**SwiftGrade: Integrating OCR and NLP in an AI-Assisted Quiz and Exam Management System for College Educators**

FRANCIS DANIEL D. DE VERA

KENNY WALTER P. DURAN

ERVIN JOHN S. LIBARDOS

UNDERGRADUATE THESIS PROPOSAL SUBMITTED TO THE FACULTY OF SOFTWARE DEVELOPMENT, THE DEPARTMENT OF INFORMATION TECHNOLOGY COLLEGE OF INFORMATION SCIENCES AND COMPUTING, CENTRAL MINDANAO UNIVERSITY, IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY

MAY 2025

# Table of Contents

[Table of Contents 2](#_Toc195367439)

[CHAPTER 1 4](#_Toc195367440)

[1.1 Background of the Study 4](#_Toc195367441)

[1.2. Statement of the Problem 5](#_Toc195367442)

[1.3. Objectives of the Study 5](#_Toc195367443)

[1.4. Scope and Limitations of the Study 6](#_Toc195367444)

[1.5. Significance of the Study 7](#_Toc195367445)

[CHAPTER 2 8](#_Toc195367446)

[2.1. Review of Related Concepts 8](#_Toc195367447)

[2.1.1. Technology Integration in Education 8](#_Toc195367448)

[2.1.2. Automated Assessment Systems 10](#_Toc195367449)

[2.2. Review of Related Systems 14](#_Toc195367450)

[2.2.1 Optical Character Recognition (OCR) in Academic Settings 14](#_Toc195367451)

[2.2.2 Natural Language Processing (NLP) for Essay and Short Answer Evaluation 16](#_Toc195367452)

[2.3. Synthesis and Gap Analysis 19](#_Toc195367453)

[CHAPTER 3 20](#_Toc195367454)

[3.1. Flutter (Dart) 20](#_Toc195367455)

[3.2. Firebase 21](#_Toc195367456)

[3.3. OCR and NLP Engines – Automated Data Recognition and Essay Evaluation 21](#_Toc195367457)

[3.4. GitHub/SourceForge 22](#_Toc195367458)

[3.5. Coding Environments – WSL, VS Code, and Sublime Text 23](#_Toc195367459)

[3.6. Python, and its Machine Learning Libraries 24](#_Toc195367460)

[CHAPTER 4 25](#_Toc195367461)

[4.1. Conceptual Framework 25](#_Toc195367462)

[4.2. Research Approach 26](#_Toc195367463)

[4.3. System Developmental Methodology 26](#_Toc195367464)

[4.3.1 Requirements Gathering and Analysis 27](#_Toc195367465)

[4.3.2 System Design 27](#_Toc195367466)

[4.3.3 Implementation 28](#_Toc195367467)

[4.3.4 Integration and Testing 28](#_Toc195367468)

[4.3.5 Deployment of System 28](#_Toc195367469)

[APPENDICES AND FIGURES 29](#_Toc195367470)

[BIBLIOGRAPHY: 29](#_Toc195367471)

# CHAPTER 1

INTRODUCTION

## 1.1 Background of the Study

In all over the world, computers and information technology are causing a revolution in education systems, these changes claiming They have numerous benefits, including lowering costs on education and learning and having global reach (Bahel et al., 2019). One such Advantage is that: Up Until now, these new generation of e-learning technologies has proven useful in a variety of fields. Interactive Lessons, Video Tutorials and even Digital Module Based Learning. Another improvement is manifested during evaluations, Web-Based Tests and Online examinations are claimed to be more accessible and flexible than traditional assessment methods. Research made by Elsayed et al. last 2013, has quoted that “...They can minimize the overall expenses of exams and in examination processing (checking and recording) especially as they save papers, storage, and materials’ costs”. This proves that reforms for Technology in Education started gaining attention and focus decades ago, and has been constantly improving itself.

Among these tools, ZipGrade stands out as a practical solution for automating the grading of multiple-choice assessments. Launched in 2013, ZipGrade quickly gained popularity among educators for its cost-effective and efficient grading capabilities. Designed for them, it leverages Optical Character Recognition (OCR) technology to scan and score answer sheets instantly, reducing manual grading time and minimizing human error. Its user-friendly interface and mobile compatibility make it accessible for both in-class and remote assessments. However, despite its advantages, ZipGrade has limitations, such as reliance on printed answer sheets, limited support for complex question formats, and the need for high-quality scans to ensure accurate results.

## 1.2. Statement of the Problem

Despite the adaptation and digitalization of lessons and tests, Our Education Systems almost failed and was tested severely during the pandemic period. This change in paradigm, made the teaching field face several tough situations. Some pain points of the ZipGrade System, pointed by (Arora & Al-Hattami, 2025) include:

1. The Grading engine is optimized for multiple-choice question (MCQ) formats making a Format Lock. As the core functionality exhibits a degree of format inflexibility.
2. Academic IT challenges like: sufficient device availability for teachers and a requirement of having a requisite level of digital literacy
3. While immediate scoring is a key feature, the inherent feedback mechanism is largely quantitative.

These points highlight key technical considerations that institutions and educators must address to ensure effective and robust implementation of the ZipGrade system.

## 1.3. Objectives of the Study

This project aims to improve the process of academic assessments by integrating current AI technologies such as Optical Character Recognition (OCR) and Natural Language Processing (NLP).

Specifically, the project does:

1. To develop an AI-assisted quiz and exam management system using Optical Character Recognition (OCR) for automatic multiple-choice grading. And Natural Language Processing (NLP) to evaluate short-answer and essay-type questions.

