

# CM50175 – Project Proposal

Oscar Dos Santos Nunes  
Student ID: JOFDSN20

28 March 2025

## Project Title

The current title for the project is ”**Automated Essay Scoring Using Deep Learning with XXX**”.

This title reflects to general aim for the work to design, implement and evaluate and automated essay scoring (AES) algorithm focusing on the utilization of deep learning techniques.

However, the title will be updated as the project progresses and the research focus becomes more defined. The current placeholder “XXX” represents a yet to be decided add-on. This may involve the implementation of a specific deep-learning architecture—such as LSTM, BERT or a hybrid transformers model or the integration of additional functionality such as expandability, feedback generation, or the processing of handwritten text.

Some possible titles include:

- *Automated Essay Scoring Using Deep Learning*
- *Neural Approaches to Automated Essay Scoring: A Deep Learning Perspective*
- *Transformer-Based Models for Automated Essay Scoring*
- *Evaluating Student Writing with Deep Learning: A Comparative Approach*
- *Automated Essay Scoring with Explainable Deep Learning Architectures*
- *Beyond Accuracy: Fair and Interpretable Deep Learning for Essay Scoring*
- *Multimodal Essay Assessment: Integrating Handwritten and Typed Text in Neural AES Systems*
- *Benchmarking Traditional and Deep Learning Models for Essay Scoring*

## Problem Statement

Essay based work is an essential part of the educational process and therefore is vital for performance evaluation. However the grading of essays is a highly subjective process and can be time consuming. Studies have shown that the grading of essays by different teachers can vary significantly, with some teachers being more lenient than others. This can lead to inconsistencies in grading and can affect the overall performance of students. These issues are especially prevalent in larger-scale assessments (e.g. GCSEs, ALevels and Uni assessments) where many markers are involved. The use of automated essay scoring (AES) systems can help to address these issues by providing a more consistent and objective grading process.

AES systems aim to address these limitations through computational techniques to evaluate the quality of written text. Early AES systems such as Project Essay Grade, relied very heavily

on manually engineered features such as grammar usage, word length, and sentence structure, combined with traditional machine learning models. While these systems showed success (e-rater was used in english as a foreign language scoring), they often lacked robustness and generalization capabilities.

Recent advancements in deep learning and natural language processing have introduced more sophisticated models and approaches to AES. Neural architectures such as LSTM networks and transformers based models such as BERT have achieved high performance across a variety of NLP tasks, including text classification and sentiment analysis. For example, BERT achieved a GLUE benchmark score of 80.5 in its original implementation, showing its ability to model contextual information and the relationships between words in a sentence. The models are well suited to AES tasks due to their ability to contextualize meaning beyond the surface-level features.

However, there are still several challenges to overcome. Deep networks often act as "black boxes", which offer very little interpretability into how they arrive at an output. This creates a concern for use in education where explainability/feedback is essential. Furthermore, these models require vast quantities of labeled-data of which is not always available. Additionally there are a number of documented risks regarding the algorithmic bias in AES particularly when dealing with non-native speakers or student with a less represented background than in the training data.

Moreover, much of the existing AES models focus solely on predicting a final holistic score, with less of a focus on providing additives that would be given by a human marker, such as feedback on specific aspects, annotated marking, or the ability to discuss the current mark and how to improve. Each of these would substantially improve the possible use of AES in education.

This project aims to explore and evaluate the application of deep learning techniques to automate essay scoring, with a focus on performance, fairness, and extensibility. Through the implementation of a deep learning model, the research aims to contribute to the development of a more accurate, explainable and general use AES system.

## **Objectives and Research Questions**

## **Background and Related Work**

## **Project Plan**

### **Task Breakdown**

### **Timeline**

### **Dependencies and Risks**

## **Resources and Limitations**

## **References**