AI-Enhanced VR Self-talk for Psychological Counseling: A Formative Qualitative Study

Moreah Zisquit, Alon Shoa, Ramon Oliva, Stav Perry, Anat Brunstein Klomek, Mel Slater, Doron Friedman

Submitted to: JMIR Formative Research
on: October 21, 2024

Table of Contents

AI-Enhanced VR Self-talk for Psychological Counseling: A Formative Qualitative Study

?Moreah Zisquit??[1]; ?Alon Shoa[2]; Ramon Oliva[3] PhD; Stav Perry[1]; Anat Brunstein Klomek[1] Prof Dr; Mel Slater[3] Prof Dr; Doron Friedman[2] Prof Dr

[1]Baruch Ivcher School of Psychology Reichman University Herzliya IL
[2]Sammy Ofer School of Communications Reichman University Herzliya IL
[3]Event Lab University of Barcelona Barcelona ES

Corresponding Author:
?Moreah Zisquit??
Baruch Ivcher School of Psychology
Reichman University
Herzliya
Herzliya
IL

Abstract

Background: Access to mental health services continues to pose a global challenge, with current services often unable to meet the growing demand. This has sparked interest in conversational artificial intelligence (AI) agents as potential solutions. Despite this, the development of a reliable virtual therapist remains challenging, and the feasibility of AI fulfilling this sensitive role is still uncertain. One promising approach involves using AI agents for psychological self-talk, particularly within virtual reality (VR) environments. Self-talk in VR allows for externalizing self-conversation by enabling individuals to embody avatars representing themselves as both patient and counselor, thus enhancing cognitive flexibility and problem-solving abilities. However, participants sometimes experience difficulties progressing in sessions, which is where AI could offer guidance and support.

Objective: This formative study aimed to assess the challenges and advantages of integrating an AI agent into self-talk in VR for psychological counseling.

Methods: We carried out an iterative design and development of a system and protocol integrating LLMs within VR self-talk. In addition, we conducted an exploratory study in which 11 participants completed a session including: identifying a problem they wanted to address, attempting to address this problem using self-talk in VR, and then continuing self-talk in VR, but this time with the assistance of an LLM-based virtual human. The sessions were carried out with a trained clinical psychologist and were followed by semi-structured interviews. We used qualitative analysis after the interviews to code and develop key themes for the participants that addressed our research objective.

Results: In total, four themes were identified regarding the quality of advice, the potential advantage of human-AI collaboration in self-help, the believability of the virtual human and other topics. The participants rated 8.3 out of 10 their desire to engage in additional such sessions, and more than half of the respondents indicated that they prefer using VR self-talk with AI rather than without it.

Conclusions: This exploratory study suggests that the VR self-talk paradigm can be enhanced by LLM-based agents, how exactly to achieve this, potential pitfalls, and additional insights.

(JMIR Preprints 21/10/2024:67782)
DOI: https://doi.org/10.2196/preprints.67782

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

AI-Enhanced VR Self-talk for Psychological Counseling: A Formative Qualitative Study

Moreah Zisquit, Alon Shoa, Ramon Oliva, Stav Perry, Anat Brunstein-Klomek, Mel Slater and Doron Friedman

Abstract

Background: Access to mental health services continues to pose a global challenge, with current services often unable to meet the growing demand. This has sparked interest in conversational artificial intelligence (AI) agents as potential solutions. Despite this, the development of a reliable virtual therapist remains challenging, and the feasibility of AI fulfilling this sensitive role is still uncertain. One promising approach involves using AI agents for psychological self-talk, particularly within virtual reality (VR) environments. Self-talk in VR allows for externalizing self-conversation by enabling individuals to embody avatars representing themselves as both patient and counselor, thus enhancing cognitive flexibility and problem-solving abilities. However, participants sometimes experience difficulties progressing in sessions, which is where AI could offer guidance and support.

Objective: This formative study aimed to assess the challenges and advantages of integrating an AI agent into self-talk in VR for psychological counseling.

Methods: We carried out an iterative design and development of a system and protocol integrating LLMs within VR self-talk. In addition, we conducted an exploratory study in which 11 participants completed a session including: identifying a problem they wanted to address, attempting to address this problem using self-talk in VR, and then continuing self-talk in VR, but this time with the assistance of an LLM-based virtual human. The sessions were carried out with a trained clinical psychologist and were followed by semi-structured interviews. We used qualitative analysis after the interviews to code and develop key themes for the participants that addressed our research objective.

Results: In total, four themes were identified regarding the quality of advice, the potential advantage of human-AI collaboration in self-help, the believability of the virtual human and other topics. The participants rated 8.3 out of 10 their desire to engage in additional such sessions, and more than half of the respondents indicated that they prefer using VR self-talk with AI rather than without it.

Conclusions: This exploratory study suggests that the VR self-talk paradigm can be enhanced by LLM-based agents, how exactly to achieve this, potential pitfalls, and additional insights.

