

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 Professor Anderson

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 a high-level AI 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

<b>Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background</b>	<b>3</b>
2.1 Artificial Intelligence in Education . . . . .	3
2.1.1 Personalized Learning and Feedback . . . . .	4
2.1.2 Administrative Efficiency . . . . .	4
2.2 Advancements in Large Language Models . . . . .	5
2.2.1 Introduction to GPT-4 . . . . .	5
2.2.2 GPT-4 in Educational Applications . . . . .	5
2.3 AI-Assisted Grading Using GPT-4 . . . . .	6
2.3.1 Automating Short Answer Grading . . . . .	6
2.3.2 Grading Handwritten Responses . . . . .	6
2.3.3 Semantic Understanding and Contextual Evaluation . . . . .	7
2.4 Challenges in AI-Assisted Grading with GPT-4 . . . . .	7

2.4.1	Accuracy and Reliability . . . . .	7
2.4.2	Bias and Fairness . . . . .	8
2.4.3	Handwriting Recognition Limitations . . . . .	8
2.4.4	Ethical and Privacy Concerns . . . . .	8
2.5	Existing Research on AI Grading Systems . . . . .	9
2.5.1	Machine Learning Approaches . . . . .	9
2.5.2	Feasibility Studies with GPT-4 . . . . .	9
2.5.3	Student Perceptions of AI Grading . . . . .	10
2.6	Potential Benefits of GPT-4 in Grading . . . . .	10
2.6.1	Efficiency and Scalability . . . . .	10
2.6.2	Consistency and Fairness . . . . .	10
2.6.3	Timely Feedback . . . . .	11
2.6.4	Enhanced Learning Outcomes . . . . .	11
2.7	Summary . . . . .	11
2.8	Conclusion . . . . .	11
<b>3</b>	<b>Proposal</b>	<b>13</b>
3.1	Summary of Requirements . . . . .	13
3.2	OICLearning Grading Process . . . . .	14
3.2.1	Requirements for Creating Assignments . . . . .	14
3.2.2	Filled-Form Assignments . . . . .	15
3.2.3	Free-Form Written Assignments . . . . .	17
3.2.4	Filled-Form Online Assignments . . . . .	18
3.2.5	Objective-Form Assignments/Assessments . . . . .	19
3.3	Proposed Approach . . . . .	19
3.3.1	Python-Based AI Grading System . . . . .	20

3.3.2	Web Application Integration . . . . .	21
3.3.3	Beta and User Testing . . . . .	22
3.4	Final Deliverables . . . . .	22
3.5	Development Prerequisites . . . . .	23
3.6	Task Delineation . . . . .	23
3.6.1	Task Breakdown . . . . .	23
3.6.2	Timeline Summary . . . . .	25
3.7	Risk Assessment and Mitigation . . . . .	25
3.7.1	Potential Risks . . . . .	25
3.7.2	Mitigation Strategies . . . . .	26
3.8	Conclusion . . . . .	26
<b>4</b>	<b>Testing and Evaluation Plan</b>	<b>29</b>
<b>5</b>	<b>Conclusion</b>	<b>31</b>
<b>A</b>	<b>Requirements Specification</b>	<b>33</b>
	<b>Bibliography</b>	<b>35</b>





## List of Figures



## List of Tables



# Chapter 1

## Introduction

In the educational landscape, grading written student work is a task of high importance, this directly influences the learning experience as well as being able to provide crucial feedback to both students and educators. While traditional grading methods are effective, they can sometimes be time-consuming and labor-intensive, particularly for teachers managing large volumes of students and their responses. The need for efficient and scalable grading solutions has become increasingly evident as educators seek to streamline their workloads without compromising the quality of their assessments.

Recent advancements in artificial intelligence, particularly in the development of Large Language Models (LLMs), offer a promising approach to this challenge. LLMs, such as OpenAI's GPT-4, have shown 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 proposes 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 an educational platform like 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, the integration of LLMs into the grading process presents certain challenges that must be carefully addressed. One of the key considerations is the model's ability to accurately differentiate between nuanced meanings and appropriately group responses that, while semantically similar, may vary in their correctness. Additionally, the effectiveness of these models in handling a diverse range of student expressions and potential errors needs to be thoroughly evaluated.

