

# Chapter 1

## INTRODUCTION

### 1.1 Introduction

With the rapid integration of technology into education, teaching has become more engaging through tools like projectors, online tutorials, and animated video lessons. However, while instructional methods have evolved, the evaluation of descriptive answer sheets remains largely manual. Faculty members still evaluate exam papers by hand, which is time-consuming, inconsistent, and susceptible to human error. Subjectivity and emotional variation during manual correction can also lead to unfair grading.

To address these challenges, the Gradex project proposes an automated evaluation system that utilizes Optical Character Recognition (OCR) and Natural Language Processing (NLP) to streamline and standardize answer sheet correction. By replacing the traditional manual evaluation process with AI-driven mechanisms, we can ensure faster, more consistent, and transparent grading.

This system uses Tesseract OCR to extract handwritten or printed text from scanned answer sheets. The extracted text is then preprocessed and evaluated using Ollama language models, which analyze the content against provided answer keys and question papers. This unified evaluation logic ensures fairness and eliminates personal bias.

The project also features a user-friendly interface where teachers can upload answer sheets, organize evaluations by exam type (e.g., CAT 1, Term 1), and access results in real-time. All evaluated data is stored in a structured database and can be exported in CSV format or PDF format for reporting or record-keeping.

### 1.2 Problem Statement

Despite advancements in teaching technologies, most educational institutions still rely on traditional, manual methods for evaluating exam papers. This manual evaluation process is labor-intensive, time-consuming, and prone to inconsistency due

to individual differences among evaluators. Additionally, the lack of standardization in marking schemes can lead to unfair assessments.

To overcome these issues, Gradex introduces an AI-based automated evaluation system. This solution reduces correction time, eliminates evaluator bias, and ensures a consistent and objective grading process. The system also provides quick and accessible results, improving the overall efficiency of academic evaluations.

### 1.3 Objectives

The objectives of the Gradex system are:

- To automate the evaluation of internal exam papers and reduce the time and manual effort required.
- To ensure consistency in marking by applying a common evaluation standard powered by AI.
- To minimize human errors and emotional bias in the correction process.
- To provide instant results and transparent grading accessible to both teachers and students.
- To categorize and manage results exam-wise (e.g., Term 1, CAT 1, CAT 2) for easier reporting and tracking.

### 1.4 Scope

The scope of this project is to develop a **fully automated answer sheet evaluation platform** for internal academic assessments. The system performs the following tasks:

- Extracts text from scanned answer sheets using Tesseract OCR.
- Uses Ollama LLMs to compare extracted answers with the answer key.
- Scores answers, calculates total marks, and stores the results in a database.
- Offers an interface for teachers to upload, categorize, and manage answer sheets.
- Provides downloadable CSV reports for easy sharing and record maintenance.
- Can be extended for use in schools, colleges, coaching centers, and competitive exam bodies.

## 1.5 Organization of the Report

### Chapter 1: Introduction

- Learn the challenges of manual grading in education.
- Understand the need for automation in answer sheet evaluation.
- Discover how OCR and NLP can standardize grading.
- Know the objectives: reduce human bias, speed up corrections, and ensure transparency.
- Understand the project scope and what the system will do.

### Chapter 2: Literature Survey

- Explore previous research on automated grading systems.
- Learn about techniques like rule-based NLP, machine learning classifiers, and transformer models.
- Identify the limitations of existing methods (lack of semantic understanding, poor OCR accuracy).
- Understand the technological gap Gradex is designed to fill.

### Chapter 3: System Analysis

- Understand how manual and basic automated systems work and their limitations.
- Learn the importance of semantic understanding in grading.
- See how Gradex combines OCR and NLP for better performance.
- Understand the system architecture and data flow across components.

### Chapter 4: System Requirements

- Know the hardware needed to run the system (processor, RAM, storage, scanner).
- Learn about the software stack (Python, Django, PostgreSQL, Tesseract, Ollama, BERT).
- Understand the functional requirements like login, upload, evaluate, and dashboard access.
- Review non-functional aspects like reliability, performance, scalability, and usability.

## **Chapter 5: System Implementation**

- Learn how the system is designed using UML diagrams.
- Understand how users interact with the system (use case diagram).
- See the relationships between data objects (class diagram).
- Follow the sequence of system operations (sequence diagram).
- Understand workflows and decisions (activity diagram).
- View how data flows and is stored (DFD and ER diagrams).

## **Chapter 6: Modules Implementation**

- Learn how OCR extracts text from answer sheets.
- Understand how the Ollama module cleans and formats the OCR text.
- See how BERT calculates semantic similarity between student answers and the answer key.
- Know how grading is done using similarity scores and how feedback is generated.
- Explore the features of the teacher dashboard (exam management, result review).
- Explore the features of the student dashboard (result viewing, feedback access).

## **Chapter 7: Experimental Results**

- Analyze OCR accuracy for handwritten and printed text.
- See how text refinement improves grading accuracy.
- Understand how BERT handles varied student responses and evaluates meaning.
- Review overall system performance in terms of accuracy, speed, and scalability.

## **Chapter 8: Conclusion**

- Recap how Gradex solves key challenges in academic evaluation.
- Learn the benefits for teachers (faster, consistent grading) and students (instant feedback).
- Understand the effectiveness and reliability of the system in real use cases.

## **Chapter 9: Appendices**

- Review sample Django source code used to build Gradex.
- View screenshots of the user interfaces for students and teachers.

## **Chapter 10: Future Enhancement**

- Learn how Gradex can automate input using mobile cameras or digital tablets.
- Explore plans for multilingual support and real-time evaluation.
- See how feedback loops can improve scoring accuracy.
- Understand plans for advanced analytics and plagiarism detection.
- Explore future integration with LMS platforms and mobile apps.

## Chapter 2

### LITERATURE SURVEY

A literature survey or a literature review in a project report is that section which shows the various analysis and research made in the field of interest and the results already published, taking into account the various parameters and the extent of the project.

#### 2.1 Introduction

In education, grading handwritten exam answers is a time-consuming and labor-intensive task. Traditional grading approaches face challenges, such as subjectivity, inconsistency, and human error, particularly with large volumes of scripts. Automating this process can not only reduce the grading workload but also provide faster and fairer feedback to students.

To address these needs, the proposed system leverages advanced technologies, integrating Optical Character Recognition (OCR) with Large Language Models (LLMs) such as BERT and LLaMA. OCR enables accurate text extraction from handwritten responses, which can then be refined and interpreted by LLMs for effective grading. These models are adept at understanding semantic meaning, ensuring that answers are evaluated based on content rather than exact phrasing. This system also enhances reliability by mitigating OCR errors and improving grading accuracy across varied subjects and response styles.

By combining OCR and LLM-based semantic analysis, this system aims to transform grading into a more efficient, scalable, and objective process. The result is a significant step forward in educational technology, offering a consistent grading solution adaptable to diverse educational environments and large-scale assessments.

