Automated AI-Based Exam Grading System

Abstract

In the field of educational assessment, manual grading of exam papers takes a lot of time and is prone to mistakes. This paper introduces an AI-Based Automated Exam Grading System that aims to simplify this process. By using Large Language Models (LLMs), specifically the Google Gemini API, the system can evaluate both text-based and image-based handwritten student answers against a solution key. The proposed solution takes a hybrid approach, using standard text parsing for structured digital submissions and multimodal vision capabilities for handwritten content. The system offers detailed feedback, gives partial credit based on a set marking scheme, and highlights alternative valid methods for manual review. This automation greatly cuts down grading time while keeping consistency high and offering detailed insights into student performance.

The code can be found here- https://github.com/vishnuchebol/AI_exam_corrector.git

Section 1: Introduction

1.1 Problem Statement

Evaluating student assessments is extremely time consuming for educators. Manual grading suffers from several inherent limitations: it is slow, prone to inconsistency due to human fatigue or bias, and often results in delayed feedback for students. Furthermore, scaling becomes challenging.

1.2 Limitations of Existing Solutions

Existing automated grading tools rely on rigid keyword matching or standard Optical Character Recognition (OCR). These approaches struggle with contextual understanding and large variations in handwriting.

1.3 Contributions

This paper proposes a robust AI-driven grading system that addresses these gaps. The key contributions are:

1. Ability to evaluate both typed text and handwritten images.
2. Semantic and context-aware scoring using Google Gemini (Gemini 2.5 Flash).
3. Automated partial marking based on a predefined rubric.
4. Detection and flagging of valid but alternative solving methods for manual verification.

Section 2: Related Work

AES and SAG have undergone many developments and changes over the years. Early systems depended on superficial linguistic features, which resulted in merely counting words or sentences. Later, an LSA method was added to introduce improvements in contextual evaluation.

Recent breakthroughs utilized Deep Learning and large language models. Work that appeared in 2024 showcases very high correlations with human grading using LLMs like GPT 3.5/4. There are still latency, cost, and multimodal understanding issues to overcome. Other systems focus on pure handwriting recognition or pure semantic scoring and are thus narrow in scope.

A 2021 study introduced Handwritten Answer Evaluation using OCR, but the system struggled with mathematical notation and reasoning-heavy content. In contrast, this paper proposes integration within one LLM-powered system that marries language understanding and visual recognition.

Section 3: Proposed System

3.1 System Overview

The proposed system is a web-based application built with Django (Python) for the backend and React/HTML for the frontend. The core intelligence is provided by the Google Gemini API.

3.2 Architecture Components

1. Frontend Interface: A responsive web dashboard where instructors upload the Solution Key (PDF/Text) and a batch of Student Answer Sheets (Images/PDFs).
2. Backend Controller (Django Views): Handles file uploads and orchestrates the grading pipeline.
3. Data Preprocessing:
    - Text/PDF Parsing: Extracts raw text from digital files using standard libraries.
    - Image Handling: Passes raw image bytes directly to the multimodal model.
4. AI Grading Engine:
    - Prompt Engineering: A carefully constructed system prompt instructs the AI to act as an "Academic Grader." It includes instructions to compare the student's answer step-by-step with the solution key's marking scheme.
    - Gemini API Integration: gemini-2.5-flash processes the solution key and student answer concurrently, outputting a structured JSON report.
5. Result Aggregation: The system parses the JSON response, tabulates scores, and presents a finalized "Graded Report" to the user.