Keywords: Virtual human; Large Language Model; Virtual Reality; Self-talk; Psychotherapy

Introduction

Access to mental health services and treatment is a major issue in all countries and cultures across the globe. Worldwide, major depression is the leading cause of years lived with disability and the fourth leading cause of disability-adjusted life years [17]. According to research conducted among the health systems of various nations, more than 20% of people will suffer from mental illness in

their lifetime [22]. Unfortunately, the current clinical workforce is insufficient to meet these needs. There are approximately nine psychiatrists per 100,000 people in developed countries [18].

This inadequacy in meeting the present or future demand for care has led to the proposal of technology as a solution. Particularly, there is a growing interest surrounding conversational agents or multipurpose virtual assistants [23]. These AI agents can be integrated with virtual reality (VR) to simulate a personal, realistic therapeutic environment to reduce feelings of unnaturalness for the patient.

One of the main benefits of using virtual therapists is that they may help to overcome some of the challenges of traditional therapy, such as lack of access to mental health care, discomfort, or embarrassment about discussing personal issues with a therapist in person. Additionally, the use of AI-controlled virtual humans in VR therapy can make the therapy experience less intimidating and more engaging for some patients, especially children or people with social phobia [2].

However, developing a virtual therapist is still challenging; despite much progress in generative AI, it is unclear whether AI can be relied on for such a sensitive role. An interesting alternative for a VR counseling experience that bypasses the need for a virtual therapist has been suggested by Osimo et al. [19] – VR self-conversation. In this paradigm, the participant switches between embodying two avatars – one representing themselves as a patient, typically a look-alike avatar, and the second representing themselves as a counselor. In this way, the participant can “experience themselves from the outside” and provide advice accordingly. This has the advantage of a meaningful VR experience with psychotherapeutic value, which nevertheless does not require either a human or an AI therapist.

Nevertheless, when using VR self-talk, participants sometimes get “stuck”. In such instances, we suggest LLMs can be useful – even if not yet qualified to replace human therapists, they can be used as conversational partners with some limited understanding of counseling and help enhance the process of self-conversation. This paper describes the process of adding such an “AI” component (an LLM-based agent) to VR self-talk. First, we describe the iterative design and development process, and lessons learned along the way. Finally, we describe the first version that we considered successful and a qualitative study in the context of motivational interviewing [11].

Background

Virtual agents for psychotherapy

Embodied AI applications in mental health care aim to improve the quality of care and control expenditures [5]. In addition, they also hold the promise of reaching underserved populations in need of mental health services and improving life opportunities for vulnerable groups. However, there is a persistent gap between current, rapid developments in AI mental health and the successful adoption of these tools into clinical environments by health professionals and patients [13].

The psychological implications of the representation of the virtual therapist have been studied for decades [6], showing mixed results: sometimes, a human appearance enhanced the effectiveness of an application, while at other times, it did not. A meta-analysis by [24] demonstrated that adding a face (as opposed to just voice or text) was more significant than the effect of realism; i.e., there was more gain in impact from having a face than from making that face more photographically or behaviorally realistic. [7] evaluated the impact of an agent's representation (in a non-immersive environment) and perceived emotion over the perceived social believability in the agent. They found that appropriate emotions conveyed through the agent's body, mainly related to the sense of competence and warmth, could lead to higher believability.

Self-talk in VR

In times of crisis, individuals may find their capacity to analyze issues and help themselves limited. Their ability to gain insight is constrained, and it often feels like there is only one solution or, in more dire circumstances, “no way out.” Interestingly, contemplating the problem as an onlooker or from a friend’s perspective can improve problem-solving capabilities, a phenomenon known as Solomon’s paradox [10]. The paradox states that people reason more wisely regarding other people’s social problems than they do about their own. Furthermore, a straightforward linguistic change from using “me/I” to “you” has also been shown to increase the psychological distance from personal problems, consequently alleviating the distress they cause and fostering cognitive flexibility [15]. Hence, a relatively straightforward shift in language and perception can enhance problem-solving capabilities.

Both findings are at the base of the VR self-conversation paradigm, first introduced by Osimo et al. [19]. The virtual environment comprises a consultation room and two avatars. One avatar represents the participants as themselves, and the other represents the participants as the counselor: the participant swaps between the avatars, thus enabling a unique experience of physically talking to oneself. The initial study compared two counselor avatars: one group interacted with an avatar resembling Sigmund Freud, while the other engaged with an additional avatar resembling themselves. The results indicated that utilizing the self-talk paradigm proved beneficial in finding more satisfying solutions and improved mood, with the Freud group exhibiting better outcomes.

Since, these results have been replicated in multiple domains. VR self-talk has been shown to improve motivation for weight loss in individuals suffering from obesity when they conversed with a future self who had lost weight [1]. Convicted offenders spoke to their future selves, reducing self-defeating behaviors such as alcohol consumption and violent behavior towards others [9]. An additional study showed that using VR self-talk when the “counselor” is an avatar of a leading athlete, such as Serina Williams or Lebron James, could improve adherence to an exercise regime [Levy et al., under review]. These examples all imply that using VR self-talk enhances cognitive flexibility, subsequently facilitating behavioral changes that previously appeared unattainable. Importantly, the design of such paradigms requires care and attention; a study using stereoscopic video revealed that besides the potential benefits of self-talk in VR, under some circumstances, the experience can also ‘backfire’ for some specific populations [16].

