

Psicoterapia e inteligência artificial: limites tecnológicos en la promoción de la salud mental

Sophia Ivantes-Rodrigues¹, Marília da Matta Silva¹, Leonardo Pestillo de Oliveira¹, Lucas França Garcia¹

1. Universidade Cesumar, Maringá/PR, Brasil.

Resumen

La integración de la inteligencia artificial en psicología ha crecido significativamente, aunque las cuestiones bioéticas relacionadas con esta práctica siguen siendo poco exploradas. Este ensayo crítico analiza los desafíos bioéticos y biojurídicos, así como aquellos relacionados con el Código de Ética del Profesional de la Asociación Americana de Psicología respecto al uso de inteligencias artificiales en psicoterapia. Se desarrollaron dos categorías principales: "Inteligencias artificiales y psicoterapia: funcionalidad, integración y aplicación" y "Desafíos y límites bioéticos en la integración de la inteligencia artificial en la psicología". Este ensayo analizó la naturaleza de estas intervenciones destacando sus posibles riesgos y beneficios, así como los límites éticos involucrados. Se concluye que la integración de inteligencias artificiales en psicoterapia presenta riesgos y limitaciones que requieren estudios más profundos, además de una mayor atención y debate por parte de la comunidad científica y los profesionales del área.

Palabras clave: Inteligencia artificial. Bioética. Psicología. Confidencialidad. Privacidad.

Resumo

Psicoterapia e inteligência artificial: limites tecnológicos na promoção da saúde

A integração da inteligência artificial (IA) na psicologia tem registrado um crescimento significativo, porém as questões bioéticas associadas ainda são pouco discutidas. Este ensaio crítico analisa os desafios bioéticos, biojurídicos e os que se relacionam com o Código de Ética do Profissional da Associação Americana de Psicologia no contexto do uso de inteligências artificiais para psicoterapia. Duas categorias principais foram desenvolvidas: "Inteligência artificial e psicoterapia: funcionalidade, integração e aplicação" e "Desafios e limites bioéticos na integração da inteligência artificial na psicologia". O ensaio examina o caráter dessas intervenções, destacando seus riscos e benefícios potenciais, bem como os limites éticos envolvidos. Conclui-se que a integração de inteligências artificiais na psicoterapia apresenta riscos e limitações que demandam estudos mais aprofundados e maior atenção e discussão por parte da comunidade científica e dos profissionais da área.

Palavras-chave: Inteligência artificial. Bioética. Psicologia. Confidencialidade. Privacidade.

Abstract

Psychotherapy and artificial intelligence: the limits of technology in the promotion of mental health

The integration of artificial intelligence in psychology has been growing significantly, although bioethical issues related to this practice remain underexplored. This study aims to analyze the bioethical and biolegal challenges, including those related to the American Psychological Association's Code of Ethics, concerning the use of artificial intelligence in psychotherapy. This critical article addresses the topic from a reflective perspective. Two main categories were developed: "Artificial intelligence and psychotherapy: functionality, integration, and application" and "Bioethical challenges and limits in the integration of artificial intelligence in psychology." The nature of these interventions was analyzed, highlighting their potential risks and benefits, as well as the ethical boundaries involved. It is concluded that the integration of artificial intelligence into psychotherapy practice presents risks and limitations that require further studies, increased scrutiny, and discussion by the scientific community and professionals in the field.

Keywords: Artificial intelligence. Bioethics. Psychology. Confidentiality. Privacy.

Los autores declaran que no existe ningún conflicto de interés.

Con el crecimiento en el uso de las redes sociales y los avances tecnológicos en el siglo XXI, la integración entre la psicología, las ciencias de la conducta y las inteligencias artificiales (IA) se ha convertido en una realidad inherente^{1,2}. A pesar de las importantes contribuciones de dicha integración, como la ampliación del acceso a la psicoterapia, surgen preocupaciones éticas y morales. Estas cuestiones incluyen posibles infracciones del Código de Ética Profesional de la Asociación Americana de Psicología (APA), además de desafíos a los principios de la bioética y el bioderecho³.

Las principales preocupaciones son los impactos positivos y negativos para los pacientes y psicólogos, así como el uso de datos para mercadotecnia y la naturaleza aún poco conocida de las intervenciones realizadas por IA, que suelen considerarse una psicoterapia no auténtica. En este escenario, los principios bioéticos, como la beneficencia y la no maleficencia, se ven amenazados por riesgos como los sesgos algorítmicos y las decisiones automatizadas imprecisas, mientras que se cuestiona la justicia en cuanto a la accesibilidad y distribución equitativa de estas tecnologías^{4,5}. Holohan y Müller⁶, en este contexto, presentan el concepto de “*significant otherness*”, que resalta la relevancia de la alteridad y el reconocimiento de la singularidad del paciente en las psicoterapias asistidas por IA.