## Chapter 2

# Background

The integration of artificial intelligence (AI) into education has revolutionized the way teaching and learning processes are conducted. 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 dives 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 an significant part of modern education, it offers tools and solutions that enhance learning experiences and streamlines administrative tasks. AI-driven educational platforms can personalize learning paths, adapt content to individual student needs, and provide real-time feedback [1]. This personalization can be important 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 [1]. 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 departments as well. Tasks such as grading, scheduling, and record-keeping can be handled efficiently by AI systems, reducing the workload on educators [1]. 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 students



that are absent. 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 [1].

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 [1]. 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 [4]. This semantic understanding allows for more accurate grading of open-ended questions.

Liu et al. (2024) [4] 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) [2] 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 [1]. 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. (2024) [4] 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 lower this issue this issue.

### **2.4.2 Bias and Fairness**

AI models can inadvertently show 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. (2024) [5] 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. (2024) [3] 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) [1] 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 blowup 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 (2024) [6] 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) [2] 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. (2024) [5] 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 [1].

### **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 transforming potential of AI-assisted grading using GPT-4. By understanding the capabilities and limitations of these

systems, teachers can implement AI tools that reflect 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 project aims to automate the grading process for different types of assignments submitted on OICLearning.com. The system will utilize GPT-4 and GPT-4 Vision for grading assignments, both scanned and online submissions, using a Python-based AI tool that will later be integrated into a web application. The following

sections describe the current manual grading process requirements and how the AI-based grading system will be developed to fulfill these needs.

## **3.2 OICLearning Grading Process**

OICLearning.com currently lacks an automated grading system for general questions such as, short answer, mathematical and scientific questions, that do not require diagrams or figures in the answer. The proposed AI-assisted grading system handle these various types of assignments, including scanned handwritten submissions and online text entries. By leveraging GPT-4 Vision's ability to interpret images and perform NLP, along with custom grading algorithms, the system will streamline grading for educators by automating tasks that would otherwise require significant manual effort.

The current manual grading process involves several steps:

1. A teacher creates an assignment on Moodle.
2. A student submits an assignment.
3. The teacher grades the assignment.
4. The teacher realizes that nothing good happens after 2 AM and ponders existential questions.

The new AI-assisted system will automate step 3 and remove step 4, significantly reducing the time and effort required from educators.

### **3.2.1 Requirements for Creating Assignments**

Assignments on OICLearning.com fall into three categories:

1. **Filled-Form Scanned Assignments:** These are written tests that students fill out manually and are later scanned and uploaded.
2. **Free-Form Written Assignments:** These assignments have flexible content, and the location of questions and answers may vary.
3. **Filled-Form Online Assignments:** These are completed online directly through the platform.

Filled-Form Scanned assignments must allow the instructor to designate areas where answers are located. Each of these areas must be labeled using a unique string (e.g. "1", "2.a", "2.b", ...). The system must also allow the instructor to input the expected answers for each question and specify the number of points each question is worth. Additionally, work may be given partial credit if, 1) the instructor tagged a part of the answer as worth partial credit with the number of points, and 2) the student's answer contains semantically equivalent work.

The AI system must be able to handle grading for all three types of assignments by utilizing GPT-4 Vision to read and interpret images directly, performing OCR and NLP in a single step, and applying grading algorithms to assign appropriate grades.

### 3.2.2 Filled-Form Assignments

For filled-form assignments, the AI system must meet the following requirements:

1. The assignment is printed as a form by the teacher for the students to fill out.
2. A PDF or image of the assignment is pre-loaded into OICLearning.

3. The system must identify the locations of Name and ID number (or Email) directly from the image using GPT-4 Vision.
4. Question and Answer locations are identified by the teacher using a user-friendly interface to draw rectangles on the pre-loaded image, specifying the question number (e.g., 1.a., 1.b., etc.) and points.
5. The system must incorporate grading algorithms that compare student responses to expected answers for each question to assign appropriate grades.