Despite the success of VR self-talk, authors report (private communication) that people occasionally tend “to get stuck” in the conversation and cannot provide a suitable solution or advice. Sometimes, participants want to end the session or do not know how to continue after a certain point. Often, people run out of advice from the counselor's point of view. This is where AI can play an important role. Utilizing large language models)LLMs(, we can provide individuals with the external help and ideas they need to continue making progress in their self-talk sessions. This use of AI can also be considered an intermediate step towards automated therapists.

Automated Dialogue

With advancements in machine learning, natural language processing, and deep neural networks, AI has become more capable of performing complex tasks and understanding human language. One of the most notable developments is the emergence of conversational agents such as ChatGPT based on LLMs [4]. LLMs have seen tremendous advances in recent years thanks to increased computational power and the availability of vast datasets for training. Models today, with many billions of parameters, can generate remarkably human-like text and engage in dialogue while demonstrating some reasoning capabilities. However, significant challenges remain in making these models more aligned with human values, interpreting instructions correctly, and generating factually accurate statements. Current models may generate plausible but incorrect or nonsensical text. Thus, while

LLMs today are impressive and valid for specific applications, they require close monitoring and oversight before being deployed into sensitive real-world settings, such as the clinical psychology domain.

LLMs can be the missing piece to the puzzle of VR in therapeutic mental health treatment. In this study, we examined whether LLMs can support participants in their self-talk when they run out of advice and understand how we can use an AI character to help people feel safe and not judged. We describe lessons learned during the development of this unique protocol, which involves a range of advanced techniques: body ownership illusions [22], self-talk in VR, and LLM-based virtual humans [20]. Finally, we evaluate the final version of the system with a qualitative study.

Iterative Design and Pilot Studies: Lessons Learned

VR Self-talk

The VR self-talk experience is based on the ability of participants to switch in and out of virtual bodies in VR. The sensorimotor contingencies yield a strong illusion of ownership of the virtual body [21]: the virtual body moves with the participant's movements, and the participants can see their virtual body in a virtual mirror. In our studies, the participants (as well as their avatars) were sitting, and the illusion was based on the upper body–hand and head tracking (head is visible only in the mirror) (Figure 2, top).

The participant is initially embodied in either a look-alike or a generic gender-matched avatar. The first part of the VR experience is intended to strengthen the virtual body ownership illusion: the participant engages in a short "embodiment" exercise, alternately moving both hands, looking around, and seeing themselves in the mirror. Following the embodiment exercise, the participant is asked to describe a current life issue. Importantly, although it was made clear to the participants that the session was experimental and should not be considered psychological counseling, our sessions were carried out with a trained clinical psychologist. When they finish, the participants press a button, transitioning them to the counselor avatar. Previous studies found that it is best to model the counselor avatar based on famous inspiring persons; we have selected Barack Obama in our studies. Next, the participants press a button and listen to a replay of what they just said, spoken from the avatar representing themselves as patients. A prompt instructs them to respond "like a counselor," and their response is recorded. The participant's pitch is changed in the recording to avoid having it sound exactly like the participant's, as described in Osimo et al. [19]. Once completed, participants return to their look-alike avatar to continue the conversation in this iterative manner.

During the different stages of the study, we have used two VR self-talk implementations; one is a commercial product called ConVRself developed by Kiin Tech, and the other is a research version with very similar design and functionality.

LLM Integration

Integrating the LLM into VR self-talk first requires text-to-speech and speech-to-text functionalities. To address this, we employed the Anonymoys system [anonymous] developed in our lab. The system transcribes participant voice responses from within the VR session to text, which then serves as input for the LLM. Once the LLM generates a response, the text is converted to speech and played through the Unity application.

Initially, the AI was represented as a "help" button with voice functionality (Figure 1). Participants were instructed that a virtual counselor was listening to their conversation, and they could receive its

input if they felt stuck and did not know how to continue in the self-conversation. We have evaluated this as part of a study comparing VR self-talk with a physical-world equivalent setup – the empty chair technique from Gestalt therapy [anonymous]. However, out of eleven participants in this pilot study, only one engaged the AI. It turned out that a short psycho-education tutorial regarding basic emotion regulation skills followed by VR self-talk resulted in an overwhelming experience; all participants were highly engaged in the session and did not remember they could ask for AI “help”. The minimal AI representation within the virtual environment – as only a button – can explain why it was ignored. In additional pilot evaluations, the combination of less-than-perfect speech recognition with less-than-perfect dialogue capabilities (see below) often resulted in low-quality generated responses from the agent.

To make the AI agent more salient in the experience, we replaced the clickable button with a virtual human (Figure 2, bottom). This required addressing the repetition of texts; a by-product of the VR self-talk paradigm is that each text is repeated twice: first, it is spoken by the participant live and recorded, and next, it is played back by the corresponding avatar to be experienced by the participant after the body switch. When there was a third character in the scene, these repetitions became confusing. Consequently, we restricted access to the AI only from the "counselor" avatar and instructed participants that interaction with the AI avatar was recommended only between speech turns.