Bajo la perspectiva del bioderecho, la creciente aplicación de la IA en la psicoterapia exige una estricta vigilancia respecto a la privacidad, confidencialidad y seguridad de los datos personales, de conformidad con los reglamentos como la Ley General de Protección de Datos Personales (LGPD)⁷ y el General Data Protection Regulation (GDPR)⁸. Aspectos cruciales, como el consentimiento informado, la protección de información sensible y la transparencia en las operaciones algorítmicas, son esenciales para garantizar los derechos de los pacientes^{3,9}. Por lo tanto, este ensayo realiza un análisis crítico de las implicaciones bioéticas del uso de IA en psicoterapia, y explora los principios bioéticos, la importancia de la alteridad y los desafíos jurídicos que acompañan a esta innovación.

Método

Tipo de estudio

Este ensayo crítico trata de las principales cuestiones bioéticas relacionadas con la integración de

IA en el contexto de la psicoterapia, con base en un análisis reflexivo de la literatura disponible. Esta metodología permite realizar una evaluación profunda de las diversas investigaciones sobre este tema, con el objetivo de discutir los desafíos y las implicaciones éticas de integrar la tecnología de IA en la psicología.

Fuentes de información y estrategia de búsqueda

Se realizaron búsquedas en las bases de datos PubMed, Scopus, Web of Science y SciELO en enero de 2025, utilizando la combinación de los siguientes términos: “*artificial intelligence*”, “*psychotherapy*” y “*bioethics OR ethics*”.

Criterios de inclusión y exclusión

El ensayo incluyó artículos publicados en revistas científicas de 2020 a 2025, en inglés, español y portugués, centrados en los aspectos éticos, bioéticos y jurídicos de la integración de IA en psicoterapia. Se descartaron los artículos que no se alineaban con el objetivo principal.

Cribado y selección de estudios

Tras recopilar los datos en las bases de datos mencionadas anteriormente, se analizaron los artículos seleccionados en la primera etapa del cribado en su totalidad. Las divergencias entre los evaluadores se solucionaron mediante consenso o, en su caso, mediante la intervención de un tercer revisor.

Análisis y síntesis de los datos

Tras concluir el cribado, se descargaron los artículos en su totalidad, y se organizaron y codificaron sus datos a través del software QSR NVivo 15 para Windows¹⁰. Se realizó el análisis con base en el método de análisis de contenido que propuso Bardin, y consistió en la categorización temática de los datos y su interpretación para identificar patrones y tendencias recurrentes¹¹. La síntesis de los datos se realizó de manera narrativa, resaltando los principales resultados y explorando sus relaciones con las cuestiones bioéticas sobre el uso de IA en la psicoterapia. Esta estrategia tuvo el objetivo de identificar patrones, contrastes y brechas en los estudios evaluados¹². Además, dos investigadores

(SRI y LFG) realizaron análisis y codificaciones independientes, y agruparon las unidades de registro en unidades de significado (conceptos clave), que se consolidaron en categorías temáticas.

Tabla 1. Categorías temáticas

Unidad temática central	Categorías temáticas	Unidades de significado (ejemplos)	Evidencias en los artículos
IA y psicoterapia: funcionalidad, integración y aplicación	Eficacia clínica	Chatbots (bots conversacionales) aumentan la adherencia terapéutica ¹³	12/17 artículos (el 70,6%)
		Limitaciones de la empatía artificial ¹⁴	
	Digital therapeutic alliance (DTA)	Replicación de vínculos mediante la personalización ¹⁵	9/17 artículos (el 52,9%)
		Falta de profundidad emocional ⁶	
Desafíos y límites bioéticos	Autonomía y consentimiento informado	Opacidad algorítmica compromete las decisiones ¹⁶	14/17 artículos (el 82,4%)
		Necesidad de un lenguaje accesible ¹⁷	
	Protección de datos y cumplimiento legal	Riesgos del <i>digital phenotyping</i> según la LGPD/GDPR ¹⁸	11/17 artículos (el 64,7%)
		Falta de criptografía robusta ¹⁹	

Resultados y discusión

En esta sección, se discutirán los principales hallazgos del ensayo crítico acerca de la implementación de inteligencias artificiales en la psicoterapia, los cuales se estructuraron en categorías temáticas resultantes del análisis de contenido de los 17 artículos analizados. Los artículos proceden de regiones como Europa, América del Norte y Asia, y no se identificó ningún estudio brasileño que cumpliera los criterios de inclusión.

