

# AI Avatar Single Video Generator – Report

## 1. Introduction

This project is a simple text-to-video generator that creates a short video based on user input. The system takes four inputs (prompt, style, camera angle, and duration) and uses a text-to-video model to generate a small MP4 clip. The whole implementation was tested and run in Google Colab.

## 2. Pipeline Design

### a. User Input

The user provides: prompt, style, camera angle, and duration.

### b. Prompt Construction

Inputs are merged into:

{subject\_action}. Style: {style}. Camera angle: {angle}.

### c. Model Loading

Uses the damo-vilab/text-to-video-ms-1.7b model with fp16, CPU offloading,

and VAE slicing.

d. Frame Generation  
Frames = duration \* fps (fps = 8).

### e. Video Assembly

Frames exported into outputs/output.mp4.

### f. Output

Notebook displays the final MP4 video.

### 3. Model/API Used and Reasoning

Model: [damo-vilab/text-to-video-ms-1.7b](#).

Reasons: open source, works with Diffusers, runs on T4 GPU, suitable for short video generation, no paid API needed.

### 4. Parameter Handling Strategy

Parameters: --prompt, --style, --angle, --duration.

Duration capped at <=10 seconds. Default values exist. All inputs merged into one prompt.

### 5. Evaluation and Observations

Simple prompts give clearer results. Complex scenes may look blurry. First run is slow due to model download. Short durations work best on Colab.

### 6. Conclusion

The project meets the task requirements: simple interface, parameter handling, model integration, and video output generation.