#### 2.2 Automated Evaluation of Student Answers using NLP Techniques

**Authors:** Kumar A., Singh R.

**Year:** 2020

This research focused on designing a rule-based Natural Language Processing (NLP) system to evaluate descriptive answers submitted by students. The system

processes both student responses and model answers through tokenization, stop-word removal, and stemming. It then performs keyword matching and uses cosine similarity to measure textual overlap between the student's answer and the model solution. The evaluation is based on the number of matched keywords and overall similarity. The system provides a lightweight solution for automating descriptive answer scoring in controlled academic environments.

**Disadvantages:**

- The method lacks semantic understanding and cannot comprehend paraphrased or grammatically restructured answers.
- It performs poorly when students use synonyms or varied sentence structures.
- It requires manual configuration of keywords and threshold similarity values for different questions

### ***2.3 Intelligent Assessment System using Machine Learning for Descriptive Answers***

**Authors:** Sharma N., Patel V.

**Year:** 2021

This system adopted a data-driven machine learning approach to score descriptive student answers. A training dataset was compiled consisting of various answers labeled with scores assigned by human examiners. Using TF-IDF for feature extraction and classifiers like Support Vector Machines (SVM) and Random Forests, the model was trained to predict scores based on patterns learned from the labeled data. This approach enabled generalization to new student answers without predefined rules, and allowed faster and automated scoring once trained.

**Disadvantages:**

- The model's performance was highly dependent on the quality and diversity of the training data.
- It was not effective when applied to questions outside the training distribution or domain.
- The system lacked semantic reasoning and contextual understanding.

### ***2.4 Automatic Grading of Short Answers using BERT***

**Authors:** Zhang Y., Liu H.

**Year:** 2022

This project introduced a transformer-based model, BERT, to perform automatic grading by capturing the deep contextual meaning of student answers. The model was fine-tuned on a dataset containing question-answer-score triples. BERT's attention mechanism allowed the system to understand the relevance of each word in context, enabling it to evaluate even paraphrased or semantically equivalent answers accurately. This approach showed significant improvement over traditional keyword or rule-based models.

**Disadvantages:**

- The system was computationally expensive, requiring GPU acceleration for real-time use.
- It required careful fine-tuning for different subjects or types of questions.
- Interpretability was a challenge, as deep learning decisions were opaque to users and teacher

## ***2.5 OCR-Based Automated Examination System***

**Authors:** Das S., Roy A.

**Year:** 2019

This work focused on using Optical Character Recognition (OCR) to digitize handwritten student answers. The system used Tesseract OCR to convert scanned answer sheets into text. After extraction, the system applied basic keyword-based matching to evaluate content correctness. This project served as an initial step towards paperless evaluations by eliminating manual data entry.

**Disadvantages:**

- OCR performance degraded significantly with poor handwriting or low image quality.
- Keyword matching lacked depth and failed to consider context or answer coherence.
- The system was not scalable for subjective, multi-paragraph answers.

## ***2.6 Assessment using AI: A Comparative Study of Rule-Based and ML Models***

**Authors:** Thomas J., Mehta R.

**Year:** 2021



This study compared different AI approaches — rule-based, traditional machine learning, and deep learning models — for automated evaluation. The authors developed small-scale prototypes for each category and analyzed them for accuracy, training time, scalability, and generalization. This comparative framework helped highlight the strengths and limitations of each technique, offering guidance for hybrid model development.

**Disadvantages:**

- No single model performed well across all question types.
- Deep models required significant resources and domain-specific tuning.
- Rule-based systems needed continuous manual maintenance and updates.

## Chapter 3

### System Analysis

#### *3.1 Existing System*

##### *3.1.1 Traditional Manual Grading and Its Limitations*

Manual evaluation of descriptive answer sheets remains the dominant approach in most educational institutions. In this process, teachers individually review each student's handwritten responses, compare them with expected answers, and assign marks based on their interpretation and experience. While this method allows for human judgment and flexibility, it suffers from several critical limitations. Firstly, it is extremely **time-consuming**, especially when dealing with large volumes of answer sheets, such as during term-end or competitive examinations. Teachers often spend long hours reading through pages of handwritten content, leading to fatigue and slower result processing.

Another key drawback is the **subjectivity and human bias** involved in manual grading. Different teachers may interpret and score the same answer differently due to personal preferences, expectations, or unconscious bias. This lack of standardization results in inconsistent and unfair scoring. Additionally, **errors and inconsistencies** frequently occur, particularly under tight deadlines. Mistakes in mark calculation, misinterpretation of answers, and overlooked content are common when teachers are under pressure or when answer sheets are not clearly written. These challenges have highlighted the urgent need for a more objective, scalable, and accurate solution for evaluating descriptive answers.

##### *3.1.2 Automated Grading for Objective-Type Questions*

Automated grading systems have already seen success in handling objective-type assessments like multiple-choice questions, true/false, and fill-in-the-blank formats. E-learning platforms such as Google Forms, Moodle, and Edmodo employ **rule-based grading mechanisms**, which rely on exact answer matching. These systems offer speed and accuracy in assessing large datasets with zero ambiguity in answers. However, their limitations become apparent when used to evaluate **descriptive or subjective answers**, as rule-based systems struggle to interpret synonyms, paraphrased responses, or

sentence-level meaning. They are unable to provide meaningful feedback or assess the creativity and reasoning skills exhibited in longer answers.

### ***3.1.3.OCR-Based Answer Extraction***

To digitize and process handwritten answer sheets, **Optical Character Recognition (OCR)** is an essential technology. Several tools and APIs like **Tesseract OCR**, **Google Vision API**, and **OpenAI's OCR models** have been applied to extract text from scanned images or photographs of answer sheets. While these tools work well for printed text, **handwritten content remains a major challenge** due to varied writing styles, poor scan quality, image noise, and inconsistencies in character formation. These limitations often result in partially or incorrectly extracted text, which negatively affects the downstream evaluation process. The current study overcomes this by employing **Ollama OCR**, a deep learning-based handwriting recognition model that improves accuracy significantly. It leverages neural networks trained on diverse handwriting samples, enabling better generalization across different student answer scripts.

### ***3.1.4.NLP and AI in Descriptive Answer Evaluation***

Recent advancements in **Natural Language Processing (NLP)** have opened the door to AI-powered evaluation systems that go beyond keyword matching and instead focus on understanding the **semantic meaning** of text. By applying **word embeddings** like Word2Vec, GloVe, and FastText, along with more advanced **transformer-based models** such as BERT, T5, LLaMA, and GPT, these systems can analyze context and evaluate the **logical structure and relevance** of student answers. Unlike traditional rule-based systems, NLP models are capable of **context-aware assessment**, recognizing correct answers written in different ways using synonyms or paraphrased structures.

One of the major strengths of AI-driven evaluation is **semantic similarity matching**, where student responses are compared with model answers not through exact word matching but through contextual alignment. These models also handle **grammatical variations** and sentence restructuring, thus offering a more flexible and accurate evaluation method. Studies show that transformer-based models like BERT and GPT consistently outperform traditional techniques in grading accuracy and

interpretation of complex answers, making them suitable for deployment in educational settings.

