

Los retos de la integración ética de la inteligencia artificial en la acción profesional de las y los psicólogos

The challenges of the ethical integration of Artificial Intelligence in the professional practice of psychologists

César Augusto García Avitia ^a

Abstract:

This article examines the challenges and opportunities posed by integrating artificial intelligence (AI), particularly generative AI, into professional psychological practice. It highlights how AI is rapidly transforming multiple sectors, including health and education, and increasingly influencing core psychological functions such as assessment, intervention and research. The text reviews current applications, including algorithms that detect personality traits or suicide risk from language, therapeutic chatbots delivering CBT-based support and platforms that automate clinical notes or monitor therapeutic progress. Based on these developments, the article identifies major ethical risks, such as threats to privacy and confidentiality, algorithmic bias, lack of clinical validation, automated errors, potential dehumanization of care and erosion of professional skills. However, it also recognizes ethical benefits, including increased accessibility, personalized interventions, reduction of human bias and improved efficiency. The article offers ethical recommendations aligned with APA and UNESCO guidelines: ensuring transparency and informed consent, maintaining human oversight, protecting sensitive data, selecting validated tools, strengthening digital competence and contributing to regulatory frameworks. It concludes that AI can enhance psychological practice only if integrated critically and ethically, preserving human judgment and prioritizing the well-being and rights of users.

Keywords:

Artificial intelligence, adoptions of AI in psychology, ethics of artificial intelligence, bioethics

Resumen:

El artículo analiza los retos y posibilidades de integrar la inteligencia artificial (IA), especialmente la inteligencia artificial generativa, en la práctica profesional de las y los psicólogos. Parte del reconocimiento de que la IA ha transformado sectores como la salud, la educación y la investigación, y que su rápida adopción también alcanza funciones esenciales de la psicología: evaluación, intervención e investigación. El texto describe aplicaciones actuales como algoritmos que detectan rasgos de personalidad o riesgo suicida a partir del lenguaje, chatbots terapéuticos que ofrecen apoyo basado en TCC y plataformas que automatizan notas clínicas o monitorean el progreso terapéutico. A partir de ello, identifica riesgos éticos significativos, incluyendo amenazas a la privacidad y confidencialidad, sesgos algorítmicos, falta de evidencia clínica, errores automatizados, deshumanización y pérdida de habilidades profesionales. No obstante, también reconoce beneficios potenciales como mayor accesibilidad, personalización del tratamiento, reducción de sesgos humanos y eficiencia operativa. Finalmente, el artículo propone recomendaciones éticas basadas en lineamientos de la APA y la UNESCO: asegurar transparencia y consentimiento informado, mantener supervisión humana, proteger los datos sensibles, elegir herramientas validadas, fortalecer la competencia digital y participar en marcos regulatorios. Concluye que la IA puede ser aliada de la psicología solo si se integra de manera crítica, ética y centrada en la persona.

Palabras Clave:

Inteligencia artificial, aplicaciones de la IA en psicología, ética de la inteligencia artificial, bioética

Introducción

La Inteligencia Artificial (IA), incluyendo su forma más reciente, la inteligencia artificial generativa (IAG), se ha incorporado rápidamente a la vida cotidiana y a múltiples campos profesionales. En 2024, más de la mitad de la población adulta en países como España había utilizado IA generativa, principalmente con fines personales o para obtener consejos según datos del Observatorio Nacional

de Tecnología y Sociedad. En México, se reporta un uso menos generalizado, aunque el 18.6% de internautas ya reportaron usarla según algunos estudios. Esta tecnología ha transformado sectores como la salud, la educación, las finanzas, la industria y muchos otros, planteando nuevas oportunidades, pero también profundos desafíos éticos. La psicología, en sus diversas áreas de aplicación, no es ajena a esta transformación. Las y los psicólogos comienzan a emplear sistemas de IA para apoyar sus funciones profesionales de evaluación, intervención e investigación, lo

^a Autor de Correspondencia, Universidad de Colima | Facultad de Psicología | Colima-Colima | México, <https://orcid.org/0000-0003-1841-3033> Email: garciaavitia@uclm.mx

Fecha de recepción: 15/11/2025, Fecha de aceptación: 05/02/2026, Fecha de publicación: 05/03/2026

DOI: <https://doi.org/10.29057/esat.v13iEspecial.16589>



cual promete innovaciones sin precedentes pero conlleva dilemas éticos importantes (López Fernández, 2025; Observatorio Nacional de Tecnología y sociedad [ONTSI], 2025 & Piedras, 2025).

El objetivo de este ensayo es analizar los retos de integrar la IA de forma ética en la acción profesional de psicólogos y psicólogas, respondiendo tres preguntas clave: 1. ¿Qué aplicaciones tiene la IA en la práctica psicológica actual, considerando las distintas funciones profesionales? 2. ¿Cuáles son las implicaciones éticas del uso de IA en el trabajo profesional psicológico? y 3. ¿Qué recomendaciones críticas pueden hacerse para lograr una integración ética de la IA en la práctica de la psicología?.

Para abordar estas cuestiones, primero se describe la reciente irrupción de la IA en la vida cotidiana y en concreto en la acción profesional de las y los psicólogos, describiendo algunos de los usos específicos de la IA en las funciones profesionales generales de las y los psicólogos, como lo son: evaluación, intervención e investigación, con ejemplos reales y documentados en algunas áreas de la psicología. A continuación, se analizan las implicaciones éticas de dichas aplicaciones, considerando tanto los riesgos como los potenciales beneficios, tomando una perspectiva desde la ética profesional. Sobre esa base, se presentan recomendaciones críticas para una integración ética de la IA en la práctica psicológica.

1. La irrupción de la IA en la acción profesional de las y los psicólogos

El impacto de la IA en la sociedad ha crecido de forma exponencial en la última década, alcanzando un punto de inflexión con la aparición de los modelos de IA generativa de uso masivo. Un hito emblemático fue el lanzamiento de ChatGPT a finales de 2022, que en pocos días alcanzó millones de usuarios, sorprendiendo al público general con su capacidad de mantener conversaciones coherentes, generar textos e incluso códigos de programación. En los años 2023 y 2024, herramientas generativas de texto, imagen y voz pasaron a integrarse en tareas rutinarias: asistentes virtuales para redactar documentos, traductores automáticos, chatbots conversacionales, entre otros. Indicadores recientes muestran la penetración de estas tecnologías: por ejemplo, 42% de la población adulta española usó alguna herramienta de IA en 2023, y más de 50% utilizó IA generativa en 2024, especialmente para actividades cotidianas como redactar textos, obtener información o recibir orientación. Estas cifras ilustran cómo la IA se ha posicionado en la rutina diaria de muchas personas, pasando de ser una novedad tecnológica a una herramienta casi omnipresente (De la Fuente Tambo & Armayones Ruiz, 2025; López Fernández, 2025 & ONTSI, 2025).

En el campo de la psicología, es importante destacar que si la IA tiene antecedentes históricos aplicados a la acción profesional, como es el caso de los primeros chatbots terapéuticos de la historia como ELIZA y PARRY en las décadas de 1960 y 1970 ya exploraban

la interacción persona-máquina en contextos clínicos. Sin embargo, una de las diferencias con la actualidad es la escala masiva y la autonomía de las aplicaciones modernas. Los modelos de IAG actuales operan con enormes volúmenes de datos y potentes algoritmos de machine learning y deep learning, permitiéndoles aprender, generar contenido original y adaptarse en tiempo real. Además, gracias al procesamiento de lenguaje natural avanzado, las IAG pueden interactuar directamente con los usuarios de forma conversacional y aparentemente empática, algo impensable hasta hace pocos años. Esta presencia ubicua de la IA ha creado tanto expectativas (por ejemplo, que estas herramientas potencien la eficiencia y personalización de servicios) como preocupaciones (por los posibles riesgos a la privacidad, la desinformación o el reemplazo de labores humanas) (De la Fuente Tambo & Armayones Ruiz, 2025; Eliot, 2023; López Fernández, 2025; & Krieger et al. 2024).