Se resaltaron las implicaciones éticas del uso de IA en la práctica psicoterapéutica, abordando sus impactos positivos y negativos para psicólogos y pacientes. Las categorías desarrolladas reflejan los principales debates y desafíos bioéticos presentes en la literatura, que incluyen: “IA y psicoterapia: funcionalidad, integración y aplicación” y “Desafíos y límites bioéticos en la integración de las IA en la psicología”. Además, también se definieron subcategorías para explorar cuestiones específicas, tales como “Posibles infracciones y directrices del Código de Ética del Profesional de la Asociación Americana de Psicología” e “Implicaciones jurídicas”. Esta estructura permite realizar un análisis más detallado de los límites normativos y legales asociados a la implementación de esta tecnología en el entorno clínico.

IA y psicoterapia: funcionalidad, integración y aplicación

El uso de IA en psicoterapia se está desarrollando rápidamente, lo que proporciona nuevas herramientas para mejorar la salud mental, al tiempo que introduce importantes desafíos. Grodniewicz y Hohol¹⁴ señalan tres obstáculos principales: el primero, el “terapeuta confundido”, resalta la dificultad de la IA para comprender matices emocionales específicas; el segundo, el “terapeuta no humano”, subraya las limitaciones en las relaciones interpersonales y la falta de empatía genuina; y el tercero, el “terapeuta de inteligencia restringida”, demuestra la dificultad de la IA para proporcionar intervenciones integrales e integradas. Alfano y colaboradores¹³, así como Khawaja y Bélisle-Pipon¹⁶, reconocen el papel de la IA en la ampliación del acceso a tratamientos psicológicos, especialmente en comunidades vulnerables, pero también advierten sobre los riesgos relacionados con dicha práctica, incluida la estandarización excesiva y el comprometimiento de la humanización en la atención.

Alfano y colaboradores¹³ señalan que, debido a la pandemia de COVID-19, hubo un crecimiento notable en la cantidad de empresas que empezaron a ofrecer servicios de psicoterapia utilizando IA. Dichas empresas flexibilizaron el acceso

a la psicoterapia, eliminando barreras como limitaciones geográficas, vulnerabilidad y costos relacionados con dicha práctica. Aunque los datos proporcionados por estas empresas son limitados, revelan que los usuarios más asiduos tienen entre 16 y 25 años, un grupo de edad que ha crecido inmerso en la tecnología y ha experimentado su adopción a gran escala durante la pandemia.

La demanda creciente de psicoterapia mediada por IA, principalmente entre los adolescentes, está relacionada con factores como la vergüenza de compartir acontecimientos y sentimientos con psicólogos humanos, la familiaridad con las plataformas digitales y la superación de barreras físicas y económicas. Sin embargo, a pesar de estos beneficios, algunos aspectos fundamentales de la práctica psicológica, como el abandono de la psicoterapia y la falta de un entorno terapéutico apropiado, plantean preguntas relevantes. Las preocupaciones incluyen intereses financieros de las empresas, escasez de pruebas que comprueben la eficacia de los servicios y falta de cualidades humanizadas en los *chatbots*, como la empatía y la confidencialidad. Además, se cuestiona el funcionamiento de los algoritmos, sobre todo con respecto a los criterios de evaluación de la calidad de los servicios y a las posibles transformaciones en el mercado laboral¹³.

De hecho, ya se están implementando *chatbots* en aplicaciones que utilizan algoritmos para detectar patrones lingüísticos que indiquen riesgos de suicidio y depresión en jóvenes adultos y adolescentes. En este escenario, la IA destaca debido a su capacidad de procesar grandes volúmenes de datos en pocos minutos, mientras que el tratamiento, por sí solo, sigue dependiendo profundamente del psicólogo y del valor intrínseco de la relación humana en la intervención y el contacto con el paciente¹³. Por lo tanto, la calidad de la relación terapéutica mediada por IA es crucial. Malouin-Lachance y colaboradores¹⁵ introducen el concepto de alianza terapéutica digital, y resaltan la tendencia a formar vínculos terapéuticos mediante interacciones digitales, siempre que estas sean personalizadas, responsivas y se adapten a la historia emocional del paciente. Sin embargo, dichas interacciones digitales son simplemente una reproducción de los mecanismos terapéuticos tradicionales, que se manifiestan de una manera significativamente diferente de las interacciones

humanas, principalmente con respecto a la profundidad emocional.