The system should support both teacher and student uploads of images or PDFs, automatically assigning them to the correct students and handling various document lengths. GPT-4 Vision will read the images directly, extracting the necessary information without the need for separate OCR processes.

#### **3.2.2.1 Process for Grading Filled-Form Assignments**

1. After uploading the assignment, GPT-4 Vision identifies the student by reading the handwritten Name and ID from the image.
2. Upon accessing the grading page, the AI lists questions and whether they are graded.
3. The teacher inputs the expected answers for each question into the system.
4. GPT-4 Vision reads and interprets student responses directly from the images.
5. The AI compares student responses to the expected answers using the grading algorithms, assigning grades based on correctness and completeness.

6. Default groupings are created based on semantic similarity and correctness. Teachers can create additional groups, adjust groupings, rename groups, and override AI decisions.
7. Once groups are finalized, the teacher reviews the assigned grades, adds comments if necessary, and confirms the grades.
8. The grades are then applied to all relevant student responses, and the teacher proceeds to the next question.
9. Once all questions are graded, a final score is recorded for each student.

### **3.2.3 Free-Form Written Assignments**

Free-form assignments require a flexible approach due to their variability.

#### **3.2.3.1 Requirements**

1. Teachers must identify Question and Answer pairs, specifying question numbers and points.
2. Assignments are uploaded as images or PDFs, and GPT-4 Vision handles varying lengths by interpreting the content directly from the images.
3. The AI system allows for manual marking of each answer within the image, flagging unanswered questions, and overwriting old data if new submissions are made before the deadline.
4. Grading algorithms compare student responses to expected answers provided by the teacher.

### **3.2.3.2 Process for Grading Free-Form Assignments**

1. The teacher uploads the assignment and inputs the expected answers for each question.
2. GPT-4 Vision processes the images to identify question locations and extract responses.
3. The AI compares student responses to the expected answers using the grading algorithms, assigning grades accordingly.
4. Answers are grouped by the AI based on semantic similarity and correctness, with manual adjustments available to the teacher.
5. Teachers can add comments and adjust grades as they review each group.
6. Grades are automatically applied to all students in each group, with a final score calculated upon completion.

### **3.2.4 Filled-Form Online Assignments**

For online submissions, the process is streamlined since the input is already in text format.

#### **3.2.4.1 Process for Grading Online Assignments**

1. Students submit their answers directly through the platform.
2. The teacher inputs the expected answers for each question into the system.
3. GPT-4 processes the text responses, comparing them to the expected answers using the grading algorithms.

4. Responses are grouped based on semantic similarity and correctness.
5. Grading and feedback generation follow the same procedure as with scanned assignments.

### **3.2.5 Objective-Form Assignments/Assessments**

Objective-form assignments, such as multiple-choice or fill-in-the-blank questions, can also benefit from GPT-4 Vision's capabilities.

#### **3.2.5.1 General Issues and Requirements**

1. GPT-4 Vision reads the selected answers directly from the images.
2. The system must handle various answer formats, including circling, underlining, or filling boxes.
3. The AI compares the selected answers to the correct options provided by the teacher, assigning grades accordingly.
4. Responses are automatically grouped under correct or incorrect categories.
5. Teachers can review groupings and adjust as necessary.

## **3.3 Proposed Approach**

The AI-assisted grading system will be developed using Python and integrated with GPT-4 and GPT-4 Vision to automate the grading process. GPT-4 Vision will read and interpret images directly, performing both OCR and NLP in a single step, while the grading algorithms will compare student responses to expected answers to assign grades. This approach simplifies the architecture and improves efficiency.

### 3.3.1 Python-Based AI Grading System

The Python-based backend will utilize GPT-4 and GPT-4 Vision for processing both scanned and online assignments. GPT-4 Vision will handle the direct interpretation of images, performing OCR and NLP simultaneously, while custom grading algorithms will assign grades based on expected answers.

