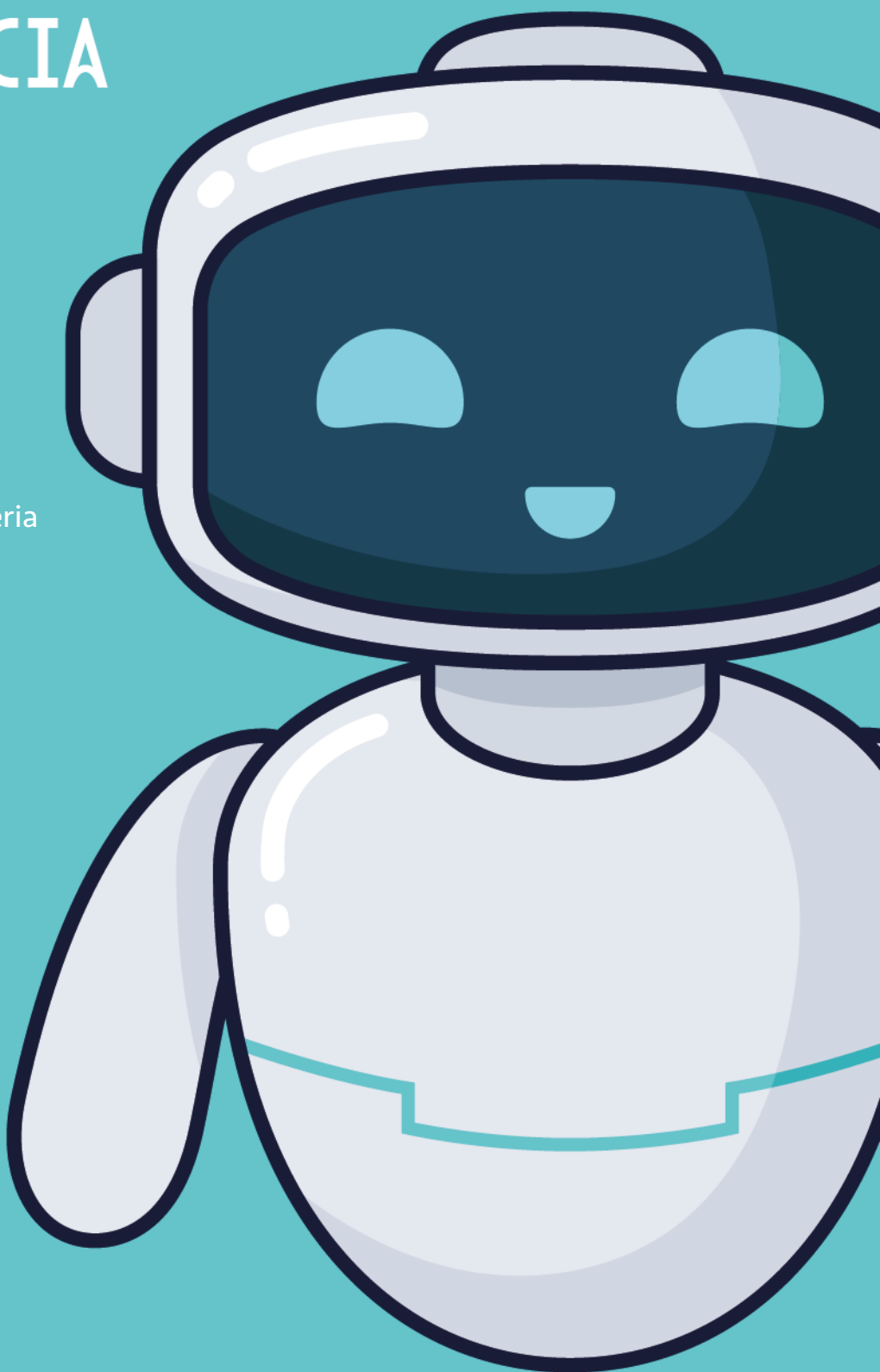


¿PUEDE UN CHATBOT SER NUESTRO AMIGO?:

