

Automatic essay grader with artificial intelligence feedback

Interim Report

DT211C

BSc in Computer Science Infrastructure

**Joseph Egan**

**C20389531**

**Supervisor**

**Dr Aneel Rahim**

School of Computer Science

Technological University, Dublin

**15/11/2024**

Abstract

Since 2016, there has been a 13.8% increase in the number of enrolled students in Irish universities, (<https://hea.ie/statistics/data-for-download-and-visualisations/access-our-data/access-our-data-students/>), reflecting a growing student body in Ireland. This has contributed to a growing lecturer to student ratio problem in Ireland. Ireland currently has a student-to-staff ratio of 22.38-to-1, (<https://www.oecd.org/en/data/indicators/students-per-teaching-staff.html?oecdcontrol-8476e61c3c-var6=TRY> ), while the OECD (Organisation for Economic and Educational development which Ireland is apart of) average is 15.58. This growing disparity of the student-to-staff ratio proses a problem for lecturers, especially for labour intensive tasks such as grading essays and creating individual feedback for students.

Detailed feedback on written work is crucial for students, as it supports their academic growth and development. Yet, with the added workload, lecturers may struggle to provide the kind of individualized feedback that helps students improve essential skills, like writing and critical analysis.

An artificial intelligence-supported essay grading and feedback tool could help bridge this gap. By automating parts of the grading process, such a tool can deliver timely, constructive feedback on areas like grammar, argument structure, and coherence. It allows lecturers to allocate their time to more complex, personalized teaching tasks while ensuring students receive consistent, relevant feedback. Implementing an AI-driven system in this context could enhance the learning experience for students and help educators better manage the demands of growing class sizes across Irish higher education.

Declaration

I hereby declare that the work described in this dissertation is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

Signed:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Joseph Egan

15/11/2024

Acknowledgements

I would like to thank my supervisor Doctor Aneel Rahim for his guidance throughout the course of the project. I am also grateful to my family for supporting me.

Table of Contents

[1. Introduction 7](#_Toc119925914)

[1.1. Project Background 7](#_Toc119925915)

[1.2. Project Description 7](#_Toc119925916)

[1.3. Project Aims and Objectives 7](#_Toc119925917)

[1.4. Project Scope 7](#_Toc119925918)

[1.5. Thesis Roadmap 7](#_Toc119925919)

[2. Literature Review 8](#_Toc119925920)

[2.1. Introduction 8](#_Toc119925921)

[2.2. Alternative Existing Solutions to Your Problem 8](#_Toc119925922)

[2.3. Technologies you’ve researched 8](#_Toc119925923)

[2.4. Other Research you’ve done 8](#_Toc119925924)

[2.5. Existing Final Year Projects 8](#_Toc119925925)

[2.6. Conclusions 8](#_Toc119925926)

[3. System Design 9](#_Toc119925927)

[3.1. Introduction 9](#_Toc119925928)

[3.2. Software Methodology 9](#_Toc119925929)

[3.3. Overview of System 9](#_Toc119925930)

[3.X. Other Sections 9](#_Toc119925931)

[3.X. Conclusions 9](#_Toc119925932)

[4. Testing and Evaluation 10](#_Toc119925933)

[4.1. Introduction 10](#_Toc119925934)

[4.2. Plan for Testing 10](#_Toc119925935)

[4.3. Plan for Evaluation 10](#_Toc119925936)

[4.4. Conclusions 10](#_Toc119925937)

[5. Prototype Development 11](#_Toc119925938)

[5.1. Introduction 11](#_Toc119925939)

[5.2. Prototype Development 11](#_Toc119925940)

[5.3. Other Sections 11](#_Toc119925941)

[5.4. Conclusions 11](#_Toc119925942)

[6. Issues and Future Work 12](#_Toc119925943)

[6.1. Introduction 12](#_Toc119925944)

[6.2. Issues and Risks 12](#_Toc119925945)

[6.3. Plans and Future Work 12](#_Toc119925946)

[6.3.1. GANTT Chart 12](#_Toc119925947)

[Bibliography 13](#_Toc119925948)

# 1. Introduction

**As least 2 pages, but as many as you like**

## Project Background

Some background and literature, start with an interesting fact or a newspaper item

Ireland was one of the founding members of the Organisation of Economic Co-operation and Education (OECD) in 1960 and currently faces significant challenge with its staff-to-student ratios with one of the worst students to lecturer ratios of the current organization members. This means that lecturers are, relatively speaking, dealing with more students compared to lecturers in other countries, for example Ireland has 22.38:1 student to lecturers whereas the United States has 13.63, and the UK has a ratio of 13.50.

This ratio imbalance creates excess strain on lectures, especially for labour intensive tasks such as grading essays, and providing detailed personalised feedback. The growing number of Irish students creates a need for tools that can solve this problem.

## Project Description

An overview of the project

My automatic essay grader with AI feedback (AEGAF) program allows students to get estimates of the quality of their essays and possibly points of feedback and advice on weaknesses and potential areas of improvements for their essays. Students will choose what question they are answering and then post their essay in a doc or pdf format into my system. This system will use a combination of the grading rubric for that question and my machine learning model. The results of this essay, the grading rubric and the essay will be passed on into an AI that will write the grade and feedback into a more human readable format, for example you scored 80% on your essay, but need to focus on grammar and giving an example to support your point. The system will them post the result and feedback to the user.