2. To enhance the efficiency and accuracy of grading for educators by Designing an intuitive and user-friendly interface accessible across platforms, potentially minimizing grading time by automating the assessment process

3. To provide transparency to students, promoting faster learning outcomes by using smart analytics on student performance and NLP-generated feedback.

4. To ensure the system's adaptability and scalability, Supporting deployment in colleges, universities, and other educational institutions. And even allowing future enhancements

## 1.4. Scope and Limitations of the Study

The scope of this project encompasses the development and deployment of an AI-assisted quiz intended for use by college educators only, not for K-to 12 education and lower years or corporate training. Utilizing Optical Character Recognition (OCR) for automatic grading of multiple-choice questions and Natural Language Processing (NLP) for Essay-like questions. The Project does not include features for live proctoring or academic integrity enforcement. Initial deployment is also limited to institutions with modern digital infrastructure, ensuring optimal performance and support.

Other Identified Limitations include:

1. The OCR functionality requires high-quality image scans for optimal accuracy, which may pose challenges in low-light or low-resolution scenarios. At the worst case, hardware constraints can hinder the performance of the system.
2. NLP evaluation may encounter difficulties in interpreting highly creative, contextually ambiguous, or culturally specific responses (contractions and vernacular corpus) as the system will be primarily designed for English-language assessments, with no support functionality for other languages.
3. The system's AI models may require periodic updates to maintain accuracy and performance.

## 1.5. Significance of the Study

The proposed system for detecting anomalous documents is crucial for academic institutions because it addresses key challenges like document authenticity, operational efficiency, and institutional reputation. This Claim itself can prove the huge significance of our project, as Academic registrars and administrative staff often face time-consuming and error-prone manual document verification processes. Outdated methods leave institutions vulnerable to document fraud and manipulation which we wanted to address. By using technologies like Optical Character Recognition (OCR) and Natural Language Processing (NLP), the system aims to make document verification more efficient, reduce errors, and ensure the authenticity of academic records.

Additionally, this study contributes to the growing field of automation and machine learning in administrative tasks. The development of this system can serve as a model for other institutions seeking to “modernize” their document verification processes. If the project gains traction, we can apply the application to scopes beyond education. Since the methods explored here could benefit industries like government and finance, where secure document handling is equally critical. Ultimately, this project not only meets the needs of academic institutions but also lays the foundation for scalable and user-friendly solutions for document authentication.

# CHAPTER 2

**REVIEW OF RELATED LITERATURE**

## 2.1. Review of Related Concepts

### 2.1.1. Technology Integration in Education

Research Findings proved that technology integration in education is significant (Lainez et al., 2024). Several other sources have discussed the development and application of automated systems for grading examinations, with the goal of reducing manual effort, improving accuracy, and providing faster feedback. Technology Stacks like Optical Mark Recognition (OMR), Optical Character Recognition (OCR), Natural Language Processing (NLP), machine learning, and even deep learning are utilized in this field (Rahaman & Mahmud, 2022).

One such technology: OMR, is highlighted as a useful data entry tool for capturing human-marked data from documents like surveys and tests. It can advance the checking of multiple-choice exams, then stores results in databases, which then display scores, and generate reports. The technology is cost-effective and accurate for processing large volumes of answer sheets, with the effect of reducing manual effort and increasing grading speed. However, challenges such as initial setup costs, technical issues, and the need for staff training, as well as poor document quality, are heavily noted in its disadvantages (Lainez et al., 2024). Webcam-based OMR solutions are also proposed to promote even more access to automated grading for institutions lacking specialized equipment (Singh K. et al., 2024).

Handwritten answers and more complex question types, the proposed approach is the combination of both OCR and NLP. OCR converts handwritten text into digital format, which can then be fed and analyzed via NLP techniques to understand the semantic meaning and assess the correctness of the answers (Kulkarni et al., 2024). NLP Techniques like cosine similarities are used to assess student answers with modeled answers based on key terms rather than just answer length (Kulkarni et al., 2024). Other Deep learning architectures are even combining CNNs and BiLSTMs are proposed for handwritten answer recognition and grading (Rahaman & Mahmud, 2022).

The development of OCR, Android-based test paper checking applications using and machine learning is also presented as a way to automate and streamline the grading process in educational institutions, providing immediate results and offline functionality (Masula et al., 2024).

AI-powered systems, such as the use of machine learning and deep learning, are being developed to automate several aspects of exam evaluation, such as grading, feedback, and plagiarism detection as pointed by Reddy & Chauhan, (2024).

Integrating technology in education, mainly through automated assessment systems, it offers numerous benefits, with increased efficiency in grading (Kulkarni et al., 2024), scores higher accuracy compared to manual grading, having a reduction in manual effort and workload for educators (Singh K. et al., 2024), faster feedback for students, having more objectives and constant evaluations, and the ability to process large volumes of answer sheets efficiently (Reddy & Chauhan, 2024).

Automated systems can also deliver detailed reports and statistical analysis of exam results, giving valuable insights for educators (Lainez et al., 2024). It can also streamline administrative tasks and enhance the overall educational experience (Kulkarni et al., 2024).

Implementing technology in field of education, especially automated assessment systems, also presents challenges. Some of it are: initial setup costs (Domeh et al., 2023), potential technical issues, the need for staff knowledge, dealing with poor document quality and variations in marks (Domeh et al., 2023).

AI-powered grading, relying in the AI's decision-making process and addressing potential biases in training data are one of its crucial considerations. It needs to incorporate human assessment together with automated systems highlighted (Domeh et al., 2023).

