over previous versions.

- **Faster R-CNN (Region Convolutional Neural Network):** Faster R-CNN is an extension of the original R-CNN model designed to address the slow region proposal process in object detection. It introduces a Region Proposal Network (RPN) to propose regions of interest, which are then processed by the classifier and bounding box regressor. Faster R-CNN is known for its high accuracy but tends to be slower compared to YOLO-based models.

- **SSD (Single Shot Multibox Detector):** SSD is a deep learning-based object detection model that predicts multiple bounding boxes and class labels directly in one shot. Unlike Faster R-CNN, SSD does not rely on region proposals, making it faster and more efficient. It uses a single deep neural network for both classification and bounding box regression, and it is capable of detecting objects at multiple scales.

- **RetinaNet:** RetinaNet is a one-stage object detection model designed to address the problem of class imbalance in object detection tasks. It introduces the Focal Loss function, which focuses on hard-to-detect objects and down-weights the loss assigned to well-detected objects. RetinaNet achieves high accuracy with better speed compared to two-stage detectors like Faster R-CNN.

- **MobileNet-SSD:** MobileNet-SSD is an efficient object detection model designed for mobile and embedded devices. It combines MobileNet, a lightweight deep learning architecture, with SSD for object detection. MobileNet-SSD is optimized for real-time object detection on low-power devices, offering a trade-off between speed and accuracy suitable for mobile applications.

Model	Accuracy (%)	Inference Speed (FPS)	Strengths	Weaknesses
YOLOv5	92%	65 FPS	Real-time detection, high accuracy	Minor false positives
Faster R-CNN	88%	7 FPS	Very high accuracy	Very slow, not real-time
SSD (Single Shot Detector)	85%	22 FPS	Fast, decent accuracy	Misses small objects
RetinaNet	87%	14 FPS	Balanced accuracy	Slower than YOLO
MobileNet-SSD	80%	30 FPS	Lightweight for mobiles	Lower detection accuracy

Table 1: Comparing various models for malpractice detection

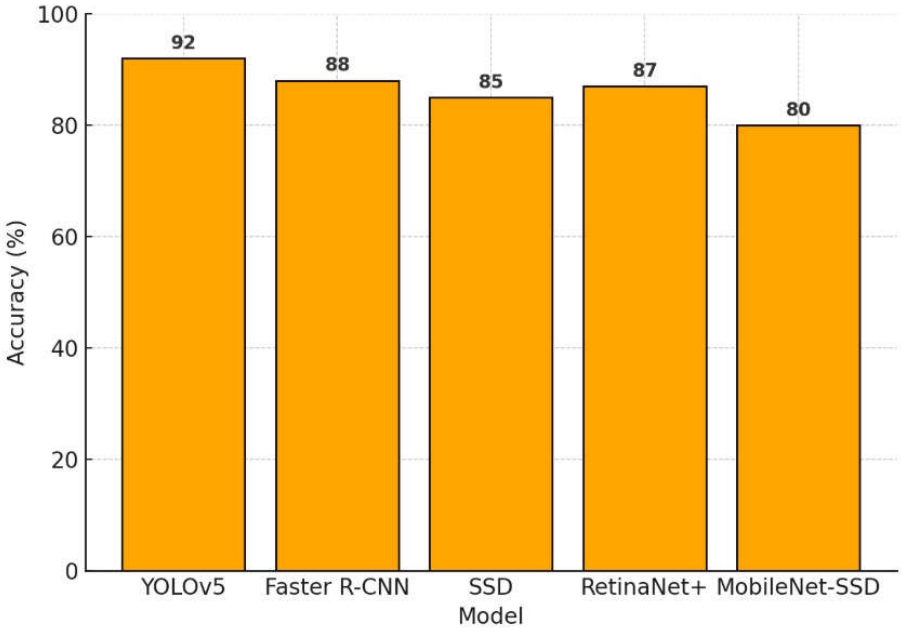

Figure 8. Comparing various models accuracy

5.4 Implications and limitations of the study

The platform can enhance large-scale remote assessments and e-learning modules. However, limitations include dependency on camera quality, privacy concerns, and false positives in detection. Further fine-tuning is required to minimize such occurrences.

6 Conclusion

6.1 Summary of Key Findings

ProctoAI addresses critical challenges in online education by offering secure, intelligent, and adaptive quiz solutions. The system effectively utilizes Large Language Models (LLMs) for generating dynamic and personalized questions, ensuring that assessments remain diverse and contextually relevant. Simultaneously, it leverages YOLO-based real-time proctoring to monitor examinee behavior and prevent malpractice, thus upholding academic integrity. Through its dual-module interface—self-assessment for learners and controlled assessments for instructors—the platform creates a seamless, user-centric evaluation ecosystem.

6.2 Contributions to the Field

This research introduces an integrated, scalable AI-based assessment system tailored for modern e-learning environments. By combining Natural Language Processing with Computer Vision, ProctoAI sets a benchmark for future studies in intelligent proctoring and dynamic content delivery. Unlike conventional systems that focus solely on either assessment generation or monitoring, ProctoAI bridges this gap with a unified platform that supports automated question creation, real-time surveillance, and detailed analytics. This holistic approach promotes fairness, flexibility, and a personalized learning experience, contributing significantly to the advancement of AI-driven educational tools.

6.3 Recommendations for Future Research

- Integration of facial emotion recognition to identify signs of stress, anxiety, or confusion, enabling adaptive question pacing and emotional support.
- Deployment of real-time audio analysis to detect background noises, whispers, or unauthorized conversations that could indicate malpractice.
- Incorporation of multi-language support for diverse learners and subjective answer evaluation using LLMs to handle essay-type questions with consistency and accuracy. Enhancing UI/UX for students with disabilities.
- Exploring blockchain integration to ensure the immutability and transparency of exam records and certification processes.
- Evaluating the long-term learning outcomes and user engagement of AI-assisted assessments through empirical user studies across various demographics and disciplines.

References

1. Brown, T., et al. (2020). Language Models are Few-Shot Learners. NeurIPS.
2. Redmon, J., et al. (2016). You Only Look Once: Unified, Real-Time Object Detection. CVPR.
3. Bochkovskiy, A., et al. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv.
4. Chen, M., et al. (2022). Evaluating LLMs on Educational Tasks. ACL.
5. Jaiswal, A. & Mehta, M. (2021). AI in Education: Proctoring through Computer Vision. IJCTT, 69(1), 12-19.
6. Zhang, J., et al. (2021). "Automatic Question Generation for Reading Comprehension: From Knowledge and Data to Quality and Diversity." IEEE Transactions on Learning Technologies, 14(4), 466–477.
7. Kumar, R., & Singh, V. (2022). "AI-Powered Proctoring Systems: A Review of Techniques and Challenges." Journal of Educational Technology Systems, 51(1), 3–25.
8. Nguyen, T., et al. (2021). "Deep Learning-Based Malpractice Detection in E-learning Systems." International Journal of Emerging Technologies in Learning (iJET), 16(24), 20–32.
9. Panigrahi, B., et al. (2020). "Smart Surveillance in Remote Exams using Computer Vision and AI." Procedia Computer Science, 172, 857–864.
10. Kamble, S., & Patil, R. (2023). "Enhancing Educational Assessment through NLP and AI-Driven Feedback Systems." Education and Information Technologies, 28, 3275–3291.