Holohan y Müller⁶ amplían esta discusión con el concepto de *significant otherness*, y alegan que las experiencias terapéuticas mediadas por IA pueden tener un impacto emocional considerable cuando esta tecnología se desarrolla con empatía digital, interactividad adaptativa y lenguaje afectivo. Sugieren que la relación entre paciente e IA, en este contexto, se puede comparar con el concepto psicoanalítico de transferencia, en el que las proyecciones emocionales del paciente sobre la tecnología pueden explorarse mejor mediante estudios en profundidad y evaluadas con una visión crítica por psicólogos e investigadores.

En el ámbito de las innovaciones prácticas, Özden³ investiga el potencial de las terapias inmersivas con IA, que abarcan la realidad aumentada, los entornos simulados y las intervenciones sensoriales personalizadas, resaltando su eficacia en el tratamiento de fobias, traumas y trastornos de ansiedad. Zhang² resalta la importancia del reconocimiento emocional en tiempo real, y detalla como los algoritmos de la IA, al analizar las variaciones en el lenguaje, los tonos de voz y las expresiones faciales, hacen más responsivas las interacciones, ajustando el diálogo según las emociones detectadas. Dehbozorgi, Zangeneh, Khooshab, Nia y Hanif¹⁷ refuerzan la necesidad de cocreación tecnológica, y alegan que es fundamental la participación activa de psicoterapeutas y pacientes en el proceso de desarrollo de estas herramientas, para garantizar mayor adherencia, confianza y seguridad, y evitar fallos éticos en el diseño de las intervenciones.

Así, los avances de la IA en psicoterapia presentan tanto oportunidades como desafíos significativos. Y esto implica no solo el desarrollo de algoritmos más atentos al contexto emocional de los pacientes, sino también la adopción de prácticas éticamente adecuadas, que deben incluir supervisión humana constante, auditoría transparente de los procesos algorítmicos y formación ética para los profesionales del área. Además, es fundamental fomentar estudios longitudinales que analicen los efectos a largo plazo de las terapias asistidas por IA, para garantizar que ellas no solo aumenten la eficiencia tecnológica, sino también preserven la seguridad emocional y la calidad de la atención psicoterapéutica.

Desafíos y límites bioéticos en la integración de las IA en la psicología

La integración de la IA en psicoterapia plantea profundos cuestionamientos bioéticos, principalmente con respecto a los principios de la ética biomédica establecidos por Beauchamp y Childress⁵. Estos principios incluyen: la autonomía, que respeta la autodeterminación del paciente y su capacidad de tomar decisiones informadas; la beneficencia, que requiere la promoción activa del bienestar del paciente; la no maleficencia, que impone la obligación de evitar causar daños; y la justicia, que requiere una distribución justa y equitativa de beneficios y riesgos.

La autonomía es un pilar fundamental en la relación terapéutica, que implica el derecho del paciente a tomar decisiones informadas, sin coacción ni manipulación. Khawaja y Bélisle-Pipon¹⁶ y Zhang² señalan que la IA puede socavar este principio al generar una percepción engañosa de vínculo terapéutico, sobre todo cuando el papel de la tecnología en la interacción no es evidente. El consentimiento informado es un aspecto crucial: Dehbozorgi y colaboradores¹⁷ resaltan que debe ser transparente, accesible y comprensible, para evitar malentendidos sobre las limitaciones y capacidades de la IA. Holohan y Müller⁶ también señalan que la opacidad de los algoritmos perjudica la comprensión del paciente sobre los procesos que guían las decisiones automatizadas. Por lo tanto, para promover la autonomía, las plataformas de IA deben adoptar prácticas como la transparencia algorítmica, el consentimiento adaptativo y el uso de un lenguaje claro. Tener este compromiso con la autodeterminación digital refuerza el papel del paciente como agente activo en su proceso terapéutico.

El principio de beneficencia implica la promoción activa del bienestar del paciente, mientras que el de no maleficencia impone el deber de evitar daños. Stoll y colaboradores⁴ y Grodniewicz y Hohol¹⁴ advierten que, aunque la IA amplía el acceso a terapias, los algoritmos sesgados o entrenados con datos limitados pueden, en realidad, causar más daños que beneficios. Y esto incluye intervenciones inadecuadas o potencialmente perjudiciales, además de perpetuar estigmas y prejuicios. Además, Özden³ resalta que la falta de supervisión adecuada en terapias inmersivas

puede resultar en malestar emocional o dependencia tecnológica.