## Project Aims and Objectives

The main aim for the project is to develop a simple tool for students to use to help with their essays.

Overall aim and some milestones along the way to achieve the aim

* Write the interim report
* Get the document input working
* Get the document processing working
* Get the rubric grading scheme of some sample papers
* Develop the machine learning model for grading papers
* Develop demo for interim
* Develop the AI API
* Develop checks for the AI
* Get the result output working
* Get save feature for the result and feedback working

## Project Scope

Project scope, what the project isn’t about

The main focus of the project is the machine learning model grading an essay off of a pre-existing question. It is not supposed to write essays for students or help them get passed Ai or plagiarising checkers. The AI is supposed to enhance the grading process, i.e. explain the results in personalised terms, not grade an essay on its own. This is not about building my own large language model artificial intelligence.

## Thesis Roadmap

One sentence explaining what each of the following chapters is about.

1 or 2 line summary of that part of the document.

# Literature Review:

Research for the project using scientific papers

# 3. System Design

The plan and the architecture of the project.

# 4. Testing and Evaluation

Validation and verification achievements and standards, and the predetermined way to check if they have been achieved.

# 5. Prototype Development

The product for the interim report and how it will help the future development of the project

# 6. Issues and Future Work

The future possible developments to the project and alternative ways that it could have been done or developed.

# Literature Review

## 2.1. Introduction

An overview of machine learning models and a brief overview of artificial intelligence is provided in this section as well as the problems and challenges related to it.

There are already tools available to testing a person’s essay that can be freely used, examples such as ChatGPT and the GitHub automated essay scoring project. These will be discussed in terms of comparison with the functionality available. There are also other suitable technologies that were researched during development including existing ai such as Chat GPT or Gemini, datasets, programming languages, and cloud providers.

## 2.2. Alternative Existing Solutions to Your Problem

Software you’ve investigated:

Artificial Intelligences – ChatGPT, Gemini, CoGrader, and GradeWrite.

GitHub automated essay scoring – a GitHub repository with a similar function to mine.

Lecturers – My project’s goal is to alleviate the effort for lecturers to grade students’ s essays, but they can still personally grade essays themselves.

Other tools – Grammarly or ETS e-reader.

## 2.3. Technologies you’ve researched

Programming languages, operating systems, etc.

Languages – I used the programming language Dart a cross-platform language last year that would allow my to build a single application that would work on a series of platforms, such as IOS, Android, Windows, or MacOS. I also researched python a language I have used heavily over the years as a student. But I currently plan on using Java due to the pre-existing supports for the language

Operating system – I have researched and come to the conclusion that I cannot develop for the MacOS or IOS systems due to lack of availability for development and testing. I am currently developing a Windows application, but I hope to add a Android mobile application if I have the opportunity to do so.

Ai – I have researched the availability of pre-existing artificial intelligences, such as ChatGPT, Gemini, and GradeWrite and come to the conclusion that ChatGPT is the most supportive ai to use as a basis for my project.

## 2.4. Other Research you’ve done

Domain specific research

## 2.5. Existing Final Year Projects

Title: Machine Learning for Predictive Analysis and Recommender System

Student: Glory Pierce Eguare

Description (brief): An application that generates a predictive algorithm for the fuel

consumption efficiency of automotives

What is complex in this project: The

What technical architecture was used: Model template view architecture

Explain key strengths and weaknesses of this project, as you see it.

Strengths: The use of a detailed dataset auto-mpg, to train the machine learning algorithm.

Weaknesses: The normalisation of the ranges of his dataset to facilitate easier training, may

have led to skewed data biases to his results. This would have disproportionally punished

more specialised vehicles, i.e. an unusually heavy car would look like it would have

unexpectedly terrible mpg, and raise the value of more average cars.

Title: Generating expenditure tracking and analysis from hard copy receipts

Student: Cillian Keohane

Description (brief): A system to scan receipts and process them into an expenditure tracker

What is complex in this project: The use of a Wikimedia API and the machine learning

algorithm to process the receipts.

What technical architecture was used: Model, View, and Controller architecture

Explain key strengths and weaknesses of this project, as you see it.

Strengths: the use of the scanner and photo processor, and the machine learning algorithm

that adapts to the individual user and improves over time.

Weakness: It is not synchronized to other banking devices to automate the process of taking

in receipts.

## 2.6. Conclusions

The literature review highlights the existing technologies and methodologies that refine my project. It demonstrates that while several AI tools and essay scoring systems exist, they often lack the tailored approach needed for individualized academic feedback. It also demonstrates other alternatives are not covering my niche topic.

# 3. System Design

## 3.1. Introduction

This section contains the design process and system architecture of the project. It contains the software methodology, core components, architecture and design overview of the project.

## 3.2. Software Methodology

The software methodology that I am currently choosing to use for my project is the Agile software methodology. I am choosing to use this methodology, because of its core principle of creating iterations of the project. The Agile methodology can be broken down into 5 phases.