### ***3.2 Disadvantages of Existing Systems***

- Time-consuming process for teachers, especially during large-scale examinations.
- Lack of scalability with increasing student population.
- High risk of human bias and subjectivity in grading.
- Inconsistencies and errors due to fatigue or oversight.
- Minimal or no feedback provided to students for improvement.
- Poor performance of automated systems on descriptive answers.
- Dependence on exact keyword matching in automated grading systems.
- OCR tools struggle with handwritten text due to variations in writing style and image quality.
- No digital storage or centralized access to student performance data.
- Limited support for semantic understanding and context-aware evaluation.
- Absence of data-driven analytics and feedback loops for institutional improvement.

### ***3.3 Proposed System***

The proposed system, Gradex, is an AI-powered answer sheet evaluation platform designed to overcome the limitations of traditional manual grading and rule-based automated systems. It combines advanced OCR, NLP, and semantic analysis techniques to accurately evaluate descriptive student answers. By digitizing handwritten responses, extracting meaningful content, and applying intelligent scoring mechanisms, Gradex aims to provide faster, more consistent, and unbiased evaluation of academic answer sheets.

#### ***3.3.1 Handwritten Text Extraction Using Ollama OCR***

At the core of the system is the use of **Ollama OCR**, a deep learning-based optical character recognition engine designed to accurately read and extract text from handwritten documents. Unlike conventional OCR tools like Tesseract or Google Vision API, which struggle with varied handwriting styles, Ollama OCR leverages advanced neural networks to process noisy, distorted, or complex handwriting. This

enables the system to accurately digitize handwritten responses from scanned answer sheets, forming a reliable input for further analysis.

### ***3.3.2 Text Preprocessing and Cleaning***

Once the text is extracted, it undergoes a robust **text preprocessing pipeline** to clean and normalize the content. This step includes removing unwanted characters, correcting common OCR errors, and structuring the response into coherent sentences. Preprocessing also involves tokenization and part-of-speech tagging, which prepares the text for semantic analysis by ensuring clarity and structure in the response data.

### ***3.3.3 Semantic Evaluation Using NLP Models***

The core evaluation mechanism of Gradex uses **state-of-the-art NLP models** to assess the content of student answers. These models—such as **BERT**, **GPT**, and **LLaMA**—are capable of understanding semantic meaning rather than relying on keyword matches. The system compares each student answer with the corresponding model answer (key) using **semantic similarity scoring**, which considers synonyms, paraphrasing, and logical structure.

This allows the system to grade answers that are correct but differently worded, a capability absent in traditional systems. The models also account for grammatical variations and logical coherence, making the grading process more holistic and context-aware.

### ***3.3.4 Score Calculation and Feedback Generation***

Based on the semantic similarity scores and logical structure of the response, Gradex assigns marks to each question. It uses a configurable rubric system to ensure flexibility across different subjects and question types. In addition to scores, the system generates **automated feedback** highlighting the strengths and weaknesses of the student's response. This feedback helps students understand their mistakes and improve in future assessments

### 3.3.5 Teacher Dashboard, Student Information, and Result Accessibility

Gradex provides a comprehensive teacher dashboard that enables educators to efficiently manage the entire evaluation workflow. Teachers can upload scanned answer sheets, create and manage various exams such as CAT 1, CAT 2, and Term 1, and monitor detailed student performance analytics. The dashboard offers valuable insights into individual and class-wide progress, including average scores, frequently missed concepts, and improvement trends. Additionally, the system organizes student results under exam-wise categories, making it easy to retrieve and review historical performance. When a teacher accesses the "Student Info" section, they can view all attempted exams by a student along with the respective scores and AI-generated feedback. Teachers can also review and, if necessary, override AI-evaluated answers, ensuring a transparent and flexible grading process that supports both automation and human oversight.

### 3.4 System Architecture

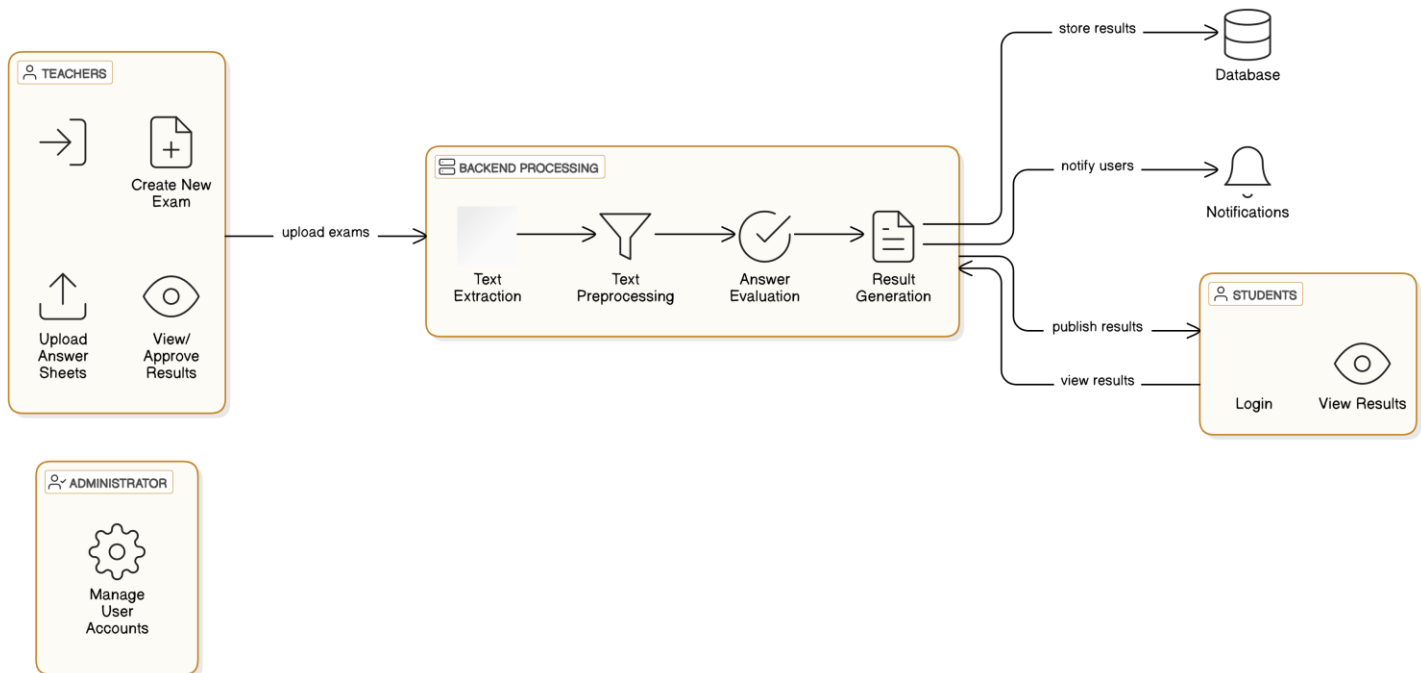


Fig: 3.1 GradeX Architecture

## Chapter 4

### System Requirements

#### *4.1 Overall Description*

This section describes the **system requirements** for the Gradex project, focusing on both **hardware** and **software** components needed to run the system effectively. These requirements are essential for ensuring that the software functions efficiently and reliably within the specified environment.

