ChatAvatar: Creating Hyper-realistic Physically-based 3D Facial Assets through AI-Driven Conversations

Qixuan Zhang
ShanghaiTech University and Deemos
Technology Co., Ltd.
Shanghai, China
zhangqx@deemos.com

Longwen Zhang
ShanghaiTech University and Deemos
Technology Co., Ltd.
Shanghai, China
zhanglw2@shanghaitech.edu.cn

Lan Xu
ShanghaiTech University
Shanghai, China
xulan1@shanghaitech.edu.cn

Di Wu
Deemos Technology Co., Ltd.
Shanghai, China
wudi@deemos.com

Jingyi Yu
ShanghaiTech University
Shanghai, China
yujingyi@shanghaitech.edu.cn

Figure 1: Preview of ChatAvatar.

ABSTRACT

ChatAvatar revolutionizes the process of creating hyper-realistic 3D facial assets with physically-based rendering (PBR) textures through AI-driven conversations, specifically text-to-avatar interactions. By leveraging advanced diffusion technology and our comprehensive Production-Ready Facial Assets dataset, we generate CG-friendly assets that adhere to industry standards and seamlessly integrate into popular platforms like Unity, Unreal Engine, and Maya. This groundbreaking technology opens up new horizons for immersive virtual experiences, pushing the boundaries of realism and interactivity.
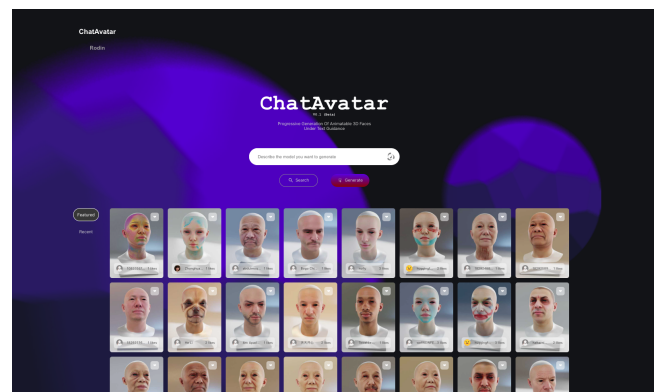
1 INTRODUCTION

Figure 3: Generated facial assets of ChatAvatar.

Figure 2: Integration into Unreal Engine.

ChatAvatar is a cutting-edge system that enables the creation of hyper-realistic 3D facial assets through AI-driven conversations. The key contributions of this paper are as follows:

- Introducing a progressive scheme for generating 3D facial assets under text guidance.
- Presenting a coarse-to-fine scheme for geometry generation and a dual-path mechanism for appearance generation.
- Demonstrating the seamless integration of ChatAvatar-generated assets into popular real-time engines.
- Showcasing the potential of ChatAvatar in creating immersive virtual experiences.

ChatAvatar advances the state of the art in real-time asset creation and open up new avenues for realistic and interactive virtual environments.

2 UNDERSTANDING CHATAVATAR

ChatAvatar employs a robust technical process centered around a coarse-to-fine scheme to generate neutral facial geometry with a consistent topology. The system utilizes a CLIP embedding space selection strategy to generate coarse geometry, followed by detailed displacements and normals optimization using Score Distillation Sampling (SDS) from the Latent Diffusion Model (LDM). To generate realistic appearances, ChatAvatar combines the generic LDM with a texture LDM, employing a dual-path mechanism that ensures diversity and precise textural specification in the UV space. Furthermore, a two-stage optimization approach performs SDS in both latent and image spaces, providing compact priors for fine-grained synthesis and mapping into physically-based textures such as diffuse albedo, specular intensity, and normal maps.

3 UTILIZING CHATAVATAR

ChatAvatar presents a novel approach to generating facial assets through user conversations. Utilizing an AI-driven framework, users engage in a chat dialogue with ChatAvatar to describe their desired facial appearance. The AI analyzes and summarizes the user's description, establishing a feedback loop that allows users to review and update the generated results within the ongoing conversation. Once the conversation concludes, ChatAvatar promptly generates the desired facial asset.
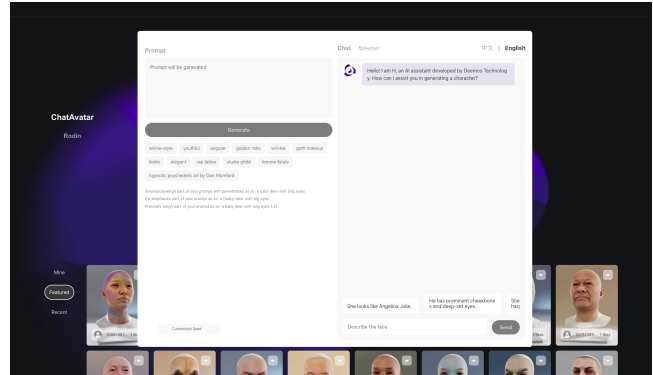

Figure 4: Prompt user interface.

The implications of ChatAvatar's work in the real-time space are significant. With an impressive asset generation time of just 20 seconds, the system is ideal for interactive live demonstrations, making it an invaluable tool for creating high-quality 3D facial assets for video games, virtual environments, and other real-time applications. Moreover, this breakthrough technology opens up possibilities for generating realistic 3D facial assets from video footage, even for unique characters in cartoons and fictional movies.

4 INTEGRATION INTO REAL-TIME ENGINES

The seamless integration of ChatAvatar-generated assets into real-time engines is a crucial factor for their practical implementation within virtual environments. With its consistent topology and physically-based textures, the facial assets produced by ChatAvatar are readily compatible with popular engines. The maintained consistency in topology facilitates immediate utilization of the generated facial assets, enabling real-time manipulation through generated expression blendshapes. Additionally, the physically-based textures ensure realistic rendering of the facial assets within real-time engines, enhancing the overall visual fidelity. By adhering to industry standards and leveraging efficient data structures, the integration process ensures the smooth operation and optimal performance of ChatAvatar assets within real-time environments.

The ChatAvatar project holds great value for the SIGGRAPH and real-time communities, offering a game-changing solution for the creation of immersive virtual experiences. Its ability to rapidly generate high-quality assets while adhering to industry standards makes it a powerful tool for professionals in computer graphics, game development, and virtual reality.