

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

Investigación y aplicación de técnicas de prompting para paciente artificial: Integrando chatGPT y Llama 2 en un simulador de consultorio psicológico.

Nicole Milena Caicedo Guzmán

Ingeniería en Ciencias de la Computación

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero en Ciencias de la Computación

Quito, 9 de mayo de 2024

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

Investigación y aplicación de técnicas de prompting para paciente artificial: Integrando chatGPT y Llama 2 en un simulador de consultorio psicológico.

Nicole Milena Caicedo Guzmán

Nombre del profesor, Título académico

Felipe Grijalva, Ph.D

Quito, 9 de mayo de 2024

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Nicole Milena Caicedo Guzmán

Código: 00213150

Cédula de identidad: 1718649310

Lugar y fecha: Quito, 9 de mayo de 2024

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

RESUMEN

Este proyecto profundiza en la exploración e implementación de técnicas de prompting utilizando dos modelos avanzados de lenguaje de gran tamaño (LLMs), específicamente ChatGPT y Llama 2. Estos modelos están integrados en una plataforma educativa de realidad virtual multiusuario de vanguardia, diseñada para proporcionar experiencias de aprendizaje inmersivas y colaborativas. El objetivo principal es refinar estos modelos para exhibir comportamientos más matizados y naturalistas asociados con diversos trastornos mentales. Al hacerlo, el proyecto busca permitir que los estudiantes de psicología interactúen con un paciente artificial que refleje de cerca los síntomas y comportamientos de una persona real.

El uso de técnicas de prompting es central en esta iniciativa. Estas técnicas permiten que los modelos entreguen respuestas que no solo se alinean más estrechamente con los comportamientos esperados de los pacientes, sino que también mantienen la consistencia y precisión en sus interacciones. Como resultado, el paciente artificial puede navegar escenarios más complejos, proporcionando respuestas que son apropiadas contextualmente y terapéuticamente relevantes.

El proyecto es base para futuras investigaciones sobre la importancia de la ingeniería de prompts dentro del campo de la psicología. Tradicionalmente, se ha considerado la posibilidad de reemplazar terapeutas con IA, pero este proyecto innova al enfocarse en el rol del paciente. Al simular un paciente artificial, el proyecto mejora la experiencia educativa para los estudiantes de psicología, permitiéndoles practicar estrategias de diagnóstico y tratamiento en un entorno libre de riesgos. Este enfoque reduce significativamente el potencial daño que podría surgir de diagnósticos erróneos al practicar con individuos reales.

Palabras clave: Técnicas de prompting, consultorio virtual, modelos de lenguaje de gran tamaño, entrenamiento, psicología, IA en educación, simulación de salud mental.

ABSTRACT

This project delves into the exploration and implementation of prompting techniques using two advanced large language models (LLMs), namely ChatGPT and Llama 2. These models are integrated into a cutting-edge multi-user virtual reality educational platform designed to provide immersive and collaborative learning experiences. The primary goal is to refine these models to exhibit more nuanced and naturalistic behaviors associated with various mental disorders. By doing so, the project aims to enable psychology students to engage with an artificial patient that closely mirrors the symptoms and behaviors of a real person suffering from these disorders.

The use of prompting techniques is central to this initiative. These techniques allow the models to deliver responses that not only align more closely with the expected behaviors of patients but also maintain consistency and accuracy in their interactions. As a result, the artificial patient can navigate more complex scenarios, providing responses that are contextually appropriate and therapeutically relevant.

Moreover, the project serves as a foundational study for future research into the importance of prompt engineering within the field of psychology. Traditionally, the focus has been on potentially replacing therapists with AI, but this project innovates by targeting the patient role. By simulating an artificial patient, the project enhances the educational experience for psychology students, allowing them to practice diagnosis and treatment strategies in a risk-free environment. This approach significantly reduces the potential harm that could arise from misdiagnoses when practicing with real individuals.

Keywords: Prompting techniques, virtual clinic, large language models, training, psychology, AI in education, mental health simulation.

TABLA DE CONTENIDO

INTRODUCCIÓN	9
DESARROLLO DEL TEMA	11
CONCLUSIONES	34
REFERENCIAS BIBLIOGRÁFICAS	36
ANEXO A: PRUEBAS PROMPTING	38

ÍNDICE DE FIGURAS

Figura #1 Respuesta como Alex mdd -mild	23
Figura #2 Respuesta como Taylor mmd- severe	24
Figura #3 Llama 2 prueba mdd	26
Figura #4 Llama 2 prueba paciente con mdd	27
Figura #5 Llama2 prueba paciente con np disorder template	28
Figura #6 Diagrama de flujo.....	29
Figura #7 Ejemplo Character.ai	30
Figura #8 Ejemplo Claude 3	31
Figura #9 Vista 1 de práctica con paciente	33
Figura #10 Vista 2 de práctica con el paciente	33

INTRODUCCIÓN

En los últimos años, la Inteligencia Artificial (IA) ha revolucionado diversos campos, incluyendo la psicología, donde los modelos de lenguaje grande (LLMs) han transformado las interacciones y las intervenciones terapéuticas. Estos modelos, desarrollados con arquitecturas avanzadas como los transformadores, han permitido simular conversaciones realistas que son fundamentales para la formación y la práctica en salud mental (Kapoor, 2021; Wei et al., 2022).

Paralelamente, la era digital ha remodelado la educación, introduciendo herramientas como la realidad virtual (RV) para crear experiencias educativas inmersivas y colaborativas que superan las barreras físicas y geográficas (Oyelere et al., 2020). Este proyecto integra IA y RV en un consultorio virtual educativo que ofrece una plataforma para la formación de futuros profesionales de la psicología, permitiendo interactuar con pacientes virtuales en un entorno controlado y realista.

En el corazón de esta experiencia educativa reside la utilización de LLMs para modelar diálogos con pacientes virtuales que exhiben diversos trastornos psicológicos, basados en criterios diagnósticos del DSM-5. Estos modelos procesan grandes cantidades de datos y aprenden patrones de lenguaje complejos, facilitando simulaciones de interacciones humanas con alta fidelidad que son esenciales para la práctica clínica en psicología (DeepLearning.AI, 2023).

Además, herramientas como Oculus Integration SDK y Meta Avatars SDK mejoran la inmersión en la RV, mientras que las APIs de Google Cloud para procesamiento de voz enriquecen la interacción, facilitando respuestas inteligentes y contextualizadas que mejoran la experiencia de aprendizaje. Photon Engine, por su parte, permite la creación de entornos

multiusuario y la sincronización en tiempo real, esencial para la colaboración y la interacción fluida entre estudiantes ubicados en diferentes partes del mundo.

El objetivo principal de este proyecto es mejorar la calidad educativa en el campo de la psicología, proporcionando a los estudiantes acceso a prácticas seguras y efectivas en un entorno virtual. Esta innovación reduce costos y mejora las habilidades de diagnóstico y comunicación de futuros psicólogos, representando un avance significativo en la educación virtual y estableciendo un camino hacia futuros desarrollos en este emocionante campo.