#### *4.2 Specific Requirements*

##### *4.2.1 Hardware Requirements*

The hardware requirements for the Gradex system outline the necessary physical components to run the software smoothly. The system should be capable of processing large batches of scanned answer sheets and executing complex AI models. The recommended hardware specifications are:

- **Processor:** A **2.0 GHz** or higher multi-core processor (such as Intel i5 or equivalent). This is required to handle the processing tasks of OCR and NLP models efficiently.
- **RAM:** A minimum of **4 GB of RAM** is recommended to allow smooth operation, especially when handling large batches of answer sheets. More memory may be required for large-scale deployments.
- **Disk Space:** The system will need **1 GB** or more of free hard disk space to store the software, processing data, and results. This space is also used for caching images and saving evaluation results.
- **Scanner:** A high-resolution scanner (at least **300 DPI**) is required to ensure that the scanned images of the answer sheets are clear and suitable for OCR processing.
- **Network Connection:** A stable internet connection is required for cloud-based functionalities, such as accessing AI models, uploading files, and synchronizing data with the server.

##### *4.2.2 Software Requirements*

The **software** requirements for Gradex are the essential tools and technologies needed to build, run, and maintain the system. These requirements include the operating system, programming languages, frameworks, and development tools.

- Operating System: Windows 10, Linux (Ubuntu), or macOS
- Programming Language: Python 3.9 or above
- Web Framework: Django
- Database: PostgreSQL (recommended) or SQLite (for testing)
- OCR Tool: Tesseract OCR model
- Evaluation Models: Ollama LLMs for semantic answer evaluation
- Development Tools: VS Code or PyCharm
- Model Testing: Google Colab Notebook
- Deployment: Pythonanywhere.com

### ***4.3 Functional Requirements***

- The system must allow teachers to log in securely and manage exams (e.g., CAT 1, CAT 2, Term 1).
- Teachers should be able to upload scanned student answer sheets in image format.
- The system must extract text from uploaded images using Ollama OCR.
- Extracted text should be evaluated by comparing it with key answers using Ollama LLMs.
- The system must automatically assign marks based on semantic similarity.
- Teachers should be able to review and override AI-evaluated scores.
- The system must store and categorize student results based on exam type.
- The system should provide detailed performance reports for each student and exam.
- Teachers should be able to search and view individual student performance history.

### ***4.4 Non-Functional Requirements***

- **Reliability:** The system should perform evaluations accurately and consistently under normal load.
- **Scalability:** It should support increasing numbers of users and answer sheets without performance degradation.
- **Security:** Only authenticated teachers can access and modify student data. Sensitive information must be protected.
- **Performance:** OCR and evaluation processes should complete within an acceptable response time.



- **Maintainability:** The codebase should be modular and well-documented to ease future updates and debugging.
- **Usability:** The interface should be intuitive and accessible for teachers with basic technical skills.
- **Availability:** The system should be available online with minimal downtime for maintenance or updates.

## Chapter 5

### System Implementation

#### 5.1 Architectural Diagram

The architectural design gives the description about the overall system design. It is specified by identifying the components defining the control and data flow between them. The arrow indicates the connection and rectangular box represents the functional units. The Fig. 5.1 shows the architectural diagram of this project which shows the overall operation from uploading question papers to the evaluation of the student answer scripts.

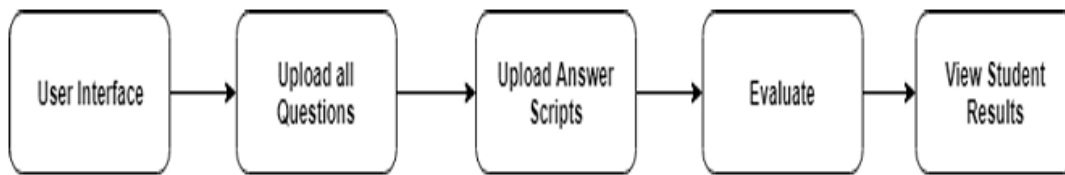


Fig:5.1 GradeX Flow Diagram

#### 5.2 UML Diagram

##### 5.2.1 Use Case Diagram

A Use Case Diagram is a behavioral diagram in UML (Unified Modeling Language) that visually represents the interactions between users (actors) and the system, showing the system's functional requirements. It outlines the various ways in which users interact with the system through specific actions known as use cases.

In the context of the Gradex project, the use case diagram identifies the core functionalities of the AI-based evaluation system and maps out the roles of its primary users—teachers and students. Teachers interact with the system to perform operations such as creating exams, uploading answer sheets, and accessing evaluation results, while students use the system to view their exam results and performance feedback.

