DEVELOPING AN AI-ASSISTED GRADING SYSTEM USING LARGE LANGUAGE MODELS (GPT-4)

by

Andrei Modiga

A PROJECT PROPOSAL

Presented to the Faculty of

The School of Computing at the Southern Adventist University

In Partial Fulfilment of Requirements

For the Degree of Master of Science

Major: Computer Science

Under the Supervision of Scot Anderson, Ph.D.

Collegedale, Tennessee

August, 2024

DEVELOPING AN AI-ASSISTED GRADING SYSTEM USING LARGE LANGUAGE MODELS (GPT-4)

Andrei Modiga, M.S.

Southern Adventist University, 2024

Adviser: Scot Anderson, Ph.D.

Grading written student work is a critical component of education, directly influencing the learning experience and providing essential feedback to both students and educators. This project focuses on building an AI-assisted grading system that leverages Large Language Models (LLMs), specifically GPT-4, to automate and enhance the grading process. By classifying student responses based on semantic similarity, the system streamlines grading and allows educators to provide more consistent and efficient feedback. The project aims to reduce the manual workload on teachers, improve grading consistency, and provide timely feedback to students.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the educational landscape, grading written student work is a task of high importance; it directly influences the learning experience and provides crucial feedback to both students and educators. Traditional grading methods, while effective, can be time-consuming and labor-intensive, especially for teachers managing large volumes of student submissions. The demand for efficient and scalable grading solutions has become increasingly evident as educators seek to streamline their workloads without compromising assessment quality.

Recent advancements in artificial intelligence, particularly in the development of Large Language Models (LLMs), offer a promising approach to this challenge. LLMs like OpenAI's GPT-4 have demonstrated remarkable proficiency in understanding and generating human-like text by analyzing vast datasets and recognizing patterns in language. These models have the potential to transform the grading process by automating the evaluation of student responses, thereby reducing the workload on teachers.

This project explores the development of an AI-assisted grading system that incorporates GPT-4 to classify student responses by grouping those that are semantically similar. By identifying clusters of answers that convey the same meaning, regardless of phrasing or structure, the system can assist educators in grading more efficiently.

The primary goal of this project is to build and implement this AI-assisted grading system, including the ability to recognize handwritten text, and integrate it into the educational platform, OICLearning.com. By focusing on building a practical solution, the project aims to enhance the grading process, making it quicker and more consistent, thereby allowing teachers to focus on more critical aspects of instruction while ensuring that students receive fair and accurate feedback.

However, integrating LLMs into the grading process presents challenges that must be carefully addressed. Key considerations include the model's ability to accurately differentiate nuanced meanings and appropriately group responses that, while semantically similar, may vary in correctness. Additionally, the effectiveness of these models in handling a diverse range of student expressions and potential errors needs thorough evaluation.

Chapter 2

Background

The integration of artificial intelligence (AI) into education has revolutionized teaching and learning processes. A significant area of impact is the automation of grading, where AI technologies, particularly Large Language Models (LLMs) like GPT-4, have shown great promise. This chapter delves into the advancements in AI-assisted grading, focusing on GPT-4's role, and discusses existing research, challenges, and the potential benefits of such systems.

2.1 Artificial Intelligence in Education

Artificial intelligence has become a significant part of modern education, offering tools and solutions that enhance learning experiences and streamline administrative tasks. AI-driven educational platforms can personalize learning paths, adapt content to individual student needs, and provide real-time feedback [**?**]. This personalization is crucial in accommodating diverse learning styles and improving student engagement.

Moreover, AI systems can analyze vast amounts of data to identify patterns and trends in student performance. Teachers can use these insights to tailor instruction, address knowledge gaps, and improve overall educational outcomes. The automation of administrative tasks, such as scheduling and attendance tracking, allows teachers to focus more on instruction and less on paperwork [?].

2.1.1 Personalized Learning and Feedback

One of the most significant contributions of AI in education is the ability to offer personalized learning experiences. Intelligent tutoring systems powered by AI can adjust the difficulty level of exercises based on a student's performance, ensuring optimal learning progression [?]. These systems can also identify areas where a student is struggling and provide targeted interventions.

In the context of grading, AI can provide detailed feedback on assignments, highlighting specific areas of strength and weakness. GPT-4, for instance, can generate personalized comments that help students understand their mistakes and learn from them. This immediate and individualized feedback is more effective than generic responses and can significantly enhance the learning process [?].

2.1.2 Administrative Efficiency

AI's ability to automate routine administrative tasks is transforming the educational landscape. Tasks such as grading, scheduling, and record-keeping can be handled efficiently by AI systems, reducing the workload on educators [?]. This automation not only saves time but also minimizes human errors that can occur in manual processes.

Furthermore, AI systems can monitor student engagement and participation, alerting educators to potential issues such as decreased performance or absenteeism. By identifying these problems early, interventions can be implemented promptly, supporting student retention and success [?].

2.2 Advancements in Large Language Models

2.2.1 Introduction to GPT-4

GPT-4, developed by OpenAI, is one of the most advanced LLMs available today. Building upon the success of its predecessors, GPT-4 has been trained on an extensive dataset comprising diverse textual content, enabling it to understand and generate human-like text with remarkable fluency and coherence [?].