In summary, the sources indicate a strong trend towards integrating various technologies, particularly AI, NLP, and OCR, into educational settings automating and improving assessment processes. The integration aims to enhance efficiency, accuracy, and the overall learning experience, while also acknowledging and addressing both its disadvantages and challenges.

### 2.1.2. Automated Assessment Systems

Automated Assessment Systems (AAS) are technologies gaining traction within the education to streamline the evaluation process, improving accuracy, and reducing the burden of manual grading (Deepak et al., 2024). These systems leverage several technologies to automatically assess student work, offering a departure from traditional, time-consuming manual methods as proven by Agarwal et al. (2021).

OMR Systems are designed for grading multiple-choice and objective-type exams where students mark bubbles or boxes on answer sheets (Reddy & Chauhan, 2024).

One such feature is functionality, OMR systems can scan answer sheets, and compare marked responses to a predefined answer key, calculate scores, store results, and create reports (Reddy & Chauhan, 2024). Some systems can even execute error correction and handle regional deformations (Singh K. et al., 2024).

One perks of OMR is speed and high accuracy for standardized tests, reducing human error in marking and allowing for the efficient processing of massive volumes of answer sheets (Reddy & Chauhan, 2024). Webcam-based OMR aims to provide a cost-effective alternative for institutions lacking specialized scanning equipment (Domeh et al., 2023).

Some downsides of OMR are primarily limited to structured answer formats like multiple-choice questions and cannot evaluate open-ended or essay-type answers (Reddy & Chauhan, 2024). One such challenge is inclusion of the needs for specific answer sheet formats, possible issues with crumpled or poorly marked sheets, and the initial cost of specialized equipment (yet webcam-based solutions address the latter) (Domeh et al., 2023).

Based Systems such as OCR and NLP, these systems are designed for evaluating handwritten and textual answers (Deepak et al., 2024). OCR technology converts handwritten or typed text into machine-readable digital text, which is then analyzed using NLP techniques (Deepak et al., 2024).

Another is its Functionality, these systems can process scanned answer sheets, then extract textual content, and compare student responses to model answers based on semantic similarity (various algorithms like cosine similarities), identify keywords, and assess word grammar and spelling. Some advanced systems could even aim to handle diagrams and other non-textual elements (Deepak et al., 2024).

The perks of OCR+NLP systems support unstructured, handwritten answer sheets, allowing for more flexible examination formats and the evaluation of open-ended questions (Reddy & Chauhan, 2024). It can provide more nuanced grading by understanding the meaning of the text, rather than just looking for particular matches (Agarwal et al., 2021). These systems also offer advantages like reduced grading time, improved consistency, and having potential for detailed feedback (Deepak et al., 2024).

However, the limitations include its accuracy. OCR can be affected by handwriting quality, requiring careful-preprocessing for reliable results. Assessing open-ended questions can also be subjective, necessitating clear grading rubrics and potentially the integration of machine learning models trained on human-graded data (Reddy & Chauhan, 2024).

Another AAS leverages Artificial Intelligence (AI) and Machine Learning (ML) Powered Systems: These systems include deep learning (CNN, RNN, LSTM, BiLSTM), with the goal of automating exam evaluation (Agarwal et al., 2021).

One of the AI-driven systems several functions is to perform tasks such as handwritten text recognition (OCR), essay scoring, grading based on semantic similarity, identifying grammatical errors (NLP), and even generating questions (LLMs) (Rahaman & Mahmud, 2022). Some systems utilize Transformer-based models like TrOCR for OCR and GPT models for answer evaluation. Parallel processing techniques (OpenMP, CUDA) can be combined to augment speed and manage workload distribution (Reddy & Chauhan, 2024).

The Advantages of AI and ML, offers the potential for highly accurate and efficient automated assessment, capable of learning complex grading rubrics from huge datasets. It can provide consistent and objective evaluations, reduce bias, and offers personalized feedback to students (Domeh et al., 2023). These systems can also handle huge datasets and scale to high demands (Reddy & Chauhan, 2024).

Developing and training AI/ML models limits to require substantial data and computational resources (Deepak et al., 2024). It includes the challenges of ensuring the accuracy of grading for open-ended questions requiring analysis and understanding of its context, addressing the variability of handwriting styles (Rahaman & Mahmud, 2022), ensuring the explainability and fairness of AI-driven evaluations by mitigating potential biases in training data. Integration of these systems with existing educational infrastructure (like Learning Management Systems - LMS) can also present challenges (Kulkarni et al., 2024).

One factor is increasing efficiency, It significantly reduces the time and effort required for grading, allowing educators to concentrate on other tasks (Singh K. et al., 2024).

Minimizing human error and subjectivity in grading, leading to more consistent and fair evaluations to improved accuracy and reliability (Domeh et al., 2023).

It faster the feedback for students which enables quicker release of results, facilitating timely learning and improvement (Rahaman & Mahmud, 2022).

Capable of handling huge volumes of answer sheets efficiently, suitable for large-scale assessments (Deepak et al., 2024).

Cost-Effectiveness: Reduces the costs associated with manual grading, such as manpower and resources (especially with webcam-based OMR and applications using readily available hardware) (Agarwal et al., 2021).

Data Analysis and Insights: Provides detailed reports and statistical analysis of exam results, offering valuable insights into student performance and question effectiveness (Kulkarni et al., 2024).

Initial Setup and Integration Costs: Implementing new systems and integrating them with existing infrastructure can involve initial financial investment (Lainez et al., 2024).