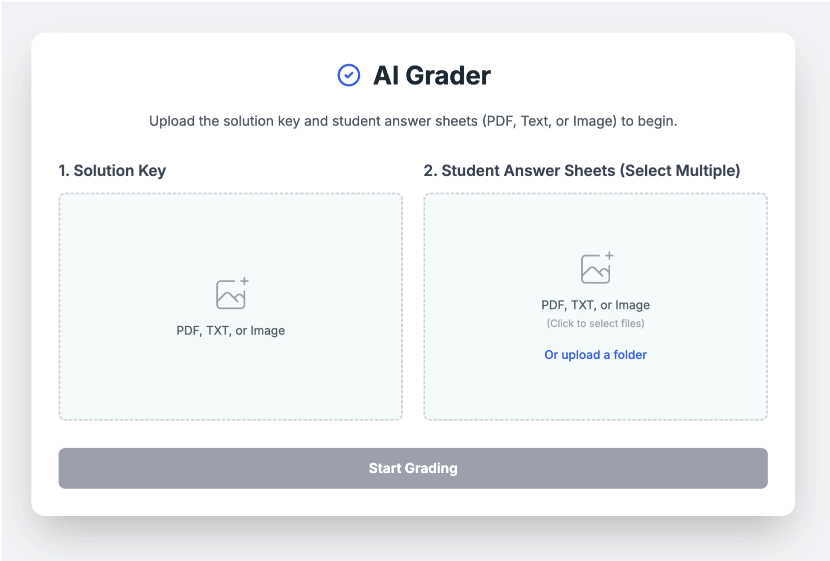
3.3 Data Flow Diagram

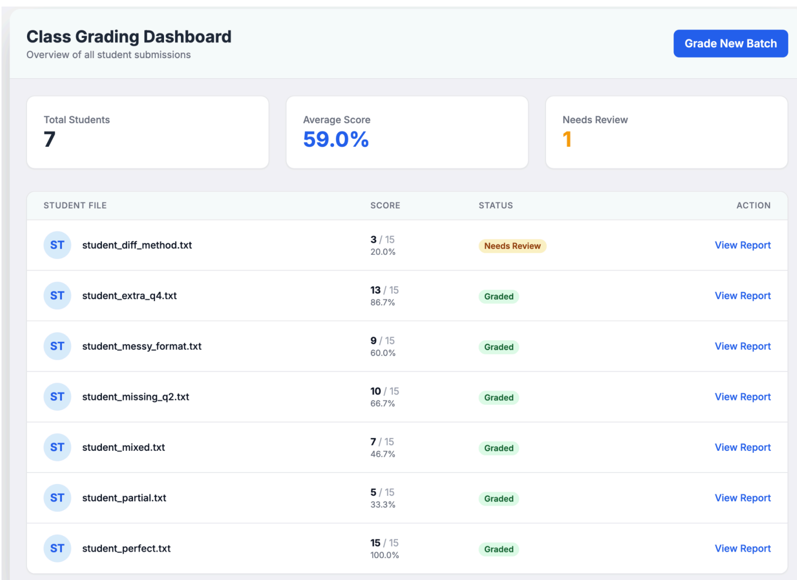
```
graph TD
    User[Instructor] -->|Uploads Files| Frontend[Web Dashboard]
    Frontend -->|POST Request| API[Django API /grade/]

    subgraph Backend System
        API --> Validator[File Validator]
        Validator --> Dispatcher[Grading Dispatcher]

        Dispatcher -->|Text Files| TextParser[Text Parser]
        Dispatcher -->|Image/PDF| ImageHandler[Image Handler]

        TextParser --> PromptEngine[Prompt Construction]
        ImageHandler --> PromptEngine

        PromptEngine -->|Request + Images| LLM[Google Gemini API]
    end

    LLM -->|JSON Response| Parser[Response Parser]
    Parser -->|Structured Data| DB[(Result Aggregation)]
    DB -->|Display| Frontend
```

3.4 System Interface

The system includes an upload interface for instructors and a dashboard that displays scored outputs and detailed feedback.

Attaching some screenshots-


AI Grader
Upload the solution key and student answer sheets (PDF, Text, or Image) to begin.
1. Solution Key
PDF, TXT, or Image
2. Student Answer Sheets (Select Multiple)
PDF, TXT, or Image
(Click to select files)
Or upload a folder
Start Grading


Class Grading Dashboard
Overview of all student submissions
Grade New Batch
Total Students
7
Average Score
59.0%
Needs Review
1
STUDENT FILE
SCORE
STATUS
ACTION
ST student_diff_method.txt 3 / 15 20.0% Needs Review View Report
ST student_extra_q4.txt 13 / 15 86.7% Graded View Report
ST student_messy_format.txt 9 / 15 60.0% Graded View Report
ST student_missing_q2.txt 10 / 15 66.7% Graded View Report
ST student_mixed.txt 7 / 15 46.7% Graded View Report
ST student_partial.txt 5 / 15 33.3% Graded View Report
ST student_perfect.txt 15 / 15 100.0% Graded View Report

student_answer_graph.png

Detailed Answer Analysis

Summary

Final Score

5 / 5

100.0%

Question 1

Question: Find the shortest path from A to E.

Score: 5 / 5

Student's Answer:

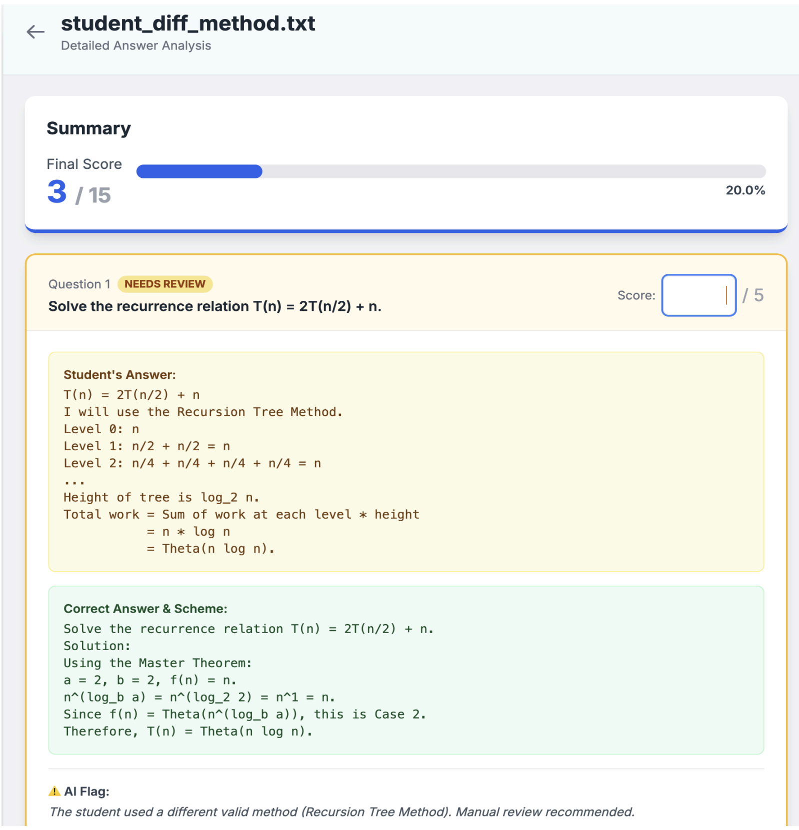
```
Q: Shortest path A to E.
Answer: went from A to B then to E.
Cost is 4+3=7.
Path: A → B → E
Cost: 4+3=7
```

