

**COLEGIO DE CIENCIAS E INGENIERÍAS**

INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

Planificación para el Desarrollo del Proyecto Integrador

Tutor: Felipe Grijalva

Autor: Nicole Caicedo Guzmán

Quito – Ecuador

2024

1. Título del Proyecto:

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

1. Relevancia y Justificación:

La psicología, como disciplina en constante evolución, se enfrenta al desafío de comprender la complejidad de la mente humana y diagnosticar trastornos mentales de manera precisa. La importancia del practicar, fundamental en cualquier carrera, toma una dimensión crítica en el ámbito de la psicología. No obstante, la práctica directa con personas conlleva riesgos, especialmente en el ámbito de diagnósticos erróneos.

La propuesta de este proyecto, estableciendo un consultorio virtual con la asistencia de inteligencia artificial, surge como una solución innovadora para facilitar el entrenamiento de estudiantes de psicología. La idea es proporcionar un entorno de simulación seguro donde los estudiantes puedan practicar sin el riesgo asociado a errores en diagnósticos, a la vez que adquieren experiencia valiosa.

El creciente desarrollo de las inteligencias artificiales agrega una capa fascinante a este proyecto. Si bien las ciencias exactas han experimentado una integración más fluida con la inteligencia artificial, el comportamiento humano sigue presentando un desafío. En el ámbito de la psicología, las inteligencias artificiales aún están lejos de replicar de manera natural la complejidad de las interacciones humanas. Esto resalta la necesidad de una investigación continua para adaptar estos recursos avanzados de manera efectiva a campos tan sensibles como la psicología.

1. Objetivos
   1. Generales

Desarrollar un módulo de lenguaje de un paciente virtual usando inteligencia artificial.

* 1. Específicos
* Investigación de simulaciones con inteligencia artificial para el área de psicología.
* Investigar y desarrollar técnicas de prompting adecuadas para emular comportamientos propios de trastornos mentales en específico.
* Aplicar las técnicas de prompting anteriormente investigadas en el modelo ya existente usando chatGPT y Llama 2.

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 ciber terapia 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

1. Metodología de Trabajo

En la etapa inicial, se llevará a cabo una exhaustiva revisión del estado del arte para obtener una comprensión detallada de los avances en el uso de inteligencia artificial en el ámbito psicológico. Paralelamente, se realizará una revisión de la literatura, centrándose en personajes literarios, con el propósito de modelar el comportamiento deseado para la inteligencia artificial (IA). Se llevará a cabo una investigación exhaustiva de papers relacionados con la psicología para fundamentar y enriquecer la propuesta.

Posteriormente, se emprenderá una investigación adicional enfocada en el desarrollo de avatares en videojuegos o entornos de realidad virtual. Esta fase implicará la exploración de cómo el modelo de IA puede reflejar con precisión los síntomas psicológicos, evaluando la eficiencia de diversas técnicas. Paralelamente, se adquirirán conocimientos específicos sobre la integración con ChatGPT, el uso de técnicas de prompting, y se buscará la orientación de estudiantes de psicología para asegurar una conexión efectiva entre la IA y la experiencia psicológica.

La fase final consistirá en la implementación práctica, donde se probará la IA con ChatGPT y Llama 2, integrándola de manera coherente con la simulación virtual y el avatar con voz. Este enfoque integrado permitirá evaluar la capacidad de la IA para simular y responder a escenarios psicológicos, representando un paso crucial hacia la validación y mejora continua del modelo

1. Sumario de Contenidos

* Introducción
* Estado del arte
* Descripción del proyecto general
* Descripción de la propuesta en el módulo del lenguaje
* Metodología
* Resultados
* Conclusiones

1. Recursos
   1. Humanos

Estudiante, tutor, profesores que colaboran con el desarrollo de este proyecto, estudiantes de psicología.

* 1. Materiales

Computadores personales, hardware del laboratorio para llama 2, si se llega a la parte de conexión con el ambiente virtual también necesitaremos los recursos de dicho proyecto.

* 1. Económicos

1. Cronograma de Actividades

A continuación, se presenta un cronograma contando desde la semana del 15 de enero a la semana de marzo 25. El último entregable es el 21 de abril y se pretende que un mes antes aprox ya se esté implementando todo lo investigado y desarrollado en el modelo del proyecto oficial.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Semanas 2024** | | | | | | | | | |
| **Actividades** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **A1:** Desarrollo del documento de planificación. | X | X | X |  |  |  |  |  |  |  |
| **A2:** Estudio del estado del arte |  | X | X |  |  |  |  |  |  |  |
| **A3:** Investigación de psicología |  |  |  | X |  |  |  |  |  |  |
| **A4:** Revisión de literatura y temática avatares |  |  |  |  | X | X |  |  |  |  |
| **A5:** Investigación y desarrollo de técnicas de prompting |  |  |  |  |  |  | X | X |  |  |
| **A6:** Conexión a chatgpt usando prompting y guiada por estudiantes de psicología. |  |  |  |  |  |  |  | X | X |  |
| **A7:** Implementación |  |  |  |  |  |  |  |  |  | X |

1. Entregables

Entregable 1: Investigación parte de psicología, revisión de literatura y creación de avatares, e investigación y desarrollo de técnicas de prompting.

Entregable 2: Conexión y uso de las técnicas anteriormente investigadas.

Entregable 3: Implementación del módulo de lenguaje en el proyecto. Es decir, se acopla a el modelo virtual.

1. Referencias
2. J. E. Rivera Estrada and D. V. Sánchez Salazar. 2016. "Inteligencia Artificial ¿Reemplazando al Humano en la Psicoterapia?," *Scielo.*
3. F. Tusa y S. Tejedor. 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 Tecnologias de Informação*,
4. A. Argüero Fonseca, J. Martínez Soto, O. U. Reynoso González, y D. P. Aguirre Ojeda. 2021. "Validación técnica-metodológica de un dispositivo de realidad virtual para la exposición de ambientes restauradores," Centro Universitario de los Altos, Universidad de Guadalajara
5. J.-F. Bonnefon, I. Rahwan y A. Shariff. 2024. "The Moral Psychology of Artificial Intelligence," *Annual Review of Psychology*, vol. 75, pp. 653-675, enero 18, 2024.
6. S. Gado, R. Kempen y T. Bipp. 2021. "Artificial intelligence in psychology: How can we enable psychology students to accept and use artificial intelligence?," Sage *Journal N*, vol. 21
7. Revisión y firma del tutor del proyecto

Yo, Felipe Grijalva, profesor de la carrera de Ingeniería en Ciencias de la Computación, hago constar que he revisado y, por lo tanto, apruebo el documento de planificación del proyecto titulado “Investigación y aplicación de técnicas de prompting para paciente artificial: Integrando chatGPT y Llama 2 en un simulador de consultorio psicológico” propuesto por el estudiante Nicole Caicedo Guzmán. Por otra parte, me comprometo a proporcionar al estudiante el soporte necesario y oportuno para el buen desarrollo del proyecto antes mencionado.

Fdo: Felipe Grijalva

Quito, 4 de febrero de 2024