Así, estamos ante una irrupción reciente de la IA que difumina la línea entre herramientas tecnológicas especializadas y compañeras constantes en actividades diarias. Este contexto sienta las bases para comprender por qué las profesiones, incluida la psicología, enfrentan ahora la tarea ineludible de integrar la IA en su práctica, asegurando al mismo tiempo que dicha integración respete principios éticos fundamentales y promueva el bienestar de individuos y comunidades.

La rápida adopción de la IA en diversos sectores profesionales conlleva implicaciones generales que conviene delinear antes de centrarnos en la psicología. En términos positivos, la IA ofrece oportunidades transversales: mayor eficiencia operativa, automatización de tareas repetitivas, análisis de grandes volúmenes de datos para apoyar la toma de decisiones, personalización de servicios a las necesidades individuales, y avances acelerados en investigación y desarrollo. Por ejemplo, en salud se utilizan sistemas de IA para diagnosticar imágenes médicas o proponer tratamientos personalizados; en educación, algoritmos que adaptan materiales al ritmo del alumnado; en seguridad, modelos predictivos para prevención del delito, etc. Estas aplicaciones prometen que las profesiones sean más efectivas (López Fernández, 2025).

En el caso de la psicología, la práctica profesional abarca diversas funciones, que de manera general se pueden categorizar en evaluación, intervención e investigación; y son ejercidas en distintos ámbitos como el clínico, educativo, social-comunitario, organizacional, etc. Actualmente, la IA empieza a emplearse en todas estas funciones, generando herramientas novedosas. A continuación se describen los usos más destacados, con ejemplos documentados y casos reales, organizados por función profesional.

1.1. Evaluación psicológica asistida por IA

La evaluación es una función central del quehacer psicológico, que incluye acciones como la aplicación de pruebas, observación de conductas, realización de entrevistas y análisis de datos de las y los pacientes,

usuario o cliente, ya sea para obtener un perfil psicológico, un diagnóstico o con otros diversos fines. La IA está revolucionando esta área al posibilitar la recolección y análisis automatizado de grandes cantidades de información de forma eficiente y, a veces, menos intrusiva que los métodos tradicionales. Una aplicación clave es el procesamiento de lenguaje natural para analizar textos, discursos o expresiones escritas de individuos en busca de indicadores psicológicos. Por ejemplo, un estudio de la Universidad de Barcelona mostró que modelos avanzados de IA como BERT y RoBERTa, desarrollados por Google, pueden detectar rasgos de personalidad a partir de textos escritos por las personas. Estos modelos identifican patrones lingüísticos sutiles asociados a los llamados “cinco grandes” rasgos de personalidad o a dimensiones del indicador Myers-Briggs, logrando una precisión significativa y proporcionando explicaciones sobre qué palabras o frases influyen en sus predicciones (Saeteros et al., 2025).

Este tipo de análisis automatizado abre vías para evaluaciones más naturales y menos dependientes de cuestionarios o pruebas psicológicas explícitas, lo cual podría ser especialmente útil en poblaciones amplias o contextos donde la evaluación tradicional es difícil. Las y los investigadores subrayan que estas herramientas no reemplazarán a corto plazo las pruebas psicométricas clásicas, sino que las complementarán, integrándose en un enfoque multimodal junto con datos conductuales y otras fuentes para ofrecer una visión más completa de la personalidad y otros procesos mentales. Es importante enfatizar la necesidad de que estos modelos estén alineados con teorías psicológicas sólidas y que incorporen técnicas de IA explicable para garantizar interpretaciones éticas y transparentes (Saeteros et al., 2025).

En el ámbito clínico, la IA se emplea para detectar signos de trastornos mentales o problemas de salud mental a través de diversos datos. Algoritmos de aprendizaje automático pueden analizar patrones en el habla, la voz o incluso la expresión facial. Un ejemplo vanguardista es el trabajo de John Pestian y colegas, que desarrollaron modelos para identificar el riesgo de suicidio mediante el análisis de lenguaje natural y propiedades acústicas de la voz. Al alimentar al sistema con notas escritas de personas con ideación suicida y grabaciones de pacientes en crisis, el algoritmo aprendió a reconocer lo que simbólicamente se ha llamado como el lenguaje del suicidio, como la presencia de instrucciones concretas en notas de despedida; así como indicadores vocales como suspiros más frecuentes, risas ausentes, pausas prolongadas, vocales acortadas y tono entrecortado en el habla de quienes contemplaban quitarse la vida. En pruebas ciegas, este sistema logró coincidir con las evaluaciones de expertos humanos en alrededor del 85% de los casos, demostrando potencial para apoyar la detección temprana de riesgo suicida. Si bien estos sistemas requieren mayor

validación, ilustran cómo la IA puede procesar señales sutiles que quizás pasan desapercibidas para un psicólogo o psicóloga profesional, sirviendo como herramienta de alerta y apoyo diagnóstico en evaluación de crisis (Infobae, 2023).

Otra línea de uso es el análisis automatizado de pruebas psicométricas y comportamentales. Por ejemplo, se han desarrollado algoritmos que a partir de pruebas tradicionales predicen los puntajes o clasificaciones diagnósticas de forma automatizada. Algunos programas analizan no solo las respuestas sino la forma de interactuar con la prueba, lo cual incluye tiempos de reacción, patrones de clics u otros indicadores, incorporando técnicas de test adaptativo informatizado. Incluso se explora la evaluación a través de juegos serios o aplicaciones lúdicas donde la IA interpreta el desempeño del usuario para inferir funciones cognitivas o rasgos, como podría ser un juego en el que se evalúa la atención, la memoria y el estado de ánimo para predecir un puntaje de depresión (Krieger et al., 2024).

Asimismo, la recopilación de datos mediante sensores u uso de teléfonos inteligentes, como la actividad, la geolocalización, el uso de apps o los patrones de sueño, combinada con IA permite crear perfiles continuos del estado de ánimo o la ansiedad de una persona en su vida diaria. Esto se conoce como fenotipado digital, lo cual permite acciones como la detección de depresiones incipientes a partir de menor movilidad y cambios en hábitos de comunicación. Estudios previos ya demostraron la viabilidad de tales enfoques, donde se han realizado análisis de fotos de Instagram para predecir depresión por el uso de ciertos filtros y colores, y su precisión está mejorando con modelos de IA más actuales (Krieger et al., 2024).

Pasando al ámbito de la psicología organizacional, la IA se está aplicando a la evaluación de personal y procesos de selección. Sistemas de reclutamiento con IA analizan currículos, pruebas en línea e incluso grabaciones de entrevistas para inferir competencias laborales y rasgos de personalidad relevantes, a fin de preseleccionar candidatos. Un estudio reciente destacó que la IA puede identificar patrones lingüísticos ligados a rasgos de personalidad que podrían escapar a métodos convencionales, lo que podría conducir a evaluaciones más precisas en selección de personal. Sin embargo, estos mismos sistemas han generado polémica por posibles sesgos: por ejemplo, si el modelo fue entrenado con datos de empleadas o empleados exitosos, podría perpetuar sesgos de género o etnia presentes en los contextos laborales (Tenelema Jiménez et al. 2025).