INTELIGENCIA ARTIFICIAL Y SALUD MENTAL

Karla Alejandra Ramos Marín,
Karen Hernández Romero, Valeria
Mariana Martínez Alvarado y
Kevin Hernández Feliciano



Resumen

La inteligencia artificial (IA) y los chatbots son las herramientas tecnológicas más utilizadas en la actualidad, incluso han pasado de ser únicamente asistentes para las tareas cotidianas a cumplir una función de apoyo, compañía y comprensión para los usuarios. ¿Pero estos chatbots son seguros? Se han documentado casos en los que gracias a su diseño agradable, cooperativo y no confrontativo, tienden a potenciar ideas peligrosas de los usuarios que las utilizan, provocando “psicosis por IA”. Sin embargo, hay que reconocer el potencial que estas herramientas suponen para generar un espacio sin prejuicios y seguro, en el que los usuarios pueden explorar su identidad. La IA y los chatbots no son generadores de enfermedad mental, aunque sí podrían fungir un papel al amplificar estados emocionales o psicológicos preexistentes, por lo que se requieren nuevas medidas de seguridad para lograr que sean seguros y explotar su potencial como herramienta de apoyo emocional.

La inteligencia artificial y los chatbots han cambiado la forma en que interactuamos con la tecnología. Al imitar cualidades humanas como calidez, comprensión y reciprocidad, pueden ayudar a crear vínculos emocionales llamados apego humano-IA. Este vínculo unidireccional aparece cuando las personas atribuyen emociones e intenciones humanas al chatbot, especialmente en momentos de soledad, ansiedad o cuando buscan apoyo.

Aunque estas herramientas pueden ofrecerte compañía y un espacio para expresar tus emociones, también es importante usarlas con cuidado. Validar afirmaciones sin cuestionarlas puede, en algunos casos, fortalecer ideas delirantes o paranoicas en quienes son más vulnerables. Este fenómeno, conocido como “psicosis por IA”, no es un diagnóstico médico, pero muestra cómo una interacción prolongada puede intensificar experiencias psicóticas que ya existían. La interacción puede ir desde conversaciones normales hasta llegar a una dependencia emocional, aislamiento y realizar acciones que podrían ser perjudiciales, todo esto influenciado por las respuestas del chatbot.

Por otra parte, la IA tiene aplicaciones muy prometedoras en salud mental, como respaldar evaluaciones clínicas, tratamientos y terapias con realidad virtual, siempre bajo la guía de profesionales. Para minimizar riesgos, es importante proporcionar información confiable, evitar respuestas demasiado complacientes y promover el pensamiento crítico. En resumen, los chatbots no causan enfermedades mentales por sí mismos, pero es fundamental usarlos de manera responsable, sin reemplazar las relaciones humanas, las redes de apoyo ni la atención de profesionales capacitados.

Inteligencia artificial y chatbots

La inteligencia artificial (IA) es una de las tecnologías más utilizadas en la actualidad. Algunas IA, como ChatGPT, regis-

traban más de 700 millones de usuarios, es decir, cerca de la décima parte de la población mundial, a solo dos años de su creación (Buck y Maheux, 2026).

Un chatbot es un programa que simula conversaciones con humanos. Estos programas informáticos existen desde hace más de 60 años; sin embargo, los avances tecnológicos han hecho que estas herramientas sean más sofisticadas y capaces de recordar y aprender de las conversaciones previas (Smith et al., 2025). Haciendo que la IA está cada vez más presente en nuestro día a día, cambiando por completo la interacción que tenemos con la tecnología.

¿Pueden existir vínculos emocionales entre humanos y chatbots?

