**STANDARDIZED GRADING SYSTEM FOR TEXT CONTENT USING LARGE LANGUAGE MODELS**

DEVELOPMENT PHASE

GRADUATE DIPLOMA IN DATA ANALYTICS

BY

Cecilia Lincon Carrillo

Jermaine Jan Calip

SUPERVISORY TEAM

Dr. Harsh Tiwari (NZSE)

Dr Amir Mohammadi (PROMPTECH)

February 2025

Contents

[Abstract (300 words) 3](#_Toc185500010)

[Keyword 3](#_Toc185500011)

[Introduction (400 words) 3](#_Toc185500012)

[Project Questions (250 words) 3](#_Toc185500013)

[Project Design (900 words) 3](#_Toc185500014)

[Project Development (600 words) 4](#_Toc185500015)

[Discussion (400 words) 4](#_Toc185500016)

[Conclusion and future directions (400 words) 4](#_Toc185500017)

[References 5](#_Toc185500018)

# Abstract (300 words)

This is a test

This is a second test

Summarize the key objectives: developing a content grading system for audiences under 15 using LLMs.

Briefly explain the technical design, including the integration of language models and criteria like cultural sensitivity and explicit content.

Highlight the main findings and how they support ethical decision-making on online platforms.

# Keyword

Large Language Models (LLMs), Natural Language Processing (NLP), Text Classification, Content Grading, Age-appropriate Content, Explicit Content Detection, Violence Detection, Child Protection, Parental Control, and Artificial Intelligence (AI).

# Introduction

The connectivity of the digital age grants us access to various types of content in an instant and accessible manner. While this accessibility brings us many benefits, it also brings with it the risk of exposing children and adolescents to potentially inappropriate, sensitive or disturbing material. Audiovisual platforms such as Netflix or YouTube have advanced systems to classify their content according to the age of the audience, however, text platforms such as media, news or blogs lack equivalent solutions. Access to content that details explicit violence or emotionally complex topics for audiences under 15 years of age is especially worrying since this age group can be affected in their emotional or psychological development.

This project aims to address this gap by developing a platform with an automated grading system for text content based on Language Large Models and Natural Language Processing. The main objective is to create a tool that guarantees access to appropriate content for audiences under 15 years old. The system analyses the text and classifies it according to a predefined scale of appropriateness based on criteria such as violence detection, explicit sexual content, cultural sensitivity and mature themes. This tool will allow parents, educators and content providers to protect young audiences from harmful content that may impact them psychologically or emotionally.

The lack of effective tools to classify text content represents a critical challenge for media, educational or blogging platforms. This project not only addresses this gap but also introduces a scalable and adaptable solution, applicable to diverse digital platforms. The implementation of the proposed system will enable platforms to meet higher standards of social responsibility while strengthening user trust. The role of advanced technologies such as LLMs and NLPs in the data analysis of this project represents an innovation in the field of content moderation, contributing to the development of more robust systems to protect young audiences, through the study carried out to choose the most appropriate model that would present the best performance in text classification for the present context.

In the following sections, the project questions and design are detailed, including diagrams and methodologies used to define the model integration. The project development is then presented, where the implementation details and the tests performed are addressed. Finally, the key findings, challenges faced, and conclusion are discussed, as well as future recommendations and the potential impact of the system on the textual content moderation industry.

# Project Questions (250 words)

The following questions are intended to address the needs of the project for its development and also guide the exploration of existing literature concerning the topic in question.

* **What is the role of NLP and LLM in text analysis?**

Explore through literature review how advanced Natural Language Processing (NLP) and Large Language Models (LLM) techniques have been used in text analysis, what their role is, and how they can be applied to the classification of inappropriate text.

* **Which LLM models have the best performance in text classification in previous studies?**

Identify through previous studies or academic articles which model(s) present the best performance in the implementation of text analysis, which will provide a solid basis for selecting the most appropriate model for the project.

# Project Design (900 words)

Frame your industry project question(s) well. Ensure they are clear and aligned with questions in the project proposal. Also, it is required to make sure the project's scope is matched with the project proposal. You should provide a breakdown of the project questions in detail and make the connection to the business needs or problems. The degree of complexity of the business needs should be evident.

Describe each design component:

* **Data Sources**: Explain the use of APIs or existing datasets.
* **Preprocessing**: Detail text cleaning and normalization.
* **LLMs**: Define their configuration and training.
* **Schematic Diagram**: Include a visual representation showing the interaction between data, model, and outputs.

# Project Development (600 words)

Explain the implementation stages:

* **Source Code**: Include screenshots with detailed descriptions.
* **Testing**: Discuss examples used to validate the model.
* **Maintenance and Scaling**: Provides a reference for future modifications, scaling the system, or debugging issues that may arise after deployment (if it is applicable).

# Discussion (400 words)

Discuss challenges related to your design and development and technical items. Align the discussion with the design and development undertaken, highlighting the practical application of the coding phase. Cover the knowledge gaps that must be closed to execute this project competently. It should also give details on the limitations of the project. *(Testing part limitations):  
- Lack of AI Models*

* *Text -> is recommendable testing with several examples (more than 100)*

# Conclusion and future directions (400 words)

Conclude with a summary of the project, highlighting key findings, achievements, and practical implications, and **make recommendations** for the project. Include the unique contribution to data analytics or the industry, emphasising innovative aspects and potential impact. Finally, outline any future issues and directions to consider.

* Conclusion: Summarize achievements, such as model accuracy and relevance to the industry.
* Future Directions: Suggest improvements, such as expanding evaluation criteria or including continuous monitoring.

.

# References

**(using APA-v7 style)**

* **OpenAI API Documents**
* **Metric Selection**