# Low-Level Design

**Freeform Text Generation for Content Creators project**

|  |  |
| --- | --- |
| Written By | Dharam Chopda |
| Email | chopdadharam143@gmail.com |
| Date | 03-11-2024 |

## Document Change Control Record

|  |  |  |
| --- | --- | --- |
| Version | Date | Author |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Reviews:

|  |  |  |
| --- | --- | --- |
| Version | Date | Reviewer |
|  |  |  |
|  |  |  |

## Approval Status:

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Review Date | Reviewed By | Approved By |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## Index

1. Introduction:  
2. Scope:  
3. Objective:  
4. System Overview & Architecture  
5. Modules Descriptions (Data Preprocessing, Model, Text Generation, Evaluation Modules)  
6. Sequence Diagrams Process Flow Diagram  
7. Data Structures  
8. Additional Details  
9. Conclusion

## Introduction

What is Low-Level Design Document. The goal of LLD or a low-level design document (LLDD) is to give the internal logical design of the actual program code for ‘FreeText Generation project’. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

## Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code, and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

## Objective

Develop a system that generates sentences based on provided concepts using the meta-llama/Llama-2-7b-chat-hf model. Each sentence should naturally incorporate the specified concepts.

## System Overview

The system leverages the pretrained Llama-2-7b-chat model for generating sentences. It consists of modules for data preprocessing, model loading, prompt formatting, text generation, and evaluation to ensure generated sentences meet project requirements.

## Modules

3.1) Data Preprocessing Module - Purpose: Prepare concept sets by cleaning, tokenizing, and formatting them for input to the model.

- Input:

Raw concept data (e.g., 'catch\_V', 'dog\_N')  
- Output:

Cleaned concept lists (e.g., ['catch', 'dog'])

- Classes/Functions:

ConceptCleaner - Prompt Creator  
  
3.2) Model Module -

Purpose: Loads and manages the meta-llama/Llama-2-7b-chat-hf model for sentence generation.

- Input:

Formatted prompt from Data Preprocessing Module.

- Output:

Generated sentence containing the required concepts

- Classes/Functions: -

LlamaModelLoader - PromptFormatter  
  
3.3) Text Generation Module -

Purpose: Generates sentences based on formatted prompts and ensures required concepts are present.

- Classes/Functions: - TextGenerator - KeywordValidator  
  
3.4) Evaluation Module - Purpose: Measures the quality using metrics like BERTSCORE

## Sequence Diagrams Process Flow Diagram

Data Preprocessing Module:  
- Cleans and formats concept data and generates prompts.  
  
Model Module:  
- Loads and configures Llama-2-7b-chat.  
  
Text Generation Module:  
- Generates sentence and validates its contents.  
  
Evaluation Module:  
- Computes quality metrics.

## Data Structures

- Inputs:  
 - Concepts: List of concepts with suffixes removed (e.g., ['catch', 'dog']).

- Prompt: Formatted string containing instruction for the model (e.g., 'Create a sentence using these concepts: catch, dog.').  
  
- Outputs:- Generated Sentence: A natural language sentence containing all specified concepts (e.g., 'The dog runs to catch the ball.').

## Additional Details

- Error Handling:  
 - Out of Memory: Monitor GPU memory usage, and implement batch processing where possible.  
 - Concept Validation: If required concepts are missing, regenerate output.  
  
- Dependencies:  
 - Python Libraries: transformers, datasets, torch, logging.

## Conclusion

This LLD document outlines a structured approach to generating sentences using the meta-llama/Llama-2-7b-chat-hf model with the Common Lite Gen Dataset. The design details cover modules for prepossessing, model management, text generation, and evaluation to ensure that generated sentences align with the project objectives.