

Technical Description of the Live Dialogue System

The avatar called Socrates, (to the right, presented in different gender and racial formats) is operated directly and live by a human hidden in the system. The actor's face is captured through a geometrical smart grid and mapped onto the avatar's face. An attached live data stream application enables the actor to operate the avatar directly and in real time, generating a lifelike mimicry of human expressions on the avatar's face. The actor's spoken words then trigger a pre-encoded pipeline of multiple AI components to enable a live conversation: The human voice output is converted simultaneously into text which prompts the second avatar, named von Kempelen. Von Kempelen is creating text through a natural language processing AI system. We use statistical based probabilistic language models called Large Language Models to generate text (such as the GPT series or PaLM 2 and others). The text is simultaneously converted into audio. Through an audio-to-face AI module, von Kempelen's machine-generated audio output also operates his own lip sync, gestures, facial expressions, and body movements.