Furthermore, significant progress was made in LLMs over the two years of development. Initially, we utilized a specific model developed in the lab. This was based on the 7B parameters version of the GPT-J model [3], fine-tuned on the two volumes of published counseling and psychotherapy data from Alexander Street Press [25]. The volumes are searchable collections of transcripts containing real counseling and therapy sessions and first-person narratives illuminating the experience of mental illness and treatment. The two volumes contain 3,500 session transcripts and more than 700,000 utterances between a counselor and a patient. We fine-tuned the model with an 80%-20% train-test split. While this model was state of the art at the time of the early sessions, rapid developments in LLMs rendered it obsolete, and we therefore replaced it with a pre-trained model by OpenAI (more details in section 3.2.2).

Our LLM interface provides an interface for a human operator, which can be used by the experimenters (for full details, see [19]). The ongoing automatically transcribed conversation appears in a text window. The operator can select parts of the text or even modify the text and send specific parts to the voice playback in the application. We have also explored the possibility of allowing the experimenter to decide when to intervene in the conversation.

To allow for a realistic experience, we designed and implemented the AI avatar to be gaze-activated, i.e., the virtual human representing the AI played an idle and silent animation loop, and only spoke after the participant stared at it for a duration of 1 second. Such gaze activation requires careful tuning: if the gaze duration was too short or the gaze area too broad, false positives might occur, whereas otherwise, the activation becomes unnatural. In our post-experimental interviews, none of the participants complained about this type of activation.

Another challenge was response latency. The delay includes: i) waiting for the participant to finish talking (indicated by clicking a controller button), ii) gaze activation of the AI avatar, iii) speech recognition in the cloud, iv) LLM: one prediction call (latency mostly depends on server configuration and context length), and v) cloud-based text-to-speech conversion. To mitigate latency, the system can be used in a continuous mode, i.e., the system is prompted to keep generating responses after each counselor utterance, based on the ongoing conversation, regardless of whether the AI agent was activated or not.

The consequence is that latency thus mostly depends on LLM prediction call latency. When using lab

based LLMs running from our own server, we could control the latency. The overall round-trip would take several seconds, and this was acceptable in the context of a counseling experience. Using OpenAI means you are dependent on their response time; in our study there were no major delays and none of the participants in the post-experiment interview (below) complained about latency.

Finally, we note that due to limitations in AI tools in many languages, we had to carry out the sessions in English, rather than the local native language. Hence the participants were selected to participate in the study based on their level of English speaking, ranging between conversational level English and mother-tongue. Language support limitations include all parts of the pipeline – recognition, dialogue, and generation quality.

Methods

The system was iteratively refined and tested as described above until pilot studies indicated that it operated smoothly. Consequently, we carried out a user study. Since the overall experience is very rich and overwhelming (as described below), we opted for an exploratory study, carrying out careful sessions including the presence of a trained clinical psychologist. Although it was made clear to the participants that the system is experimental and this was not a real counseling session, it is clear from the responses that participants behaved almost like in a “real” counseling session, and most participants went through a psychologically meaningful session, as evident from in-depth post-experiment interviews (see Section Results).

Participants

The study population comprised 14 ANON University students (5 female) aged 19 to 26 (M =23.43, SD =2.21) who received credit for their participation and signed an informed consent form. All participants had conversational-level English. Three participants were not included in the final sample due to miscellaneous technical problems that came up during their session.

Materials and Equipment

The VR environment featured a consultation room. The VR simulation was developed with the Unity game engine (Version 2020.3.20f). The VR headset Quest 2 (Oculus, California, US), including its hand-held controllers for embodiment (upper body tracking), was used. Quest 2 has a single LCD panel for each eye, with a display resolution of 1832*1920. The refresh rate of the panel is 120Hz. Its weight is approximately 500g, and it has a head strap that ensures comfort during prolonged use. In addition, Oculus Quest 2 delivers a comprehensive 6 degrees of freedom, providing participants with both rotational and positional tracking capabilities.

The software was described above. It includes a VR self-talk experience, and integration of an LLM-based agent based on MILO. In the exploratory study we used Barack and Michelle Obama, gender-matched, as the counselor avatars. In this study we used generic gender-matched avatars for the participant-as-patient avatar, rather than look-alike avatars. The AI avatar was Einstein. [Figure 2].

For LLM, we used GPT 3.5 using the following prompt: "You will now act as a motivational interviewer with a lot of experience in interpersonal psychotherapy. I am a counselor and will turn to you for advice while speaking to my patient. Sometimes, I will get stuck and address you, and in this case, you should try to act as an expert counselor and say something that would help me progress the session as best as possible. Please do not mention that you are an artificial intelligence in your replies. Act as a real person."