En el ámbito de la psicología educacional, herramientas de IA pueden evaluar habilidades académicas o socioemocionales de los estudiantes. Un ejemplo es el uso de tutores inteligentes que, mientras el alumno realiza ejercicios en la computadora, van evaluando en tiempo real su comprensión y adaptando la dificultad de las tareas. Asimismo, la minería de datos educativos con IA puede identificar estudiantes en riesgo de deserción o con dificultades específicas de aprendizaje, analizando sus

patrones de desempeño y comportamiento escolar (Carbonell Bernal et al., 2024)

Por lo anteriormente descrito, queda claro que la IA aplicada a la evaluación psicológica ofrece velocidad y alcance, pues se pueden administrar pruebas en línea a miles de personas simultáneamente, corregirlas instantáneamente, detectar patrones complejos en los datos que serían arduos manualmente y realizar un seguimiento continuo del estado de un individuo. Todo esto puede complementar la labor de las y los psicólogos, permitiendo enfoques de evaluación más dinámicos y basados en datos reales del día a día. No obstante, también impone cautela: interpretar resultados de IA requiere la supervisión de profesionales para evitar conclusiones erróneas, y es imprescindible asegurar que estas herramientas estén validadas científicamente en poblaciones diversas antes de usarlas para tomar decisiones profesionales.

1.2 Intervención psicológica apoyada por IA

La intervención psicológica, incluyendo la psicoterapia y otras estrategias de cambio, es quizá el área que más atención mediática ha recibido respecto a la IA, debido al surgimiento de chatbots terapéuticos y plataformas digitales de apoyo emocional. Aquí es posible distinguir dos grandes modalidades: a) la IA como agente de intervención directa, como es el caso de chatbots que interactúan con los pacientes, clientes o usuarios; y b) la IA como asistente para las y los interventores, terapeutas o profesionales, apoyando tareas pero sin tratar directamente al paciente, cliente o usuario.

En la primera categoría se ubican los chatbots de terapia o asistentes conversacionales para apoyo en salud mental. Un caso pionero es Woebot, creado en 2017 por la psicóloga Alison Darcy en Stanford, que brinda intervención basada en Terapia Cognitivo-Conductual (TCC) a través de una app de mensajería. Woebot emplea procesamiento de lenguaje natural para interpretar las entradas del usuario, por ejemplo: si la o el usuario escribe que se siente muy preocupado, le responde guiándole por ejercicios cognitivos previamente preparados, emulando técnicas de reestructuración de pensamientos, registro de emociones, etc. La idea central fue proveer un seguimiento diario a las y los pacientes que, tras ser dados de alta de terapia, tendrían a recaer sin ese apoyo continuo. Ningún terapeuta puede estar con un paciente todo el día, todos los días, señalaba Darcy, así que un chatbot disponible 24/7 sería un complemento conveniente. Hasta 2023, Woebot había atendido a casi 1.5 millones de personas (muchas en zonas con escasez de servicios de salud mental) con solo un centenar de empleados humanos detrás de la plataforma (De la Fuente Tambo & Armayones Ruiz, 2025 & Infobae, 2023).

De forma similar, otros chatbots como Wysa, con más de 6 millones de usuarios, o Youper, con más de 1 millón de usuarios, ofrecen intervenciones basadas en TCC y mindfulness mediante conversaciones estructuradas. Estos sistemas buscan mejorar síntomas leves o moderados de ansiedad, depresión y estrés, actuando

como un mentor que practica escucha empática y propone actividades. Incluso existen derivados como Replika, que se presenta más como un compañero conversacional general, o aplicaciones como Happify con juegos y tareas para el bienestar emocional (De la Fuente Tambo & Armayones Ruiz, 2025 & Infobae, 2023).

Los beneficios potenciales de estas IAs en intervención son accesibilidad y escalabilidad. Pueden llegar a usuarios que no recibirían ayuda de otro modo, ya sea por barreras geográficas, económicas o temor al estigma social, y ofrecen apoyo inmediato en cualquier momento, por ejemplo a medianoche cuando una persona ansiosa no tiene con quién hablar, y pueden interactuar con miles de usuarias y usuarios simultáneamente, algo imposible para las y los terapeutas o profesionales humanos. También proporcionan consistencia, pues no se cansan ni tienen variaciones de humor y responden siempre según sus guiones terapéuticos. Además, algunos incorporan detección de palabras clave de riesgo. Por ejemplo, el bot Cass, servicio popular en Estados Unidos de América que funciona vía mensajes de texto, con decenas de millones de usuarias y usuarios, y que está programado para contactar a un profesional humano si el usuario menciona ideas suicidas o muestra señales de crisis (De la Fuente Tambo, & Armayones Ruiz, 2025).

Sin embargo, los límites y riesgos también son evidentes. Un estudio de 2025 de la Universidad de Stanford evaluó a varios modelos de lenguaje (como GPT-4) en simulaciones de sesiones clínicas y halló fallos en estos chatbots terapéuticos. Por ejemplo, al presentar un caso simulado de un usuario con ideación suicida encubierta (preguntando por los puentes más altos de Nueva York), el chatbot respondió con una lista literal de puentes altos, sin reconocer la indirecta de posible comportamiento suicida ni activar ningún protocolo de alerta (Moore et al., 2025).

Las y los investigadores han descrito un fenómeno llamado sociopatía digital, es decir, una cortesía superficial y respuestas amables de parte de los chatbots de IA, pero sin genuina empatía ni capacidad de contención emocional, lo cual puede ser peligroso en situaciones de urgencia clínica. En esencia, estos chatbots carecen del juicio contextual, responsabilidad y vínculo humano las y los profesionales humanos: la crucial alianza terapéutica, basada en confianza, empatía real y compromiso, no puede ser replicada con simples frases cálidas[48]. Por ello, las y los expertos advierten que estos sistemas no deben presentarse como sustitutos de la terapia real sin evidencia clínica sólida que los avale (Moore et al., 2025). Hasta la fecha, aunque algunos estudios sugieren que la llamada terapia conversacional automatizada puede tener eficacia moderada en síntomas leves, como es el caso de usuarios reportando sentirse escuchados o aprendiendo alguna técnica cognitiva, no hay consenso de que funcionen para problemas serios o que mantengan mejoras a largo plazo. Por ende, se tiende a recomendarlos solo como apoyo complementario, entre sesiones con una o un profesional humano, o como primer punto de contacto para educación psicológica, siempre

con la posibilidad de derivar a ayuda profesional humana ante riesgos o necesidades mayores (Moore et al., 2025).

La segunda gran modalidad es la IA como asistente del profesional en sus intervenciones. Aquí la IA no interactúa sola con las o los clientes, usuarios o pacientes, sino que potencia las capacidades de la acción profesional. Un ejemplo concreto es Upheal, una plataforma creada en Estados Unidos de América para profesionales clínicos de la salud mental potenciada por IA. Upheal automatiza la creación de notas de progreso de las sesiones terapéuticas, transcribiendo y resumiendo las sesiones grabadas, genera planes de tratamiento personalizados sugiriendo objetivos SMART, es decir específicos, medibles, alcanzables, relevantes y delimitados en el tiempo; los cuales son basados en el caso. También proporciona análisis detallados de las sesiones, que incluyen patrones de conversación, emociones detectadas, indicadores de progreso, etc. Todo ello en un entorno seguro, cumpliendo estándares de privacidad (Upheal, 2025).