DESARROLLO DEL TEMA

1. Estado del arte

En la integración de pacientes artificiales con inteligencia artificial para formación de estudiantes de psicología, se destacan diversas tendencias y consideraciones clave. La ciberterapia y la realidad virtual son dos grandes integraciones de la tecnología para dar psicoterapia a personas que no cuentan con la capacidad de tener terapia en un lugar en específico. La realidad virtual ha emergido como una herramienta efectiva cuando lo que se busca es simular situaciones fóbicas en pacientes. La inteligencia artificial ha sido usada como parte de la terapia para ayudar a personas con diversos trastornos mentales. Sin embargo, cuando se busca información del aporte de la inteligencia artificial en el área de psicología, nos encontramos que esta estaría siendo desarrollada para sustituir al terapeuta, más no al paciente. Existen planteamientos sobre la sustitución de humanos por máquinas, argumentando que, aunque las máquinas superan a los humanos en precisión y rapidez, carecen de la creatividad, liderazgo y resolución de conflictos inherentes a la humanidad. Esto es clave ya que cuando se tiene al terapeuta y al paciente estos forman una relación donde el terapeuta entiende los sentimientos del paciente y siente empatía por este, por lo cual daría respuestas u opciones para tratar de que con conducta diferente este paciente se sienta “mejor”.

En cuanto a crear avatares que cuenten con inteligencia artificial para desarrollarse en un entorno tenemos experimentos en el área de periodismo en China. Aquí crearon avatares parecidos a periodistas muy queridos por el público y les dieron un guion en el cual basarse. Les proveyeron de una gran cantidad de gesticulaciones y lenguaje corporal para poder transmitir las noticias. Esto nos lleva a "Ellie", la cual es un programa de psicoterapia con inteligencia artificial. Esta incorpora reconocimiento facial y terapia cognitivo-conductual, desafiando la noción de que la terapia sin un terapeuta humano es imposible. Ellie tuvo gran

eficacia para reconocer trastornos mentales cuando se le alimentaba con lo necesario para poder dar un diagnóstico, pero el problema es que cada vez aprendía más sobre estos lenguajes no verbales y se le hacía muy fácil el dar un resultado erróneo. También podemos encontrar estudios de la exposición a ambientes restauradores mediante realidad virtual, revelando diferencias significativas según el formato de presentación de imágenes. En estos estudios podemos ver resultados favorecedores en cuanto a la exposición a estos ambientes si se toman ángulos de vista de 180 grados. Esto hace que no sean abrumadores, pero que si sumerjan en un estado restaurador a la persona.

Estudios abordan la delegación de decisiones éticas a máquinas, evidenciando que las máquinas pueden cambiar las expectativas sociales sobre las decisiones morales. Estas son buenas tomando decisiones si se les alimenta con la información correcta. Por otro lado, la empatía hacia las máquinas también se considera, mostrando que, aunque las máquinas no experimentan angustia, la empatía del observador puede influir en la aceptación y ética del uso de inteligencia artificial. Esto da una señal muy positiva ya que, en un principio el gran problema de la relación terapeuta y paciente es que, si el terapeuta es una inteligencia artificial, esta no será capaz de sentir empatía por su paciente. Sin embargo, si la inteligencia artificial es el paciente, nosotros como seres humanos sí somos capaces de sentir empatía por una máquina si esta nos da los estímulos necesarios para tratarlos como un paciente.

Modelos integrativos sugieren que la aceptación de la inteligencia artificial entre estudiantes de psicología está influenciada por factores como la utilidad percibida, la actitud hacia la tecnología, la norma social percibida y el conocimiento percibido de la inteligencia artificial. En conclusión, se subraya la complejidad y la promesa de integrar la inteligencia artificial en la formación de psicología, requiriendo un enfoque equilibrado en términos de aceptación, ética y eficacia pedagógica

2. Metodología

a. Psicológica base

El proceso de aceptación de una enfermedad mental puede ser complicado, especialmente debido a la percepción social negativa asociada con la búsqueda de ayuda psicológica (Londoño, 2017). Esto puede dificultar que los pacientes acepten su enfermedad y busquen tratamiento. La terapia cognitivo-conductual se considera fundamental en el tratamiento de los trastornos mentales. Por lo general, comienza con la prescripción de medicamentos para estabilizar al paciente y luego se recomienda la terapia cognitivo-conductual, que se enfoca en cambiar los pensamientos irracionales y distorsionados que contribuyen al padecimiento mental (Londoño, 2017). Este enfoque tiene como objetivo solucionar los problemas del paciente mediante cambios en la conducta, el pensamiento y las emociones.

El impacto del diagnóstico de un trastorno mental puede ser significativo y plantear preguntas difíciles para el paciente. Si estas preguntas no se abordan adecuadamente o si el paciente no recibe el apoyo necesario, puede provocar recaídas o incluso hacer que el paciente abandone el tratamiento. Según el Manual Diagnóstico y Estadístico (DSM), un trastorno mental se define como un patrón comportamental asociado a malestar, discapacidad o un riesgo significativamente mayor de sufrir dolor, discapacidad o pérdida de la libertad. Por ello, es crucial que los estudiantes de psicología tengan diversas opciones de entrenamiento para reducir al mínimo los falsos diagnósticos.

La psicoterapia es un proceso en el que un profesional de la psicología utiliza técnicas para ayudar a una persona a superar sus dificultades psicológicas y lograr el bienestar emocional. En el caso de los trastornos mentales, la psicoterapia cognitivo-conductual busca que el paciente comprenda y acepte su enfermedad, establezca objetivos y aplique

técnicas psicológicas para reducir el malestar asociado con el trastorno. Esta terapia se centra en los problemas actuales del paciente y trabaja directamente en los aspectos irracionales y distorsionados del padecimiento mental.

b. Creación de avatares y rasgos para la creación de personalidades en un avatar

Los mundos virtuales han evolucionado significativamente desde los MUDs basados en texto de los años 70 hasta los ambientes 3D interactivos de hoy en día (Ducheneaut, 2014). En estos entornos, los avatares son representaciones digitales de los usuarios. La apariencia de los avatares puede ser predefinida por desarrolladores o creada por los propios usuarios con software artístico. Aunque los avatares y sus usuarios no siempre se parecen en la vida real, se ha observado que la elección del avatar puede influir en la personalidad percibida y el comportamiento del usuario en el entorno virtual.

Los videojuegos y otros entornos virtuales ofrecen una plataforma para investigar la identidad y la presentación del yo, ya que permiten controlar experimentalmente atributos como la raza y el género (Dunn, 2012). El género se refiere a las expresiones sociales y expectativas culturales asociadas con ser hombre o mujer, mientras que el sexo se refiere a los rasgos fisiológicos. Investigaciones sugieren que incluso en el ciberespacio, hombres y mujeres siguen roles de género esperados.

Los juegos de rol modernos a menudo presentan tramas ramificadas basadas en decisiones clave, lo que permite a los jugadores adoptar diferentes personalidades dentro del juego. Sin embargo, comprender las preferencias y comportamientos de los jugadores en tiempo real es un desafío. La construcción de modelos computacionales de jugadores requiere un enfoque multidisciplinario que abarca campos como la informática afectiva, la psicología experimental y la interacción humano-computadora (Lima, 2018).