Además, también surgen preocupaciones éticas relacionadas con el uso de la IA: Dehbozorgi y colaboradores¹⁴ señalan que los sistemas automatizados que proporcionan diagnósticos sin una validación humana pueden generar falsas expectativas o tratamientos inadecuados. Otra cuestión es que la personalización inadecuada de las interacciones con IA puede dar lugar a respuestas genéricas y comprometer la atención, según sugieren Holohan y Müller⁶. Para mitigar estos riesgos, es fundamental implementar prácticas de auditoría algorítmica, pruebas continuas y el monitoreo humano de las decisiones automatizadas. Por lo tanto, integrar protocolos de seguridad, entrenar la IA con datos representativos y validar constantemente la eficacia de las terapias digitales son prácticas esenciales para garantizar el cumplimiento de los principios de beneficencia y no maleficencia.

La justicia en la bioética requiere que los beneficios y riesgos de la IA en psicoterapia se distribuyan de manera equitativa, para promover accesibilidad e inclusión². En este sentido, Bhatt²⁰ y Gutierrez y colaboradores²¹ señalan que la justicia se cumple cuando las innovaciones tecnológicas llegan a diferentes grupos sociales, evitando disparidades en el acceso. Grodniewicz y Hohol¹⁴ advierten que los algoritmos entrenados con datos homogéneos pueden perpetuar desigualdades raciales y culturales, y resaltan la necesidad de conjuntos de datos diversos y auditables. Stoll y colaboradores⁴ también resaltan que desigualdades digitales, como la falta de acceso a internet, intensifican la exclusión de grupos vulnerables. Por lo tanto, la justicia requiere una perspectiva multidimensional que tenga en cuenta los impactos sociales, culturales y económicos de la IA en psicoterapia.

La alteridad, que significa reconocer la singularidad del otro, es esencial para crear relaciones terapéuticas significativas⁶. Tavory⁹ también defiende que se debe desarrollar la IA con *responsability*, es decir, con la capacidad de responder con sensibilidad a las subjetividades de cada persona. Dehbozorgi y colaboradores¹⁷ añaden que la falta de diversidad cultural en los algoritmos puede resultar en interacciones insensibles, lo que refuerza estereotipos y estigmas sociales en lugar de proporcionar una atención verdaderamente inclusiva.

Posibles infracciones y directrices del Código de Ética del Profesional de Psicología

El Código de Principios Éticos de la APA orienta que los psicólogos solo deben diagnosticar después de evaluar minuciosamente al paciente. Y esto crea una zona gris, visto que no especifica si el uso de IA en la práctica clínica es éticamente adecuado o no. Sin embargo, la APA permite a los psicólogos adoptar estas tecnologías, siempre que aclaren las limitaciones de estos métodos y cómo pueden influir en la confiabilidad y validez de sus conclusiones¹.

En cuanto a la confidencialidad y privacidad del consentimiento informado, el psicólogo debe informar al paciente que su privacidad puede estar en riesgo, sobre todo si los datos son para fines de mercadotecnia. Además, es crucial explicar cuáles son los posibles daños y el objetivo del uso de esta información y obtener el consentimiento del paciente de forma clara. Cabe recordar que el uso de estas prácticas no cumple con la *Health Insurance Portability and Accountability Act of 1996*, lo que puede comprometer la privacidad del paciente. La APA también resalta que el uso de las redes sociales para evaluar a los pacientes puede llevar a un uso excesivo de información que no es necesaria para la evaluación, sobre todo cuando se exploran dichas redes a través de IA¹.

Aunque proporcionan servicios personalizados, como atención ilimitada en cuanto a horarios y lugares, consideración de características culturales y personales e integración de enfoques como la terapia cognitivo-conductual (TCC) y el psicoanálisis, los *chatbots* tienen limitaciones. Dichas limitaciones —como la posible explotación de usuarios, consejos con sesgos algorítmicos y recursos terapéuticos que pueden ser perjudiciales— pueden llevar al malentendido terapéutico (MT) cuando se malinterpretan o se ignoran¹⁴.

El MT, un tema ampliamente discutido en la bioética, se refiere a la subestimación de las capacidades de la IA para proporcionar un apoyo terapéutico adecuado, ético y eficaz. Este malentendido puede surgir de cuatro maneras: a través de mercadotecnia imprecisa, formación de una alianza terapéutica digital, diseño inadecuado de los *chatbots* y limitación de la autonomía del usuario¹⁴.

