Enhancing Narratives with SayMotion's text-to-3D animation and LLMs

Kevin He
DeepMotion, Inc.
San Mateo, California, USA
kevin@deepmotion.com

Annette Lapham
DeepMotion, Inc.
San Mateo, California, USA
annette@deepmotion.com

Zenan Li
DeepMotion, Inc.
San Mateo, California, USA
zenan.li@deepmotion.com

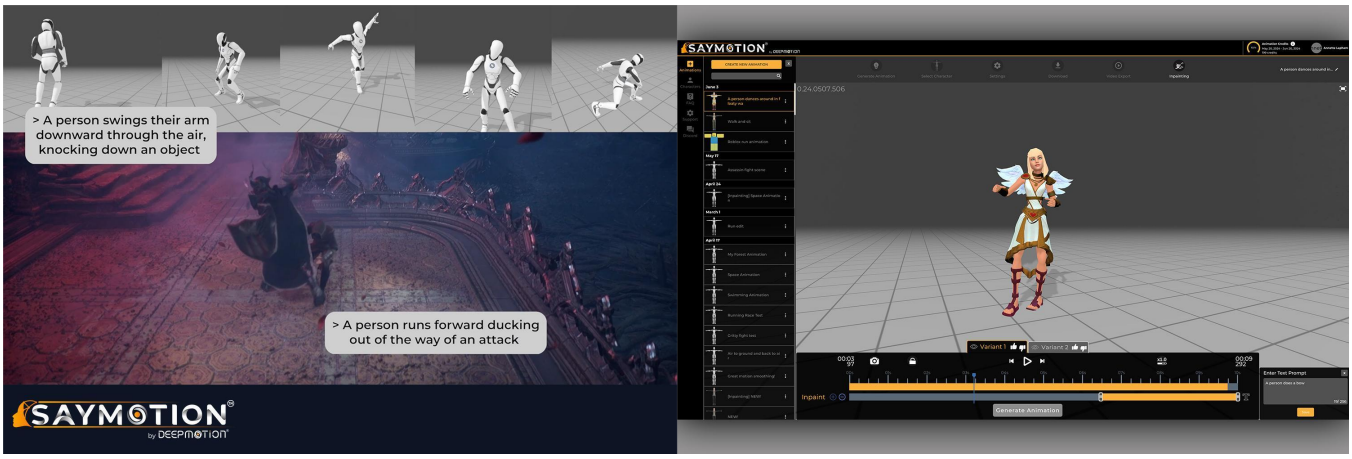

Figure 1: Preview of SayMotion

ABSTRACT

SayMotion, a generative AI text-to-3D animation platform, utilizes deep generative learning and advanced physics simulation to transform text descriptions into realistic 3D human motions for applications in gaming, extended reality (XR), film production, education and interactive media. SayMotion addresses challenges due to the complexities of animation creation by employing a Large Language Model (LLM) fine-tuned to human motion with further AI-based animation editing components including spatial-temporal Inpainting via a proprietary Large Motion Model (LMM). SayMotion is a pioneer in the animation market by offering a comprehensive set of AI generation and AI editing functions for creating 3D animations efficiently and intuitively. With an LMM at its core, SayMotion aims to democratize 3D animations for everyone through language and generative motion.

ACM Reference Format:
Kevin He, Annette Lapham, and Zenan Li. 2024. Enhancing Narratives with SayMotion's text-to-3D animation and LLMs. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Real-Time Live! (SIGGRAPH Real-Time Live! '24), July 27–August 01, 2024, Denver, CO, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3641520.3665309

SIGGRAPH Real-Time Live! '24, July 27–August 01, 2024, Denver, CO, USA

ACM ISBN 979-8-4007-0526-7/24/07
https://doi.org/10.1145/3641520.3665309

1 INTRODUCTION

Text-to-3D animation (also referred to as text-to-motion or language-to-motion) generation is a promising multimodal generative AI technology that holds vast potential to transform gaming, extended reality (XR), film production, education and interactive media. . The currently available techniques are limited to the generation of short individual animation clips that can be difficult to refine to production quality and combined into a coherent animated story. Additionally,the modality gap between natural language descriptions and humanoid motions makes it challenging for end users to articulate the movements of their characters in an intuitive and effective way. To address these challenges, SayMotion offers a comprehensive AI-based animation editing feature set including Multi-Variants, Inpainting, Outpainting, Merging, Refining, …, etc.. These tools empower artists to intuitively edit their animations and build their narratives towards their artistic vision by using natural language as the interface. This results in animations that capture the detailed and expressive body language required for narrative storytelling.We further simplify the text-to-3D animation prompt engineering process by equipping the platform with a LLM fine-tuned to the motion modality to automatically translate natural narrative languages into human movement-oriented text prompts to enhance generation quality. As a result we demonstrated how the SayMotion platform can be used to convert a text-based short story into an animated short film in a matter of 6 minutes.