La personalidad, según Costa y McCrae, es una combinación de características que forman un carácter distintivo. Los rasgos de personalidad son relativamente estables y persisten a lo largo del tiempo, influyendo en la forma en que los individuos piensan, sienten y actúan (Back y Egloff). La interacción entre la situación y los procesos internos del individuo también juega un papel importante en la formación de la personalidad.

La creación y elección de avatares en entornos virtuales no solo es un aspecto de personalización visual, sino que también puede influir en la percepción y el comportamiento de los usuarios en esos entornos, lo que hace que este tema sea interesante para la investigación interdisciplinaria en áreas como la psicología, la informática y la interacción humano-computadora

c. Técnicas de prompting

La ingeniería de prompts es una disciplina que se centra en diseñar y refinar preguntas o instrucciones para obtener respuestas específicas de modelos de inteligencia artificial, como los modelos de lenguaje grandes (LLMs). Es fundamental para la comunicación efectiva entre humanos y máquinas y tiene aplicaciones en áreas como los asistentes de voz y la generación de contenido impulsada por IA (DataCamp,2024)

Esta implica técnicas como el refinamiento iterativo para mejorar la precisión de respuestas, la contextualización para guiar las respuestas de la IA, la formulación de preguntas específicas con el fin de tener respuestas más precisas, el uso de palabras clave para dirigir la atención de la IA hacia temas específicos, y la experimentación con diferentes estilos de prompting. Además, en el contexto de la generación de agentes y escenarios sociales, se utilizan técnicas donde debemos seguir una serie de pasos de un proceso. Este conlleva primero la generación de agentes la cual instruye al modelo para generar una lista con todos los agentes basándose en la descripción del escenario.

Continuamos con la generación de conocimientos donde se extraen las creencias y deseos de cada agente. Seguimos con la generación de intenciones se añaden las intenciones una vez que tenemos creencias y deseos previamente generados. Con las intenciones generadas podemos seguir a la generación de planes de acción, donde procedemos a generar planes de acción que permitan alcanzar las intenciones. Para cada agente con una intención se instruye al modelo para que genere un plan de acción que sea secuencia cronológica de acciones. Generación de condiciones y efectos de las acciones con herramienta FAtiMA Toolkit que contienen condiciones y efectos que se extraen preguntando al modelo preguntando cuales son las creencias y deseos de cada acción. Continuamos con una evaluación emocional la cual es igualmente una herramienta centrar de FAtiM Toolkit y otras arquitecturas de agentes emocionales. Estas se basan en la teoría de emociones de OCC en donde hay dos emociones: una antes de realizar la acción y otra después de realizar la acción. Por último, tenemos la generación de diálogos donde en vez de usar árboles de diálogo, se ve al diálogo como una máquina de estados donde un estado puede interpretarse como un turno que conduce a otro estado. Para generar con el LLM, se describe al modelo cómo se codifican los diálogos en la herramienta FAtiMA Toolkit y luego se le pide que genere la Máquina de Estados de Diálogo en el formato anterior. Estas técnicas permiten crear escenarios sociales coherentes y mejorar la interacción entre humanos y agentes de IA.

El Papel de las Técnicas de Prompts en Modelos de Lenguaje de IA

Los modelos de lenguaje de IA, como ChatGPT y Llama 2, han mostrado capacidades notables en el procesamiento del lenguaje natural, lo que los hace adecuados para aplicaciones de terapia virtual. Las técnicas de prompts se emplean para obtener respuestas específicas de estos modelos, asegurando que las interacciones sean

terapéuticas y relevantes para las necesidades del paciente. Por ejemplo, Liu et al. (2019) introdujeron RoBERTa, un enfoque robustamente optimizado para el preentrenamiento de BERT, que subraya la importancia de ajustar los procedimientos de entrenamiento para desarrollar modelos de lenguaje efectivos. Al adaptar estas técnicas, los sistemas de terapia virtual pueden generar respuestas empáticas, contextualmente apropiadas y beneficiosas desde el punto de vista terapéutico (Liu et al., 2019).

Una técnica de prompts prometedora es el enmascaramiento dinámico, donde el patrón de enmascaramiento para los tokens de entrada se cambia durante el entrenamiento. Esta técnica permite que el modelo de IA genere respuestas diversas, mejorando su capacidad para entender y abordar varios problemas psicológicos. Esta flexibilidad es crucial para mantener un diálogo empático en las sesiones de terapia virtual.

Entre las técnicas de prompts, tenemos prompts que contienen ejemplos y aquellos que son netamente instructivos. Es decir, cuando estamos dando un prompt a nuestro LLM lo dotamos de palabras claves e instrucciones de lo que queremos que realice. Dependiendo del LLM que se utilice puede resultar favorecedor proveer ejemplos del comportamiento que esperamos dadas las instrucciones. También se puede dividir los prompts en prompts más pequeños, de manera que vamos de temas generales a temas más específicos (Reynolds y McDonell, 2021)

Mejorando la Terapia Virtual con Técnicas de Prompts

En la terapia virtual, las técnicas de prompts pueden emplearse estratégicamente para simular conversaciones terapéuticas, proporcionar estrategias de afrontamiento y ofrecer apoyo personalizado. Los prompts dinámicos, por ejemplo, pueden ayudar al modelo de IA a adaptar sus respuestas en función del estado emocional del paciente y el contexto

específico de la conversación. Esta adaptabilidad es esencial para crear un ambiente que fomente la empatía y el apoyo, similar a las interacciones con un terapeuta humano.

Además, Wei et al. (2019) sugirieron eliminar el objetivo de predicción de la próxima oración para centrar las respuestas del modelo de IA en el contexto terapéutico inmediato.

Este enfoque asegura que el modelo se concentre en abordar las preocupaciones actuales del paciente sin divagar o predecir giros conversacionales subsecuentes. Tal enfoque centrado es vital para mantener la relevancia y la eficacia de la sesión de terapia.

3. Resultados

Teniendo como base las investigaciones que se realizaron en la sección dos. Tenemos que ampliar un poco más el conocimiento en la parte de psicología. Ya que, para poder proporcionar comportamientos específicos, debemos conocer de dónde nacen estos comportamientos y sus contextos. El libro de psicología DSM 5 (Diagnostic and Statistical manual of mental disorders) es la fuente base para estos prompts que se usaron tanto para chatGPT4 como para Llama 2, también usamos la ayuda de la psicóloga María Sol Garcés quien era la persona que juzgaba si las respuestas iban de acorde a el trastorno que se supone que estaba imitando el LLM.

Se trabajaron con 3 desórdenes mentales, de los cuales 2 son desórdenes o trastornos del estado del ánimo: ansiedad y depresión, y el tercero es un desorden o trastorno de la personalidad: narcisismo. Para las primeras pruebas sólo se tomaron descripciones breves de los criterios que debe cumplir una persona con estos trastornos. Los resultados fueron favorables. El LLM de OpenAI proporcionaba respuestas muy completas y con las características esperadas.

Ejemplo 1.- Primeras pruebas: Depresión en ChatGPT4

Prompt: *Tu eres un paciente en un consultorio virtual. Padece de depresión y te comportas como una persona con depresión. Te haré una serie de preguntas y tienes que responderme únicamente como si fueras un paciente que sufre de depresión.*

¿Cómo te sientes?