Semi-structured Interview

Participant experiences and thoughts regarding the AI were assessed using a semi-structured interview that lasted approximately fifteen minutes. Due to the exploratory nature of this complex experience, we opted for the flexibility offered by semi-structured interviews. We constructed 14 questions in advance, including: “Please describe this experience as if you were describing it to a friend who was not here,”; “What did you think of the facilitator's advice?” and “What part of the experience was beneficial?”. The questions were presented individually to the participants, and additional questions were added based on the participant's responses.

The actual conversation of the participants is logged by default as part of the system, including the AI comments. However, we opted not to analyze the conversations due to the privacy of the participants, and the content was removed.

Procedure

Upon arrival at the lab, the participants were given instructions about the experiment and filled out a consent form. They were told that they would have the opportunity to talk about a personal problem that causes them an average level of distress (on a scale of 1 to 10, a problem rated between 4 and 7). Then, they wrote a short sentence describing the problem in their words. Next, the participants donned the VR headset and performed the VR experience in which they conversed freely between their gender-matched avatars and matching Obama characters (Michelle for female participants and Barack for male participants). After a back-and-forth conversation between the participant and Obama's avatar, the participants choose when to end the session.

After the first session, the participants answered two questions regarding their experience: i) Please rate this experience on a scale of 1 to 10, 1=not useful at all, 10=very useful, and ii) On a scale of 1 to 10, how helpful was this self-conversation for solving your problem?

Next, it was explained to them that they would perform a similar session, only this time another third avatar, depicted as Albert Einstein, would be seated in the virtual consultation room. They were told that he would be listening to their conversation and that they should turn to him during the conversation to receive his input. The gaze activation was explained, and they were instructed to turn to Einstein only from the Obama avatar. The participants proceeded to conduct the AI-enhanced iterative VR self-conversation, and when they were done, the semi-structured interview was conducted.

Results

Five of the participants chose to speak about a problem related to work or school. Five of the participants chose to speak about a relationship problem and one participant chose to speak about difficulty managing stress.

The semi-structured interviews were transcribed and analyzed using the thematic coding method by two of the co-authors. Five participants preferred the session with the AI, 4 preferred the session without it, and the rest (2) did not address this. Seven out of 11 participants said that the AI's attitude towards them was positive, 1 said it was neutral, and the rest (3) did not address this.

The participants were asked two questions in the break between the first session (without AI) and the second (with AI). The usefulness of the session was rated 6.91 (SD=0.54) and the degree to which it helped solve their problem was rated 6.1 (SD=1.58). At the end of the interview the participants were asked to rate on a 1-10 scale the extent to which they would like to come back for another session; the mean response was 8.32 (std = 1.55), indicating a very high level of satisfaction from the experience.

The qualitative analysis revealed the following high-level themes.

Quality of advice

Several of the participants indicated that the advice provided by the automated agent were useful and insightful, for example:

> *S6: "...Einstein brought a lot of emotional considerations to the two-way problem I didn't think about. It was interesting… It sounds like if I would listen to him like everything would work out."*
>
> *S8: "For my questions it was very, very beneficial. It was. It was able to provide very thorough and clear and insightful points and information regarding specifically related to my problem related to the career path and choosing a career path. And it was really cool to see how the interaction with the computer algorithm was able to really contribute to my own self-reflection."*

Some had mixed opinions, but it is difficult to disentangle what may only be a result of previous unrealistic expectations from "AI", e.g.:

> *S4: "I think that on the one hand it was good, and he was saying smart things, but on the other hand it was a little bit too logical and technical, and sometimes emotional problems are a little more complex."*
>
> *S7: "...Good as I said, but it was a little bit surface level."*
>
> *S13: Like robotic smart. Like human, but inhuman smart. You know what I'm saying? Like he was too good of a psychologist. I don't know how to say it. It felt like a nice thing to say, but feels like I don't know how to describe it too, like, kitsch… The AI surprised me... Yeah, the AI was was kind of freaking yeah. Like it felt. It felt like too good. I don't know how even to describe it…Yeah, but also, I mean it was like in between. It was like really compassionate and understanding and like what the perfect answer from a psychologist should be. But the perfect answer is like not, not human to some degree. It also felt like he just blabbered a bunch of information, like instantly to me, like, Oh yeah, it's OK to be compassionate and it's OK to be. You shouldn't be perfect all the time, like, in one sentence. "*

Some of the stereotypic responses towards AI could be over-trusting it; such phenomena had been

already alerted [12]. This may be reflected in:

> *S11: "I think he just had a way of wording the same ideas that I had in my head, but sort of organizing them to a more sophisticated or organized manner which made them more believable or reliable, seem more reliable."*

Also, some of the "complaints" could also be a reasonable approach to a human counselor, e.g.:

> *S6: "...that it was not really practical, but it was like another part that you need to think about."*
>
> *Q: What wasn't practical about it?*
>
> *S6: Like for my problem. I need an answer, it's like a yes or no question and he really talked about like what it made you feel and maybe rephrase the problem, and there's no way to really rephrase it."*

Self-AI collaboration

In some cases, participants felt like the AI was a complementary advisor to themselves, and the combination was better than each one alone.

