

# **AUTOMATED GRADING OF FREE-TEXT STUDENT SUBMISSIONS USING LARGE LANGUAGE MODELS.**

Subtitle

**School of Computer Science & Applied Mathematics  
University of the Witwatersrand**

**Sphamandla Mbuyazi  
2618115**

**Supervised by Prof. Richard Klein**

**April 13, 2025**



Ethics Clearance Number: XX/XX/XX

A proposal submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg,  
in partial fulfilment of the requirements for the degree of Bachelor of Science with Honours

## **Abstract**

Abstract things....

### **Declaration**

I, Sphamandla Mbuyazi, hereby declare the contents of this research proposal to be my own work. This proposal is submitted for the degree of Bachelor of Science with Honours in Computer Science at the University of the Witwatersrand. This work has not been submitted to any other university, or for any other degree.

## **Acknowledgements**

Thanks World.

# Contents

|   |          |
|---|----------|
| <b>Preface</b>                                    |          |
| Abstract . . . . .                                | i        |
| Declaration . . . . .                             | ii       |
| Acknowledgements . . . . .                        | iii      |
| Table of Contents . . . . .                       | iv       |
| List of Figures . . . . .                         | v        |
| List of Tables . . . . .                          | vi       |
| <br>  |          |
| <b>1 Introduction</b>                             | <b>1</b> |
| 1.1 Introduction . . . . .                        | 1        |
| 1.2 Purpose of this review . . . . .              | 1        |
| 1.3 Brief overview of the key themes . . . . .    | 2        |
| 1.3.1 A Subsection about Citation Style . . . . . | 2        |
| 1.3.2 Compiling . . . . .                         | 2        |
| <br>  |          |
| <b>2 Floats</b>                                   | <b>3</b> |
| <br>  |          |
| <b>3 Some Referencing Tricks</b>                  | <b>5</b> |
| <br>  |          |
| <b>4 IDE/Editors</b>                              | <b>6</b> |
| <br>  |          |
| <b>A Extra Stuff</b>                              | <b>7</b> |
| A.1 What is an appendix? . . . . .                | 7        |
| <br>  |          |
| <b>References</b>                                 | <b>8</b> |

# List of Figures

|                                |   |
|--------------------------------|---|
| 2.1 This is an image . . . . . | 3 |
|--------------------------------|---|

# List of Tables

|     |  |   |
|-----|--|---|
| 1.1 | Themes and Observations in AI-based Educational Assessment . . . . . | 2 |
| 2.1 | Table Name . . . . .   | 4 |

# Chapter 1

## Introduction

### 1.1 Introduction

In education technology, auto grading of free text has long been a sought-for solution, particularly in large-scale assessments and technical subjects like Computer Science , where assessments cannot be easily reduced to multiple choice questions. As the number of students enrolled in a course increases per instructor and the curriculum becomes more demanding, educators face a mounting pressure to provide fair, consistent and fast feedback.

This poses a need to try to automate the process of marking and giving feedback. In the state of art of education right now, your instructor grades your assessments and 4 years later provides you with a feedback, as we can argue without proof that early feedback helps students perform better in their studies.

At the same time, Large Language Models(LLMs) like GPT-4, Claude, and open-source alternative such as LLaMA and DeepSeek are proving to be not just the conversational tools but they are capable evaluators, summarizers, and even tutors. Their impressive capacity for natural Language understanding and generation raises an exciting possibility:

Can these models be harnessed to grade student responses automatically and meaningfully? –and perhaps even provide personalized feedback that guides learning?

### 1.2 Purpose of this review

The aim of this review is to synthesize and critically examine recent research on the use of LLMs for autograding and feedback generation in educational contexts. We shall look at how these models have been applied to assess student submissions, what techniques have been used to improve reliability and fairness, and how feedback mechanisms are integrated to support and elevate student learning.



By comparing methodologies, models and findings we aim to infer what is working, what remains challenging, and where there's space for improvement. This should ultimately inform how we design our own LLM-based grading system for technical subjects, one that minimizes educators workload while enhancing the feedback experientia

## 1.3 Brief overview of the key themes

Across the literature I have examined, several key themes and technical strategies emerge.

| Theme                       | Approaches / Observations  |
|-----------------------------|--|
| Grading Accuracy            | LLMs like GPT-4 and fine-tuned BERT models show strong agreement with human graders.                 |
| Feedback Generation         | Models like BeGrading and GPT-4 in science writing generate formative feedback students find useful. |
| Explainability              | SHAP explanations and linguistic feature analysis offer insight into deep model decisions.           |
| Data Efficiency             | Active learning (uncertainty, topology, hybrid methods) helps reduce labeling needs.                 |
| Prompt Engineering          | Careful prompt design significantly improves LLM grading consistency and relevance.                  |
| Peer/Human-in-the-loop      | Hybrid systems using peer grading or instructor scaffolding improve fairness and training.           |
| Open-Source Model Potential | LLaMA 2 and Falcon perform comparably to commercial models in bioinformatics grading tasks.          |

Table 1.1: Themes and Observations in AI-based Educational Assessment

### 1.3.1 A Subsection about Citation Style

Citations are important. Citation style for Computer Science is:

- When used in the text, use the authors with the date in brackets:  
Klein and Celik [2017] say very important things.
- When used as a reference after a face, put everything in brackets:  
Import things are true [Klein and Celik 2017].

### 1.3.2 Compiling

Remember to compile multiple times to resolve references. Usually:

```
pdflatex file.tex
bibtex file
pdflatex file.tex
pdflatex file.tex
```

# Chapter 2

## Floats

$\text{\LaTeX}$  decides how to place images. It also does the referencing for you as seen in Figure 2.1. If you have subimages, they should have their own captions and labels – look into the subfig or subfigure packages.



Figure 2.1: This is an image

Figure captions are at the bottom. Table title are at the top of the table as seen in Table 2.1 on the following page. There is a package called BookTabs which is *way* better for tables and you should learn how to use that instead.

Usually let  $\text{\LaTeX}$  handle the placement of floats unless you *really* need to force it to do something else. The float package used above allows you to use H as the placement which means *here and only here*. When using the float package, the placement options are:

1. h – a gentle nudge to place it here if possible
2. t – top of a page
3. b – bottom of a page
4. H – here and only here, do not move it at all
5. p – on its own page

Table 2.1: Table Name

| Col1  | Col2  |
|-------|-------|
| R0,C0 | R0,C1 |
| R1,C0 | R1,C1 |

# Chapter 3

## Some Referencing Tricks

CleverRef and VarioRef are helpful:

- Normal Ref: See Figure [2.1](#)
- CleverRef: See Figure [2.1](#) and Table [2.1](#)
- CleverRef+VarioRef: See Figure [2.1](#) on page [3](#) and Table [2.1](#) on the facing page

# Chapter 4

## IDE/Editors

Overleaf has a great online editor for latex. Use it.

# Appendix A

## Extra Stuff

### A.1 What is an appendix?

An appendix is useful when there is information that you need to include, but breaks the flow of your document, e.g. a large number of figures/tables may need to be shown, but maybe only one needs to be in the text and the rest are just included for completeness.

# References

- [Klein and Celik 2017] R. Klein and T. Celik. The Wits Intelligent Teaching System: Detecting student engagement during lectures using Convolutional Neural Networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2856–2860, Sep. 2017.