The architecture of GPT-4 allows it to capture complex language patterns, understand context, and infer meaning beyond surface-level interpretations. This deep understanding is particularly beneficial in educational applications where more than simple comprehension of student responses is required.

2.2.2 GPT-4 in Educational Applications

GPT-4's capabilities extend to various educational applications. It can assist in content creation, such as generating lesson plans, educational materials, and practice questions tailored to specific learning objectives [?]. In language learning, GPT-4 can provide conversational practice, corrections, and explanations to learners.

In grading, GPT-4's ability to understand and evaluate open-ended responses makes it a valuable tool. It can analyze essays, short answers, and problem-solving explanations, providing not only grades but also constructive feedback. This level of assessment requires a deep understanding of the subject matter and the ability to interpret diverse student expressions [?].

2.3 AI-Assisted Grading Using GPT-4

2.3.1 Automating Short Answer Grading

Short answer questions are challenging to grade automatically due to the variability in student responses. Traditional grading systems rely on keyword matching or predefined answer patterns, which can miss correct answers phrased differently or accept incorrect answers that contain the right keywords.

GPT-4 overcomes these limitations by understanding the semantic meaning behind the text. It can recognize correct answers even when they are expressed in unconventional ways and identify incorrect answers that superficially appear correct [**?**]. This semantic understanding allows for more accurate grading of open-ended questions.

Liu et al. (2023) [**?**] conducted a study demonstrating GPT-4's effectiveness in grading university-level mathematics exams. The AI model was able to evaluate complex mathematical reasoning and provide grades that closely aligned with human graders. This study highlights GPT-4's potential in handling subjects that require critical thinking and problem-solving skills.

2.3.2 Grading Handwritten Responses

Grading handwritten assignments poses additional challenges due to the need for accurate handwriting recognition. Optical Character Recognition (OCR) technologies have improved, but they still struggle with illegible handwriting or complex symbols commonly used in subjects like mathematics and physics.

Kortemeyer (2023) [**?**] explored the feasibility of using AI to grade handwritten physics solutions. By integrating OCR technologies with GPT-4, the study achieved a significant correlation between AI-generated grades and human grading. However, the study also noted limitations in accurately interpreting diagrams and notations, indicating areas where further improvement is needed.

2.3.3 Semantic Understanding and Contextual Evaluation

GPT-4's advanced natural language processing capabilities enable it to understand context and semantics deeply. It can evaluate not just the correctness of an answer but also the reasoning process behind it. This is important in subjects where the method is as important as the final answer, such as mathematics and science.

For example, GPT-4 can assess a student's problem-solving approach, identify logical errors, and provide specific feedback on where the reasoning went astray [**?**]. This level of detailed evaluation helps students understand their mistakes and learn more effectively.

2.4 Challenges in AI-Assisted Grading with GPT-4

2.4.1 Accuracy and Reliability

While GPT-4 demonstrates high proficiency in understanding and evaluating text, it is not immune to errors. Misinterpretations can occur, especially with ambiguous or poorly constructed responses. Liu et al. (2023) [**?**] emphasized the importance of human oversight to ensure the reliability of AI-generated grades.

Moreover, AI models may sometimes be overconfident in their assessments, potentially leading to incorrect grading. Implementing mechanisms for uncertainty estimation and flagging ambiguous cases for human review can mitigate this issue.

2.4.2 Bias and Fairness

AI models can inadvertently exhibit biases present in their training data. This raises concerns about fairness, as certain groups of students might be disadvantaged by biased grading. Tossell et al. (2023) [?] highlighted the importance of addressing these biases to ensure equitable treatment of all students.

Strategies to mitigate bias include using diverse and representative training data, implementing fairness-aware algorithms, and regularly auditing AI systems for discriminatory patterns [?].

2.4.3 Handwriting Recognition Limitations

Despite advancements in OCR technologies, accurately recognizing handwritten text remains challenging. Variations in handwriting styles, the use of non-standard symbols, and poor scan quality can hinder recognition accuracy.

Kortemeyer et al. (2023) [?] found that AI struggled with interpreting hand-drawn diagrams and complex notations, which are common in STEM subjects. Improving OCR algorithms and combining them with context-aware models like GPT-4 can enhance performance, but human intervention may still be necessary in some cases.

2.4.4 Ethical and Privacy Concerns

The use of AI in grading involves processing sensitive student data, raising ethical and privacy considerations. Compliance with regulations such as the Family Educational Rights and Privacy Act (FERPA) [?] is essential.

Alto (2023) [?] emphasizes the need for robust data protection measures, transparency in AI operations, and obtaining informed consent from users. Ensuring that AI systems are secure and that data is handled responsibly is crucial for maintaining trust.

2.5 Existing Research on AI Grading Systems

2.5.1 Machine Learning Approaches

Before the advent of LLMs like GPT-4, machine learning approaches to grading primarily involved training models on labeled datasets to recognize correct answers. Weegar and Idestam-Almquist (2022) [?] explored such methods for grading short answers in computer science exams.

Their research demonstrated that machine learning could significantly reduce grading workload by clustering similar answers and automating scoring. However, these systems often required extensive preprocessing and were limited in handling the full variability of human language.

2.5.2 Feasibility Studies with GPT-4

Studies utilizing GPT-4 have shown promising results in various subjects. Kortemeyer (2023) [?] conducted a feasibility study on using GPT-4 for grading physics problems. The AI's grades correlated highly with human graders, indicating its potential effectiveness.