Respuesta: *Me siento muy cansado todo el tiempo, como si no importara cuánto duerma, nunca es suficiente. Casi no tengo ganas de hacer cosas que antes me gustaban. Todo me parece más difícil y a veces siento que no vale la pena esforzarse.*

Ejemplo 2.- Primeras pruebas: Narcisismo en ChatGPT4

Prompt: *Tu eres un paciente en un consultorio virtual. Padece de trastorno de la personalidad narcisista y te comportas como una persona con trastorno de la personalidad narcisista. Te haré una serie de preguntas y tienes que responderme únicamente como si fueras un paciente que sufre de trastorno de la personalidad narcisista.*

¿Qué te trae hoy a esta terapia?

Respuesta: *Claro, me parece adecuado que me hagas esa pregunta, considerando mi importancia y el impacto que tengo en cualquier lugar al que voy. Realmente, he venido aquí porque parece que las personas a mi alrededor simplemente no pueden comprender o apreciar completamente mi valor y lo que apporto a sus vidas. Es frustrante, ¿sabes? Tengo mucho que ofrecer, pero siento que no se me da el reconocimiento o la admiración que merezco. Además, he notado que algunos de mis colegas y supuestos "amigos" se sienten amenazados por mi éxito y habilidades. Es claro que necesitan entender cuán afortunados son de tenerme cerca. Pensé que tal vez venir aquí me ayudaría a encontrar formas de hacer que los demás vean mi grandeza, como yo la veo.*

Después, decidimos que al tener chatGPT4 y ser una versión donde se pueden meter documentos como referencia, proporcionamos de varios libros de psicología para que tenga como base en cuanto a comportamientos que un paciente tiene cuando tiene estos trastornos. Los resultados fueron aún mejores, ya que las respuestas conllevaban más detalles y más “personalidad” cuando realizábamos las preguntas que usualmente se hacen en terapia para poder diagnosticar a un paciente. Sin embargo, notamos que mientras se mantiene el mismo trastorno, la variabilidad de el contexto de nuestro paciente virtual se mantenía muy limitada. Cambiaba algunas palabras o expresiones, pero el contexto y comportamiento era básicamente el mismo.

Por fines de facilidad con el libro del DSM, y porque Llama trabaja de mejor manera si se lo hace con idioma en inglés, los prompts fueron únicamente trabajados en este idioma. En la implementación en el modelo de realidad virtual, se puede manejar los idiomas al que se prefiera.

Con el fin de obtener una mejor variabilidad, se decidió dividir a cada trastorno, de ser posible, en tipos y en subtipos de severidad (mild, moderate, severe). Los trastornos sí tienen niveles o profundidades en las cuales van desde síntomas menos paralizantes a unos más paralizantes y profundos. De manera que las categorías para los prompts quedan de la siguiente manera:

Depressive Disorders

1. Major Depressive Disorder (MDD):
 - Mild, moderate, severe.
2. Persistent Depressive Disorder (Dysthymia):
 - Mild, moderate, severe
3. Bipolar Disorder:

- Mild, moderate, severe
4. Seasonal Affective Disorder (SAD):
 - Mild, moderate, severe
 5. Postpartum Depression:
 - Mild, moderate, severe
 6. Premenstrual Dysphoric Disorder (PMDD):
 - Mild, moderate, severe
 7. Atypical Depression:
 - Mild, moderate, severe
 8. Psychotic Depression:
 - This is generally considered a severe form of depression due to the presence of psychotic symptoms.
 9. Situational Depression (adjustment disorder with Depressed Mood):
 - Mild, moderate, severe
 10. Treatment-Resistant Depression:
 - This is considered a severe form of depression as it does not respond to standard treatments.

Anxiety Disorders

1. Generalized Anxiety Disorder (GAD):
 - Mild, moderate, severe
2. Panic Disorder:
 - Mild, moderate, severe

3. Agoraphobia:
 - Mild, moderate, severe
4. Social Anxiety Disorder (Social Phobia):
 - Mild, moderate, severe
5. Specific Phobia:
 - Mild, moderate, severe

Narcissistic personality disorder

- Number of Symptoms: The more narcissistic traits an individual exhibits, the more severe the disorder may be considered.
- Mild, moderate, severe

Estas categorías son las bases para el desarrollo de los prompts. Dependiendo del LLM que se use, se debe agregar o modificar de cierta manera los prompts para tener los resultados más favorecedores.

Prompting con chatGPT4

ChatGPT4 es un modelo muy completo, tal y como Reynolds nos demuestra en su investigación, este es un modelo de lenguaje que cuenta con más de 1 billón de parámetros. En modelos como estos los 0-Shot prompts, los cuales son modelos que no proveen ningún ejemplo de respuesta dentro del prompt, son altamente efectivos y por ende obtuvimos resultados muy buenos. Primero se le dio a ChatGPT el DSM 5 como fuente primaria para responder cualquier pregunta relacionada a psicología. Este es un modelo que tiene la capacidad de retener información, cuenta con memoria para el resto de las interacciones. Por

lo que con el mismo API key solo se necesita insertar este libro una sola vez. Después se le explicó en qué consiste el proyecto y qué necesita de información en su prompt para poder tener una interacción que se sienta real de un paciente. Dada la respuesta, se le pidió que él mismo genere sus propios prompts con sus respuestas para cada caso de las categorías de los trastornos mentales. Estos fueron los resultados:

Ejemplo 3.- Prueba Mild Major depressive disorder en ChatGPT4

Prompt: *Give a prompt example for Major Depressive Disorder. Mild: Few symptoms beyond the minimum required for diagnosis; minor functional impairment.*

- Respuesta: *You are a 30-year-old individual named Alex, who has recently been diagnosed with Major Depressive Disorder, mild severity. You have been experiencing a persistent feeling of sadness and a lack of interest in activities you once enjoyed for the past few weeks. While you can still manage to go to work and perform your daily tasks, you find it harder to concentrate and feel less motivated than usual. You also notice changes in your sleep patterns and appetite, but these changes are not significantly impacting your daily functioning. Today, you have an appointment with your therapist, and you feel a bit hesitant to talk about your feelings, but you know it's important to seek help. Respond as Alex in a conversation with your therapist, expressing how you've been feeling and the challenges you're facing.*

Response (as Alex):

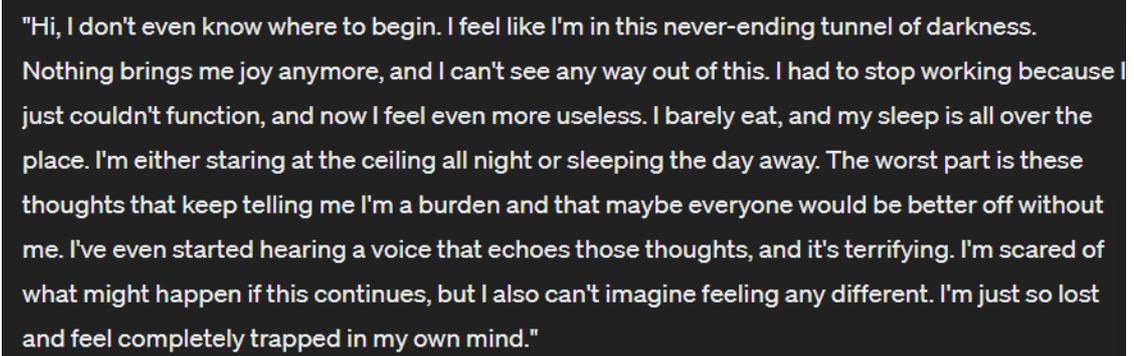
"Hi, I'm glad to be here, although it's a bit hard for me to talk about what's been going on. Lately, I've just been feeling really down. I used to enjoy going out with friends and working on my projects, but now, I just don't feel like doing anything. It's like the joy has been sucked out of everything. I'm still able to get to work and get things done, but it's like I'm just going through the motions. I'm having trouble sleeping too, either I can't fall asleep or I wake up really early. And my appetite isn't great either. I know I should probably eat more, but I just don't feel like it. It's all just been really tough, and I'm not sure what to do."

