



COLEGIO DE CIENCIAS E INGENIERÍAS

INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

Estado del arte

En la actualidad la inteligencia artificial (IA) ha tenido un impacto fuerte en las diferentes áreas de la vida, en especial en el campo laboral; volviéndose una de las herramientas más grandes y útiles para los profesionales, incluyendo, los profesionales de salud mental (Salazar & Velastegui, 2024). En el área de la Psicología Clínica, la IA se ha considerado una de las herramientas más útiles para el diagnóstico y la creación de tratamientos e intervenciones para pacientes de psicoterapia, lo cual ha puesto en marcha muchas preguntas y curiosidades sobre su uso en el campo profesional (Salazar & Velastegui, 2024).

La IA, a través de algoritmos de aprendizaje automático, puede analizar grandes cantidades de datos de pacientes, identificar patrones y, en algunos casos, predecir resultados terapéuticos. Además, ha surgido el uso de agentes conversacionales basados en IA para apoyar tanto a los pacientes como a los profesionales de la salud mental en diversas etapas del tratamiento, desde la evaluación hasta la intervención psicoterapéutica.

En este sentido, la IA automatiza los procesos y es una herramienta muy útil a la hora de tomar una decisión sobre el caso del paciente, identificando tendencias y respuestas de los pacientes con diversos trastornos mentales (Salazar & Velastegui, 2024). Sin embargo, es importante señalar que la IA no reemplaza la interacción humana en la psicoterapia, sino que, complementa y expande las capacidades de



los terapeutas para realizar análisis más profundos y personalizados, mejorando la precisión diagnóstica y el tratamiento basado en datos empíricos (Salazar & Velastegui, 2024).

Por otro lado, los LLMs han avanzado significativamente en el tratamiento del lenguaje natural, lo que ha permitido su implementación en varias áreas, incluida la psicología. Los LLMs, como los modelos GPT (Generative Pre-trained Transformer), son algoritmos entrenados con grandes cantidades de texto para generar respuestas coherentes a preguntas o interactuar con los usuarios en un contexto conversacional. Estos modelos tienen el potencial de asistir en la simulación de pacientes artificiales, proporcionando respuestas que imitan el comportamiento humano en sesiones de psicoterapia.

De acuerdo con un estudio reciente, los LLMs se utilizan para simular interacciones entre terapeuta y paciente, generando respuestas que reflejen un análisis psicoterapéutico (Kovač et al., 2024). Esta capacidad abre nuevas oportunidades en la investigación de la psicoterapia, puesto que, los modelos pueden ser evaluados por terapeutas humanos para medir su efectividad en la simulación de comportamientos clínicos y en la generación de respuestas adaptativas y pertinentes a las necesidades del paciente (Kovač et al., 2024).

Además, los LLMs han demostrado un potencial extraordinario en la práctica clínica, superando en algunos casos a los profesionales humanos en tareas específicas. Un estudio reciente reveló que los LLMs superaron a los profesionales de la salud mental y médicos en la identificación del trastorno obsesivo-compulsivo (TOC), destacando su capacidad para asistir en el diagnóstico oportuno y preciso de trastornos mentales (Kim, J., et al., 2023). Este hallazgo subraya el potencial de los LLMs para mejorar la eficiencia y precisión en el diagnóstico de enfermedades mentales, especialmente en contextos donde los recursos humanos son limitados.

De igual manera, los chatbots basados en IA han comenzado a ser utilizados como una herramienta de apoyo en el tratamiento de la salud mental. Estos sistemas permiten a los pacientes interactuar con un agente virtual que puede proporcionar respuestas terapéuticas inmediatas, ofrecer recursos educativos y



ayudar en la gestión emocional diaria. A través del uso de algoritmos de procesamiento del lenguaje natural, los chatbots pueden adaptarse a las necesidades individuales de los pacientes, ofreciendo una forma de apoyo continuo y accesible fuera del horario de las consultas tradicionales.