Technical Issues and Maintenance: Requires technical expertise for setup, maintenance, and troubleshooting (Domeh et al., 2023). Handwriting Variability: OCR accuracy can be significantly affected by the diversity of handwriting styles (Deepak et al., 2024).

Subjectivity in Open-Ended Questions: Fully automating the grading of complex, open-ended questions remains challenging and may require human oversight (Agarwal et al., 2021).

Bias in AI Models: AI-powered systems can inherit biases from their training data, potentially leading to unfair evaluations (Domeh et al., 2023).

Need for Clear Rubrics and Training Data: Effective automated assessment, especially for subjective questions and AI-based systems, relies on well-defined grading criteria and large, high-quality training datasets (Reddy & Chauhan, 2024).

Security and Data Integrity: Ensuring the security and integrity of the examination process and student data is crucial (Kulkarni et al., 2024).

In conclusion, automated assessment systems represent a significant advancement in educational technology, offering numerous benefits for educators and students. While challenges exist in their implementation and further development, the continued progress in technologies like OMR, OCR, NLP, and AI holds great promise for transforming the landscape of educational assessment (Domeh et al., 2023).

## 2.2. Review of Related Systems

This section explores existing research and applications of Optical Character Recognition (OCR) in academic settings and Natural Language Processing (NLP) for essay and short answer evaluation, drawing upon the provided sources.

### 2.2.1 Optical Character Recognition (OCR) in Academic Settings

Optical Character Recognition (OCR) technology plays a crucial role in modernizing academic processes, particularly in the context of assessment and data management (Kulkarni et al., 2024). It involves converting images of handwritten or printed text into machine-readable digital text (Deepak et al., 2024). In academic settings, OCR is primarily used for processing answer sheets, thereby enabling automated evaluation and reducing the reliance on manual grading (Deepak et al., 2024).

Automated Grading of Examinations: OCR is a fundamental component of automated examination correction systems (Reddy & Chauhan, 2024). It allows for the extraction of text from scanned answer sheets, whether handwritten or printed, making it possible for subsequent analysis and grading by software (Deepak et al., 2024).

Processing Multiple-Choice Questions (MCQs): While Optical Mark Recognition (OMR) is traditionally used for MCQs, OCR can also play a role, particularly when student identification numbers or other textual information are handwritten on OMR sheets (Maniar et al., 2021). Some systems integrate both OCR and image processing for comprehensive MCQ evaluation (Maniar et al., 2021).

Digitization of Answer Scripts: OCR facilitates the conversion of physical handwritten answer sheets into digital formats, which can then be stored, analyzed, and managed electronically (Deepak et al., 2024). This addresses the logistical challenges of storing and handling large volumes of physical papers and mitigates risks like damage (Singh K. et al., 2024).

Integration with NLP for Comprehensive Assessment: Combining OCR with Natural Language Processing (NLP) allows for the evaluation of not just the presence of keywords but also the semantic understanding and coherence of student responses, particularly for short answer and essay questions (Deepak et al., 2024).

Accuracy with Handwritten Text: The accuracy of OCR can be significantly impacted by the variability in handwriting styles and the quality of the scanned images (Deepak et al., 2024). Poor handwriting, variations in letter formation and size, and noise in the scanned document can lead to misinterpretations (Rahaman & Mahmud, 2022).

Preprocessing Requirements: To improve OCR accuracy, significant preprocessing of the scanned images may be required. This can include noise reduction, skew correction, binarization, and character segmentation (Masula et al., 2024).

Handling Non-Textual Elements: Traditional OCR systems may struggle with non-textual elements present in answer scripts, such as diagrams, mathematical formulas, and figures. More advanced systems are being developed to address this (Deepak et al., 2024).

Integration with Existing Systems: Integrating OCR technology with existing educational databases and Learning Management Systems (LMS) can present technical challenges and may require careful planning (Reddy & Chauhan, 2024).

Deep Learning-Based OCR: Recent advancements in deep learning have led to significant improvements in OCR accuracy, particularly for handwritten text recognition (Masula et al., 2024). Architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks, are being used to build more robust OCR models (Masula et al., 2024).

### 2.2.2 Natural Language Processing (NLP) for Essay and Short Answer Evaluation

Natural Language Processing (NLP) offers powerful techniques for the automated evaluation of textual responses in academic assessments, going beyond simple keyword matching to understand the meaning and quality of student writing (Deepak et al., 2024).

Semantic Similarity Measures: Algorithms like cosine similarity, Jaccard similarity, bigram similarity, and synonym similarity are used to determine how semantically similar a student's answer is to a model answer (Agarwal et al., 2021). These methods can assess if the student has conveyed the correct meaning, even if the exact wording differs (Agarwal et al., 2021).

Keyword Extraction and Matching: NLP techniques can identify key concepts and terms in both student answers and model answers, and evaluate the student's response based on the presence and context of these keywords (Agarwal et al., 2021).

Grammatical and Spelling Error Detection: NLP tools can analyse student writing for grammatical errors, spelling mistakes, and punctuation issues, providing feedback and contributing to the overall score (Rahaman & Mahmud, 2022).

Coherence and Entailment Analysis: More advanced NLP models can assess the logical flow, coherence, and understanding of the underlying concepts demonstrated in a student's response (Deepak et al., 2024).

Sentiment Analysis: While less directly applicable to factual answers, sentiment analysis can be used in evaluating argumentative or opinion-based essays to understand the student's stance and reasoning (Szabó Nagy & Kapusta, 2023).

Text Summarization: NLP techniques can summaries student answers and compare them to summaries of model answers to assess comprehension and conciseness (Agarwal et al., 2021).