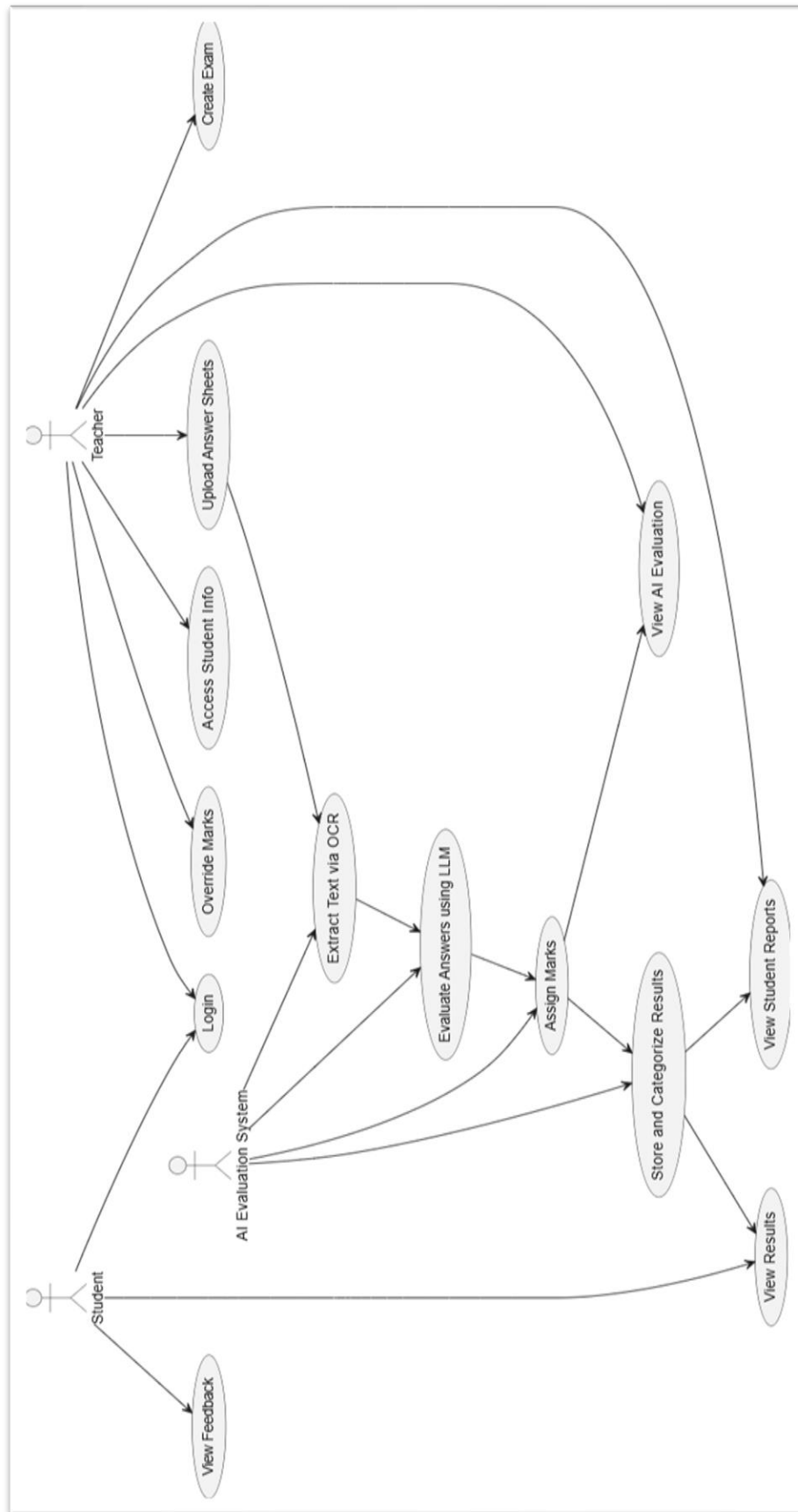


Fig 5.2 Use Case Diagram

### 5.2.2 *Class Diagram*

A Class Diagram is a type of static structure diagram in the Unified Modeling Language (UML) that describes the structure of a system by showing its classes, attributes, methods, and the relationships among the classes. It serves as a blueprint of the system's object-oriented design and forms the foundation for coding and system implementation.

In the context of the Gradex project, the class diagram models the key components of the AI-based answer sheet evaluation system. It includes entities such as Student, Teacher, Exam, AnswerSheet, and EvaluationResult, each represented as a class with defined attributes (e.g., student ID, exam name) and methods (e.g., uploadSheet(), evaluateAnswer()). The relationships between these classes—such as association, aggregation, or inheritance—are also illustrated to show how different components of the system interact.

Class diagrams help developers and stakeholders understand the internal structure and logic of the application. They are essential during the design phase for identifying objects, organizing code, and ensuring the system supports scalability, maintainability, and reusability.

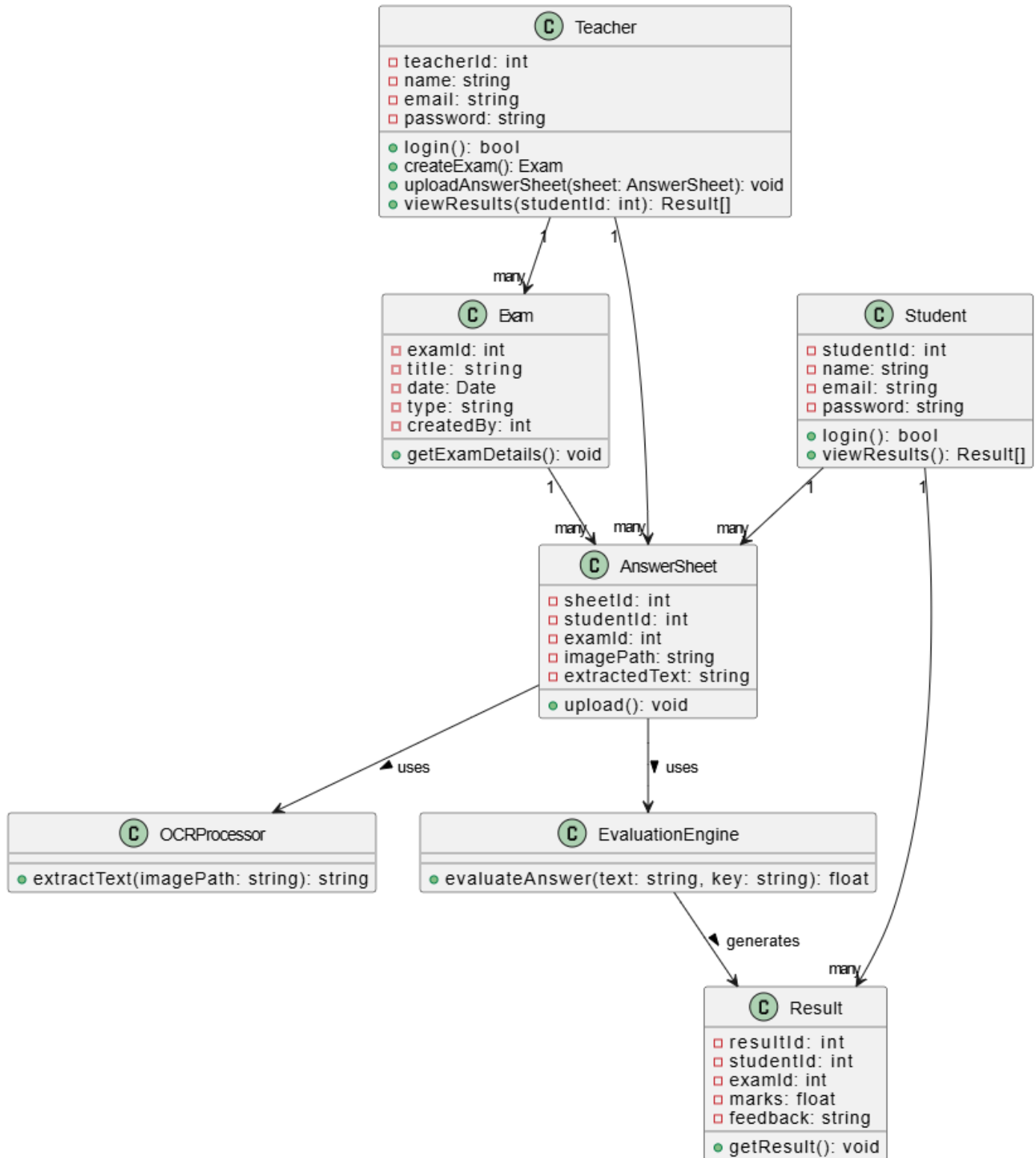


Fig 5.3 Class Diagram

### ***5.2.3 Sequence Diagram***

A Sequence Diagram is a type of interaction diagram in the Unified Modeling Language (UML) that illustrates how objects in a system interact with one another over time. It focuses on the order of messages exchanged between various system components (objects or actors) to accomplish a specific task or use case.