Un estudio reciente destacó que los chatbots pueden ser una herramienta muy útil en el contexto de los trastornos depresivos y ansiosos, a la hora de suplantar por un corto periodo de tiempo al profesional, es decir, cuando sean casos emergentes que el profesional no se encuentre disponible. Sin embargo, se enfatiza que estos sistemas deben ser supervisados y evaluados regularmente para garantizar que no se generen respuestas inadecuadas o que puedan causar daño emocional (Pham, Nabizadeh, & Selek, 2022).

Para esto, la evaluación humana, o "Human Evaluation", es un proceso clave, puesto que, ayuda de gran manera a la evaluación del rendimiento de los LLMs en contextos clínicos. Este tipo de evaluación implica que terapeutas y psicólogos humanos revisen las respuestas generadas por los modelos para asegurarse de que sean coherentes, pertinentes y útiles en un contexto terapéutico. Dado que la simulación de pacientes con IA en psicoterapia implica que las respuestas sean lo suficientemente realistas como para que los terapeutas puedan ofrecer intervenciones significativas, en este punto podemos decir que, la evaluación humana es crucial para medir la calidad y efectividad de los LLMs en estos escenarios (Kämmer et al., 2023).

El uso de evaluaciones humanas no solo ayuda a mejorar la precisión de los LLMs, sino que también garantiza las interacciones entre los pacientes y agentes, verificando el cumplimiento de los estándares clínicos y la ética del proceso (Mehandru et al., 2024). Al emplear evaluadores humanos, profesionales en el área, es posible determinar si las respuestas generadas son terapéuticamente válidas o si necesitan ajustes para alinearse mejor con las necesidades emocionales y psicológicas de los pacientes.

Esto es esencial para el proyecto porque permite evaluar si las respuestas de los pacientes artificiales son lo suficientemente realistas y útiles para el aprendizaje de los estudiantes. La retroalimentación constante de expertos permitirá que los LLMs sean legibles y comprensibles, mejorando su capacidad para



actuar en un contexto terapéutico. Al hacerlo, se asegura que los futuros terapeutas reciban una formación que refleje con mayor precisión las interacciones en la vida real (Kämmer et al., 2023).

Ergo, otra forma de evaluar es mediante Test de Turing, este es un test que fue propuesto por Alan Turing en el año de 1950. El Test de Turing busca evaluar la capacidad de una máquina para exhibir un comportamiento humano. En el contexto de la psicoterapia asistida por IA, el Test de Turing podría ser una herramienta útil para medir el grado en que los pacientes artificiales creados con LLMs, simulando la interacción humana en una sesión terapéutica. Si un terapeuta humano no puede diferenciar entre las respuestas de un paciente real y las de un paciente artificial, se podría argumentar que el modelo ha superado con éxito el Test de Turing. Pero, a pesar de su relevancia teórica, el Test de Turing tiene limitaciones prácticas cuando se aplica al campo de la psicoterapia. Esto se debe a que "las interacciones terapéuticas no solo se basan en la coherencia verbal, sino también en la empatía, el lenguaje no verbal y la intuición clínica" (Pham, Nabizadeh, & Selek, 2022), áreas donde las IA actuales aún están rezagadas.

Sin embargo, como punto de referencia, el Test de Turing sigue siendo una métrica útil para evaluar los avances en la simulación de pacientes artificiales. Ya que, si bien el Test de Turing no puede evaluar completamente la efectividad de los pacientes artificiales en un contexto terapéutico, su aplicación permite establecer un punto de comparación para medir la calidad de las interacciones generadas por los LLMs. Esto ayudará a determinar hasta qué punto los pacientes artificiales pueden imitar el comportamiento humano, proporcionando a los estudiantes de psicología un entorno de práctica lo más cercano posible a la realidad.

Con esta información previa se ha observado la posibilidad de diseñar un sistema de pacientes artificiales que pueda ser utilizado por los estudiantes de psicología para practicar antes de atender a pacientes reales. Con la ayuda de IA, es posible simular una amplia variedad de perfiles psicológicos, en especial el trastorno depresivo, ansioso y narcisista, permitiendo a los estudiantes adquirir experiencia en situaciones diversas y prepararse mejor para la práctica clínica. Esto no solo tiene el potencial de mejorar la calidad de la educación, sino también de reducir el riesgo de errores en la atención a pacientes reales.