Además, la falta de empatía genuina de los *chatbots*, debido a la ausencia de emociones, puede causar daños al usuario e impedir que se

establezca un vínculo terapéutico, algo fundamental en la psicoterapia. Aunque las IA son capaces de identificar y manipular emociones y vulnerabilidades, pueden poner a grupos vulnerables y minoritarios en riesgo de daños y expectativas poco realistas⁹. Sin la presencia de un psicólogo humano, no se desarrolla la confianza ni se toman decisiones compartidas, lo que puede comprometer la autonomía del paciente^{3,14}. Además, aunque las IA demuestran eficacia a corto plazo, no presentan resultados positivos en servicios a largo plazo. Cuestiones como la retención de memoria pueden afectar la continuidad del proceso terapéutico³.

Implicaciones jurídicas

En cuanto a los desafíos jurídicos y de protección de datos, Gerke y colaboradores¹⁸ resaltan que el uso de IA en salud mental debe seguir marcos normativos como la LGPD y el GDPR. También resaltan la importancia de tener protocolos rigurosos para recopilar, almacenar y procesar datos sensibles. Jin y colaboradores¹⁹ señalan que son altos los riesgos de la recopilación continua de datos por agentes conversacionales y plataformas de telepsicología, sobre todo cuando existe un intercambio no autorizado de información. Stoll y colaboradores⁴ señalan la necesidad de políticas claras de consentimiento y de criptografía robusta para proteger la confidencialidad del paciente.

Dehbozorgi y colaboradores¹⁷ advierten sobre los riesgos del *digital phenotyping* —el análisis de patrones digitales para inferir el estado emocional del paciente— y recomiendan que estos sistemas cumplan con las directrices de la LGPD y del GDPR. Gerke y colaboradores¹⁸ defienden la realización de auditorías algorítmicas para garantizar que el uso secundario de datos esté debidamente regulado. Mientras tanto, Holohan y Müller⁶, así como Grodniewicz y Hohol¹⁴, resaltan que la complejidad de los sistemas de IA puede hacer menos transparente el proceso de consentimiento, lo que dificulta la comprensión de los riesgos que conlleva el uso de estas tecnologías en terapias digitales. Stoll y colaboradores⁴ sugieren adoptar formularios de consentimiento interactivos y adaptativos, que sean capaces de informar a los pacientes, de manera clara, sobre cómo funcionan las tecnologías utilizadas.

Estas reflexiones demuestran que el uso ético de la IA en psicoterapia va más allá de la simple

innovación tecnológica; requiere prácticas centradas en los derechos humanos, con enfoque en responsabilidad, empatía e inclusión. Es esencial desarrollar sistemas de IA que incluyan procesos participativos, involucrando a psicoterapeutas, pacientes y expertos en bioética, para garantizar que los principios bioéticos se apliquen de forma práctica y efectiva. Los desafíos jurídicos y de protección de datos resaltan la necesidad de equilibrar la innovación tecnológica con el respeto a los derechos fundamentales del paciente. Alinear estos avances tecnológicos con el cumplimiento normativo es crucial para garantizar ética y seguridad en las prácticas psicoterapéuticas mediadas por IA.

El éxito de estas tecnologías se evaluará no solo por su eficiencia técnica, sino también por su capacidad de promover dignidad, seguridad y atención centrada en el paciente. Además, una regulación clara y la creación de comités interdisciplinarios que evalúen de forma continua los impactos éticos de la IA son pasos esenciales. E esto asegura que los avances tecnológicos contribuyan a una práctica psicoterapéutica verdaderamente ética.

Consideraciones finales

Las implicaciones del uso de la inteligencia artificial en psicoterapia, analizadas bajo las dimensiones éticas, jurídicas y relacionales, revelan la complejidad de este avance y sus desafíos fundamentales. La IA tiene el potencial de democratizar el acceso a la psicoterapia y personalizar las intervenciones. Sin embargo, Grodniewicz¹⁴ y Dehbozorgi y colaboradores¹⁷ advierten sobre el riesgo de vínculos ilusorios y las limitaciones de los algoritmos para comprender la complejidad de las experiencias humanas. El análisis bajo los principios de la bioética expone los siguientes dilemas: la amenaza a la autonomía ante consentimientos opacos¹⁶, los riesgos a la beneficencia con sesgos algorítmicos⁴ y las desigualdades en el acceso a la tecnología, lo que compromete la justicia²⁰. A la vez, Gerke y colaboradores¹⁸ y Jin y colaboradores¹⁹ resaltan la urgencia de adecuación a la LGPD y al GDPR, subrayando que el consentimiento informado y la protección de datos son pilares innegociables para la ética en el uso de IA en salud mental.