In the context of the Gradex project, the sequence diagram models the flow of interactions between the key actors—Teacher, Student, and the Gradex system components such as the OCR engine, evaluation module, and database. It visually represents how, for example, a teacher uploads an answer sheet, the system extracts text using OCR, evaluates the content using AI models, and stores the results. Similarly, it shows how a student can later access the evaluated results.

The sequence diagram emphasizes temporal ordering, showing which interactions happen first and how control flows among the system parts. It is crucial for understanding runtime behavior, debugging logic flow, and aligning system operations with user expectations.

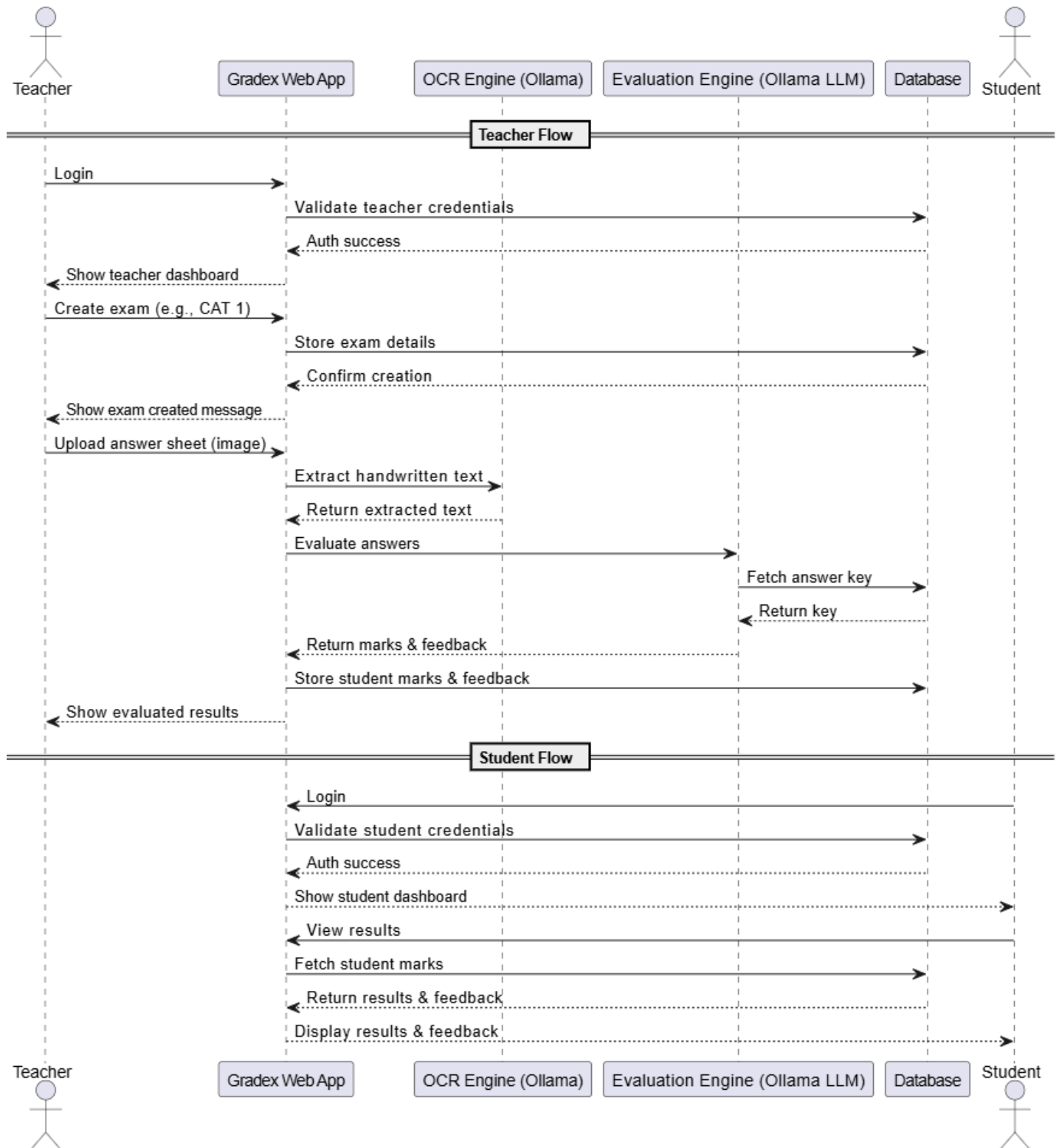


Fig 5.4 Sequence Diagram

### 5.2.4 Activity Diagram

An Activity Diagram is a type of behavioral diagram in the Unified Modeling Language (UML) that represents the workflow or sequence of activities in a system. It focuses on dynamic aspects of the system by modeling the flow of control from one activity to another, including decisions, parallel processes, and loops.

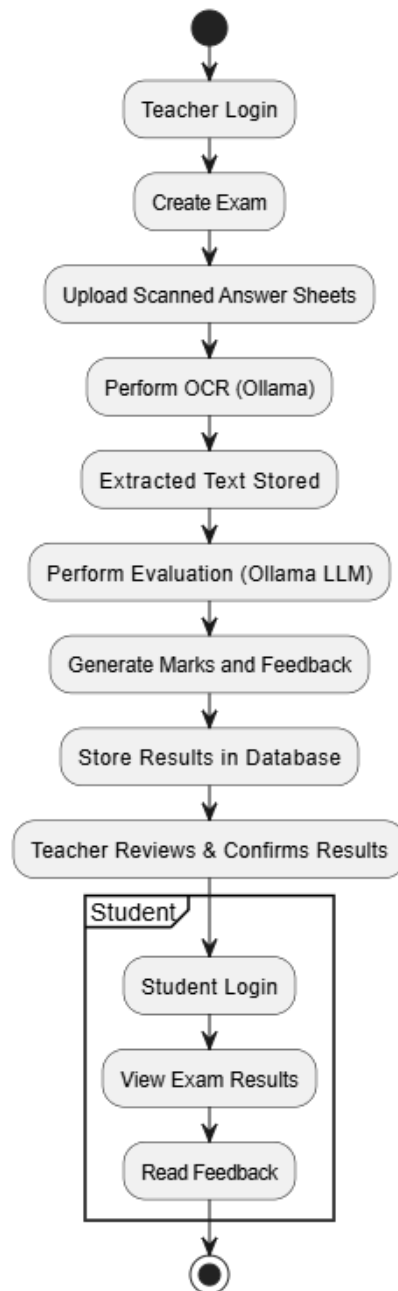


Fig 5.5 Activity Diagram



### 5.2.5 DFD Diagram

A **Data Flow Diagram (DFD)** is a graphical representation used to visualize how **data moves through a system**. It shows the flow of information between various **processes, data stores, external entities**, and the system itself. DFDs are widely used during the system analysis phase to understand the functional aspects of a system and how it handles data.