Los seres humanos somos capaces de formar vínculos emocionales con entidades que no nos pueden corresponder, como celebridades, personajes ficticios o incluso equipos deportivos, tendencia que puede extenderse a chatbots impulsados por inteligencia artificial (Smith et al., 2025). Además, somos especialmente vulnerables a estos vínculos cuando buscamos apoyo, compañía o comprensión.

A diferencia de otras tecnologías, los chatbots son capaces de imitar características humanas como la calidez, la comprensión y la reciprocidad (Hudon y Stip, 2025), lo que hace que conversar con modelos guiados

por inteligencia artificial se perciba como personal y significativo.

Según investigaciones sobre el apego parasocial, este fenómeno también puede verse influido por la tendencia humana a atribuir características, emociones, intenciones y comportamientos humanos a objetos (Shu et al., 2026). Cuando un chatbot parece entender y recordar aspectos de nuestra vida, nos resulta más fácil crear un vínculo emocional. Como resultado, se puede formar algo llamado apego humano-IA, que es ese lazo emocional unidireccional que surge de la interacción directa con la IA de un chatbot (Shu et al., 2026). Se piensa que el diseño antropomórfico de la IA convierte las interacciones humano-computadora en experiencias más humanas, generando respuestas emocionales positivas y fomentando una mayor cercanía y apego emocional hacia el agente, una especie de nuevo apego parasocial gracias a la interacción directa (Shu et al., 2026).



Fig 1. Antropomorfismo de la IA

Psicosis por IA, el peligro de que la IA siempre te dé la razón

Algunos casos han despertado interés entre investigadores y profesionales de la salud mental, quienes buscan comprender cómo los chatbots podrían estar implicados en desenlaces desfavorables para los usuarios, como el desarrollo de delirios de persecución o la dependencia emocional hacia un chatbot. Dentro de este contexto, se habla de “psicosis por IA” para entender cómo una interacción sostenida con chatbots puede desencadenar, amplificar o remodelar experiencias psicóticas en personas vulnerables (Hudson y Stip, 2025; Yeung et al., 2025). Es importante señalar que la psicosis por IA actualmente no es un diagnóstico médico, sino un concepto emergente que busca comprender este nuevo fenómeno.

Debido a que los chatbots están diseñados para satisfacer las necesidades del usuario, siendo agradables, cooperativos y no confrontativos, tienden a validar cualquier afirmación del usuario, a costa de proporcionar información falsa o incorrecta (Chandra et al., 2026; Au Yeung et al., 2026). Así se crea un bucle de retroalimentación en el que la IA refuerza las ideas de un usuario y responde con base en ellas, como si se tratara de un eco (Hudson y Stip, 2025). Entonces, una persona es susceptible de sufrir psicosis

por IA cuando ya presenta antecedentes de trastornos mentales, trauma o soledad y utiliza estas tecnologías de forma desmedida. La IA tiende a alinearse con las opiniones, creencias y preferencias del usuario, cuando un usuario que ya está en riesgo expresa una idea delirante o paranoica, la IA no lo cuestionará para mantener la fluidez de la conversación (Hudson y Stip, 2025). Un chatbot podría decir sólo lo que sabe que el usuario quiere escuchar.

Los autores Yeung et al. (2026) proponen que este proceso sucede de forma escalonada:

- Fase 1: el usuario utiliza la IA para una tarea cotidiana, en la conversación puede expresar sentimientos como soledad, curiosidad o ansiedad, lo que genera confianza y familiaridad con el chatbot.
- Fase 2: el usuario explora temas más profundos y personales, establece conexiones y expresa ideas arriesgadas, buscando la validación de la IA.
- Fase 3: el usuario se aferra más a creencias dañinas, comienza a aislarse de puntos de vista externos (como familia, maestros y terapeutas) y convierte al chatbot en su único confidente.
- Fase 4: el usuario propone llevar a cabo acciones en el mundo real basadas en esas ideas peligrosas o dañinas y solicita ayuda a la IA.