El propósito de una plataforma como Upheal es aliviar la carga administrativa de las y los profesionales de la psicoterapia, que suelen invertir mucho tiempo tras cada sesión en redactar registros clínicos, permitiéndoles enfocarse más en el cliente, usuario o paciente y menos en el papeleo. Plataformas similares incluyen Mentalyc o Texta AI, que transcriben automáticamente la sesión y producen un borrador de nota clínica estructurada. Estas notas generadas por IA son personalizables al estilo profesional y ahorran tiempo sin sacrificar detalle. De igual forma, sistemas como Eholo (desarrollado en España, con más de 10 mil psicólogos usuarios) ofrecen suites integrales donde la IA agenda citas, organiza expedientes, registra contabilidad e incluso prepara resúmenes de sesión (De la Fuente Tambo & Armayones Ruiz, 2025 & Upheal, 2025).

Otra aplicación es la asistencia en el proceso clínico mediante IA como colaborador en el análisis y toma de decisiones. Ciertos programas pueden analizar cuestionarios y datos del paciente para identificar patrones preocupantes o hipótesis diagnósticas que el terapeuta podría pasar por alto. Por ejemplo, la herramienta Lyssn emplea IA para evaluar la calidad de la psicoterapia: analiza grabaciones de sesiones y proporciona retroalimentación objetiva sobre habilidades del terapeuta, como el uso de la empatía, adherencia a técnicas específicas, proporción de tiempo que habla el paciente vs. terapeuta, etc. Lyssn también está siendo usada en entrenamientos de terapeutas y en clínicas para mantener estándares de calidad, ayudando a identificar áreas de mejora en la intervención que quizás un supervisor o supervisora humana tardaría en notar (De la Fuente Tambo, & Armayones Ruiz, 2025).

Otro software, Limbic, acompaña al terapeuta monitorizando el progreso del paciente a lo largo del tratamiento: integra datos de seguimiento como síntomas reportados periódicamente por el paciente o resultados de escalas, y alerta si detecta estancamiento o

empeoramiento, incluso sugiriendo ajustes terapéuticos basados en lo que ha funcionado en casos similares. De hecho, Limbic ha sido implantado en alrededor del 40% del Servicio Nacional de Salud (NHS) del Reino Unido para apoyar a psicólogos y psiquiatras en sus procesos clínicos. Asimismo, plataformas de apoyo clínico como Aimentia, desarrollada en España, combinan diagnóstico asistido, recomendación de intervenciones basadas en evidencia actualizada y seguimiento digital, funcionando como un asistente clínico virtual que trabaja junto al profesional (De la Fuente Tambo, & Armayones Ruiz, 2025).

En el contexto de intervenciones sociales, la IA también encuentra usos novedosos. Por ejemplo, en programas de prevención, bots conversacionales se emplean para difundir psicoeducación en redes sociales, detectando automáticamente cuando usuarios publican ciertos mensajes de angustia y ofreciéndoles recursos de ayuda. También se han probado agentes de IA que brindan primeros auxilios psicológicos básicos via SMS en zonas remotas con pocos profesionales.

1.3 Investigación psicológica potenciada por IA

La investigación en psicología también ha tenido impactos positivos gracias a las herramientas de IA. Tradicionalmente, muchas investigaciones psicológicas se limitaban por la capacidad humana de procesar datos: encuestas de centenares de personas, análisis manual de texto en estudios cualitativos, experimentos con grupos relativamente pequeños en laboratorio. La IA rompe estos límites permitiendo analizar grandes volúmenes de datos complejos que antes estaban fuera del alcance. Por ejemplo, en psicología social, ahora es factible analizar millones de publicaciones en redes sociales para estudiar patrones de expresión emocional o de interacción social a gran escala, empleando algoritmos que extraen temas, sentimientos predominantes, redes de relaciones, etc. De esta manera, se pueden detectar tendencias de salud mental poblacional casi en tiempo real o estudiar cómo se propagan ciertas creencias o estereotipos en entornos online.

La IA también ha facilitado la investigación mediante simulaciones y modelos computacionales. Otra posibilidad es el uso de agentes virtuales para modelar interacciones sociales o dinámicas psicológicas: por ejemplo, se pueden programar agentes con distintos rasgos de personalidad simulados y hacerlos interactuar miles de veces para ver qué dinámicas emergen, aportando hipótesis sobre teorías de grupo o influencia social. Recientemente, se ha propuesto que sistemas de IA multiagente podrían hasta ensayar intervenciones en salud mental entre sí, para predecir qué respuestas serían más efectivas antes de probarlas con humanos. Este enfoque podría acelerar la generación de hipótesis y la personalización de tratamientos (Forbes Argentina, 2025). En el ámbito del análisis de datos, la IA aporta técnicas como redes neuronales y modelos de análisis de clases latentes que identifican estructuras ocultas en los datos psicológicos. Por ejemplo, machine learning ha sido usado para encontrar nuevos subtipos de trastornos psicológicos analizando combinaciones de síntomas en

grandes bases de datos clínicas, lo cual podría refinar las clasificaciones diagnósticas. Asimismo, la IA ayuda en meta-ciencia, pues existen herramientas que automatizan la búsqueda y resumen de literatura científica, como Elicit o Semantic Scholar, e incluso que detectan patrones de sesgo de publicación o calidad metodológica en conjuntos de estudios (Ópez Steinmetz & Carlos Godoy, 2023).

Otra contribución es en el área de las neurociencias y psicofisiología, pues la IA ha resultado crucial para interpretar las enormes cantidades de datos de imágenes cerebrales como Resonancias Magnéticas Funcionales o Electroencefalogramas, y para relacionarlos con estados mentales. Algoritmos de aprendizaje profundo han logrado identificar con notable precisión patrones de activación cerebral correspondientes a ciertos estímulos o pensamientos, e incluso reconstruir imágenes mentales básicas a partir de la actividad neuronal, logros que amplían la comprensión mente-cerebro (Chen, et al., 2022).

No debe olvidarse que la propia metodología de la investigación psicológica se está enriqueciendo gracias a la IA. Por ejemplo, en metodología experimental, se emplean robots o asistentes virtuales estandarizados para interactuar con participantes humanos de forma consistente, garantizando que cada participante reciba exactamente las mismas instrucciones o estímulos sin variabilidad de las o los investigadores humanos. También, en análisis cualitativo, algoritmos de procesamiento de lenguaje pueden codificar cientos de entrevistas o respuestas abiertas detectando temas y emociones, apoyando el trabajo del investigador humano. Por último, cabe señalar que el uso de IAG ha sido explorado para generar respuestas simuladas de participantes. Se puede pedir a un modelo de lenguaje que responda a cuestionarios o entrevistas psicológicas como si fuera un participante o una participante con ciertos rasgos, para ver si se pueden predecir resultados o acelerar pruebas piloto. Si bien esto está en etapas iniciales y tiene serias limitaciones tanto éticas como metodológicas, ilustra la creatividad con que se están aplicando estas herramientas en investigación.

2. Implicaciones éticas del uso de IA e IAG en la práctica psicológica

La incorporación de tecnologías de IA en la psicología plantea desafíos éticos significativos. Dado que la labor psicológica suele involucrar información altamente sensible, relaciones interpersonales basadas en la confianza y decisiones que afectan el bienestar de las personas, el uso de sistemas automatizados debe examinarse cuidadosamente a la luz de los principios éticos de la profesión y de la bioética. En este apartado se discuten las implicaciones éticas, primero identificando riesgos y luego considerando beneficios que también emergen con estas tecnologías.