In the context of the Gradex project, a DFD illustrates how data such as scanned answer sheets, extracted text, evaluation results, and student records flow between key components. These include external entities like Teachers and Students, internal processes like OCR extraction and AI-based evaluation, and data stores such as exam records and student databases.

The DFD helps in identifying the inputs, outputs, and transformations of data within the system. It ensures that all data interactions are accounted for, helping developers and analysts spot bottlenecks, redundancies, or missing components in the system design.

Gradex's DFD provides a logical view of the system without focusing on implementation details, making it easier to communicate system functionality with both technical and non-technical stakeholders.

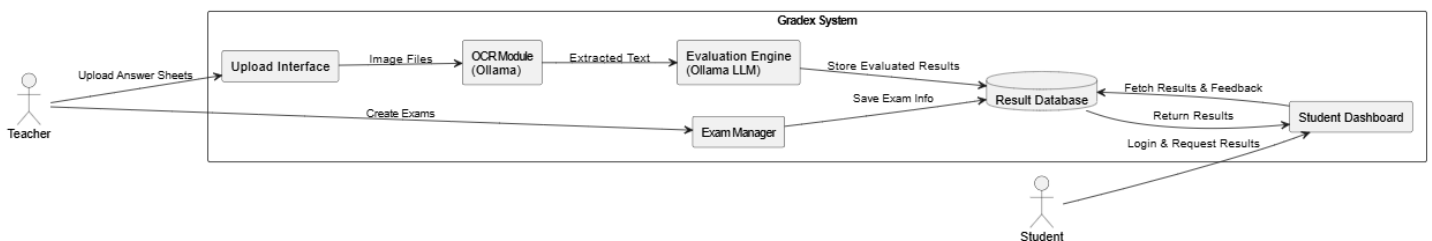


Fig 5.6 DFD Diagram

### 5.2.6 ER Diagram

An **Entity-Relationship (ER) Diagram** is a data modeling technique used to visually represent the **entities** within a system and the **relationships** between them. It provides a high-level conceptual view of how data is structured and connected in a database, forming the foundation for designing relational databases.

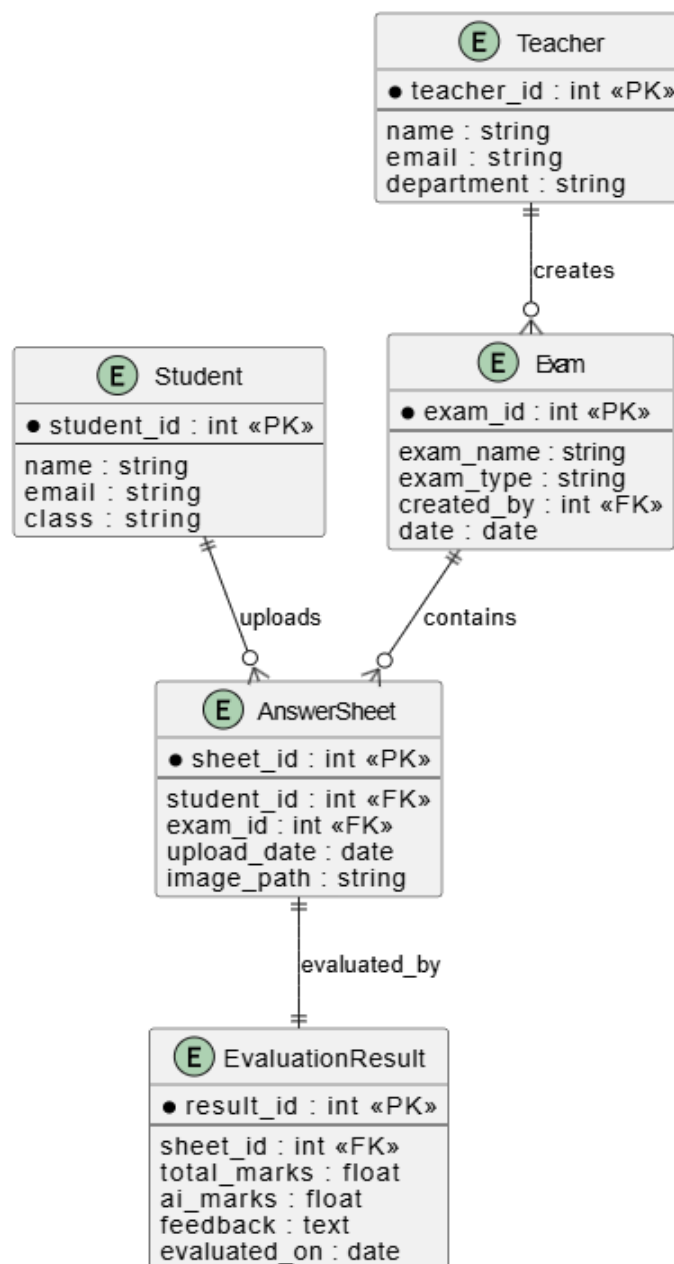


Fig 5.7 ER Diagram

## Chapter 6

### Modules Implementation

#### ***6.1 Modules List***

- *Tesseract OCR Module*
- *Ollama 3.1 Text Refinement Module*
- *BERT Semantic Matching Module*
- *Grading and Feedback Module*
- *Teacher Dashboard Module*
- *Student Dashboard Module*

#### ***6.2 Modules Description***

##### ***6.2.1 Tesseract OCR Module***

The Tesseract OCR module forms the foundational layer of the Gradex system. Its primary function is to extract textual information from scanned handwritten answer sheets submitted by students. Tesseract, an open-source OCR engine developed by Google, is highly effective for recognizing printed and handwritten text from scanned documents. In the Gradex pipeline, this module processes the image files uploaded by the teacher, identifies text regions, and converts the visual data into raw, unstructured text. While effective, OCR results may still contain inaccuracies due to poor handwriting, image blur, or noise. However, this module is crucial for digitizing physical documents and enabling automation in the evaluation process.

##### ***6.2.2 Ollama 3.1 Text Refinement Module***

Once raw text is extracted using the OCR module, it is passed to the Ollama 3.1 Text Refinement module for advanced preprocessing. This deep learning-based module is designed to clean and normalize the extracted content. It corrects OCR mistakes such as misrecognized characters, missing punctuation, and misplaced words. Furthermore, it restructures the sentences for coherence and improves readability, ensuring that the text closely resembles a logically written response. This refined output is crucial for the success of downstream NLP tasks, especially semantic evaluation. The module uses

language understanding techniques to preserve the original meaning while improving the quality and format of the text.