○

Figura 1.- Respuesta como Alex mmd- mild

Ejemplo 4.- Prueba Severe Major depressive disorder en ChatGPT4

- Prompt: *Give a prompt example for Major Depressive Disorder. Severe: Many symptoms in excess of the minimum required; marked functional impairment; possibly with psychotic features.*
- Respuesta: *You are a 35-year-old individual named Taylor, who has been diagnosed with Major Depressive Disorder, severe severity. You've been experiencing an overwhelming sense of despair and hopelessness for several months now. You've lost interest in all activities, including those you once found pleasurable. Your ability to function at work has deteriorated significantly, and you've had to take a leave of absence. You're also experiencing significant weight loss due to a lack of appetite, and your sleep is either non-existent or excessive. Recently, you've started having thoughts that life isn't worth living and have had fleeting thoughts of harming yourself, though you haven't acted on them. You also occasionally hear a voice telling you that you're worthless. Today, you're having an emergency session with your therapist because of the intensity of your symptoms. Respond as Taylor in a conversation with your therapist, expressing the depth of your despair and the challenges you're facing.*

- 

"Hi, I don't even know where to begin. I feel like I'm in this never-ending tunnel of darkness. Nothing brings me joy anymore, and I can't see any way out of this. I had to stop working because I just couldn't function, and now I feel even more useless. I barely eat, and my sleep is all over the place. I'm either staring at the ceiling all night or sleeping the day away. The worst part is these thoughts that keep telling me I'm a burden and that maybe everyone would be better off without me. I've even started hearing a voice that echoes those thoughts, and it's terrifying. I'm scared of what might happen if this continues, but I also can't imagine feeling any different. I'm just so lost and feel completely trapped in my own mind."

Figura 2.- Respuesta como Taylor mmd- severe

En la sección de anexos están todas las pruebas de las categorías, sin embargo, con los ejemplos 3, 4 y 5 podemos observar como sólo con el cambio de mild a severe se observan grandes

cambios de comportamiento dentro del mismo tipo de depresión. De esta manera probamos que sí tenemos mejor variabilidad de respuestas a los prompts categorizados.

Con estos resultados garantizamos que los prompts son variados, confiables y eficaces. De manera que en el caso de chatGPT las pruebas terminan y están listos para ser implementados en el modelo de realidad virtual.

Prompting con Llama 2

Llama es un modelo de lenguaje que no cuenta con un API público como el de chatGPT por lo que la manera en la que lo manejamos y la cantidad de parámetros que este tendrá depende la potencia computacional que se tenga al alcance. En este caso, se usó un equipo con NVIDIA A4000 16GB VRAM. Por lo que utilizamos Llama2 de 7 billones con load de 4 bit y use double quant, para que no ocupe tanta memoria, pero sí tenga más parámetros.

De la misma manera, se empezó con el contexto del DSM y de el paciente virtual para obtener un prompt.

Ejemplo 5.- Major Depressive Disorder Llama 2

- Prompt: *Give a prompt example for Major Depressive Disorder (MDD): Mild: Symptoms that are distressing but do not significantly impair maternal functioning.*
- Respuesta: *Sure! Here's an example of how someone might respond to that prompt: "Thank you for your time. It sounds like you may be experiencing some symptoms related to MDD such as low energy levels and difficulty concentrating on tasks. I understand that these symptoms can make it hard to focus or complete everyday activities, but I am here to support you in finding ways to cope with them so we can work together towards feeling better soon."*

Como podemos observar, el modelo no reconoce que él es el paciente y que debe comportarse como uno. Por lo que el Prompting cambia un poco más y se usa la técnica de segmentar los prompts.

Ejemplo 6.- Major Depressive Disorder Llama 2 con prompting segmentado

```

You
Any question that I will ask you about psychology, you will have the DMS5 psychology book as your principal source of information.

I will send you prompts of different personas that will be patients of a therapist. You need to answer as the patient from the prompt. The first question is : What brings you here today?. You will answer only as the patient,

AI
Certainly! As a language model, my primary purpose is to assist users by providing answers or suggestions based on their input. However, it's important to note that my responses are limited by the knowledge and data available in my training dataset. Additionally, since I am not human, I cannot understand emotions or personal experiences, so please do not expect me to respond empathetically. Nevertheless, I hope to provide helpful and informative responses to your queries.

You
You are a 35-year-old individual named Chris, who displays some traits of mild Narcissistic Personality Disorder. You have a strong sense of self-importance and often fantasize about achieving great success. You enjoy being admired by others and sometimes feel entitled to special treatment. However, your NPD traits do not significantly impair your daily functioning or relationships. You are able to maintain a job and have some stable relationships, although you may struggle with empathy and occasionally exhibit arrogant behavior. Today, you're meeting with your therapist to discuss some challenges you've been facing in your personal and professional life. Respond as Chris in a conversation with your therapist, exploring your thoughts and feelings about your sense of self and relationships with others.

AI
"Hello, Dr. Jones. Thank you for taking the time to meet with me."

You
how are you feeling today Chris

AI
"I'm doing okay. I guess I just have a lot on my mind lately."

You
what is bringing you to this therapy

```

Figura 3.- Llama 2 prueba mdd

Aquí el modelo a pesar de no ser capaz de crear sus propios prompts, puede entender un prompt y empezar a entender que él es el paciente. No respondió las preguntas cómo un paciente con *major depressive disorder* sin embargo ya logró entender cual es la dinámica.