However, the study also noted that GPT-4 sometimes missed nuanced aspects of the solutions, particularly in assessing the quality of reasoning and the appropriateness of assumptions. This suggests that while GPT-4 can assist in grading, it should complement rather than replace human judgment.

2.5.3 Student Perceptions of AI Grading

The acceptance of AI-assisted grading by students is crucial for its successful implementation. Tossell et al. (2023) [?] examined student perceptions of using AI tools like ChatGPT in academic settings.

While students appreciated the promptness and consistency of AI feedback, they expressed concerns about the transparency of the grading process and the AI's ability to understand their unique perspectives. Building trust requires clear communication about how the AI operates and opportunities for students to contest or discuss their grades [?].

2.6 Potential Benefits of GPT-4 in Grading

2.6.1 Efficiency and Scalability

Implementing GPT-4 in grading systems can significantly reduce the time educators spend on evaluating assignments. This is especially beneficial in large classes where manual grading is impractical. AI systems can handle vast amounts of data quickly, providing timely feedback to students [?].

2.6.2 Consistency and Fairness

AI grading systems apply the same criteria uniformly, reducing inconsistencies that can arise from human graders' subjective judgments or fatigue. This consistency contributes to fairness, ensuring that all students are evaluated on the same basis [?].

2.6.3 Timely Feedback

Prompt feedback is essential for effective learning. AI systems can provide immediate evaluations, allowing students to understand their performance and address any issues while the material is still fresh in their minds [**?**].

2.6.4 Enhanced Learning Outcomes

By analyzing patterns in student responses, AI systems can identify common misconceptions and areas where many students struggle. Educators can use this information to adjust their teaching strategies, focus on problematic topics, and improve overall learning outcomes [**?**].

2.7 Summary

The integration of GPT-4 into grading systems offers significant advantages, including increased efficiency, consistency, and the ability to provide personalized feedback. While challenges such as accuracy, bias, and ethical considerations exist, ongoing research and development are addressing these issues.

By combining GPT-4's capabilities with human oversight and ethical practices, AI-assisted grading can enhance educational experiences for both students and educators. The potential for improved learning outcomes and reduced workloads makes this an important area of development.

2.8 Conclusion

The existing body of research highlights the transformative potential of AI-assisted grading using GPT-4. By understanding the capabilities and limitations of these systems, educators can implement AI tools that complement their teaching practices.

The proposed AI-assisted grading system aims to build upon this foundation, addressing current challenges and harnessing GPT-4's full potential. By doing so, it seeks to improve the grading process, enhance learning experiences, and contribute to the advancement of AI in education.

Chapter 3

Proposal

This chapter outlines the key requirements for developing an AI-assisted grading system using Large Language Models (LLMs) like GPT-4 and GPT-4 Vision, implemented in Python. The goal is to automate the grading process for various types of assignments on OICLearning.com by utilizing GPT-4 Vision to read images directly, performing both Optical Character Recognition (OCR) and Natural Language Processing (NLP) in a unified model. Additionally, the system will incorporate grading algorithms to assign grades based on expected answers for each question. This integration simplifies the workflow and enhances the system's ability to accurately interpret and grade student submissions.

3.1 Summary of Requirements

The AI-assisted grading system will require the development of both backend and frontend components. The backend will handle the processing of student submissions, including images and text, and communicate with GPT-4 and GPT-4 Vision via API calls. The frontend will provide an interface for instructors to review, sort, and manage graded assignments.

Key requirements include:

- **Backend API Development:** Implement a backend API in Python that can receive images and text submissions from students and send them to GPT-4 and GPT-4 Vision for processing.
- **Frontend Interface:** Create a user-friendly frontend interface that allows instructors to view, sort, and manage the process of grading assignments.
- **Data Security and Privacy:** Ensure the system complies with data protection regulations, implementing necessary security measures.
- **Scalability:** Design the system to handle a large number of student submissions and various types of assignments.
- **User Authentication:** Integrate with the secure login and access control mechanisms avaialable in the OICLearning.com project.

3.2 Proposed Approach

The proposed AI-assisted grading system will be implemented using a modular architecture, separating the backend processing from the frontend user interface. The backend will be developed in Python and will utilize the OpenAI API to interact with GPT-4 and GPT-4 Vision. The system will process images submitted by students, perform OCR to extract text, and then use NLP to interpret the content.

3.2.1 Backend Implementation

The backend API will handle the following tasks:

- Receive images and text submissions from the frontend.
- Send images to GPT-4 Vision via API calls for OCR and NLP processing.
- Parse and handle responses from GPT-4.
- Compare the interpreted responses to predefined expected answers.
- Store the results in a database.

3.2.2 Frontend Development

The frontend will be a web-based application that allows users to:

- Securely log in to the system.
- Upload and submit assignments.
- View graded assignments and student responses.
- Sort and filter submissions based on various criteria (e.g., grade, date submitted).
- Calculate grades based on a grading rubric or algorithm.
- Override grades or make manual adjustments.

3.2.3 System Integration

By leveraging GPT-4 Vision's capabilities, the system simplifies the grading workflow by handling both OCR and NLP tasks within a single model, reducing complexity and potential points of failure. The modular design allows for scalability and ease of maintenance.