Holohany Müller⁶, así como Tavory⁹, resaltan que, para respetar la alteridad, se debe diseñar la IA con *response-ability*, es decir, con la capacidad de responder con sensibilidad a las singularidades culturales y emocionales del usuario. Bhatt²⁰ refuerza la importancia de la diversidad en los conjuntos de datos, para evitar sesgos culturales.

Las IA aplicadas a los servicios de salud mental aún se encuentran en etapa inicial de desarrollo, por lo que no deben reemplazar a los psicólogos humanos. Más bien deben utilizarse como herramientas para ampliar las capacidades de intervención, sobre todo en cuestiones como la memoria humana, la sobrecarga de profesionales y la rápida identificación de señales tempranas de problemas de salud mental, como ideas suicidas y episodios de depresión^{2,3,16}. Así, para evitar la normalización de una dependencia excesiva de estas tecnologías y reducir los impactos relacionados con la búsqueda de apoyo y soporte social, es esencial resaltar repetidas veces que las IA no reemplazan al psicólogo humano¹³.

La inversión gubernamental en servicios de salud mental, junto con la colaboración multiprofesional entre psicólogos, desarrolladores de IA y profesionales del derecho, es fundamental para crear normativas que garanticen los derechos de los usuarios de forma transparente y clara. Además, es fundamental mejorar las IA, para que se adapten a las diferencias culturales, sociales y económicas. Se sugiere que estos servicios ofrezcan la opción de atención humana en situaciones de crisis y necesidades de apoyo más complejas.

La integración de la IA en psicoterapia solo será éticamente adecuada y eficaz si combina innovación tecnológica con responsabilidad ética y legal. Es imprescindible que las prácticas estén reguladas, sean transparentes y se centren en el respeto a la dignidad humana, además de buscar un equilibrio entre el potencial técnico y el cuidado humano que caracteriza la psicoterapia. Así, se concluye que las IA no deben sustituir los servicios prestados por profesionales de la psicología, sino que deben utilizarse como herramientas complementarias. Hay que respetar los principios éticos, los derechos del paciente y garantizar la humanización en las atenciones clínicas.