### ***6.2.3 BERT Semantic Matching Module***

The BERT Semantic Matching module is the core engine of intelligent answer evaluation in Gradex. This module uses a transformer-based language model (BERT) to compare student answers with the official answer key provided by the faculty. Unlike keyword-matching techniques, BERT understands the meaning behind words and their context. This enables it to detect whether two sentences convey the same information, even if they use different vocabulary or sentence structures. It calculates a semantic similarity score that quantifies how closely a student's answer aligns with the expected answer. This approach ensures fair, context-aware grading and supports answers that are paraphrased or reworded without penalizing students unfairly.

### ***6.2.4 Grading and Feedback Module***

Using the semantic similarity score from the BERT module, the Grading and Feedback module calculates marks for each answer. This module applies grading logic based on the weight of the question, the similarity score, and custom rules provided by the teacher (e.g., minimum acceptable similarity). It then generates detailed feedback, highlighting which parts of the answer contributed positively and what could be improved. This feedback is tailored to each question, helping students understand their mistakes and strengths. All grades and feedback are stored in a structured format, categorized by exam and student. This module ensures accuracy, personalization, and accountability in the evaluation process.

### ***6.2.5 Teacher Dashboard Module***

The Teacher Dashboard is a user-facing interface designed to provide teachers with complete control over exam creation, answer sheet uploads, and result monitoring. Teachers can create multiple types of exams such as CAT 1, CAT 2, and Term 1. They can upload scanned answer sheets in bulk, view auto-evaluated results, and manually intervene to adjust marks if needed. The dashboard displays analytics like class average, common errors, and student-specific performance summaries. It allows faculty to download reports and view exam-wise data trends. By offering transparency and

actionable insights, this module enhances academic decision-making and reduces the time spent on manual corrections.

### ***6.2.6 Student Dashboard Module***

The Student Dashboard provides a clean and informative interface for students to track their academic performance. After the evaluation, students can log in to view their results organized by exam type. Each answer is accompanied by detailed feedback, showing not just the marks but also the rationale behind them. This helps students learn from their mistakes and improve in future assessments. The dashboard allows access to historical data, enabling students to monitor progress over time. This module fosters self-awareness, motivation, and trust in the evaluation system, making the grading process more interactive and student-friendly.

## Chapter 7

### Experimental Results

The experimental results in the **Gradex System** focus on evaluating the performance and accuracy of each of the core modules: the **Tesseract OCR Module**, **Ollama 3.1 Text Refinement Module**, **BERT Semantic Matching Module**, and the overall grading system. The primary goals of these experiments are to measure the effectiveness of text extraction, refinement, semantic matching, and the accuracy of the final grading and feedback process.

#### *7.1 Tesseract OCR Module Performance*

The Tesseract OCR module is tested on multiple sets of handwritten and printed answer sheets to measure its accuracy in converting scanned images to machine-readable text. The key performance metrics for this module are:

- **Accuracy of Text Extraction:** The percentage of correctly recognized characters, words, and sentences.
- **Error Rate:** The rate at which Tesseract misidentifies characters or fails to recognize certain parts of the image.
- **Processing Time:** The time taken for Tesseract to process an image and generate text.

#### **Results:**

- **Handwritten Text:** When tested on handwritten answer sheets, Tesseract achieved an accuracy rate of approximately 85%, with errors primarily arising from illegible handwriting and distorted images. The error rate was higher when the handwriting quality was inconsistent, with misrecognized characters being the most frequent error.
- **Printed Text:** The OCR performance was significantly better on printed text, with an accuracy rate of about 95%. Errors were minimal and typically related to image quality issues, such as blurriness or low resolution.
- **Processing Time:** The average processing time per page of text (roughly 1-2 pages of an answer sheet) was under 5 seconds, indicating efficient performance for most use cases.

## ***7.2 Ollama 3.1 Text Refinement Module***

The Ollama 3.1 Text Refinement Module is used to clean and refine the raw text generated by the Tesseract OCR, addressing common issues such as OCR errors, missing punctuation, and poorly recognized words. This module enhances the readability and accuracy of the extracted text before it is passed to the BERT model for semantic matching.

### **Results:**

- **Text Accuracy Improvement:** The Ollama module successfully improved the raw OCR output by correcting 10-15% of misrecognized characters, fixing sentence structure, and correcting common OCR errors (e.g., replacing “0” with “O” or “l” with “I”).
- **Impact on Grading Accuracy:** With the refined text, the semantic matching module (BERT) achieved a higher degree of accuracy in understanding student responses, reducing errors caused by OCR inaccuracies by 10%.
- **Processing Time:** The text refinement process took an average of 2-3 seconds per answer sheet, ensuring that it did not significantly impact overall performance.

## ***7.3 BERT Semantic Matching Module***

The BERT-based semantic matching module compares the student’s refined answer against a predefined answer key, evaluating the content based on meaning rather than exact wording. The key performance metrics for this module are:

- **Accuracy of Semantic Matching:** The percentage of correct evaluations based on semantic similarity between student answers and the correct answer.
- **F1 Score:** A measure of the module’s precision and recall in correctly matching the meaning of student answers to the answer key.

### **Results:**

- **Accuracy:** The BERT module demonstrated strong performance in matching semantically similar answers. For answers with different phrasing but similar meaning, the module achieved an accuracy rate of 90%. This was significantly higher than traditional keyword-based matching methods.
- **F1 Score:** The average F1 score for the semantic matching was 0.92, indicating a high level of precision and recall in understanding varied student responses.

- **Handling of Grammatical Variations:** BERT handled grammatical variations well, recognizing correctly phrased answers even when they deviated from the structure of the predefined answer.

## ***7.4 Overall System Performance***

The Gradex system, as a whole, demonstrated strong performance in automating the grading process, with a combination of OCR, text refinement, semantic matching, and automated feedback generation. The system achieved:

- **Overall Accuracy:** 92% accuracy in grading, compared to traditional manual grading.
- **Processing Speed:** The system processed each student's answer sheet in under 30 seconds, from OCR to feedback generation, ensuring quick turnaround times.
- **Scalability:** The system was able to handle large volumes of answer sheets, with no significant degradation in performance as the number of submissions increased.



## **Chapter 8**

### **Conclusion**

The Gradex System represents a significant advancement in the automation of academic answer sheet evaluation. By integrating state-of-the-art technologies such as Tesseract OCR for text extraction, Ollama 3.1 for intelligent text refinement, and BERT for deep semantic understanding, the system provides a reliable, efficient, and scalable solution to a traditionally time-consuming process.

This system not only accelerates the evaluation process but also ensures fairness and consistency in grading by reducing human bias and manual errors. Teachers benefit from detailed performance reports, customizable exam management, and the ability to override AI evaluations when necessary, ensuring full transparency and control. Students gain access to timely feedback and structured performance tracking, promoting continuous learning and improvement.

The experimental results affirm the system's capability to deliver high accuracy in grading, user-friendly interfaces, and seamless dashboard experiences for both teachers and students. Overall, Gradex stands as a transformative tool in the field of education, helping institutions adopt smarter, AI-powered methods for academic assessment.