* Phase 1: Planning and Requirements Analysis
* Phase 2: Design
* Phase 3: Development
* Phase 4: Testing
* Phase 5: Deployment

I am choosing to use the Agile methodology, because of the flexibility and continuous development aspects of the methodology. In order to improve and adapt the project over time during the completion of the project.

## 3.3. Overview of System

My automatic essay grader with AI feedback (AEGAF) program allows students to get estimates of the quality of their essays and possibly points of feedback and advice on weaknesses and potential areas of improvements for their essays. Students will choose what question they are answering and then post their essay in a doc or pdf format into my system. This system will use a combination of the grading rubric for that question and my machine learning model. The results of this essay, the grading rubric and the essay will be passed on into an AI that will write the grade and feedback into a more human readable format, for example you scored 80% on your essay, but need to focus on grammar and giving an example to support your point. The system will them post the result and feedback to the user.

## 3.X. Other Sections

(

## 3.5. Conclusions

The system design section lays a solid foundation for my project. This combines established software methodologies and a well-structured architecture. By adopting Agile practices, the design process ensures flexibility and continuous improvement, aligning with my project's iterative nature. The integration of machine learning models and user-centric artificial intelligence feedback generation illustrates the system's alignment with its goals of enhancing academic support through artificial intelligence.

# 4. Testing and Evaluation

## 4.1. Introduction

This section contains the testing and evaluation plan, process, and development of the project. It covers the methods used to test each module, details evaluation metrics, and outlines the approaches for assessing the overall performance and user experience.

## 4.2. Plan for Testing

The methodology for the development of the project is an Agile approach. This involves performing development cycles over a development period, for example get a component done in one week. I aim to get units of the system, for example document input, essay analysis, rubric application, and feedback generation, programmed and then validation tested. I will then do system testing to validate the integration of the components. Automated test cases will be written to verify functional requirements, and simulated user scenarios will assess the system.

## 4.3. Plan for Evaluation

Evaluation will focus on two key aspects:

1. Performance Metrics: The system’s speed, accuracy of grading, and relevance of feedback will be compared against manual grading benchmarks.
2. UserFeedback: I hope to get students and lecturers to provide input and performance evaluation on the system’s usability and the value of its feedback. This input will guide iterative improvements.

## 4.4. Conclusions

The testing and evaluation process will hopefully confirm the system's ability and effectiveness in generating personalised essay feedback. By addressing potential shortcomings during this phase, the project aims to deliver a practical tool to enhances both teaching and learning, for both students and lecturers.

# 5. Prototype Development

**As least 2 pages, but as many as you like (but lots of code samples).**

(take screenshots and explain the functionality)

## 5.1. Introduction

This section documents the initial implementation of the project, showcasing the development of core features and highlighting challenges encountered. It will show an early proof of concept of the project and demonstrates its complexity and values. This will provide code samples, design diagrams, and explanations of the chosen solutions.

## 5.2. Prototype Development

## 5.3. Other Sections

## 5.4. Conclusions

The prototype successfully demonstrates the core capabilities of the AEGAF system, laying the foundation for further enhancements. The next phase will focus on integrating additional features, optimizing AI models, and refining user experience.

# 6. Issues and Future Work

## 6.1. Introduction

This section highlights the challenges faced during development, outlines their implications, and proposes solutions. It also discusses potential extensions to the project that could enhance its impact and utility.

## 6.2. Issues and Risks

Scalability: One of the biggest challenges my project faces is the development of grading rubrics with examples for my project to learn how to grade new papers. For example, to add in an new paper that my system can grade, Advanced Security 1 2025/2026, an administrator would have to input the paper, a grading rubric with the answers for the questions and examples of answers so that my system can learn how to grade the paper. This must be done for each and every question in every exam paper in advance, which generates labour for lectors and teacher assistants to perform in order for my system to work.

Reliability: There needs to be a dataset of each question being answered to teach the model how to grade the questions. There are risks however of the system grading the wrong way, i.e. grading biases, struggling with grammar or spelling, and giving incorrect answers and feedback.

Maintenance: The system is only useful if it is up to date with the latest exam papers and will require regular updates and the ability to fix any errors in the system.

## 6.3. Plans and Future Work

Platform expansion: Currently the scope of the project only includes functionality on Windows and possibly on Android. A path for future development is to expand the project’s scope to include operation on Android, IOS, and on MacOS platforms.

Multilingual support: The current scope of the project is on traditional English only. The hope for possible future development is to include support for simple English (US) and other additional languages, such as French, German, or Arabic.

Advanced artificial intelligence assistance: Training a specialised artificial intelligence for directly grading the essays, instead of the current system of the machine learning model grading with the artificial intelligence only offering feedback.

### 6.3.1. GANTT Chart

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Nov/Dec | Jan | Feb | March | April |
| Planning |  |  |  |  |  |
| Development |  |  |  |  |  |
| Testing |  |  |  |  |  |
| Deployment |  |  |  |  |  |
| Iterations |  |  |  |  |  |
| Presentation |  |  |  |  |  |

(2 paragraphs to explain the chart^)

# Bibliography