> *S4: "...I liked the mix of the AI and myself as the counselor…I think the advice I gave myself was taking into account more like an emotional feeling and maybe even a little bit too much. And he was a little too technical. So maybe a mixture of them? Yeah, a mixture of them would be perfect".*
>
> *S8: "I think in terms of advice, it was similar to what I said, although the AI was able to provide a more detailed and well-structured answer. The AI's examples and ideas were a bit more beneficial"*
>
> *S9: "And I liked the mix of him and Michelle together, that I, as Michelle gave a bit of a softer input and he was like, do this and this and this and this and that together was like, they filled each other…No, I think they were really complementary to each other. I benefited very much from both of them. I'm really grateful. I got to talk to myself…OK, more minds better."*

Believability and attitude towards AI

It seems that the content of the dialogue was more believable than the non-verbal behavior; subjects "complained" about the realism of several aspects such as voice and "body language":

> *S8: "... yeah, the sound and obviously the, the graphics obviously. But in terms of the syntax and the structure of how the answers were provided and delivered then it was fairly good. Fairly satisfying."*
>
> *S4: "I feel like a mixture, like the words that he was saying felt like they made sense, they were human-like, but the way that he was telling them and like it was very technical…"*
>
> *S5: "The conversation felt like very human and Einstein Avatar was a little bit robotic and felt like scripted and just a little bit like not related... No, the look was OK."*
>
> *S7: "I'm asking a question and then we'll just speak like we talk and talk and talk and talk… Humans usually wait a little bit."*
>
> *S7: "Technical stuff, which is like the character will move a little bit differently than human, and that's pretty much."*

Gradually overcoming the "breaks in presence" [8] is a well-known phenomenon, especially in

immersive VR; the place illusion [21] is so powerful that participants “want to suspend disbelief”; arguably more so than in the case of non-immersive virtual agents.

> *S8: “I think it's it's still not 100% human like but it did feel like I was talking to a very insightful person because at the end of the day I was I was focusing more on the information that was given to me by the algorithm rather than the actual feeling and and the the sound of the the interaction itself.”*

> *S9: “... at once I got over the metallic voice, yeah, it felt a bit more natural… It's like a fake character. It's an avatar, but I think when you get in, in into it, it's become less and less weird. Less and less, Yeah.”*

The use of celebrity avatars

Some comments were made about the selection of celebrities for the counselor and the AI assistant personas:

> *S5: “And I think Obama is quite a good character to use, actually. I feel that she's confident. She's a woman, I guess like a strong woman. She's very powerful. It seems like it's very fun to sit around her and like look at her and talk to her.”*

> *S7: “That's the point, Talking to Obama, even if, like, I don't agree with him that much. But like, yeah, it's a cool person to talk to.”*

> *S10: “I think I would change Barack Obama's character to someone that I love. As I said, one of the reasons I felt more comfortable with Einstein is because that's a character I can relate to and love more.”*

> *S13: “Again, maybe it's just because it's Einstein. I mean if like a human being would say something like that to me but would phrase it differently and, you know, have different facial expressions like not like the computer Einstein, then I would feel differently about the answer.”*

> *S14: “... Maybe because it was Einstein. Like maybe if it was someone else maybe I wouldn't think about like it was but just because it was him and I know that he was like really big scientist so he might have like great things to say.”*

Discussion

Mental health and well-being are a major challenge worldwide, and the demand for counseling is much greater than the supply. Technologies such as XR and AI may be a part of the solution. VR self-talk is unique in suggesting counseling without the need of either a human in the loop nor AI and has been shown in the past to be beneficial and effective. Adding an LLM assistant is a natural next step; nevertheless, introducing an LLM-based AI counselor assistant into XR is not trivial, and our iterative design process has revealed several important lessons.

Integrating multiple challenging technologies such as VR, body-tracking, dialogue, voice recognition, and speech generation into a meaningful psychological experience proved arduous. While each of these technologies separately has made impressive progress recently, accumulated problems or errors easily resulted in a non-usable experience. In the case described here, it seems that during 2023 the underlying AI technologies, specifically speech recognition and LLM-based dialogue, passed an essential threshold, increasing the probability of a successful and meaningful experience regardless of the complex nature of the technological design.

Based on the themes that emerged from the semi-structured interviews, we can conclude that participants generally responded positively to the AI advice, with reactions ranging from mixed to favorable. None of the participants evaluated it as poor or as obstructive to the self-talk process. Considering the complexity of the experience, this result is very encouraging. VR self-talk, when performed correctly, is a very powerful experience, though it can be confusing even without the addition of AI. To address this, we implemented a gradual protocol, allowing the participants to accustom themselves to the VR self-talk before introducing the AI agent in a subsequent session. This protocol was necessary for our preliminary research. However, future studies could benefit from a research design such as randomized control trials that could compare self-talk with and without AI, providing a deeper understanding of the intricate nature of this experience and the added value of AI to it.

The overarching goal of this endeavor was to design and implement an AI agent within a self-talk experience. We envisioned the AI advice as complementing the participants' own self-guidance, rather than replacing their need to think, reflect, and reason with themselves regarding the issue at hand. This vision was in alignment with the participants' responses. They found it straightforward to implement the AI advice into their self-dialogue and quickly adapted to the flow of the conversation. Our results further support the perspective that the integration of technology into counseling has the potential to enhance skills and abilities that humans already possess [26].