The key contributions of SayMotion are the following:

- Intuitive text-to-3D animation platform accessible in a no-code format in a web browser, democratizing 3D animation for creators of any level of experience.

Table 1: SayMotion features and their supported inputs

Feature	Text	Motion
Prompt Generation		
Prompt Craft	✓	
Motion Generation		
Text-to-motion	✓	
Motion Editing		
AI Inpainting	✓	✓
AI Merge	✓	✓
AI Refinement	✓	✓

- Robust AI-editing feature set that empowers an intuitive human-in-the-loop animation creation process.
- An LLM fine-tuned to the motion modality to address prompt engineering challenges faced by text-to-3d animation users.

2 UNDERSTANDING SAYMOTION

DeepMotion recently unveiled SayMotion, their new generative AI text-to-3D animation platform that leverages generative deep learning and physics simulation to simplify the process of animating 3D digital humans. This innovation allows for the intuitive translation of textual descriptions into complex human motions, bridging the gap between conceptualization and visualization with minimal user input. Artists can now perform AI Inpainting with text prompts to intuitively and efficiently edit and refine their AI-generated animation to achieve their artistic vision. Early adopters utilizing the language-to-3D animation tool powered with DeepMotion's Large Motion Model confirmed a prolific challenge; accurately describing human motions in exact detail the way they imagine it while also staying within the bounds of how SayMotion's AI models are trained. To solve this problem, SayMotion integrated an LLM fine-tuned to the 3D humanoid motion modality to generate detailed, accurate prompts, from users' input for optimal results. Another enhancement that addresses the core value proposition of SayMotion is the Merge feature, allowing one to seamlessly merge multiple animation clips together with generative connecting motion segments. This allows for the creation of continuous, story-driven 3D motion narratives. Seamless narrative creation pushes the boundaries of storytelling within digital 3D environments, enhancing experiences requiring true 3D motion. This includes applications in gaming, XR, film production, education, interactive media and animation projects providing more lifelike and varied character movement.

3 UTILIZING SAYMOTION

In this section, we introduce SayMotion based on its functions. We delineate five primary functions and categorizes them into three modules according to their respective purposes, as illustrated in Table 1

3.1 Prompt Generation

SayMotion's Prompt Craft tool allows the input of a text-based story and will automatically break the story into a sequence of action-based shots and generate a text prompt for each action shot. Users can fine-tune the text prompts for eachshot and instruct SayMotion to generate animation clips for some, or all the shots, using them as the building blocks for the animated story.

3.2 Motion Generation

Given a text prompt in a structure of subject-action-details (e.g. "a girl skips happily in a circle") generates multiple variants of 3D human animation clips of maximum 10 seconds each. Users select the best one to further edit or polish its quality and content.

3.3 Motion Editing

Given the raw animation clips generated with SayMotion, users can utilize various AI editing tools as the following to improve the quality of the clips and merge them into a longer clip to tell the entire story. The editing process is conducted with text prompts and a simple graphical user interface for arranging and editing the animation clips.

3.3.1 AI Inpainting. With Inpainting users can selectively edit a portion of the animation clip with AI-generated replacement motion conditioned on an Inpainting text prompt. Temporal inpainting allows the user to select the Inpainting section along the temporal axis of the animation. Spatial Inpainting lets users select the Inpainting section by the body parts of the underlying humanoid rig.

3.3.2 AI Merge. Users can use the Merge function to combine multiple individual animation clips into one extended clip using AI to generate the transition clips to make the result one seamless animation.

3.3.3 AI Refinement. Empowers users to edit their animations into different styles, providing smoothing and auto-correction features to refine flawed motions. Users can fine-tune their animations using the creativity hyperparameter and the Refining text prompt.

4 INTEGRATION INTO REAL-TIME ENGINES

SayMotion enables seamless integration of its generated animation assets into real-time engines by exporting in industry-standard .FBX, .BVH, and .GLB file formats, compatible with standard humanoid rigs in all prevalent 3D engines and editor tools. Users can also import custom character models into SayMotion for animation via automatic character model import.

SayMotion also provides a REST API for generating and editing 3D human animations live in real-time in games or applications. This opens an entire new horizon for developers to develop interactive applications that render a vibrant animated virtual world with dynamically generated animations driven by conversational natural language.

5 FUTURE WORK

Besides body animation generation and editing, facial and hand animations are also important to portray a complete character-centric animated story. SayMotion will be extended to cover the generation and editing of body, hand and facial animations through AI. Additionally, the support for multi-modal prompts including text, image, video, and motion will be supported to further enrich the creative interface available to users.