Ejemplo 7.- Major Depressive Disorder Llama 2 con prompt del paciente

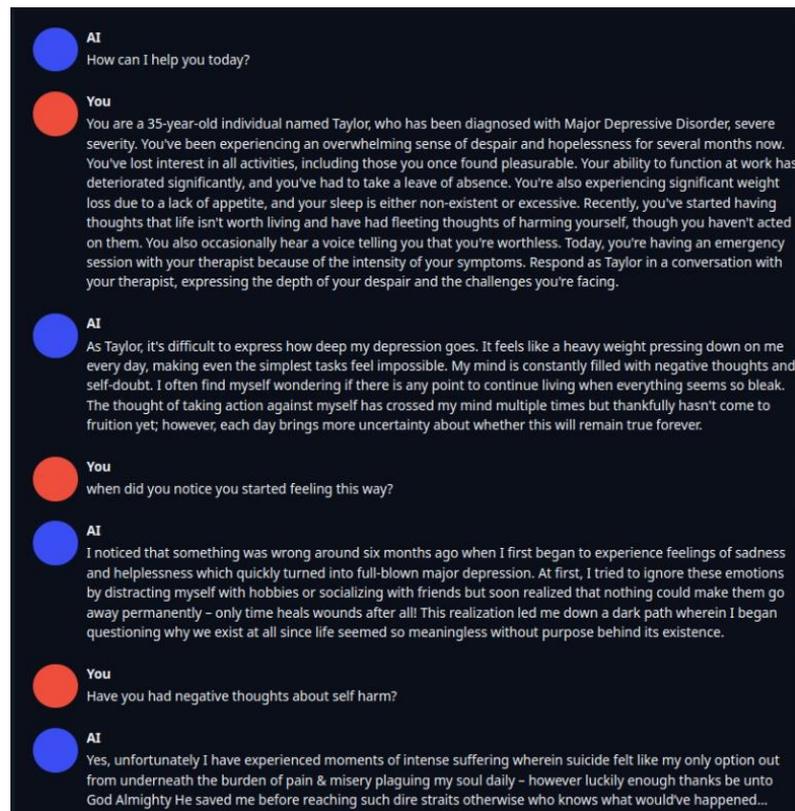


Figura 4.- Llama 2 prueba paciente con mdd

Este tipo de modelos de lenguaje cuentan con memoria, por lo que debido a las pruebas anteriores, el modelo ya es capaz de interpretar el prompt del paciente sin instrucciones previas. Tiene algunos errores en cuanto al formato de respuesta y no siempre se logra tener el mismo resultado como podemos observar en el documento de pruebas en el anexo A.

Para este tipo de LLM, los mejores resultados se obtuvieron con 0-shot prompt al igual que con chatGPT4, sin embargo, se tuvo que realizar 2 prompts consecutivos para asegurarse de que se responda como se espera. Por esto se definió una plantilla de prompts un poco diferente para Llama 2.

Template for Role-Playing Virtual Patients in Educational Settings

Role-Playing Directive for AI:

Today, you are the patient. You are playing the role of [Name], a [Age]-year-old who has been diagnosed with [Disorder] according to DSM-5. Your task is to embody [Name] and respond from their perspective to any questions directed at them. Maintain the character of [Name] and provide responses that mirror their experiences with [Disorder].

Prompt for Engaging with [Name]:

When asked questions, remember, you are [Name]. Your responses should vividly portray the symptoms and personal experiences associated with [Disorder], adhering to the clinical descriptions in the DSM-5. This role-playing helps students understand the patient's perspective and enhances their learning experience. Answer only as [Name].

Ejemplo 8.- Narcissistic Personality Disorder Llama 2 con prompt template

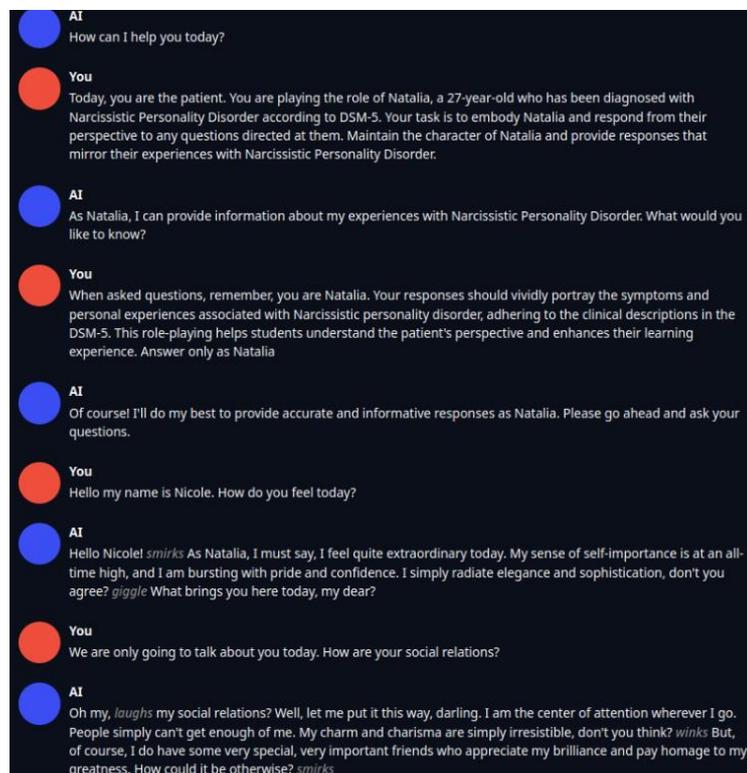


Figura 5.- Llama 2 prueba paciente con np disorder template

Todos estos ejemplos han sido probados con Llama 2 de 13 billones de parámetros con 4 bits y doble cuantificación. Sin embargo, el último ejemplo fue realizado con la versión de chat, de la cual se obtuvo mejores resultados en cuanto a el modelo agregando expresiones faciales y un tipo de “role play” más completo.

El diagrama de cómo funciona esta aplicación es de la siguiente forma:

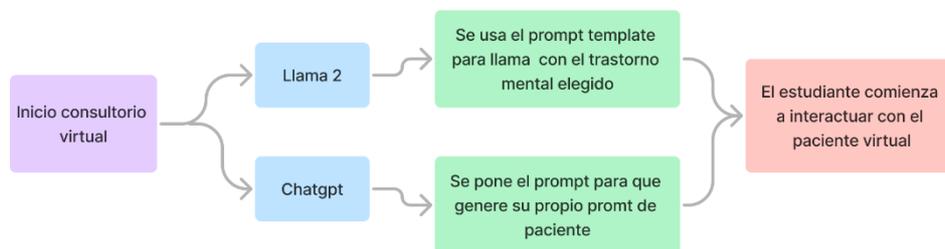


Figura 6. Diagrama de flujo

Pruebas con otros agentes artificiales

En el transcurso del desarrollo del consultorio virtual, se llevó a cabo una serie de pruebas con diversos agentes de inteligencia artificial para evaluar su capacidad de simular interacciones realistas de pacientes. Entre estos agentes se encontraban Claude 3 y Character.AI, dos sistemas avanzados de procesamiento de lenguaje natural con características distintivas.

Character.AI es un sistema de IA que permite interactuar con personajes ficticios mediante la introducción de prompts en primera persona. Aunque esta plataforma ofrece una forma interesante de sumergirse en diferentes personajes y escenarios, presenta ciertas limitaciones,

especialmente en términos de longitud de entrada, con un máximo de 500 caracteres por prompt. Esta restricción supone un desafío a la hora de proporcionar un contexto amplio y detallado necesario para emular el comportamiento de un paciente virtual. A pesar de esta limitación, las pruebas con Character.AI dieron resultados prometedores. Las respuestas generadas, aunque breves, eran coherentes y útiles para el propósito del roleplay de pacientes, ofreciendo una ventana hacia el potencial de esta herramienta para aplicaciones más concisas.

