Software Requirement Specification for Text-to-Video Generation using Generative AI with Diffusion Model

Presented By:

Mallarapu yagnesh naidu (210304124030)

Modugula gopinath (210304124079)

Midde Venkat Sai (210304124074)

Y.V.Niranjan Reddy(210304124449)

Guided by:

HIREN RAITHATHA

CSE DEPARTMENT

PIET, PARUL UNIVERSITY

Project no:cse 493

1. Introduction

1.1 Purpose

The purpose of this document is to outline the software requirements for developing a text-to-video generation system utilizing Generative AI with the Diffusion Model. This system aims to generate realistic videos based on input textual descriptions using state-of-the-art generative models.

1.2 Scope

The system will take textual input descriptions as input and generate corresponding videos that closely match the provided descriptions. The focus will be on leveraging Generative AI techniques, particularly the Diffusion Model, to produce high-quality and diverse video content.

1.3 Intended Audience

This document is intended for software developers, designers, stakeholders, and any other individuals involved in the development, testing, and deployment of the text-to-video generation system.

2. Functional Requirements

2.1 Text Input Processing

2.1.1 Accept Text Input

The system shall accept textual input descriptions from users or external sources.

2.1.2 Preprocess Text Input

The system shall preprocess the input text to remove any irrelevant characters, punctuation, or formatting.

2.2 Video Generation

2.2.1 Diffusion Model Integration

The system shall integrate the Diffusion Model for video generation.

2.2.2 Generate Video from Text

The system shall use the Diffusion Model to generate videos based on the preprocessed textual input.

2.3 Quality Control

2.3.1 Realism Assessment

The system shall include mechanisms to assess the realism of generated videos, ensuring they closely match the input descriptions.

2.3.2 Diverse Output

The system shall ensure diversity in generated video outputs to provide users with varied results for the same input text.

2.4 User Interface

2.4.1 Input Interface

The system shall provide an interface for users to input textual descriptions easily.

2.4.2 Output Interface

The system shall present generated videos to users through an intuitive interface.

3. Non-Functional Requirements

3.1 Performance

3.1.1 Speed

The system shall generate videos within a reasonable timeframe to ensure a responsive user experience.

3.2 Reliability

3.2.1 Error Handling

The system shall handle errors gracefully and provide informative error messages to users.

3.3 Security

3.3.1 Data Privacy

The system shall ensure the privacy and security of user data, adhering to relevant data protection regulations.

4. System Constraints

4.1 Computational Resources

4.1.1 Hardware Requirements

- The system may require substantial computational resources, including high-performance GPUs, for efficient operation.

5. Glossary

Generative AI:Artificial intelligence techniques used to generate data that is similar to, but not the same as, the input data.

Diffusion Model: A generative model capable of synthesizing high-quality images or videos by iteratively applying diffusion processes.

6. References

[1] Grathwohl, W., et al. (2021). *Your Local GAN: Designing Two Dimensional Local Attention Mechanisms for Generative Models*. arXiv preprint arXiv:2105.08050.

[2] Nakkiran, P., et al. (2021). *The Anatomy of an Excellent Image Generation Model*. arXiv preprint arXiv:2108.10343.

7. Revision History

- Version 1.0: Initial Draft (Date: [INSERT DATE])
- Version 1.1: Revised based on feedback (Date: [INSERT DATE])

This Software Requirement Specification provides a comprehensive overview of the requirements for developing a text-to-video generation system utilizing Generative AI with the Diffusion Model.