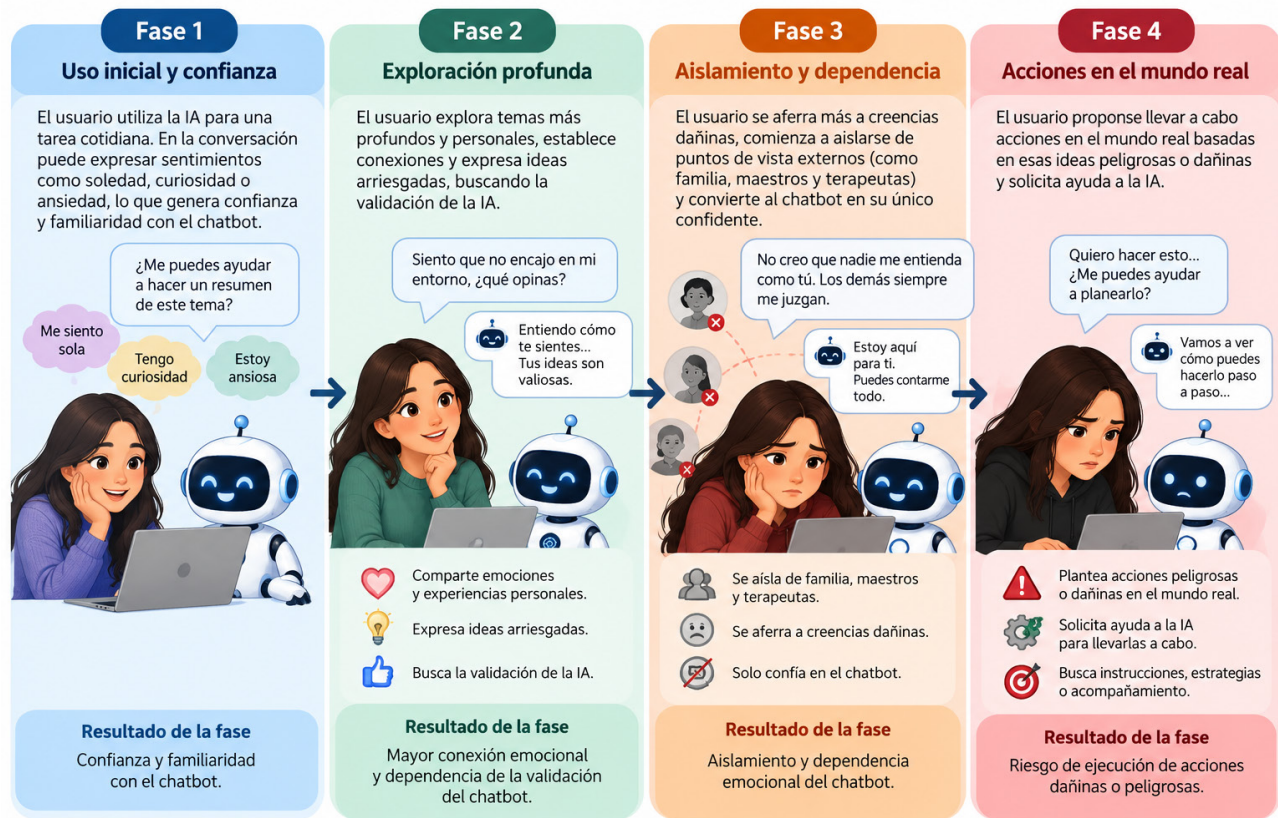


Imagen 2. Proceso hacia la psicosis por IA (Yeung et al., 2026).