Word Embeddings and Deep Learning: Techniques like Word2Vec, ELMo, and Transformer models (like GPT) are used to create vector representations of words and sentences, capturing semantic relationships and allowing for more sophisticated similarity comparisons and essay scoring (Telenyk et al., 2021). Recurrent Neural Networks (RNNs), including LSTMs and BiLSTMs, are particularly useful for processing sequential text data and evaluating the overall structure and content of essays (Telenyk et al., 2021).

Improved Accuracy and Objectivity: NLP can provide a more consistent and objective evaluation of textual answers compared to manual grading, reducing biases and inconsistencies (Deepak et al., 2024).

Efficiency and Scalability: Automated NLP-based systems can process a large number of answer scripts quickly, significantly reducing the time and effort required for grading (Singh K. et al., 2024).

Detailed Feedback: NLP tools can provide students with specific feedback on grammar, spelling, and the presence of key concepts, facilitating learning and improvement (Rahaman & Mahmud, 2022).

Assessment of Deeper Understanding: NLP techniques can go beyond surface-level keyword matching to assess the semantic understanding and coherence of student responses (Deepak et al., 2024).

Requirement for Large Training Datasets: Developing accurate NLP models for essay scoring and short answer evaluation often requires large datasets of human-graded answers for training (Deepak et al., 2024).

Defining Effective Grading Rubrics: Clear and well-defined grading rubrics are essential for developing and evaluating NLP-based assessment systems (Deepak et al., 2024).

Addressing Different Writing Styles: NLP models need to be robust enough to handle the diverse writing styles and grammatical variations present in student responses (Kulkarni et al., 2024).

Integration with OCR: When evaluating handwritten answers, the accuracy of the preceding OCR process directly impacts the effectiveness of the NLP analysis. Errors in OCR can cascade and affect the NLP results (Deepak et al., 2024).

In conclusion, both OCR and NLP are critical technologies in the development of automated assessment systems. OCR provides the means to digitise textual data from answer scripts, while NLP offers the tools to analyse and evaluate the content of these responses in a more sophisticated and automated manner. The integration of these technologies, along with advancements in machine learning and deep learning, is driving the evolution of more efficient, accurate, and comprehensive automated assessment solutions in academic settings (Reddy & Chauhan, 2024).

## 2.3. Synthesis and Gap Analysis

Current literature and existing systems have settled for good enough—basic answer-checking through bubbles, fragmented app performance across devices, and a near-total disregard for open-ended evaluation.

ZipGrade, once considered cutting-edge, now shows its age through its stagnation: no built-in NLP intelligence, no real-time cloud integration, and no significant innovation in adapting to the dynamic needs of modern college instruction. What’s worse—many of these tools remain tied to rigid infrastructures that suffocate scalability and experimentation.

This project exists to do what others haven’t: to unify machine-read assessment with intelligent essay evaluation, and to wrap it all inside a sleek, multi-platform system built for speed, scale, and extensibility. Where others stop at scanning sheets, we inject machine learning. Where others fear multi-device complications, we embrace Flutter. We are not just proposing a research solution—we are planting the first iteration of a system that can outpace, outscale, and outthink its predecessors. The academic sector may call it a prototype, but we already know it as infrastructure for a disruption.

# CHAPTER 3

**TECHNICAL BACKGROUND**

In the development of the project, the proponents are carefully selected from set of tools and technologies to ensure efficient and effective development. This section provides an overview of the critical components of our technical stack. Both Primary (Back, Middleware and Front-Ends), and the Tools used/Secondary Stack:

These are to be our Primary Tech Stack:

### 3.1. Flutter (Dart)

Flutter, developed by Google, is a modern UI framework that allows developers to build natively compiled applications across mobile, web, and desktop from a single codebase. Utilizing Dart as its primary programming language, Flutter supports reactive-style programming and offers rich libraries for building clean, responsive, and scalable user interfaces. This framework simplifies development across multiple platforms, reducing the need for maintaining separate codebases for each environment. Its built-in widgets and performance-optimized rendering engine make it ideal for applications requiring smooth, real-time interaction.

Flutter was selected due to its maturity, active open-source community, and proven scalability in building multi-platform educational tools which making it a practical choice for a project intended to evolve into a deployable suite across various devices.

Its relevance to the project lies in its ability to maintain a consistent user experience across instructor devices, enabling broader accessibility and adoption of the system within college-level academic settings.

### 3.2. Firebase

Firebase is another Google-managed Backend-as-a-Service (BaaS) that provides scalable cloud functions, real-time databases, authentication services, analytics, and hosting which all integrated into one secure environment. It supports modular integration for mobile and web platforms.

Firebase is selected for its modular, cloud-based architecture and robust documentation which enables rapid development and security compliance without sacrificing flexibility.

It is particularly relevant to the project’s requirement for secure user authentication, fast data retrieval, and seamless synchronization of exam records and results, thus improving both the efficiency and reliability of exam processing in an academic environment.

### 3.3. OCR and NLP Engines – Automated Data Recognition and Essay Evaluation

The Optical Character Recognition (OCR) module is critical for automating the recognition of handwritten or printed answers from scanned exam sheets. Which, Paired with Natural Language Processing (NLP) engines, the system is designed to semantically analyze and grade student essays based on coherence, grammar, and rubric-based scoring mechanisms.

The project considers the following models for its OCR+NLP Engines for the OCR, we consider Tesseract (v5.3.1 LTS) as this Open-source OCR engine has strong community support and multilingual recognition, this is to anticipate both Bubble Answer Sheet Assessment and Essay Assessment Module. Additional sideloading with the model includes *tesseract.js* or Flutter wrapper flutter\_tesseract\_ocr (v1.0.0) to integrate with the UI part of the project.