Es por esto por lo que es relevante conocer los antecedentes y las potenciales ventajas que tienen los LLMs en el campo de la psicoterapia, ya que esto permite que los estudiantes interactúen con pacientes virtuales que reaccionan igual a como lo haría un ser humano. Los LLMs pueden generar respuestas realistas que ayuden a los estudiantes a aprender a manejar situaciones terapéuticas complejas y que dan cuenta de la realidad de esta profesión. Además, la capacidad de ajustar los LLMs para simular distintos tipos de pacientes ofrece un ambiente de aprendizaje seguro y controlado, donde los estudiantes pueden practicar sin el riesgo de perjudicar a un paciente real. Al mismo tiempo, esto puede aportar a la psicoterapia y la salud mental, mejorando el acceso a recursos de práctica para futuros terapeutas.

Referencias bibliográficas

- Bird, J. J., Wright, D., & Sumich, A. (2024). Generative AI in psychological therapy: Perspectives on computational linguistics and large language models in written behaviour monitoring. *Proceedings of the 17th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '24)*, 322-328. <https://doi.org/10.1145/3652037.3663893>
- Kämmer, J. E., Hautz, W. E., Krummrey, G., Sauter, T. C., Penders, D., Birrenbach, T., & Bienenfeld, N. (2023). Effects of interacting with a large language model compared with a human coach on the clinical diagnostic process and outcomes among fourth-year medical students: Study protocol for a prospective, randomised experiment using patient vignettes. *BMJ Open*. <https://doi.org/10.1136/bmjopen-2023-072517>
- Kim, J., Leonte, K.G., Chen, M.L. et al. Large language models outperform mental and medical health care professionals in identifying obsessive-compulsive disorder. *npj Digit. Med.* 7, 193 (2024). <https://doi.org/10.1038/s41746-024-01181-x>
- Kovač, G., Sawayama, M., Ford Dominey, P., & Oudeyer, P. Y. (2024). Stick to your role! Stability of personal values expressed in large language models. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0309114>
- Kim, J., Leonte, K.G., Chen, M.L. et al. Large language models outperform mental and medical health care professionals in identifying obsessive-compulsive disorder. *npj Digit. Med.* 7, 193 (2024). <https://doi.org/10.1038/s41746-024-01181-x>
- Landers, Richard & Auer, Elena & Helms, Adrian & Marin, Sebastian & Armstrong, Michael. (2020). Gamification of Adult Learning: Gamifying Employee Training and Development. <https://doi.org/10.1017/9781108649636.012>.



- Mehandru, N., Miao, B. Y., Rodriguez Almaraz, E., Sushil, M., Butte, A. J., & Alaa, A. (2024). Evaluating large language models as agents in the clinic. *npj Digital Medicine*, 7(84). <https://doi.org/10.1038/s41746-024-01083-y>
- Norzagaray , C. C., Sevillano , M. L., & Valenzuela, B. A. (2021). Estrategias de aprendizaje y rendimiento académico: La perspectiva del estudiante de psicología. *Raíces. Revista Iberoamericana de la Educación Superior*, 3(1), 59-68. <https://doi.org/10.17811/ria.3.1.2021.59-68>
- Pham, K. T., Nabizadeh, A., & Selek, S. (2022). Artificial intelligence and chatbots in psychiatry. *Psychiatric Quarterly*, 93(1), 249-253. <https://doi.org/10.1007/s11126-021-09974-5>
- Salazar Garcés, L. F., & Velastegui Hernandez, D. C. (2024). Inteligencia artificial y su impacto en la psicología humana: Mini revisión. *Revista Universitaria con Proyección Científica, Académica y Social*, 8(1). <https://doi.org/10.31243/mdc.uta.v8i1.2306.2024>