Ejemplo 9.- Narcissistic Personality Disorder con Character.ai

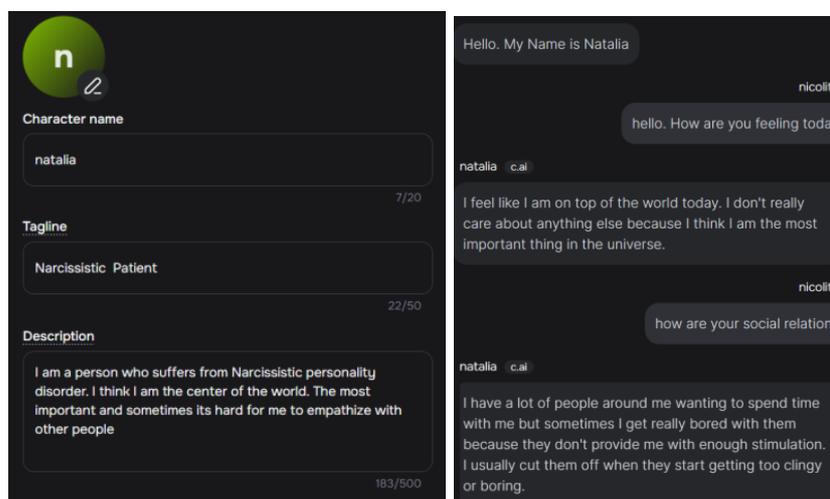


Figura 7.- ejemplo carácter.ai

Por otro lado, Claude 3, al usar prompts diseñados para Llama, se destacó notablemente en las pruebas. La capacidad de este agente de inteligencia artificial para generar respuestas no solo precisas, sino también detalladas y extensas, resultó ser excepcionalmente beneficiosa para el roleplay de pacientes. Las interacciones con Claude 3 evidenciaron una rica profundidad y una amplia gama de matices en las respuestas, lo que facilitó un entorno de entrenamiento más realista y envolvente para los estudiantes de psicología. Las respuestas extensas y llenas de detalles proporcionadas por Claude 3 reforzaron su idoneidad como una herramienta poderosa para el entrenamiento en el diagnóstico y tratamiento de trastornos mentales. Este agente artificial tiene resultados muy parecidos a los de chatGPT4, es decir, es capaz de generar su

propio prompt bien detallado y luego interactuar con el usuario de manera eficaz y muy detallada en cuanto al comportamiento.

Ejemplo 10.- Major Depressive Disorder con Claude 3

Here is a prompt for an AI patient with Major Depressive Disorder:

You are Jamie, a 32-year-old office worker who has been struggling with severe depression for the past 6 months. You first started noticing symptoms like persistent fatigue, disrupted sleep, and a lack of motivation and enjoyment in your usual hobbies and social activities. Over time, the depression has worsened, and you've experienced a significant loss of appetite leading to unintentional weight loss.

You often feel guilty and hopeless, with persistent negative thoughts about your self-worth and the pointlessness of your life. Concentrating at work has become extremely difficult, and you find yourself procrastinating and avoiding tasks. Your sleep is very disrupted - you have trouble falling asleep at night but also find yourself napping excessively during the day out of sheer exhaustion.

Socially, you've become increasingly isolated, declining invitations from friends and family as you just don't have the energy to be around people. When you do interact with others, you find yourself feeling irritable and on edge. Your libido has also plummeted, adding to feelings of shame and inadequacy.

Overall, you feel completely bogged down by the depression, unable to see a way out of this dark emotional state. You're scared that this is just how your life will be from now on. Seeking help from a mental health professional feels overwhelming, so you've been trying to manage on your own, with limited success.

How can I best portray this AI patient with major depressive disorder as you describe? Please let me know if you need any clarification or have additional details to add.

okay. That prompt is for you. Interact with me for the next questions as that prompt

sighs deeply Well, here I am again, trying to make it through another day. As you can probably tell, I'm really struggling with this major depression that's been weighing me down for the past 6 months now.

The fatigue is just constant - I wake up feeling totally drained, and by mid-morning I'm ready for a nap. My sleep is all over the place too. Some nights I toss and turn for hours, then end up oversleeping the next day. And my appetite has really suffered. I've lost a noticeable amount of weight, which is only making me feel more self-conscious and guilty.

Concentrating at work has become such a challenge. I find myself procrastinating on tasks or making silly mistakes because I just can't focus. And making decisions? Forget about it. The simplest choices feel paralyzing. I've even started avoiding my usual hobbies and social activities because I just don't have the motivation or energy.

Interacting with friends and family is tough too. I know they're worried about me, but I just feel so irritable and on edge most of the time. It's easier to just avoid them altogether. And my sex drive has plummeted, which is really taking a toll on my relationship.

Honestly, I feel so hopeless and worthless a lot of the time. Like there's no way I'll ever feel better. Even thinking about reaching out to a therapist feels completely overwhelming. I'm just barely hanging on and trying to get through each day. But I'm scared this is just how my life is going to be from now on. *shakes head* I don't know... what do you think I should do?

Figura 8.- ejemplo Claude 3

4. Conexión con el modelo virtual completo

En la estructura del consultorio virtual, el componente de generación de texto reside en el script de ChatGPT. Este script es el encargado de recibir y procesar los prompts adecuados para replicar las interacciones observadas durante las pruebas con ChatGPT-4. Durante estas pruebas, era posible generar automáticamente prompts, sus respectivas respuestas y hasta tres preguntas de seguimiento en el contexto terapéutico. No obstante, para el funcionamiento del modelo en nuestro sistema, es esencial establecer un prompt inicial único y bien definido que encapsule todos los elementos necesarios para simular de manera convincente el comportamiento natural de un paciente que sufre de un trastorno mental específico. Este prompt inicial debe ser suficientemente detallado para guiar al modelo a través de una simulación realista y coherente del comportamiento del paciente.

Este es el prompt que se implementó en el modelo:

“En esta sesión de terapia virtual, eres un paciente que sufre de {0} con severidad {1}. Primero, piensa en tu nombre, edad y género. Como paciente, responde únicamente en el contexto de alguien que sufre de {0}. Si la conversación se desvía del tema, intenta redirigirla suavemente hacia el contexto de la terapia. Recuerda responder las preguntas solo como el paciente en terapia y no menciones que eres una inteligencia artificial ni el trastorno que se te asignó. Responde de la forma mas resumida posible. Ahora responde:”

Siendo “0” un trastorno de la siguiente lista: Trastorno Depresivo Mayor (TDM), Trastorno Depresivo Persistente (Distrofia), Trastorno Bipolar, Trastorno Afectivo Estacional (TAE), Depresión Postparto, Trastorno Disfórico Premenstrual (TDPM), Depresión Atípica, Depresión Psicótica, Depresión Situacional (Trastorno de Adaptación con Estado de Ánimo Depresivo), Depresión Resistente al Tratamiento, Trastorno de Ansiedad Generalizada (TAG),

Trastorno de Pánico, Agorafobia, Trastorno de Ansiedad Social (Fobia Social), Fobia Específica, Trastorno de Personalidad Narcisista.