In our case, an "AI" button was not enough, and we opted to integrate a gaze-activated avatar to embody the AI. While participants commented that the voice and animation of the avatar were not completely realistic, several of the participants indicated that they were able to overcome such limitations and focus on the social interaction as well as the content of the conversation.

Finally, most of the participants pointed to the need to improve design elements such as voice, body language, avatar appearance, and character selection. There was a relative consensus among participants regarding the need for improvements in this area to enhance the overall experience. This is not surprising as the level of realism has been shown to have a strong impact on affective responses of participants [27-28]. Now that we have successfully integrated multiple technologies and created a seamless psychological experience, future development, and testing should focus on refining and improving these design aspects.

Limitations

This study has several limitations that should be addressed. Although multiple pilot studies were conducted throughout the iterative design process, including approximately 30 additional participants, the final sample consisted of only 11 participants. The relatively small sample size underscores the necessity for further research with larger and more diverse populations that could confirm these initial results. However, as emphasized throughout this paper, this is a novel and promising application of technology, and every exploration must begin somewhere. Due to the psychological focus of this study, further research should explore specific participant characteristics that could influence the effectiveness of self-talk in VR. Furthermore, research could target particular symptoms of psychopathology or specific diagnoses to improve understanding of potential clinical applications.

Another limitation regards the bias of novelty. The use of VR and AI in a psychological setting could have potentially fascinated the participants leading to an inflated sense of efficacy and distorting the true impact of the experience. Additionally, semi-structured interviews could have contributed to social desirability and acquiescence bias, as participants might have felt inclined to provide responses, they expected would be favorable to the researcher. Further research could explore ways to mitigate these identified biases.

Ethical considerations

The ethical considerations regarding implementing AI in psychological settings are paramount [27]. The development process and exploratory study outlined in this paper aimed to address preliminary questions concerning effectiveness and validation. In addition, in our study, a clinical psychologist provided professional oversight, which is typically recommended for AI implementation. Other critical issues for future research include rigorous testing for quality and reliability, given the sensitive nature of therapeutic applications. Additionally, confidentiality and privacy must be carefully addressed before real-world deployment.

Conclusions

The introduction of new technologies could potentially transform psychotherapy giving rise to numerous potential challenges, limitations and ethical considerations that should be addressed. Yet, given our encouraging results, we suggest that the paradigm of AI-enhanced VR self-talk may be ready for further research with the general population, and/or studies with clinical populations. Additionally, our work suggests there are numerous potential opportunities for integrating AI into XR wellness, extending beyond just “automated therapist agents”.

Acknowledgements

TBD

Conflicts of Interest

None

Abbreviations

AI: Artificial intelligence

VR: Virtual reality

References

1. Anastasiadou D, Herrero P, Vázquez-De Sebastián J, Garcia-Royo P, Spanlang B, Álvarez de la Campa E, Slater M, Ciudin A, Comas M, Ramos-Quiroga A, and Lusilla-Palacios P. Virtual self-conversation using motivational interviewing techniques to promote healthy eating and physical activity: A usability study. Frontiers Psychiatry; 2023 (14): 999656 https://doi.org/10.3389/fpsyt.2023.999656
2. Bell I.H, Nicholas J, Alvarez-Jimenez M, Thompson A, and Valmaggia L. Virtual reality as a clinical tool in mental health research and practice. Dialogues Clin. Neurosci.; 2020 (22.2): 169-177 https://doi.org/10.31887/DCNS.2020.22.2/lvalmaggia
3. Black S, Biderman S, Hallahan E, et al. Gpt-neox-20b: An open-source autoregressive language model. arXiv preprint; 2022: 2204.06745. https://doi.org/10.48550/arXiv.2204.06745
4. Brown T.B, Mann B, Ryder N, et al. Language models are few-shot learners. In Advances in Neural Information Processing Systems; 2020 .
5. Cresswell K, Cunningham-Burley S, and Sheikh A. Health care robotics: Qualitative exploration of key challenges and future directions. J. Med. Internet Res; 2018 (20.7): e10410. https://doi.org/10.2196/10410
6. Dehn D.M and Van Mulken S. Impact of animated interface agents: a review of empirical

research. Int. J. Hum. Comput. Stud ; 2000 (52.1): 1-22. https://doi.org/10.1006/ijhc.1999.0325

