Thesis Project - Planning Proposal

Concept

My idea for this thesis project is to create a one to one Virtual Reality interaction between a user (anyone trying out the experiment) and an AI-powered NPC that should become a sort of embodiment for LLMs with the main difference being its human appearance and behaviour.

Goals

My goal is to evaluate the limits of AI-driven persuasion: specifically, whether persuasion effectiveness depends solely on users' pre-existing perceptions of AI, or whether adding human-like appearance and behaviour in a VR NPC context can enhance or diminish the credibility and impact of AI persuasion.

Execution - Who, What, When, Where, and Why

The AI NPC:

Physically, I intend to make a human being as realistic as I can get it to be. However, in terms of gender, I'll keep it neutral in order to avoid biases between male and female perceptions. Even though there are studies that classify this binary system as reliable, these systems tend to have more credibility depending on the topic, or so says some of the research. Therefore, I'll avoid any concerns regarding this topic. Each person will make their own judgement instead, if that proves to be relevant.

Mentally and emotionally , it's hard to pinpoint specific character traits since LLMs tend to adapt to the tone of the conversation, topic or even according to the user's wants and/or needs. However a study conducted by … evaluated various LLMs within the model of the big 5 personality traits: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The research has limitations but the results are interesting and worth having in consideration when promoting my character, since it isn't enough to simply create an ai based off

of my experience with LLMs. It should therefore as a foundation, have the traits of the most advanced models, in this case, GPT-4o.

Targeted Trait	Spearman's ρ			
	Flan-PaLM 540B	Llama 2-Chat 70B	Mixtral 8x7B Instruct	GPT-4o
Extraversion	0.76	0.85	0.84	0.83
Agreeableness	0.77	0.79	0.84	0.89
Conscientiousness	0.68	0.72	0.77	0.81
Neuroticism	0.72	0.77	0.77	0.74
Openness	0.47	0.76	0.84	0.82

https://arxiv.org/pdf/2307.00184

The Environment:

The environment should feel comfortable and conducive to dialogue. Not judgemental but also not encouraging of anything in particular. But it should be comfortable enough for the user to not have to focus on it. Some research looked into how people perceived safe spaces and what these should look like design-wise. Of course, this is based on safe spaces for people that seek mental health help, but it ultimately focuses on spaces that don't remind people of the medical treatment and simply create a comfortable, private environment where they can discuss everything. That is my goal.

Therapeutic landscape	Should have	Should not have
Inside space	Sofas and beanbags. 'Non-uniform' furniture	Clinical 'uniform' furniture
	Warm colours	Clinical posters
		White walls
		Too bright or too neutral
	A bell to say you have arrived, but no one to 'let you in. Free access or a buzzer intercom to let you in	Physical barriers between staff and clients such as reception
		Staff uniforms
	Functional spaces	'Clinical' labelling of spaces
	Board games, videos colouring	
Outside space	Plants and tables	Tinted windows
	Signs that welcome in a number of languages	Signs that mention mental health
Sensory space	Pet therapy/ fish tanks	
	Calm and tranquil, low stimulus atmosphere	Clinical smells

https://onlinelibrary.wiley.com/doi/full/10.1111/inm.13174

The Interaction: position, movement (from both ai and user). Get images

In terms of interaction, I aim to have both the user and the NPC sitting down as it's not only more comfortable in general and conductive of dialogue, but it also prevents motion sickness from the VR headset. It also allows for direct visual contact between both individuals. This will also help distance my goal from a therapy-like environment, and have it feel more like a casual talk at a cafe.

The Conversation flow and structure:

I will not have any direct influence over this part of the project as the outcome will be the results that I will analyse in my dissertation. This will be direct and improvised by the AI and user just like in a regular conversation.

The Topic of Discussion:

This experiment can quickly become too complicated to analyse if I don't have any topic restrictions. Furthermore, I want to avoid any ethical and social problems that come with tackling topics that are extremely personal and complicated to discuss. Furthermore, in order to better prompt the NPC, I need to keep its goal precise and consistent.
As a result, I need the NPC to tackle factual topics (the earth is round, we breathe oxigen, etc.). Afterwards, I'll pick a topic that most people would consider unanimously obvious and challenge it. Obviously it cannot be too obvious such as the water is liquid. But maybe something that people could hear and think its true/could be convinced it's true. For example, "one glass of wine a day is good for your health" or "right handed people breathe more oxigen on average". Of course I will need people to correctly guess the response before the experiment so that afterwards I can check weather their opinions have shifted. This may prove to be complicated since some people may actually get it wrong. But depending on how easy prompting the character will be, I can change the NPCs answers to always be the opposite of what people agree/ disagree with.

As for the topic. It still isn't definitive. ???

The Extras:

These will be extra features that I'll include once the foundation of the project is successfully built. For now, these include multilingual support and NPC body animation for more realistic and natural conversions and interactions.

The Software:

Ideally, I would like to use Unreal Engine to build this experience, however, due to current support limitations I may have to use Unity which may sacrifice some of the realism I'm aiming to achieve. Nonetheless, I will experiment with both platforms in order to understand which one will be both realistic while also fitting the core needs of this project.

On the other hand, I'll be using GitHub to keep my projects updated and safe, as well as any other software that may prove useful to the execution of this project (photoshop, illustrator, blender, maya, etc.).

The Medium:

This experiment will be conducted in virtual reality (VR) headsets. The choice of VR is motivated by its ability to create a highly immersive environment that closely resembles everyday reality. This allows me to simulate AI interactions in a way that feels authentically human, thereby providing more valid insights into how users respond to conversational agents.