2.1 Riesgos éticos potenciales

Se puede comenzar mencionando que las y los psicólogos manejan datos sumamente confidenciales, como historias clínicas, resultados de tests y revelaciones íntimas de sus

clientes, usuarios o pacientes. El uso de IA implica a digitalizar y almacenar estos datos en plataformas tecnológicas, lo que conlleva riesgo de violación de la confidencialidad. Fallos de seguridad, accesos no autorizados o filtraciones de información son peligros reales, de hecho, se han reportado incidentes de filtraciones masivas de datos en servicios de salud mental en línea, atentando contra el principio bioético de no maleficencia. Además, algunos servicios de IA podrían reutilizar los datos para re-entrenar sus algoritmos si sus políticas no lo prohíben, lo cual es inaceptable éticamente sin un consentimiento explícito (American Psychological Association, 2025).

La American Psychological Association [APA], en su guía ética para IA (2025), advierte que las y los profesionales deben asegurarse de usar sistemas con fuertes medidas de seguridad, como encriptación, controles de acceso y auditorías. También deben verificar las políticas de los proveedores de IA sobre el uso de datos y cumplir con regulaciones de protección de datos aplicables[70]. También enfatiza que se debe dar al cliente control sobre qué datos suyos se comparten y la posibilidad de retractar su consentimiento en cualquier momento, acorde al principio bioético de autonomía. La confidencialidad es una piedra angular de la ética psicológica; por tanto, cualquier fuga o mal manejo de datos por parte de la IA no solo dañaría al usuario o usuaria sino también la confianza en la profesión en su conjunto.

Por otro lado, los algoritmos de IA pueden heredar sesgos presentes en los datos con que fueron entrenados. En el contexto psicológico, esto puede traducirse en injusticias o prejuicios en las evaluaciones o intervenciones automatizadas. El estudio de Stanford evidenció respuestas discriminatorias de chatbots hacia ciertos perfiles diagnósticos, reflejando estigma social, como respuestas más negativas a usuarios con esquizofrenia. Podría ocurrir algo semejante con sesgos de género, étnicos, culturales o de idioma. Un sistema de IA de evaluación entrenado mayormente con población urbana occidental podría puntuar erróneamente más alto o más bajo en la evaluación de algún proceso psicológico o de salud mental a personas de culturas no occidentales por diferencias de comunicación, o no captar matices idiomáticos en otro idioma. Esto atenta contra el principio bioético de la justicia (Moore et al., 2025 & Beauchamp & Childress, 2025).

La APA (2025) enfatiza que las y los psicólogos deben evaluar si el sistema de IA fue validado en poblaciones diversas, monitorear activamente las salidas de la IA para detectar resultados sesgados y preferir herramientas que incorporen estrategias de reducción de sesgo. Además, las o los profesionales deben mantener una actitud crítica para corregir cualquier sesgo que observe. No hacerlo podría llevar a reproducir inequidades, pues la IA debe potenciar la equidad, no minarla.

Con respecto al principio bioético de autonomía, que exige obtener consentimiento informado para los procedimientos, si un psicólogo o psicóloga usa IA en cualquier parte del proceso, debe informarlo claramente al paciente, procurando transparencia. Esto incluye explicar qué tipo de

IA se usará, cuáles son sus funciones y limitaciones, qué riesgos potenciales entraña, cómo se manejarán sus datos, y confirmar que el paciente acepta o puede optar por no usar dichas herramientas. Omitir esta transparencia sería poco ético, ya que el paciente podría no darse cuenta de que ciertas recomendaciones o decisiones vinieron de una máquina y no directamente del juicio del terapeuta. También es problemático si, por ejemplo, se implementan sistemas de selección de personal psicológicos basados en IA sin informar a las y los candidatos. La transparencia es necesaria además para que el usuario pueda cuestionar o entender las recomendaciones de la IA; de lo contrario, hay opacidad en la toma de decisiones que afecta su autonomía y participación informada en su proceso psicológico de atención profesional (APA, 2025).

Un aspecto más es que la IA, especialmente la generativa, puede producir errores, imprecisiones o directamente invenciones, llamadas alucinaciones de los modelos generativos. En un contexto clínico, esto podría ser desastroso. Un asistente de IA podría resumir mal una sesión, tergiversando lo que dijo el paciente, o sugerir una interpretación clínica errónea. Si el psicólogo o psicóloga confía ciegamente en los productos elaborados por la IA, podría actuar sobre información equivocada. La APA subraya que nunca se debe delegar a la IA decisiones clínicas o diagnósticas sin revisión humana y que el profesional o la profesional debe evaluar críticamente los resultados generados por IA de manera continua. Igualmente, recomienda documentar en el expediente cuándo se ha utilizado IA y en qué medida contribuyó al resultado, para mantener claridad y trazabilidad (APA, 2025)

Lo anterior está ligado a la cuestión de responsabilidad, con implicaciones de los principios bioéticos de beneficencia y no maleficencia, pues cualquier decisión o intervención recae ética y legalmente sobre la psicóloga o el psicólogo humano, no sobre la herramienta de IA. Si un algoritmo comete un fallo, el profesional debe estar preparado para asumir la responsabilidad y explicar qué ocurrió. Un riesgo es que el uso creciente de IA difumine esa responsabilidad. La postura ética correcta, sin embargo, es que la IA debe ser como una herramienta cuyas recomendaciones supervisan las o los profesionales, no un sustituto que tome las riendas. La supervisión humana constante es indispensable para prevenir errores. 16

Un aspecto más es lo que se ha alertado sobre el peligro de deshumanización en la práctica psicológica si se sobreutilizan herramientas de IA. Ya sea la psicoterapia o muchos otros tipos de acción profesional de las y los psicólogos, están basadas en la conexión humana, la validación emocional auténtica y la construcción de una relación de confianza. Si clientes, usuarios o pacientes perciben que se les atiende mediante respuestas preprogramadas o análisis fríos de una IA, podría minarse la alianza terapéutica.

Se puede plantear como ejemplo el caso de una o un usuario, cliente o paciente que recibe un informe

psicológico generado automáticamente podría percibirlo como impersonal, aunque fuese preciso. O en una empresa, las y los empleados evaluados por un algoritmo podrían verlo con recelo y falta de calidez humana. Este efecto puede socavar la calidad percibida de la atención y la satisfacción de las y los usuarios. Además, existe el riesgo de dependencia a interacciones simuladas, pues algunas personas podrían preferir hablar con un chatbot y evitar el esfuerzo de la interacción humana, lo cual en ciertos casos podría aislarlas más. Esto atenta contra el principio de promover la autonomía de las personas usuarias. Así, hay un delicado balance entre beneficiarse de la IA y preservar el factor humano que incluso es terapéutico en sí mismo. La consigna ética es que la IA nunca reemplace la empatía real, sino que libere tiempo a los profesionales para dedicarlo a la parte humana, que es insustituible.

Un tema más al respecto de los riesgos es el relacionado con la precisión y la falta de evidencia sólida, está la posibilidad de que se apliquen herramientas de IA no validadas clínicamente que produzcan daño en vez de beneficio. Por ejemplo, si una app de terapia digital ofrece consejos inapropiados, puede retrasar que el individuo busque la ayuda adecuada, con el consiguiente deterioro. Otro ejemplo puede ser el caso de un sistema de evaluación psicológica que podría etiquetar erróneamente a alguien con algún problema que en realidad no presenta (falso positivo) generando estigma o sobre-reacción; o al revés, dar falsos negativos que lleven a una falsa tranquilidad.