The NLP module for the essay assessment will be that of the BERT (Bidirectional Encoder Representations from Transformers) Famly, specifically: bert-base-uncased (v4.40.0).

Other Modules together with BERT is is tokenizers module(v0.19.1).. This module helps eliminate those commonly connecting words like liking verbs (A, An, The, But…).

Lexical Modules, like understanding Sentence similarity and rubric scoring are supported via sentence-transformers (v2.6.1), these helps in the interpretation and Sematic Understanding of Sentences.

While the specific engines are under evaluation (with open-source models such as Tesseract OCR and BERT-based language models as strong candidates), the combined implementation will allow for robust, automated exam evaluation for both objective and subjective question formats.

This hybrid automation stack is chosen due to its proven feasibility in academic contexts and the strong open-source ecosystem surrounding them that offers pre-trained support, extensive documentation, and reproducible pipelines which critical for a research project where explainability and benchmarking matter.

In Tandem with these Modules, our Secondary Tech Stack for the project also includes that of the Applications used to code and debug the modules, they include:

## 3.4. GitHub/SourceForge

GitHub serves as the primary code repository and version control platform for the project. Its capacity to provide robust source control via Git, allows collaborative development, branching, and a faster issue tracking.

The repository will be configured, with protected main branches, GitHub Actions for CI/CD (continuous integration and deployment), and automated test runners for critical modules. It also ensures reproducibility and traceability of commits, which is crucial in academic development settings.

The researches opted to use this repository because of its GitHub was chosen for its mature ecosystem, institutional familiarity, and strong integration with tools like Visual Studio Code, GitHub Classroom, and GitHub Copilot which making it ideal for both academic tracking and professional scalability. ensuring structured, collaborative development within the team and allows secure code backup, easy rollbacks, and development pipeline automation. Which becomes important especially when dealing with multiple grading modules being integrated asynchronously.

In addition to the project's primary repository hosted on GitHub, SourceForge is considered as a complementary distribution and archival platform. SourceForge has long been recognized as a reliable repository for open-source software, offering robust version control, download tracking, and project visibility in academic and enterprise environments. Its long-standing presence in the open-source ecosystem allows for broader reach and platform-agnostic collaboration, especially in regions or institutions where GitHub access is restricted or discouraged. By utilizing SourceForge, the project ensures an additional layer of repository redundancy and accessibility for both current collaborators and future stakeholders. This platform acts as an alternative fallback that supports community access and project mirroring.

The decision to include SourceForge as part of the repository infrastructure reflects our vision of making the system openly available to a wider audience, including educators and developers who operate within more traditional or compliance-heavy institutions. In tandem with GitHub, it reinforces a commitment to transparency, longevity, and adaptability—traits that are not only valuable in academic research but also vital in preparing the system for potential startup-grade scaling. Through this dual-repository approach, SwiftGrade positions itself as both academically grounded and operationally future-proof.

## 3.5. Coding Environments – WSL, VS Code, and Sublime Text

The main development environment revolves around Windows Subsystem for Linux (WSL) integrated with Visual Studio Code (VS Code).

WSL provides a lightweight Ubuntu-like environment for Python and machine learning dependencies without leaving Windows. And, with Visual Studio Code is used as the primary editor due to its wide extension support, integrated terminal, and debugging tools.

In case of system-specific performance limitations or minimalist coding needs, Sublime Text (v4.0) serves as the backup environment.

This tri-environment configuration offers flexibility and resilience. VS Code and WSL allow running Python-based ML models without virtual machines, while Sublime serves as a lightweight fallback for quick edits or unstable systems. The combination allows the project to achieve seamless integration of Python-based ML models, cross-platform debugging, and even low-resource development on older systems—which is practical in an academic context where developer machines vary.

## 3.6. Python, and its Machine Learning Libraries

Several Python-based scientific computing and machine learning libraries are included for OCR, NLP, and general data handling in the SwiftGrade project. These include the standard scientific stack: NumPy, SciPy, and Pandas for numerical operations, signal processing, and data analysis.

For deep learning components, TensorFlow, PyTorch, and optionally TensorRT (TensorCore APIs) are considered to optimize model inference on supported hardware.

These libraries are industry-standard, widely documented, and open-source which makes them ideal for reproducible academic research and experimental model tuning. They also enable efficient data preprocessing, statistical grading insights, essay evaluation, and performance predictions. With GPU support via TensorFlow and PyTorch, the project can leverage local or cloud-based acceleration for large datasets, such as institutional exam archives or bulk student submissions.

# CHAPTER 4

**METHODOLOGY**

## 4.1. Conceptual Framework

The conceptual framework of the project summarizes the entire workflow of the system, from input data collection to the final generation of actionable outputs. At its core, it highlights the seamless integration of machine learning, such as OCR for multiple-choice grading and NLP for essay evaluation, coupled with efficient data management processes. These components collaboratively enable a robust platform tailored to the needs of college educators. A detailed breakdown of these elements and their interplay is presented in Figure 1.

|  |  |  |
| --- | --- | --- |
| **Input** | **Process** | **Output** |
| * Teacher/Student Information * Exam Sheets * Essays * Answer Keys * Rubrics | * Firebase Authentication * CRUD of Teacher/Student Information * CRUD of Exam Sheets, Essays, Answer Keys, Rubrics * OCR Module for multiple choice grading * NLP Semantic Scoring for Essay evaluation | * Teacher/Student Management * Exam Summary Management * Graded Results * Performance Analytics (Item Analysis) on Students * Visualized Reports on Student Performance |

*Figure 1. Input-Process-Output Model of the System*

The Figure illustrates the Input-Process-Output layers of the system, as illustrated the inputs are Teacher/Student Information which will be primary input source. Learning materials like exam sheets, essays, answer keys and rubrics will be also considered as inputs once authentication is approved.