3.3 Detailed Requirements and User Stories

Four user stories describe the processes that drive the requirements. These stories include: the teacher creating an assignment, teachers or students uploading submissions, the AI assisted grading of assignments and finally the teacher transfering the grades to an Learning Management system (LMS). These stories and their associated requirements describe a series of webpages shown in Figure 3.1 The following sections detail each of these stories.

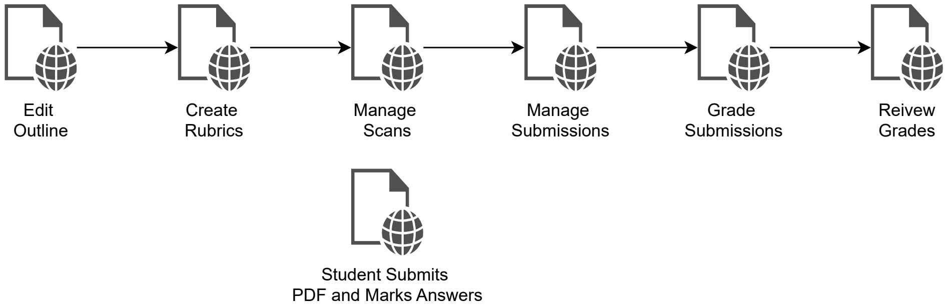


Figure 3.1: Page flow diagram showing the sequence of pages in the system.

3.3.1 A teacher creates an assignment

When a teacher creates an assignment, they choose from different kinds of AI assisted types. These types include:

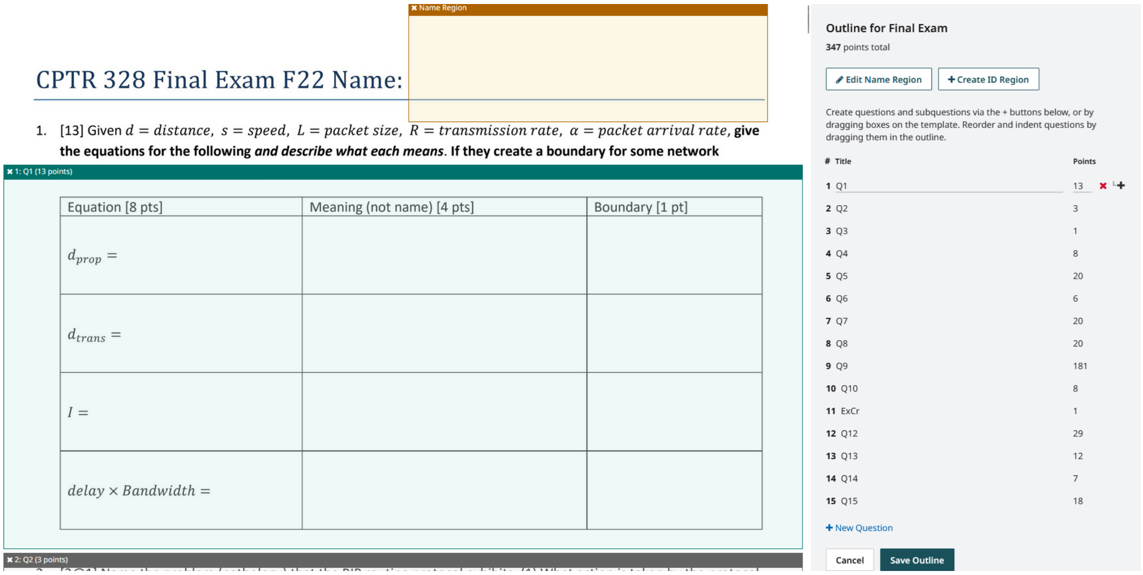
- Filled-Form that students write on and are later scanned in by the teacher. E.g. a written test.

- Free-Form typed or written assignments where the location of answers may vary. E.g. an assignment that is given to the students in a word document and the length of their answer may vary, and hence their location may vary.

- Online Filled-form assignments. E.g. a quiz that is taken online where answers that contains digital responses from students. This is not within the scope of this project.

For Filled-Form type assignments, the teacher needs to prompt the system by uploading a blank of the form. Figure 3.2 shows an example of what this page might look like. The teacher first identifies the location of the student identity (name and/or ID number). In the Figure, this is shown by the tan box added by the teacher around the area where a student is instructed to place their name. To add a question, the teacher clicks on the "New Question" link at the bottom right. The system adds a new question to the list and an associated box to go with it. Then the teach must: name the question, give the points , locate and size the associated box (shown in green) of the answer area over the blank spot they expect students to write in, and give the expected answer (not shown). Figure 3.2 shows the name area and a single question where the teacher can accomplish all these things. Here when a teacher clicks on a question in the list, they can edit the box size/location along with the name or points. The plus sign next to the red X, allows the teacher to add a sub-question. Sub-questions work the same way as questions except they appear as a sublist under the question.

This process is called editing the outline and includes everything the system needs to define the assignment. The teacher can then save the outline and the



CPTR 328 Final Exam F22 Name:

1. [13] Given $d = distance,\ s = speed,\ L = packet\ size,\ R = transmission\ rate,\ \alpha = packet\ arrival\ rate$, **give the equations for the following *and describe what each means*. If they create a boundary for some network**

Equation [8 pts]	Meaning (not name) [4 pts]	Boundary [1 pt]
$d_{prop} =$		
$d_{trans} =$		
$I =$		
$delay \times Bandwidth =$		

Outline for Final Exam

347 points total

Create questions and subquestions via the + buttons below, or by dragging boxes on the template. Reorder and indent questions by dragging them in the outline.