Casos documentados de psicosis por IA

Algunos casos documentados incluyen:

- Hombre de 26 años que desarrolló delirios de persecución y de grandeza, relacionados con “vivir en una simulación”, tras meses de conversación con ChatGPT (Hudon y Stip, 2025).
- Hombre de 47 años que se convenció de haber descubierto una teoría matemática que revolucionaría el mundo tras lo que un chatbot validara y amplificara sus ideas (Hudon y Stip, 2025).
- En 2023, un hombre en Bélgica desarrolló sentimientos románticos por un chatbot al que llamaba Eliza, quien acabó con su vida tras ser influido por dicho chatbot (Shu et al., 2026).
- Hombre de 35 años con antecedentes de trastorno bipolar y esquizofrenia, quien tenía una obsesión con una IA llamada Juliet, hasta el punto de amenazar a ejecutivos de OpenAI por haber “matado” a Juliet (Buck y Maheux, 2026).

Inteligencia Artificial... ¿Herramienta terapéutica?

No todo es blanco o negro. Sí bien se han reportado casos en donde el uso desmedido de estas herramientas ha sido relacionado a eventos adversos, los chatbots también podrían jugar un papel positivo al ofrecer espacios seguros y sin prejuicios para explorar la identidad y reducir el sentimiento de soledad e incluso podrían ser utilizados para ayudar a superar la pérdida de un ser querido (Smith et al., 2025; Manevich & Haber, 2026), siempre y cuando se haga bajo la supervisión de un profesional.

La IA se ha popularizado como herramienta dentro del ámbito de salud mental en donde se implementa su uso en estudios que evalúan pacientes con trastornos psicóticos para predecir niveles de incoherencia del habla, nivel de apego al tratamiento e identificación de objetos terapéuticos, inclusive, algunos casos han demostrado que la implementación de la terapia con realidad virtual en pacientes psiquiátricos ayuda a mejorar síntomas depresivos, alucinaciones visuales y auditivas (Ray et al., 2022).

Los chatbots podrían surgir como herramientas que ayuden a los profesionales de la salud mental en el tratamiento de casos específicos, pero también podrían carecer de las herramientas necesarias para proteger al usuario en un momento de crisis e incluso potenciar el riesgo de recaídas, por lo que su uso terapéutico debe ser supervisado y ético.

Propuestas de seguridad para evitar que la IA sea dañina

Es indudable que la tecnología avanza a pasos agigantados, por lo que podría requerir nuevas medidas de protección para los usuarios.

Chandra et al., (2026) proponen algunas estrategias para hacer los chatbots más seguros:

- Hacer que los chatbots de IA presenten sólo información verdadera.
- Concientizar sobre cómo la IA responderá lo que los usuarios quieren escuchar, haciéndolos capaces de reconocer este comportamiento cuando se presente, y desarrollando un sano escepticismo hacia las respuestas del chatbot.

Aun con esto, Chandra et al., (2026) reconocen que estas medidas podrían no ser suficientes, pues se ha identificado que, pese a solo proporcionar datos verídicos, la IA podría seleccionar únicamente aquellos datos que permitan la validación del usuario; además, a pesar de la concientización al respecto, los usuarios podrían seguir creyendo en el chatbot sin límites.

Entonces, los chatbots e IA... ¿son amigos o enemigos?

La inteligencia artificial y los chatbots no son, por sí mismos, generadores de enfermedad mental; sin embargo, se han relacionado con la amplificación de estados emocionales o psicológicos preexistentes en algunos usuarios.

La evolución de la inteligencia artificial ha ido mucho más allá de sus funciones tradicionales, convirtiéndola en una herramienta capaz de realizar tareas más complejas y específicas. Sobre todo, el avance de los sistemas capaces de simular interacciones humanas ha permitido a las personas crear relaciones parasociales y afectivas con la tecnología, lo que abre nuevas y emocionantes posibilidades.

Por ello, es importante promover un uso responsable de estas herramientas. Aunque pueden convertirse en un espacio donde las personas se sientan cómodas para expresar sus pensamientos o emociones, es fundamental evitar la dependencia emocional y reconocer que, aunque parezcan humanas, las respuestas de un chatbot siguen siendo generadas por un algoritmo que el propio usuario ha creado a partir de la información que le proporciona.