Afterward, the system checks if teacher/student information is recorded in database (Firebase Authentication), once confirmed all other process will be available. Learning materials can now be added, removed, changed and/or recorded, to which the system also offers. It creates, reads, updates and deletes commands are available to accredited staff (teachers, instructors, professors, teaching personnel), and just like the original system,

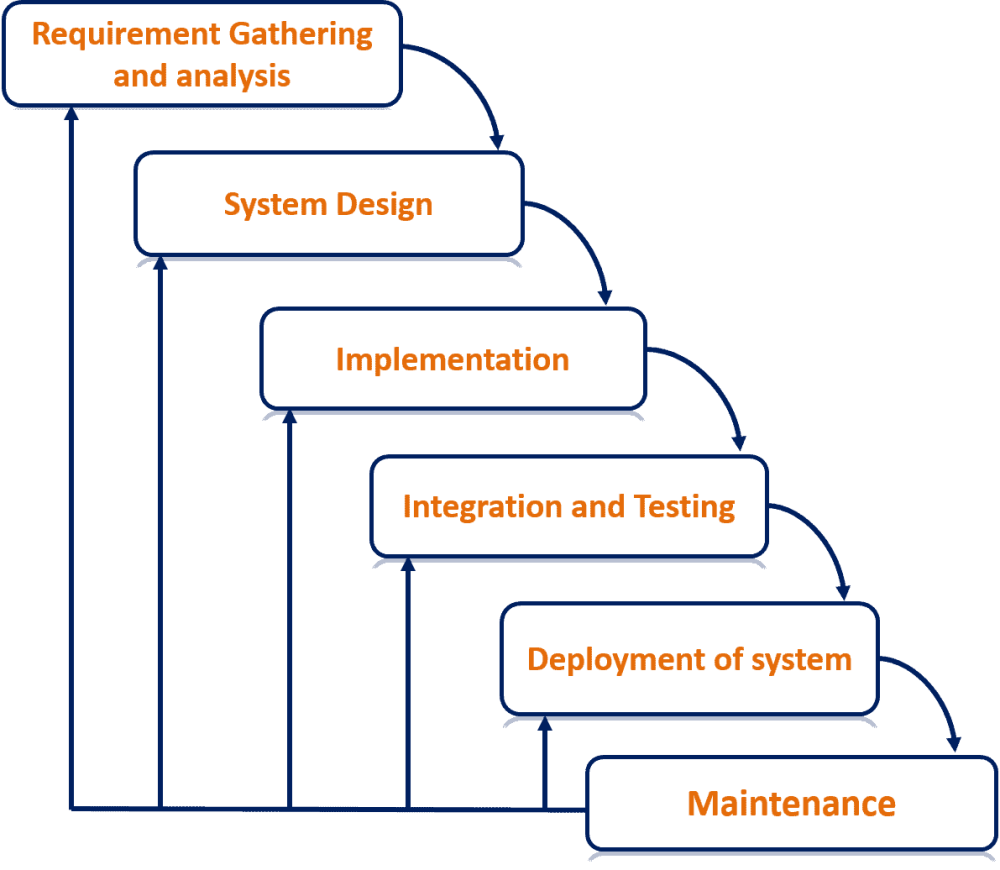
Management Tools are one of the Outputs of the System, For the Accredited Staff: Analytics and Item Analysis on a certain material, or their course history are presented in an analytic lens.

## 4.2. Research Approach

The interview approach is used by the proponents to determine the problems of the stakeholders in the processes of managing their quizzes and exams. While on the secondary users, the proponents used written opinions questionnaire to gather insights on the proposed system.

## 4.3. System Developmental Methodology

The proponents will utilize a modified waterfall model, known for its structured and sequential approach. Each phase is completed in full before moving to the next, with iterative feedback loops for refinement. This balance of clarity and flexibility minimizes risks and improves overall quality of the project. A detailed illustration and their interconnections are provided in Figure 2.



*Figure 2. Modified Waterfall Model  
Source: https://radhikaclasses.com/waterfall-model-in-software-engineering/*

## 4.3.1 Requirements Gathering and Analysis

In this phase, the proponents will gather requirements from the stakeholders through a 10-15-minute open-ended questionnaire. This will identify key problems and determine the essential features and software requirements needed for effective project implementation.

## 4.3.2 System Design

After gathering the requirements, the proponents will plan on the system architecture design, database schema and data flow diagram to structure properly the system to provide a detailed guide on implementation. It also designs the integration of OCR and NLP models for the essay evaluation feature. The proponents will create a wireframe to have a detailed blueprint on its features and modules of the system, which later basis for the development.

## 4.3.3 Implementation

In this implementation phase, the proponents will convert the detailed plans and designs into a functional AI-assisted quiz and exam management system. It will begin by setting up the development environment and adding the necessary hardware and software tools. Then, the proponents will start the development by coding and training the model algorithm that will be applied in the system.

The development process is divided into two groups, the first proponent will train or transfer learn pre-trained models with the required datasets which are later applied to the features of its system, while the other proponents will code the user interface and other core functionalities based on the requirements specified by the stakeholders. This development approach provides faster and more efficient progress in its system development.