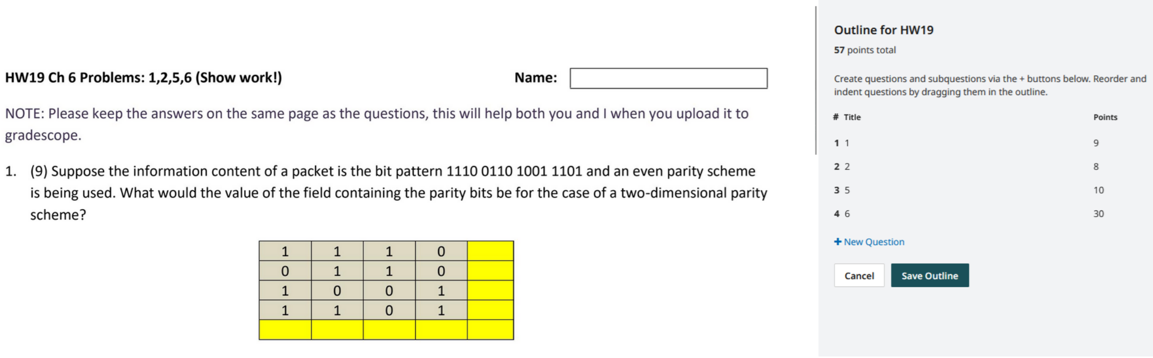
#	Title	Points
1	Q1	13
2	Q2	3
3	Q3	1
4	Q4	8
5	Q5	20
6	Q6	6
7	Q7	20
8	Q8	20
9	Q9	181
10	Q10	8
11	ExCr	1
12	Q12	29
13	Q13	12
14	Q14	7
15	Q15	18

Figure 3.2: The filled-form assignment creation page.

system will be ready to accept student submissions. The teacher can create rubrics at this point or, more likely, defer this step grading. If the teacher decides to create the rubrics now, they progress to the webpage called “Create Rubrics” (see Figure 3.1). Otherwise, they can add rubrics during grading.

A rubric defines a value subtracted from a score when applied, and a description of the reason for the deduction. The teacher can add up to 36 rubrics and they shall be numbered 0-9, and a-z. Teachers assign one or more rubrics to an answer (or group of answers when grading groups) during grading.

In the case of Free-formed typed or written assignments, the teacher gives the students questions, and the students turn in their answers as a PDF. Students may or may not include the questions, so the answer’s location will vary from student to student. Here the teacher does not need to identify the location of an answer, but they do need to provide question identifiers and points for students to use when turning in the document. Figure 3.3 shows an example of the teacher creating a Free-Form assignment. Notice that there is no need to draw a rectangle around the answer area and except for that task the teacher's work is identical to the filled-form assignment above.



HW19 Ch 6 Problems: 1,2,5,6 (Show work!) **Name:**

NOTE: Please keep the answers on the same page as the questions, this will help both you and I when you upload it to gradescope.

1. (9) Suppose the information content of a packet is the bit pattern 1110 0110 1001 1101 and an even parity scheme is being used. What would the value of the field containing the parity bits be for the case of a two-dimensional parity scheme?

1	1	1	0	
0	1	1	0	
1	0	0	1	
1	1	0	1	

Outline for HW19

57 points total

Create questions and subquestions via the + buttons below. Reorder and indent questions by dragging them in the outline.

#	Title	Points
1	1	9
2	2	8
3	5	10
4	6	30

Figure 3.3: The free-form assignment creation page.

3.3.2 Preparing Submissions to be graded

The two tasks at this stage are uploading submissions and assigning names to them. This is done in two steps. First, the assignments are uploaded by the teacher or by the individual student. Second, the teacher approves/assigns names to the assignments. This takes two screeens Manage Scans and Manage Submissions, or one screen when uploaded by the student (refer to the pages list in Figure 3.1).

Manage scans: After a teacher has logged in and gone to a filled-in assignment, they upload all the assignments as a single PDF. The system already knows how many pages are in each assignment because it identified that number in the creation process above. The system separates the PDF into individual documents and attempts to assign the names of students based on the identity. This is accomplished on a screen called Manage. However, this page does not show the individual assignments or the names it has assigned. It only shows the number of assignments it assigned [1].

Manage Submissions: In this page, a teacher can view a portion of each submission that contains the identity of the student. The system knows the students in the class and populates drop-down selection boxes for the names of students that it could not identify. The teacher can then select the correct student name from the drop-down. Of course, this is only necessary for Filled-form assignments uploaded by the teacher. Free-form assignments that students upload are automatically assigned to the student that uploaded them. Consequently, Free-form assignments will not have this page in the list of pages teachers see to manage the assignment.