7. Demeure V, Niewiadomski R, and Pelachaud C. How is believability of a virtual agent related to warmth, competence, personification, and embodiment? Presence Teleoperators Virtual Environ.; 2011 (20.5): 431-448. https://doi.org/10.1162/PRES_a_00065
8. Garau M, Widenfeld H.R, Antley A, Friedman D, Brogni A, and Slater M. Temporal and spatial variations in presence: A qualitative analysis. In The 7th Annual International Presence Workshop; 2004: 293–309.
9. van Gelder J.L, Cornet L.J, Zwalua N.P, Mertens E.C and van der Schalk J. Interaction with the future self in virtual reality reduces self-defeating behavior in a sample of convicted offenders. Scientific Reports; 2022 (12.1): 2254. https://doi.org/10.1038/s41598-022-06305-5
10. Grossmann I and Kross E. Exploring Solomon's Paradox: Self-Distancing Eliminates the Self-Other Asymmetry in Wise Reasoning About Close Relationships in Younger and Older Adults. Psychological Science; 2014 (25.8): 1571-1580/ https://doi.org/10.1177/0956797614535400
11. Hettema J, Steele J, and Miller W.R. Motivational interviewing. Annual Review of Clinical Psychology; 2005 (1.1): 91-111. https://doi.org/10.1146/annurev.clinpsy.1.102803.143833
12. Hitron T, Morag N, and Erel H. Implications of AI bias in HRI: Risks (and opportunities) when interacting with a biased robot. In ACM/IEEE International Conference on Human-Robot Interaction; 2023, March: 83-92. https://doi.org/10.1145/3568162.3576977
13. Ienca M, Wangmo T, Jotterand F, Kressig R.W, and Elger B. Ethical Design of Intelligent Assistive Technologies for Dementia: A Descriptive Review. Science and Engineering Ethics; 2018 (24): 1035-1055. https://doi.org/10.1007/s11948-017-9976-1
14. Kross E and Ayduk O. Self-Distancing: Theory, Research, and Current Directions. In Advances in Experimental Social Psychology; 2017 (55): 81-136. https://doi.org/10.1016/bs.aesp.2016.10.002
15. Landau D, Hasler B, Golland Y, Huebbe B, Idan O, Magnat M, Magidov E, and Friedman D. A Self-Compassion Experience in Immersive Video: Opportunities and Pitfalls. PsyArxiv Prepr. (2022).
16. Murray C.J, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study. The Lancet; 2012 (380.9859): 2197-2223 https://doi.org/10.1016/S0140-6736(12)61689-4
17. Oladeji B.D and Gureje O. Brain drain: a challenge to global mental health. BJPsych. Int; 2016 (13.3): 61-63. https://doi.org/10.1192/s2056474000001240
18. Osimo S.A, Pizarro R, Spanlang B, and Slater M. 2015. Conversations between self and self as Sigmund Freud - A virtual body ownership paradigm for self counselling. Scientific Reports; 2015 (5.1): 13899. https://doi.org/10.1038/srep13899
19. Shoa A, Oliva R, Slater M, and Friedman D. Sushi with Einstein: Enhancing hybrid live events with LLM-based virtual humans. In Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents; 2023, September, Germany. https://doi.org/10.1145/3570945.3607317
20. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences; 2009 (364.1535): 3549-3557. https://doi.org/10.1098/rstb.2009.0138
21. Spanlang B, Normand J.M, Borland D, Kilteni K, Giannopoulos E, Pomés A, et al. How to build an embodiment lab: Achieving body representation illusions in virtual reality. Frontiers

in Robotics and AI; 2014 (1): 1-22. https://doi.org/10.3389/frobt.2014.00009

22. Trautmann S, Rehm J, and Wittchen H.U. The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? EMBO Rep; 2016 (17.9): 1245-1249. https://doi.org/10.15252/embr.201642951
23. Vaidyam A.N, Wisniewski H, Halamka J.D, Kashavan M.S, and Torous J.B. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. The Canadian Journal of Psychiatry; 2019 (64.7): 456-464. https://doi.org/10.1177/0706743719828977
24. Yee N, Bailenson J.N, and Rickertsen K. A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In Conference on Human Factors in Computing Systems - Proceedings; 2007 https://doi.org/10.1145/1240624.1240626
25 2008. Counseling and Psychotherapy Transcripts, Client Narratives, and Reference Works. *Choice Rev. Online* (2008). https://doi.org/10.5860/choice.46-1766
26. Imel Z.E, Caperton D.D, Tanana M, and Atkins D.C Technology-enhanced human interaction in psychotherapy. Journal of counseling psychology; 2017 (64.4), 385-393. https://doi.org/10.1037/cou0000213
27. Newman M, Gatersleben B., Wyles K. J, and Ratcliffe E. (2022). The use of virtual reality in environment experiences and the importance of realism. Journal of environmental psychology; 2022 (79): 101733. https://doi.org/10.1016/j.jenvp.2021.101733
28. Weber S, Weibel D, & Mast F.W. How to get there when you are there already? Defining presence in virtual reality and the importance of perceived realism. Frontiers in psychology; 2021 (12): 628298. https://doi.org/10.3389/fpsyg.2021.628298
29. Fiske A, Henningsen P, and Buyx A. Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. Journal of Medical Internet Research; 2019 (21.5): e13216 doi: 10.2196/13216

Supplementary Files

Figures

A screenshot from the counseling scenario with AI as a help button on a virtual panel.

Screenshots from the counseling scenario. Top: Self-talk VR as seen through a virtual mirror. Bottom: Self-talk VR enhanced by an AI agent (Einstein avatar).