#### 3.3.1.1 Development Steps

1. **Data Ingestion:** Develop modules to handle the ingestion of images and online submissions.
2. **Image Interpretation:** Utilize GPT-4 Vision to read images, extracting text and understanding content directly.
3. **Expected Answer Input:** Implement interfaces for teachers to input expected answers and grading criteria for each question.
4. **Response Analysis:** Use GPT-4 and GPT-4 Vision to interpret student responses and compare them to expected answers.
5. **Grading Algorithms:** Develop algorithms that assign grades based on the comparison between student responses and expected answers, considering partial credit and common misconceptions.
6. **Response Grouping:** Implement algorithms to group responses based on correctness and semantic similarity to facilitate efficient grading.
7. **User Interface Development:** Create a user-friendly interface for educators to interact with the system, input expected answers, review AI-generated grades, and make necessary adjustments.



8. **Database Integration:** Set up a secure database to store student submissions, grades, expected answers, and feedback.
9. **Testing and Validation:** Test the system with sample data to ensure accuracy and reliability, focusing on the performance of GPT-4 Vision and the grading algorithms.

### 3.3.2 Web Application Integration

Following the backend development, the system will be integrated into a web application to provide an accessible interface for users.

#### 3.3.2.1 Integration Steps

1. **Frontend Development:** Use modern web technologies like React or Angular to build an interactive user interface.
2. **API Development:** Expose backend functionalities through secure APIs for frontend consumption.
3. **Authentication and Security:** Implement robust authentication mechanisms to protect user data and ensure secure access.
4. **Responsive Design:** Ensure the application is accessible across various devices, including desktops, tablets, and smartphones.
5. **User Experience (UX) Optimization:** Design intuitive workflows for teachers to input expected answers, review AI-generated grades, and for students to view feedback.

### 3.3.3 Beta and User Testing

To ensure the system meets user needs and functions correctly:

1. **Pilot Program:** Conduct a pilot with select educators and students to gather initial feedback.
2. **Iterative Improvements:** Use feedback to refine system functionalities, fix bugs, and enhance usability.
3. **Performance Monitoring:** Assess system performance under real-world conditions, making adjustments as necessary.
4. **Documentation and Training:** Provide comprehensive documentation and training materials to support users during the beta phase.

## 3.4 Final Deliverables

Once the project is complete, the following deliverables will be provided:

- **Application Source Code:** The full source code for the AI-assisted grading system, including grading algorithms.
- **Administrative Access to Project Source Control:** Access to the code repository for ongoing maintenance and updates.
- **User and Developer Documentation:** Detailed guides on the application's general structure, codebase, usage instructions, and grading algorithm specifications.
- **Testing and Evaluation Report:** A comprehensive report on all testing phases, including beta testing results and user feedback.

- **Final Project Report:** A detailed report summarizing the project's goals, implementation, challenges, outcomes, and future recommendations.

## 3.5 Development Prerequisites

To develop and test the AI-assisted grading system, the following hardware and software are required:

- **Hardware:** computer for development
- **Software:** Python for AI model implementation, React/Angular for front-end development, and Git for version control.
- **Tools:** Database management systems (e.g., PostgreSQL), cloud services for deployment and scaling, and tools for developing and testing grading algorithms.
- **Access to GPT-4 and GPT-4 Vision:** Licenses or API access to OpenAI's GPT-4 and GPT-4 Vision models for integration and testing.

## 3.6 Task Delineation

This section outlines the specific tasks and timelines required to complete the project.

### 3.6.1 Task Breakdown

1. **Requirement Analysis and Planning (2 weeks)**
  - Define system requirements and specifications.

- Design system architecture and data models.
- Outline grading algorithm requirements.

## 2. **Backend Development** (2.5 months)

- Implement data ingestion and processing modules.
- Integrate GPT-4 and GPT-4 Vision APIs for image and text interpretation.
- Develop grading algorithms to assign grades based on expected answers.
- Set up databases and ensure data security.
- Implement feedback generation mechanisms.