## 4.3.4 Integration and Testing

In this phase, the proponents will conduct precise testing to ensure the system meets its functional and performance requirements. Testing process will begin with alpha testing, where the developers evaluate the system to identify and resolve any immediate issues. Unit testing will be applied to examine individual features and modules in isolation, ensuring that each component performs as intended.

The system will undergo integration testing if only the unit testing is successfully completed which later verifies that all modules work together flawlessly, mostly focusing on the OCR and NLP components. This iterative approach allows for immediate identification of errors and correction of bugs, ensuring system stability and reliability before proceeding to the next stage.

## 4.3.5 Deployment of System

# APPENDICES AND FIGURES

## BIBLIOGRAPHY:

*9 ways digital document checking beats manual every time | GBG Identity*. (n.d.). Retrieved November 6, 2024, from https://www.gbgplc.com/en/identity-verification/9-ways-digital-document-checking-beats-manual-every-time/

Agarwal, M., Kalia, R., Bahel, V., & Thomas, A. (2021). AutoEval: A NLP Approach for Automatic Test Evaluation System. *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, 1–6. https://doi.org/10.1109/GUCON50781.2021.9573769

Al-kfairy, M. (2024). Factors Impacting the Adoption and Acceptance of ChatGPT in Educational Settings: A Narrative Review of Empirical Studies. *Applied System Innovation*, *7*(6), Article 6. https://doi.org/10.3390/asi7060110

Bashir, A., & Fadlalla, Y. (2021). A Review: Strategies for Recognizing Forgery in Identity Documents. *Journal of Imaging*.

Deepak, P., Rohan, R., Rohith, R., & R., R. (2024). NLP and OCR based Automatic Answer Script Evaluation System. *International Journal of Computer Applications*, *186*(42), 22–27. https://doi.org/10.5120/ijca2024924038

Domeh, J., Akowuah, E. K., Apeatu, B., & Brako Paa. (2023). *VisioMark: An AI-powered multiple-choice sheet grading system*. https://doi.org/10.13140/RG.2.2.12796.31368

Jaswal, B. A. (2016). Digitization of Archival Collections by Libraries in Pakistan: Issues, Strategies, Challenges and Opportunities. *Pakistan Journal of Information Management and Libraries*, *17*, 97–108. https://doi.org/10.47657/201617897

Kulkarni, M., Adhav, G., Wadile, K., Chavan, R., & Deshmukh, V. (2024). *Digital Handwritten Answer Sheet Evaluation System*. In Review. https://doi.org/10.21203/rs.3.rs-3978232/v1

Lainez, S. M. D., Gresos, L. J. Q., & Maganggo, V. K. A. (2024). An Automated Entrance Examination Checker Using Optical Mark Recognition. *Journal of Computer, Software, and Program*, *1*(1), Article 1. https://doi.org/10.69739/jcsp.v1i1.43

Maniar, S., Parmani, J., Bodke, M., & Saxena, K. (2021). Generation and grading of arduous MCQs using NLP and OMR detection using OpenCV. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–7. https://doi.org/10.1109/ICCCNT51525.2021.9580089

Masula, J., Arsenio, C. J., Flores, N., Nacasabug, E., Tamboong, M., & Geralde, K. (2024). Development of an Offline Android-based Test Paper Checker Application for Northwestern Visayan Colleges. *Journal of Innovative Technology Convergence*, *6*(3), 31–40. https://doi.org/10.69478/JITC2024v6n3a04

Mohammed, M., & Bamasoud, D. M. (2022). THE IMPACT OF ENHANCING AWARENESS OF CYBERSECURITY ON UNIVERSITIES STUDENTS: A SURVEY PAPER. . *. Vol.*, *15*.

Pannu, A. (2015). *Artificial Intelligence and its Application in Different Areas*. *4*(10).

Rahaman, A., & Mahmud, H. (2022). Automated Evaluation of Handwritten Answer Script Using Deep Learning Approach. *Department of CSE,Hamdard University Bangladesh*, 16.

Reddy, A., & Chauhan, V. (2024). *<p>Examination Correction System</p>* (SSRN Scholarly Paper No. 4990316). Social Science Research Network. https://doi.org/10.2139/ssrn.4990316

Singh K., J., Kulkarni, S., B Patil, S., M, S., & Un, S. (2024). OMR Automated Grading. *International Journal of Innovative Science and Research Technology (IJISRT)*, 3757–3761. https://doi.org/10.38124/ijisrt/IJISRT24MAY1072

Szabó Nagy, K., & Kapusta, J. (2023). TwIdw—A Novel Method for Feature Extraction from Unstructured Texts. *Applied Sciences*, *13*(11), Article 11. https://doi.org/10.3390/app13116438

Telenyk, S., Pogorilyy, S., & Kramov, A. (2021). Evaluation of the Coherence of Polish Texts Using Neural Network Models. *Applied Sciences*, *11*(7), 3210. https://doi.org/10.3390/app11073210

Ulven, J. B., & Wangen, G. (2021). A Systematic Review of Cybersecurity Risks in Higher Education. *Future Internet*, *13*(2), Article 2. https://doi.org/10.3390/fi13020039

Wangsa, K., Karim, S., Gide, E., & Elkhodr, M. (2024). A Systematic Review and Comprehensive Analysis of Pioneering AI Chatbot Models from Education to Healthcare: ChatGPT, Bard, Llama, Ernie and Grok. *Future Internet*, *16*(7), Article 7. https://doi.org/10.3390/fi16070219