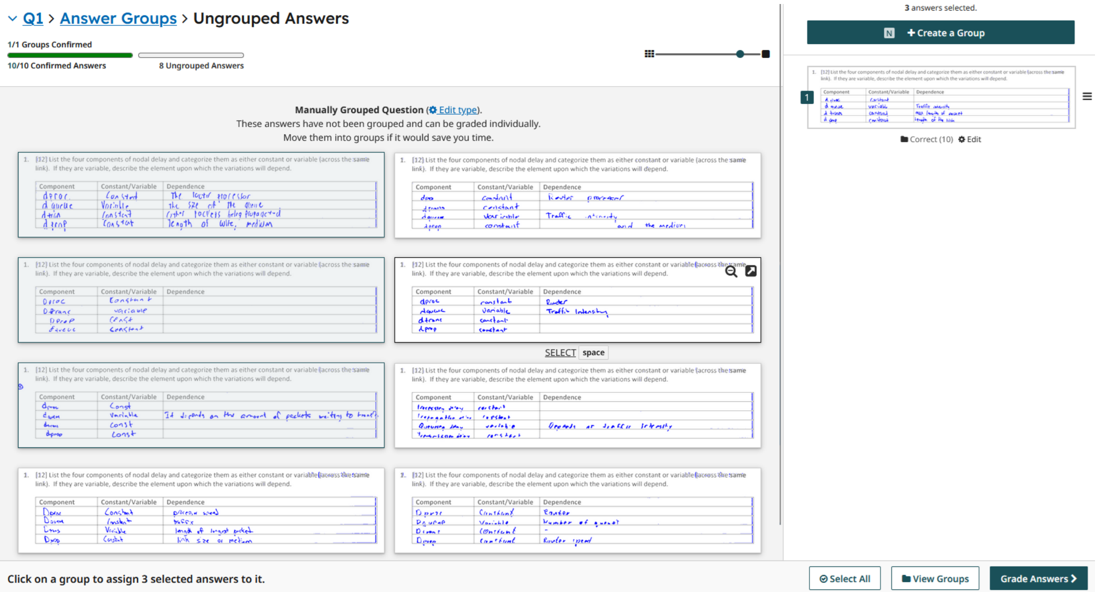
3.3.3 Grading Submissions

The process starts with a list of questions for an assignment that needs grading. Each question is a link to a page where grouping can occur. The AI assists in this process by default, but the teacher can override the AI's grouping, or stop the process and manually group answers.

In the AI assisted process, the screen shows the list of groups on the right with the main screen showing the student answers as images in a grid. Figure 3.4 shows an example with eight answers. The first three in the left column have been selected by clicking on the images, and at this point would form a new group if the user clicks on the "+Create a Group" button. The AI should group the results first, but may leave some in an ungrouped "group." That is, single submissions that don't fit into a group need not be put in a separate group, but should be kept in an "ungrouped" group. *All* groups should be listed at the right. When clicking on the group, the system should de-select any images currently selected and then show

[1] In other grading systems, PDFs are converted to images. Although we would like to deal with PDFs natively, this is an acceptable solution

the newly selected group's images in the main page. When images are selected and the "Create a Group" button is clicked, the system should create the group and move the images from the current group to the new group, then display the new group and it's images.



Manually Grouped Question (Edit type).
These answers have not been grouped and can be graded individually.
Move them into groups if it would save you time.

1. [12] List the four components of nodal delay and categorize them as either constant or variable (across the same link). If they are variable, describe the element upon which the variations will depend.

Component	Constant/Variable	Dependence

Click on a group to assign 3 selected answers to it.

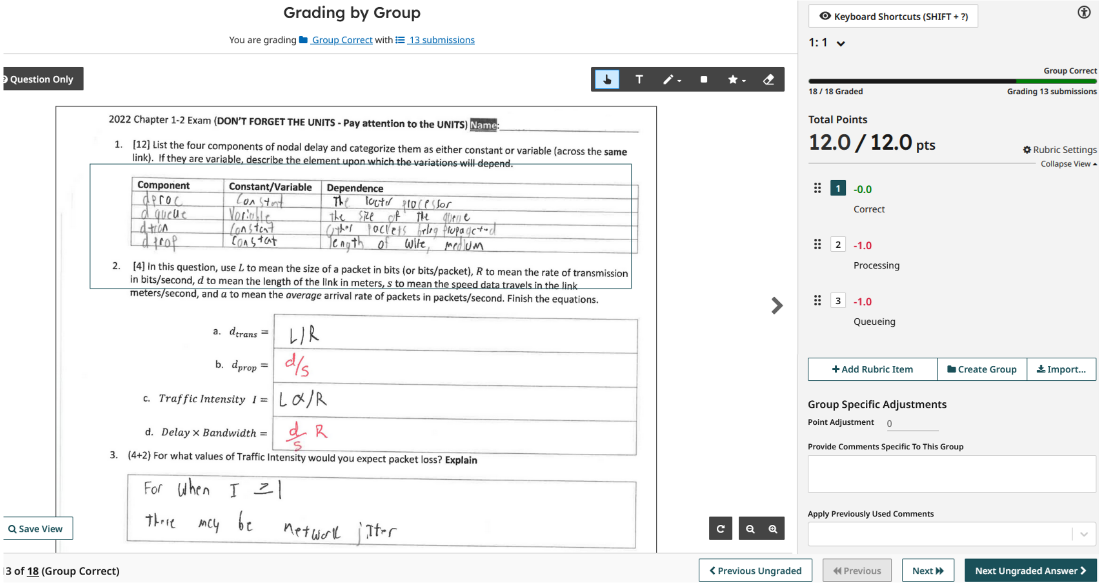
Figure 3.4: Grouping answers.