## 3. **Frontend Development** (2 months, overlapping with Backend Development)

- Build user interfaces for teachers to input expected answers and review grades.
- Develop interfaces for students to view grades and feedback.
- Implement authentication and user management.
- Design intuitive workflows for grading review and adjustments.

## 4. **Integration and Testing** (1.5 months)

- Integrate frontend and backend components.
- Perform unit, integration, and system testing.
- Optimize performance and scalability.
- Validate grading algorithms with test data.

## 5. **Beta Testing and Deployment** (1 month)

- Deploy the system for beta users.

- Collect feedback and make necessary refinements.
- Prepare for full-scale deployment.

## 6. Documentation and Final Reporting (2 weeks)

- Finalize all documentation.
- Develop training materials and conduct training sessions.
- Compile final project report.

### 3.6.2 Timeline Summary

Task	Duration
Requirement Analysis and Planning	2 weeks
Backend Development	2.5 months
Frontend Development	2 months
Integration and Testing	1.5 months
Beta Testing and Deployment	1 month
Documentation and Final Reporting	2 weeks

## 3.7 Risk Assessment and Mitigation

### 3.7.1 Potential Risks

- **API Limitations:** Changes in GPT-4 or GPT-4 Vision APIs could impact development.
- **Data Privacy Concerns:** Handling student data requires strict compliance with privacy laws.

- **Technical Challenges:** Developing accurate grading algorithms that handle a variety of correct responses.
- **User Adoption:** Resistance from educators or students in adopting the new system.

### 3.7.2 Mitigation Strategies

- **API Contingency Planning:** Stay updated with API changes and maintain communication with OpenAI.
- **Privacy Compliance:** Implement robust security measures and ensure compliance with regulations like FERPA and GDPR.
- **Algorithm Validation:** Rigorously test grading algorithms with diverse data sets and involve educators in the development process.
- **User Engagement:** Involve users early in the development process; provide training and highlight the benefits of the system.

## 3.8 Conclusion

The proposed AI-assisted grading system aims to revolutionize the grading process on OICLearning.com by integrating GPT-4 and GPT-4 Vision along with custom grading algorithms that assign grades based on expected answers. By automating the grading of both text-based and handwritten assignments through direct image interpretation, the system will significantly reduce educators' workload, enhance grading consistency, and provide timely, personalized feedback to students.

The project's success will demonstrate the practical application of advanced AI technologies in education, paving the way for future innovations.





## **Chapter 4**

# **Testing and Evaluation Plan**



## **Chapter 5**

## **Conclusion**



## **Appendix A**

### **Requirements Specification**



# Bibliography

- [1] Valentina Alto. *Modern Generative AI with ChatGPT and OpenAI Models: Leverage the capabilities of OpenAI's LLM for productivity and innovation with GPT3 and GPT4*. Packt Publishing Ltd, 2023. [2.1](#), [2.1.1](#), [2.1.2](#), [2.2.1](#), [2.2.2](#), [2.3.3](#), [2.4.4](#), [2.6.1](#)
- [2] Gerd Kortemeyer. Toward ai grading of student problem solutions in introductory physics: A feasibility study, November 2023. [2.3.2](#), [2.5.2](#)
- [3] Gerd Kortemeyer, Julian Noh, and Daria Onishchuk. Grading assistance for a handwritten thermodynamics exam using artificial intelligence: An exploratory study, June 2024. [2.4.3](#)
- [4] Tianyi Liu, Julia Chatain, Laura Kobel-Keller, Gerd Kortemeyer, Thomas Willwacher, and Mrinmaya Sachan. Ai-assisted automated short answer grading of handwritten university level mathematics exams, August 2024. [2.3.1](#), [2.4.1](#)
- [5] Chad C. Tossell, Nathan L. Tenhundfeld, Ali Momen, Katrina Cooley, and Ewart J. de Visser. Student perceptions of chatgpt use in a collegeessay assignment: Implications for learning,grading, and trust in artificial intelligence, 2024. [2.4.2](#), [2.5.3](#)

- [6] Rebecka Weegar and Peter Idestam-Almquist. Reducing workload in short answer grading using machine learning, June 2024. [2.5.1](#)