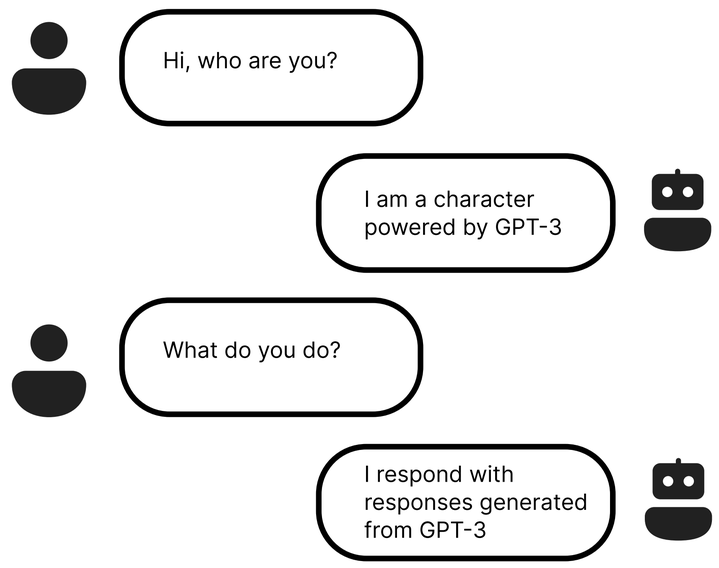
Personalizable Virtual Experiences: Creating Virtual Reality Content From Text Using Generative Artificial Intelligence

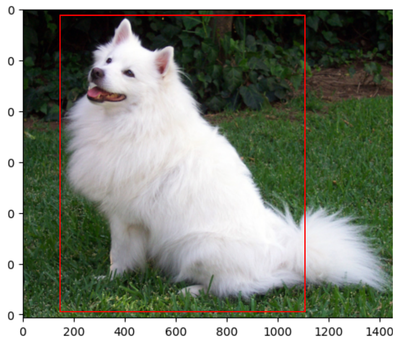
Brian Liang

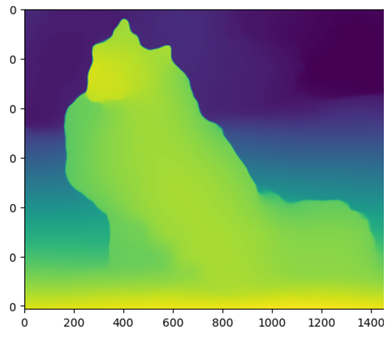
We aim to personalize virtual experiences such that virtual reality systems will have the capability to provide users with any experience they want.

A galaxy far far away	
A relaxing cruise to the bahamas	
Ten thousand leagues under the sea	

Addressing the challenge of generalizing virtual experiences may seem unachievable. We attempt to take a step in the right direction with a **chain generation and composition approach**. We begin by taking in as input a text description of a digital experience from the user. The user shouldn't have to write too much, but the system also needs to have enough details to be able to create the experience. We generate these details by feeding the text description through a language model. From this, we obtain viable descriptions of elements that make up the experience. We then take these descriptions and use a text to 3D model to create graphics for each element. The graphics are then composed into a 3D scene. Elements in the scene are given interactive attributes from their descriptions using language models. The composed scene can then be converted into a virtual reality experience.







To gain some intuition behind the system, we can begin by making the assumption that for all experiences, there is a way to represent each one digitally. This must be true because any digital experience is just some combination of pixels, and every combination of pixels can be shown on a screen. We use large generative models as models of the world; because they are built on massive quantities of data, we can make the assumption that they accurately represent a distribution of the digital world. Through prompting, we can then sample the parts of this distribution that represent the user's description, and convert this sample into a 3D representation to create a virtual experience matching the description.

Our results were reasonable, but left a lot to be desired. The generated graphics were low quality due to both the limitations of the model and to our limited computational resources. The model often contained deformities. Dynamics were also challenging to simulate. The components would move around with seemingly random behaviors, instead of moving cohesively to push the plot of the generated experience. Scene layouts were also not the best, due to deformities in generated images. Given additional resources, a better way to generate 3D scenes would be to use data driven methods. In general, the textual aspects of the system worked well. Generated intermediary descriptions were accurate to the inputs. Generated responses from components were always accurate to the context of the interaction. Because of this, our system was able to generalize fairly well. One of the reasons our system may have had challenges with quality but performed well in generalization is because using large AI models breaks the assumption that we have a perfect model of the world; generative AI models by nature will make mistakes at the expense of generalizing well.