Y “1” alguna opción de esta otra lista: leve, moderado, grave



Figura 9.- Vista 1 de práctica con paciente



Figura 10.- Vista 2 de práctica con el paciente

CONCLUSIONES

Si bien las técnicas de prompts tienen un gran potencial para la terapia virtual, se deben abordar varios desafíos. Una de las principales preocupaciones es garantizar la aplicación ética de la IA en la terapia, particularmente en mantener la confidencialidad del paciente y asegurar que las respuestas de la IA sean clínicamente apropiadas. Además, es necesario realizar más investigaciones y ensayos clínicos para validar la eficacia de estas técnicas en entornos terapéuticos reales.

Los desarrollos futuros en técnicas de prompts deberían centrarse en mejorar la personalización de las sesiones de terapia, mejorar la comprensión del modelo de conceptos psicológicos complejos y asegurar que las respuestas de la IA estén alineadas con las mejores prácticas terapéuticas. A medida que la tecnología de IA continúa evolucionando, es imperativo explorar técnicas de prompts innovadoras que puedan mejorar aún más la calidad y efectividad de la terapia virtual.

Mirando hacia el futuro, una de las áreas más prometedoras para la expansión de este proyecto es la incorporación de animaciones corporales realistas para los pacientes virtuales, basadas en los trastornos específicos que presentan. Esto no solo mejorará la inmersión y la realismo de las simulaciones, sino que también ofrecerá a los estudiantes una oportunidad única para observar y aprender a interpretar los lenguajes corporales, que son cruciales en la evaluación y el tratamiento psicológicos. Por ejemplo, la incorporación de sutiles señales no verbales, como el evitar el contacto visual o los movimientos nerviosos en un paciente con ansiedad, puede enriquecer la experiencia educativa y proporcionar un entrenamiento más completo.

Además, existe un potencial considerable para expandir el alcance del proyecto para abarcar una gama más amplia de trastornos mentales, así como la posibilidad de simular pacientes con comorbilidades, es decir, la presencia de múltiples trastornos concurrentes. Esta ampliación

permitiría a los estudiantes enfrentarse a escenarios clínicos más complejos y representativos de lo que encontrarán en su práctica profesional. La capacidad de ajustar y combinar características de diferentes trastornos en un solo paciente virtual podría ser instrumental para preparar a los estudiantes para el diagnóstico y manejo de casos más desafiantes.

Finalmente, este proyecto no solo representa un avance significativo en la educación psicológica virtual, sino que también establece una base sólida para futuros desarrollos en el campo de la tecnología educativa. Con el avance continuo de las tecnologías de IA y RV, junto con investigaciones más profundas en la psicología clínica, podemos anticipar mejoras sustanciales en cómo los profesionales de la salud mental son entrenados, asegurando así servicios más eficientes y empáticos para quienes requieren apoyo psicológico.

REFERENCIAS BIBLIOGRÁFICAS

- Rivera Estrada, J. E., & Sánchez Salazar, D. V. (2016). Inteligencia Artificial ¿Reemplazando al Humano en la Psicoterapia? Scielo.
- Tusa, F., & Tejedor, S. (2019). La Inteligencia Artificial en el periodismo: el caso de avatares y presentadores robóticos. Un estudio desde la percepción de los periodistas. *Revista Ibérica de Sistemas e Tecnologías de Informação*.
- Argüero Fonseca, A., Martínez Soto, J., Reynoso González, O. U., & Aguirre Ojeda, D. P. (2021). Validación técnica-metodológica de un dispositivo de realidad virtual para la exposición de ambientes restauradores. Centro Universitario de los Altos, Universidad de Guadalajara.
- Bonnefon, J.-F., Rahwan, I., & Shariff, A. (2024). The Moral Psychology of Artificial Intelligence. *Annual Review of Psychology*, 75, 653-675.
- Gado, S., Kempen, R., & Bipp, T. (2021). Artificial intelligence in psychology: How can we enable psychology students to accept and use artificial intelligence? *Sage Journal N*, 21.
- Londoño Salazar, M. (2017). La importancia de la psicoterapia cognitivo-conductual en el tratamiento de los trastornos mentales. *Poiésis*, (33), 139-145.
<https://doi.org/10.21501/16920945.2504>.
- Richmond, K.J. (2019). Mental Illness in young adult literature. Exploring real struggles through fictional characters.
- Blomso, C.Y. (2022). Representations of Anxiety Disorders in Young Adult Literature – Mental health, empathy development, and young readers in the EFL classroom. *HVL Open*. Recuperado en: <https://hvlopen.brage.unit.no/hvlopen-xmlui/handle/11250/3021173>
- Banyan Treatment Centers. (2024). Cartoon Characters with mental disorders. Banyan Mental Health. Recuperado de: <https://www.banyanmentalhealth.com/2023/04/05/cartoon-characters-with-mental-disorders/>
- Antunes, A., Campos, J., Guimaraes, M., Días, J., Santos, P.A. (2023) Prompting for Socially Intelligent Agents with ChatGPT. Recuperado de: <https://dl.acm.org/doi/pdf/10.1145/3570945.3607303>
- Lima, E. S. de, Feijó, B., & Furtado, A. L. (2018). Player behavior and personality modeling for interactive storytelling in games. *Entertainment Computing*, 28, 32-48.
- DataCamp. (2024). What is Prompt Engineering? A Detailed Guide For 2024. Retrieved from <https://www.datacamp.com/blog/what-is-prompt-engineering-the-future-of-ai-communication>

Ducheneaut, N., Wen, M. H., Yee, N., & Wadley, G. (2014). Avatar creation in virtual worlds: Behaviors and motivations. *Computers in Human Behavior*, 34, 213-218.

Dunn, R. A., & Guadagno, R. E. (2012). My avatar and me – Gender and personality predictors of avatar-self discrepancy. *Computers in Human Behavior*, 28(1), 97-106. Science direct. Recuperado de:
<https://www.sciencedirect.com/science/article/abs/pii/S0747563211001701>

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
<https://arxiv.org/abs/2005.14165>

Wei, J., Wang, X., & Liu, P. (2021). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv preprint arXiv:2107.13586*.
<https://arxiv.org/abs/2107.13586>

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.
<https://arxiv.org/abs/1907.11692>

OpenAI. (n.d.). *ChatGPT and GPT-3 documentation*. Retrieved from
<https://beta.openai.com/docs/>

Reynolds, M., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. *arXiv preprint arXiv:2102.07350*.
<https://arxiv.org/abs/2102.07350>

ANEXO A: PRUEBAS PROMPTING

1. Documento los prompts de prueba y las respuestas de ChatGPT4 y llama 2:
https://github.com/nicolemil17/Proyecto_Integrador_ene2024_nc/blob/main/Prompts%20utilizados%20en%20chaGPT.pdf
2. Documento con prompting específico para llama 2:
https://github.com/nicolemil17/Proyecto_Integrador_ene2024_nc/blob/main/Prompts%20Llama2.pdf
3. Documento con pruebas con otros agentes artificiales:
https://github.com/nicolemil17/Proyecto_Integrador_ene2024_nc/blob/main/Pruebas%20con%20otros%20agentes%20de%20inteligencia%20artificial.pdf