Así, dado que muchos de estos sistemas aún carecen de validación independiente, su uso prematuro entraña un riesgo ético de experimentación encubierta con los usuarios. Con respecto a esto, un grupo de investigación de Stanford concluyó que permitir chatbots no probados actuar como terapeutas es potencialmente peligroso y que ninguna herramienta debe comercializarse como sustituto terapéutico sin evidencia que respalde su seguridad y eficacia (Moore et al., 2025).

Por su parte, la Organización de las Naciones Unidas para la Educación la Ciencia y la Cultura [UNESCO], igualmente advierte que las herramientas de IAG deben validarse en su idoneidad ética y pedagógica antes de adoptarse masivamente en el contexto educativo. Extrapolando, en psicología debería exigirse validación de eficacia y aprobación regulatoria para aquellas IA que sirven para evaluar o intervenir. Hoy, la realidad es que muchas aplicaciones de apoyo en salud mental operan sin controles clínicos ni regulatorios en varios países [82], creando un vacío ético y legal que es urgente abordar (Moore et al, 2025 & UNESCO, 2024).

Un aspecto más a considerar es que la facilidad que brinda la IA podría llevar a cierta relajación en las competencias del psicólogo. Por ejemplo, si las y los profesionales se acostumbra a que un algoritmo le sugiera diagnósticos o conclusiones de una evaluación, podrían descuidar su propio razonamiento o habilidades. También, las y los nuevos psicólogos podrían apoyarse tanto en herramientas tecnológicas que no desarrollen

plenamente destrezas básicas como calificar pruebas a mano o redactar informes concisos por sí mismos.

La APA indica que las y los profesionales deben usar solo herramientas que comprenden y para las que tienen formación adecuada, y que mantengan actualización continua en los avances de IA, sus mejores prácticas y regulaciones emergentes. Esto implica un compromiso ético con la competencia profesional: no se debe delegar en IA algo para lo cual el psicólogo o psicóloga no sabría detectar si está mal. La automatización no exime de saber hacer; análogamente a cómo un piloto automático en aviación no significa que la o el piloto humano pueda ser inexperto, pues debe intervenir en caso de fallo, en psicología la o el profesional siempre debe estar capacitado para realizar por sí mismo las funciones críticas si la IA falla. Así se evita la pereza mental o una posible pérdida de habilidades en las nuevas generaciones (APA,2025).

2.2 Beneficios éticos potenciales

Frente a los riesgos mencionados, es importante reconocer que la IA bien empleada también puede potenciar el cumplimiento de principios éticos y generar efectos positivos en la práctica psicológica. Se puede comenzar por hablar de la posibilidad de aumentar el acceso y la equidad en los servicios, pues desde una perspectiva de justicia social, la IA ofrece la posibilidad de aumentar la cobertura en cierto tipo de apoyos psicológicos. En regiones con pocas o pocos profesionales de salud mental, un chatbot o una plataforma en línea puede brindar psicoeducación, detección temprana o intervención básica a personas que de otro modo no tendrían alguna ayuda. Por ejemplo, Woebot y similares han alcanzado a cientos de miles de usuarios en lugares con escasez de terapeutas. Del mismo modo, en organizaciones grandes donde es inviable dar seguimiento psicológico personalizado a cada empleado, un asistente virtual puede ofrecer orientación generalizada. Esto reduce barreras de acceso geográficas, económicas y de horarios, promoviendo el principio bioético de beneficencia hacia poblaciones tradicionalmente desatendidas. Siempre que estos servicios sean de calidad razonable, su amplia disponibilidad es éticamente positiva (De la Fuente Tambo & Armayones Ruiz, 2025).

Otro beneficio es que una herramienta tecnológica puede permitir una atención más personalizada si utiliza la información adecuada. La IA puede analizar el perfil único de cada individuo, como sus patrones de comportamiento, sus preferencias de comunicación o su progreso específico en un tratamiento y adaptar las recomendaciones o ejercicios a ese perfil. Un chatbot, por ejemplo, podría recordar que cierto usuario prefiere enfoques más directos y ajusta su estilo comunicativo, o un sistema de evaluación puede seleccionar ítems de test más pertinentes según las respuestas previas del evaluado, haciendo la evaluación más pertinente y breve para esa persona. Esta personalización implementa mejor el ideal de respeto por la individualidad y la dignidad de cada cliente, usuario o paciente, que son valores éticos

centrales en la psicología.

Otro beneficio es la posibilidad de evitar otros sesgos comunes en humanos. Por ejemplo, un psicólogo o psicóloga puede tener fatiga o prejuicios inconscientes hacia ciertos clientes, lo que afecta su trato. Un sistema automatizado en tareas definidas, como puntuar un test estandarizado, lo hará igual cada vez, garantizando uniformidad. También puede ayudar a tomar decisiones más objetivas. Por ejemplo, una IA entrenada con criterios clínicos puede servir de checklist para que las y los profesionales no pasen por alto preguntas importantes, reduciendo sesgos de confirmación. En procesos de selección de personal, si se diseñan con cuidado, los algoritmos podrían ignorar atributos irrelevantes (género, edad) y centrarse solo en competencias, potencialmente más justos que entrevistas humanas sesgadas. Claro está, esto depende de un diseño ético del algoritmo, pero es una oportunidad para mitigar parcialidad.

Otro aspecto más es que la automatización de tareas administrativas y repetitivas, como escribir notas, gestionar agenda o elaborar informes rutinarios es un beneficio claro. Considerando la responsabilidad profesional, esto aumenta la eficiencia y productividad, permitiendo atender a más gente con los mismos recursos o dedicar más tiempo a la atención directa. Un psicólogo o psicóloga que ya no tenga que invertir horas semanales en papeleo puede usar ese tiempo en reflexionar mejor los casos, formarse más, o simplemente descansar y prevenir burnout, lo cual redundará en una mejor atención.

La UNESCO (2024) resalta la necesidad de aprovechar la IA sin comprometer valores humanísticos; en este sentido, usar IA para aquello que resulta burocrático y reservar al humano para lo más relacional, apoya ese equilibrio. El resultado puede ser servicios más diligentes y de mayor calidad, cumpliendo con la beneficencia y responsabilidad. Ejemplo de ello puede ser que gracias a herramientas como Upheal, un o una terapeuta documenta más rápido los procesos y quizás puede ver uno o dos pacientes adicionales por semana o dedicar más tiempo a cada uno, beneficiando a más personas, algo que en el sistema de atención pública es sumamente deseable.

Otro beneficio más es que la IA puede actuar como una herramienta de apoyo que reduzca errores y aumente la precisión de las evaluaciones e intervenciones. Un sistema que recuerde a las y los profesionales información relevante sobre algún cliente, usuario o paciente, o sobre algún proceso realizado, funciona como una segunda opinión o un recordatorio constante de información crucial, cosa que humanamente es difícil mantener en la mente cuando se trabaja con muchas personas a la vez. Esto puede llevar a intervenciones más acertadas, mejor ajuste de técnicas y monitoreo continuo de eficacia. Así, apoyando el juicio profesional con datos objetivos y alertas, la IA puede elevar la calidad ética de la atención, aumentando la beneficencia.