Referencias

1. Fleming MN. Considerations for the ethical implementation of psychological assessment through social media via machine learning. *Ethics Behav* [Internet]. 2021 [acceso 11 ago 2025];31(3):181-92. DOI: 10.1080/10508422.2020.1817026
2. Zhang Z, Wang J. Can AI replace psychotherapists? Exploring the future of mental health care. *Front Psychiatry* [Internet]. 2024 [acceso 11 ago 2025];15:1-7. DOI: 10.3389/fpsyt.2024.1444382
3. Özden HC. The use of artificial intelligence in psychotherapy: practical and ethical aspects. *Turk J Psychiatry* [Internet]. 2024 [acceso 11 ago 2025];35(4):336-9. Disponível: <https://bit.ly/46WIOTr>
4. Stoll J, Müller JA, Trachsel M. Ethical issues in online psychotherapy: a narrative review. *Front Psychiatry* [Internet]. 2020 [acceso 11 ago 2025];10:1-16. DOI: 10.3389/fpsyt.2019.00993
5. Beauchamp TL, Childress JF. Principles of biomedical ethics. 5ª ed. New York: Oxford University Press; 2001.
6. Holohan M, Müller R. Beyond humanism: telling response-able stories about significant otherness in human-chatbot relations. *Front Psychol* [Internet]. 2024 [acceso 11 ago 2025];15. DOI: 10.3389/fpsyg.2024.1357572
7. Brasil. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). *Diário Oficial da União* [Internet]. Brasília, 15 ago 2018 [acceso 11 ago 2025]; Disponível: <https://bit.ly/4mZaduf>
8. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union* [Internet]. Brussels, 27 abr 2016 [acceso 11 ago 2025]; Disponível: <https://bit.ly/3VVXT2F>
9. Tavory T. Regulating AI in mental health: ethics of care perspective. *JMIR Ment Health* [Internet]. 2024 [acceso 11 ago 2025];11. DOI: 10.2196/58493
10. Houghton C, Murphy K, Meehan B, Thomas J, Brooker D, Casey D. From screening to synthesis: using NVivo to enhance transparency in qualitative evidence synthesis. *J Clin Nurs* [Internet]. 2017 [acceso 11 ago 2025];26(5-6):873-81. DOI: 10.1111/jocn.13443
11. Bardin L. Análise de conteúdo. São Paulo: Edições 70; 2008.
12. Khurshid Z, De Brún A, Martin J, McAuliffe E. A systematic review and narrative synthesis: determinants of the effectiveness and sustainability of measurement-focused quality improvement trainings. *J Contin Educ Health Prof* [Internet]. 2021 [acceso 11 ago 2025];41(3):210-20. DOI: 10.1097/CEH.0000000000000331
13. Alfano L, Malcotti I, Ciliberti R. Psychotherapy, artificial intelligence and adolescents: ethical aspects. *J Prev Med Hyg* [Internet]. 2023 [acceso 11 ago 2025];64(4):438-42. Disponível: <https://bit.ly/4mU8AOk>
14. Grodniewicz JP, Hohol M. Waiting for a digital therapist: three challenges on the path to psychotherapy delivered by artificial intelligence. *Front Psychiatry* [Internet]. 2023 Jun [acceso 11 ago 2025];14. DOI: 10.3389/fpsyt.2023.1190084
15. Malouin-Lachance A, Capolupo J, Laplante C, Hudon A. Does the digital therapeutic alliance exist?: integrative review. *JMIR Ment Health* [Internet]. 2025 Feb 7 [acceso 11 ago 2025];12. DOI: 10.2196/69294
16. Khawaja Z, Bélisle-Pipon JC. Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. *Front Digit Health* [Internet]. 2023 [acceso 11 ago 2025];5. DOI: 10.3389/fdgth.2023.1278186
17. Dehbozorgi R, Zangeneh S, Khooshab E, Nia DH, Hanif HR, Samian P *et al*. The application of artificial intelligence in the field of mental health: a systematic review. *BMC Psychiatry* [Internet]. 2025 [acceso 11 ago 2025];25(1):132. DOI: 10.1186/s12888-025-06483-2
18. Gerke S, Minssen T, Cohen IG. Ethical and legal challenges of artificial intelligence-driven healthcare. In: Bohr A, Memarzadeh K. Artificial intelligence in healthcare. London: Academic Press; 2020. p. 295-336. DOI: 10.1016/B978-0-12-818438-7.00012-5
19. Jin KW, Li Q, Xie Y, Xiao G. Artificial intelligence in mental healthcare: an overview and future perspectives. *Br J Radiol* [Internet]. 2023 [acceso 11 ago 2025];96(1150):20230213. DOI: 10.1259/bjr.20230213

20. Bhatt S. Digital mental health: role of artificial intelligence in psychotherapy. Ann Neurosci [Internet]. 2024 [acceso 11 ago 2025];32(2):117-27. DOI: 10.1177/09727531231221612
21. Gutierrez G, Stephenson C, Eadie J, Asadpour K, Alavi N. Examining the role of AI technology in online mental healthcare: opportunities, challenges, and implications, a mixed-methods review. Front Psychiatry [Internet]. 2024 [acceso 11 ago 2025];15. DOI: 10.3389/fpsyt.2024.1356773

Sophia Ivantes Rodrigues – Estudiante de grado – sophiaivantes@hotmail.com

 0009-0008-9967-6049

Marília da Mata Silva – Estudiante de doctorado – marilia.mata1991@gmail.com

 0000-0001-7346-2386

Leonardo Pestillo de Oliveira – Doctor – leopestillo@gmail.com

 0000-0001-5278-0676

Lucas França Garcia – Doctor – lucasfgarcia@gmail.com

 0000-0002-5815-6150

Correspondencia

Lucas França Garcia – Programa de Pós-Graduação em Promoção da Saúde, Unicesumar. Av. Guedner, 1.610, bloco 7. CEP 87050-390. Maringá, Paraná, Brasil.

Contribución de los autores (CRediT)

Sophia Ivantes Rodrigues participo en el diseño del estudio, recopilación de datos y redacción del manuscrito. Marília da Mata Silva participo en el diseño del estudio. Lucas França Garcia participo en el diseño del estudio y redacción del manuscrito. Leonardo Pestillo de Oliveira participo en el redacción del manuscrito. Todos los autores participaron de la análisis de los datos y revisión crítica del manuscrito.

Disponibilidad de los datos: Todos los datos utilizados o generados en la investigación se describen y presentan íntegramente en el cuerpo del artículo.

Editores responsable: Dilza Teresinha Ambrós Ribeiro

Recibido: 25.2.2025

Revisado: 11.8.2025

Aprobado: 22.8.2025