Selecting a group shows the group's Information. This information includes the images in the group (in the main screen), the name of the group, the number of images in the group, and a description of the group. All of these are assigned by the AI and the teacher may edit these items as appropriate.

Before moving to the next step it is important that the teacher review all the groups and that they are satisfied that the groups are correct. If the teacher decides that the grouping is incorrect, they must come back to this page to regroup them.

Once the system has groups created, the teach can proceed to grade the answers. The teacher does this by clicking on the Grade Answers button at the lower right in Figure 3.4. Groups are graded in order that they appear on the screen with the sole exception that no matter where the ungrouped group is, it is always graded last.

When grading a group, the teacher is shown the group name, the number of images in the group, a sample answer(s) from the group, and the description of the group. The teacher can then assign rubrics that subtract points and provide feedback to students. The description of the group is part of the feedback that the student will see. Figure 3.5 shows and example of what the grading page might look like. Several important requirements come into play for ease of use, specifically key presses and mouse movements have the following meaning:



Grading by Group

You are grading Group Correct with 13 submissions

2022 Chapter 1-2 Exam (**DON'T FORGET THE UNITS - Pay attention to the UNITS**) Name:

1. [12] List the four components of nodal delay and categorize them as either constant or variable (across the **same** link). If they are variable, describe the element upon which the variations will depend.

Component	Constant/Variable	Dependence
dproc	Constant	The router processor
d queue	Variable	the size of the queue
dtran	Constant	Other packets being propagated
dprop	Constat	length of wire, medium

2. [4] In this question, use L to mean the size of a packet in bits (or bits/packet), R to mean the rate of transmission in bits/second, d to mean the length of the link in meters, s to mean the speed data travels in the link meters/second, and a to mean the *average* arrival rate of packets in packets/second. Finish the equations.

a. $d_{trans} =$ L/R

b. $d_{prop} =$ d/s

c. $Traffic\ Intensity\ \ I =$ L α/R

d. $Delay \times Bandwidth =$ $\frac{d}{s}$ R

3. (4+2) For what values of Traffic Intensity would you expect packet loss? **Explain**

For when I ≥ 1
there may be network jitter

3 of 18 (Group Correct)

Keyboard Shortcuts (SHIFT + ?)

1: 1

Group Correct

18 / 18 Graded

Grading 13 submissions

Total Points

12.0 / 12.0 pts

1	-0.0	Correct
2	-1.0	Processing
3	-1.0	Queueing

Group Specific Adjustments

Point Adjustment 0

Provide Comments Specific To This Group

Apply Previously Used Comments

Figure 3.5: The grading page.

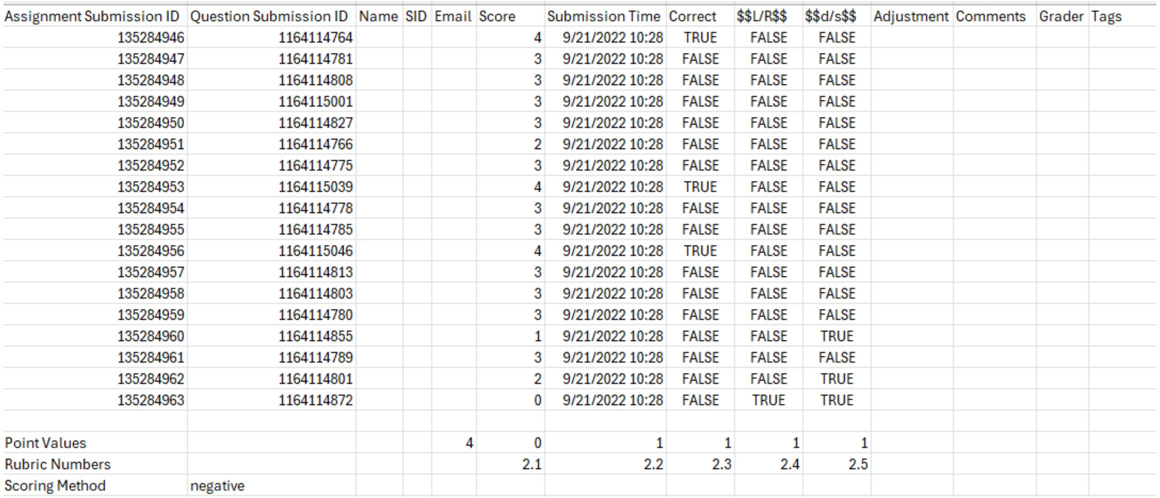
- Pressing 0–9 or some key combination with a–z toggles the rubric as assigned or not-assigned to the group.
- Left arrow takes the teacher to the previous group or individual submission.
- Right arrow takes the teacher to the next group or individual submission.

- Scrolling up or down over the rubrics scrolls through the available rubrics.
- Scrolling on the image zooms in and out.
- Clicking and dragging the image moves it around in the viewing area.

Suppose a teacher decides to change the value of a rubric. They can simple edit the rubric while grading and all grades where this rubric applies are updated with the new value.