Finalmente, la introducción de IA al campo de la psicología ha obligado a la profesión a revisar y reforzar sus marcos

éticos. Por ello, organizaciones como la APA han desarrollado directrices específicas en 2025 y están en proceso de integrarlo a la nueva versión de sus principios éticos y código de conducta, lo cual mantiene vivo el diálogo ético y la conciencia de las y los profesionales sobre su responsabilidad al usar nuevas herramientas. Este ejercicio es positivo, pues al articular lineamientos sobre transparencia, sesgo, privacidad, etc., la psicología como campo se adapta proactivamente. Además, con la IA muchos psicólogos han debido aprender sobre ciberseguridad, sobre derechos digitales, sobre sesgos algorítmicos, etc., ampliando su formación ética más allá de lo tradicional. Esto redundará en personas profesionales más preparadas para los retos del siglo XXI (APA, 2025). Se puede concluir esta sección mencionando que los efectos de la IA en la ética profesional de la psicología no son unívocos: presentan riesgos serios, pero también oportunidades de mejora si se encarar responsablemente. La clave estará en maximizar los beneficios éticos (más acceso, más precisión, más eficiencia, más personalización) minimizando a la vez los perjuicios (sesgos, violaciones de privacidad, deshumanización, etc.). Para ello, se requieren estrategias y lineamientos claros, tema que se aborda en la siguiente sección.

3. Recomendaciones para una integración ética de la IA en la práctica psicológica

Dada la complejidad de las implicaciones analizadas, es necesario establecer estrategias que orienten a las y los psicólogos en una integración responsable de la IA. Como ya se ha revisado, diversos organismos ya han emitido lineamientos, como la APA con su Ethical Guidance for AI in the Professional Practice of Health Service Psychology y la UNESCO con sus marcos éticos sobre IA. A partir de estas fuentes, entre otras, se presentan las siguientes recomendaciones fundamentales (APA, 2025 & UNESCO, 2024).

Es necesario asegurar la transparencia y el consentimiento informado. Todo uso de IA en la práctica psicológica debe ser comunicado desde el inicio. La o el profesional debe informar al paciente si empleará herramientas de IA para almacenar datos, analizarlos o proponer intervenciones, explicando funciones, limitaciones, riesgos, implicaciones para la privacidad y beneficios esperados. Solo así el consentimiento informado es válido. También debe ofrecerse la opción de no utilizar IA, sin consecuencias discriminatorias. La UNESCO insiste en que las y los usuarios deben saber cuándo interactúan con IA y poder decidir al respecto. Mantener esta claridad fortalece la autonomía, evita la sensación de una opacidad en la relación terapéutica y exige dejar constancia documental de las partes generadas o asistidas por IA (APA, 2025 & UNESCO, 2024).

También se debe priorizar la ética del diseño y elegir herramientas confiables. Antes de integrar una IA, la o el psicólogo debe evaluar críticamente su validez científica, origen, datos de entrenamiento, sesgos, manejo de privacidad y cumplimiento regulatorio, especialmente en

contextos de alto riesgo como salud mental. La APA recomienda seleccionar herramientas con estrategias de minimización de sesgos y validadas en poblaciones similares[86], además de preferir sistemas con cifrado, anonimización y sin reutilización de datos sin permiso. En coherencia con la UNESCO, se requieren estándares éticos y sellos de calidad para IA en salud, como en la educación. Mientras estos marcos se consolidan, corresponde a cada psicólogo y psicóloga comprender los fundamentos de las herramientas y emplearlas solo dentro de su competencia profesional. Asimismo, es necesario exigir transparencia y responsabilidad a las empresas desarrolladoras, incluyendo soportes para reportar errores y documentación clara sobre el entrenamiento de los modelos (APA, 2025 & UNESCO, 2024).

Se requiere mantener la primacía del juicio humano y la supervisión constante. Ninguna IA debe actuar de forma autónoma en decisiones que afecten significativamente a las personas, pues el juicio profesional humano tiene la última palabra. Las y los psicólogos debe revisar cualquier resultado generado por IA antes de integrarlo en el proceso evaluativo, tomando sugerencias diagnósticas como hipótesis y verificando resúmenes o recomendaciones antes de utilizarlos. La supervisión humana también implica reconocer señales de dependencia excesiva de la IA y recuperar el criterio profesional cuando sea necesario. Deben existir planes de contingencia para continuar la atención en caso de fallos del sistema. En la formación, los supervisores deben guiar el uso crítico de IA, garantizando que los profesionales jóvenes no sustituyan el aprendizaje de bases profesionales por una confianza excesiva en estas herramientas, y documentar adecuadamente las decisiones asistidas por IA para fines éticos y legales[80] (APA, 2025). Se debe proteger la confidencialidad y seguridad de datos rigurosamente. Las prácticas éticas exigen el uso de plataformas con cifrado de extremo a extremo, almacenamiento seguro y la prohibición de ingresar datos identificables en herramientas públicas sin garantías de privacidad. Debe anonimizarse la información antes de su análisis y actualizar los consentimientos explicando ubicación de los datos, tiempos de resguardo y posibles accesos. Las y los profesionales deben cumplir con normativas aplicables, verificar requisitos para transferencias internacionales de datos y permitir que el cliente decida qué información compartir, con posibilidad de revocar su autorización. Asimismo, requieren formación básica en ciberseguridad, pues la protección tecnológica es insuficiente si la o el profesional tiene prácticas inseguras (APA, 2025).

Es necesario fortalecer la competencia digital y la formación continua. La rápida evolución de la IA obliga a las y los psicólogos a mantenerse actualizados en desarrollos técnicos, evidencias, regulaciones y buenas prácticas. La APA subraya la necesidad de actualización constante. La alfabetización digital debe incluir comprensión básica del funcionamiento algorítmico, puntos ciegos, IA explicable y principios éticos. Los programas universitarios y la capacitación profesional deben integrar estos contenidos, promoviendo habilidades interdisciplinarias como

programación elemental o análisis de datos para colaborar de manera efectiva con expertos técnicos (APA,2025).

Hay que delimitar los roles: la IA como complemento, no sustituto. Es necesario definir qué funciones pueden delegarse a IA y cuáles permanecen exclusivamente en manos del psicólogo. En práctica clínica o educativa, la IA puede apoyar en tareas administrativas o de monitoreo, pero las decisiones críticas, como diagnósticos, intervenciones en crisis o devoluciones de resultados, deben ser responsabilidad humana. La UNESCO propone proteger la intervención humana en procesos clave. También es importante gestionar expectativas ante el público, evitando presentar estas herramientas como sustitutos terapéuticos y manteniendo una comunicación clara y honesta sobre sus alcances reales (UNESCO, 2024).

Se requiere participar activamente en la regulación y desarrollo de IA. La ética profesional implica una responsabilidad social: las y los psicólogos deben involucrarse en debates regulatorios, colaborar con comités, producir posicionamientos y participar en grupos interdisciplinarios que discutan el impacto de la IA. Su conocimiento sobre comportamiento humano es esencial para anticipar riesgos psicosociales y orientar el diseño de tecnologías más justas. A nivel internacional, apoyar iniciativas como la Recomendación Ética de la UNESCO (2021) y las guías de IA generativa (2024) contribuye a construir marcos de protección global (UNESCO, 2021 & UNESCO,2024).