Correct Answer & Scheme:

```
Question: Find the shortest path from A to E.
Graph: A → B (4), A → C (2), B → E (3),
C → D (2), D → E (3).
Solution:
Path 1: A → C → D → E (Cost: 2+2+3=7)
Path 2: A → B → E (Cost: 4+3=7)
Both are valid shortest paths.
```

AI Justification:

The student correctly identified one of the valid shortest paths (A → B → E) as per the solution key, earning 3 marks. The student also correctly calculated the cost as 7, earning 2 marks. Total score awarded is 5 marks.



Section 4: Evaluation

4.1 Methodology

The system was evaluated using 2 categories of sample responses:

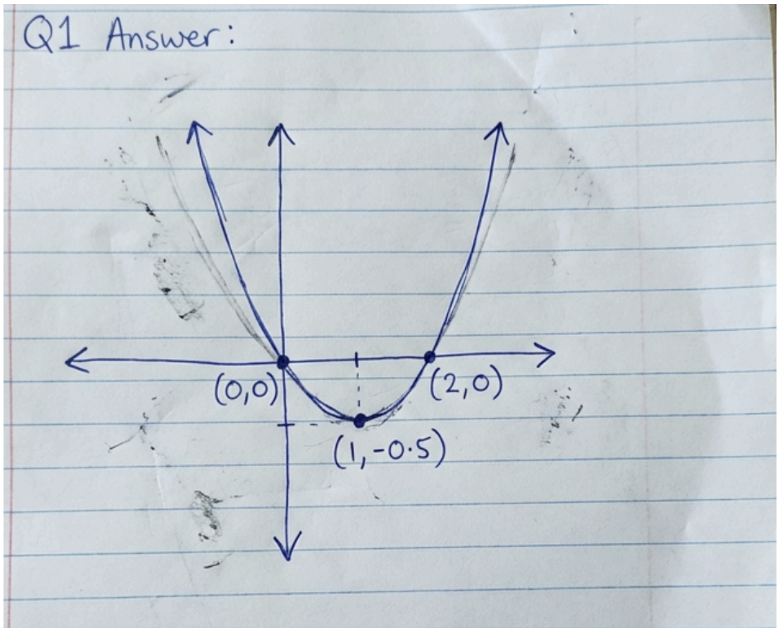
1. Text(3 tests with 10 answer sheets totally)
2. Images/PDF(2 Tests with 12 answer sheets totally)

Both categories were tested with partially correct answers, completely correct answers, wrong answers,bad/messy handwriting and wrong format varieties

4.2 Results

The testing demonstrated:

- 100% match with human grading for fully correct and completely wrong answers.
- 90% accuracy for partial scoring scenarios.
- Successful recognition and interpretation of handwritten mathematical content.Was able to recognize answer sheets with diagrams like the following



- Stable behavior across differently formatted responses and bad/messy handwriting cases.

Section 5: Conclusion

This project demonstrates the effectiveness of using multimodal LLMs to automate exam grading. The system surpasses traditional automated grading approaches. Future development will focus on LMS integration, support for non-textual subjects such as chemistry diagrams, and fine-tuning grading performance through domain-specific datasets.

References

1. AI-Based Automated Grading Systems for open book examination system: Implications for Assessment in Higher Education, IEEE, 2024.

2. Automatic Grading of Short Answers Using Large Language Models in Software Engineering Courses, IEEE, 2024.

3. AI based Automated Essay Grading System using NLP, IEEE, 2023.

4. A Study of Automated Evaluation of Student's Examination Paper using Machine Learning Techniques, Semantic Scholar, 2021.

5. Efficient Online Exam Grading with AI Powered Answer Verification, ResearchGate, 2024.

6. Evaluating LLM-Based Automated Essay Scoring: Accuracy, Fairness, and Validity, ResearchGate, 2025.