When the teacher completes all of the groups and each individual assignment in the "ungroupped" group, they can move on to the next question in the assignment. When they are finished with all the questions, they can publish the results so that students can see them. This "Review Grades" page allows the teacher to see a list of all the grades has buttons to do the following actions:

- Publish or Unpublish grades (Students see grades or not)
- Download Grades (export grades to an Excel or CSV file)
- Export submissions (to a zip file). Submissions are in PDF format and include the following:
  - Name and Identifying information of the student along with the assignment score.
  - A list of question names with the points assigned and the rubrics applied to that question.
  - After the grade information, The students original work is shown.
- Export Evaluations (to a zip file of CSV files see Figure 3.6 for an example of a single question evaluation).



Assignment Submission ID	Question Submission ID	Name	SID	Email	Score	Submission Time	Correct	$$L/R$$	$$d/s$$	Adjustment	Comments	Grader	Tags
135284946	1164114764				4	9/21/2022 10:28	TRUE	FALSE	FALSE				
135284947	1164114781				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284948	1164114808				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284949	1164115001				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284950	1164114827				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284951	1164114766				2	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284952	1164114775				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284953	1164115039				4	9/21/2022 10:28	TRUE	FALSE	FALSE				
135284954	1164114778				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284955	1164114785				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284956	1164115046				4	9/21/2022 10:28	TRUE	FALSE	FALSE				
135284957	1164114813				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284958	1164114803				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284959	1164114780				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284960	1164114855				1	9/21/2022 10:28	FALSE	FALSE	TRUE				
135284961	1164114789				3	9/21/2022 10:28	FALSE	FALSE	FALSE				
135284962	1164114801				2	9/21/2022 10:28	FALSE	FALSE	TRUE				
135284963	1164114872				0	9/21/2022 10:28	FALSE	TRUE	TRUE				
Point Values				4	0	1	1	1	1				
Rubric Numbers					2.1	2.2	2.3	2.4	2.5				
Scoring Method	negative												

Figure 3.6: Comma separated values (CSV) file example.

This naturally brings us to the last step, transferring grades to an LMS.

3.3.4 Transferring Grades to an LMS

For the time being, this means downloading an Excel or CSV file that includes the student identifying information and the grade (e.g. 11/15) and a percentage (e.g. 73) without the percent sign. The teacher can then upload this file to the LMS.

3.4 Final Deliverables

Upon completion of the project, the following deliverables will be provided:

- Fully functional AI-assisted grading system integrated with OICLearning project.
- Source code for both backend and frontend components.
- Technical documentation detailing system architecture, setup, and usage.

- User guide for instructors on how to use the system.
- Test cases and testing reports.
- Deployment scripts and configuration files.

Chapter 4

Testing

While a full testing methodology is not required at this stage, we plan to conduct targeted testing activities to ensure that the system meets its key objectives. The testing will focus on verifying functionality, accuracy, and usability.

4.1 Planned Testing Activities

- **Basic Functional Tests:** Verify that the backend API correctly processes images and text submissions, interacts with GPT-4 Vision, and returns the expected output. Confirm that the frontend can display graded results, allow for rubric application, and handle user authentication.
- **OCR and NLP Accuracy Checks:** Provide known inputs (e.g., a set of handwritten and typed responses with known correct answers) to ensure the OCR accurately extracts text and GPT-4 Vision semantically interprets answers as intended.
- **Semantic Grouping Verification:** Input multiple variations of a correct response and a few incorrect ones. Check that the AI groups similar answers

together and isolates outliers. Teachers reviewing these groups should find the clustering logical and helpful for grading.

- **Rubric and Grading Validation:** Apply rubrics to test questions and confirm that the final scores are computed as expected. Check that changing a rubric value updates all affected scores consistently.
- **User Interface and Workflow Checks:** Have a small set of instructors perform typical grading tasks. Gather feedback on the interface clarity, ease of navigation, and overall efficiency. Any confusing elements or inefficiencies will be noted for improvement.
- **Export and Integration Tests:** Generate sample CSV/Excel output and ensure that the data is correctly formatted, including student identifiers, scores, and comments. Test the upload of these files into a mock LMS environment to confirm interoperability.

4.2 Expected Outcomes

By focusing on these testing activities, we aim to:

- Confirm that the system's core functionality (OCR, NLP interpretation, semantic grouping, and grading) operates as intended.
- Identify any major usability issues and address them early.
- Ensure that the grading results are consistent, transparent, and easily transferable to external systems.
- Build confidence that the system will reduce manual grading effort and maintain or improve grading quality.

These initial tests will inform further refinements and pave the way for more comprehensive evaluations in a future development phase.

Chapter 5

Conclusion

This project proposes the development of an AI-assisted grading system integrated with OICLearning.com, leveraging GPT-4 and GPT-4 Vision capabilities. By automating OCR, NLP interpretation, and semantic grouping of student answers, the system aims to streamline the grading process, ensure greater consistency, and provide timely, meaningful feedback to students.

While challenges exist—such as ensuring accuracy, handling diverse input formats, and maintaining fairness—thoughtful design, targeted testing, and human oversight can help address these concerns. The planned testing activities will help validate functionality and usability, guiding improvements prior to deployment.

Overall, this project sets the stage for a more efficient and equitable grading process. By reducing teachers' manual workload, improving feedback quality, and helping identify common misconceptions, the system can ultimately enhance educational experiences for instructors and students alike.