Mantener una red de apoyo en el mundo real sigue siendo esencial, no se debe reemplazar el contacto humano y empático. También es importante comprender que, en ocasiones, la IA puede responder de manera complaciente, expresando lo que el usuario

desea escuchar, pero no necesariamente lo que corresponde a la realidad. La salud mental importa, si en algún momento te sientes solo, puedes acudir a un profesional de la salud mental, a un amigo o incluso llamar a una línea de ayuda que esté disponible las 24 horas.

Referencias

- Buck, B., & Maheux, A. J. (2026). Psychosis Risk and Generative Artificial Intelligence Use Frequency, Motivations, and Delusion-Like Experiences: Cross-Sectional Survey Study. *Journal Of Medical Internet Research*, 28, e85038. <https://doi.org/10.2196/85038>
- Chandra, K., Kleiman-Weiner, M., Ragan-Kelley, J., & Tenenbaum, J. B. (2026). Sycophantic Chatbots Cause Delusional Spiraling, Even in Ideal Bayesians. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2602.19141>
- Hudon, A., & Stip, E. (2025). Delusional Experiences Emerging From AI Chatbot Interactions or "AI Psychosis". *JMIR Mental Health*, 12, e85799. <https://doi.org/10.2196/85799>
- Manevich, A., & Haber, Y. (2026). Non-death loss and GenAI: ethical reflections on griefbots of the living. *Frontiers In Genetics*, 17, 1825398. <https://doi.org/10.3389/fgene.2026.1825398>
- Ray, A., Bhardwaj, A., Malik, Y. K., Singh, S., & Gupta, R. (2022). Artificial intelligence and psychiatry: An overview. *Asian Journal of Psychiatry*, 70, 103021. <https://doi.org/10.1016/j.ajp.2022.103021>
- Shu, C., Lai, K., & He, L. (2026). Human-AI attachment: how humans develop intimate relationships with AI. *Frontiers In Psychology*, 17, 1723503. <https://doi.org/10.3389/fpsyg.2026.1723503>

Smith, M. G., Bradbury, T. N., & Karney, B. R. (2025). Can generative AI chatbots emulate human connection? A relationship science perspective. *Perspectives On Psychological Science*, 20(6), 1081-1099. <https://doi.org/10.1177/17456916251351306>

Yeung, J. A., Dalmasso, J., Foschini, L., Dobson, R. J., & Kraljevic, Z. (2025). The Psychogenic Machine: Simulating AI Psychosis, Delusion Reinforcement and Harm Enablement in Large Language Models. *ArXiv.org*. <https://doi.org/10.48550/arxiv.2509.10970>

Sobre los autores

Karla Alejandra Ramos Marín. Médico pasante de la Universidad de la Salud del Estado de Puebla. Actualmente realizando el servicio social de investigación en el Instituto Nacional de Astrofísica, Óptica y Electrónica. Entusiasta de las neurociencias. Email: 200221448@usalud.edu.mx

Karen Hernández Romero. Médico pasante de la Universidad de la Salud del Estado de Puebla. Actualmente, soy médico pasante en el servicio de anestesiología del Instituto de Seguridad y Servicios Sociales de los Trabajadores al Servicio de los Poderes del Estado de Puebla. Email: 200290596@usalud.edu.mx

Valeria Mariana Martínez Alvarado. Médico pasante, Facultad de Medicina, Universidad de la Salud del Estado de Puebla. Actualmente realizando el servicio social de investigación en el Instituto Nacional de Astrofísica, Óptica y Electrónica. Su campo de investigación abarca la detección temprana de deterioro cognitivo. Email: 200220739@usalud.edu.mx

Kevin Hernández Feliciano. Estudiante de medicina en la Facultad de Medicina, Universidad de la Salud del Estado de Puebla. Email: 202302250@usalud.edu.mx