Finalmente, hay que evaluar y monitorear continuamente los impactos de integrar IA, lo cual exige revisiones periódicas de sus efectos en usuarios, clientes o pacientes. Es necesario recoger retroalimentación, detectar problemas de uso, brechas de seguridad o falta de beneficio, y ajustar o retirar la herramienta según corresponda. Esta dinámica de mejora continua permite responder a riesgos emergentes y adaptarse a actualizaciones o nuevos hallazgos científicos.

Todas estas recomendaciones buscan mantener a las y los psicólogos en el centro y utilizar la IA como herramienta de apoyo crítico. Como resume la APA, en cada dimensión (transparencia, equidad, privacidad, exactitud, supervisión y competencia) la IA debe siempre apoyar a las y los profesionales, y no reemplazar su juicio [89]. Esta orientación permite aprovechar las innovaciones tecnológicas sin perder la ética humanista que sustenta a la psicología como profesión de ayuda (APA,2025).

Conclusiones

La revolución de la inteligencia artificial y particularmente de la IA generativa ha irrumpido con fuerza en la vida cotidiana y en el ejercicio de la psicología, ofreciendo herramientas potentes para enriquecer la evaluación, la intervención, la investigación y la educación en este campo. A lo largo de este artículo se han revisado numerosas aplicaciones de la IA en la acción profesional de las y los psicólogos: desde algoritmos que detectan rasgos de personalidad en textos o señales de riesgo

suicida en la voz, pasando por asistentes virtuales que ayudan a terapeutas a redactar notas y planificar tratamientos (como Upheal), hasta chatbots que brindan apoyo emocional basado en TCC a millones de personas (Woebot, Wysa, etc.), sin olvidar los sistemas de análisis de datos que potencian la investigación psicológica. Estos ejemplos reales muestran un panorama de innovación acelerada, donde la IA se perfila como un aliado transformador de la práctica psicológica.

Sin embargo, también se han identificado profundas implicaciones éticas. El uso de IA en psicología entraña riesgos de violar la privacidad y confidencialidad de datos sensibles, de introducir sesgos y discriminación en decisiones que deben ser justas, de opacar la transparencia requerida para el consentimiento informado, e incluso de comprometer la calidad de la atención mediante errores automatizados o la erosión del contacto humano empático. Asimismo, surge la necesidad de redefinir la responsabilidad profesional en un contexto donde las decisiones son asistidas por herramientas tecnológicas, y de asegurar que las y los psicólogos mantengan sus competencias y criterio ante la tentación de delegar demasiado a la IA. No obstante, también se han reconocido potenciales beneficios éticos si la IA se integra correctamente: podría ampliar el acceso a servicios psicológicos a poblaciones desatendidas, personalizar las intervenciones respetando la individualidad de cada usuario, eliminar algunos sesgos humanos e incrementar la eficiencia, permitiendo a los profesionales dedicar más tiempo a lo esencialmente humano de su labor. Para maximizar esos beneficios y mitigar los riesgos, se han propuesto recomendaciones críticas hacia una integración ética de la IA en psicología, apoyadas en directrices de organismos como la APA (2025) y la UNESCO (2024). Entre ellas destacan: garantizar siempre la transparencia y el consentimiento informado al usar IA; escoger herramientas cuyo diseño priorice la equidad, la privacidad y la seguridad; mantener al psicólogo en control, supervisando y validando las acciones de la IA; proteger con rigor la confidencialidad mediante medidas técnicas y legales; actualizar la capacitación profesional en competencias digitales y éticas; y trazar límites claros donde la IA sea complemento y no reemplazo del juicio clínico ni de la relación humana. Adicionalmente, se recomienda que la comunidad psicológica participe activamente en la elaboración de marcos regulatorios y en el diálogo multidisciplinario, para asegurar que las normas y desarrollos tecnológicos futuros incorporen los valores humanistas fundamentales de la disciplina psicológica.

En conclusión, la IA no es ni panacea milagrosa ni amenaza distópica inevitable para la psicología, sino una herramienta poderosa cuyo impacto dependerá de cómo se use. Como toda herramienta, puede hacer el bien o el mal: puede ayudar a las y los psicólogos a llegar donde antes no podían o puede causar daño si se usa sin cuidado. Por ello, los retos de su integración ética son, ante todo, un llamado a que la profesión psicológica reafirme sus valores en un contexto nuevo: poniendo la tecnología al servicio de las personas, y no viceversa. Con guía ética, perspectiva crítica y colaboración entre todos los actores, es posible aprovechar

las ventajas de la IA para potenciar la acción profesional de psicólogos y psicólogas, mejorando la calidad y alcance de sus servicios, a la vez que se salvaguardan los derechos de aquellas y aquellos a quienes se les brindan servicios psicológicos.

Referencias

- American Psychological Association. (2025). Ethical guidance for AI in the professional practice of health service psychology.
- Beauchamp, T. L., & Childress, J. F. (2025). Principios de ética biomédica.
- Carbonell Bernal, N., & Hernández Prados, M. (2024). Impacto de los sistemas de tutoría inteligente: Una revisión sistemática. *EduTec. Revista Electrónica de Tecnología Educativa*, 89, 121–143.
- Chen, Z., Kulkarni, P., Galatzer-Levy, I., Bigio, B., Nasca, C., & Zhang, Y. (2022). Modern views of machine learning for precision psychiatry. *Patterns*, 3(11), Article 100602.
- De la Fuente Tambo, D., & Armayones Ruiz, M. (2025). La IA en la práctica psicológica: ¿Qué existe y cómo puede ayudar en psicología asistencial? *Papeles del Psicólogo*, 46(1), 18–24.
- Eliot, L. (2023). La legendaria ELIZA y PARRY se enfrentan cara a cara con ChatGPT en una reveladora batalla sobre el uso de IA generativa para la salud mental. *Forbes*.
- Forbes Argentina. (2025). Así es como la inteligencia artificial multiagente podría revolucionar la terapia en salud mental. *Forbes Argentina*.
- Infobae. (2023). ¿La inteligencia artificial puede ayudar a mejorar la salud mental? *Infobae*.
- Krieger, V., Magallón-Neri, E., & Amador-Campos, J. (2024). Ética e inteligencia artificial en evaluación psicológica. *Facultad de Psicología, Universidad de Barcelona*.
- López Fernández, M. (2025). Ética y regulación de la inteligencia artificial: El debate más urgente de 2025.
- López Steinmetz, L., & Carlos Godoy, J. (2023). Posibles aplicaciones prácticas del uso de machine learning en la investigación y práctica de la clínica psicológica. *Acta Psiquiátrica y Psicológica de América Latina*, 69(4), 266–272.
- Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D., & Haber, N. (2025). Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (pp. 599–627).
- Observatorio Nacional de Tecnología y Sociedad. (2025). Indicadores de uso de inteligencia artificial en España 2024.
- Piedras, E. (2025). #IA en México: Conocimiento, uso y barreras. *El Economista*.
- Saeteros, D., Gallardo-Pujol, D., & Ortiz-Martínez, D. (2025). Text speaks louder: Insights into personality from natural language processing. *PLoS One*, 20(6), 4–52.
- Tenelema Jiménez, I., Cabrera Moreira, J., Vinza Coronel, C., & Plua Panta, K. (2025). Aplicación de la inteligencia artificial en los procesos de reclutamiento y selección de personal. *Ciencia Digital*, 9(3.1), 103–126.
- UNESCO. (2021). Recomendación sobre la ética de la inteligencia artificial.
- UNESCO. (2024). Guía para el uso de IA generativa en educación e investigación.
- Upheal. (2025). Upheal. <https://www.upheal.io/2025>.