

# Text-to-Avatar

2022 - Current

Project Leader: Dr. Yajie Zhao and Yunxuan Cai

## Background

The creation of realistic digital human faces has long been a key objective in both industry and academia. As digital human applications continue to advance, there is a growing need for an efficient creation tool, especially for users with limited expertise in the field. Such a tool should enable users to generate high-quality, customizable digital humans that are compatible with modern physically-based rendering pipelines. It should also allow precise control over facial attributes and semantics, as well as easy editing after asset generation. A framework like this could significantly impact industries such as gaming, film, and AR/VR teleconferencing, and has significant value for forensic science, soldier training with personalized doubles, and deepfake detection.

## Objectives

To overcome the limitations of human face diversity in controlled scan-based datasets, we introduce a novel, industry-quality virtual human generator, Text2Avatar. This tool translates user inputs into high-fidelity 3D avatars, enabling extensive customization and realism. Trained on a diverse dataset of 44,000 high-quality face models with detailed annotations—including age, gender, and ethnicity—our model ensures broad representation of human features across global populations. Through integration with web-based software, we make digital human creation more accessible and efficient for downstream tasks.

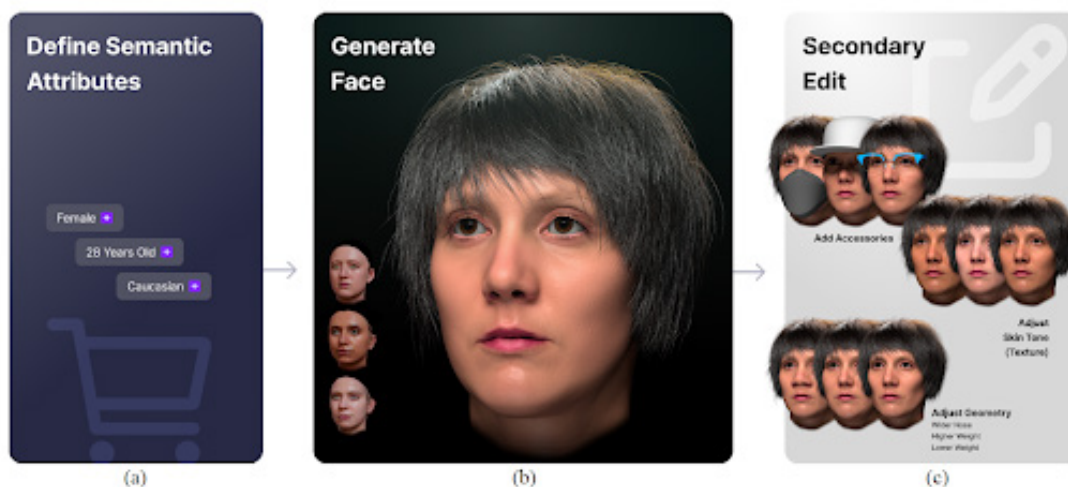


Image: Text2Avatar Generator: (a) User-defined semantic labels. (b) Avatars generated and rendered using all the assets. Users can generate multiple avatars and select their preferred subject. (c) Our model also supports texture space and geometry space editing, and offers a handcrafted accessory database (e.g., hairstyles, hats, glasses) for users to choose from.

## Results

The Text2Avatar platform offers precise semantic control, allowing users to create faces based on attributes such as age, gender, and ethnicity. It generates comprehensive PBR-based face assets, including base geometry, albedo, specular, and displacement maps, while also integrating secondary assets like eyeballs, teeth, and gums for fully realized models. Users can modify geometry and textures post-generation while preserving identity, ensuring flexibility in customization. The system produces high-quality, movie-grade 3D face models with complete texture maps and leverages a GAN-based approach for efficient asset creation. With access to a diverse database of over 44,000 high-quality albedo/geometry pairs annotated with demographic labels, our tool supports a broad range of representations. Each model comes fully rigged and ready for animation, ensuring compatibility with industry-standard 3D software and game engines. Additionally, an interactive web-based UI provides a user-friendly experience for exploring and generating face models.

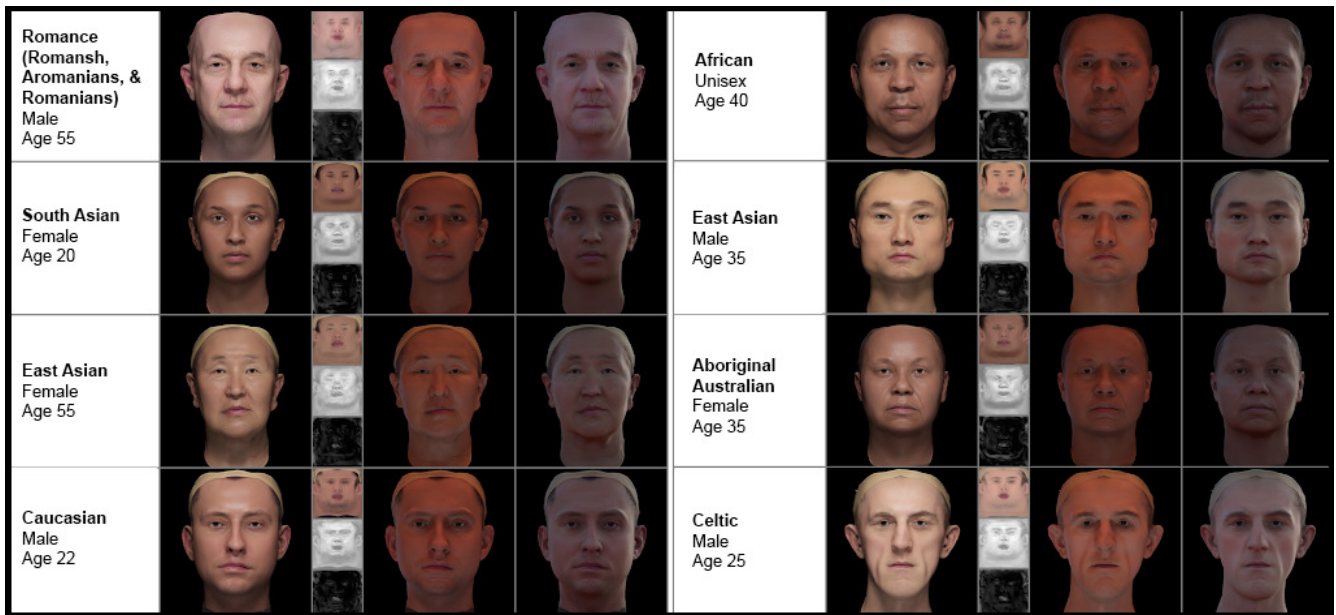


Image: Virtual humans generated by our Text2Avatar system. For each user input, we show the generated avatar under different lighting conditions. Each result displays, from left to right: input semantic labels, generated avatar in flat lighting, texture maps, and the avatar rendered in dark and partial lighting.

For more information, please visit <https://vgl.ict.usc.edu/>

**Project Leader: Dr. Yajie Zhao and Yunxuan Cai**

Established in 1999, the USC Institute for Creative Technologies (ICT) is a Department of Defense (DoD) University Affiliated Research Center (UARC), sponsored by the US Army. Harnessing technology, creativity, academic innovation and military-domain expertise, ICT conducts award-winning R&D in Artificial Intelligence (AI), Computer Graphics, Geospatial Sciences, Human Performance, Learning Sciences, Modeling, Simulation & Gaming, Mixed Reality (MxR), Medical VR, Narrative, and Virtual Humans.